Phenotypic Pliancy and the Breakdown of Epigenetic Polycomb Mechanisms

Maryl Lambros[1], Yehonatan Sella[1], Aviv Bergman[1,2,3,4,*],

[1]Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA
[2]Dominick P. Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA
[3]Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461, USA
[4]Santa Fe Institute, Santa Fe, NM 87501, USA

[*]To whom correspondence should be addressed. E-mail: aviv@einsteinmed.edu

Abstract

Epigenetic regulatory mechanisms allow multicellular organisms to develop distinct specialized cell identities despite having the same total genome. Cell-fate choices are based on gene expression programs and environmental cues that cells experience during embryonic development, and are usually maintained throughout the life of the organism despite new environmental cues. The evolutionarily conserved Polycomb group (PcG) proteins form Polycomb Repressive Complexes (PRCs) that help orchestrate these developmental choices. Post-development, these complexes actively maintain the resulting cell fate, even in the face of environmental perturbations. Given the crucial role of these polycomb mechanisms in providing phenotypic fidelity (i.e. maintenance of cell fate), we hypothesize that their dysregulation after development will lead to decreased phenotypic fidelity allowing dysregulated cells to sustainably switch their phenotype in response to environmental changes. We call this abnormal phenotypic switching *phenotypic pliancy*. We introduce a computational evolutionary model that allows us to test our systems-level phenotypic pliancy hypothesis *in-silico* and in a context-independent manner. We find that 1) phenotypic fidelity is an emergent systems-level property of PcG-like mechanisms, and 2) phenotypic pliancy is an emergent systems-level property resulting from this mechanism's dysregulation. Since there is evidence that metastatic cells behave in a phenotypically pliant manner and PcG dysregulation is common in cancer, we hypothesize that progression to metastasis is driven by the emergence of phenotypic pliancy in cancer cells as a result of PcG mechanism dysregulation. We corroborate our hypothesis and the results of our computational model using single-cell RNA-sequencing data from metastatic cancers. We find that metastatic cancer cells are phenotypically pliant in the same manner as predicted by our model.

Significance Statement

We introduce the concept of cellular phenotypic pliancy– sustained abnormal phenotypic switching in response to environmental changes– and demonstrate that such behavior can be caused by dysregulation of Polycomb mechanisms. To overcome the incomplete knowledge about these mechanisms in higher organisms, we develop an abstract computational model to study the emergence of phenotypic pliancy from a general systems-level view by confirming our hypothesis over a wide range of simulated

gene-regulatory networks and Polycomb mechanisms' patterns. We corroborate our hypothesis and model predictions using single-cell RNA-seq metastatic cancer datasets. Our hypothesis has the potential to shed light on a general phenomenon for complex diseases where abnormal phenotypic switching is relevant.

Introduction

During early development in multicellular organisms, the differentiated phenotype of cells, i.e. cell identity, is sensitive to and determined by the environmental influences [1, 2]. However, after development, cells tend to maintain their identity, becoming markedly less sensitive to environmental changes even as they may exhibit some limited phenotypic plasticity [3, 4]. We will refer to the maintenance of cellular identity post-development as *phenotypic fidelity*.

Because all cells in a multicellular organism possess the same genome, a key regulator of cell-fate choices and cell identity maintenance is epigenetic regulation mechanisms, which emerged during the expansion of the Metazoa [5–7]. One of the most prominent and enigmatic epigenetic regulatory mechanisms is the evolutionarily conserved polycomb mechanisms involving the Polycomb group (PcG) and Trithorax group (TrxG) proteins that form complexes (called Polycomb Repressive Complexes (PRCs) for PcG proteins) thought to have evolved at the dawn of multicellularity [8, 9]. PcG and TrxG proteins act antagonistically to cause heritable alterations in gene expression or function, without changes in DNA sequence. PcG proteins act by continually repressing a set of target genes if their expression, as influenced by the environment and maternal gene expression levels, falls below a certain threshold at a critical time point during development [3, 10–14]. Conversely, TrxG proteins act to counteract PcG-mediated repression and to continually maintain the memory of the transcriptional states of the target genes [14, 15]. The polycomb mechanism promotes cellular differentiation during development in response to environmental inputs, and *actively maintains* cell identity post-development even in the face of environmental perturbations (see Figure 1) [3, 9, 13–15]. It has also been demonstrated computationally that a PcG-like mechanism allows dynamic altering of a gene regulatory network to create multiple gene regulatory sub-networks and genotype-phenotype mappings out of a single network (see Figure 1) [16].

Given the crucial role of polycomb mechanisms in providing phenotypic fidelity, we hypothesize that dysregulation of polycomb mechanisms after development will lead to decreased phenotypic fidelity, which will allow dysregulated cells to directly and sustainably switch their phenotype in response to environmental changes. We call this abnormal phenotypic switching in response to environmental changes *phenotypic pliancy*. Here, the environment encompasses both the external environment, such as availability of oxygen, temperature, and perhaps a local microbiome, and the signaling environment created by cellular communication networks. We further hypothesize that cells with dysregulated polycomb mechanisms, when introduced into a new environment will switch into a phenotype that more closely, *but not fully*, resembles that of the already evolved phenotype in this environment since we expect that the environmental cues will dominate once the cellular memory disappears. For example, a dysregulated breast cell placed in a lung environment would alter its phenotype to more resemble that of a lung cell than a breast cell if the PRC(s) dysregulated in the breast cell derepressed the target genes important for lung cell determination.

While much knowledge has been gained about the mechanisms and functions of the various PRCs and their PcG protein components, much still remains unknown owing to the complexity of the polycomb system. For humans and other complex multicellular organisms, the set of all possible combinations of PcG proteins that can form different

PRCs and how they function is still largely unknown (there are estimated to be more than 100 different PRC variants) [3, 14, 17, 18]. Research is ongoing for ways of determining PRCs' target gene sets, which are thought to be different but possibly overlapping between the various PRCs [18, 19]. Additionally, despite much research, there is no determined DNA sequence that distinguishes a target gene (called Polycomb Repressive Elements (PREs)) like in *Drosophila Melanogaster* [20–22] and no way to reliably classify PRC target genes in humans [18, 20, 23].

Given the lack of target gene knowledge and the immense number of possible combinations of PcG proteins combining to form different PRCs with potentially differing functions and target gene sets in different contexts (developmental stages, cell types, and environments), it is difficult to study from a systems-level a breakdown of the polycomb mechanisms and its consequences experimentally in humans. As an alternative, we introduce a new conceptual computational model that allows us to test our hypothesis *in-silico* with full control over the model's parameters. By using a computational model, we can investigate the *general* functions of polycomb mechanisms, and emergent properties upon their evolution and breakdown that are independent of specific contexts.

One possible occurrence of phenotypic pliancy is in metastatic cancer. In contrast to the phenotypic fidelity exhibited by normally functioning cells, metastatic cancer cells seemingly exhibit phenotypic pliancy [24–29]. The wide variety of environments encountered is highlighted throughout the course of the disease as the metastatic cells progress through invasion, intravasation, survival in the circulatory system, extravasation, arrival at the metastatic site, and, finally, colonization of the metastatic site [30–32]. Each of these occur in a different environment and seemingly requires the metastatic cells to assume a different phenotype in order to adapt to and survive in that environment [24–29, 31], while non-metastatic cancer cells do not switch their phenotype in response to environmental change [31]. Additionally, it has been observed that sustained abnormal phenotypic switching is not limited to cancer stem cells but is the property of a wider cell population [24, 28, 29]. Attaining a deeper understanding of the underlying mechanism of such abnormal phenotypic pliancy is therefore paramount.

As a special case of our general hypothesis, we hypothesize that progression to metastasis is driven by the emergence of phenotypic pliancy in cancer cells as a result of polycomb mechanism dysregulation.

Indeed, current literature demonstrates that dysregulation of polycomb mechanisms plays an important role in cancer and metastasis [3, 9, 14, 15, 17, 18, 33], but how and why is still contradictory. A potential explanation of the discrepancies in polycomb mechanism involvement in primary and metastatic cancers, is that PcG proteins, when not involved in a PRC, have different functions when acting in isolation versus as part of a PRC [17]. For example, EZH2 is a PcG protein component of the Polycomb Repressive Complex 2 (PRC2) that on its own is thought to be involved in cell cycle progression [4]. So EZH2 is often up regulated in primary tumor cancers, which may help cancer cells proliferate [34–36]. However, the other PRC2 PcG protein components are not up regulated, hinting that EZH2 up regulation may not be affecting the PRC2 function. When EZH2 is found to be down regulated, it is typically in conjunction with the other PRC2 components and in metastatic cancers [4].

Previous hypotheses of disease progression behaviors that could possibly explain metastatic cancer cells' phenotypic pliancy are de- and re-differentiation [37–39]; cells' accumulation of appropriate mutations while dormant [40, 41]; or exosome-mediated metastasis [42, 43]. However, there are important commonalities captured by our hypothesis that are involved in these three seemingly different behaviors: environmental influences, epigenetic modifications, and abnormal phenotypic switching. Our *systems-level* hypothesis does not contradict these more circumscribed hypotheses, but

rather may provide a more parsimonious and unifying conceptual mechanistic framework that may shed light on the mechanisms underlying all the above previous hypotheses.

In-silico experiments using our computational model demonstrate, in a general and context-independent way, that the breakage of polycomb mechanisms leads to the emergence of phenotypic pliancy. As hypothesized, the observed pliancy is sustainable, with phenotypic movement towards the evolved normal phenotype in the surrounding environment, and the degree of phenotypic movement is correlated with the extent of polycomb dysregulation. These in-silico results are corroborated by our analysis of single-cell RNA-sequencing data of solid tumor metastatic cancers since phenotypic pliancy is at the cellular level, which shows that metastatic cancer cells display phenotypic pliancy behavior correlated with the breakage of polycomb mechanisms.

Results

Model Description

We develop a holistic computational model of the developmental process including the action of PcG-like mechanisms, as well as post-developmental gene-regulatory dynamics, and the evolution of these dynamics for populations of individuals. Our model is built upon a well-established computational gene regulatory network model. The original model was used to study the evolution of canalization [44]. Later, it was used to study the role of a PRC in decoupling genetic and environmental robustness during development [16]. Briefly (see Materials and Methods for details), the main features of our current, extended computational model are:

1) Explicit environment-gene interactions in addition to gene-gene interactions. We model a set of abstract "environmental factors" as well as a set of abstract genes, with an environmental state vector in addition to a gene-expression state vector, and a matrix that determines the effect of each environmental factor on each gene as well as a matrix that determines gene-gene interaction.

2) Multicellular evolution and development. Since PcG-like mechanisms are integral to the evolution of multicellularity, we explicitly evolve individuals composed of two cells with the same genotype but in two different environments simultaneously. We choose a fitness function which simultaneously selects for a distinct optimal phenotype in each environment. This allows the evolving PcG-like mechanisms to facilitate differentiated genotype-phenotype mappings.

3) Evolution of two PRCs. We model dysregulation as complete loss of function of one or both PRCs. We also model the two PRCs such that each complex can affect different, possibly overlapping, sets of genes without needing the biological specifics for how these target genes are recognized and repressed. Note, TrxG complexes are implicitly modeled such that the targets that are not repressed by PRCs are available to be transcribed.

In total, these features allow us to carry out multicellular evolution experiments with polycomb-like epigenetic regulation to compare the phenotype of a cell, when post-developmentally transferred to a new environment, to that cell's evolved phenotype in both the old and new environments.

Emergence of Phenotypic Pliancy during Evolution: Model Results

We investigate the role of PcG-like mechanisms in the evolution towards phenotypic fidelity to differentiated phenotypes, and the general consequences of dysregulating these evolved PcG-like mechanisms post-developmentally. We hypothesize that

post-developmental dysregulation of PcG-like mechanisms allows cells to become phenotypically pliant. We use our computational model described above to generate artificial "single-cell gene expression" data. With this generated data, we visually and quantitatively demonstrate that intact PcG-like mechanisms ensure phenotypic fidelity behavior, while PcG-like mechanism dysregulation causes phenotypic pliancy to emerge.

Using our computational model, we run evolution on 1,000 different initial populations, each evolving under 10 different evolutionary trajectories, for a total of 10,000 evolved populations, each with 1,000 individuals.

Throughout the course of evolution, the total population fitness results indeed tend to increase as expected (see Supplementary Figure S1). Increasing fitness means that the populations successfully evolve toward the selected-for differentiated phenotypes. Note, the fitness in environment 1 is expected to decrease since all individuals start at the optimum in environment 1 but not in environment 2, so it must evolve towards environment 2, i.e. demonstrating the evolution towards the capacity for differentiation (see Materials and Methods for details).

Moreover, we see that during the course of evolution, PcG-like mechanisms take an increasingly active role, with an increasing number of genes under PcG-like mechanism control (see Supplementary Figure S2), and correspondingly a decreasing effective connectivity of the gene-regulatory network (see Supplementary Figure S3). The effective connectivity is the result of removing the genes which are repressed by PRCs during development in each environment and their connections in the network, similar to the representation in Figure 1. Interestingly, our model predicts that there is a preference for a higher incoming connectivity among PcG-like mechanism target genes that are repressed by any PRC, relative to non-repressed target and non-target genes (see Supplementary Figure S3).

These trends suggest that PcG-like mechanisms play an important role in facilitating the evolved differentiation. However, the mutation-selection balance must be considered before reaching such a conclusion more definitively. Moreover, it is possible that PcG-like mechanisms play a secondary role in evolution, by facilitating genetic robustness, an inquiry outside the scope of this paper.

Next, we observe that the evolution of populations with PcG promotes phenotypic fidelity as an emergent property, without being directly selected for. To see this, we apply principal component analysis to the gene expression vectors of cells in a representative population. We carry out this analysis at two evolutionary time points, after 50 and 1,000 generations, respectively. After 1,000 generations (see Figure 3A), we observe phenotypic fidelity. Indeed, when PcG-like mechanisms are left intact post-developmentally and the cells which developed in environment 1 are transferred to environment 2, the resulting stable gene expression patterns (cyan) largely remain close to the original gene expression patterns of cells developed in environment 1 (green) while largely staying far from the gene expression patterns of cells developed in environment 2 (blue X's).

In contrast, the population at an early stage of evolution (see Figure 3B) does not exhibit phenotypic fidelity. When the same experiment is carried out, the cells which developed in environment 1 and transferred to environment 2 with PcG-like mechanisms intact adopt gene expression patterns much closer to those of cells developed in environment 2.

This result confirms biological understanding of the role of an evolved polycomb mechanism in maintaining cell identity, and shows that our computational model, while general and abstract, is sufficiently true to biological reality to manifest this emergent property.

Finally, we observe that, when evolved populations have PcG-like mechanisms broken post-development, they exhibit phenotypic pliancy. This can be seen both

visually in a PCA, as well as numerically (see Figure 3A and 3C). In the PCA, when PcG-like mechanisms are broken post-developmentally and the population is then transferred from environment 1 to environment 2, the resulting stable gene expression patterns (red) move closer to the normal phenotype in environment 2 (blue X's), and away from their original phenotype in environment 1 (green). This observation illustrates that, when a PcG-like mechanism is broken and the population is transferred to a different environment, cells will switch their phenotypes to phenotypes that more closely resemble the normal phenotypes that evolved in the environment transferred to.

For numerical evidence, we devise a measure of pliancy that measures, for each individuals, the proportion of genes which, when PcG-like mechanisms are broken post-development and upon transferring to a different environment, move away from the individual's phenotype in the original environment and toward its phenotype in the new environment. We see that this phenotypic pliancy score tends to increase substantially over the course of evolution, indicating that phenotypic pliancy upon broken PcG-like mechanisms is an emergent property of evolution (see Figure 3C). Specifically, we compute a phenotypic pliancy score upon breaking both PcG-like mechanisms, and when breaking each PRC separately (see Materials and Methods). We find that the level of phenotypic pliancy corresponds to the degree of polycomb mechanisms dysregulation, as the average phenotypic pliancy of all 10 million simulated individuals is much greater when both PRC1 and PRC2 are dysregulated simultaneously than when PRC1 or PRC2 is dysregulated separately (see Figure 3C). The dysregulation of these PRCs does not affect the stability of the simulated cells, where around 0 to 0.1% of cells are unstable upon dysregulation of the PRCs separately or together (see Supplementary Figure S4).

Importantly, we find that phenotypic pliancy is sustained, meaning that cells can keep switching their phenotype when placed in a different environment and not get stuck in a certain phenotype. This sustainability can be visualized using PCA for the case when switch back to environment 1 after switching to environment 2 (see Supplementary Figure S5), such that the phenotype switches back to resemble that of environment 1 (green circles) when moved back (purple X's). Lastly, as a control, we test the case when PcG-like mechanisms are broken but the cells are kept in the same environment in which they had developed. We find the gene expression patterns do move away from their normal expression, but not substantially (see tan circles in Supplementary Figure S5).

To verify our model results are robust, we test for phenotypic pliancy over a wide range of parameters, demonstrating that phenotypic pliancy is a general, systems-level phenomenon (see Supplementary Figure S6 and Supplementary Information).

In sum, our model results demonstrate that: 1) evolution of PcG-like mechanisms causes post-developmental phenotypic fidelity to evolve, 2) a breakdown of PcG-like mechanisms lead to the emergence of phenotypic pliancy, 3) the level of phenotypic pliancy corresponds to the degree breakdown of PcG-like mechanisms, 4) phenotypic pliancy is sustainable, and 5) not only do phenotypically pliant cells move away from their evolved phenotype in the environment they move from, but they also move closer to the normal evolved phenotype in their new environment.

Assessment of Phenotypic Pliancy in Metastatic Cancer

To assess our hypothesis and our model's predictions, we use publicly available single-cell RNA-sequencing data from matching primary tumor and metastatic cancer sites, and normal cells at each site. This data departs from our in-silico data in various ways. In this data we do not have access to identical cells placed in different environments. Instead, we analyze populations of individual cells and make use of the fact that metastatic cells have undergone a change of environment from the primary to the metastatic site. Moreover, unlike with our model data, where the only experimental

conditions involve polycomb dysregulation and a change in environment, here we are faced with additional mutations and abnormalities coming from cancer cells. To control for that, we compare metastatic cells both to the primary tumor cells and to the normal cells at both the primary and metastatic sites. Driven by our findings from the model in the above results section, we address the following three sub-hypotheses: 1) PcG genes are differentially expressed in cancer cells relative to normal cells in their environment; 2) metastatic cells exhibit elements of phenotypic pliancy, such that their phenotype (i.e. gene expression profile) moves away from that of the primary tumor, and closer to the normal cells in the metastatic site; and 3) PcG genes are implicated in the observed phenotypic pliancy of metastatic cells, such that there is a positive correlation between the extent of dysregulation of PcG genes and the degree of phenotypic movement in the direction of normal phenotype.

Here we emphasize that metastatic cells' phenotypic movement in the direction of normal does not imply taking on the identity of the surrounding normal cells.

We use two different single-cell RNA-sequencing data sets from patients with metastatic head and neck cancer (H&N) obtained by Puram et. al. [45] and patients with metastatic serous epithelial ovarian cancer (ovarian) obtained by Shih et. al [46] (see Materials and Methods). We utilize a list we curated with 69 genes that are involved in polycomb mechanisms including 22 PcG genes, 35 TrxG genes, and 12 genes that have been verified to be controlled by PRCs (see Supplementary Table 1) [14, 15, 19, 47]. Preprocessing and all further analysis of the Puram et. al. and Shin et. al. data sets is performed using the R software package Seurat (see Materials and Methods) [48–50].

Metastatic cells are enriched with dysregulated PcG mechanisms

We test sub-hypothesis 1 by performing differential expression analysis on the preprocessed data using the default method (Wilcoxon rank sum test) of the R Seurat package [48–50]. For the H&N data set, differential gene expression analysis is done separately for each of the two pairwise comparisons: metastatic cells versus non-cancer cells in the lymph nodes, and primary tumor cells versus non-cancer cells in the oral cavity. For the ovarian data set, differential gene expression analysis is done between primary tumor cells versus normal cells in the ovary, and metastatic cells versus these same normal cells in the ovary since normal samples were not taken in the omentum.

For both H&N and ovarian data, our results for primary versus normal cells (column 2, row 1 and 3 in Supplementary Table 2, respectively) and metastatic versus normal cells (column 2, row 2 and 4 in Supplementary Table 2) show no significant trend in the number of polycomb mechanism genes that are differentially expressed. Since down regulation is more indicative of a breakdown in function, we investigate the extent to which these differentially expressed genes are up regulated or down regulated. To do this, we calculate the log-fold change sum for primary and normal cells at the primary site (column 3, row 1 and 3 in Supplementary Table 2) and the sum for metastatic and normal cells at the metastatic and primary site, respectively (column 3, row 2 and 4 in Supplementary Table 2). To calculate the significance in the difference in the log-fold change sums between primary and metastatic for each H&N and ovarian, we test against the null hypothesis that polycomb dysregulation is the same in primary and metastatic, by using a permutation test randomly permuting, over 1000 trials, the identity of primary and metastatic for each PcG gene's expression data, and using the generated histogram to extract p-values.

For the H&N results, a decrease in log-fold change sum at the metastatic site (p-value < 0.15) gives partial evidence that metastatic cells' PcG mechanism is being down regulated to a greater extent than primary cells. For the ovarian results, despite a decrease in the number of differentially expressed PcG genes in metastatic cells compared to normal, we observe strong evidence that metastatic cells' PcG mechanism

is being highly down regulated while primary cells' mechanism is actually being up regulated (p-value < 0.05). Since down regulation is more indicative of breakdown in function, this supports our hypothesis of increased polycomb dysregulation in metastasis. Up regulation in primary cells may be explained by many findings that some polycomb mechanism genes have independent functions from their PRC(s) in many cancers, such as cell cycle progression roles, as discussed for EZH2 [4, 34–36].

Metastatic cells are phenotypically pliant

To test sub-hypothesis 2 and assess our model prediction that cells with broken polycomb mechanisms move toward the normal phenotype of the new environment in which they are placed (see Figure 3A and B), we further narrow the set of genes under consideration to only those which are differentially expressed between primary and metastatic cancer cells resulting in a total of 2,246 genes for the H&N and 3,393 genes for the ovarian data. This restriction is in order to not drown out the signal of the difference between these two cellular categories.

In order to visualize phenotypic pliancy, we use 2 different methods to allow us to capture critical features of the data to extract knowledge of phenotypic pliancy behavior without having dynamical data. First, we perform a principal component analysis (PCA) on the two data sets. The first two principal components are plotted in Supplementary Figure S7, showing a clear clustering of the data for both the H&N and ovarian data. Importantly, especially along the first principal component, metastatic cells tend to be closer to normal cells than primary tumor cells are, lending credence to our hypothesis of increased phenotypic pliancy in metastatic cells and prediction from our model. Moreover, Supplementary Figure S7 reveals that primary tumor cells can be divided into two sub-clusters based on the sign of PC2 (y-axis), where one sub-cluster substantially overlaps with metastatic cells (in cyan). Secondly, we apply Uniform Manifold Approximation and Projection (UMAP) [51] to the two data sets and see an even clearer progression from primary tumor cells to metastatic cells and then to normal cells (see Figure 4). In addition, these UMAP findings also reveal that a significant number of primary tumor cells (labeled cyan in Supplementary Figure S7) possess a gene-expression profile closer to that of metastatic, which suggests these primary cells may have transitioned to a more metastatic cell type.

We quantify if metastatic cells are phenotypically pliant using two methods. Firstly, we measure whether the distance of metastatic phenotype (gene expression profile) to the normal phenotype is less than the distance of primary phenotype to normal phenotype. We compute the distance between two populations of cells as the Euclidean distance between the centroids of their log-scale gene expression data. With this measure, the distance between primary tumor and normal cells is 55.1, which is larger than the distance between metastatic and lymph normal cells, which is 50.2 for the H&N data. For the ovarian data, the distance between primary tumor and normal cells is 34.2, which is larger than the distance between metastatic and normal ovary cells of 32.5. These distances are further increased when only considering the "primary2" cells that are clustered separately from the metastatic cells (green in Supplementary Figure 5), with "primary2" to normal distance of 57.4 compared to metastatic lymph normal distance 50.2 for H&N, and primary2 to normal distance of 39.8 compared to metastatic normal distance 32.5 for ovarian data.

Secondly, we calculate pliancy at the gene level, such that a gene is pliant if the distance of its expression in metastatic cells to normal cells is less than the distance of its expression in primary cells to normal cells. We calculate the gene-by-gene differences between primary and normal at the primary site (x-axis) versus the differences between metastatic and normal at the lymph node site or ovary site for H&N and ovarian data, respectively (y-axis) (see Supplementary Figure S8), and find, of the genes differentially

expressed between primary tumor versus metastatic cells, 68% are pliant (p-value $< 10^{-8}$) for H&N and 72% (p-value $< 10^{-9}$) for ovarian. All these measures lend strong evidence to sub-hypothesis 2 of increased phenotypic pliancy of metastatic cells for two different cancer types and our model predictions.

Degree of PcG mechanism dysregulation correlates with level phenotypic pliancy

To test our model prediction that the degree of phenotypic pliancy is correlated with the level of polycomb mechanisms disruption (see Figure 3C), we assign a measure of phenotypic pliancy to each metastatic cell, and test the correlation of this measure to the level of PcG mechanism dysregulation.

We measure a metastatic cell's phenotypic pliancy as the *pliancy z-score*, that measures the relative extent of movement of a particular metastatic cell's phenotype toward the normal phenotype and away from the mean of all metastatic cells (see Materials and Methods for details). We use the genes differentially expressed between metastatic and normal cells, excluding the 53 PcG and TrxG genes because we use these PcG mechanism genes as predictors in the regression analysis below (these 53 genes do not include the 12 known PcG target genes) for the H&N data and excluding the 56 PcG and TrxG genes for the ovarian data (these do not include the 9 known PcG target genes). We then fit a linear regression to predict the pliancy z-score as a function of all the PcG mechanism genes' expressions ($r^2 = 0.22$ for H&N and $r^2 = 0.56$ for ovarian data). For the H&N data, out of the 53 PcG and TrxG genes, 30 have positive coefficients in the linear model. Positive coefficients mean that decreased PcG and TrxG expression is positively correlated with metastatic cells' lower distance to normal cells, thus indicating that dysregulation of the PcG-like mechanism drives pliancy. For the ovarian data, out of the 56 PcG and TrxG genes, 21 have positive coefficients.

As a control for this result, we perform a bootstrap, randomly picking 1000 samples of 53 genes for H&N and 56 genes for ovarian and performing the same linear regression of the score as a function of these 53 and 56 genes' expression profiles for H&N and ovarian data, respectively. For each such regression, we record the number of genes with a positive coefficient and compute the distribution of this number (see Supplementary Figure S9). The mean is 22.3 out of 53 for H&N data, and mean is 10.6 out of 56 for the ovarian data. Relative to these distributions, 30 for H&N data and 21 for ovarian data is statistically significant (p-value < 0.05 and p-value $< 10^{-5}$, respectively) (see Supplementary Figure S9). So, in both the H&N and ovarian data there is a high level of dysregulation of PcG mechanism genes that is associated with metastatic cells moving closer toward the normal phenotype.

Discussion

In this paper, we propose and validate a framework of phenotypic pliancy, and its emergence as a result of PcG mechanism dysregulation.

We demonstrate using a computational model that evolution of PcG mechanisms results in phenotypic fidelity to differentiated phenotypes while dysregulation of PcG mechanisms leads to decreased phenotypic fidelity, which allows dysregulated cells to become phenotypically pliant. Additionally, we show that cells with dysregulated PcG mechanisms, when introduced into a new environment, switch into phenotypes that more closely resemble that of the already evolved phenotypes in this environment, which we expect since the environmental cues should dominate once the cellular memory disappears due to loss of polycomb mechanisms.

These results demonstrate that phenotypic fidelity and phenotypic pliancy are both general emergent properties over a wide range of simulated gene-regulatory networks and polycomb mechanism patterns, suggesting that these are general systems-level phenomena independent of specific contexts or selection. Similarly, previous research using abstract biologically inspired models demonstrated the emergence of systems-level properties that resulted in important biological insights [16, 44, 52–54].

We begin to test if the model conclusions hold true in biological contexts by utilizing publicly available single-cell RNA-sequencing data from metastatic cancer. We find preliminary evidence that 1) metastatic cells have enrichment of PcG mechanism genes differentially expressed when compared to normal, non-cancer cells at the respective site relative to primary tumor cells; 2) metastatic cells behave like phenotypically pliant cells, such that their phenotypes move away from that of the primary tumor, and closer to the normal, non-cancer cells in their surrounding site; and 3) PcG mechanism dysregulation is positively correlated with the degree of phenotypic movement in the direction of normal phenotype, i.e. level of phenotypic pliancy.

The metastatic cancer data findings corroborate the results from our computational model. While our metastatic cancer data results are not definitive since many PRCs, PcG mechanism genes, and PcG target genes are still unknown, our results do provide initial evidence of a general hypothesis that can be further tested and validated with experimental studies. If correct, these findings can explain a mechanism driving development and progression of metastatic cancer. Additionally, our computational model can be used in future studies to help guide the criteria for identifying polycomb mechanism target genes that contribute most to phenotypic pliancy. For example, our model predicts that these target genes will have higher incoming connections relative to the rest of the genes in the gene regulatory network under investigation.

Our findings related to metastatic cancer cells being phenotypically pliant do not contradict but rather may provide a more parsimonious and unifying conceptual mechanistic framework for metastatic disease emergence and progression than the existing more circumscribed hypothesizes, i.e. de- and re-differentiation; appropriate mutation accumulation during dormancy; and exosome-mediated niche-construction. Future work can investigate the interplay between polycomb mechanisms dysregulation and the above hypotheses in driving phenotypic pliancy and metastatic cancer.

There are additional unknowns pertaining to the primary tumor that need to be addressed in future work to understand the initial stages of metastatic cancer development. Further research is needed to assess whether polycomb mechanisms evolve to provide phenotypic fidelity to the generated tumor microenvironment of primary tumor cells. Additionally, our computational model does not address the interplay between polycomb mechanisms dysregulation and the accumulated mutations that lead to development of primary cancer. Thus, further detailed analysis is needed to investigate the implications of this interplay.

We demonstrate that the emergence of phenotypic pliancy is not contingent on selection pressures, contexts, or levels of investigation and therefore could hold true in diverse and broad scopes, like other complex diseases. Thus, our proposed mechanism of phenotypic pliancy has the capacity to be a general phenomenon not unique in cancer metastasis, and may shed light on complex diseases where sustained abnormal phenotypic switching have been observed [55]. Even though our results are reproducible in both our computational model and two different solid tumor metastatic cancers, further experimental work and testing is needed to assess phenotypic pliancy's broader implications.

Materials and methods

Computational Model Description

Our computational model operates on two levels: gene regulatory network (cell) dynamics and population dynamics (see Figure 2). At the detailed level of gene-regulatory network dynamics, a simulated cell's state at any given time is given by a gene expression state vector, $\mathrm{S_W}$, together with an environment state vector, $\mathrm{S_E}$. The gene expression vector has dimension N – the number of genes – and its coefficients represent gene expression levels. The environment vector's coefficients represent abstract "environmental factors" which can potentially affect gene expression. The dimensionality of these environmental factors is a parameter, which we choose to be equal to N. Next, the basic gene-regulatory dynamics are described by two matrices (see Figure 2). The first matrix, W_{GG}, is a $N \times N$ matrix encoding interactions among genes, where the ij entry of the W_{GG} matrix represents the effect of gene i on the product of gene j. The second matrix, denoted W_{GE}, similarly encodes the effect of the environmental factors on gene products in the cell. Please see Supplementary Table 3 for a list of model parameters and their values.

Finally, we incorporate the effects of PcG-like proteins in a developing individual into the model. Each simulated "cell" in this model is assigned multiple simulated PRC's. The information of which genes are targets of a given PRC is encoded in a PRE matrix (denoted θ), with entry $\theta_{\mathrm{i,j}}$ equal to 1 if gene i is under the control of PRC j, and equal to 0 otherwise. During the developmental step in our model, a gene possessing PREs ($\theta_{\mathrm{i,j}} = 1$ for some j) is permanently silenced by a PRC if that gene fails to be expressed above a threshold γ by a predetermined critical time point $\mathrm{t_c}$ during development.

Letting $\mathrm{S_G^t}$ denote the gene-expression state vector at time t and assuming a fixed environment state vector $\mathrm{S_E}$, the gene expression dynamics of each simulated cell are then given by $(\mathrm{S_G^{t+1}})_i = \sigma((W_{GG} \cdot S_G^t) + (W_{GE}) \cdot S_E)_i)$, where σ is the standard sigmoid, unless gene i is under the control of a PcG-like mechanism, $t > t_c$, and $(S_G^{t_c})_i < \gamma$, in which case $(S_G^{t+1})_i = 0$. As a short-hand, we denote this piecewise function by: $S_G^{t+1} = \sigma((W \cdot S^t)_G)[\theta]$.

Development is defined as the process of starting from some initial state and iterating the gene expression dynamics as described above until steady state is reached, denoted S_G. Steady state is reached when a stability measurement, a normalized variance of gene-expression pattern within the last 10 developmental time-steps, is smaller than the error term $\epsilon = 10^{-4}$. Cells that reach steady state are deemed developmentally stable [52, 56], otherwise they are considered lethal.

At the detailed level of population dynamics, a population of M individuals undergo iterations of mutation, reproduction, development, and selection, where each iteration represents a generation (see Figure 2). In total, the population dynamics step models the evolution of the population through 1,000 generations in the presence of stabilizing selection. A population is initially generated from one developmentally stable individual, known as a founder. The matrices W_{GG} and W_{GE}, together with the PRE matrix θ, constitute the individual's genotype. The W_{GG} and W_{GE} matrices of the founder are randomly generated matrices with a fixed ratio c (the matrix connectivity) of non-zero, Gaussian-distributed entries. The matrix θ is initially set to 0, with a fixed number of potential PRCs, so that in the beginning of evolution no gene is under PcG-like mechanism control. We add the ability to model the evolution of a population of multicellular individuals (see Figure 2), each individual's cells sharing the same genotype but developing in two different environments. We randomly generate 2 different environment binary state vectors which are pairwise different by a certain percentage difference, Δ_e, and such that the founder individual is stable in these environments. We set the optimal phenotype in environment 1 to be equal to the stable

state of the founder in environment 1, and we randomly pick the optimal phenotypic vector in environment 2, $S_{G_2}^{opt}$, such that it is different by a threshold, Δ_S, from environment 1. We pick initial state vectors S_{G_i} which are environmentally stable in their respective environments but may differ from $S_{G_i}^{opt}$, allowing for evolution toward the optimum. We measure developmentally stable individuals' fitness Ω with stabilizing and directional selection components: $\Omega = \frac{(e^{-(D/s)} + max(0, 1 - a \cdot D))}{2}$, where $D = (|S_{G_1} - S_{G_1}^{opt}| + |S_{G_2} - S_{G_2}^{opt}|)/(2N)$ measures distance from optimum, s is the stabilizing selection strength, and a is a parameter of the directional selection strength (see Supplementary Table 3).

This fitness function selects for individuals for which S_{G_i} moves closer to $S_{G_i}^{opt}$. During each generation, reproduction and selection are carried out by setting each simulated cell's probability of reproduction to be proportional to its fitness. Again, developmentally unstable individuals are considered lethal, thus not included into the next generation. The population of cells reproduces sexually.

Mutation is carried out by allowing each of the nonzero entries of the individual's gene interaction network, W_{GG}, and gene-environment interaction, W_{GE} to mutate according to a Gaussian distribution, with mutation rate μ. The entries of the PRE matrix, θ, are also free to switch between 0 and 1 with a set mutation rate, allowing evolution of susceptibility to PRCs. The mutation rate is set such that during each generation a fraction μ of a cell's N genes can be mutated to either become susceptible to PRCs ($\theta_i j$ goes from 0 to 1), or lose susceptibility to PRCs ($\theta_i j$ goes from 1 to 0). A detailed analysis of the previous model shows robust behavior to a wide range of the model's parameters [16], which we also observe in parameter testing for our current model (see Supplementary Figure S6 and Supplementary Information).

Phenotypic Pliancy Score

Using our computational model, we investigate the degree of phenotypic pliancy upon transferring a population of individuals from the first environmental condition to the second when the PcG-like mechanisms are intact versus broken. We simulate 1,000 different populations that each evolve under 10 different evolutionary trajectories, where each of these populations is comprised of 1,000 individuals to generate populations with varied intact PcG-like mechanisms.

At the end of and during the simulated evolution, we measure phenotypic pliancy for each of the 10-million individuals in each population as follows. We consider four environmental conditions: a) the cell is developed in environment 1 and has its polycomb mechanisms intact; b) the cell is developed in environment 2 and has its polycomb mechanisms intact; c) the cell is developed in environment 1, then post-developmentally transferred to environment 2 with intact polycomb-like mechanism(s); and d) the cell is developed in environment 1, then post-developmentally transferred to environment 2 with broken polycomb-like mechanism(s). We only consider those individuals for which the resulting gene expressions are all stable. Each experimental condition then gives rise to a different stable gene expression vector which we denote $S_G^a, S_G^b, S_G^c, S_G^d$, respectively.

We employ the following heuristic to determine when an individual is pliant at gene i– if it meets two criteria: $|(S_G^d)_i - (S_G^b)_i| < \frac{1}{2}|(S_G^a)_i - (S_G^b)_i|$ and $|(S_G^d)_i - (S_G^b)_i| < |(S_G^c)_i - (S_G^b)_i|$. We then calculate the frequency of genes at which that individual is pliant, forming an individual's pliancy score (see Figure 2).

Descriptions of Head and Neck Single-Cell RNA-Sequencing Data

The entire data set consists of expression data for single cells obtained from 18 patients (5 patients with matching primary and metastatic samples) in two locations: a primary tumor site (oral cavity) with 1,426 cancer cells and 2,817 non-cancer cells, and a metastatic site (lymph node) with 788 cancer cells and 546 non-cancer cells. The non-cancer (normal) cells include fibroblasts, endothelial cells, and B and T cells, amongst others; however, we only consider fibroblasts and endothelial cells. In our analysis, there are four cellular categories: metastatic in lymph node, normal in lymph node, primary tumor in oral cavity, and normal in oral cavity.

Descriptions of Ovarian Single-Cell RNA-Sequencing Data

The data set consists of expression data for single cells obtained from 9 patients (4 with matching primary and metastatic samples) in three locations and thus cellular categories: a primary tumor site (ovary or fallopian tube) with 1,649 cells, metastatic site (omentum) with 1,062 cells, and normal site (ovary) with 355 cells. The normal cells are comprised of fibroblasts, stromal cells, and mesothelial cells. When we perform PCA on all the cells, the normal cells do not cluster by cell type and the primary cells do not cluster by location.

Preprocessing of single-cell RNA-sequencing data

Preprocessing and all further analysis of the Puram et. al. and Shin et. al. data sets is performed using the R software package Seurat (version 3.2.3) [48–50]. First, we eliminate cells with fewer than 200 genes expressed in the cellular categories (primary, metastatic, and normal at either both primary and metastatic site or just primary site) and eliminate genes which have non-zero expression in only two cells or less. For H&N data set, 1,426 cancer and 2,817 non-cancer cells from the primary site, and 788 cancer and 546 non-cancer cells from the metastatic site remain. For the ovarian data set, 1,555 cancer and 345 non-cancer cells from the primary site and 1,028 cancer from the metastatic site remain. As for genes after preprocessing, we retain 21,294 genes in the H&N and 18,973 genes in the ovarian data set for further analysis. For H&N data, 65 of these genes are involved in a PcG-like mechanism including 22 PcG genes, 31 TrxG genes, and 12 genes that have been verified to be controlled by PRCs. For ovarian data, 65 of these genes are involved in a PcG-like mechanism including 21 PcG genes, 35 TrxG genes, and 9 PRC target genes. We then normalize by log transforming and centering the data and by using SCTransform in the Seurat package [48, 49], finding that the genes' estimated count-depth relationships were indeed zero for all the cellular categories in each data set. Finally, we regress out cell cycle genes to eliminate any variance due to difference in cell cycles.

Pliancy z-score for metastatic cells

To measure a metastatic cell's phenotypic pliancy, we develop a score, referred to as the *pliancy z-score*, that measures the relative extent of movement of a particular metastatic cell's phenotype toward the normal phenotype and away from the mean of all metastatic cells. This score is calculated as follows. For each gene g, let μ_g^n, σ_g^n and μ_g^m, σ_g^m denote the mean and standard deviations of the expression of gene g across normal (n) and metastatic (m) cells, respectively. For each metastatic cell c, let $x_{g,c}$ denote the expression of gene g at cell c and let $z_{g,c} = \frac{x_{g,c} - \mu_g^m}{\sigma_g^m}$ be its z-score relative to the distribution across all metastatic cells. Finally let $z'_{gc} = sign(\mu_g^m - \mu_g^n) \cdot z_{g,c}$. Thus,

a positive score means moving away from normal relative to the average, and a negative score means moving toward normal. The cell's overall score is then devined as $z_c = \frac{1}{N}\sum_g z'_{g,c}$, where the sum is taken over the N genes included in the analysis.

Data Availability Statement

The code for both our computational model data generation, model data analysis, and metastatic cancer data analysis can be accessed at:
https://github.com/AvivLab/Phenotypic-Pliancy.

Acknowledgments

The authors would like to acknowledge Eduardo Fajardo for help with writing perl scripts for pre-processing the sc-RNA-seq datasets. Carlos Madrid-Aliste for help with HPC. The authors also would like to thank Moshe Sadofsky, Matthew Scharff, Ruben Coen-Cagli, and Ian Willis for many helpful discussions. Support was provided by NIH CMBG training grant T32-GM007491 to ML, and NIH R01-CA164468 and R01-DA033788 to AB.

Figures

604

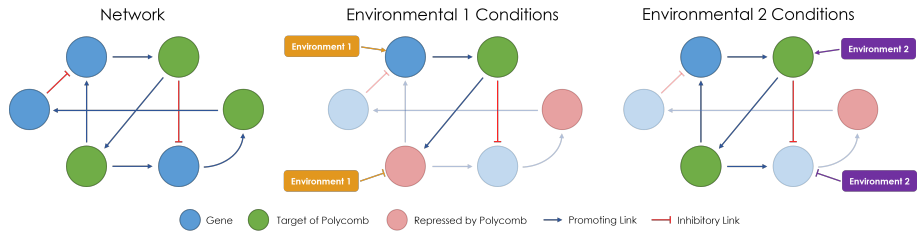

Fig 1. Schematic of Polycomb Mechanisms During Development: The evolution of Polycomb mechanisms carves the overall gene regulatory network into subnetworks based on the environments encountered during development, which leads to different differentiated phenotypes at the end of development.

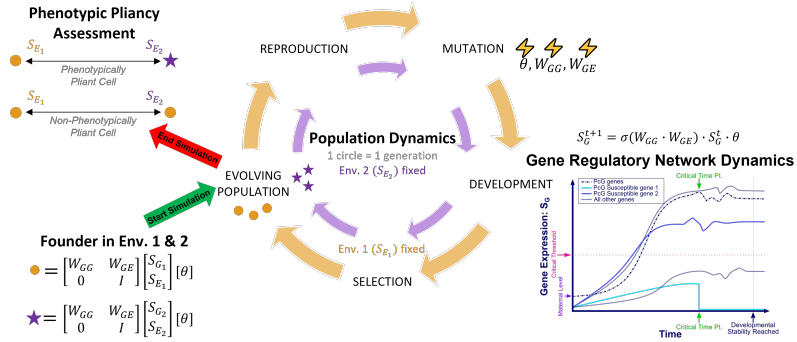

Fig 2. Schematic of our Computational Model: Schematic of both the population dynamics and gene regulatory network dynamics components of our model under environment 1, where each simulated cell's phenotype in the population is represented by a blue sphere. Not shown in this schematic but simultaneously happening in our model is the evolution under environment 2, where each simulated cell's phenotype is represented as a purple star. This figure also shows a schematic of phenotypic pliancy assessment when a simulated cell switches from environment 1 to environment 2 after evolution.

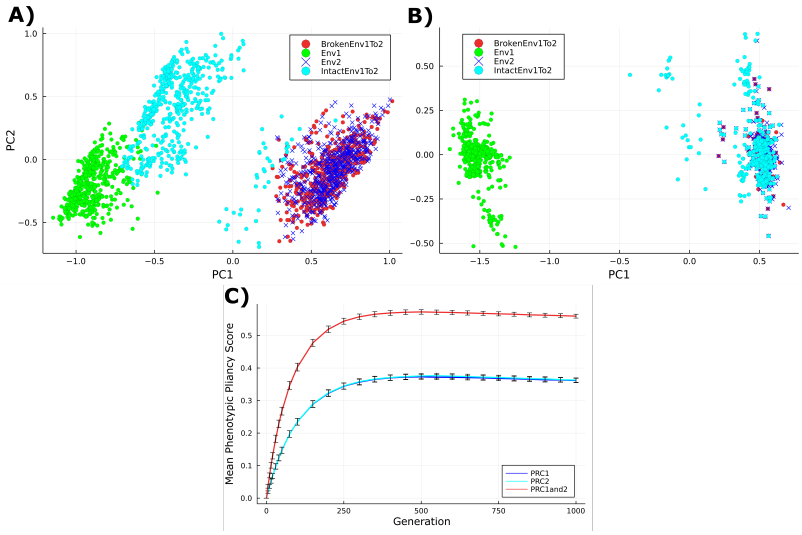

Fig 3. Phenotypic Pliancy Results from Computational Model: A. PCA results for all individuals in population at the end of evolution (after generation 1,000) when polycomb mechanisms have more fully evolved to visually assess phenotypic pliancy. When transfer from environment 1 to environment 2, only the case when polycomb mechanisms are broken and switched to environment 2 (red) alter their phenotypes to resemble that of environment 2 that moved to (blue X's). The individuals with polycomb mechanisms left intact and moved to environment 2 (cyan) have phenotypes that do not switch and resemble more closely environment 1 that moved from (green). **B.** Principle Componenet Analysis (PCA) results for all individuals in population right after generation 50 during evolution when polycomb mechanisms have not more fully evolved to visually assess phenotypic pliancy. When transfer from environment 1 to environment 2, both the cases when polycomb mechanisms are left intact (cyan) and broken (red) alter their phenotypes to resemble that of environment 2 that moved to (blue X's). **C.** Average phenotypic pliancy score when vary degree of PcG-like mechanisms dsyregulation during evolution for all 10,000 populations when move from environment 1 to environment 2 to quantitatively assess pliancy. We vary degree of dsyregulation by breaking PRC1 alone (blue), PRC2 alone (cyan), or both PRC1 and 2 (red) for all 10,000 populations for a total of 10 million simulated cells pliancy score averaged (y-axis) for different generations throughout evolution (x-axis).

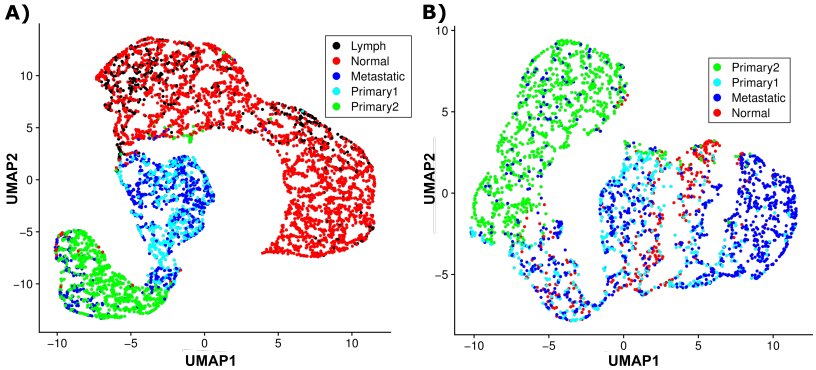

Fig 4. UMAP Results for Metastatic Cancer SC-RNA-Seq Datasets: UMAP results for the two single-cell RNA-sequencing data sets: **A.** H&N data and **B.** ovarian data. Note: the primary tumor cells are colored based on the PCA sub-clustering in Supplementary Figure 5.

References

1. Cepko CL, Austin CP, Yang X, Alexiades M, Ezzeddine D. Cell fate deternination in the vertebrate retina. PNAS. 1996;93:589–595.

2. Gurdon JB, Laskey RA, Reeves OR. The developmental capacity of nuclei transplanted from keratinized skin cells of adult frogs. Journal of Embryology and Experimental Morphology. 1975;34(1). doi:10.1242/dev.34.1.93.

3. Piunti A, Shilatifard A. The roles of Polycomb repressive complexes in mammalian development and cancer. Nature Reviews Molecular Cell Biology. 2021;22(5):326–345. doi:10.1038/s41580-021-00341-1.

4. Comet I, Riising EM, Leblanc B, Helin K. Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. Nature Publishing Group. 2016;doi:10.1038/nrc.2016.83.

5. Ng RK, Gurdon JB. Epigenetic inheritance of cell differentiation status. Cell Cycle. 2008;7(9):1173–1177. doi:10.4161/cc.7.9.5791.

6. Prohaska SJ, Stadler PF, Krakauer DC. Innovation in gene regulation: The case of chromatin computation. Journal of Theoretical Biology. 2010;265:27–44. doi:10.1016/j.jtbi.2010.03.011.

7. Whitcomb SJ, Basu A, Allis CD, Bernstein E. Polycomb Group proteins: an evolutionary perspective. Trends in Genetics. 2007;doi:10.1016/j.tig.2007.08.006.

8. Whitcomb SJ, Basu A, Allis CD, Bernstein E. Polycomb Group proteins: an evolutionary perspective. Trends in Genetics. 2007;doi:10.1016/j.tig.2007.08.006.

9. Sawarkar R, Paro R. Interpretation of Developmental Signaling at Chromatin: The Polycomb Perspective. Developmental Cell. 2010;19:651–661. doi:10.1016/j.devcel.2010.10.012.

10. Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. The Core of the Polycomb Repressive Complex Is Compositionally and Functionally Conserved in Flies and Humans. MOLECULAR AND CELLULAR BIOLOGY. 2002;22(17):6070–6078. doi:10.1128/MCB.22.17.6070-6078.2002.

11. Schwartz YB, Pirrotta V. Polycomb silencing mechanisms and the management of genomic programmes. Nature Reviews Genetics. 2007;8(1):9–22. doi:10.1038/nrg1981.

12. Oktaba K, Gutié Rrez L, Gagneur J, Girardot C, Sengupta AK, Furlong EEM, et al. Dynamic Regulation by Polycomb Group Protein Complexes Controls Pattern Formation and the Cell Cycle in Drosophila. Developmental Cell. 2008;15:877–889. doi:10.1016/j.devcel.2008.10.005.

13. Petruk S, Sedkov Y, Johnston DM, Hodgson JW, Black KL, Kovermann SK, et al. TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell. 2012;150(5). doi:10.1016/j.cell.2012.06.046.

14. Schuettengruber B, Bourbon HM, Croce LD, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell. 2017;171. doi:10.1016/j.cell.2017.08.002.

15. Schuettengruber B, Martinez AM, Iovino N, Cavalli G. Trithorax group proteins: switching genes on and keeping them active. Nature Reviews Molecular Cell Biology. 2011;12:799–814. doi:10.1038/nrm3230.

16. Gombar S, MacCarthy T, Bergman A. Epigenetics Decouples Mutational from Environmental Robustness. Did It Also Facilitate Multicellularity? PLoS Computational Biology. 2014;10(3):e1003450. doi:10.1371/journal.pcbi.1003450.

17. Sauvageau M, Sauvageau G. Polycomb Group Proteins: Multi-Faceted Regulators of Somatic Stem Cells and Cancer. Cell Stem Cell Review. 2010;7:299–313. doi:10.1016/j.stem.2010.08.002.

18. Simon JA, Kingston RE. Mechanisms of Polycomb gene silencing: knowns and unknowns. Nature Reviews Molecular Cell Biology. 2009;10:697–708. doi:10.1038/nrm2763.

19. Di Croce L, Helin K. Transcriptional regulation by Polycomb group proteins. Nature Publishing Group. 2013;20(10):1–4. doi:10.1038/nsmb.2669.

20. Bauer DC, Zadoorian A, Wilson LOW, Thorne NP. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. Briefings in Bioinformatics. 2016;19(July 2016):bbw097. doi:10.1093/bib/bbw097.

21. Entrevan M, Schuettengruber B, Cavalli G. Regulation of Genome Architecture and Function by Polycomb Proteins. Trends in Cell Biology. 2016;26(7). doi:10.1016/j.tcb.2016.04.009.

22. Grossniklaus U, Paro R. Transcriptional Silencing by Polycomb-Group Proteins. Cold Spring Harbor perspectives in biology. 2014;6. doi:10.1101/cshperspect.a019331.

23. Simon B, Fletcher JA, Doebeli M. Towards a general theory of group selection. Evolution. 2013;67(6):1561–1572. doi:10.1111/j.1558-5646.2012.01835.x.

24. Gligorijevic B, Bergman A, Condeelis J. Multiparametric Classification Links Tumor Microenvironments with Tumor Cell Phenotype. PLoS Biol. 2014;12(11). doi:10.1371/journal.pbio.1001995.

25. Ferrao PT, Behren A, Anderson RL, Thompson EW. Cellular and phenotypic plasticity in cancer. Frontiers in Oncology. 2015;5(171). doi:10.3389/fonc.2015.00171.

26. Yuan S, Norgard RJ, Stanger BZ. Cellular plasticity in cancer. Cancer Discovery. 2019;9(7). doi:10.1158/2159-8290.CD-19-0015.

27. Teeuwssen M, Fodde R. Cell heterogeneity and phenotypic plasticity in metastasis formation: The case of colon cancer. Cancers. 2019;11(9). doi:10.3390/cancers11091368.

28. Aw Yong KM, Sun Y, Merajver SD, Fu J. Mechanotransduction-Induced Reversible Phenotypic Switching in Prostate Cancer Cells. Biophysical Journal. 2017;112(6). doi:10.1016/j.bpj.2017.02.012.

29. Pourfarhangi KE, Bergman A, Gligorijevic B. ECM Cross-Linking Regulates Invadopodia Dynamics. Biophysical Journal. 2018;114(6). doi:10.1016/j.bpj.2018.01.027.

30. Klein CA. Selection and adaptation during metastatic cancer progression. Nature. 2013;501(7467):365–372. doi:10.1038/nature12628.

31. Stoletov K, Kato H, Zardouzian E, Kelber J, Yang J, Shattil S, et al. Visualizing extravasation dynamics of metastatic tumor cells. Journal of Cell Science. 2010;123:2332–2341. doi:10.1242/jcs.069443.

32. Yokota J. Tumor progression and metastasis. Carcinogenesis. 2000;21(3):497–503.

33. Bracken AP, Helin K. Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nature Reviews Cancer. 2009;9. doi:10.1038/nrc2736.

34. Yu J, Yu J, Rhodes DR, Tomlins SA, Cao X, Chen G, et al. A Polycomb Repression Signature in Metastatic Prostate Cancer Predicts Cancer Outcome. Cancer Res. 2007;67(22):10657–63. doi:10.1158/0008-5472.CAN-07-2498.

35. Kleer CG, Cao Q, Varambally S, Shen R, Ota I, Tomlins SA, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. PNAS. 2003;100(20):11606–11611.

36. Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO Journal. 2003;22(20):5323–5335. doi:10.1093/EMBOJ/CDG542.

37. Brabletz T. EMT and MET in Metastasis: Where Are the Cancer Stem Cells? Cancer Cell. 2012;22(6):699–701.

38. Barnhart BC, Simon MC. Metastasis and stem cell pathways. Cancer Metastasis Review. 2007;26:261–271. doi:10.1007/s10555-007-9053-3.

39. Brabletz T, Jung A, Reu S, Porzner M, Hlubek F, Kunz-Schughart LA, et al. Variable-catenin expression in colorectal cancers indicates tumor progression driven by the tumor environment. PNAS. 2001;98(18):10356–10361.

40. Rossari F, Zucchinetti C, Buda G, Orciuolo E. Tumor dormancy as an alternative step in the development of chemoresistance and metastasis-clinical implications. Cellular Oncology. 2020;43:155–176. doi:10.1007/s13402-019-00467-7.

41. Crea F, Ridzwan N, Saidy N, Collins CC, Wang Y. The epigenetic/noncoding origin of tumor dormancy. Trends in Molecular Medicine. 2015;21(4). doi:10.1016/j.molmed.2015.02.005.

42. Wortzel I, Dror S, Kenific CM, Lyden D. Exosome-Mediated Metastasis: Communication from a Distance. Developmental Cell. 2019;49. doi:10.1016/j.devcel.2019.04.011.

43. Syn N, Wang L, Sethi G, Thiery JP, Goh BC. Exosome-Mediated Metastasis: From Epithelial-Mesenchymal Transition to Escape from Immunosurveillance. Trends in Pharmacological Sciences. 2016;37(7). doi:10.1016/j.tips.2016.04.006.

44. Siegal ML, Bergman A. Waddington's canalization revisited: Developmental stability and evolution. PNAS. 2002;99(16):10528–10532. doi:10.1073[U+517E]pnas.102303999.

45. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 2017;171(7):1611–1624.e24. doi:10.1016/j.cell.2017.10.044.

46. Shih ID AJ, Menzin A, Whyte J, Lovecchio J, Liew A, Khalili H, et al. Identification of grade and origin specific cell populations in serous epithelial ovarian cancer by single cell RNA-seq. 2018;doi:10.1371/journal.pone.0206785.

47. Kolybaba A, Classen AK. Sensing cellular states—signaling to chromatin pathways targeting Polycomb and Trithorax group function. Cell Tissue Research. 2014;365:477–493. doi:10.1007/s00441-014-1824-x.

48. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature Biotechnology. 2018;36(5). doi:10.1038/nbt.4096.

49. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888–1902.e21. doi:10.1016/j.cell.2019.05.031.

50. Lytal N, Ran D, An L. Normalization Methods on Single-Cell RNA-seq Data: An Empirical Survey. Frontiers in Genetics. 2020;11(41). doi:10.3389/fgene.2020.00041.

51. Becht E, Mcinnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnology. 2019;38(1). doi:10.1038/nbt.4314.

52. Bergman A, Siegal ML. Evolutionary capacitance as a general feature of complex gene networks. Nature. 2003;424(6948):549–52. doi:10.1038/nature01765.

53. Kacser H, Burns JA. The molecular basis of dominance. Genetics. 1981;97(3-4):639–666. doi:10.1056/NEJM199402033300507.

54. MacCarthy T, Bergman A. Coevolution of robustness, epistasis, and recombination favors asexual reproduction. PNAS. 2007;104(31):12801–12806. doi:10.1073[U+517E]pnas.0705455104.

55. Portela A, Esteller M. Epigenetic modifications and human disease. Nature Biotechnology. 2010;28(10). doi:10.1038/nbt.1685.

56. Featherstone DE, Broadie K. Wrestling with pleiotropy: Genomic and topological analysis of the yeast gene expression network. BioEssays. 2002;24(3):267–274. doi:10.1002/bies.10054.

Supporting information

Model Parameter Testing

We test the sensitivity of the model's behavior, and the generality of our results, by measuring the change in the percent of cells exhibiting overall phenotypic pliancy when PcG-like mechanism is intact versus broken over a wide range of parameters and 10 randomly chosen starting gene-regulatory network architectures. Specifically, we vary the following five parameters: gene-regulatory network connectivity density (=0.1, 0.3, 0.5), selection strength (=0.5, 1, 2), gene-activation sigmoidal strength (=1, 4, 6), gene threshold level that triggers PRC repression (=0.1, 0.15, 0.2), and mutation rates (gene mutation rate per genome, environment interaction mutation rate, and PcG-like mechanism mutation rate) (=0.1, 0.2, 0.3). Therefore, we obtain results for 243 different parameter settings. We vary these five parameters independently, and we use the same starting population that undergoes the same 100 different evolutionary trajectories for each parameter setting. We then calculate the percent of phenotypically pliant individuals in each evolved population when PcG-like mechanism is left intact or is broken, then calculate the average for each of these two cases over the 100 different evolved populations for each parameter setting. Given that our phenotypic pliancy score (see Materials and Methods) can only be measured for the individuals with broken PcG-like mechanism, we measure overall phenotypic pliancy for each case when PcG-like mechanism is broken vs. left intact to better compare parameter effects. We measure overall phenotypic pliancy by calculating the distance between its stable phenotype after development in its original environment, S_W, and the resulting stable phenotype upon transferring it post-developmentally to another environment, S'_W. Phenotypic pliancy for a given cell corresponds to a large Euclidean distance $|S_W - S'_W|$, and phenotypic fidelity to a small distance. In our preliminary work, we used a threshold distance of 0.05 to determine phenotypic pliancy, such that if the Euclidean distance is greater than 0.05 then that cell is considered phenotypically pliant. We see a drastic increase in the average phenotypic pliancy when PcG-like mechanism is broken for each of the 243 different parameter settings (p-value = 10^{-16}) (see Supplementary Figure 8). Our parameter testing results show the different parameter settings and network architectures do not change our phenotypic pliancy results, strongly suggesting that, while our model lacks biological specificity, it is still biologically relevant in its general implications.

Supplementary Figures and Tables

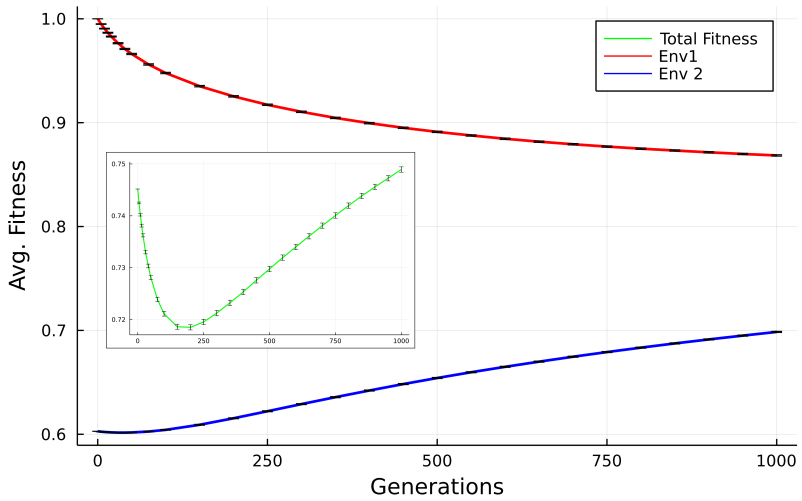

Figure S 1. Average Fitness Throughout Evolution: The average combined fitness for both environment 1 and environment 2 shown in the insert due to scale (green), and the average fitness for environment 1 (red) and environment 2 (blue) fitness results.

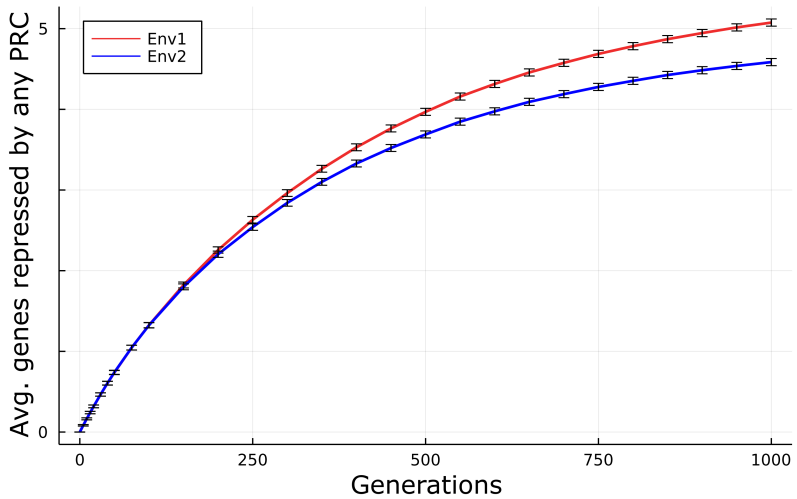

Figure S 2. Genes Repressed by PcG-like mechanisms During Evolution: The average number of genes that are repressed by any PRC during evolution in environment 1 (red) or environment 2 (blue).

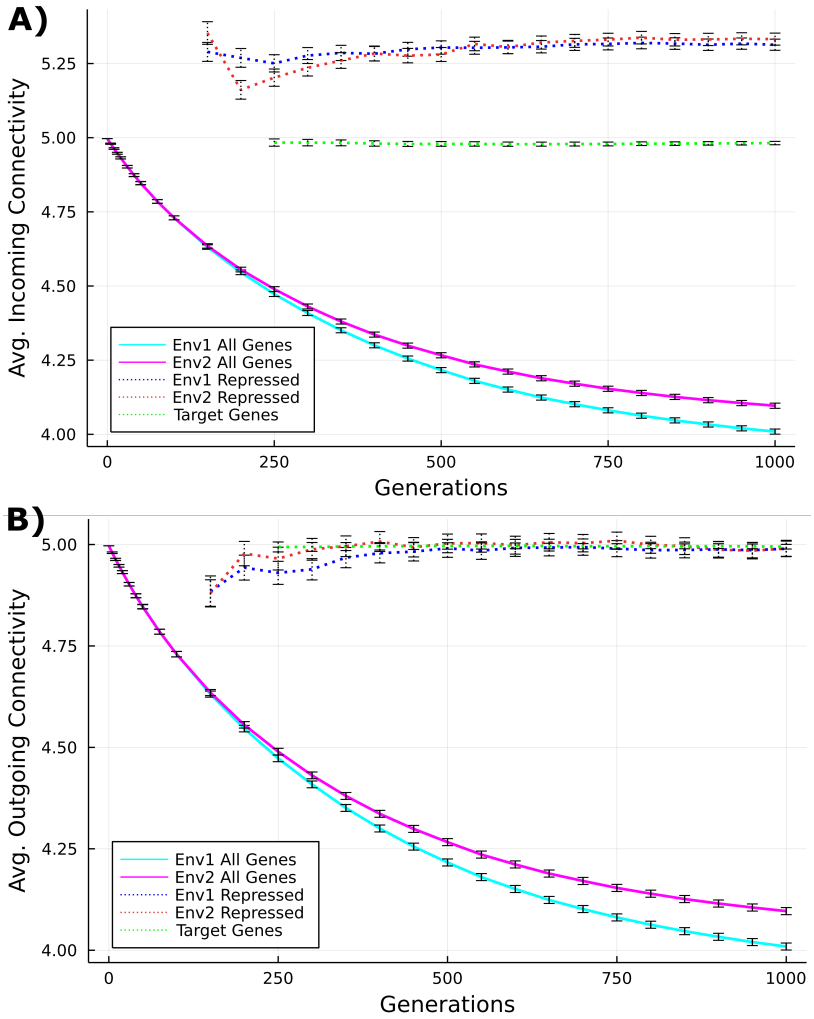

Figure S 3. Connectivity of the Gene Regulatory Networks Throughout Evolution: The average incoming (A) and outgoing (B) effective connectivity of the gene regulatory networks for all 10 million cells (solid lines). The effective connectivity averaged across all the genes in the network for environment 1 and environment 2 decreases throughout evolution as would expect (magenta and cyan solid lines). The average connectivity of the target PRC genes (green dotted lines) and genes repressed by any PRC (red and blue dotted lines) are shown as well. Genes that are repressed by any PRC have a higher average incoming connectivity as compared to all the target genes. Note, we look at average connectivity, as opposed to average effective connectivity, for PcG-like mechanism target genes that are repressed or not to show the connections the repressed target had before repression.

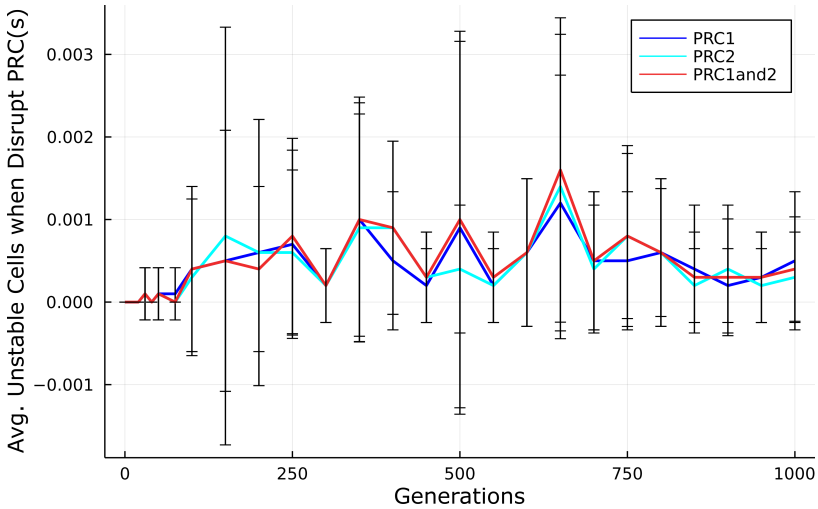

Figure S 4. Average Number of Unstable Cells For Varying Degree of PcG-like Mechanisms Breakage: The average number of unstable cells in each of the 10,000 different populations when break PRC1 alone (blue), PRC2 alone (cyan), and PRC1 and PRC2 together (red).

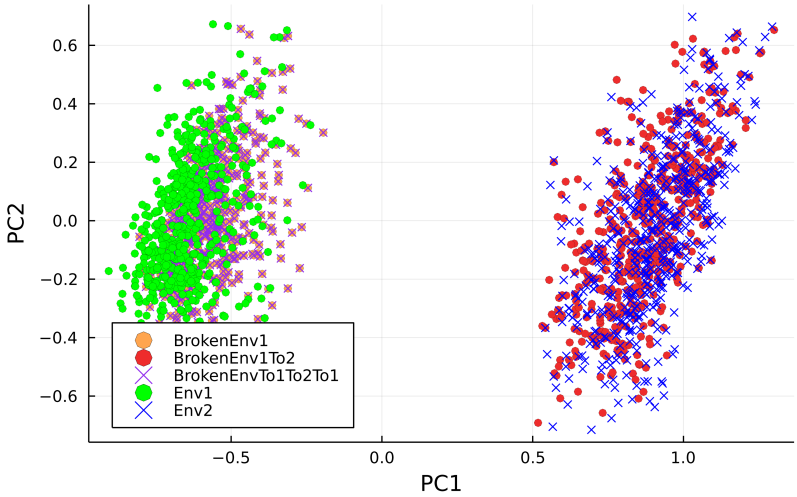

Figure S 5. Sustained Phenotypic Pliancy due to PcG-like Breakdown in Model: Principle Component Analysis (PCA) showing that when PcG-like mechanisms are dysregulated in our model that phenotypic switching is sustainable, such that after switching from environment 1 to environment 2 (red circles) and then back to environment 1 (purple X's) that the phenotypes more closely resemble what environment they were switched to. As a control, we also see that if break PcG-like mechanisms but keep in environment 1 (tan circles) then phenotype stays close to that of the evolved phenotype in environment 1 when PcG-like mechanisms were intact.

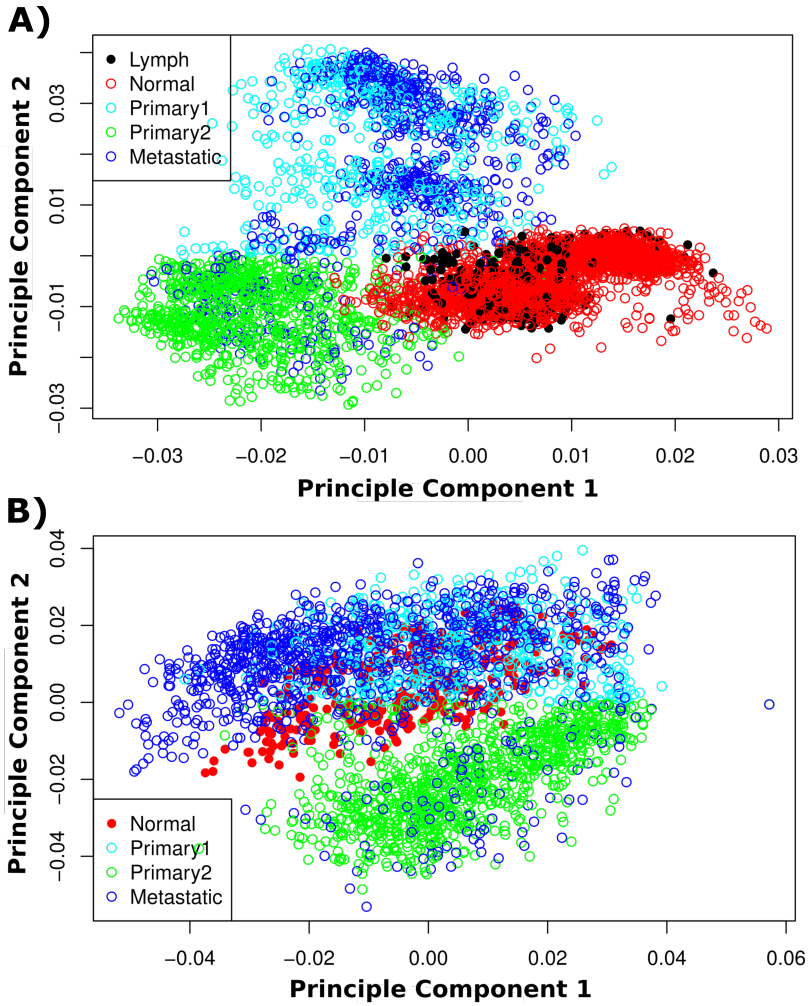

Figure S 6. PCA Results for Metastatic Cancer Datasets: Principle Component Analysis (PCA) results for **A.** Head and Neck and **B.** Ovarian metastatic cancer dataset, where metastatic cells' phenotypes (given by their gene expression patterns) are each represented by dark blue circles, primary cancer cell's phenotypes are represented by both green and cyan, and the normal non-cancer cells' phenotypes at the metastatic site and primary site are represented by black and red circles, respectively. Note, the primary cells are split by principle component 2 (y-axis), such that primary cells with PC2 values greater than zero are represented by cyan circles and primary cells with values less than zero are represented by green circles.

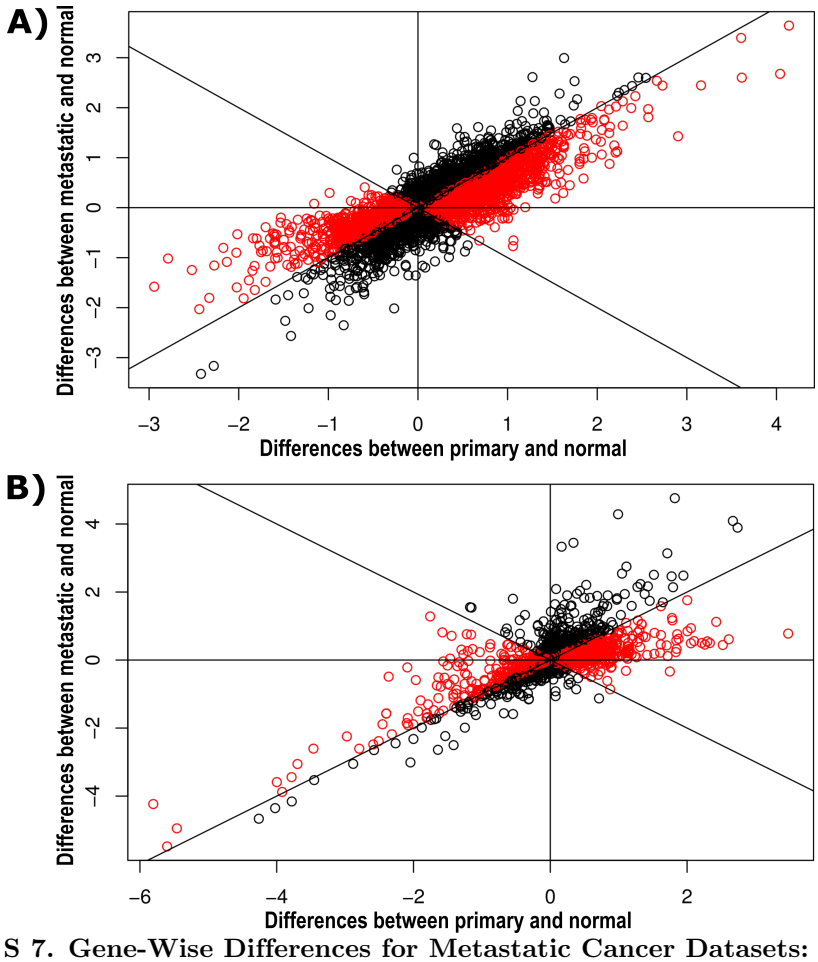

Figure S 7. Gene-Wise Differences for Metastatic Cancer Datasets: Plot of the gene-by-gene differences between primary and normal at the primary site (x-axis) versus the differences between metastatic and normal at the lymph node site or ovary site for the **A.** head and neck and **B.** ovarian metastatic cancer dataset, respectively (y-axis). We have included lines $y = x$ and $y = -x$ for comparison. There is a trend for the gene-wise differences between metastatic and normal at the lymph node site being smaller in absolute value than gene-wise differences between primary and normal at the primary site, which is the case for **A.** 68% of the genes differentially expressed between primary tumor versus metastatic cells (p-value $< 10^{-8}$) for H&N (shown in red), and **B.** 72% (p-value $< 10^{-9}$) for ovarian (shown in red).

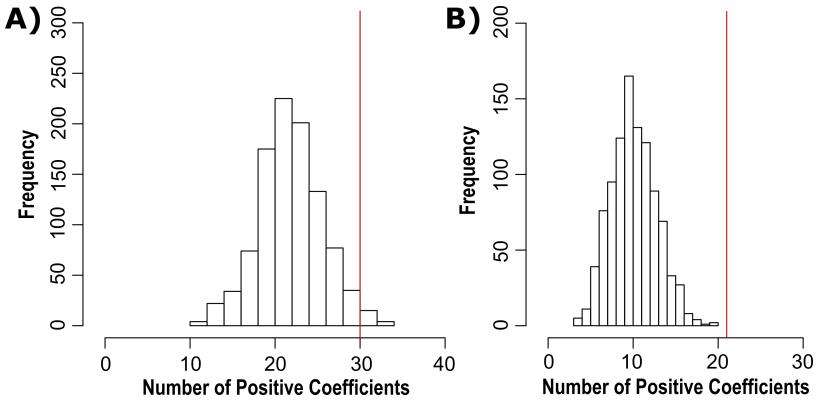

Figure S 8. Significance Assessment of PcG Genes as Source of Pliancy: Histogram of the bootstrapping results, which we perform by randomly picking 1000 samples of PcG mechanism genes and performing linear regression of the score as a function of these genes' expression profiles for **A.** H&N and **B.** ovarian data, respectively. Each plot is the distribution of the positive coefficients computed for each random sample. The red line is the number of PcG mechanism genes with positive coefficients, which is 30 for H&N data and 21 for ovarian data. The mean is 22.3 out of 53 for H&N data, and mean is 10.6 out of 56 for the ovarian data, so the mean for both **A.** and **B.** (red line) is statistically significant (p-value < 0.05 and p-value $< 10^{-5}$, respectively).

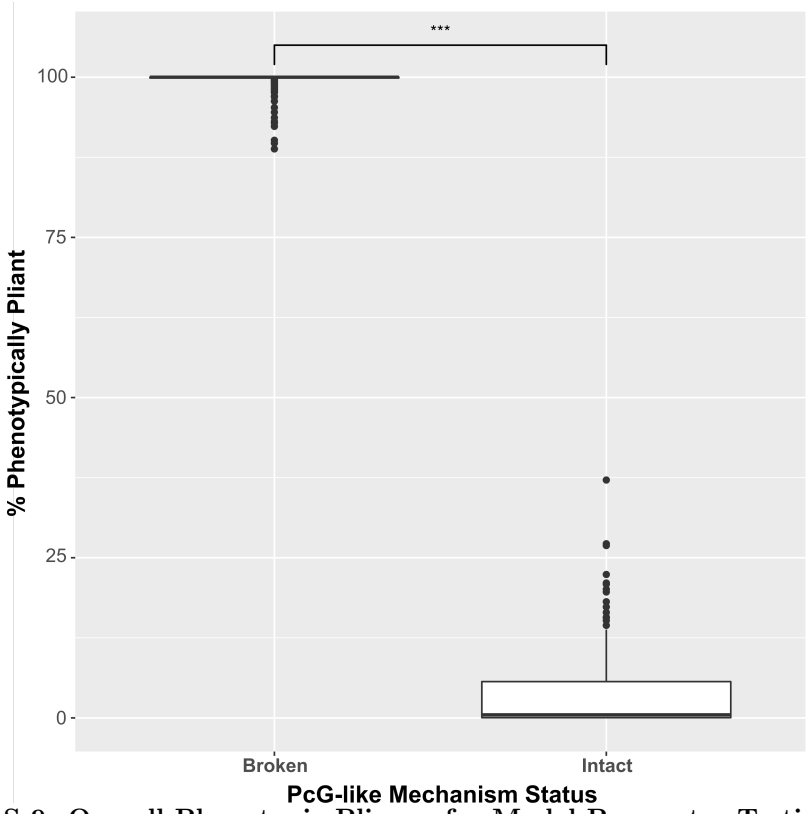

Figure S 9. Overall Phenotypic Pliancy for Model Parameter Testing: Phenotypic pliancy when PcG-like mechanism is intact versus broken over a wide range of parameters and 10 randomly chosen starting gene-regulatory network architectures. The percent of cells that are phenotypically pliant when PcG-like mechanism is broken is statistically significantly greater than when the PcG-like mechanism remains intact (p-value < 0.001).

Table 1. Polycomb Mechanism Genes Used in Single-Cell RNA-Sequencing Data Analysis:

PcG Protein	TrxG Protein	Controlled by PRC
RING1	CHD8	PCDH15
RNF2	ASH2L	PCDHB1
CBX2	SMARCA1	PCDHB4
CBX4	SMARCA2	PCDHB15
CBX6	RBBP5	CDH8
CBX7	WDR5	CDH13
CBX8	CHD3	CDH18
PCGF1	CHD4	CDH19
PCGF2	CHD1	CDH23
PCGF3	CHD2	RAP1A
BMI1	MTA1	RAP1B
PCGF5	MTA2	RAP1GAP
PCGF6	MTA3	
SCMH1	HDAC1	
RYBP	HDAC2	
YAF2	MBD2	
EZH1	MBD3	
EZH2	POLD3	
EED	BPTF	
SUZ12	SMARCB1	
RBBP4	DPF1	
RBBP7	KMT2A	
	KMT2D	
	KMT2C	
	KMT2B	
	KAT8	
	DPY30	
	SETD1A	
	SETD1B	
	CXXC1	
	WDR82	
	KDM6A	
	NCOA6	
	PAXIP1	
	PAGR1	

Table 2. Differential gene expression analysis for metastatic cancer data sets.

Data (Tumor)	DE PcG Genes	Log-Fold Change Sum
H&N (Primary)	42 genes	-20.1
H&N (Metastatic)	45 genes	-29.2
Ovarian (Primary)	25 genes	20.2
Ovarian (Metastatic)	16 genes	-14.3

Table 3. Model Parameters and Values

Parameter	Value
Genes	50
Population Size	1,000
Gene Regulatory Network Connectivity	0.1
Generations	1,000
Environment Components	50
Maximum Iterations	100
Gene Mutation Rate	0.1
Environment Interaction Mutation Rate	0.1
PcG-like Mechanism Mutation Rate	0.1
Selection Strength	0.5
Gene-Activation Sigmoidal Strength	1.0
PRC Repression Gene Threshold Level	0.15
PRC Critical Time Point	2
Number Different Environments	2
Number of Different PRCs	2
Proportion of Env. Components Affecting Cell > 0	0.4
Proportion of Genes Able to be Affected by Envs.	0.4
Proportion of Env. Components Affecting Gene	0.4
Minimum Difference Between Env. 1 and Env. 2	70%
Minimum Difference in Env. Optimum States	40%
Individual Weights	Gaussian
Sexual Reproduction Flag	true