

# SCIENTIFIC REPORTS



OPEN

## Discovering novel phenotypes with automatically inferred dynamic models: a partial melanocyte conversion in *Xenopus*

Daniel Lobo<sup>1</sup>, Maria Lobikin<sup>2</sup> & Michael Levin<sup>2</sup>

Received: 28 October 2016  
 Accepted: 16 December 2016  
 Published: 27 January 2017

Progress in regenerative medicine requires reverse-engineering cellular control networks to infer perturbations with desired systems-level outcomes. Such dynamic models allow phenotypic predictions for novel perturbations to be rapidly assessed *in silico*. Here, we analyzed a *Xenopus* model of conversion of melanocytes to a metastatic-like phenotype only previously observed in an all-or-none manner. Prior *in vivo* genetic and pharmacological experiments showed that individual animals either fully convert or remain normal, at some characteristic frequency after a given perturbation. We developed a Machine Learning method which inferred a model explaining this complex, stochastic all-or-none dataset. We then used this model to ask how a new phenotype could be generated: animals in which only some of the melanocytes converted. Systematically performing *in silico* perturbations, the model predicted that a combination of altanserin (5HTR2 inhibitor), reserpine (VMAT inhibitor), and VP16-XICreb1 (constitutively active CREB) would break the all-or-none concordance. Remarkably, applying the predicted combination of three reagents *in vivo* revealed precisely the expected novel outcome, resulting in partial conversion of melanocytes within individuals. This work demonstrates the capability of automated analysis of dynamic models of signaling networks to discover novel phenotypes and predictively identify specific manipulations that can reach them.

One of the key areas in which artificial intelligence and the information sciences can contribute to biology is by helping human scientists understand cellular behavior in the context of a complex organism<sup>1,2</sup>. Computational methods can reverse-engineer mechanistic models of tissue patterning and shape formation from expression data and experimental phenotypes<sup>3–9</sup>. The utility of these methods is their ability to find novel regulatory interactions<sup>10</sup> and even novel necessary regulatory genes<sup>11</sup>. These methods are indeed becoming indispensable for understanding the complex coordination of signals necessary to develop and maintain correct body shapes and organs. Moreover, such methods are required in order to develop interventions to make rational changes to complex anatomy and physiology, in the context of regenerative medicine and systems-level diseases such as cancer<sup>12</sup>.

The coordination of cellular behavior towards the anatomical needs of the host organism, and away from tumorigenesis, is achieved in part via bioelectrical communication among many cell types<sup>13–19</sup>. Recent work showed that depolarization of resting potential in a special cell population in *Xenopus* embryos, so-called instructor cells, results in a metastatic-like conversion of normal melanocytes<sup>20</sup>. Throughout the body (including at considerable distance from the depolarized region), these pigment cells become hyper-proliferative, acquire a highly arborized morphology, and invade organs and blood vessels, colonizing ectopic regions of the body. This conversion recapitulates the main features of a melanoma-like phenotype (including up-regulation of cancer-related genes)<sup>21</sup>. It is quite distinct from pigment changes occurring due to light exposure<sup>22–24</sup> because it affects not the pigment level of individual melanocytes but instead triggers drastic invasiveness that does not significantly change as a function of light level. The molecular pathway involved is only partially understood, and comprises voltage-dependent movement of serotonin and subsequent serotonergic signaling<sup>25</sup>. This pathway is an important proof-of-principle of activating a melanoma-like invasive phenotype in the absence of oncogenic mutation or carcinogens<sup>26</sup>.

<sup>1</sup>Department of Biological Sciences, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA. <sup>2</sup>Allen Discovery Center at Tufts University, and Department of Biology, Tufts University, 200 Boston Avenue, Suite 4600, Medford, MA 02155, USA. Correspondence and requests for materials should be addressed to M.L. (email: michael.levin@tufts.edu)

Although the background rate in control populations is very low (1% of the animals), the incidence of conversion from a depolarization treatment (such as the chloride channel opener ivermectin, which in *Xenopus* depolarizes cells bearing the GlyR channel<sup>20,27</sup>) is ~98%<sup>21</sup>. Other treatments (e.g., those interfering with serotonergic steps that mediate bioelectric control of melanocytes) result in intermediate penetrance of the phenotype, with each of them occurring in various characteristic percentages of a treated population<sup>21</sup>. However, a most striking aspect is that it is all-or-none at the level of the individual. While the decision to convert or not is stochastic within the treated cohort<sup>21</sup>, it is always made at the level of the whole animal: no partially converted tadpoles have been seen in any of the experiments (including dozens of different reagents and means of altering serotonergic or bioelectric properties of cells *in vivo*). Each animal is either fully covered with the ectopic, highly-arborized melanocytes, or it looks completely normal (Fig. 1). The cellular network either reaches a fully-converted outcome or not, but the entire larva does it as a coordinated unit. For any given treatment, each individual makes a stochastic decision to convert or not (so that the population converts at a rate specific for that perturbation), but that decision is coordinated among all the cells. A salt-and-pepper pattern, which would be expected from cells that made independent stochastic decisions, has not been reported in this context. Despite the detailed molecular and cell-level dissection of the signaling pathway involved, the stochastic nature and whole-body coordination in this system remains largely mysterious. We thus turned to Machine Learning as a way to discover models that could help understand the global systems properties of this event.

Attempting to understand this synchronized stochasticity, a recent computational approach inferred a fully specified signaling network that quantitatively explained the dynamics of this system<sup>21</sup>. Using incidence of conversion from each perturbation experiment as input, together with a basic pathway that had been uncovered between the relevant signaling components<sup>20,25</sup>, the method reverse-engineered a complete signaling network including all the necessary and sufficient regulatory elements, interactions, and parameters that could recapitulate the frequencies of converted vs. wild-type larvae observed *in vivo* in each experiment.

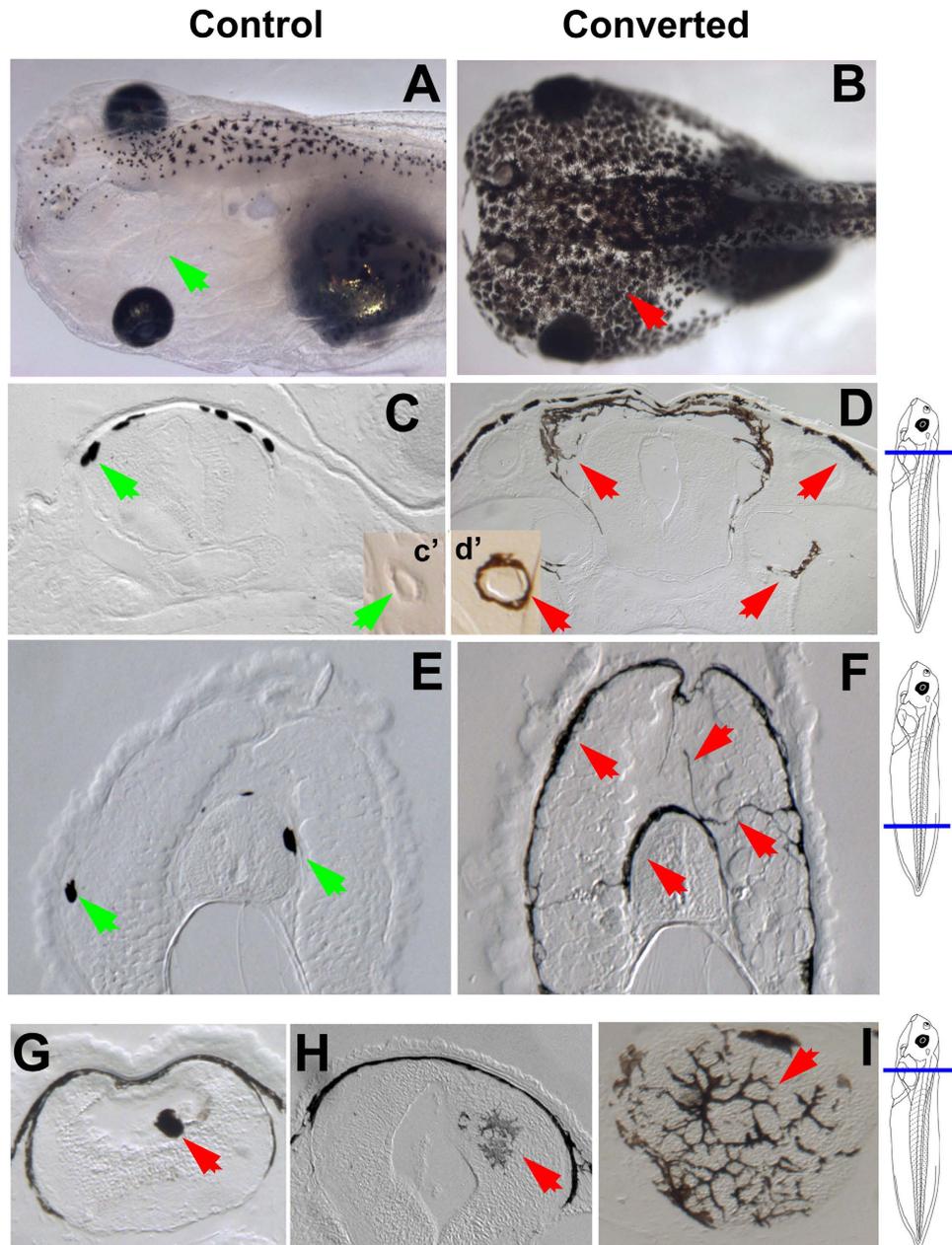
The use of machine learning to understand complex dynamics from molecular functional data is an important strategy for addressing the gap between increasingly available genetic data and the control of large-scale system behaviors in biomedicine<sup>28–31</sup>. The ultimate goal for such systems is not only to explain available data, but to be used to rapidly identify treatments that would give rise to desired outcomes, especially in situations too complex for human scientists to infer by hand what course of action will lead to a desired systems-level response. Thus, we sought to validate the identified network and approach using a most stringent test: interrogating the network to discover a treatment that produced a capability not reached in any of the previous experiments in this field. Specifically, we asked what experimental condition could break the synchrony and achieve partially-converted tadpoles. Our results validate the approach, identifying a treatment that for the first time produced partial melanocyte conversion within individual animals.

## Results

***In silico* discovery of a novel phenotype.** We employed a stochastic model of the regulation of melanocyte conversion in *Xenopus*, uncovered by a Machine Learning approach<sup>21</sup>, to computationally discover a novel phenotype and the pharmacological perturbations necessary to produce it. Figure 2A shows the inferred dynamic model, including the entities used to perturb the system in past work<sup>21</sup>, the signaling elements participating in the pathway<sup>20,25</sup>, the incidence of conversion produced by various phenotypes, unknown required components inferred (predicted) by the system, and all the interactions between all the components. The perturbation experiments consist in the application of a precise combination of reagents during *Xenopus* development, which stochastically results in a wild-type phenotype (correct numbers of round melanocytes in their normal locations) or a completely converted phenotype (embryo covered extensively with highly-arborized melanocytes that invade many internal tissues and blood vessels) (Fig. 1). Interestingly, no intermediate phenotype was observed in the training data set (20 experiments with none, one, or two reagents applied) or the validation data set (seven experiments with two or three reagents applied)—tadpoles converted or not, stochastically but always in an all-or-none manner for each animal<sup>21</sup>. Could this coordination be broken, and if so, how?

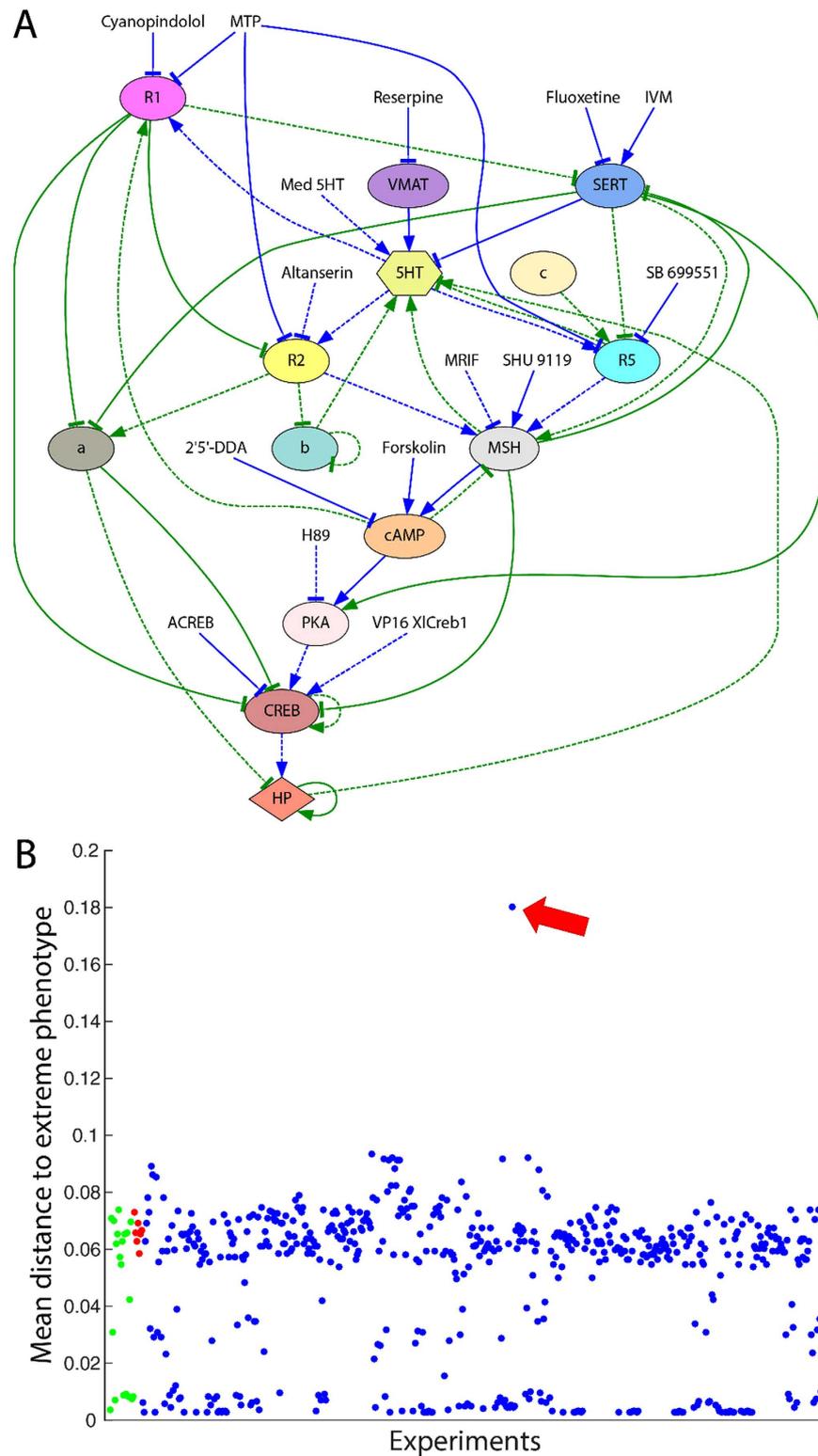
We then performed *in silico* all possible experiments involving the application of up to three reagents, for a total of 576 different experiments, each repeated 100 times. Figure 2B shows, for each experiment, the average distance from the resultant phenotype (degree of conversion) to either of the extreme phenotypes: normal or complete conversion (see methods section for the exact formulation of the distance metric). In green are those experiments (combination of reagents) used in the training dataset to infer the model, in red are those used in the validation dataset, and in blue are all the experiments that had never been performed *in vivo*. The results show that all except one experiment (red arrow) are below the 0.1 threshold to an extreme phenotype (normal or converted). The experiment above the 0.1 threshold indicates an intermediate phenotype between the previously observed wild-type normal pigmentation and the converted phenotype (i.e., a novel situation in which only some melanocytes within a single animal convert to the metastatic-like phenotype). This phenotype is predicted by the inferred model to result from a combined application of altanserin (an R2 inhibitor), reserpine (a VMAT inhibitor), and VP16-XlCreb1 (constitutively active CREB).

**Dynamics of the partially converted phenotype.** We first investigated *in silico* the dynamic characteristics of the predicted partially converted phenotype with the inferred model using the combined application of the drugs altanserin and reserpine, and constitutively active CREB. Figure 3 shows the phase portraits of 100 simulations of the wild type and the treated embryos with the trajectories of the state of two serotonin receptors (5HT-R1 and 5HT-R2) and the degree of conversion through time. The simulation starts in the unstable initial state (yellow dot) and, due to the inherent stochasticity of the model, ends in one of several stable attractors, representing a final partial converted state. Without treatment (wild type, Fig. 3A), the embryos stochastically develop normally (blue dot, 99% of simulations) or high conversion levels (red dot, 1% of simulations). In

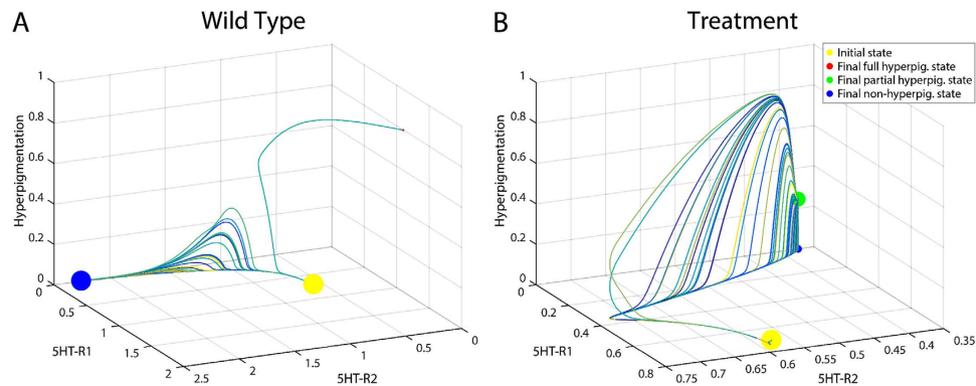


**Figure 1. Conversion of melanocytes to a metastatic-like state.** (A) Dorsal view of a wild-type st. 45 *Xenopus laevis* tadpole; note the small melanocytes, absent from large areas of the head (green arrowhead). (B) In contrast, animals resulting from any of several treatments<sup>20,39</sup> that depolarize instructor (GlyR-expressing) cells or perturb their downstream serotonergic signaling exhibit extensive overabundance and uniform coverage with highly arborized melanocytes (red arrowhead). This occurs in an all-or-none manner in some percentage of the animals (frequency depending on the specific manipulation)<sup>21,25</sup>. Sectioning (level of section shown in schematics to the right) reveals the main features of melanoma-like phenotype: over-proliferation, arborization, and invasiveness. (C) Normally round melanocytes dorsal to the neural tube (green arrow) become highly arborized and drop down over the neural tube itself (D, red arrows). Inset panels show blood vessels, normal in c' and covered by invasive melanocytes in d', as occurs in melanoma. Sections further along the tail likewise show small numbers of round melanocytes in control larvae (E, green arrows) compared to the excess of long, abnormally extended and ectopically localized melanocytes in converted animals (F, red arrows). In converted animals, cells can be seen invading the lumen (G) or neural tissue (H) of the neural tube, also often forming networks (I) as has been observed in vasculogenic mimicry of mammalian melanoma<sup>40</sup>. The authors thank Vaibhav Pai for allowing us to use his drawing of a st. 45 *Xenopus laevis* tadpole in this figure.

contrast, applying the discovered treatment with three drugs, the embryos stochastically develop with an intermediate conversion level (green dot, 75% of simulations) or a normal conversion level like the wild type (blue dot, 25% of simulations).



**Figure 2. Inferred signaling network model and *in silico* predictions.** (A) A computational method reverse-engineered a signaling network able to recapitulate the level of conversion stochasticity of a series of pharmacological experiments. (B) Phenotypic predictions of the model for all experimental combinations of up to three reagents. Each point represents the mean distance of 100 simulations of a specific experiment (i.e., a specific combination of reagents) to the two extreme tadpole phenotypes (normal and total conversion). Training dataset in green, validation dataset in red, new experiments in blue. Only one combination of three reagents (red arrow) is far from the two extreme phenotypes, indicating a partially converted phenotype.



**Figure 3. Phase space of the wild type and the experiment predicted to produce a partially pigmented phenotype.** The trajectories of the state of two serotonin receptors (5HT-R1 and 5HT-R2) and the degree of conversion is shown among 100 simulations. **(A)** Without any treatment (wild type), the dynamical system transitions from the initial instable state (embryo, yellow dot) stochastically towards a final stable state with a normal conversion level (0, blue dot) or towards a final state with a high conversion level (0.92, red dot). **(B)** Applying the treatment combining the discovered three reagents (altanserin, reserpine, and constitutively active CREB), the dynamics reveal a bifurcation in the dynamical system and the appearance of a new attractor with a partial conversion level (0.25, green dot) in addition to another attractor with a normal conversion level (0, blue dot). The attractor dot size is proportional to the number of converging trajectories to that final state.

The “partially converted” attractor (green dot) was not observed in any previous simulation or experiment *in vivo*<sup>21</sup>, representing a novel bifurcation in the signaling network’s dynamics due to the specific perturbation caused by the application of the three reagents. In this way, the inferred model predicted a partially converted phenotype at about 25% between wild-type and totally converted.

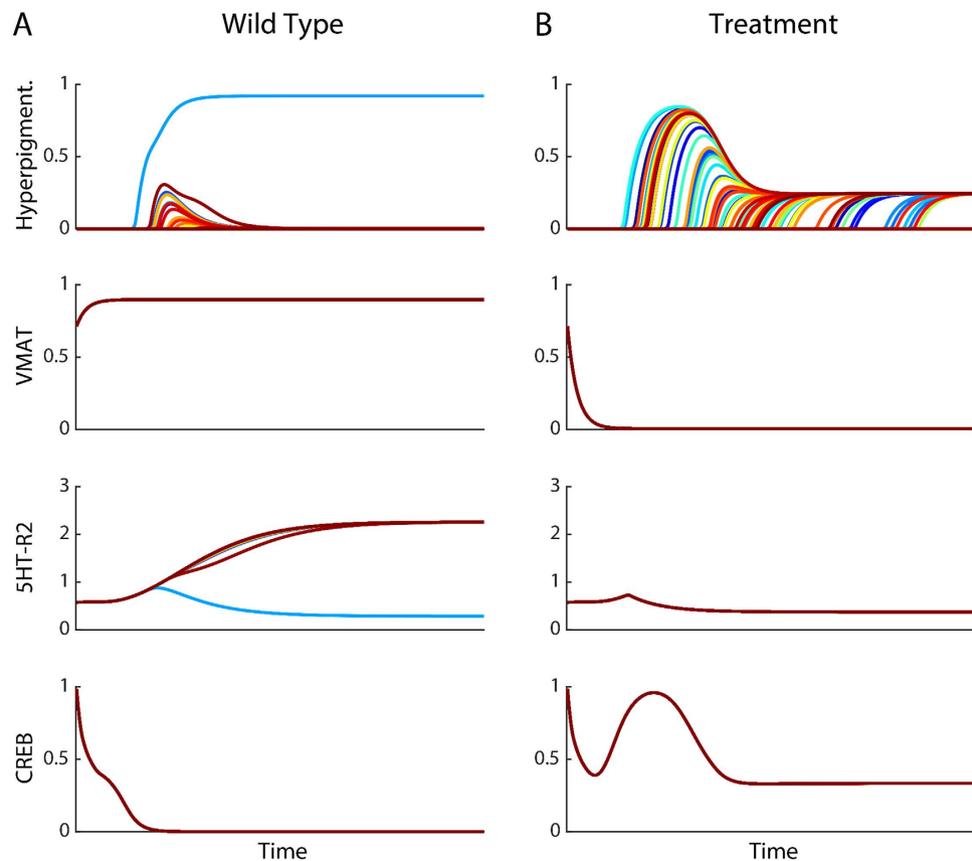
The effects of the three reagents in the treatment can be observed in the temporal dynamics of the level of conversion and pathway components targeted by the drugs (Fig. 4). Embryos with no treatment (wild type, Fig. 4A) stochastically develop normal levels of conversion (99% of the simulations) or high levels of conversion (1% of the simulations). These embryos always have high levels of VMAT and low levels of CREB. However, the levels of serotonin receptor R2 are stochastically either low (blue line, corresponding to high levels of conversion) or high (all other colors, corresponding to normal levels of conversion). Embryos under the three drugs treatment (Fig. 4B), stochastically develop either the normal level of conversion (25% of the simulations) or the partial conversion phenotype (75% of simulations). The plots show how the inhibition drugs reserpine and altanserin causes VMAT and R2, respectively, to converge to a low level, whereas the constitutively active CREB causes CREB levels to reach a high transient state and then converge to a lower level, but overall staying higher than the wild type.

***In vivo* validation of the computationally discovered treatment and phenotype.** We then sought to validate the partially converted phenotype predicted by the inferred model by applying *in vivo* the newly discovered combination of reagents. Control embryos exposed to vehicle alone were 99% normal (N = 258, Fig. 5A–A’). Embryos exposed to ivermectin (which depolarizes instructor cells; the effect has been characterized extensively<sup>20,25</sup>) exhibited 97% conversion (N = 313, Fig. 5B–B’). In contrast, embryos expressing constitutively active CREB (tom-VP16-XlCreb1) and simultaneously treated with a cocktail of Altanserin (R2 blocker, 10  $\mu$ M) and Reserpine (VMAT blocker, 100  $\mu$ M) exhibited the predicted never before-seen partial conversion in 22% of the animals (N = 167, Fig. 5C–D’), while 24% were completely normal, and 54% exhibited the standard full conversion phenotype.

## Discussion

Here, we have presented the discovery of a novel *in vivo* phenotype, and the precise intervention required to obtain it, by means of a computational methodology. From a reverse-engineered signaling network able to recapitulate the stochasticity of two phenotypes representing wild-type and extreme levels of melanocyte conversion in *Xenopus laevis* larvae, we sought to discover a pharmacological intervention that could generate a never-seen partially converted phenotype. We performed *in silico* all the possible experiments combining one, two, or three reagents and found that only one specific combination that was able to produce an intermediate phenotype. The model dynamics of this intervention showed that this precise combination produced a bifurcation in the dynamical system, generating an attractor corresponding to a phenotype 25% in between the known wild type and totally converted phenotypes. Almost a decade of highly diverse experiments focused on this pathway had not revealed such an outcome<sup>20,25</sup>. While detailed mechanistic dissection of the pathway explained signaling on a single cell level, explanations for the stochastic concordance throughout each animal, the possibility of partial conversion, or the reagents that could achieve it, had all remained unknown without the described automated approach that strongly augmented human analysis of this complex dynamical system.

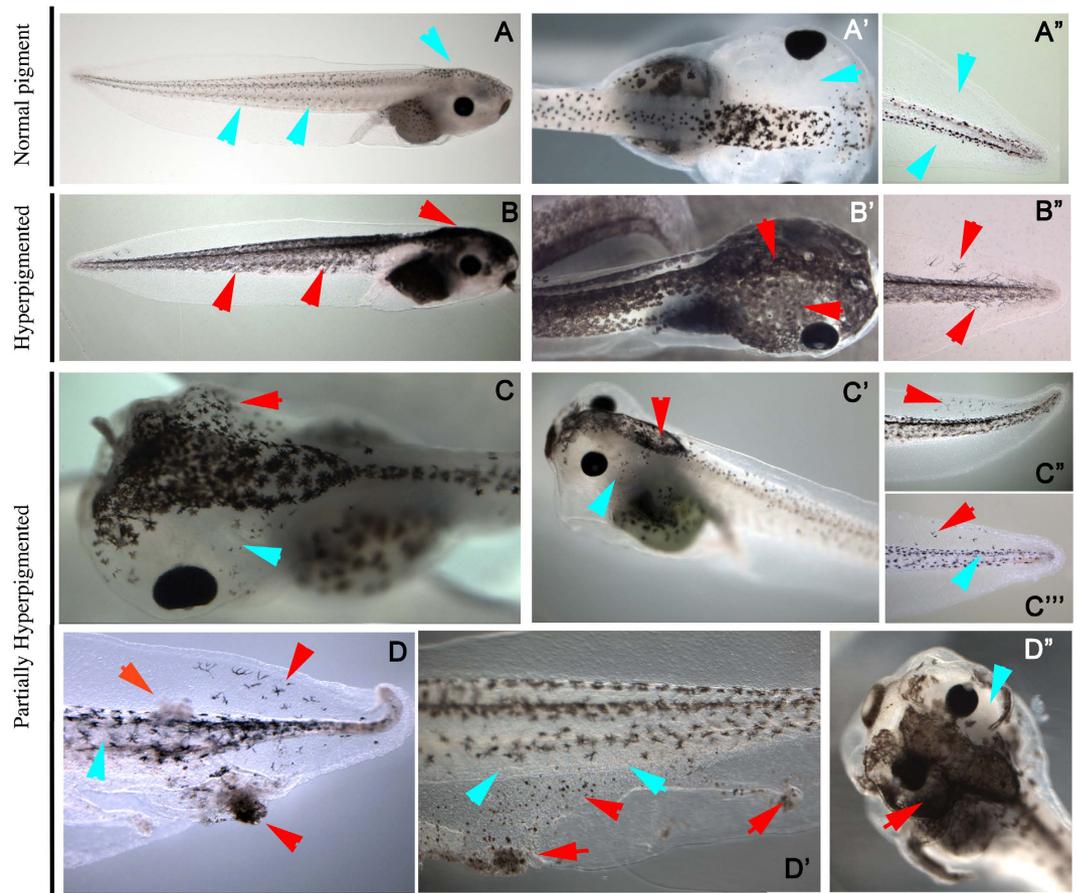
We validated these predictions *in vivo*, confirming the *in silico* discovery of the novel phenotype. When we applied the discovered combination of three reagents, two drugs and constitutively active CREB, we observed for the first time the predicted partial converted phenotype—animals in which some melanocytes and melanocyte-free regions were normal, while others were converted and colonizing ectopic sites. It is crucial to note that each of



**Figure 4. Dynamics of the degree of conversion and perturbed pathway components in the untreated and treated experiments.** The panels show the dynamics of the degree of conversion and levels of the affected components by the drugs in the treatment through time over 100 simulations. (A) With no treatment (wild type), only extreme phenotypes (normal or high conversion levels) are stochastically produced. VMAT converges to high levels, R2 to stochastically either high or low levels, and CREB to low levels. (B) The treatment combining reserpine (inhibiting VMAT), altanserin (inhibiting R2), and constitutively active CREB (increasing CREB) produces the expected changes in the levels of the affected pathway components, lowering the levels in the case of inhibitors (VMAT and R2) or increasing them in the case of adding constitutively active CREB. Notice that in both A and B there are trajectories converging at 0 conversion level (normal).

these reagents alone causes some incidence of full conversion—the individual reagents have already been validated to work in *Xenopus* (i.e., they do penetrate the embryo, reach target cells, and affect this pathway sufficiently to fully convert some percentage of each treated sample)<sup>21</sup>; however it is of an all-or-none character. Moreover, we have previously shown that injection of just a few cells in the embryo (for example, with depolarizing reagents) is sufficient to cause complete conversion<sup>21,25</sup>—the partial conversion here is *not* due to mosaic expression of the CREB construct, as injected alone, in one cell of the 32-cell stage embryos as above, causes an incidence of ~45% complete conversion. The model predicted the wild-type phenotype within 1% (24% *in vivo* vs 25% *in silico*), yet it aggregated the partially and fully converted phenotypes observed *in vivo* (22% partial, 54% fully converted) into a single partial phenotype (75% partial). This difference in the frequency of partial and fully converted phenotypes between the *in silico* and *in vivo* results may be explained by the lack of time information in the training dataset and the observed dynamics of the system (Figs 3 and 4). In particular, a number of trajectories ending in the partially converted phenotype actually are crossing the levels of highly converted phenotypes. This observation may indicate that the full conversion that we observed *in vivo* may be transient. Further investigations, and methods that enabled the tadpoles to survive to older stages for tracking any subsequent changes in melanocyte status, would be necessary to validate predictions of transient (time-dependent) changes of the phenotypes predicted by the model. Indeed, the training dataset does not contain any time information, and hence the current system does not yet capture realistic timing, which would allow a finer-course analysis of trajectories and comparison with *in vivo* results; in future work we will extend this platform to include time-series data, so that these transient phenotypes can be better investigated.

Identification of interventions that alter system-level physiology or pattern formation is a crucial nascent field with many implications for biomedicine<sup>32,33</sup>. The complexity of biological regulation makes it very hard to completely and mechanistically understand signaling networks, the prediction of resultant experimental outcomes, and hence the discovery of novel phenotypes. Computational dynamic modeling of complex regulatory networks is essential for the discovery of new capabilities of biological systems and prediction of specific perturbations



**Figure 5.** *In vivo* validation of the computationally discovered partially converted phenotype. Normal melanocytes are indicated with blue arrowheads, while areas populated with converted melanocytes are indicated with red arrowheads. **(A)** Control animals exhibit small, round melanocytes confined to stripes around the spinal cord and over the brain **(A)**. Note the absence of melanocytes in the periocular region **(A')**. In normal animals' tails, melanocytes never exit the stripes of axial muscle (do not venture into the fin) **(A'')**. **(B)** Converted animals produced by Ivermectin-induced depolarization exhibit spread-out (arborized) melanocytes all along the flank **(B)** and around the eyes **(B')**—they colonize every region of the head. The arborized melanocytes also leave the mid-body and colonize the fin **(B'')**. **(C)** In contrast, embryos treated with VP16-XlCreb1 and exposed to 5HT-R2 and VMAT blockers revealed a never before-seen phenotype where some regions of the animal were converted **(C)**, periocular region of the right eye, red arrowhead), and some were wild-type **(C)**, periocular region of left eye, blue arrowhead). This can be also seen in the transformation of cell shape over the brain **(C')**, red arrowhead) but normal melanocyte morphology elsewhere **(C)**; blue arrowhead). In the tail **(C'')**, **(C''')** we observed some cells invading the fin, but they had normal morphology. We observed a great diversity of different combinations that were intermediate between normal and converted outcomes (only a representative sample is shown here), which otherwise never occurred in the same animal. **(D)** Furthermore, we noticed a new, cancer-like behavior that had not occurred in any previous experiment. This included the induction of individual nodules, which were both pigmented and unpigmented **(D, D')**; this occurred in ~10% of the animals that likewise exhibited partial conversion **(D'')**, red arrowhead region vs. blue arrowhead region).

that result in desired outcomes. Using these models, we can computationally and systematically interrogate the system to find the precise interventions necessary for the development of specific phenotypes<sup>33</sup>. The computational methodology presented here shows how a reverse-engineered signaling system can be used in this way to precisely find a set of reagents that produce a never before seen outcome at the intersection of embryogenesis and metastatic-like conversion. These computational methods are paving the way for the discovery and further understanding and manipulations of multicellular, stochastic, and multidimensional biological phenomena, as well as for the discovery of novel and effective therapies that target cooperative or group aspects of cell function *in vivo*.

## Methods

**Animal husbandry.** *Xenopus* embryos were maintained according to standard protocols<sup>34</sup> in 0.1 × Marc's Modified Ringers (MMR; pH 7.8). *Xenopus* embryos were staged according to Nieuwkoop and Faber<sup>35</sup>. All experimental procedures involving *Xenopus* embryos were approved by the Institutional Animal Care and Use Committees (IACUC) and Tufts University Department of Laboratory Animal Medicine under protocol

M2014-79. All experiments were performed in accordance with relevant guidelines and regulations. Although we have not observed light to make any difference to the converted phenotype, all groups were subjected to the same light conditions.

**Microinjection, drug exposure, sectioning, and image collection.** Embryos were injected with mRNA encoding constitutively active CREB (tom-VP16-XICreb1<sup>36</sup>) at 1/32 cell stage and then exposed to a mix of Altanserin (5HT-R2 blocker, 10  $\mu$ M) and Reserpine (VMAT blocker, 100  $\mu$ M) using the same methods described previously<sup>20,21,37,38</sup>. Sections were made at 40  $\mu$ m on a Leica vibratome after JB4 embedding. Images were collected using a Nikon SMZ-1500 microscope and each phenotype was visually characterized as having normal melanocyte status, partial conversion, or full conversion; these three classes are discrete and readily distinguished as categorical data; the quantification of the definition of these states was described in refs 20 and 39.

**Dynamic signaling model of melanocyte conversion.** We used the dynamic signaling model of *Xenopus* melanocyte conversion presented in ref. 21. The model is based on Hill-kinetics and includes 14 stochastic ordinary differential equations describing the interactions of signaling molecules, pharmacological compounds (drugs, set to 0 when not applied and 1 when applied), and the level of melanocyte conversion (the phenotype, equivalent to the pigmentation level) of a developing embryo. The interaction between signaling molecules and drugs are modeled with a combination of Hill functions, each of them representing the interaction between two products. When multiple interactions regulate a product, they can be grouped as sufficient, necessary, or a combination of them. Sufficient interactions can regulate the production of the target product independently of other sufficient interactions (modeled with a max operator, similar to a logic OR interaction; hence, activation has preference over inhibition). In contrast, all the necessary interactions need to be coordinated to regulate the production of the target product (modeled with a min operator, similar to a logic AND interaction; hence, inhibition has preference over activation). This scheme allows the modeling of a rich variety of regulatory interaction combinations. In addition, each product includes a decay term (with a specific decay constant) and a Gaussian random noise of zero mean to model the stochasticity of the system.

The following set of differential equations, including all the kinetic parameters, define the complete model ( $\xi(t)$  represents a Gaussian random noise of zero mean):

$$\begin{aligned} \frac{dHP}{dt} &= 0.53 \cdot \min\left(\frac{HP^{1.6}}{0.028^{1.6} + HP^{1.6}}, \max\left(\frac{CREB^{3.6}}{0.44^{3.6} + CREB^{3.6}}, \frac{0.28^{4.6}}{0.28^{4.6} + a^{4.6}}\right)\right) \\ &\quad - 0.57 \cdot HP + \xi(t) \\ \frac{d5HT}{dt} &= 0.65 \cdot \min\left(\frac{0.51^3}{0.51^3 + SERT^3}, \frac{VMAT^{1.7}}{0.23^{1.7} + VMAT^{1.7}}, \right. \\ &\quad \max\left(\frac{HP^{4.3}}{0.44^{4.3} + HP^{4.3}}, \frac{0.37^{4.5}}{0.37^{4.5} + R5^{4.5}}, \frac{MSH^{2.2}}{0.21^{2.2} + MSH^{2.2}}, \right. \\ &\quad \left. \left. \frac{Med5HT^{2.1}}{0.84^{2.1} + Med5HT^{2.1}}, \frac{b^{3.6}}{0.44^{3.6} + b^{3.6}}\right)\right) - 0.78 \cdot 5HT + \xi(t) \\ \frac{dSERT}{dt} &= 0.86 \cdot \min\left(\frac{IVM^{1.6}}{0.028^{1.6} + IVM^{1.6}}, \frac{0.24^{1.8}}{0.24^{1.8} + MSH^{1.8}}, \right. \\ &\quad \left. \frac{0.28^{3.3}}{0.28^{3.3} + Fluoxetine^{3.3}}, \frac{0.24^{2.2}}{0.24^{2.2} + R1^{2.2}}\right) - 0.43 \cdot SERT + \xi(t) \\ \frac{dR1}{dt} &= 0.34 \cdot \min\left(\frac{0.76^4}{0.76^4 + Cyanopindolol^4}, \frac{0.38^{4.3}}{0.38^{4.3} + MTP^{4.3}}, \right. \\ &\quad \left. \max\left(\frac{5HT^{1.9}}{0.27^{1.9} + 5HT^{1.9}}, \frac{cAMP^{1.4}}{0.21^{1.4} + cAMP^{1.4}}\right)\right) - 0.19 \cdot R1 + \xi(t) \\ \frac{dR2}{dt} &= 0.56 \cdot \min\left(\frac{0.29^{1.2}}{0.29^{1.2} + R1^{1.2}}, \frac{0.17^{4.9}}{0.17^{4.9} + MTP^{4.9}}, \right. \\ &\quad \left. \max\left(\frac{5HT^{3.7}}{0.0095^{3.7} + 5HT^{3.7}}, \frac{0.64^{3.6}}{0.64^{3.6} + Altanserin^{3.6}}\right)\right) - 0.25 \cdot R2 + \xi(t) \\ \frac{dR5}{dt} &= 0.43 \cdot \min\left(\frac{0.76^{2.7}}{0.76^{2.7} + SB699551^{2.7}}, \frac{0.27^{3.4}}{0.27^{3.4} + MTP^{3.4}}, \right. \\ &\quad \left. \max\left(\frac{5HT^{2.4}}{0.54^{2.4} + 5HT^{2.4}}, \frac{0.67^{4.1}}{0.67^{4.1} + SERT^{4.1}}, \frac{c^{3.7}}{0.0095^{3.7} + c^{3.7}}\right)\right) \\ &\quad - 0.23 \cdot R5 + \xi(t) \end{aligned}$$

$$\begin{aligned}
\frac{dMSH}{dt} &= 0.077 \cdot \min \left( \frac{SHU9119^{4.8}}{0.061^{4.8} + SHU9119^{4.8}}, \max \left( \frac{SERT^{3.4}}{0.88^{3.4} + SERT^{3.4}}, \right. \right. \\
&\quad \left. \left. \frac{R2^{2.6}}{0.69^{2.6} + R2^{2.6}}, \frac{R5^{1.7}}{0.43^{1.7} + R5^{1.7}}, \frac{0.28^{2.9}}{0.28^{2.9} + MRIF^{2.9}}, \right. \right. \\
&\quad \left. \left. \frac{0.65^{2.9}}{0.65^{2.9} + cAMP^{2.9}} \right) \right) - 0.68 \cdot MSH + \xi(t) \\
\frac{dVMAT}{dt} &= 0.74 \cdot \frac{0.23^{3.5}}{0.23^{3.5} + Reserpine^{3.5}} - 0.83 \cdot VMAT + \xi(t) \\
\frac{dcAMP}{dt} &= 0.48 \cdot \min \left( \frac{MSH^{4.4}}{0.00028^{4.4} + MSH^{4.4}}, \frac{Forskolin^{1.2}}{0.79^{1.2} + Forskolin^{1.2}}, \right. \\
&\quad \left. \frac{0.53^{2.5}}{0.53^{2.5} + 2'5' - DDA^{2.5}} \right) - 0.5 \cdot cAMP + \xi(t) \\
\frac{dCREB}{dt} &= 0.81 \cdot \min \left( \frac{0.65^5}{0.65^5 + R1^5}, \frac{0.24^{2.7}}{0.24^{2.7} + MSH^{2.7}}, \frac{0.027^{1.4}}{0.027^{1.4} + ACREB^{1.4}}, \right. \\
&\quad \frac{0.72^{4.3}}{0.72^{4.3} + a^{4.3}}, \max \left( \frac{CREB^{2.7}}{0.49^{2.7} + CREB^{2.7}}, \frac{PKA^{3.3}}{0.41^{3.3} + PKA^{3.3}}, \right. \\
&\quad \left. \left. \frac{VP16^{2.7}}{0.026^{2.7} + VP16^{2.7}} \right) \right) - 0.82 \cdot CREB + \xi(t) \\
\frac{dPKA}{dt} &= 0.31 \cdot \min \left( \frac{SERT^{4.1}}{0.31^{4.1} + SERT^{4.1}}, \frac{cAMP^{2.4}}{0.21^{2.4} + cAMP^{2.4}}, \frac{0.12^{1.1}}{0.12^{1.1} + H89^{1.1}} \right) \\
&\quad - 0.66 \cdot PKA + \xi(t) \\
\frac{da}{dt} &= 0.96 \cdot \min \left( \frac{0.13^{2.6}}{0.13^{2.6} + SERT^{2.6}}, \frac{0.15^{1.8}}{0.15^{1.8} + R1^{1.8}}, \frac{R2^{1.4}}{0.092^{1.4} + R2^{1.4}} \right) \\
&\quad - 1 \cdot a + \xi(t) \\
\frac{db}{dt} &= 0.041 \cdot \max \left( \frac{0.27^{2.8}}{0.27^{2.8} + R2^{2.8}}, \frac{0.14^{4.4}}{0.14^{4.4} + b^{4.4}} \right) - 0.31 \cdot b + \xi(t) \\
\frac{dc}{dt} &= 0.25 - 0.48 \cdot c + \xi(t)
\end{aligned} \tag{1}$$

This set of equations (defined in the model) together with a set of initial conditions of the signaling products (defined in the model) and the drugs (defined in the experiment) form an initial value problem which is then numerically solved until reaching a steady state. Due to the stochastic nature of the equations, different runs with the same initial conditions (same experiment) can result in different phenotypic outcomes (level of melanocyte conversion) as observed *in vivo*. The model was reverse-engineered directly from a dataset of 20 experiments combining different pharmacological drugs and resulting in a specific frequency of two different and extreme phenotypes (within each phenotype, the animal-to-animal variation in pigmentation level was not visually detectable): either normal (pigmentation level 0) or high level (pigmentation level 1) of melanocyte conversion, never an intermediate level as described in ref. 21. The model was further validated with an additional dataset of 7 experiments not used during the training<sup>21</sup>. The model kinetic constants, also inferred from the experimental data, are in arbitrary units.

**Phenotypic distance metric.** To discover phenotypes different from the two extreme levels of conversion (normal or high) previously observed *in silico* with the model and *in vivo* with *Xenopus* experiments, we defined a distance metric for simulated phenotypes obtained with the model. The metric quantifies the level of intermediate conversion obtained in a set of resultant *in silico* phenotypes simulated with the model for a specific experiment (combination of reagents). The distance metric is defined with the following equation:

$$d(E) = \frac{1}{100} \sum_{i=1}^{100} 0.5 - |0.5 - p_i| \tag{2}$$

where  $d(E)$  is the mean distance of experiment  $E$  (a combination of reagents) when simulated 100 times with the model and resulting in phenotypes with level of conversion  $p_i$  (the level of conversion is a dynamic variable in the model, with values from 0 for the wild type to 1 for a total conversion).

## References

- King, R. D. *et al.* The automation of science. *Science* **324**, 85–89, doi: 10.1126/science.1165620 (2009).
- Soldatova, L. N., Clare, A., Sparkes, A. & King, R. D. An ontology for a Robot Scientist. *Bioinformatics* **22**, e464–471, doi: 10.1093/bioinformatics/btl207 (2006).
- Reinitz, J., Mjolsness, E. & Sharp, D. H. Model for Cooperative Control of Positional Information in *Drosophila* by Bicoid and Maternal Hunchback. *Journal of Experimental Zoology* **271**, 47–56, doi: 10.1002/jez.1402710106 (1995).
- Reinitz, J., Kosman, D., Vanario-Alonso, C. E. & Sharp, D. H. Stripe forming architecture of the gap gene system. *Dev Genet* **23**, 11–27, doi: 10.1002/(Sici)1520-6408(1998)23:1<11::Aid-Dvg2>3.0.Co;2-9 (1998).
- Jaeger, J. *et al.* Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**, 368–371, doi: 10.1038/Nature02678 (2004).
- Perkins, T. J., Jaeger, J., Reinitz, J. & Glass, L. Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Computational Biology* **2**, 417–428, doi: 10.1371/journal.pcbi.0020051 (2006).
- Manu *et al.* Canalization of gene expression and domain shifts in the *Drosophila* blastoderm by dynamical attractors. *PLoS Computational Biology* **5**, e1000303, doi: 10.1371/journal.pcbi.1000303 (2009).
- Crombach, A., Wotton, K. R., Cicin-Sain, D., Ashyraliyev, M. & Jaeger, J. Efficient reverse-engineering of a developmental gene regulatory network. *PLoS Computational Biology* **8**, e1002589, doi: 10.1371/journal.pcbi.1002589 (2012).
- Becker, K. *et al.* Reverse-Engineering Post-Transcriptional Regulation of Gap Genes in *Drosophila melanogaster*. *PLoS Computational Biology* **9**, e1003281, doi: 10.1371/journal.pcbi.1003281 (2013).
- Lobo, D. & Levin, M. Inferring Regulatory Networks from Experimental Morphological Phenotypes: A Computational Method Reverse-Engineers Planarian Regeneration. *PLoS computational biology* **11**, e1004295, doi: 10.1371/journal.pcbi.1004295 (2015).
- Lobo, D., Morokuma, J. & Levin, M. Computational discovery and *in vivo* validation of *hnf4* as a regulatory gene in planarian regeneration. *Bioinformatics* **32**, 2681–2685, doi: 10.1093/bioinformatics/btw299 (2016).
- Levin, M. Reprogramming cells and tissue patterning via bioelectrical pathways: molecular mechanisms and biomedical opportunities. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **5**, 657–676, doi: 10.1002/wsbm.1236 (2013).
- Levin, M. Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. *The Journal of Physiology* **592**, 2295–2305, doi: 10.1113/jphysiol.2014.271940 (2014).
- Levin, M. Molecular bioelectricity: how endogenous voltage potentials control cell behavior and instruct pattern regulation *in vivo*. *Mol Biol Cell* **25**, 3835–3850, doi: 10.1091/mbc.E13-12-0708 (2014).
- Levin, M. Molecular bioelectricity in developmental biology: new tools and recent discoveries: control of cell behavior and pattern formation by transmembrane potential gradients. *BioEssays* **34**, 205–217, doi: 10.1002/bies.201100136 (2012).
- Bates, E. Ion Channels in Development and Cancer. *Annu Rev Cell Dev Biol* **31**, 231–247, doi: 10.1146/annurev-cellbio-100814-125338 (2015).
- Lang, F. & Stournaras, C. Ion channels in cancer: future perspectives and clinical potential. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130108, doi: 10.1098/rstb.2013.0108 (2014).
- Yang, M. & Brackenbury, W. J. Membrane potential and cancer progression. *Front Physiol* **4**, 185, doi: 10.3389/fphys.2013.00185 (2013).
- Arcangeli, A. *et al.* Targeting ion channels in cancer: a novel frontier in antineoplastic therapy. *Current medicinal chemistry* **16**, 66–93 (2009).
- Blackiston, D., Adams, D. S., Lemire, J. M., Lobikin, M. & Levin, M. Transmembrane potential of GlyCl-expressing instructor cells induces a neoplastic-like conversion of melanocytes via a serotonergic pathway. *Disease models & mechanisms* **4**, 67–85, doi: 10.1242/dmm.005561 (2011).
- Lobikin, M. *et al.* Serotonergic regulation of melanocyte conversion: A bioelectrically regulated network for stochastic all-or-none hyperpigmentation. *Sci Signal* **8**, ra99, doi: 10.1126/scisignal.aac6609 (2015).
- Roubos, E. W. Background adaptation by *Xenopus laevis*: a model for studying neuronal information processing in the pituitary pars intermedia. *Comp. Biochem. Physiol. A. Physiol.* **118**, 533–550 (1997).
- Daniolos, A., Lerner, A. B. & Lerner, M. R. Action of light on frog pigment cells in culture. *Pigment Cell Res* **3**, 38–43 (1990).
- Dulcis, D. & Spitzer, N. C. Illumination controls differentiation of dopamine neurons regulating behaviour. *Nature* **456**, 195–201, doi: 10.1038/nature07569 (2008).
- Lobikin, M., Chernet, B., Lobo, D. & Levin, M. Resting potential, oncogene-induced tumorigenesis, and metastasis: the bioelectric basis of cancer *in vivo*. *Physical biology* **9**, 065002, doi: 10.1088/1478-3975/9/6/065002 (2012).
- Chernet, B. & Levin, M. Endogenous Voltage Potentials and the Microenvironment: Bioelectric Signals that Reveal, Induce and Normalize Cancer. *J Clin Exp Oncol Suppl* **1**, doi: 10.4172/2324-9110.S1-002 (2013).
- Blackiston, D. J., Anderson, G. M., Rahman, N., Bieck, C. & Levin, M. A novel method for inducing nerve growth via modulation of host resting potential: gap junction-mediated and serotonergic signaling mechanisms. *Neurotherapeutics* **12**, 170–184, doi: 10.1007/s13311-014-0317-7 (2015).
- Liu, Z. P. Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data. *Curr Genomics* **16**, 3–22 (2015).
- Uzkudun, M., Marcon, L. & Sharpe, J. Data-driven modelling of a gene regulatory network for cell fate decisions in the growing limb bud. *Mol Syst Biol* **11** (2015).
- Villaverde, A. F. & Banga, J. R. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *Journal of the Royal Society Interface* **11**, doi: 10.1098/rsif.2013.0505 (2014).
- Lobo, D. & Levin, M. Computing a worm: reverse-engineering planarian regeneration. In *Advances in Unconventional Computing: Volume 2: Prototypes, Models and Algorithms* (ed Andrew Adamatzky) 637–654 (Springer International Publishing, 2016).
- Marguet, S. L. *et al.* Treatment during a vulnerable developmental period rescues a genetic epilepsy. *Nat Med* **21**, 1436–1444, doi: 10.1038/nm.3987 (2015).
- Smolen, P., Baxter, D. A. & Byrne, J. H. Simulations suggest pharmacological methods for rescuing long-term potentiation. *J Theor Biol* **360**, 243–250, doi: 10.1016/j.jtbi.2014.07.006 (2014).
- Sive, H. L., Grainger, R. M. & Harland, R. M. *Early development of Xenopus laevis: a laboratory manual* (Cold Spring Harbor Laboratory Press, 2000).
- Nieuwkoop, P. & Faber, J. *Normal table of Xenopus laevis (Daudin): a systematic and chronological survey of the development from the fertilized egg till the end of metamorphosis* (North-Holland, 1967).
- Barco, A., Alarcon, J. M. & Kandel, E. R. Expression of constitutively active CREB protein facilitates the late phase of long-term potentiation by enhancing synaptic capture. *Cell* **108**, 689–703, doi: 10.1016/S0092-8674(02)00657-8 (2002).
- Chernet, B. T., Adams, D. S., Lobikin, M. & Levin, M. Use of genetically encoded, light-gated ion translocators to control tumorigenesis. *Oncotarget*, doi: 10.18632/oncotarget.8036 (2016).
- Lobikin, M., Pare, J. F., Kaplan, D. L. & Levin, M. Selective depolarization of transmembrane potential alters muscle patterning and muscle cell localization in *Xenopus laevis* embryos. *The International journal of developmental biology*, doi: 10.1387/ijdb.150198ml (2015).
- Morokuma, J. *et al.* Modulation of potassium channel function confers a hyperproliferative invasive phenotype on embryonic stem cells. *Proc Natl Acad Sci USA* **105**, 16608–16613 (2008).
- Hendrix, M. J., Seftor, E. A., Hess, A. R. & Seftor, R. E. Vasculogenic mimicry and tumour-cell plasticity: lessons from melanoma. *Nature reviews. Cancer* **3**, 411–421, doi: 10.1038/nrc1092 (2003).

## Acknowledgements

This work has been supported by an Allen Discovery Center award from the Paul G. Allen Frontiers Group (12171), NSF grant EF-1124651, and The G. Harold and Leila Y. Mathers Charitable Foundation (TFU141). Computation used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant ACI-1053575, and a cluster computer awarded by Silicon Mechanics.

## Author Contributions

M. Levin prepared Figures 1 and 5; D. Lobo prepared the other figures and was responsible for the computational portion of this manuscript. M. Lobikin was responsible for all experiments involving *Xenopus* embryos. All authors wrote the manuscript together.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Lobo, D. *et al.* Discovering novel phenotypes with automatically inferred dynamic models: a partial melanocyte conversion in *Xenopus*. *Sci. Rep.* 7, 41339; doi: 10.1038/srep41339 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017