Biology Direct

CrossMark

Modeling epigenetic regulation of PRC1 protein accumulation in the cell cycle

Marzena Dolbniak[1], Marek Kimmel[2,1*] and Jaroslaw Smieja[1]

Abstract

Background: Epigenetic regulation contributes to many important processes in biological cells. Examples include developmental processes, differentiation and maturation of stem cells, evolution of malignancy and other. Cell cycle regulation has been subject of mathematical modeling by a number of authors that resulted in many interesting models and application of analytic techniques ranging from stochastic processes to partial differential equations and to integral, functional and operator equations. In this paper we address the question of how the regulation of protein contents influences the long-term dynamics of the population. To accomplish this, we follow the philosophy of a 1984 model by Kimmel et al., but adjust the details to fit the experimental data on protein PRC1 from a more recent paper.

Results: We built a model of cell cycle dynamics of the PRC1 and fitted it to the data made available by Cohen and his co-authors. We have run the model for a large number of cell generations, recording the PRC1 contents in all cells of the resulting pedigree, at constant time intervals. During cell division the PRC1 is unequally divided between daughter cells. The picture emerging from simulations of Data set 1 is that of a very well-tuned regulatory circuit that provides a stable distribution of PRC1 contents and interdivision times. Data set 2 seems qualitatively different, with more variation in cell cycle duration.

Conclusions: The main question we address is whether the regulatory feedbacks deduced from single cell cycle data provide epigenetic regulation of cell characteristics in long run. PRC1 is a good candidate because of its role in setting timing of division. Findings of the current paper include tight regulation of the cell cycle (particularly the timing of the cell cycle) even that PRC1 is only one of the players in cell dynamics. Understanding that association, even close, does not necessarily imply causation, we consider this an interesting and important result.

Reviewers: This article was reviewed by Ollivier Hyrien, Anna Marciniak-Czochra and Alberto d'Onofrio.

Keywords: Cell cycle, Mathematical model, PRC1 protein, Dynamics, Stochastic fluctuations, Asymmetric division

Background

Epigenetic regulation contributes to many important processes in biological cells. Examples include developmental processes, differentiation and maturation of stem cells, evolution of malignancy and other [1]. One of the processes, which have been studied for at least several decades, is regulation of cell size and cell cycle duration. More specifically, how the dynamics of protein production and the manner in which proteins are split between the two progeny cells leads to preservation of cell population age and size structure (homeostasis). A related question is under what circumstances these dynamics lead to phenomena such as bimodality and, as a consequence, separation of distinct cell subpopulations.

Cell cycle regulation has been subject of mathematical modeling by a number of authors which resulted in many interesting models and application of analytic techniques ranging from stochastic processes to partial differential equations and to integral, functional and operator equations [2, 3]. Some of the models have been applied to data on bacterial and eukaryotic cells, also in the context of cancer modeling [4].

An example of an early model devised to capture cell cycle regulation and unequal division of mass among progeny cells, is the model by Kimmel et al. [5] (Fig. 1), which

* Correspondence: kimmel@rice.edu
[2]Departments of Statistics and Bioengineering, Rice University, MS 138, 6100 Main, Houston, TX 77005, USA
[1]Systems Engineering Group, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 2 of 15

considers the dynamics of the distribution of the total mass of cell RNA in a growing cell population. In that model, the birth-mass of a cell, represented by random variable (rv) X_0, determines both the mass at division, X_2, and the time, T, to division (cell-cycle duration).

$$X_2 = \phi(X_0), \quad T = \psi(X_0) \tag{1}$$

At division, the parent-cell mass is randomly split between the two progeny cells, according to the expression

$$X_0' = UX_2, \quad X_0'' = (1-U)X_2 \tag{2}$$

in which the rv U is independent of rv X_2, and it is distributed symmetrically over the interval (0, 1), so that $E(U) = 1/2$. Uneven partition of mass among progeny cells is the only source of randomness in the basic model. A more general version of the model retains the deterministic mass growth, but includes stochastic time to division [6]. In other models, the growth of mass is deterministic or
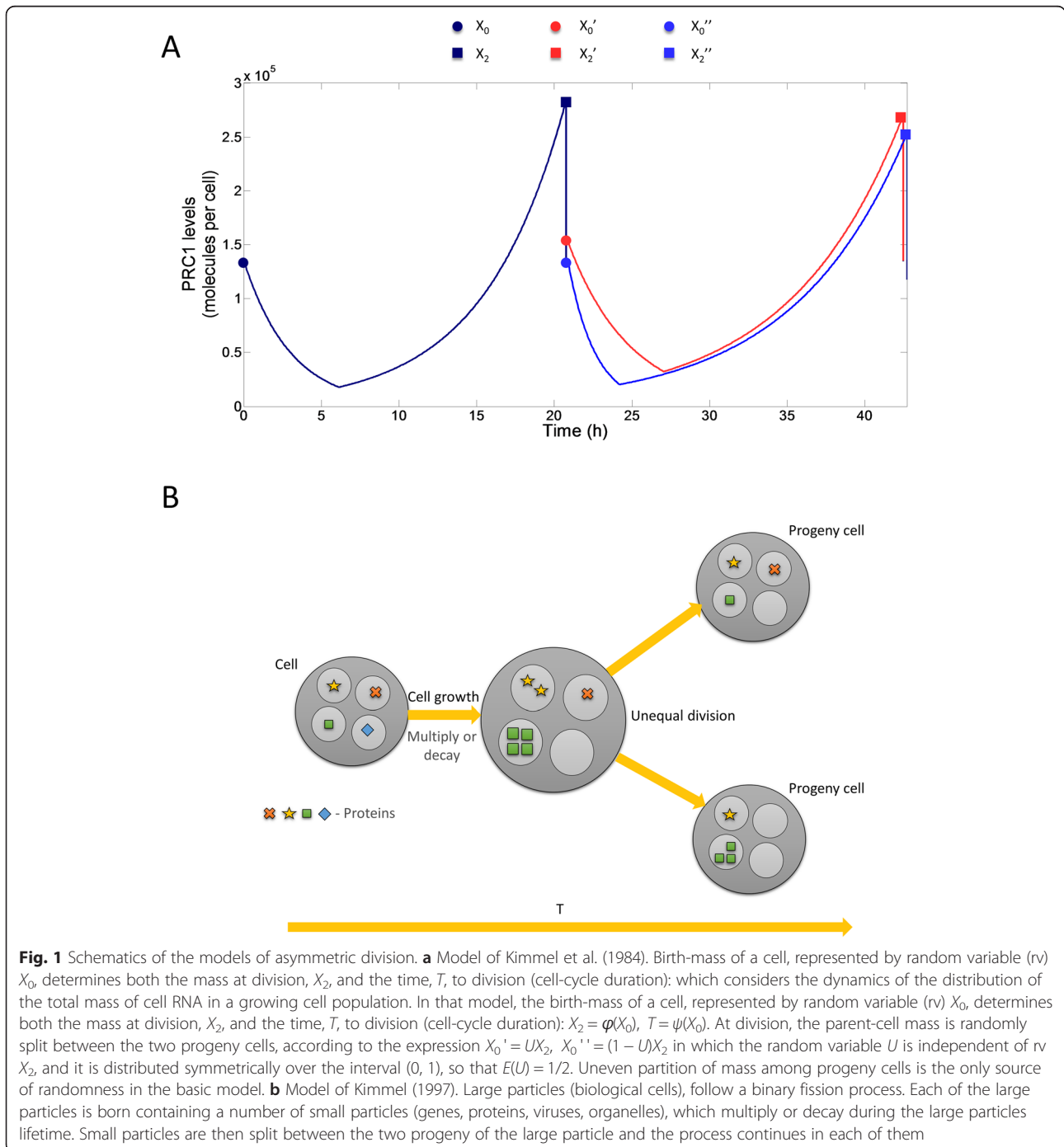

Fig. 1 Schematics of the models of asymmetric division. **a** Model of Kimmel et al. (1984). Birth-mass of a cell, represented by random variable (rv) X_0, determines both the mass at division, X_2, and the time, T, to division (cell-cycle duration): which considers the dynamics of the distribution of the total mass of cell RNA in a growing cell population. In that model, the birth-mass of a cell, represented by random variable (rv) X_0, determines both the mass at division, X_2, and the time, T, to division (cell-cycle duration): $X_2 = \varphi(X_0)$, $T = \psi(X_0)$. At division, the parent-cell mass is randomly split between the two progeny cells, according to the expression $X_0' = UX_2$, $X_0'' = (1 - U)X_2$ in which the random variable U is independent of rv X_2, and it is distributed symmetrically over the interval (0, 1), so that $E(U) = 1/2$. Uneven partition of mass among progeny cells is the only source of randomness in the basic model. **b** Model of Kimmel (1997). Large particles (biological cells), follow a binary fission process. Each of the large particles is born containing a number of small particles (genes, proteins, viruses, organelles), which multiply or decay during the large particles lifetime. Small particles are then split between the two progeny of the large particle and the process continues in each of them

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 3 of 15

stochastic and division occurs when a randomly assigned mass threshold is reached [3]. The model leads to stable exponential growth, a process in which the numbers of cells in all possible subsets of (X_0, X_2, T) values grow exponentially at a rate defined by a Malthusian parameter λ. Please see the Conclusions Section for further remarks.

Cell-to-cell differences are present in any cell population. The sources of variation in population include the extrinsic and intrinsic noise and are well characterized for many cell types (e.g., [7, 8]). Non-genetic intrinsic heterogeneity stems from the random (thermal) nature of interaction of individual molecules, such as mRNA and proteins. Since some of these biomolecules are present in a relatively small number in a cell, their stochastic fluctuations are, unlike in classical test-tube chemistry, not averaged out [1]. As stated above, an equally important source of heterogeneity is the unequal distribution of cellular mRNA and proteins between two daughter cells after cell division.

In this paper we address the question of how the regulation of protein contents influences the long-term dynamics of the population. To accomplish this, we follow the philosophy of Kimmel et al. model [5], but adjust the details to fit the experimental data on protein PRC1 from the paper by Cohen et al. [9]. PRC1 protein is expressed at relatively high levels during S and G2/M phases of the cell cycle before dropping dramatically after mitotic exit and entrance into G1 phase. PRC1 is a substrate of several cyclin-dependent kinases (CDKs) and it has become a novel human protein of cytokinetic importance since its identification [10]. PRC1 takes part in midzone microtubule formation and is essential to the cytokinetic machinery of mammals, via collaboration with Kinesin-4 in setting up a controlled zone of overlapping, antiparallel microtubules at the spindle midzone [11]. Upon anaphase onset and removal of inhibitory CDK1 phosphorylation, PRC1 dimers form, which recruit Kinesin-4, a plus-end directed motor protein that inhibits microtubule dynamics, helps stabilize and regulate spindle microtubule assembly within cytokinesis. The PRC1-Kinesin-4 complex identifies and regulates the spindle midzone microtubules during cell division, which is crucial in order for cytokinesis to progress properly. Our model assumes that PRC1 dynamics contributes to determination of the duration of the cell cycle, with influence of other factors represented as noise.

Further information concerning the role of PRC1 is found in the papers [12–16]: The role of PRC1 in cancer has been considered in references [12, 17] and the role in radiation resistance and stemness in reference [18].

Methods

Hypotheses

As already mentioned, the model we use is patterned after the model by Kimmel et al. [5]. Dynamics of PRC1 during a single cell cycle are separated into two phases:

degradation and accumulation. The following variables are used to describe the two phases:

X_0 – the number of PRC1 molecules at the beginning of the cell cycle,

X_1 – minimum number of molecules, at the end of the degradation phase and the beginning of the accumulation phase,

X_2 – number of molecules at the end of the cell cycle,

T_1 - duration of the degradation phase,

T_2 – duration of the accumulation phase,

a – protein degradation rate, and

b – protein production rate.

We assume that in the degradation phase the dynamics of the protein are described as exponential decay, while in the accumulation phase they are described as exponential growth

$$X_1 = X_0 \exp(-aT_1), \quad X_2 = X_1 \exp(bT_2) \qquad (3)$$

We assume that following cell division, the protein is split unequally among the two progeny cells according to expressions (2). In the balanced exponential growth, X_0, X_2, and U are distributed identically in each cell and so we have $X_0 = UX_2$ which by independence of X_2 and U implies

$$\begin{aligned} V(X_0) &= E(X_2^2)E(U^2) - E(X_2)^2 E(U)^2 \\ &= V(X_2)E(U^2) + E(X_2)^2 V(U) \end{aligned}$$

$$\frac{V(X_0)}{E(X_0)^2} = \frac{V(X_2)E(U^2)}{E(X_0)^2} + \frac{E(X_2)^2 V(U)}{E(X_0)^2}$$

Considering that because of symmetry $E(U) = 1/2$ and $E(X_0) = E(X_2)/2$ we obtain

$$\frac{V(X_0)}{E(X_0)^2} = 4\frac{V(X_2)E(U^2)}{E(X_2)^2} + 4V(U)$$

and passing to the coefficients of variation ($cv_X = \sqrt{V(X)}/E(X)$ for rv X), we obtain

$$cv_{X_0}^2 = cv_U^2\left(cv_{X_2}^2 + 1\right) + cv_{X_2}^2$$

Solving the above for cv_U^2 results in

$$cv_U^2 = \frac{cv_{X_0}^2 - cv_{X_2}^2}{cv_{X_2}^2 + 1} \qquad (4)$$

Data-based model building

We have at our disposal the four PRC1 data sets available from the Supplemental Data to reference [9]. Each of the data sets consists of individual-cell measurements of PRC1 contents collected at constant time intervals. The trajectories are depicted in Fig. 2, the legend of which contains the relevant details. Briefly, after division, the level of PRC1 decreases and then increases, to reach a maximum value immediately before division.

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 4 of 15

Fig. 2 Empirical and simulated trajectories of the PRC1 protein content for the 4 data sets from reference [9]. **a** *Data set 1*. Published in the main body of [9] The composite picture includes illustration of outlier removal, empirical trajectories excluding outliers, and simulated trajectories. **b** *Data set 2*. We use it for comparison with Data set 1. (**c** and **d**) *Data sets 3 and 4*

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 5 of 15

Fig. 2 (continued)

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 6 of 15

Detailed model description and estimation

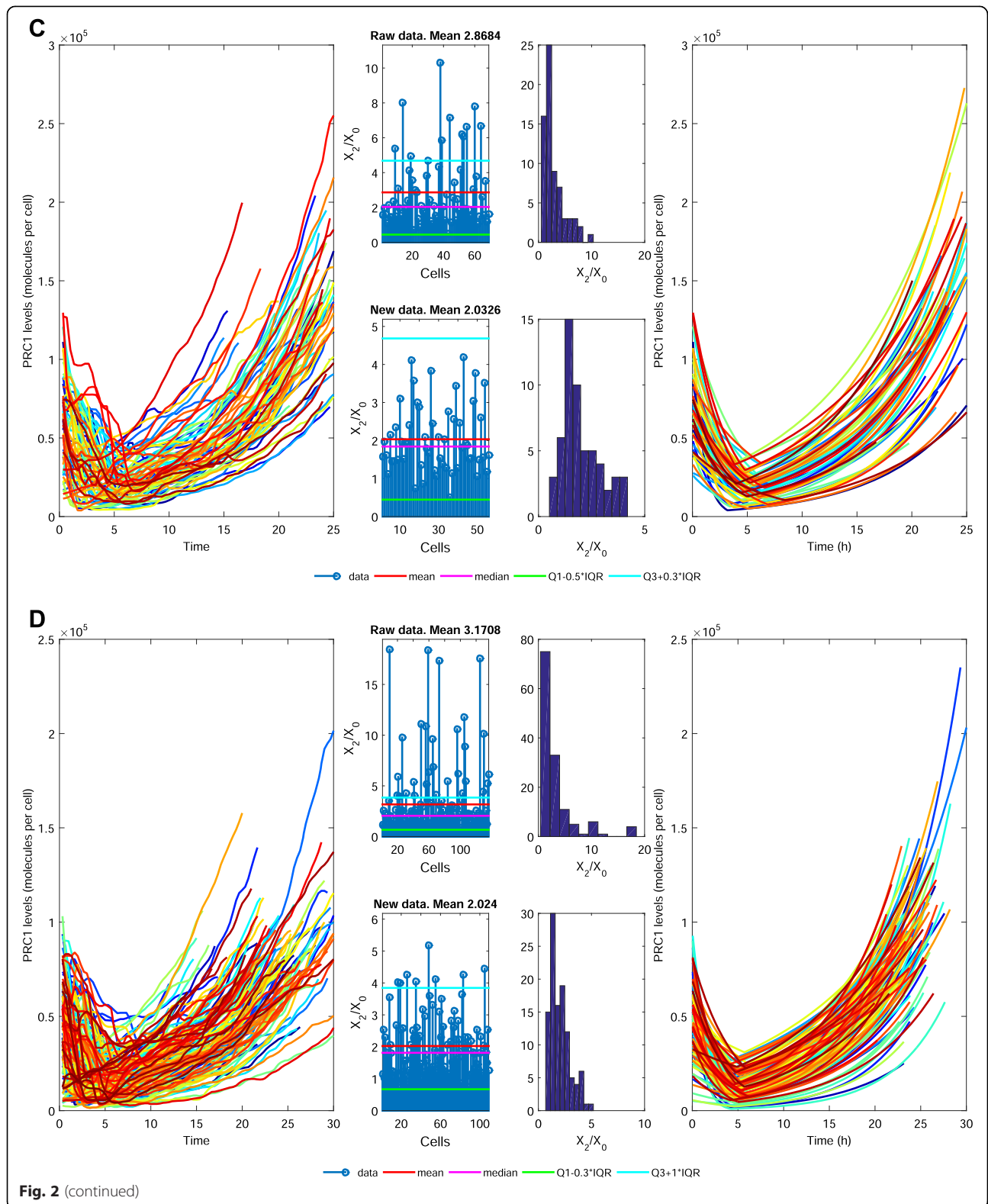Based on data analysis, the model entails the following detailed principles:

1. $\ln(a)$ depends linearly on X_0 with additive Gaussian noise,
2. $\ln(T_1)$ (T_1 in case of Data set 2) depends linearly on $\ln(a)$ with additive Gaussian noise,
3.

   $$X_1 = X_0 \exp(-aT_1),$$

4. $\ln(b)$ depends linearly on $\ln(a)$ and X_1 with additive Gaussian noise,
5. T_2 depends linearly on T_1 with additive Gaussian noise,
6.

   $$X_2 = X_1 \exp(bT_2),$$

7. $X_0' = UX_2$, $X_0'' = (1 - U)X_2$, i.e., in multi-generation simulations, the next-generation starting protein contents are modeled using Eq. 2, where the distribution of the random variable U is assumed to belong to the Beta-family.

Table 1 depicts the correlations computed from the data. They indicate a strong positive correlation between $\ln(a)$ and X_0, which supports item 1 above. Similarly, there exists a strong negative correlation between $\ln(T_1)$ and $\ln(a)$, which supports item 2. For Data set 2, the correlation is slightly better for T_1 and $\ln(a)$. Then, X_1 can be computed from the exponential decay expression as in item 3. Further, there exists a strong positive correlation between $\ln(b)$ and $\ln(a)$, and a strong negative correlation between $\ln(b)$ and X_1, which supports item 4. Finally, there exists a strong negative correlation between T_1 and T_2, which supports item 5. Then, X_2 can be computed from the exponential growth expression as in item 6.

Detection and removal of outliers in the data

The first step was finding and removing the outlier trajectories of PRC1 protein. In all 4 data sets the individual measurements differed with respect to the dynamic of this protein. We decided that the most important is how much the amount of PRC1 increases during cell cycle. When the X_2/X_0, ratio was calculated, in raw data in some cells the number of protein molecules at the end of the cell cycle was up to fifty times higher than at the beginning, which is biologically unlikely. We use a modification of Tukey test to eliminate the outliers [19]. Figure 2 shows the X_2/X_0 ratio and distributions before and after removal of outliers.

We believe that outliers result from measurement errors. The beginning or the end of the cell cycle might have been identified incorrectly.

Table 1 Data-derived correlations between pairs of variables characterizing the PRC1 trajectories in Data sets 1–4

Data set 1

	X_0	X_1	X_2	T_1	T_2	$T_1 + T_2$	a	ln(a)	b
X_1	−0.02								
X_2	0.30	0.11							
T_1	−0.08	0.25	−0.05						
T_2	0.05	−0.25	0.15	−0.62					
$T_1 + T_2$	−0.03	−0.02	0.13	0.39	0.48				
a	0.24	−0.58	0.06	−0.75	0.57	−0.17			
ln(a)	0.31	−0.71	0.05	−0.76	0.55	−0.21	0.93		
b	0.05	−0.71	0.18	0.11	−0.29	−0.22	0.29	0.38	
ln(b)	0.05	−0.75	0.19	0.10	−0.27	−0.21	0.30	0.41	0.98

Data set 2

	X_0	X_1	X_2	T_1	T_2	$T_1 + T_2$	a	ln(a)	b
X_1	0.06								
X_2	0.35	0.19							
T_1	0.22	0.13	0.00						
T_2	0.07	−0.22	0.32	−0.42					
$T_1 + T_2$	0.26	−0.11	0.31	0.45	0.62				
a	−0.03	−0.63	−0.04	−0.59	0.24	−0.27			
ln(a)	0.10	−0.77	−0.01	−0.66	0.43	−0.12	0.87		
b	−0.02	−0.78	0.11	−0.04	−0.12	−0.14	0.63	0.60	
ln(b)	−0.01	−0.81	0.12	0.02	−0.12	−0.10	0.60	0.59	0.99

Data set 3

	X_0	X_1	X_2	T_1	T_2	$T_1 + T_2$	a	ln(a)	b
X_1	0.06								
X_2	0.26	0.29							
T_1	0.23	0.12	0.02						
T_2	0.09	−0.19	0.33	−0.34					
$T_1 + T_2$	0.27	−0.08	0.32	0.47	0.67				
a	0.03	−0.63	−0.10	−0.55	0.22	−0.23			
ln(a)	0.17	−0.76	−0.11	−0.61	0.38	−0.10	0.88		
b	−0.04	−0.77	0.04	−0.06	−0.15	−0.19	0.62	0.59	
ln(b)	−0.03	−0.79	0.06	0.00	−0.15	−0.14	0.58	0.57	0.99

Data set 4

	X_0	X_1	X_2	T_1	T_2	$T_1 + T_2$	a	ln(a)	b
X_1	0.21								
X_2	0.31	0.18							
T_1	0.31	0.19	0.06						
T_2	0.26	−0.39	0.50	−0.11					
$T_1 + T_2$	0.41	−0.25	0.48	0.44	0.84				
a	0.17	−0.58	0.03	−0.40	0.41	0.16			
ln(a)	0.23	−0.66	0.01	−0.36	0.48	0.27	0.84		
b	−0.22	−0.73	−0.07	−0.03	−0.04	−0.06	0.49	0.46	
ln(b)	−0.25	−0.79	−0.06	−0.06	0.00	−0.03	0.48	0.46	0.97

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 7 of 15

Estimation of model parameters

Following determination of the structure of the model, estimation of the coefficients of the linear relationships and the variances of noise has been accomplished using standard regression techniques [19]. Final relationships for Data set 1 are presented in Table 2. Table 3 presents the simulation-based counterparts of experimental correlations from Table 1 (Data set 1). Each cell has different values of all parameters.

Comparison of the scatterplots of experimental vs. model-based relationships among model variables is depicted in Fig. 3 (Data sets 1–4). Comparison demonstrates a very good agreement of experiment and model-based data, especially for Data set 1, which was used to construct the model.

We have not observed many cases where calculated parameter was negative. Nevertheless, when such a case

Table 2 Regression-based coefficients for the 4 versions of the model based on Data sets 1–4, respectively

Data set 1

	1	X_0	$\ln(a)$	T_1	X_1	ε
$\ln(a)$	−1.600	3.9×10^{-6}	-	-	-	$N(0, 0.62)$
$\ln(T_1)$	1.110	-	−0.514	-	-	$N(0, 0.124)$
$\ln(b)$	−1.570	-	−0.100	-	$−1.78 \times 10^{-5}$	$N(0, 0.07)$
T_2	19.970	-	-	−0.6564	-	$N(0, 1.62)$

Data set 2

	1	X_0	$\ln(a)$	T_1	X_1	ε
$\ln(a)$	−3.37	7.82×10^{-5}	-	-	-	$N(0,0.45)$
T_1	4.81	-	0.0465	-	-	$N(0,2.00)$
$\ln(b)$	−1.7	-	−0.0157	-	$−9.89 \times 10^{-5}$	$N(0,0.08)$
T_2	18.93	-	-	−0.75	-	$N(0,2.54)$

Data set 3

	1	X_0	$\ln(a)$	T_1	X_1	ε
$\ln(a)$	−1.75	5.01×10^{-6}	-	-	-	$N(0,0.35)$
$\ln(T_1)$	1.18	-	−0.349	-	-	$N(0,0.15)$
$\ln(b)$	−1.8	-	−0.0151	-	$−2.12 \times 10^{-5}$	$N(0,0.09)$
T_2	19.91	-	-	−0.40	-	$N(0,2.53)$

Data set 4

	1	X_0	$\ln(a)$	T_1	X_1	ε
$\ln(a)$	−2.10	1.39×10^{-5}	-	-	-	$N(0,0.485)$
$\ln(T_1)$	1.41	-	−0.1123	-	-	$N(0,0.11)$
$\ln(b)$	−1.9	-	−0.0437	-	$−3.16 \times 10^{-5}$	$N(0,0.11)$
T_2	20.90	-	-	−0.37	-	$N(0,2.16)$

Table 3 Simulation-based correlations between pairs of variables characterizing the PRC1 trajectories in Data sets 1–4

Data set 1

	X_0	X_1	X_2	T_1	T_2	T_1+T_2	a	$\ln(a)$	b
X_1	0.23								
X_2	−0.04	0.08							
T_1	−0.25	0.55	−0.06						
T_2	0.16	−0.34	0.56	−0.62					
T_1+T_2	−0.08	0.20	0.60	0.37	0.60				
a	0.34	−0.63	−0.20	−0.75	−0.20	−0.29			
$\ln(a)$	0.30	−0.74	−0.09	−0.91	−0.09	−0.33	0.91		
b	−0.34	−0.90	0.09	−0.36	0.09	−0.13	0.46	0.54	
$\ln(b)$	−0.36	−0.91	0.10	−0.35	0.10	−0.13	0.44	0.53	0.99

Data set 2

	X_0	X_1	X_2	T_1	T_2	T_1+T_2	a	$\ln(a)$	b
X_1	−0.03								
X_2	−0.28	0.33							
T_1	0.04	−0.58	−0.42						
T_2	−0.04	0.27	0.79	−0.48					
T_1+T_2	−0.02	−0.12	0.56	0.21	0.56				
a	0.75	−0.45	−0.34	0.04	−0.34	−0.01			
$\ln(a)$	0.77	−0.40	−0.29	0.06	−0.29	0.00	0.89		
b	0.12	−0.94	−0.31	0.54	−0.31	0.11	0.53	0.45	
$\ln(b)$	0.00	−0.97	−0.25	0.56	−0.25	0.12	0.42	0.36	0.97

Data set 3

	X_0	X_1	X_2	T_1	T_2	T_1+T_2	a	$\ln(a)$	b
X_1	0.57								
X_2	0.16	0.25							
T_1	−0.19	−0.03	−0.11						
T_2	0.03	0.01	0.75	−0.17					
T_1+T_2	−0.05	0.00	0.68	0.25	0.68				
a	0.31	−0.43	−0.09	−0.60	−0.09	−0.16			
$\ln(a)$	0.31	−0.43	−0.07	−0.63	−0.07	−0.17	0.97		
b	−0.50	−0.86	0.01	0.04	0.01	0.01	0.38	0.37	
$\ln(b)$	−0.51	−0.88	0.01	0.04	0.01	0.01	0.37	0.37	0.99

Data set 4

	X_0	X_1	X_2	T_1	T_2	T_1+T_2	a	$\ln(a)$	b
X_1	0.33								
X_2	0.01	0.31							
T_1	−0.20	0.10	0.06						
T_2	0.01	−0.01	0.57	−0.10					
T_1+T_2	−0.05	0.02	0.58	0.19	0.58				
a	0.43	−0.59	−0.35	−0.44	−0.35	−0.10			
$\ln(a)$	0.45	−0.60	−0.29	−0.48	−0.29	−0.10	0.93		
b	−0.34	−0.87	−0.03	−0.05	−0.03	−0.01	0.50	0.48	
$\ln(b)$	−0.35	−0.90	−0.02	−0.05	−0.02	−0.01	0.48	0.48	0.98

Dolbniak *et al. Biology Direct* (2015) 10:62
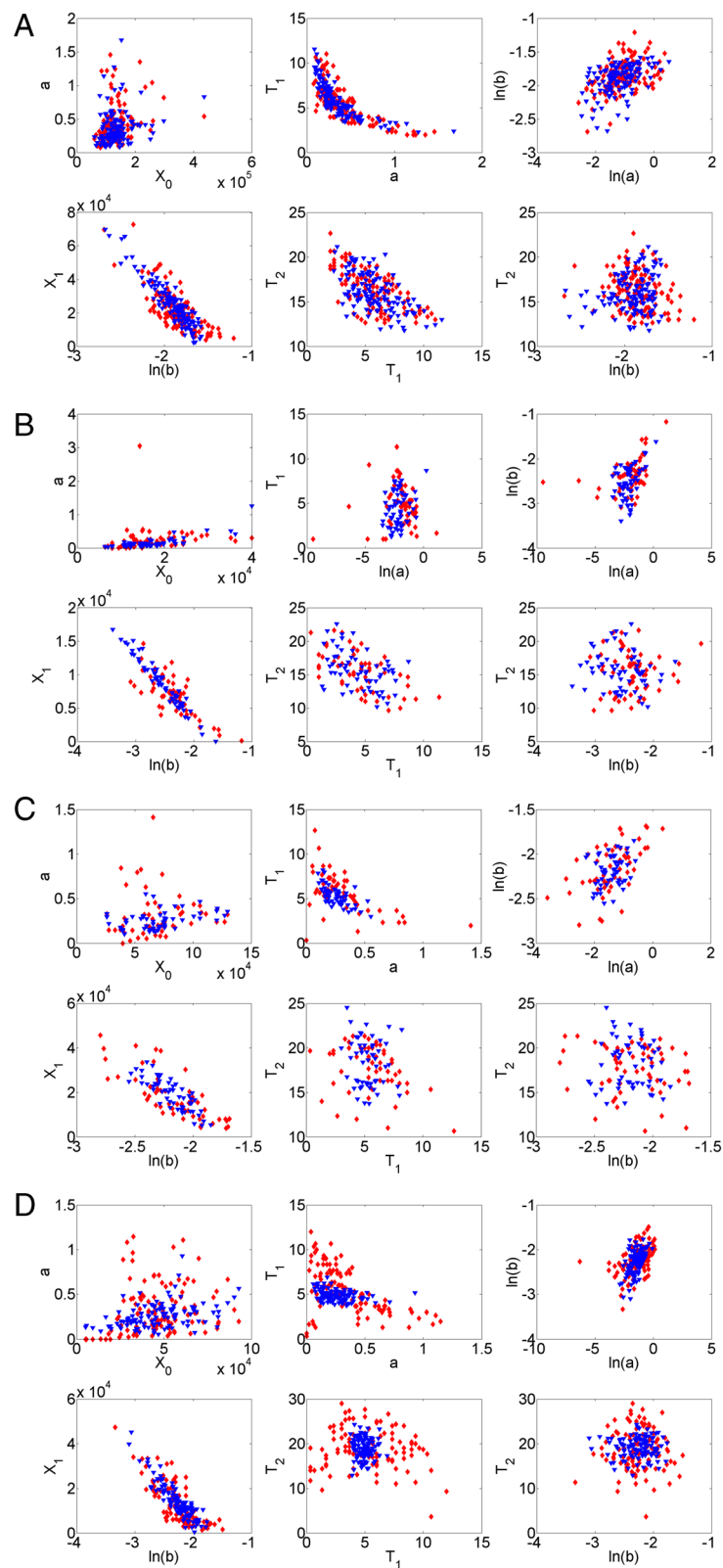
Page 8 of 15

Fig. 3 Data-based vs. simulation-based scatterplots of relationships among principal model variables for Data sets 1–4 (**a**–**d**). Symbols: data-based, red diamonds; simulation-based, blue triangles

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 9 of 15

happened we rejected the calculated parameter and draw another ε value. Effectively, this means we use Gaussian noise, conditional on nonnegativity.

As for the unequal division, the coefficient of variation of the random variable U, characterizing the asymmetry of division, is estimated from Eq. (4). In simulations, the amount of proteins received by daughter cells was sampled from a symmetric beta distribution, which has the variance equal to $V(U) = (8\alpha + 4)^{-1}$. Values of parameters of beta for all data sets are depicted in Table 4.

Results and discussion
Simulations of population dynamics

We have run the model for a large number of cell generations, recording the PRC1 contents in all cells of the resulting pedigree, at constant time intervals. To generate the population of cells, simulation was started with a single ancestor cell. During cell division the PRC1 is unequally divided between daughter cells as described above.

Results of long-term simulations of the model based on Data sets 1 and 2 are presented in Fig. 4. The picture emerging from simulations of Data set 1 is that of a very well-tuned regulatory circuit that provides a stable distribution of PRC1 contents and interdivision times. Outliers, being usually particularly high values of X_2 appear sporadically and are eliminated in the succeeding 1 or 2 generations. Data set 2 seems qualitatively different, with more variation in cell cycle duration.

Simplified mathematical model explaining the cell cycle regulation

The equations of the model can be written explicitly, including the noise terms, based on the detailed model description (items 1–6 from the list earlier on), with

Table 4 Parameters of the beta distribution of the random variable U, characterizing the asymmetry of division

Data set 1	
α	29.29
β	29.29
Data set 2	
α	15.79
β	15.79
Data set 3	
α	35.41
β	35.41
Data set 4	
α	25.95
β	25.95
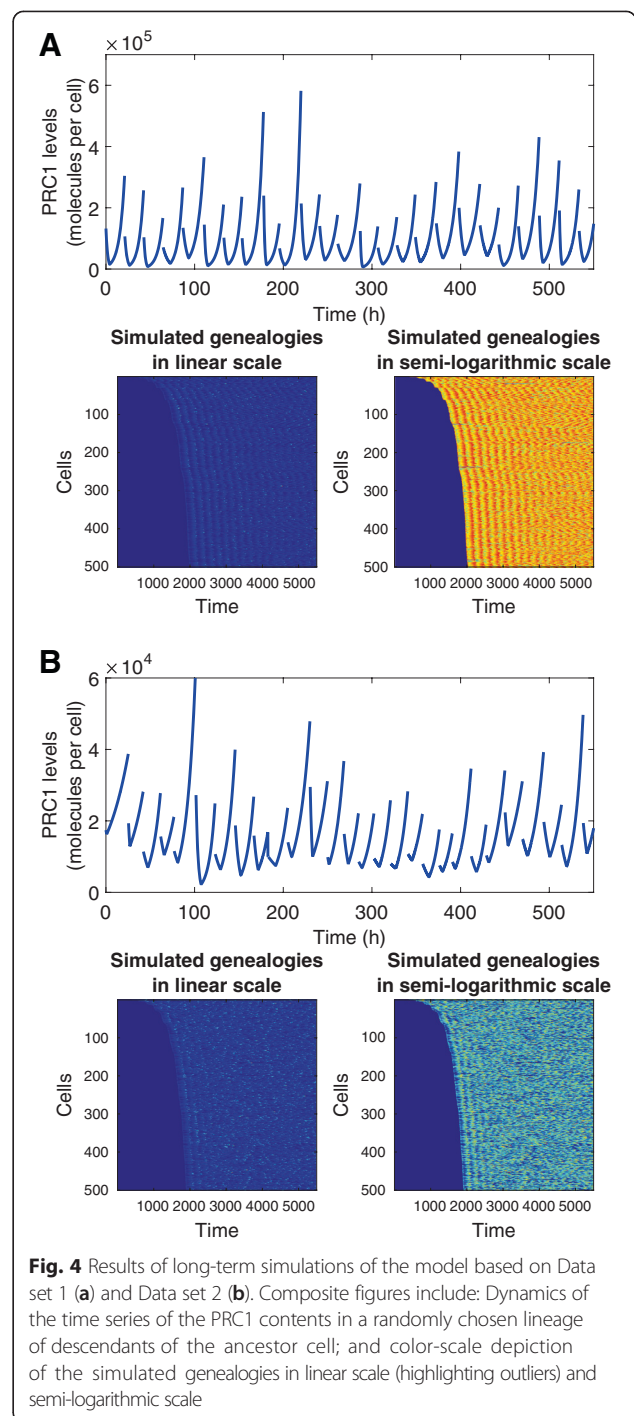
Fig. 4 Results of long-term simulations of the model based on Data set 1 (**a**) and Data set 2 (**b**). Composite figures include: Dynamics of the time series of the PRC1 contents in a randomly chosen lineage of descendants of the ancestor cell; and color-scale depiction of the simulated genealogies in linear scale (highlighting outliers) and semi-logarithmic scale

parameter values depending on the data set, as listed in Table 2. It is instructive to consider a model stripped of noise, which very clearly shows the straightforward nature of cell cycle regulation as estimated from the data. The model will be shown to be very robust to noise introduced by asymmetric division and, at least

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 10 of 15

numerically, this can be extended to any source of noise. The equations of the model without noise are as follows:

$$\ln(a) = b_1 + b_{10}X_0$$
$$\ln(T_1) = b_2 + b_{21}\ln(a) \quad (Data\ sets\ 1,\ 3,\ 4)$$
$$T_1 = b_2 + b_{21}\ln(a) \quad (Data\ set\ 2)$$
$$\ln(X_1) = \ln(X_0) - aT_1 \qquad\qquad (5)$$
$$\ln(b) = b_4 + b_{41}\ln(a) + b_{43}X_1$$
$$T_2 = b_5 + b_{52}T_1$$
$$\ln(X_2) = \ln(X_1) + bT_2$$

Explicit expressions for all the variables can be found, although they are cumbersome. In particular, we obtain the functions $X_2 = \phi(X_0)$, $T = \psi(X_0)$, with $T = T_1 + T_2$, which define the Kimmel et al. model [5]. Figure 5 depicts X_1, X_2, T_1, and T_2, as functions of X_0, for Data sets 1 and 2. The relationships are somewhat different in both cases; Data set 1 exhibits monotonous increasing dependence of X_2 on X_0,

while for the Data set 2, the X_2 graph attains a maximum and then decays to 0. Remarkably, T_1, and T_2 are practically constant as functions of X_0 in both cases.

In the case of the deterministic model (5), if in addition the divisions are symmetric, the equilibrium value of X_0 satisfies the equation

$$X_0 = \phi(X_0)/2 \qquad\qquad (6)$$

as shown in Fig. 6. However, with asymmetric division, the equilibrium is disrupted since

$$X_0' = U\phi(X_0) \qquad\qquad (7)$$

where X_0' is the initial PRC1 contents in the randomly chosen progeny. This results in the values of X_0 oscillating from one generation to another, while the values of $X_2 = \phi(X_0)$ are much less affected (Fig. 7), which illustrates the efficiency of the regulatory mechanism. This is exactly the

Fig. 5 Relationships derived from the deterministic version of the model based on Data sets 1 (**a**) and 2 (**b**). Horizontal axis, X_0, graphs depict variables X_1, X_2, T_1 and T_2, as functions of X_0

Fig. 6 Graphical depiction of the equilibria based on the model and Data sets 1 (**a**) and 2 (**b**). Horizontal axis, X_0, graph of X_2 as a function of X_0 (continuous line), intersected with the graph of $2X_0$

Fig. 7 Oscillations of the PRC1 levels before and after division, generated by models based on Data sets 1 (**a**) and 2 (**b**). For several generations of a randomly chosen lineage started by an ancestral cell, series of values of X_0 and X_2 (interpolated by smooth lines for optical convenience) are depicted. Continuous lines, X_0; dashed lines X_2. For the case of a deterministic model with asymmetry of divisions being the only source of randomness, the continuous line is superimposed on the reference deterministic equilibrium

case considered in [5] and in [20] and the equilibrium distribution can be computed using methods of these papers. Finally, if the full stochastic model is used, then both X_0 and X_2 oscillate, since the uncertainty embedded in the model counteracts the regulatory feedback. This latter case has not been studied analytically.

Parent-progeny and sib-sib correlations

These correlations were computed in [9] for Data set 1. We present model-based correlations in Fig. 8. They are somewhat different from those in the original paper, which however may be the question of scaling and color-coding.

Conclusions

It is interesting and important to understand the mechanisms of epigenetic regulation in proliferating eukaryotic cells. There exist a number of models, with very strong experimental background, which explain the interplay of signaling pathways underlying the timing of cell division including stochastic effects [21]. In addition to this, there exists a very large body of literature addressing experimental relationships among cell size at birth, duration of the cell cycle and asymmetry of division. Idiosyncratically, we may mention models of Kimmel et al. [5], Dyson et al. [3], and Di Talia et al. [22]. One of these models [23] based on observations on embryonic cells led to a bimodal distribution of cell sizes in the population.

Asymmetry of division has been deemed to play a major role in generation of variability in cell populations. Various molecular mechanisms may underlie asymmetric cell division. For example, in our previous works, we use a stochastic model based on branching processes, which qualitatively describe new wave of single cell-based

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 12 of 15
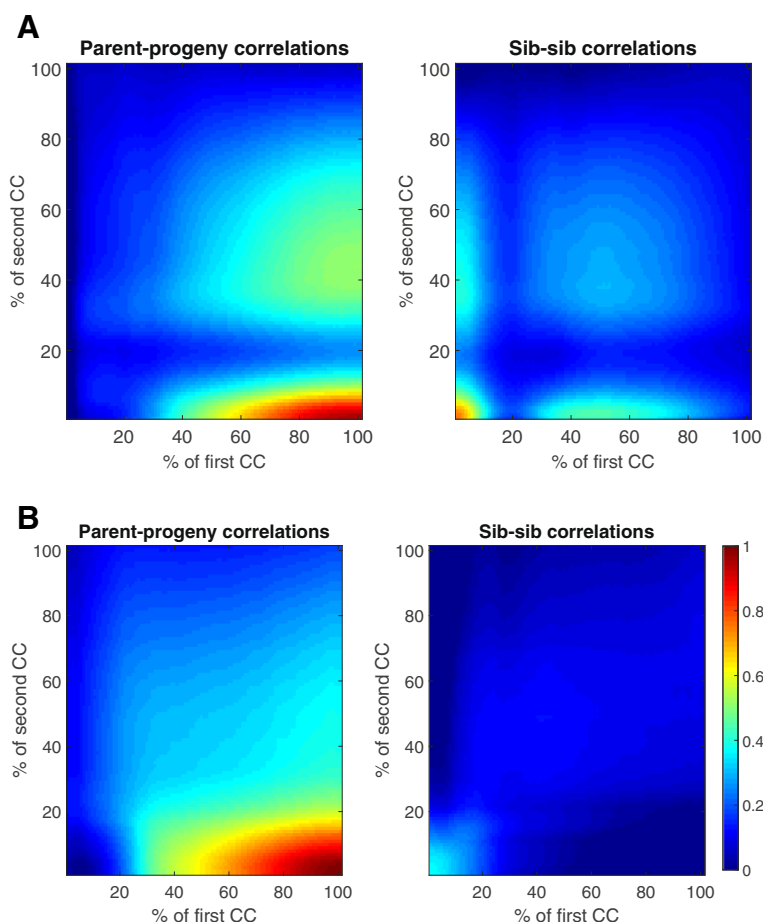
Fig. 8 Simulation-based correlations of PRC1 protein for each time point of the cell cycle (percentage) on (**a**) Data set 1, and (**b**) Data set 2

observation. The model, originally devised in [24] to model evolution of unstable gene amplification and then analyzed mathematically by other, is presented in Fig. 1b. We consider a set of large particles (biological cells), following a binary fission process. Each of the large particles is born containing a number of small particles (genes, proteins, viruses, organelles), which multiply or decay during the large particle's lifetime. The arising population of small particles is then split between the two progeny of the large particle and the process continues in each of the progeny. This "division-within-division" or "branching-within-branching" occurs in various settings in cell and molecular biology. Examples include tightly regulated phenomena such as replication of chromosomal DNA, but also processes in which the number of objects produced in each biological cell is a random variable.

Recent progress in single-cell measurement techniques enabled a much more precise look at cell cycle kinetic in individual cells. We based our modeling up-to-date on the publicly available data from Alon's laboratory. They tracked levels of a number of cell-cycle related and other proteins, some of them over a number of cell cycles and presented synthetic statistics for some of them [9].

Methodologically, we developed a relatively simple model allowing peeling off layers of stochasticity, related to intermediate stages in PRC1 regulation. The initial variability of PRC1 is party cancelled by resetting it to a low level and then increasing its contents until division. We do not know how the timing of the minimum of PRC1 is related to the cell-cycle phases; this is an interesting question in itself. The model, when stripped of stochasticity except for asymmetry of division, reduces to the old model of Kimmel et al. [5], which has been completely characterized mathematically [20] using tools of the operator semigroup theory.

It may be mentioned that in another paper, Arino and Kimmel [25] analyzed a model, which included more stochastic elements than the original model in ref. [5]. In that model, in addition to the asymmetric division, the time the progeny cell spends in the cycle is a random variable with conditional distribution density given its birth size. It has been demonstrated that this approach, originating in the theory of branching processes, is essentially equivalent

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 13 of 15

to the more typical (at the time) formulation in the form of a transport partial differential equation with a nonlocal feedback through boundary condition. For further discussion, see ref. [25].

In the current paper, we used a mathematical model to reproduce experimental trajectories of the PRC1 protein published in [9] and extend the results to model long-range dynamics of the cell population. The main question we address is whether the regulatory feedbacks deduced from single cell cycle data provide epigenetic regulation of cell characteristics in long run. PRC1 protein is regulated by the cell cycle. This protein is absolutely required in cytokinesis, without it cell cannot divide to form two daughter cells [26]. PRC1 is a good candidate because of its role in setting timing of division. Findings of the current paper include tight regulation of the cell cycle (particularly the timing of the cell cycle) even that PRC1 is only one of the players in cell dynamics. Understanding that association, even close, does not necessarily imply causation, we consider this an interesting and important result.

In recent publications authors analyzed single-cell data. Authors of the first paper [27], used the Fucci system (the first marker indicates G0/G1 phases, and the second one the S/G2/M phases) to calculate the length of the cell cycle and the cell cycle phases. They calculated correlations between parent-progeny (no correlation), siblings and cousins (high correlations) cell cycle lengths. Obtained results can be explained by circadian clock control over the mammalian cell cycle in cell. Dynamic of the two Fucci makers was not analyzed, so it is difficult to compare ref. [28] with our work, which is mainly focused on protein dynamics.

In another paper [28] authors analyzed what impact on cell signal response intrinsic and extrinsic noise has and how cells can eliminate variability causes by extrinsic noise. They performed single-cell measurements of three key signal pathways: extracellular signal-regulated kinase, calcium and nuclear factor kappa-B. Again, this paper has a different focus.

Also recently, single-cell expression of cell cycle regulators was analyzed in ref. [29], but the authors explained variability in cell cycle length in the terms of a mammalian clock control. To confirm their theory they proposed a simple linear mathematical model. We used a more parsimonious paradigm of correlation and regression methods to predict what directed influences on a cell cycle are caused by number of protein. The novelty of the present study is the combination of single cell experimental data, correlation analysis and mathematical modeling of individual cell dynamics.

Reviewers' Comments

First of all we would like to thank the referees for their comments and suggestions that were addressed as follows:

Reviewer's report 1: Prof. Ollivier Hyrien, University of Rochester

This manuscript deals with studying the contribution of PRC1 to the regulation of the cell cycle. A mathematical framework is proposed that describes (1) the dynamics of PRC1 during the cell cycle, (2) the random allocation of the protein at division, and (3) cell kinetics. The model is interesting and developed based on an earlier stochastic model proposed by Kimmel and colleagues (1984). An application of this model is presented in which the authors analyze data on the protein dynamics in H1299 non-small cell lung cancer cell lines published by Cohen et al. (2009). Simulations indicate that the model achieves a good description of experimental data.

Authors' response: *Thank you for a positive overview.*

Page 5, Eq. 3. Are the parameters a and b constrained to be positive? Are they random (i.e., cell-specific) or are they identical across cells?

Authors' response: *Parameters a and b are always positive. Every cell has different values of these parameters. Based on initial number of molecules (X_0) we use linear regression to calculate* $\log(a)$

$$\log(a) = b_1 + b_{10}X_0 + \varepsilon$$

b_1, b_{10} *and* ε *are as described in* Table 2 *(similar principles are used to estimate* $\log(b)$*). They differ among data sets.*

As you can see, we do not have to constrain a and b to be positive.

We add additional information on page 5.

Page 6: "T_2 depends linearly on T_1 with additive Gaussian noise". Since T_2 is a duration, perhaps what is meant here is simply that T_2 is linearly associated with T_1, without making any distributional assumption about the noise. This would preserve the positivity of T_2.

Authors' response: *The noise term is needed to obtain agreement with the data. We have not observed many cases where calculated T_2 was a negative value. Nevertheless, when such case happened we rejected calculated T_2 and drew another ε value. Effectively, this means we use Gaussian noise, conditional on nonnegative T_2.*

Page 7: "… at the end of the cell cycle was up to 50 times higher than at the beginning, which is biologically unlikely". Is it the beginning of the cell cycle of the beginning of the accumulation phase. Also, could the authors comment briefly on possible explanations for why the ratio X_2/X_1 was so high in some cells? For example, could this be due to the nonlinearity of the relationship between fluorescence intensity and number of molecules, or to measurement errors?

Authors' response: *Apologies for misprint. We calculate the X_2/X_0 ratio, so the ratio between number of molecules at the beginning and at the end of the cell cycle.*

We believe that outliers result from measurement errors. The beginning or the end of the cell cycle might have been identified wrongly. The measurements were performed before 2009, when cell tracking was less well developed.

Page 7: Was parameter estimation performed on the trajectory of protein concentration for each cell individually? Pages 8–9: In running model simulations, were the random times T1 and T2 assumed to follow specific distributions?

Authors' response: *Parameters in equations (5) were estimated using data from all cells. In this paper, we have not made any distributional assumptions, except for Gaussian distributions of noise terms.*

Page 9: The dependence structure induced by the assumed mechanism of protein dynamics is an interesting feature of the model. Could the authors elaborate on the comparison between model-based correlations and those obtained from experimental data?

Authors' response: *Model-based (*Table 3*) and data-based correlations (*Table 1*) are in good agreement, particularly when their absolute values are high. The agreement is best for Data set 1, which was used as a reference to create the model.*

Reviewer's report 2: Prof. Anna Marciniak-Czochra, University of Heidelberg

Authors consider a mathematical model of epigenetic control of the cell cycle, taking into account stochastic effects and resulting heterogeneity of cell population. Such models have been conceived in the past (see Kimmel's own model published in 1984), but have been largely abandoned for the lack of precise measurements of biomolecules at a single-cell scale. The topic has become important in part because of progress in quantification and in part because of recent emphasis on epigenetic controls of the cell cycle.

The model employs publicly available data from Alon's laboratory, in particular, single-cell trajectories of the PRC1 protein involved in cell cycle controls. Using a multistep estimation procedure, authors successfully build a model that reconstructs the stochastic dynamics at the single-cell level and marginal and joint distributions of most of the meaningful parameters. Authors also demonstrate that if stripped of various layers of dynamics, the model can be reduced to Kimmel et al. 1984 model of cell cycle regulation. Also, it leads to cell population homeostasis when run for extended times.

There are some interesting points that the authors should address before the paper becomes suitable for Biology Direct:

The model in its mathematical framework considers a factor (protein) that may be an active regulator of the cell cycle. It might be worthwhile to discuss if PRC1 qualifies as such factor.

Authors' response: *PRC1 protein is not an active regulator of the cell cycle per se, but it is regulated by the cell cycle. This protein is absolutely required in cytokinesis, without it cell cannot divide to form two daughter cells. More precisely, the central spindle bundle is not formed and this prevents the final abscission event [26]. We think that because of a strong correlation to cell-division events, PRC1 protein qualifies to be used in our model.*

Authors provide simulations of the long-term dynamics of the model under variable levels of stochasticity. What is missing is a discussion of mathematical results that might be relevant for establishing long-term homeostasis of the model.

Authors' response: *Recently, single-cell expression of cell cycle regulators was analyzed in ref. [29], but the authors explained variability in cell cycle length in the terms of a mammalian clock control. To confirm their theory they proposed a simple linear mathematical model.*

We used a more parsimonious paradigm of correlation and regression methods to predict what directed influences on a cell cycle are caused by number of protein. The novelty of the present study is the combination of single cell experimental data, correlation analysis and mathematical modeling of individual cell dynamics.

Finally, recently, there have been a number of new papers published, which either involve similar models, or show new techniques for obtaining data at single-cell level (see Nature 2015, 519, 468–471, or Science 2014, 346, 1370–1373). Enhanced discussion of these models, compared to the model in the present manuscript, is desirable.

Authors' response: *In these two publications authors analyzed single-cell data. Authors of the first paper [27], used the Fucci system (the first marker indicates G0/G1 phases, and the second one the S/G2/M phases) to calculate length of cell cycle and cell cycle phases. They calculated correlations between parent-progeny (no correlation), siblings and cousins (high correlations) cell cycle lengths. Obtained results can be explained by circadian clock control over the mammalian cell cycle in cell. Proteins dynamic of two makers was not analyzed, so it is hard to compare that with our work, which is mainly focused on protein dynamics.*

In the second paper [28] authors analyses what impact on cell signal response have intrinsic and extrinsic noise and how cell can eliminate variability causes by extrinsic noise. They performed single-cell measurements of three key signal pathways: extracellular signal-regulated kinase, calcium and nuclear factor kappa-B.

Reviewer's report 3: Prof. Alberto d'Onofrio, International Prevention Research Institute

In this computational epigenetics works the authors investigate how the regulation of the PRC1 protein content influence the long term behaviour of a cellular population.

Dolbniak *et al. Biology Direct* (2015) 10:62

Page 15 of 15

The general topic of how epigenetic changes impact on a population is one of the most important of molecular biology, and it is at the interface between systems biology and population dynamics. I think that the idea of the manuscript is very good, the work is well written (apart an important minor detail) and the results are of interest.

Authors' response: *Thank you for a very positive comment.*

I recommend its acceptance upon a minor but important change is implemented. The changes concerns that fact the in this work the authors do not provide enough mathematical details of the original model of unequal divisions by Kimmel et al. (1984), which makes the paper more difficult to be read for those who, differently form myself, did not read it.

Authors' response: *We included additional information about the model of Kimmel* et al. *1984.*

I suggest of inserting in the full text the supplemental figure S1.

Authors' response: *We included Figure S1 in the main body of the paper (currently* Fig. 3*)*

There is some typos. For example in the abstract "we follow the philosophy of a 1984 model by Kimmel model"

Authors' response: *Apologies for the typos. We corrected all of them.*

Abbreviations
PRC1: Protein Regulator of cytokinesis 1; CDKs: Cyclin-dependent kinases; H1299: Human non-small cell lung carcinoma cell line.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
MD took part in model building, carried out parameter estimation and simulation, prepared the manuscript. MK designed the model, supervised parameter estimation and simulation, prepared the manuscript. JS initiated the project, researched literature, supervised simulations, prepared the manuscript. All authors read and approved the manuscript.

References
1. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development. 2009. doi:10.1242/dev.035139.
2. Patsy H, Jagers P, Vatutin VA. Branching processes: variation, growth, and extinction of populations, Vol. 5. Cambridge: Cambridge University Press; 2005.
3. Janet D, Villella-Bressan R, Webb G. A nonlinear age and maturity structured model of population dynamics: I Basic theory. J Math Anal Appl. 2000;242(1):93–104.
4. Kimmel M, Axelrod D. Branching processes in biology. 2nd ed. New York: Springer; 2015.
5. Kimmel M, Darzynkiewicz Z, Arino O, Traganos F. Analysis of a cell cycle model based on unequal division of metabolic constituents to daughter cells during cytokinesis. J Theor Biol. 1984;110(4):637–64.
6. Arino O, Kimmel M, Zerner M. Analysis of a cell population model with unequal division and random transition. In: Arino O, Axelrod DE, Kimmel M, editors. Mathematical population dynamics. New York: Marcel Dekker; 1991. p. 3–12.
7. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297:1183–6.
8. Altschuler SJ, Wu LF. Cellular heterogeneity: Do differences make a difference? Cell. 2010;141:559–63.
9. Cohen AA, Kalisky T, Mayo A, Geva-Zatorsky N, Danon T. Protein dynamics in individual human cells: experiment and theory. PLoS ONE. 2009. doi:10.1371/journal.pone.0004901.
10. Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, Hunter T, et al. PRC1: a human mitotic spindleassociated CDK substrate protein required for cytokinesis. Mol Cell. 1998. doi:10.1016/S1097-2765(00)80302-0.
11. Bechstedt S, Brouhard GJ. Motors and MAPs collaborate to size Up microtubules. Dev Cell. 2013;26(2):118–20. doi:10.1016/j.devcel.2013.07.010.
12. Piunti A, Rossi A, Cerutti A, Albert M, Jammula S, Scelfo A, et al. Polycomb proteins control proliferation and transformation independently of cell cycle checkpoints by regulating DNA replication. Nat Commun. 2014. doi:10.1038/ncomms4649.
13. Hu B, Li S, Zhang X, Zheng X. HSCARG, a novel regulator of H2A ubiquitination by downregulating PRC1 ubiquitin E3 ligase activity, is essential for cell proliferation. Nucleic Acids Res. 2014. doi:10.1093/nar/gku230.
14. Subramanian R, Ti SC, Tan L, Darst SA, Kapoor TM. Marking and measuring single microtubules by PRC1 and kinesin-4. Cell. 2013. doi:10.1016/j.cell.2013.06.021.
15. van den Boom V, Rozenveld-Geugien M, Bonardi F, Malanga D, van Gosliga D, Heijink A, et al. Nonredundant and locus-specific gene repression functions of PRC1 paralog family members in human hematopoietic stem/progenitor cells. Blood. 2013. doi:10.1182/blood-2012-08-451666.
16. Rayess H, Wang MB, Srivatsan ES. Cellular senescence and tumor suppressor gene p16. Int J Cancer. 2012. doi:10.1002/ijc.27316.
17. Boukarabila H, Saurin AJ, Batsche E, Mossadegh N, van Lohuizen M, Otte AP, et al. The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation. Gene Dev. 2009. doi:10.1101/gad.512009.
18. Gieni RS, Ismail IH, Campbell S, Hendzel MJ. Polycomb group proteins in the DNA damage response: a link between radiation resistance and "stemness". Cell Cycle. 2011;10(6):883–94.
19. van de Geer SA. Least squares estimation, Encyclopedia of statistics in behavioral science. 2005. p. 1041–5.
20. Arino O, Kimmel M. Asymptotic analysis of a cell cycle model based on unequal division. SIAM J Appl Math. 1987;47(1):128–45.
21. Tyson JJ, Novak B. Control of cell growth, division and death: information processing in living cells. Interface Focus. 2014. doi:10.1098/rsfs.2013.0070.
22. Di Talia S, Skotheim JM, Bean JM, Siggia ED, Cross FR. The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature. 2007. doi:10.1038/nature06072.
23. Kimmel M, Arino O. Cell cycle kinetics with supramitotic control, two cell types, and unequal division: a model of transformed embryonic cells. Math Biosci. 1991;105(1):47–79.
24. Kimmel M. Quasistationarity in a branching model of division-within-division, Classical and modern branching processes. New York: Springer; 1997. p. 157–64.
25. Arino O, Kimmel M. Comparison of approaches to modeling of cell population dynamics. SIAM J Appl Math. 1993;53(5):1480–504.
26. Mollinori C, Kleman JP, Saoudi Y, Jablonski SA, Perard J, Yen TJ, et al. Ablation of PRC1 by small interfering RNA demonstrates that cytokinetic abscission requires a central spindle bundle in mammalian cells, whereas completion of furrowing does Not. Mol Biol Cell. 2005;16:1043–55.
27. Sandler O, Mizrahi SP, Weiss N, Agam O, Simon I, Balaban NQ. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. Nature. 2015;519:468–47.
28. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, et al. Accurate information transmission through dynamic biochemical signaling networks. Science. 2014;346(6215):1370–3.
29. Feillet C, Krusche P, Tamanini F, Janssens RC, Downey MJ, Martin P, et al. Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. Proc Natl Acad Sci U S A. 2014;111(27):9828–33.