Title	Predicting disease progression from short biomarker series using expert advice algorithm
Author(s)	Morino, Kai; Hirata, Yoshito; Tomioka, Ryota; Kashima, Hisashi; Yamanishi, Kenji; Hayashi, Norihiro; Egawa, Shin; Aihara, Kazuyuki
Citation	Scientific Reports (2015), 5
Issue Date	2015-05-20
URL	http://hdl.handle.net/2433/216227
Right	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/
Type	Journal Article
Textversion	publisher

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MACHINE LEARNING
TUMOUR BIOMARKERS
APPLIED MATHEMATICS

Received
24 September 2014

Accepted
6 February 2015

Published
20 May 2015

Correspondence and requests for materials should be addressed to K.M. (morino@mist.i.u-tokyo.ac.jp)

Predicting disease progression from short biomarker series using expert advice algorithm

Kai Morino[1], Yoshito Hirata[1,2], Ryota Tomioka[3], Hisashi Kashima[4], Kenji Yamanishi[1,5], Norihiro Hayashi[6], Shin Egawa[6] & Kazuyuki Aihara[1,2]

[1]Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan, [2]Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan, [3]Toyota Technological Institute at Chicago, Chicago, Illinois 60637, USA, [4]Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, [5]CREST, JST, Honcho, Kawaguchi, Saitama 332-0012, Japan, [6]Department of Urology, Jikei University School of Medicine, Tokyo 105-8461, Japan.

Well-trained clinicians may be able to provide diagnosis and prognosis from very short biomarker series using information and experience gained from previous patients. Although mathematical methods can potentially help clinicians to predict the progression of diseases, there is no method so far that estimates the patient state from very short time-series of a biomarker for making diagnosis and/or prognosis by employing the information of previous patients. Here, we propose a mathematical framework for integrating other patients' datasets to infer and predict the state of the disease in the current patient based on their short history. We extend a machine-learning framework of "prediction with expert advice" to deal with unstable dynamics. We construct this mathematical framework by combining expert advice with a mathematical model of prostate cancer. Our model predicted well the individual biomarker series of patients with prostate cancer that are used as clinical samples.

Mathematical models of diseases have been constructed to understand the mechanisms of diseases[1–7], provide diagnosis and prognosis[8–10], and determine treatment options[11–14]. When we focus on a clinical setting, it is crucial that we can estimate the state of a disease from short biomarker observations. Clinicians make such estimations using their experience with previous patients (see Fig. 1a). To the best of our knowledge, such estimations have not been realized mathematically thus far. If such mathematical estimation is possible, then we can optimize a treatment option in a personalized way. The difficulty of estimations stems not only from a lack of information, but also the instability of biomarkers' time-series, such as those for cancer volumes. Thus, our goal is to infer the state of a disease from both short, unstable time-series data of biomarkers obtained from a target patient and longer time-series data of biomarkers from previous patients who suffered from the same disease. We adopt the machine-learning framework of "online prediction", which integrates "experts' advice"[15] to make accurate predictions, where experts are short-term patterns of previous patients' histories, which are conformed to the target patient time-series.

A series of samples from a patient contains information on (often unstable) disease dynamics[8,10] such as rapid increase. By considering the time series observed from the unstable dynamics, we may be able to better understand the current disease state. Employing past patients' time-series as experts and the target patient's time-series as observations, we can predict a time-series with the standard *expert advice* method[15]. However, this cannot be used directly, because we must deal with the unstable dynamics in which the value of a biomarker increases rapidly. In this paper, we propose an approach that couples an existing machine-learning technique with the instability possessed by the temporal disease datasets. Our method is based on the standard expert advice[15], but deals with the instability of the underlying dynamics[8,10] by integrating trajectories in a database with weights that increase exponentially in time.

Results

The proposed method: temporal expert advice (TEA). We extend the standard expert advice method[15] to one that emphasizes near-past information. This *temporal expert advice*, or the TEA algorithm, consists of three steps (see Fig. 1b–d). TEA uses a collection of time-series, which we call experts, and weights each expert based on its agreement with the target time-series. The algorithm outputs a prediction by combining these experts.

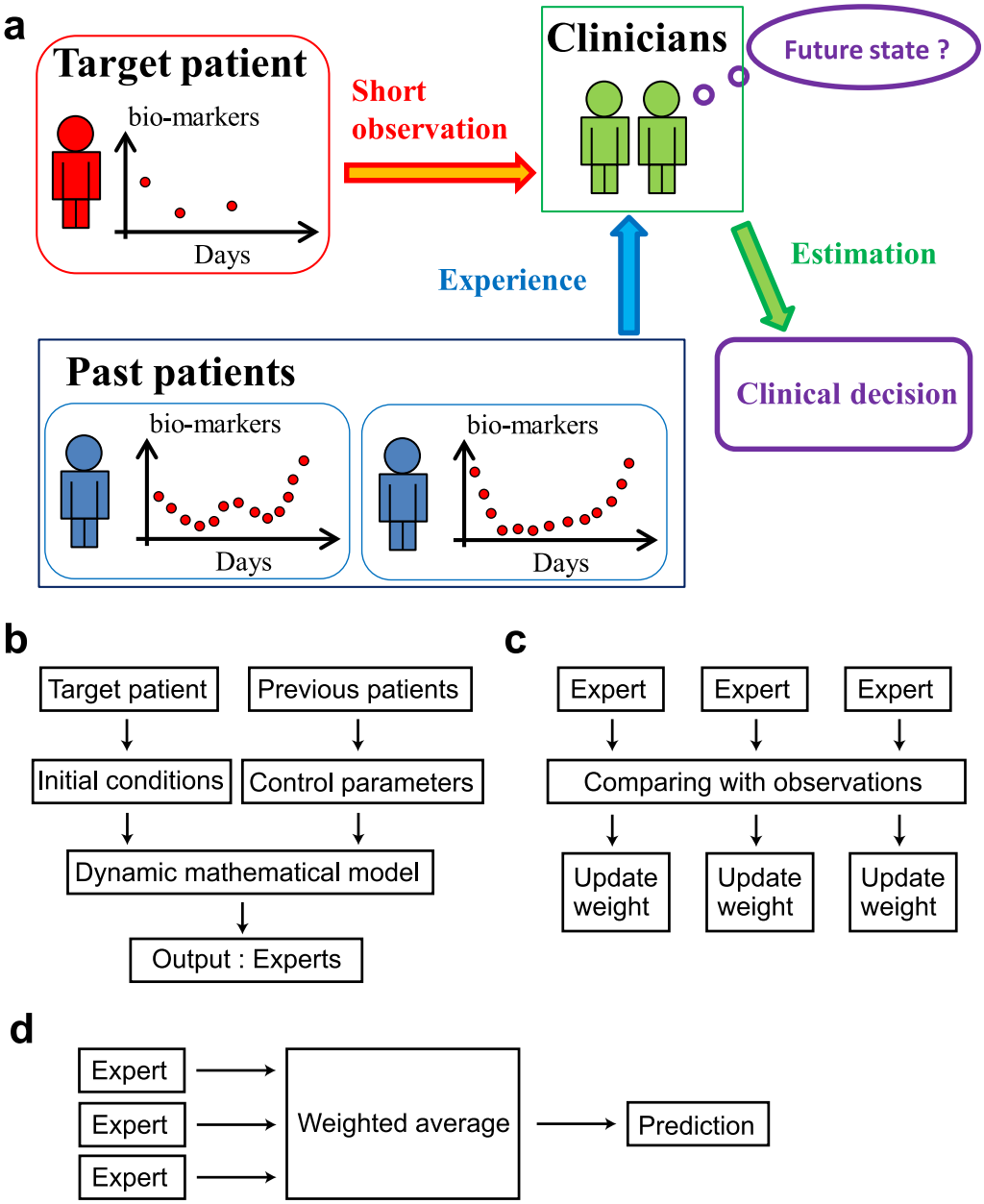

Figure 1 | Schematic illustration of the estimation from short observations of biomarkers. (a) The prediction of clinicians. (b) The first step of TEA. (c) The second step of TEA. (d) The third step of TEA.

The first step constructs an expert of a target system. There are two options. The first option is to use long time-series observed in the past as they are. We construct experts by simply inserting previous parts of the time-series or the datasets of previous patients. Let $x_{j,l}$ be the lth point of the jth time-series in a database ($j = 1, 2, \ldots, J, l = 1, 2, \ldots, L$) and $f_{i,t}$ be the ith expert's advice at time t. Numbers J and L are the number of time-series and the number of points in each time-series, respectively; We assume that the lengths of the time-series are equal, but it is easy to extend to cases of different lengths. Let P be the number of points related to each expert. Then, we can define an expert $f_{(L-P+1)(j-1)+i,k} = x_{j,i+k-1}$ for $i = 1, 2, \ldots, L - P + 1, k = 1, 2, \ldots, P$. The second option is roughly to fit a mathematical model that has a set of parameters to a very short time-series, obtain the initial conditions for each set of parameters, and prepare the set of experts with these parameters. The details of this second rough option are discussed after we introduce a mathematical model of prostate cancer in the later section.

In the second step, TEA weights the trajectories $f_{i,t}$ in the database to generate an appropriate weighting for the most current state. Let y_k denote the observation at time k, and let $l(\cdot,\cdot)$ be a loss function. When we include the next point, the weights $w_{i,t}$ are updated according to a formula obtained by modifying the standard expert advice[15]. To achieve this, we sum the loss at each time step with a coefficient as follows:

$$L_{i,t-1} = \sum_{k=1}^{t-1} a_k(t) l(f_{i,k}, y_k), \quad (1)$$

$$L_{t-1} = \sum_{k=1}^{t-1} a_k(t) l(p_k, y_k). \quad (2)$$

The modified form considers the instability of the underlying dynamics by introducing a coefficient $a_k(t)$, which measures the reliability of the prediction at, and increases exponentially with, time k such that $a_k(t) = \lambda^{k-1}$ or λ^k with $\lambda > 1$. Thus, the real values $L_{i,t}$ and

L_t are the exponentially weighted losses of the ith expert and the predictor up to time t, respectively. We define $L_0 = L_{i,0} = 0$ for simplicity. The weight of the expert is updated as

$$w_{i,t-1} = \exp(-\eta L_{i,t-1}), \tag{3}$$

where η is a learning rate. Chernov and Zhdanov[16] proposed a modified expert advice method in which they defined $a_k(t) = \rho^{t-k-1}$ and $0 < \rho < 1$. We call this the "CZ method".

The third step predicts future states of the target system by applying the obtained weighting to the future trajectories in the database[15]. We can generate a point prediction by simply adding the q steps ahead of the trajectories with the weights obtained in the second step as follows:

$$p_{t+q-1} = \frac{\sum_{i=1}^{N} w_{i,t-1} f_{i,t+q-1}}{\sum_{i=1}^{N} w_{i,t-1}}, \tag{4}$$

where N denotes the total number of experts. We can also generate a distributional prediction by assuming the distribution of observational errors, and summing this error distribution with the weights. To make these predictions online, we repeat the second and third steps iteratively.

Upper bound of the regret of TEA. We derive an upper bound of the regret of TEA. The primary property peculiar to TEA lies in our definition of $a_k(t)$. We define the coefficient as $a_k(t) = \lambda^{k-1}$ or λ^k, where $\lambda > 1$. The choice of these two options depends on each situation. When a given data is too short, we choose the latter, e.g. our prediction about the biomarker of prostate cancer, or PSA (prostate-specific antigen). Let $\breve{L}_t$ and $\breve{L}_{i,t}$ be the accumulated losses for the proposed method. We call these the *exponential accumulated losses* to distinguish them from the standard accumulated losses. In addition, we define the regret as $\breve{R}_t = \breve{L}_t - \min_{i=1,\ldots,N} \breve{L}_{i,t}$. The upper bound of the regret with $a_k(t) = \lambda^{k-1}$ is then given by

$$\breve{R}_t \le \frac{\ln N}{\eta} + \frac{\varepsilon^2 \eta}{8} \cdot \frac{\lambda^{2t} - 1}{\lambda^2 - 1}. \tag{5}$$

Proof of the upper bound in our proposed method. We give a proof of Eq. (5) in a similar way to the Proof of Theorem 2.2 in Ref. 15. Define a new variable $W_t \equiv \sum_{k=1}^{N} \breve{w}_{k,t} = \sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t}\right)$. We will consider the upper and lower bounds of $\ln(W_t/W_0)$ to construct the upper bound of the regret. First, we obtain the lower bound of $\ln(W_t/W_0)$ as

$$\begin{aligned} &\ln \frac{W_t}{W_0} \\ &= \ln W_t - \ln W_0 \\ &= \ln\left(\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t}\right)\right) - \ln N \\ &\ge \ln\left(\max_{k=1,\ldots,N} \exp\left(-\eta \breve{L}_{k,t}\right)\right) - \ln N \\ &= -\eta \min_{k=1,\ldots,N} \breve{L}_{k,t} - \ln N. \end{aligned} \tag{6}$$

Second, we derive the upper bound of $\ln(W_t/W_0)$. Observe that $\breve{L}_{k,t} = \breve{L}_{k,t-1} + \lambda^{t-1} l(f_{k,t}, y_t)$. Then, we can reformulate $\ln(W_t/W_{t-1})$ as follows:

$$\begin{aligned} \ln \frac{W_t}{W_{t-1}} &= \ln \frac{\sum_{k=1}^{N} \exp(-\eta \breve{L}_{k,t})}{\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t-1}\right)} = \\ &\ln \frac{\sum_{k=1}^{N} \exp(-\eta \breve{L}_{k,t-1}) \exp\left(-\eta \lambda^{t-1} l(f_{k,t}, y_t)\right)}{\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t-1}\right)}. \end{aligned} \tag{7}$$

Equation (7) can be regarded as the average of random variable $\exp(-\eta\lambda^{t-1} l(f_{k,t}, y_t))$ with a probability mass function proportional to $\exp(-\eta \breve{L}_{k,t-1})$. Lemma 2.2 of Ref. 15 states that

$$\ln E(\exp(sx)) \le sE(x) + \frac{s^2 (b-a)^2}{8}. \tag{8}$$

Here, we assume that x is a random variable satisfying $a \le x \le b$, and that the inequality holds when s is any real number.

Replace s by $-\eta\lambda^{t-1}$ and x by $l(f_{k,t}, y_t)$. Then, the upper bound of Eq. (7) can be found using Eq. (8) as follows:

$$\begin{aligned} &\ln \frac{W_t}{W_{t-1}} \\ &= \ln E\left(\exp\left(-\eta \lambda^{t-1} l(f_{k,t}, y_t)\right)\right) \\ &\le -\eta \lambda^{t-1} E(l(f_{k,t}, y_t)) + \frac{\eta^2 \lambda^{2(t-1)} \varepsilon^2}{8} \\ &= -\eta \lambda^{t-1} \frac{\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t-1}\right) l(f_{k,t}, y_t)}{\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t-1}\right)} + \frac{\eta^2 \lambda^{2(t-1)} \varepsilon^2}{8} \\ &\le -\eta \lambda^{t-1} l\left(\frac{\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t-1}\right) f_{k,t}}{\sum_{k=1}^{N} \exp\left(-\eta \breve{L}_{k,t-1}\right)}, y_t\right) + \frac{\eta^2 \lambda^{2(t-1)} \varepsilon^2}{8} \\ &= -\eta \lambda^{t-1} l\left(\frac{\sum_{k=1}^{N} \breve{w}_{k,t-1} f_{k,t}}{\sum_{k=1}^{N} \breve{w}_{k,t-1}}, y_t\right) + \frac{\eta^2 \lambda^{2(t-1)} \varepsilon^2}{8} \\ &= -\eta \lambda^{t-1} l\left(\breve{p}_t, y_t\right) + \frac{\eta^2 \lambda^{2(t-1)} \varepsilon^2}{8}. \end{aligned} \tag{9}$$

We assume that $l(\cdot,\cdot)$ is convex as described above. Then, the upper bound of $\ln(W_t/W_0)$ can be derived as

$$\begin{aligned} &\ln \frac{W_t}{W_0} = \sum_{k=1}^{t} \ln \frac{W_k}{W_{k-1}} \\ &\le \sum_{k=1}^{t} \left(-\eta \lambda^{k-1} l\left(\breve{p}_k, y_k\right)\right) + \sum_{k=1}^{t} \frac{\eta^2 \lambda^{2(k-1)} \varepsilon^2}{8} \\ &= -\eta \breve{L}_t + \frac{\eta^2 \varepsilon^2}{8} \cdot \frac{\lambda^{2t} - 1}{\lambda^2 - 1}. \end{aligned} \tag{10}$$

Because Eqs. (6) and (10) provide lower and upper bounds of $\ln(W_t/W_0)$, respectively, the following inequality is obtained:

$$-\eta \min_{k=1,\ldots,N} \breve{L}_{k,t} - \ln N \le -\eta \breve{L}_t + \frac{\eta^2 \varepsilon^2}{8} \cdot \frac{\lambda^{2t} - 1}{\lambda^2 - 1}. \tag{11}$$

By substituting the regret $\breve{R}_t = \breve{L}_t - \min_{k=1,\ldots,N} \breve{L}_{k,t}$ into Eq. (11), we finally reach the following inequality:

$$\breve{R}_t \le \frac{\ln N}{\eta} + \frac{\eta \varepsilon^2}{8} \cdot \frac{\lambda^{2t} - 1}{\lambda^2 - 1}. \tag{12}$$

(Proof end)

Optimization of the upper bound of TEA. We minimize the upper bound of Eq. (12) over η. First, we differentiate the upper bound with respect to η as follows:

$$-\frac{\ln N}{\eta^2}+\frac{\varepsilon^2}{8}\cdot\frac{\lambda^{2t}-1}{\lambda^2-1}=0. \tag{13}$$

The solution is

$$\eta_*=\frac{2\sqrt{2}}{\varepsilon}\cdot\sqrt{\frac{\lambda^2-1}{\lambda^{2t}-1}\cdot\ln N}, \tag{14}$$

which gives the smallest upper bound. Replacing η in the upper bound of $\breve{R}_t$ with η_*, we obtain the following optimal upper bound $\breve{A}(t)$:

$$\breve{A}(t)=\frac{\varepsilon}{\sqrt{2}}\cdot\sqrt{\frac{\lambda^{2t}-1}{\lambda^2-1}\cdot\ln N}. \tag{15}$$

Although this optimal upper bound perhaps seems to be curious at a glance due to its exponential increase with t, this is caused by the definition of the accumulated losses Eqs. (1) and (2) with $a_k(t) = \lambda^{k-1}$. This regret can be compared with the normal types of regrets using relationship described in the next section. When $\varepsilon = 1$ and $\lambda \to 1$, the optimal upper bound $\breve{A}(t)$ coincides with $\sqrt{t\ln N/2}$, which is the upper bound obtained in the standard expert advice method[15].

Comparison between the proposed method and the Chernov–Zhdanov method. Here, we highlight the difference between the CZ method and the proposed TEA method. The first point of difference is the optimal upper bound of the regret. We briefly introduce the optimal upper bound of the CZ method[16]. Let $\hat{L}_{i,t}$ and $\hat{L}_t$ be the accumulated losses for the ith expert and the predictor for the CZ method, respectively. Then, the optimal upper bound $\hat{A}_c(t)$ of the regret $\hat{R}_t=\hat{L}_t-\min_{i=1,\ldots,N}\hat{L}_{i,t}$ for the CZ method is given by

$$\hat{A}_c(t)=\varepsilon\sqrt{\frac{1-\rho^t}{1-\rho}\cdot\ln N}. \tag{16}$$

Note that we assume the case where the value of the decay rate ρ does not depend on t or k. See Ref. 16 for the proof.

Although we cannot directly compare these regrets, we can compare them after normalization. Assuming that the decay rates are equal, namely $\lambda = \rho^{-1}$, the regrets $\breve{R}_t$ and $\hat{R}_t$ have the following relation:

$$\begin{aligned}\hat{R}_t&=\sum_{k=1}^{t}\rho^{t-k}\left(l(\hat{p}_k,y_k)-\min_{i=1,\ldots,N}l(f_{i,k},y_k)\right)\\&=\rho^{t-1}\sum_{k=1}^{t}\lambda^{k-1}\left(l(\hat{p}_k,y_k)-\min_{i=1,\ldots,N}l(f_{i,k},y_k)\right)\\&=\rho^{t-1}\breve{R}_t.\end{aligned} \tag{17}$$

Using this relation, a comparison between the two upper bounds is feasible. Multiplying the optimal upper bound $\breve{A}(t)$ by ρ^{t-1}, we obtain the normalized optimal upper bound $\breve{A}_m(t)$ as

$$\begin{aligned}\breve{A}_m(t)&=\rho^{t-1}\breve{A}(t)\\&=\rho^{t-1}\frac{\varepsilon}{\sqrt{2}}\cdot\sqrt{\frac{1-\rho^{-2t}}{1-\rho^{-2}}\cdot\ln N}\\&=\varepsilon\sqrt{\frac{1-\rho^{2t}}{1-\rho^2}\cdot\frac{\ln N}{2}}.\end{aligned} \tag{18}$$

Then, the following relation is obtained:

$$\frac{\breve{A}_m(t)}{\hat{A}_c(t)}=\sqrt{\frac{1}{2}\cdot\frac{1+\rho^t}{1+\rho}}<1. \tag{19}$$

This result means that the normalized optimal upper bound of the proposed method is always smaller than that of the CZ method when $0 < \rho < 1$.

Next, we compare the weights produced by the two methods. Let $\hat{w}_{i,t}$ and $\breve{w}_{i,t}$ be the weights of the ith expert at time t in the CZ and TEA methods, respectively. Similarly to the derivation of Eq. (17), the accumulated losses of both methods are related by $\hat{L}_{i,t}=\rho^{t-1}\breve{L}_{i,t}$. Substituting this relation into $\breve{w}_{i,t}$, we have

$$\begin{aligned}\breve{w}_{i,t}&=\exp\left(-\eta\breve{L}_{i,t}\right)\\&=\exp\left(-\eta\hat{L}_{i,t}\cdot\rho^{-t+1}\right)\\&=\left(\exp\left(-\eta\hat{L}_{i,t}\right)\right)^{\rho^{-t+1}}\\&=\left(\hat{w}_{i,t}\right)^{\rho^{-t+1}}.\end{aligned} \tag{20}$$

Equality (20) means that the proposed TEA method tends to assign reliable experts with heavier weights than the CZ method. This implies that $\breve{w}_{i,t}\Big/\sum_{j=1}^{N}\breve{w}_{j,t}\geq\hat{w}_{i,t}\Big/\sum_{j=1}^{N}\hat{w}_{j,t}$ for reliable experts because $\rho^{-t+1}\geq 1$.

Examples of time-series prediction for mathematical models. We demonstrate the superiority of the TEA method to both the CZ method and the standard expert advice in online time-series prediction using toy examples. We use the Hénon map[17] and the Ikeda map[18] for our demonstration. These two models are commonly used to test nonlinear time-series analysis methods, which exhibit typical unstable chaotic dynamics. First, we generate time-series for the database using various values of parameters. We then generate a target time-series for prediction using a set of parameter values that is different from those used to generate the database. We prepare $M \times S$ experts for the database, where M is the number of parameter sets. For each parameter set, we generate S experts with different initial conditions. In numerical simulations of TEA, we set $\varepsilon = 1$ and $\eta=\sqrt{8\cdot\frac{\lambda^2-1}{\lambda^8-1}\cdot\ln MS}$. We also set $\lambda = \rho^{-1}$ and $\rho = 0.9$. See Algorithm 2 in Ref. 16 for the implementation of the CZ method, and Ref. 15 for that of the standard expert advice.

The Hénon map[17] is a two-dimensional map defined as

$$\begin{aligned}x_{n+1}&=1-ax_n^2+y_n,\\y_{n+1}&=bx_n.\end{aligned} \tag{21}$$

We set the parameters at $a = 1.35$ and $b = 0.15$ to generate the target time-series. Note that the dynamics produced by this parameter set is of deterministic chaos. The experts' parameters are uniformly chosen from $a \in [1.3,1.4]$ and $b \in [0.1,0.2]$. The initial conditions x_0 and y_0 are randomly chosen in $[-0.02, 0.02] \times [-0.02, 0.02]$, and the map is iterated for 1,000 steps to eliminate transient effects. We assume that we observe and predict the value of $x + y$. We use this assumption because we can observe a scalar biomarker of PSA in the prostate cancer application discussed later. The results presented in Figs. 2a, 2b, and 2c show that the proposed TEA achieves better online time-series prediction than the standard expert advice and the CZ method. We choose $M = 100$ and $S = 1{,}000$ in Figs. 2a, 2f, and 2g. Another example of the Ikeda map is shown in Supplementary Fig. S1 (see also Supplementary Information).

The proposed TEA method provides the best online prediction in different toy examples. The more experts we use, the smaller the prediction errors become. When a large number of experts are used, the proposed TEA tends to achieve the best online time-series prediction. We need to decay the past information in these examples, because the unstable chaotic dynamics rapidly loses the memories.

Examples of time-series prediction for real datasets. We now consider two real datasets: violin sounds[19] and the membrane potential of squid giant axons[20]. The violin sounds are RWC-MDB-I-2001 No. 15 in the RWC Music Database (Musical Instrument Sound). Previous studies on squid giant axons have demonstrated the chaotic nature of the underlying dynamics[20–23]. These time-series are both scalar and real-valued. We divide each time series into two. The first part is used to build the database, and the second constructs the targets for online prediction. We use $M = 1{,}000$ and $M = 120$ targets for the analysis of violin sounds and squid giant axon data, respectively. The lengths of the target data are 311 for the violin data and 51 for the squid giant axon data; numbers and lengths of target data are determined by the lengths of the original datasets.

We compare five methods using these real data. These are our TEA method, the CZ method, the standard expert advice, the persistence prediction, and the average prediction. The persistence prediction is a method that we let the current value to be the prediction for the next time point. We compare each pair of the method individually, and count the number of points at which the prediction by one method is better than the other for each target time-series. If one method is superior at more than half the data points, we declare that method the winner on the target data. We exclude the initial ten points from the analysis, because we cannot prepare the learning part. Finally, we count the number of wins and losses for each pair among the five methods. In the TEA numerical simulations, we set $\varepsilon = 1$ and $\eta = \sqrt{8 \cdot \frac{\lambda^2 - 1}{\lambda^8 - 1} \cdot \ln M}$. We also set $\lambda = \rho^{-1}$ and $\rho = 0.9$. The violin sound[19] results are shown in Figs. 2d and 2e, and Tables 1. For this dataset, our method and the persistence prediction produce much better results than the other methods. Therefore, we next compare our TEA method with the persistence prediction with respect to the number of experts. We use the binominal test for the analysis, i.e., if the number of wins is greater (smaller) than 531 (469), the method is significantly superior (inferior) to the other method with respect to the 95% confidence level two-sided binominal test. When the number of experts is large, our TEA method is significantly superior to the persistence prediction, as shown in Fig. 2e and Table 1. In the example of squid giant axon[20], the proposed TEA is also better than the other four methods when the number of experts is large, especially when greater than or equal to 87, as shown in Fig. 3 and Table 1.

In conclusion, our TEA method tends to provide the best prediction when the number of experts is large. The precise number of experts for which this is the case may change depending on the given data, the length of targets, and the decay parameter.

Distribution prediction to the mathematical models. We applied the distribution prediction to time-series of the Hénon map. The distribution prediction will be explained in the later Method section. The setup is similar to that for the point prediction, except that we provide the prediction as a distribution. The results are presented in Figs. 2f and 2g. The width of the distribution prediction is narrow immediately after the learning period (Fig. 2f), then grows gradually as the number of prediction steps increases because of the instability of the underlying dynamics. The predicted confidence interval tends to contain the actual values. When we increase the number of points used for prediction, the width of the distribution prediction becomes narrower (Fig. 2g). We use values of $S = 1{,}000$ and $M = 100$ in Figs. 2f and 2g. In the TEA numerical simulations, we set $\varepsilon = 1$ and $\eta = \frac{1}{\lambda}\sqrt{8 \cdot \frac{\lambda^2 - 1}{\lambda^8 - 1} \cdot \ln MS}$. The number of trials is 40 in each box in Fig. 2g. Restricting the range of λ to $1 < \lambda < 2$ gives a better prediction. We generate the target and experts' time-series as in the previous section. We also obtain the distribution prediction of the Ikeda map using $S = 1{,}000$ and $M = 100$, as shown in Fig. S1f. The result is very similar to that of the Hénon map. Again, restricting the range of λ to $1 < \lambda < 2$ gives a better prediction.

Mathematical models of prostate cancer. TEA can be applied to clinical problems, such as the prediction of prostate-specific antigen (PSA) after some initial treatments, while waiting to start an additional treatment. We apply TEA to the prediction of tumor markers for prostate cancer PSA. Before the technical details, we introduce a mathematical model of prostate cancers in this section.

Patients had already received radical prostatectomy as an initial treatment. Then, clinicians followed postoperative PSA levels to determine when to commence salvage treatment. Although the timing at which patients start salvage treatment is an important problem, there is no definitive agreement on when this to be started. Currently, clinicians are determining the start of salvage treatment based on their discretion. The clinical part of this study was approved by the ethics committees of Jikei University School of Medicine and The University of Tokyo. All patients provided written informed consent. Cancer cells tend to thrive under an androgen-rich environment. Meanwhile, lowering androgen levels makes cancer cells grow slowly or rather decline. Because of this characteristics, clinicians suppress the androgen concentration with hormone therapy. However, when cancer cells remain exposed to an androgen-poor environment, they often acquire the ability to grow without androgen. This growth signals a cancer relapse. Intermittent androgen suppression was proposed to delay the relapse of cancer[24]. In intermittent androgen suppression, we start hormone therapy, but stop when PSA levels have decreased sufficiently. Then, we wait until PSA increases and reaches a threshold value. After reaching this threshold, we resume hormone therapy. We repeat this process to delay the relapse. However, clinical trials show that the effects of intermittent androgen suppression depend on individual patients, and are limited[25,26].

Here, we use a mathematical model[8] of intermittent androgen suppression for prostate cancer[24–26]. This model was constructed based on data of Canadian patients[25,26] whose PSA had increased to some extent after radiation therapy, and were later treated by intermittent androgen suppression. Because the model of Ref. 8 has a small number of parameters, it is reasonable to predict the future PSA values with this simple model and very short time-series, although several mathematical models have been proposed to describe dynamics under intermittent androgen suppression[4–6,8,10,27–31]. In the model described in Ref. 8, we assume that there are three classes of cancer cells: androgen dependent cancer cells x_1, androgen independent cancer cells generated through reversible changes x_2, and androgen independent cancer cells generated through irreversible changes x_3. When the hormone therapy is underway, x_1 may change to x_2 or x_3. When the hormone therapy is stopped, x_2 may return to x_1, whereas x_3 cannot return to x_1 or x_2 because of genetic mutation. We previously verified two important properties of this model: namely, a piecewise linear model is sufficient to describe the dynamics of PSA, and the androgen concentration need not be explicitly included in the model[8]. Based on these verified properties, we can simply construct the mathematical model as

$$\frac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} = \begin{pmatrix} d_1 & 0 & 0 \\ d_2 & d_3 & 0 \\ d_4 & d_5 & d_6 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}, \qquad (22)$$

for the on-treatment period, and

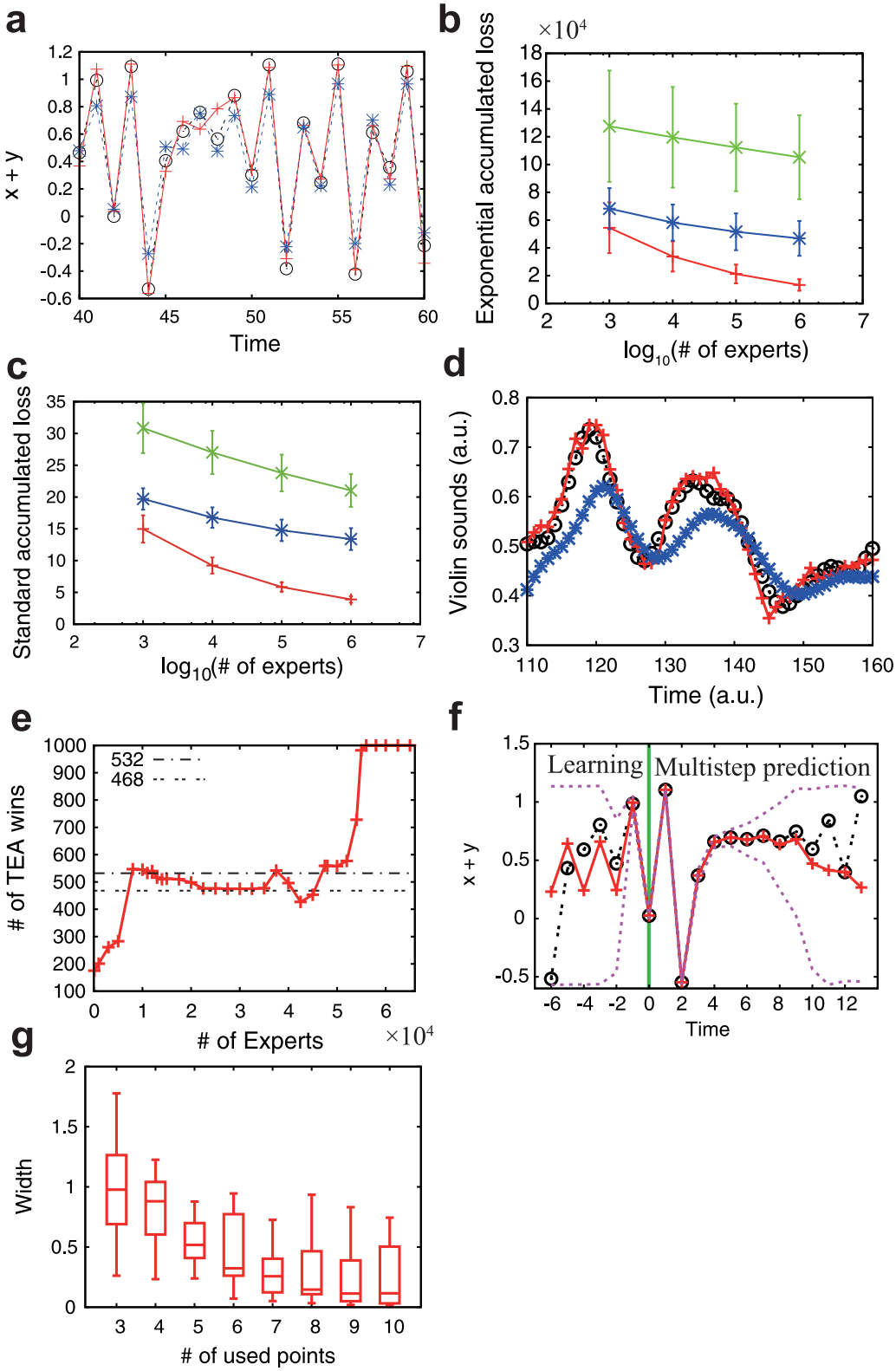

Figure 2 | Examples of prediction by TEA. (a) Online point prediction of the Hénon map. (b, c) The number of experts versus the exponential accumulated loss of errors and the standard accumulated loss of errors in the prediction of the Hénon map. (d) Online point prediction of violin sounds. (e) The number of experts versus the number of wins out of 1,000 online prediction trials against the persistence prediction in the violin example. (f) Distribution prediction of the Hénon map. (g) Confidence interval width against the number of used points shown in box plots. In panels (a) and (f), actual observations are shown with black ○ and dotted lines, TEA predictions are shown with red + and solid lines, and the prediction given by the method of Chernov and Zhdanov is shown with blue * and dashed lines. In panel (f), the purple dotted lines show the 95% confidence interval of the distribution prediction, and the green line divides the learning period and the multistep prediction period. In panels (b) and (c), red, blue, and green error bars correspond to TEA, the method of Chernov and Zhdanov, and the standard expert advice, respectively. In panel (e), the dotted lines show the 95% confidence interval under the null-hypothesis of even chance.

Table 1 | Analysis of violin sounds and the membrane potential of squid giant axon. The number of wins between each pair of the five prediction methods is shown. The each number indicates the number of experts in each case.

Method	Method of Comparison									
#	100 (Violin sounds)					1000 (Violin sounds)				
	TEA	CZ	Exist	Persistence	Average	TEA	CZ	Exist	Persistence	Average
TEA	—	1000	509	175	1000	—	1000	785	201	1000
CZ	0	—	0	0	1000	0	—	21	0	1000
Exist	491	1000	—	19	1000	215	979	—	79	1000
Persistence	825	1000	981	—	1000	799	1000	921	—	1000
Average	0	0	0	0	—	0	0	0	0	—
#	30000 (Violin sounds)					55000 (Violin sounds)				
	TEA	CZ	Exist	Persistence	Average	TEA	CZ	Exist	Persistence	Average
TEA	—	1000	1000	475	1000	—	1000	998	982	1000
CZ	0	—	0	0	1000	0	—	0	0	1000
Exist	0	1000	—	0	1000	2	1000	—	2	1000
Persistence	525	1000	1000	—	1000	18	1000	998	—	1000
Average	0	0	0	0	—	0	0	0	0	—
#	30 (Squid axon)					90 (Squid axon)				
	TEA	CZ	Exist	Persistence	Average	TEA	CZ	Exist	Persistence	Average
TEA	—	17	50	120	88	—	80	115	120	120
CZ	103	—	120	120	115	40	—	103	120	120
Exist	70	0	—	120	2	5	17	—	120	59
Persistence	0	0	0	—	0	0	0	0	—	0
Average	32	5	118	120	—	0	0	61	120	—
#	120 (Squid axon)					180 (Squid axon)				
	TEA	CZ	Exist	Persistence	Average	TEA	CZ	Exist	Persistence	Average
TEA	—	109	120	120	120	—	120	120	120	120
CZ	11	—	105	120	120	0	—	113	120	120
Exist	0	15	—	120	89	0	7	—	120	117
Persistence	0	0	0	—	0	0	0	0	—	0
Average	0	0	31	120	—	0	0	3	120	—

$$\frac{d}{dt}\begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} = \begin{pmatrix} e_1 & e_2 & 0 \\ 0 & e_3 & 0 \\ 0 & 0 & e_4 \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}, \qquad (23)$$

for the off-treatment period[8]. Here, d_1, d_2, d_3, d_4, d_5, d_6, e_1, e_2, e_3, and e_4 are model parameters. We assume that a PSA measurement is represented by $x_1 + x_2 + x_3$ for simplicity. Thus, we must specify these 10 parameters for the dynamics and three other parameters for the initial conditions of x_1, x_2, and x_3. If we try to find these 13 parameters directly only from a single target patient, we would need to obtain a long time-series. The application of the proposed TEA algorithm makes the required observation period of PSA measurements shorter by integrating observations from the target patient with the long time-series data of PSA measurements obtained from previous prostate cancer patients. We note that we only analyze the off-treatment period in this paper, because the target dataset is about the follow-up period after an initial treatment. Therefore, we need 4 control parameters and initial conditions.

Construction of experts for prediction of PSA for prostate cancer. In this paper, we have two datasets; one is a dataset of Canadian patients with many data points; the other is a dataset of Japanese patients with short time points. We need a long time-series to efficiently estimate model parameters. Therefore, we select Canadian datasets for estimation of parameters and Japanese datasets for predicting targets.

In applying TEA to prostate cancer, we first prepared 72 sets of model parameters, each of which was obtained from one of 72 Canadian prostate cancer patients treated with intermittent androgen suppression. These parameters were obtained from Ref. 8. We note that our prediction target dataset corresponds to the off-treatment period in the model[8]. Second, we chose the number of observation points to use as known data points. This must be at least three because of the model dimensions[8]. Third, using each set of parameters, we determined the initial model state to minimize the fitting error between the initial three or more PSA measurements and the model output. The optimal initial conditions were selected by minimizing the following cost function:

$$\sum_{k=1}^{K}\left(\left|\sum_{j=1}^{3} x_j(t_k) - y(t_k)\right| + \sum_{j=1}^{3} h\left(x_j(t_k)\right)\right), \qquad (24)$$

where $h(x) = 10^{15}(1-x)$ for $x < 0$ and $h(x) = 0$ for $x \geq 0$, where t_k is the kth observation time. We denote the number of observation points used for learning by K. The method of obtaining the initial conditions was similar to that in Ref. 8. Fourth, we ran the model with each set of parameters and the corresponding initial conditions to construct the database of experts $f_{i,t}$; thus, we have 72 experts.

Estimation of learning parameters. We applied the second step of the TEA algorithm to determine the weights of the PSA measurements. Then, we applied the third step of the TEA algorithm to obtain the distribution prediction. We determined the optimal decay rate λ by minimizing the error between the last learning observation and the prediction. We restricted the range of λ to $1 < \lambda < 2$ to obtain better predictions. The standard deviation σ is estimated as follows. We ran the distribution prediction with the obtained initial conditions and the decay parameter. We set the standard deviation σ to the mean of the absolute errors between the median of the distribution prediction and the corresponding observation when the mean is taken during the learning period.

Application of TEA to prediction of PSA for prostate cancer. We predict the values of PSA with distribution prediction. The distribution prediction of PSA with TEA is shown in Fig. 4. Here we evaluate the larger side of the predicted distribution, because overlooking high PSA is highly undesirable in a real clinical setting. We show seven points $u_t(Q)$ of the predicted distribution (97.5%, 87.5%, 75%, 65%, 60%, 55%, and 52.5%) in these figures, where $u_t(Q)$ is defined as

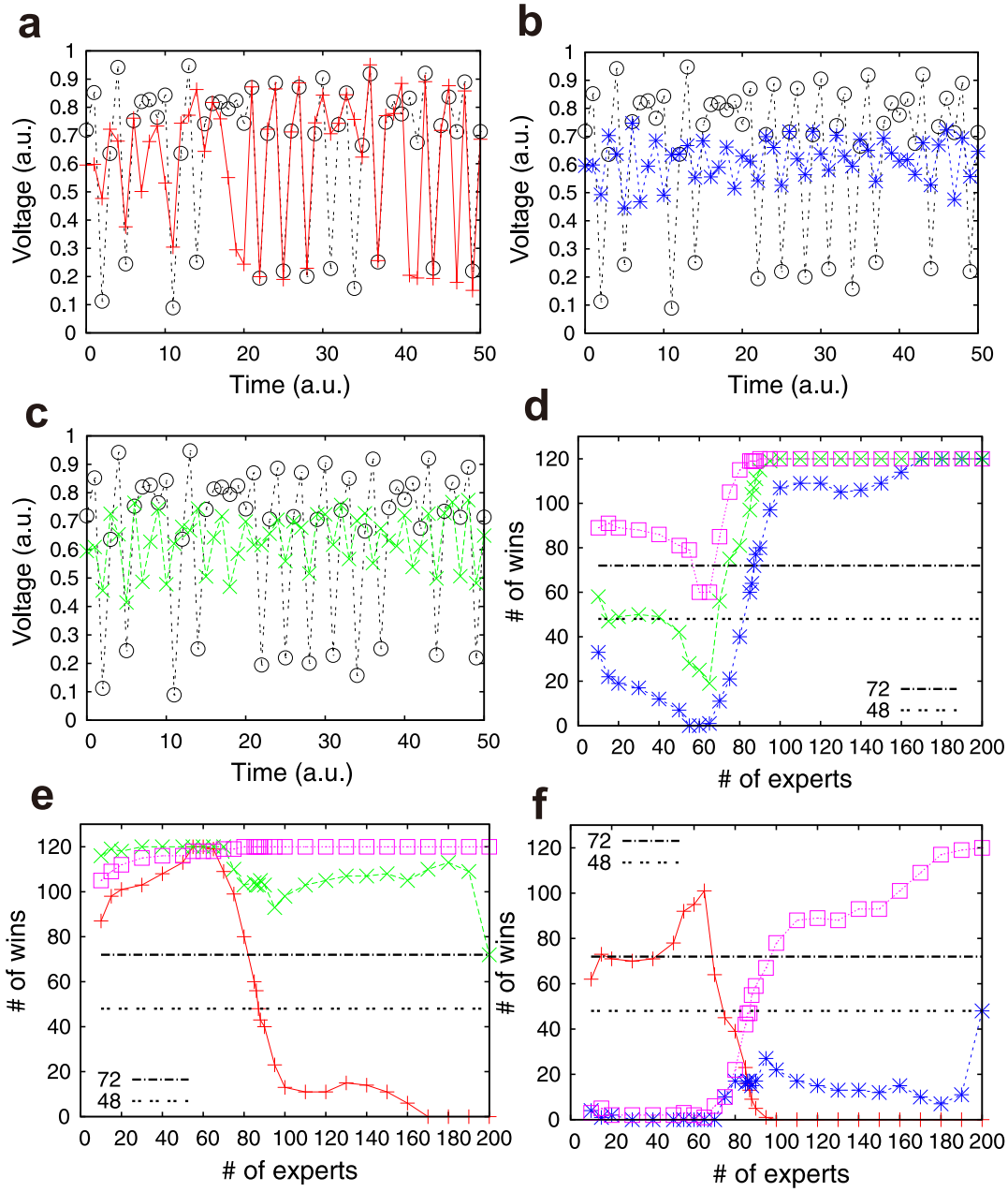

Figure 3 | Online time-series prediction of membrane potential of squid giant axon. The decay parameter was fixed at $\rho = 0.9$. (a, b, c) The observed time-series is represented by black lines. The number of experts is 160. The predicted time-series represented by a red dashed line in (a), a blue dashed line in (b), and a green dashed line in (c) are obtained by our method, CZ, and the standard expert advice, respectively. (d) The number of times our method outperforms CZ (blue *), the standard expert advice (green ×), and the average prediction (purple □) (out of 120) are shown. When the number of wins for a method is greater than 71 (the dashed-dotted horizontal line), it is significantly better than the other method in terms of the binominal test. When it is smaller than 49 (the double-dotted line), the opposite is true. (e) The number of times the CZ method outperforms our method (red +), the standard expert advice (green ×), and the average prediction (purple □). (f) The number of times the standard expert advice outperforms our method (red +), CZ (blue *), and the average prediction (purple □).

$$\int_0^{u_t(Q)} \breve{p}_t(x)dx = Q. \tag{25}$$

Note that Q is the intended value of the probability, i.e. 0.975, 0.875, 0.75, 0.65, 0.6, 0.55, and 0.525, respectively, in this situation. We obtained the proportion of PSA values that are less than the intended probability for each Q, and counted the PSA data points that are next to the final data points in the learning period; namely, if we are using three data points for learning, we count the fourth data point. In this paper, we focus on the predictability of the next point. The results are summarized in Table 2. Note that TEA can predict not only the next data point, but also those far in the future. We predicted the future PSA values for 88, 86, 80, and 69 patients when we used the first three, four, five and six time points, respectively. We also conducted numerical simulations using CZ and the standard expert advice. The predicted distributions were different for each method, as shown in Fig. 5. In numerical simulations, we set $\varepsilon = 1$. We arrange the learning rate as four constant values $\eta = \frac{1}{\lambda}\sqrt{8\frac{1-\lambda^2}{1-\lambda^{2\nu}}\ln M}$ with $\nu = 1, 2, 3$, and 4. In addition, we increase ν as the number of the learning points increases. We note that $M = 72$ is the number of experts. We also arrange the learning

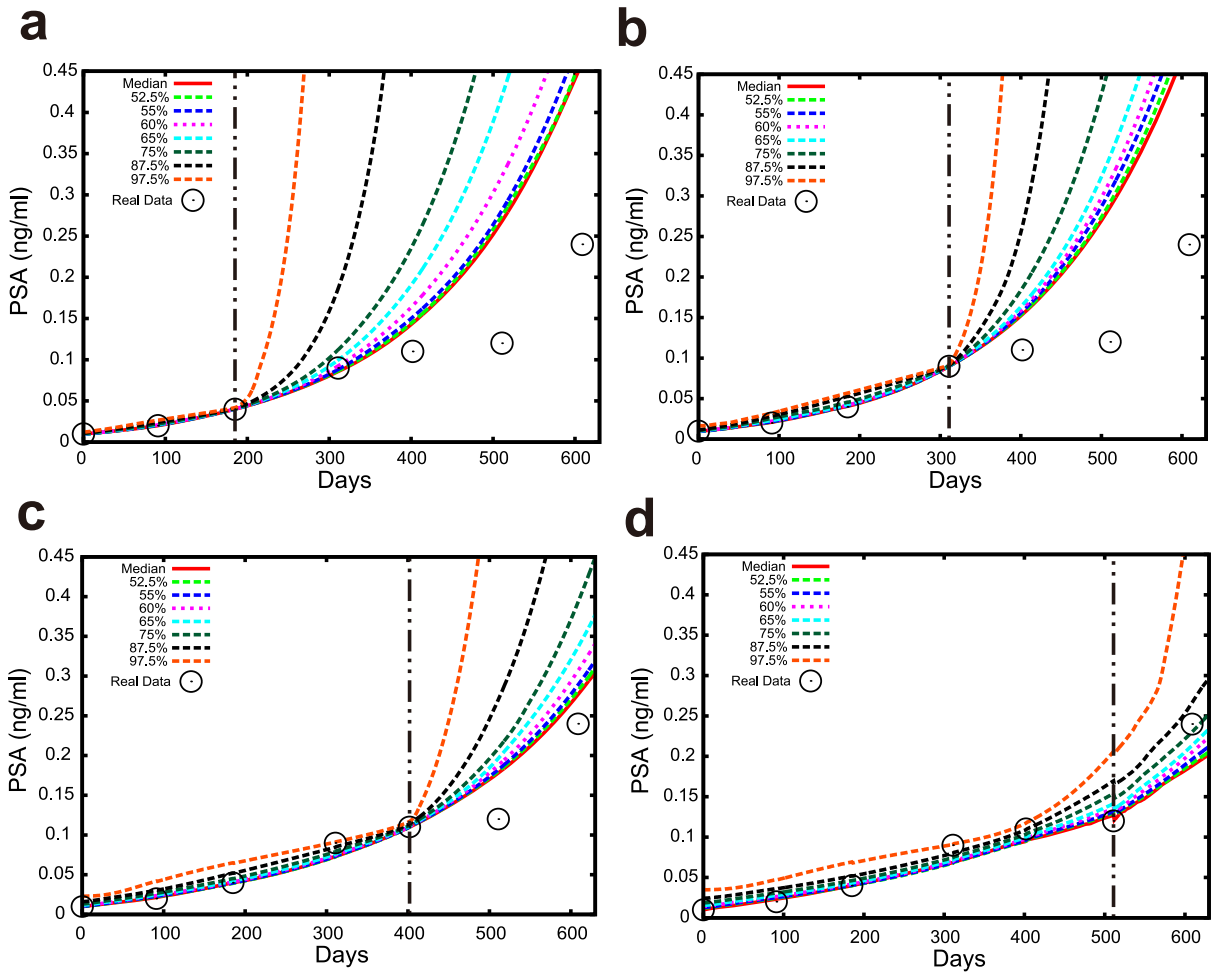

Figure 4 | Results of the distribution prediction of PSA in a patient. The black ○ shows the actual PSA observations. We used the (a) first three, (b) four, (c) five, and (d) six observation points to predict the next data points for the patient with the TEA($v = 1$). The red curve shows the median distribution prediction. Other dotted lines indicate as denoted in each figure. The same patient data are used in all the figures.

rate as $\eta = \sqrt{8\frac{\ln M}{v}}$ for the standard expert advice and $\eta = 2\sqrt{\frac{1-\rho}{1-\rho^v}\ln M}$ for the CZ method.

TEA exhibits the best performance among the three methods, because each proportion tends to be closest to the specified value of Q. These results imply that our proposed prediction method may be reasonable for real applications in a clinical setting. We also checked the prediction performance in terms of the median using the mean absolute error (MAE) as summarized in Supplementary Table S1. As a result, TEA shows the best performance in the meaning of the average MAE among the four cases. We note that the CZ method showed good performance in terms of the root mean square error (RMSE), however, we believe that the MAE suits our situation because we employed the absolute error function for the learning period.

Discussion

In general, clinicians provide a salvage treatment with patients who had recurrence after surgery. Although many studies show the clinical benefit of a salvage treatment for patients with prostate cancer, current studies have reported that an earlier salvage treatment, especially for local recurrence, could improve clinical outcomes[32]. These results suggest that post-operative patients with lower PSA values may have a higher frequency of local recurrence that could be efficiently treated by radiotherapy. If clinicians can accurately assess the PSA failure at an earlier stage than the present standard criterion of PSA failure which is that the PSA value increases to 0.2 (ng/ml) or more after surgery, salvage treatments could be more effectively scheduled for each patient, improving the final clinical outcome[33]. However, there is still no standardized criterion to determine the best timing of salvage treatments[32,33]. Combined with a mathematical model[8], TEA or its further extensions may be able to potentially predict the PSA dynamics in patients before PSA failure. Therefore, the proposed TEA could become the basis of a new standard index for earlier prediction of PSA failure using a simple mathematical solution, that offers important information for a suitable salvage treatment after surgery[7,34,35].

The more experts we use, the more (numerically) accurate the prediction tends to become (Figs. 2b, 2c, and 2e); in this sense, the accumulation of datasets is important. Additionally, the longer the learning period, the more accurate the TEA prediction tends to become in the toy examples (Fig. 2g). This could be because the toy examples have bounded unstable dynamics. The prediction error does not monotonically decrease with an increase of the learning data points in the example of prostate cancer (Tables 2 and S1), because PSA tends to increase monotonously in time. TEA exhibits the best performance in our analyses. The proposed combination of the expert advice with a predicted distribution enhances the reliability of prediction. This is important in many applications, and especially in medicine.

In summary, we have demonstrated that TEA can infer the state of a target system, by combining its short time-series and the expert

Table 2 | Prediction results for real PSA datasets. The proportions of PSA data points that are followed by the TEA, CZ, and Existing prediction are shown against Q%, points of the predicted distribution from below. A learning rate η is changed with v. An abbreviation t.v. indicates time-varying. We consider data points that are next to the final point used for the learning period. We underline the best method in each case.

		TEA					CZ					Exist				
CI Q	Used points	t.v.	$v=1$	2	3	4	t.v.	1	2	3	4	t.v.	1	2	3	4
97.5	3	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	4	100	**98.8**	100	100	100	100	100	100	100	100	100	100	100	100	100
	5	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	6	**100**	92.8	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
87.5	3	**89.8**	92	90.9	**89.8**	**89.8**	**89.8**	**89.8**	**89.8**	**89.8**	**89.8**	**89.8**	92	**89.8**	**89.8**	**89.8**
	4	**93**	94.2	94.2	**93**	**93**	**93**	**93**	94.2	**93**	**93**	**93**	94.2	94.2	**93**	**93**
	5	98.8	**96.2**	97.5	98.8	**96.2**	98.8	98.8	98.8	98.8	98.8	98.8	97.5	97.5	98.8	98.8
	6	98.6	**87**	91.3	97.1	97.1	98.6	98.6	98.6	98.6	98.6	98.6	95.7	95.7	97.1	97.1
75	3	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**	**75**
	4	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**	**84.9**
	5	91.2	**87.5**	90	90	90	91.2	91.2	91.2	91.2	91.2	90	90	90	90	90
	6	89.9	**75.4**	78.3	84.1	85.5	91.3	89.9	89.9	89.9	91.3	87	87	87	87	87
65	3	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**	**69.3**
	4	**79.1**	80.2	80.2	80.2	80.2	**79.1**	**79.1**	**79.1**	**79.1**	**79.1**	80.2	80.2	80.2	80.2	80.2
	5	88.8	**82.5**	87.5	88.8	88.8	88.8	88.8	88.8	88.8	88.8	88.8	87.5	87.5	88.8	88.8
	6	84.1	**72.5**	**72.5**	76.8	78.3	82.6	82.6	84.1	84.1	84.1	84.1	78.3	79.7	81.2	81.2
60	3	**67**	68.2	68.2	**67**	**67**	**67**	**67**	**67**	**67**	**67**	**67**	68.2	**67**	**67**	**67**
	4	79.1	**77.9**	79.1	79.1	80.2	**77.9**	**77.9**	**77.9**	**77.9**	**77.9**	80.2	79.1	79.1	79.1	79.1
	5	85	**81.2**	83.8	85	85	86.2	85	85	86.2	86.2	85	83.8	85	85	85
	6	78.3	**71**	72.5	75.4	76.8	79.7	78.3	78.3	79.7	79.7	78.3	75.4	76.8	78.3	78.3
55	3	65.9	**64.8**	**64.8**	**64.8**	65.9	65.9	65.9	65.9	65.9	65.9	65.9	**64.8**	**64.8**	**64.8**	65.9
	4	**73.3**	**73.3**	**73.3**	**73.3**	74.4	**73.3**	74.4	**73.3**	**73.3**	**73.3**	**73.3**	74.4	74.4	74.4	74.4
	5	82.5	**78.8**	**78.8**	82.5	82.5	81.2	82.5	82.5	82.5	82.5	82.5	81.2	81.2	82.5	82.5
	6	75.4	**71**	72.5	72.5	73.9	76.8	72.5	73.9	73.9	75.4	75.4	75.4	73.9	73.9	73.9
52.5	3	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**	**64.8**
	4	70.9	73.3	72.1	70.9	**69.8**	70.9	70.9	70.9	70.9	70.9	72.1	74.4	73.3	72.1	72.1
	5	80	**76.2**	77.5	80	80	78.8	80	80	80	78.8	80	78.8	80	80	80
	6	72.5	**69.6**	**69.6**	71	72.5	75.4	71	72.5	72.5	73.9	72.5	73.9	72.5	72.5	73.9

advice constructed as a collection of longer time-series. The proposed TEA may be applied to any problems where a short time-series and its database are given, as demonstrated in the violin and squid giant axon examples, although we primarily intend to apply TEA in clinical settings, such as inferring the state of a disease using a short time-series from the target patient and longer time-series from previous patients with the same disease. We hope that TEA improves the overall survival and/or quality of life for patients.

Methods

Standard expert advice method. The expert advice method[15] is an online predictor in machine learning. We briefly introduce the standard expert advice method in this section. See the book of Cesa-Bianchi and Lugosi[15] for a more detailed introduction. The expert advice consists of experts and a predictor. At each time step, each expert gives the prediction on the future. The predictor makes a prediction for the future by weighting these pieces of advice based on the experts' prediction history. After a new outcome is observed, the predictor updates the experts' weights using the losses produced in the current step. We iterate these steps to realize online prediction. Let $f_{i,t}$ be the ith expert's advice at time t and N be the number of experts. We assign each expert the weight $w_{i,t}$ at time t, and obtain the prediction by averaging the experts' advice as

$$p_t = \frac{\sum_{i=1}^{N} w_{i,t-1} f_{i,t}}{\sum_{i=1}^{N} w_{i,t-1}}, \tag{26}$$

$$w_{i,t} = \exp(-\eta L_{i,t}), \tag{27}$$

where p_t is the prediction at time t, η is a constant, and $L_{i,t}$ is the accumulated loss for the ith expert at time t. Better experts have smaller accumulated losses, and hence have larger weights. The accumulated losses for the ith expert and the predictor at time t are

$$L_{i,t} = \sum_{k=1}^{t} l(f_{i,k}, y_k), \tag{28}$$

$$L_t = \sum_{k=1}^{t} l(p_k, y_k), \tag{29}$$

where y_k is the observation at time k, and $l(x, y)$ is a convex loss function, typically the absolute error $|x-y|$ or squared error $(x-y)^2$. We evaluate the performance of the predictor by a regret, which is defined as the predictor's accumulated loss minus the accumulated loss for the best expert. Mathematically, the regret R_t is defined as

$$R_t \equiv L_t - \min_{i=1,\ldots,N} L_{i,t} \le \frac{\ln N}{\eta} + \frac{\varepsilon^2 \eta}{8} t, \tag{30}$$

where ε is the maximum value of $|l(\cdot,\cdot)|$. Namely, the regret is bounded above by the right-hand side of Eq. (30) (see Ref. 15 for the derivation). We obtain the following optimal constant η_* by minimizing the upper bound over η:

$$\eta_* = \frac{1}{\varepsilon}\sqrt{8 \ln N / t}. \tag{31}$$

When replacing η with η_*, we obtain the optimal upper bound of the regret R_t as $\varepsilon\sqrt{t \ln N/2}$. We call the accumulated losses defined in Eqs. (28) and (29) the *standard accumulated losses*.

Although the standard expert advice can be applied in many cases, the method is not suited to the prediction of unstable systems, in which the recent history should be emphasized to predict the future more accurately. Thus, we extended the standard expert advice by placing greater weights on recent past information. We call our extension the temporal expert advice, or TEA.

Distribution prediction. Here, we extend the TEA method for point prediction to the prediction of distribution, so that we can handle the prediction of biomarkers. For this purpose, we introduce the distribution prediction of the ith expert at time t $\breve{p}_{i,t}(x)$ as follows:

$$\breve{p}_{i,t}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-f_{i,t})^2}{2\sigma^2}\right), \tag{32}$$

where σ is the standard deviation. This distribution is given under the assumption that a point prediction $f_{i,t}$ is disturbed by various errors and that the error is normally

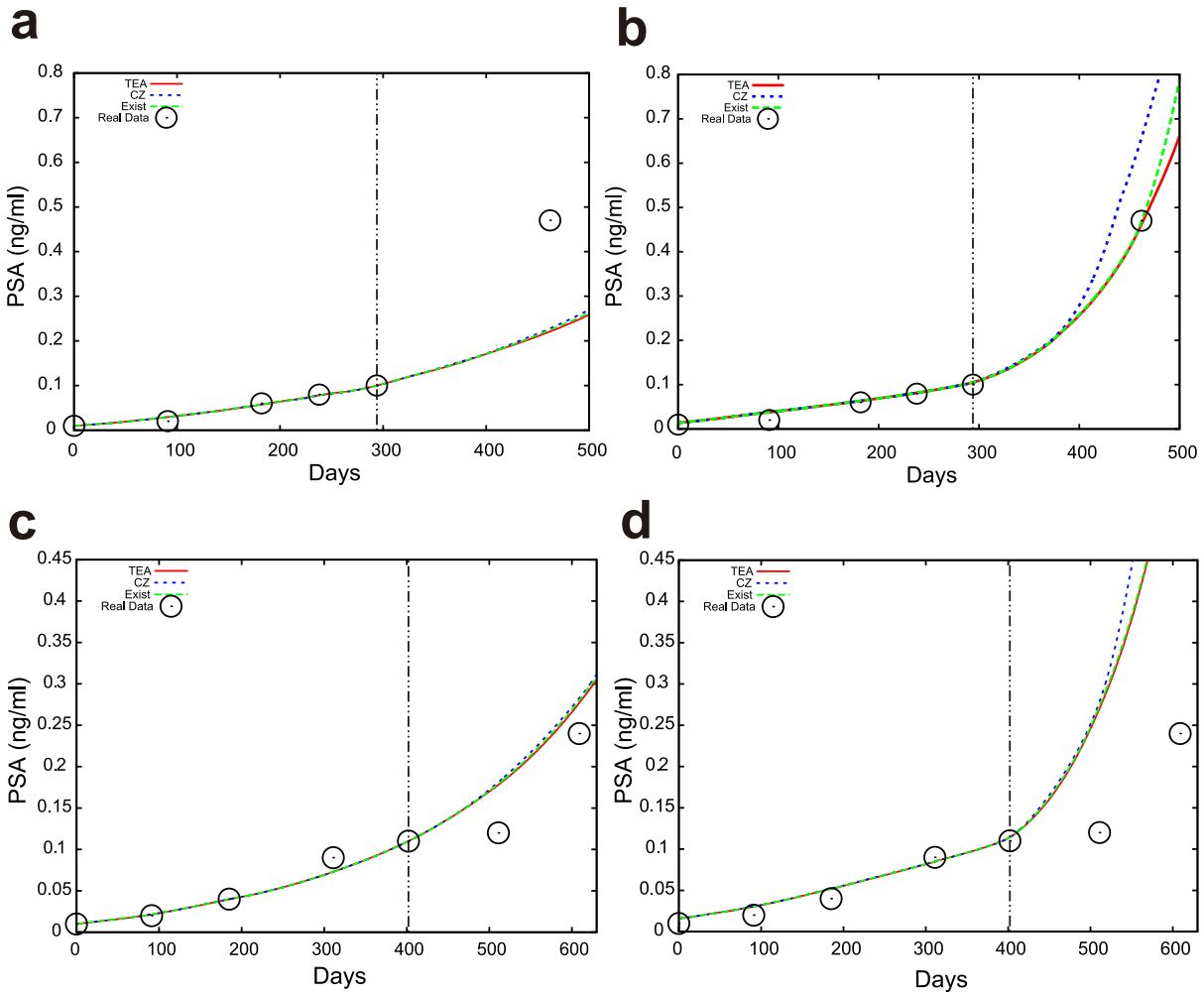

Figure 5 | Examples of PSA predictions using TEA (red), CZ (blue), and the standard expert advice (green). The black ○ shows the actual PSA observations. The red, blue, and green curves were obtained from TEA, the Chernov-Zhdanov method, and the standard expert advice, respectively. We show the median in (a) and (c). We chose 87.5% points of the predicted distributions in (b) and (d). The same patient data are used in (a) and (b). Other patient data are used in (c) and (d).

distributed around the point prediction. Then, the predictor for the distribution prediction is given by

$$\hat{p}_t(x) = \frac{\sum_{i=1}^{N} w_{i,t-1}\breve{p}_{i,t}(x)}{\sum_{i=1}^{N} w_{i,t-1}}. \quad (33)$$

We determine the optimal decay rate λ by minimizing the absolute error between the final learning point and the corresponding observation point. The standard deviation σ is set to be the absolute difference between the point prediction $\hat{f}_\tau$ and the observation y_τ at the final learning point under the estimated λ above as $\sigma = |y_\tau - \hat{f}_\tau|$, where τ is the number of the learning points. We note that we determined these parameters with a modified way in the prediction of PSA because of its instability.

1. Nowak, M. A. *et al.* Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963–969 (1991).
2. Jackson, T. L. A mathematical model of prostate tumor growth and androgen-independent relapse. *Discrete Contin. Dyn. Syst. Ser. B* **4**, 187–201 (2004).
3. Michor, F. *et al.* Dynamics of chronic myeloid leukaemia. *Nature* **435**, 1267–1270 (2005).
4. Ideta, A. M., Tanaka, G., Takeuchi, T. & Aihara, K. A mathematical model of intermittent androgen suppression for prostate cancer. *J. Nonlinear Sci.* **18**, 593–614 (2008).
5. Jain, H. V., Clinton, S. K., Bhinder, A. & Friedman, A. Mathematical modeling of prostate cancer progression in response to androgen ablation therapy. *Proc. Natl. Acad. Sci. USA* **108**, 19701–19706 (2011).
6. Portz, T., Kuang, Y. & Nagy, J. D. A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy. *AIP Adv.* **2**, 011002 (2012).
7. Draisma, G. *et al.* Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J. Natl. Cancer Inst.* **101**, 374–383 (2009).
8. Hirata, Y., Bruchovsky, N. & Aihara, K. Development of a mathematical model that predicts the outcome of hormone therapy for prostate cancer. *J. Theor. Biol.* **264**, 517–527 (2010).
9. Kronik, N. *et al.* Predicting outcomes of prostate cancer immunotherapy by personalized mathematical models. *PLoS ONE* **5**, e15482 (2010).
10. Hirata, Y., Akakura, K., Higano, C. S., Bruchovsky, N. & Aihara, K. Quantitative mathematical modeling of PSA dynamics of prostate cancer patients treated with intermittent androgen suppression. *J. Mol. Cell Biol.* **4**, 127–132 (2012).
11. Gorelik, B. *et al.* Efficacy of weekly docetaxel and bevacizumab in mesenchymal chondrosarcoma: a new theranostic method combining xenografted biopsies with a mathematical model. *Cancer Res.* **68**, 9033–9040 (2008).
12. Suzuki, T., Bruchovsky, N. & Aihara, K. Piecewise affine systems modelling for optimizing hormone therapy of prostate cancer. *Philos. Trans. R. Soc. Lond. A* **368**, 5045–5059 (2010).
13. Hirata, Y., di Bernardo, M., Bruchovsky, N. & Aihara, K. Hybrid optimal scheduling for intermittent androgen suppression of prostate cancer. *Chaos* **20**, 045125 (2010).
14. Chmielecki, J. *et al.* Optimization of dosing for EGFR-mutant non-small cell lung cancer with evolutionary cancer modeling. *Sci. Transl. Med.* **3**, 90ra59 (2011).
15. Cesa-Bianchi, N. & Lugosi, G. *Prediction, Learning, and Games* (Cambridge Univ. Press, New York, 2006).
16. Chernov, A. & Zhdanov, F. Prediction with expert advice under discounted loss. *Proc. of* ALT 2010, *Lecture Notes in Artificial Intelligence* **6331**, 255–269 (2010).
17. Hénon, M. A two-dimensional mapping with a strange attractor. *Commun. Math. Phys.* **50**, 69–77 (1976).
18. Ikeda, K. Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Opt. Commun.* **30**, 257–261 (1979).

19. Goto, M. Development of the RWC Music Database. *Proc. 18th Int. Congress on Acoustics (ICA 2004)*, I-553-556 (2004).
20. Mees, A. *et al.* Deterministic prediction and chaos in squid axon response. *Phys. Lett. A* **169**, 41–45 (1992).
21. Hirata, Y., Judd, K. & Aihara, K. Characterizing chaotic response of a squid axon through generating partitions. *Phys. Lett. A* **346**, 141–147 (2005).
22. Hirata, Y. & Aihara, K. Devaney's chaos on recurrence plots. *Phys. Rev. E* **82**, 036209 (2010).
23. Hirata, Y., Oku, M. & Aihara, K. Chaos in neurons and its application: perspective of chaos engineering. *Chaos* **22**, 047511 (2012).
24. Akakura, K. *et al.* Effects of intermittent androgen suppression on androgen-dependent tumors. *Cancer* **71**, 2782–2790 (1993).
25. Bruchovsky, N. *et al.* Final results of the Canadian prospective phase II trial of intermittent androgen suppression for men in biochemical recurrence after radiotherapy for locally advanced prostate cancer: clinical parameters. *Cancer* **107**, 389–395 (2006).
26. Bruchovsky, N., Klotz, L., Crook, J. & Goldenberg, S. L. Locally advanced prostate cancer: biochemical results from a prospective phase II study of intermittent androgen suppression for men with evidence of prostate-specific antigen recurrence after radiotherapy. *Cancer* **109**, 858–867 (2007).
27. Tanaka, G., Hirata, Y., Goldenberg, S. L., Bruchovsky, N. & Aihara, K. Mathematical modelling of prostate cancer growth and its application to hormone therapy. *Philos. Trans. R. Soc. Lond. A* **368**, 5029–5044 (2010).
28. Tanaka, G., Tsumoto, K., Tsuji, S. & Aihara, K. Bifurcation analysis on a hybrid systems model of intermittent hormonal therapy for prostate cancer. *Physica D* **237**, 2616–2627 (2008).
29. Guo, Q., Tao, Y. & Aihara, K. Mathematical modeling of prostate tumor growth under intermittent androgen suppression with partial differential equations. *Int. J. Bifurcat. Chaos* **18**, 3789–3797 (2008).
30. Tao, Y., Guo, Q. & Aihara, K. A model at the macroscopic scale of prostate tumor growth under intermittent androgen suppression. *Math. Models Meth. Appl. Sci.* **19**, 2177–2201 (2009).
31. Tao, Y., Guo, Q. & Aihara, K. A mathematical model of prostate tumor growth under hormone therapy with mutation inhibitor. *J. Nonlinear Sci.* **20**, 219–240 (2010).
32. Pfister, D. *et al.* Early salvage radiotherapy following radical prostatectomy. *Eur. Urol.* **65**, 1034–1043 (2014).
33. King, C. R. The timing of salvage radiotherapy after radical prostatectomy: a systematic review. *Int. J. Radiat. Oncol. Biol. Phys.* 84, 104–111 (2012).
34. Hazelton, W. D. & Luebeck, E. G. Biomarker-based early cancer detection: is it achievable? *Sci. Transl. Med.* **3**, 109fs9 (2011).
35. The U. S. Preventive Services Task Force, Screening for Prostate Cancer: U. S. Preventive Services Task Force Recommendation Statement. http://www.uspreventiveservicestaskforce.org/uspstf12/prostate/prostateart.htm (2012), Date of access: 04/01/2015.

Acknowledgments

We would like to express our appreciation to Dr. Nicholas Bruchovsky for valuable discussions and sharing published clinical data. This work is partially supported by JSPS KAKENHI Grant Number 11J07088, by MEXT KAKENHI Grant Number 23240019, by JST-CREST, and by the Aihara Innovative Mathematical Modelling Project, the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)", initiated by the Council for Science and Technology Policy (CSTP). The violin data used in this study is available in the RWC Music Database (Musical Instrument Sound).

Author contributions

N.H. and S.E. designed the clinical study. K.M., Y.H. and K.A. designed the rest of the study. K.M., Y.H., R.T., H.K., K.Y. and K.A. created the theoretical method. K.M. and Y.H. analyzed the data. N.H. and S.E. obtained the clinical data and suggested the clinical implications. K.M., Y.H. and K.A. wrote the manuscript. All authors checked the manuscript and agreed to submit the final version of the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: S.E. declares competing financial interests: supports from Takeda Pharmaceutical Co., Astellas, and AstraZeneca. The other authors declare no competing financial interests.

How to cite this article: Morino, K. *et al.* Predicting disease progression from short biomarker series using expert advice algorithm. *Sci. Rep.* **5**, 8953; DOI:10.1038/srep08953 (2015).