Inverse signal importance in real exposome: How do biological systems dynamically prioritize multiple environmental signals?

Thoma Itoh[1,2,3], Yohei Kondo[*4,5], Tomoya Nakayama[6], Ai Shinomiya[2], Kazuhiro Aoki[2,7,8,9], Takashi Yoshimura[5,6,10], Honda Naoki [*3,4,5,11]

[1] Department of Basic Biology, School of Life Science, SOKENDAI (The Graduate University for Advanced Studies), Myodaiji-cho, Okazaki, Aichi, Japan

[2] National Institute for Basic Biology, National Institutes of Natural Sciences, Myodaiji-cho, Okazaki, Aichi, Japan

[3] Graduate School of Integrated Sciences for Life, Hiroshima University, Higashihiroshima, Hiroshima, Japan

[4]Laboratory for Data-driven Biology, Graduate School of Medicine, Nagoya University, Nagoya, Aichi, Japan

[5]Center for One Medicine Innovative Translational Research (COMIT), Nagoya University, Nagoya, Aichi, Japan

[6] Laboratory of Animal Integrative Physiology, Department of Animal Sciences, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Aichi, Japan

[7]Quantitative Biology Research Group, Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, Myodaiji-cho, Okazaki, Aichi, Japan

[8]Laboratory of Cell Cycle Regulation, Graduate School of Biostudies, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto, Japan

[9]Center for Living Systems Information Science, Graduate School of Biostudies, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto, Japan

[10] Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya, Aichi, Japan

[11]Theoretical Biology Research Group, Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, Myodaiji-cho, Okazaki, Aichi, Japan

[*]Corresponding authors: Yohei Kondo, Honda Naoki
E-mail: kondo.yohei.e8@f.mail.nagoya-u.ac.jp, honda.naoki.t1@f.mail.nagoya-u.ac.jp,

Abstract
Living organisms integrate multiple signals from their exposome—the totality of environmental influences experienced throughout life—to adapt to complex, non-stationary environments. While organisms are thought to flexibly prioritize relevant signals depending on context, its regulatory mechanisms remain largely unknown. Laboratory studies with precisely controlled conditions fail to capture this adaptability by isolating organisms from the complex exposome. Here, we developed a machine learning framework, Inverse Signal Importance (ISI), to infer how organisms prioritize external cues from time-series data of environmental factors and physiological responses. We applied ISI to analyze gonadal development in medaka fish under natural outdoor conditions, tracking gonadosomatic index alongside environmental signals including water temperature, day length, and solar radiation over two years. Our analysis revealed that signal importance levels exhibit complex dynamics distinct from simple environmental periodicity and correlates significantly with specific gene expression patterns. Notably, genes associated with temperature-related signal importance display differential expression between outdoor and controlled laboratory conditions, suggesting their role in environmental adaptation. These findings indicate that ISI effectively captures latent physiological dynamics in adaptation of exposome. By decomposing biological responses into deterministic and adaptive components, ISI provides a novel approach to uncover mechanisms of organismal adaptation in natural environments.

Introduction
Living organisms integrate multiple signals and make decisions in an uncertain world. The overwhelming complexity these organisms confront is increasingly characterized through the lens of "exposome", a concept that describes the totality of environmental influences an organism experiences throughout its lifetime [1]. For example, animals demonstrate behavioral flexibility by selectively attending to diverse sensory inputs (e.g., auditory, visual, and olfactory signals) based on their past experiences [2, 3, 4]. In addition, metabolic systems flexibly allocate resources towards processes such as organ growth and immune response to infection depending on the metabolome [5, 6, 7]. Those observations have motivated biologists to elucidate how organisms optimally integrate and/or select information sources from the exposome in an ever-changing world [8, 9, 10, 11].

A vast amount of knowledge has been accumulated about neurologic, metabolic, and molecular mechanisms of the information processing by living organisms [12, 7, 13]. For instance, the primary visual area receives various attributes of visual information and respond to a specific stimulus such as spatial pattern, movement, and color [14]. At the molecular level, cells can detect not only ligand concentrations but also their spatial and temporal patterns and molecular composition [15, 16, 17]. Most studies, however, rely on precisely controlled environments in laboratories, which treat biological systems as something that responds in a uniform way (Fig. 1**a**). In other words, this approach ignores the inherent flexibility of living organisms and their

dynamic adaptation to the exposome, real-world complexities, such as seasonal temperature changes, rainfall, and predators to name a few [18]. Although a few exceptions came from ecology or evolutionary study, there is a scarcity of attempts that have quantitatively assessed the flexibility [19].

Recently we have proposed a data-driven approach to extract how biological decision-making depends on real-world environmental signals [20]. We demonstrated it upon analyzing the robust annual dynamics of gonad size in medaka fish (*Oryzias latipes*) under natural, uncontrolled outdoor conditions. Then we successfully identified the regulatory mechanism of the gonad size, although the prediction for seasonable dynamics was imperfect primarily due to the time-invariant nature of our modeling of the signal-response relationship. In other words, our previous framework assumed that the decision-making system itself does not change over time (Fig. 1**a**). This limitation prevents the framework from accounting for how medaka fish process photoperiod and temperature information differently across seasons, a well-established pattern observed in their natural behavior. [20, 21].

To overcome the limitation, we proposed a novel framework, Inverse Signal Importance (ISI), to address how the system modulates the integration of the multiple signals over time. The ISI method introduces time-varying internal states that represent the importance of corresponding signal modalities (Fig. 1**b**). However, the inference of the signal importance poses a significant challenge because the dynamics of internal states are not directly observable and caused by the complex interplay among the physiology and environment. The ISI method resolves the inverse problem by combining linear regression with a Kalman filter to infer a hybrid model comprising the internal state dynamics and time-invariant filters processing temperature, solar radiation, etc. Here we apply the ISI method to deepen the understanding of the circannual rhythm of gonad size in medaka, aiming to uncover the seasonable change of latent signal importance.

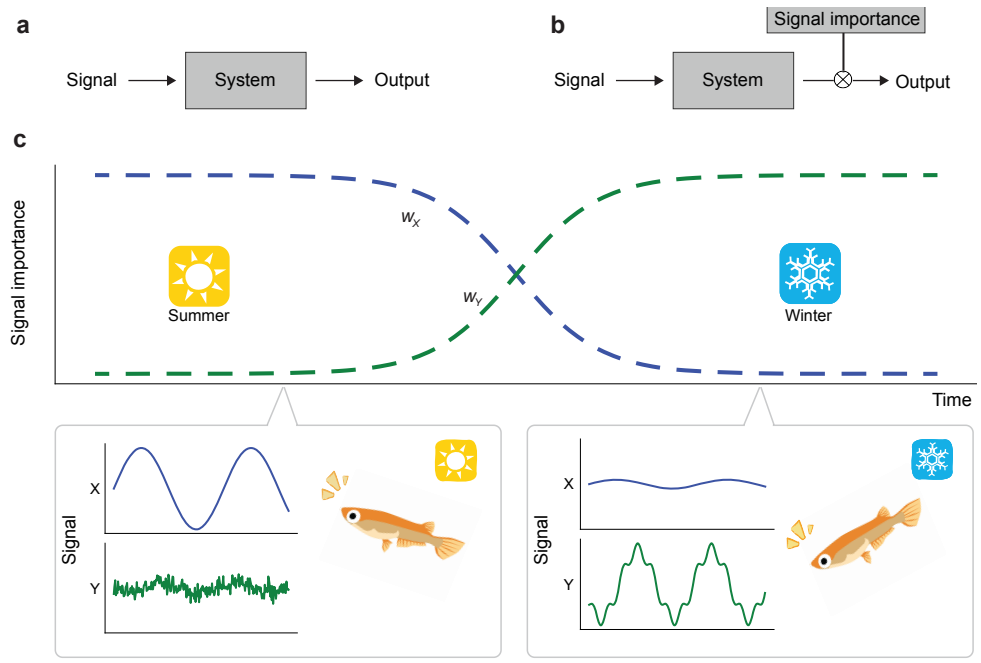

Figure 1. **a** Schematic of basic response function model. **b** Schematic of the proposed model with internal states of signal importance. The signal processed through the linear filter is further modulated by the latent factor. **c** Schematics illustrate the role of signal importance, w, in our framework. The upper panel depicts temporal changes in the signal importance for two signals, X and Y, from summer to winter. We assume that the biological system prioritizes signals containing more beneficial information by amplifying their corresponding signal importance. The lower panels show the dynamics of signals X and Y in summer (left) and winter (right). Based on the seasonal shift in w_X and w_Y, the system prioritizes signal X during summer, while signal Y becomes dominant during winter.

Results

This study aims to quantify the importance of signals by focusing on the gonadal development of medaka fish through the inverse signal importance (ISI) method. Previously, we have characterized living organisms through static models that quantify the relationship between physiological inputs and outputs (Fig 1**a**) [20]; the ISI method extends the framework by introducing internal states for signal importance that weight outputs of mechanistic models (Fig 1**b**). As illustrated in Figure 1**c**, the extension enables us to infer how living systems change which signals to focus on.

Gonad size in medaka under natural outdoor conditions.

We previously measured the gonad size of medaka fish every two weeks over two years (October 2015 to October 2017) under natural outdoor conditions in Okazaki, Japan (Fig.

2**a**) [20]. The time evolution of the gonadosomatic index (GSI), defined as gonad weight as a percentage of body weight, exhibited a noticeable annual rhythm (Fig. 2**a**, bottom panel). In the current study, we trained a predictive model, illustrated in Fig 2**b**, for the differentiation of GSI (dGSI) as the output variable (Fig. 2**c**), while the environmental data (solar radiation, water temperature, and day length) and past GSI values are used as input variables (Fig. 2**a**). Notably, the inclusion of the past GSI value as an input for prediction aligns with widely established practices in machine learning.

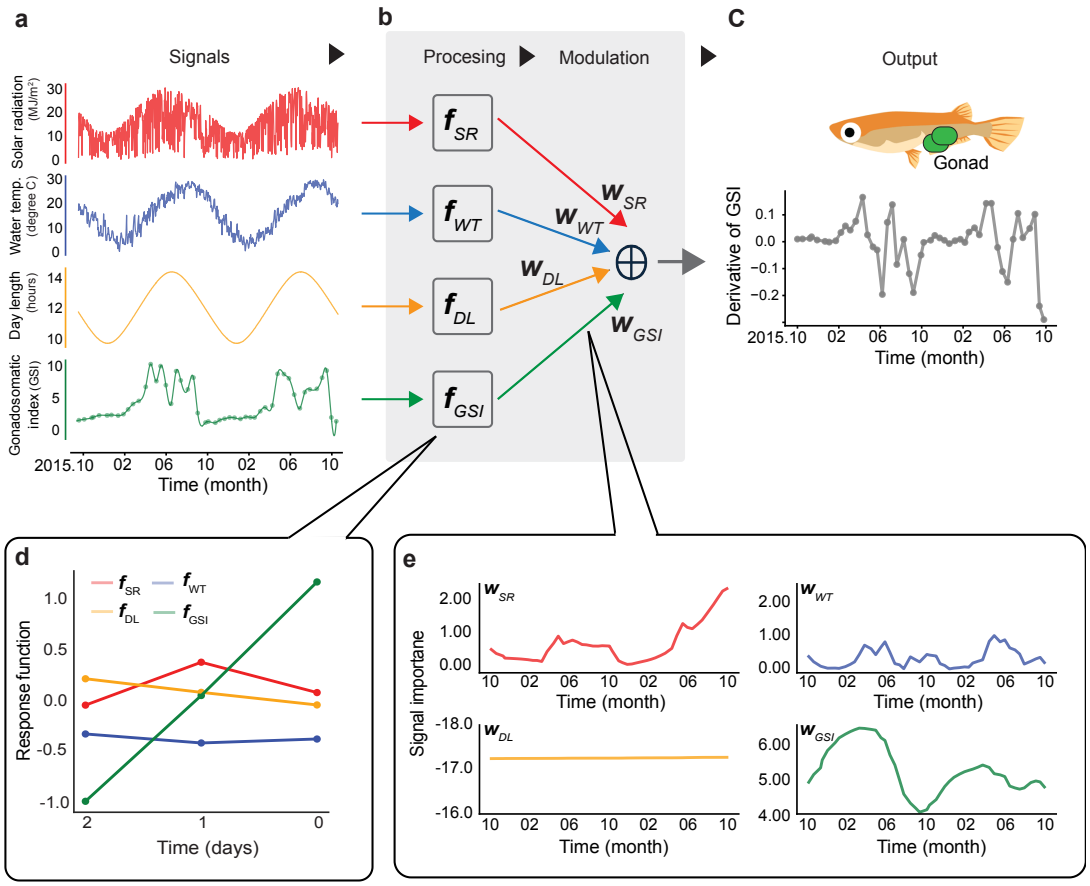

Figure 2. a-c Framework overview. Environmental data and gonad somatic index (**a**) are processed through a linear filter $\boldsymbol{f}$, followed by the modulation of signal importance by multiplication with $\boldsymbol{w}$ (**b**). Each signal has corresponding $\boldsymbol{f}$ and $\boldsymbol{w}$ (SR: solar radiation, WT: water temperature, DL: day length, GSI: gonad somatic index), which are integrated to predict the derivative of GSI at the immediate future time point (**c**). Panel **a** shows the temporal changes in environmental data and gonad somatic index (GSI) as input variables from October 2015 to October 2017, while Panel **c** shows the temporal changes in the derivative of GSI as output variables over the same period. **d**, **e** The estimated variables: the optimal response linear filter (**d**) and the optimal signal importance (**e**). Note the distinct temporal scales between the response function (days) and the signal importance (month).

Signal importance on gonad size regulation of medaka

The ISI framework consists of time-invariant linear filter $\boldsymbol{f}$ and time-variable signal importance $\boldsymbol{w}$(t) (Fig. 2). Each signal modality has corresponding $\boldsymbol{f}$ and $\boldsymbol{w}$ (See Methods and Materials for details). Through the iterative algorithm combining the ridge regression and Kalman filter, we successfully identified the optimal linear filter and signal importance (Supplementary Fig. 2)

The shape of the linear filter is consistent with our previous study [20] showing flat waveforms except for the $\boldsymbol{f}_{GSI}$ (Fig. 2**d**). The $\boldsymbol{f}_{GSI}$ exhibits the characteristics of a differentiator, negatively referencing value of 2 days ago, but positively referencing values of 0 days ago, suggesting that the change of gonad somatic size is highly affected by the past trajectory of gonad somatic size. The optimized trajectories of $\boldsymbol{w}_{SR}$ and $\boldsymbol{w}_{WT}$ show deviations from the near-perfect periodicity of environmental signals, (Fig. 2**e**), suggesting the complexity of information processing. On the other hand, $\boldsymbol{w}_{DL}$ shows a flat waveform, and $\boldsymbol{w}_{GSI}$ shows a simple sinusoidal waveform (Fig. 2**e**). Surprisingly, this estimation revealed that the periodic pattern of the temporal derivative of GSI emerges through asynchronous dynamics of $\boldsymbol{w}$. This indicates that the regulatory system continuously recalibrates its signal integration in response to environmental fluctuation, demonstrating a homeostatic capacity for maintaining periodic rhythm of gonadal development under variable conditions.

Concordance between gene expression patterns and signal importance.

Previously we collected monthly transcriptome for over a two-year period, which was simultaneously sampled with the GSI and environmental data. Based on our hypothesis that the estimated $\boldsymbol{w}_i$ (i=SR, WT, DL, GSI) values reflect environmental adaptation, we expected to find genes exhibiting similar dynamics with $\boldsymbol{w}_i$**.** Hence, we compared the time-series transcriptome with each $\boldsymbol{w}_i$ except for the $\boldsymbol{w}_{DL}$ which remained constant the whole time. As a result, we found several genes showing high concordance with the $\boldsymbol{w}$ (Fig. 3**a**). The similarity between gene expression pattern and $\boldsymbol{w}$ was quantified with the correlation coefficient. The correlation coefficients across all genes displayed a trapezoid distribution (Fig. 3**b**).

To investigate the functional tendency of the highly correlated genes, we extracted the genes which show correlation coefficients higher than 0.6 and then applied them for the enrichment analysis (clusterProfiler in R package) [22]. The genes correlated with $\boldsymbol{w}_{SR}$ are enriched in infection response pathways (e.g., Pathogenic *Escherichia coli* infection) in KEGG Orthology (KO) enrichment (Fig. 3**c,** top) and in damage response categories (e.g., double-strand binding) in Gene Ontology (GO) enrichment

(Supplementary Fig. 3**a**). The genes correlated with $\boldsymbol{w}_{WT}$ are enriched in motor protein-related pathways or Huntington's disease in KO enrichment analysis (Fig. 3**c,** middle) and in energy generation-related terms (e.g., NADH activity, cellular respiration) in GO enrichment analysis (Supplementary Fig. 3**bc**). The genes correlated with $\boldsymbol{w}_{GSI}$ are enriched in neuron-related functions, such as those involved in spinocerebellar ataxia in KO enrichment (Fig. 3**c,** bottom), and in peptidase activity in GO enrichment (Supplementary Fig. 3**d**).

To further narrow down the candidate genes, we excluded genes that showed spurious high correlations due to random chance. To this end, we computed correlation coefficients between each gene expression pattern and randomized trajectories of the signal importance, which allowed us to assess the significance of the observed correlation to randomized ones (Supplementary Fig. 4). Genes with $p < 0.05$ after Bonferroni correction were selected.

As a result, four genes were selected as significantly correlated with $\boldsymbol{w}_{WT}$ (hereafter referred to as signal importance genes). Note that any genes did not show the significant correlation with $\boldsymbol{w}_{SR}$ and $\boldsymbol{w}_{GSI}$. Signal importance genes include two known genes (LOC101160627, LOC101162833) and two novel genes (BGI_novel_G000005, BGI_novel_G000075). Notably, GO enrichment analysis revealed these genes to be associated with energy generation (Supplementary Fig3. **ef**), while KO enrichment highlighted their involvement in thermogenesis (Fig. 2**d**). The enrichment analyses indicate synchronization between thermoregulation-related genes and the signal importance of water temperature ($\boldsymbol{w}_{WT}$), suggesting seasonal physiological adaptation in medaka. These findings support that signal importance $\boldsymbol{w}$ reflects change of the internal state of the living system.

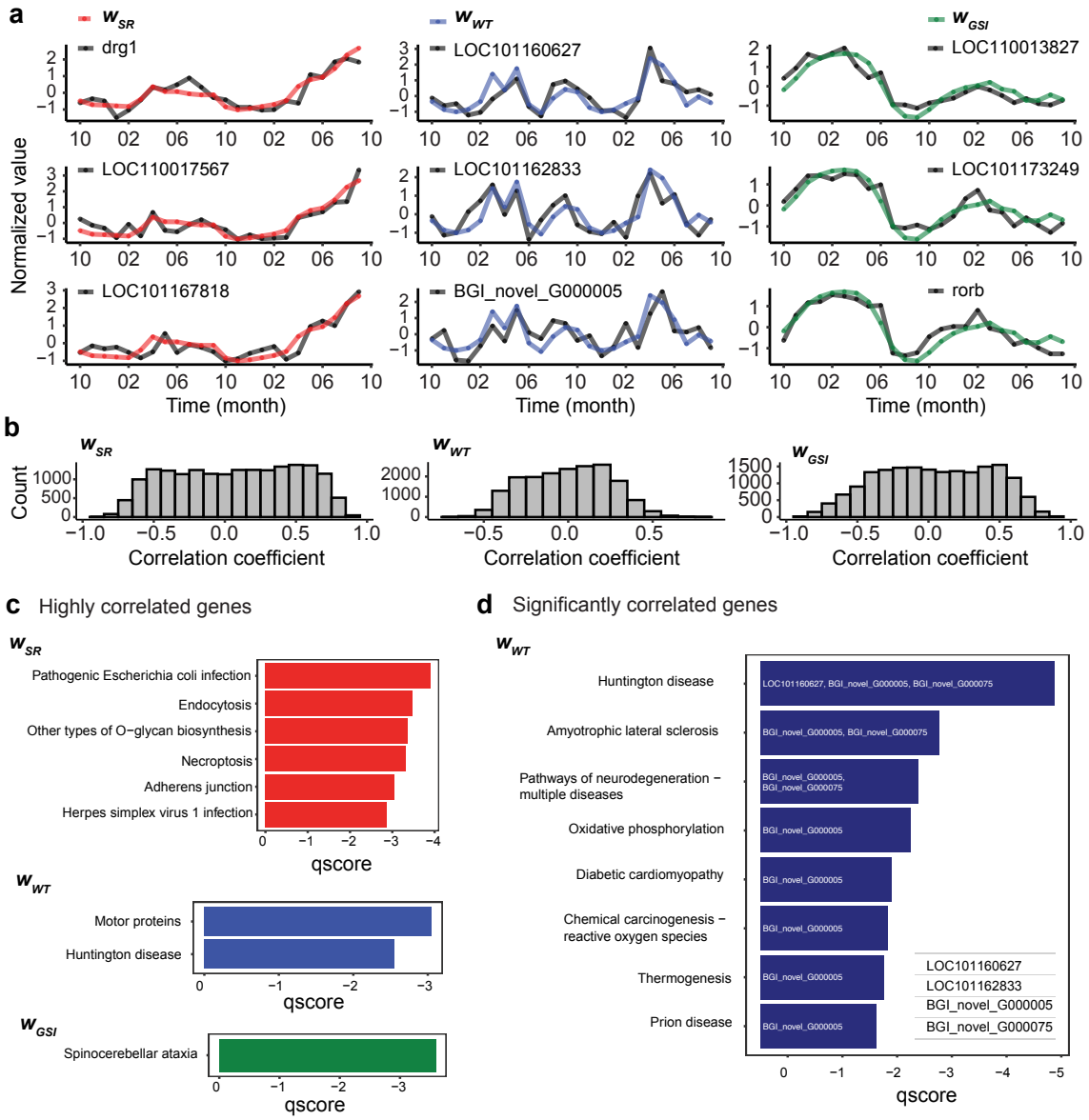

Figure 3. a The genes synchronized with the signal importance. The expression level (black) and signal importance ($\boldsymbol{w}_{SR}$: red, $\boldsymbol{w}_{WT}$: blue, $\boldsymbol{w}_{GSI}$: green) are normalized. The expression levels represent the average of two technical replicates. **b** Distribution of correlation coefficient between w and all genes (The genes which have missing value are excluded). **c,d** KO enrichment analysis. The horizontal axis shows the significance of the enrichment with qscore ($\log_{10}$p-value) and the vertical axis shows the enriched function (qscore < -1). **c** KO enrichment analysis for genes highly correlated (coef. > 0.6) with $\boldsymbol{w}_{SR}$ (red), $\boldsymbol{w}_{WT}$(blue), and $\boldsymbol{w}_{GSI}$ (green). **d** KO enrichment analysis was performed on highly correlated genes that showed statistical significance (p.adj < 0.05; Bonferroni correction). Among the tested genes, only those correlated with $\boldsymbol{w}_{WT}$ passed the statistical test. The list of significantly correlated genes are listed in the lower right table, and the gene names corresponding to each function are annotated on the bar.

Signal importance genes are involved in environmental adaptation.

The signal importance genes – the genes significantly correlated with the seasonal change of signal importance – are hypothesized to play roles in environmental adaptation. To test the hypothesis, we compared their expression dynamics under the non-stationary environment (outdoor condition) and the stationary environment (room condition). The expression data in the stationary environment were collected over 6 months under controlled room conditions (14h daylight and 26°C water temperature; Fig. 4**a**) in our previous study [20]. Although the GSI pattern shows the similar bimodality in the room and outdoor condition (Nakayama et al 2023, see Fig. 4E), the expression patterns show clear separation between the room and outdoor conditions in the signal importance genes for water temperature (Fig. 4**b**). In the room condition, the signal importance genes except LOC101162833 show low expression levels (Fig. 4**c**). The correlation of expression patterns between room and outdoor conditions are low and non-significant, in three of four genes, showing differences in expression patterns (Fig. 4**d**). The three genes (LOC101160627, BGI_novel_G000005, and BGI_novel_G000075) show distinct expression patterns – both in their absolute levels and temporal changes – suggesting that these genes operate in distinct modes under stationary versus non-stationary environments (Fig. 4**d**).

To assess how the signal importance genes differ in their expression patterns, we analyzed all genes using the same procedure applied to signal importance genes in Figures 4**c** and **d**. We quantified both the similarity in absolute expression levels between conditions (Figure 4**e**, x-axis) and the differences in expression trajectories (Figure 4**e**, y-axis). Genes in the upper left region of Figure 4**e** (Wilcoxon rank-sum test, $p < 0.01$; Spearman correlation, $p > 0.01$; Bonferroni-corrected) exhibit significant differences in both expression levels and temporal changes between conditions. Conversely, genes in the lower right region of Figure 4**e** exhibit similar dynamics under both conditions, including the 'circannual genes' that display free-running oscillation regardless of condition, as previously identified (Nakayama et al, see Fig.4GH) [20]. Among genes significantly correlated with signal importance, 75% (3/4) are in the upper left region, compared to only 30% (5,364/17,865) of all genes (Fig. 4**f**). The difference between the ratios is marginally significant (Fisher's Exact Test; $p = 0.08$), suggesting that the latent dynamics deduced from our ISI method captures the dynamics of genes involved in the environment adaptation.

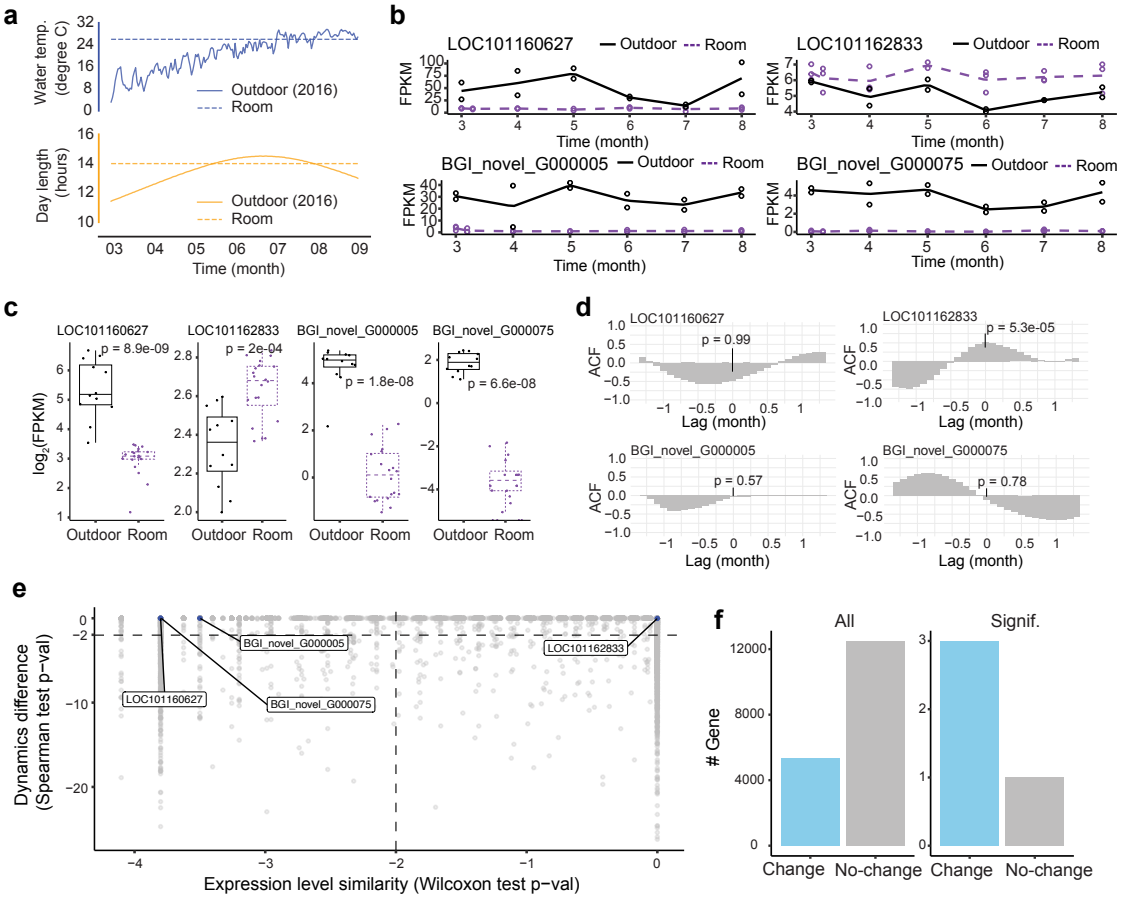

Figure 4. Expression dynamics of signal importance genes differ between outdoor and room conditions. **a** Environmental signals in outdoor (solid line) and room (dashed line) conditions. The upper panel represents water temperature, and the lower panel shows day length. The room condition was maintained at a constant 14-hour day length and 26°C water temperature. **b** Temporal patterns of expression (FPKM) in the outdoor (black) and room (purple) on the signal importance genes that are significantly correlated with water temperature. Dots show RNA-seq samples, and lines show averages in outdoor (solid: 2 samples) and room (dashed: 3 samples). **c** Expression level distributions across time series and samples. Box plots compare expression levels between outdoor (left) and room (right) conditions. Wilcoxon signed-rank test p-values are shown above.

d Cross correlation between the outdoor and room expression levels. Spearman rank correlation p-value at lag 0 indicated (upper side test). **e** Comparison between expression level similarity and temporal pattern difference across conditions. X-axis shows Wilcoxon test p-values indicating expression level similarity between conditions (as in **c**); Y-axis shows Spearman test p-values reflecting differences in temporal expression patterns between conditions (as in **d**). Points represent genes. **f** The ratio of the genes

located in the upper left region of panel **e**. Comparison between all genes (left) and signal importance genes (right). Fisher's exact test suggests a marginal difference (p=0.08).

Discussion

Our study introduces the inverse signal importance (ISI), a novel approach to infer how living organisms dynamically prioritize multi-modal environmental signals. We applied the method to the gonadal development of medaka fish. The inferred signal importances exhibit complex waveforms that deviate from simple annual periodicity, and these waveforms show strong correlations with expression patterns of several genes. In particular, the signal importance on water temperature shows a significant correlation with four genes, three of which display distinct expression dynamics between outdoor and room conditions, indicating their potential role in environmental adaptation.

The dynamics of signal importance indicates the seasonal modulation of the system, which may reflect physiological changes, but also could rise from the confounding factors and/or overfitting. The inherent uncertainty regarding the nature of signal importance necessitated further investigation. For example, four genes show significant correlation with $\boldsymbol{w}_{WT}$, the signal importance for water temperature. This indicates that $\boldsymbol{w}_{WT}$ reflects the physiological dynamics involved in thermogenesis as temperature adaptation (Fig. 3). For the rest of three signal importance, since none shows a significant correlation with any genes, it cannot exclude the possibility that those came from confounding factors. For example, genes highly (though not significantly) correlated with w_{SR} exhibit monotonic increases over the two-year observation and are enriched in *E. coli* infection terms. This suggest an increase in E. coli density, which may be linked to deteriorating water quality and the resulting reduction in sunlight reaching beneath the water. Consequently, the increasing importance of solar radiation may represent a compensation mechanism for the reduced sunlight penetration.

Among the four $\boldsymbol{w}_{WT}$ -correlated genes, we identified three that exhibit differential expression patterns between indoor and outdoor conditions, suggesting their involvement in environmental adaptation. Notably, although around 5, 000 genes in our dataset showed differential expression patterns between indoor and outdoor conditions (Fig. 4**f**), the ISI framework successfully extracted a subset of the genes associated with adaptation to water temperature changes, by screening with the predicted dynamics ($\boldsymbol{w}_{WT}$). Our procedure contrasts with our previous analysis which focused on the circannual genes showing similar oscillatory patterns between indoor and outdoor

conditions. Our results exemplify that the ISI method allows us to efficiently identify key candidate genes for specific adaptation mechanisms from large transcriptome data.

Living systems demonstrate robustness across multiple scales, from gene expression [23] to annual rhythms [20]. Hence, the superficial traits (i.e., rhythms or morphology) often appear unchanged even under perturbations due to internal compensatory mechanisms. For example, the circadian rhythm of *Arabidopsis thaliana* maintains robust oscillations against temperature fluctuations through the temperature-responsive genes, which counterbalance the effects of temperature changes on biochemistry [24]. The ISI framework offers a valuable lens to understand such internal mechanisms by separate modeling of signal processing systems and its modulators.

Our ISI framework effectively decomposes complex time series observed in natural environments into deterministic elements and variable elements. While the variations that could not be explained by deterministic systems have been often dismissed as noise, this study reinterprets these variations as environmental adaptations of deterministic systems, referred to as "signal importance," and elucidates their physiological nature. Living systems, as demonstrated by the gonadal development of medaka, can adjust their deterministic systems in response to external influences. Therefore, the ISI framework paves the way to reveal organismal adaptation to the natural environment.

Methods

A. Data processing

Meteorological and GSI data

We used the preprocessed meteorological and GSI data from the previous study (Nakayama et al., 2023). The data set was collected at the National Institute for Basic Biology in Okazaki from October 1, 2015, to October 15, 2017, with sampling conducted twice per month. In each sampling, the GSI of ten female medaka was recorded.

For the calculation of derivative of GSI ($\dot{GSI}_t$), the time series of GSI are first averaged over medaka individuals. Subsequently, the derivative at each sample point was computed through the linear fitting with a window size 3(Supplementary Fig. 1).

Data splitting

To prevent overfitting, we divided the GSI data from 10 female specimens into three subsets: Data 1 (n = 3), Data 2 (n = 3), and Data 3 (n = 4). During hyperparameter

tuning, we employed cross validation, i.e., using two subsets for training and the remaining one for testing. Once the optimal hyperparameters were determined, the entire dataset was utilized to train the final model and estimate the parameters.

B. Mathematical model

Inverse Signal Importance

The ISI method is based on the following model:

$$y_t = \sum_{i=1}^{4} w_i(t) \int_0^T f_i(\tau) x_i(t-\tau) d\tau \quad (eq.1)$$

where y_t is the temporal differentiation of GSI at time = t $(G\dot{S}I_t)$ in this study. The input signal $x_i(t)$ indicates solar radiation (SR, $x_1 \equiv x_{SR}$), water temperature (WT, $x_2 \equiv x_{WT}$), day length (DL, $x_3 \equiv x_{DL}$), and Gonadosomatic index (GSI, $x_4 \equiv x_{GSI}$), at time = t. We include the past GSI value as input to predict output $(G\dot{S}I_t)$ as widely utilized in machine learning. $f_i(\tau)$ is a time-invariant regression parameter for the input i and delay τ, and $w_i(t)$ is a time-varying coefficients for the input i at time t. The convolution of past signal information, $\int_0^T f_i(\tau) x_i(t-\tau) d\tau$, is known as an ordinary response function with kernel size T. The response function is further modified by the factor $w_i(t)$ which vary over time to describe seasonal modulation.

Parameter inference framework: hard EM algorithm

For parameter inference, let us consider the time discretization of eq. 1 as

$$y_t = \sum_{i=1}^{4} w_{i,t} \sum_{\tau=0}^{T} f_{i,\tau} \cdot x_{i,t-\tau} \quad (eq.2)$$

where $x_{i,t}$, $f_{i,t}$, and $w_{i,t}$ represent $x_i(t)$, $f_i(t)$, and $w_i(t)$. For brevity, we denote by $\boldsymbol{x}$ the collection of all $x_{i,t}$ variables for all *i* and *t*. The same notation applies to $\boldsymbol{f}$ and $\boldsymbol{w}$.

The parameters $\boldsymbol{f}$ and $\boldsymbol{w}$ are inferred to maximize the following log-likelihood:

$$\ln p(\mathcal{D}|\boldsymbol{f},\boldsymbol{w}) = \sum_{t=1} \ln \frac{1}{\sqrt{2\pi k^2}} \exp\left[-\frac{\left[y_t(\boldsymbol{x},\boldsymbol{f},\boldsymbol{w}) - G\dot{S}I_t\right]^2}{2k^2}\right] \quad (eq.3)$$

where $y_t(\boldsymbol{x},\boldsymbol{f},\boldsymbol{w})$ is a predicted $G\dot{S}I_t$ from our model (eq. 2), and $\mathcal{D}$ is a dataset including $\boldsymbol{x}$ and $\mathbf{G\dot{S}I}$.

We adopted an alternating optimization method for the log-likelihood, which is inspired by the EM algorithm (For complete derivation, see Supplementary Note 1) [25].

E step: Optimization of w

Signal importance $\boldsymbol{w}$ is inferred with the following state-space model under fixed $\boldsymbol{f}^*$:

$$\begin{cases} \boldsymbol{w}_t = \boldsymbol{w}_{t-dt} + \boldsymbol{s} & \boldsymbol{s} \sim N(\boldsymbol{0}, \sigma\, dt\, \boldsymbol{I}) \\ y_t = \boldsymbol{H}(\boldsymbol{x}_t, \boldsymbol{f}^*)\boldsymbol{w}_t + \boldsymbol{v} & \boldsymbol{v} \sim N(\boldsymbol{0}, \boldsymbol{R}) \end{cases} \qquad (eq.4)$$

where $\boldsymbol{H}_t(\boldsymbol{x}_t, \boldsymbol{f}^*)$ is the set of processed signals through a linear filter $\boldsymbol{f}^*$, $\boldsymbol{s}$ is process noise, and $\boldsymbol{v}$ is observation noise (Supplementary Note2). Here we used $\boldsymbol{x}_t$ to represent $[x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}]$. Based on *eq.* 4, the optimal $\boldsymbol{w}$ is inferred by the Kalman filter. The variance of $\boldsymbol{v}$ is estimated from the data (Supplementary Note 2), while the strength of process noise σ is determined through the grid search (Supplementary Fig. 2).

M step: Optimization of f

The linear filters $\boldsymbol{f}$ are inferred with the ridge regression where the regularization coefficient is selected as $\lambda = 1.0$ [26] to avoid overfitting.
Under fixed $\boldsymbol{w}^*$, $\boldsymbol{f}$ is updated as:

$$\boldsymbol{f}^* = \arg\min_{\boldsymbol{f}} \sum_{t=1}^{F} \left(\boldsymbol{f} \cdot \boldsymbol{A}(\boldsymbol{x}_t, \boldsymbol{w}_t^*) - \boldsymbol{G}\dot{\boldsymbol{S}}\boldsymbol{I}_t\right)^2 + \lambda \|\boldsymbol{f}\|^2 \qquad (eq.5)$$

where design matrix $\boldsymbol{A}(\boldsymbol{x}_t, \boldsymbol{w}_t^*)$ is the set of input signals modulated by $\boldsymbol{w}^*$ (Supplementary Note 3). Here we used $\boldsymbol{x}_t$ to represent $[x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}]$.

C. Statistical test

The significance test of correlation between the signal importance and gene expression.

For estimating the significance of correlation between the signal importances and gene expression patterns, we performed a statistical test with the null hypothesis that the correlation is due to random effects. First, 100,000 randomized waveforms of $\boldsymbol{w}_i (= [\boldsymbol{w}_{i,1}, \boldsymbol{w}_{i,2}, \ldots, \boldsymbol{w}_{i,t}])$ were generated while preserving the autocorrelation of the original $\boldsymbol{w}_i$ using the iterated amplitude adjusted Fourier transformation [27]. The correlation coefficients between the expression pattern of the target gene and each randomized waveform were then computed, and their histogram was constructed. The null distribution was inferred by applying kernel density estimation. Finally, the p-value was calculated as the probability of obtaining a correlation coefficient greater than the observed value (Supplementary Fig. 4).

Code availability
ISI is developed under python 3.9.13, and statistical analysis is performed under R 4.4.1. The ISI and the statistical analysis code will be distributed through GitHub after publication.

Data availability
This study was a re-analysis of existing data (Nakayama et al., PNAS, 2024). Time-course expression data of medaka were downloaded from the Gene Expression Omnibus (GEO) under accession numbers GSE234401 for the outdoor condition and GSE234565 for the room condition.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234401.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234565.

Competing interests
The authors declare no competing interests.

Author Contributions
H.N., Y.K., T.I., K.A. and T.Y. conceived the project. T.I., H.N. and Y.K. developed the model. A.S., T.N., and T.Y. conducted the experiments. T.I. analyzed the data. T.I., Y.K. and H.N. wrote the manuscript with input from all the authors.

Acknowledgments
We are grateful to Prof. Nen Saito and all members of the Yoshimura laboratory for the valuable discussions. This study was supported in part by Grant-in-Aid for Transformative Research Areas (B) [grant number 21H05170 to H.N.], Moonshot R&D–MILLENNIA Program [grant number JPMJMS2024-9 to H.N. and Y.K.] by JST, the Cooperative Study Program of Exploratory Research Center on Life and Living Systems (ExCELLS) [program number 21-102 to H.N.], and Japan Society for the Promotion of Science [grant number 23KJ1009 to T.I.].

References

1. *Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology.* **Wild, C. P.** 2005, Cancer Epidemiol Biomarkers Prev., Vol. 14, pp. 1847-1850.

2. *Multimodal sensory processing in Caenorhabditis elegans.* **Metaxakis, A., Petratou, D., & Tavernarakis, N.** 2018, Open Biology, Vol. 8.

3. *Decoding reward–curiosity conflict in decision-making from irrational behaviors.* **Konaka, Y., & Honda., N.** 2023, Nature Computational Science, Vol. 3, pp. 418-432.

4. *Dopamine modulation of sensory processing and adaptive behavior in flies.* **Siju, K. P., De Backer, J. F., & Grunwald Kadow, I. C.** 2021, Cell and tissue research, Vol. 383, pp. 207-225.

5. *Flexibility in metabolic rate confers a growth advantage under changing food availability.* **Auer, S. K., Salin, K., Rudolf, A. M., Anderson, G. J., & Metcalfe, N. B.** 2015, The Journal of animal ecology, Vol. 84, pp. 1405-1411.

6. *Leptin in Teleost Fishes: An Argument for Comparative Study.* **Copeland, D., Duff, R., Liu, Q., Prokop, J., & Londraville, R.** 2011, Frontiers in Physiology, Vol. 2.

7. *Metabolic Flexibility as an Adaptation to Energy Resources and Requirements in Health and Disease.* **Smith, R. L., Soeters, M. R., Wüst, R. C. I., & Houtkooper, R. H.** 2018, Endocrine reviews, Vol. 39, pp. 489-517.

8. *Multikinase Networks: Two-Component Signaling Networks Integrating Multiple Stimuli.* **Francis, V., & Porter, S.** 2019, Annual review of microbiology., Vol. 73, pp. 199-223.

9. *A neural mechanism for conserved value computations integrating information and rewards.* **Bromberg-Martin, E., Feng, Y., Ogasawara, T., White, J., Zhang, K., & Monosov, I.** 2022, Nature Neuroscience, Vol. 27, pp. 159-175.

10. *One gene, multiple ecological strategies: A biofilm regulator is a capacitor for sustainable diversity.* **Mhatre, E., Snyder, D., Sileo, E., Turner, C., Buskirk, S., Fernandez, N., Neiditch, M., Waters, C., & Cooper, V.** Proceedings of the National Academy of Sciences of the United States of America, 2020, Vol. 117, pp. 21647-21657.

11. *Revisiting the evolution of bow-tie architecture in signaling networks.* **Itoh, T., Kondo, Y., Aoki, K., & Saito, N.** 2024, NPJ systems biology and applications, Vol. 10, p. 70.

12. *Neural Circuits Underlying Behavioral Flexibility: Insights From Drosophila.* **Devineni, A. V., & Scaplen, K. M.** 2022, Frontiers in behavioral neuroscience, Vol. 15.

13. *Signal integration through blending, bolstering and bifurcating of intracellular information.* **Pawson, C. T., & Scott, J. D.** 2010, Nature structural & molecular biology, Vol. 17, pp. 653-658.

14. *Spatial summation in the receptive fields of simple cells in the cat's striate cortex.* **Movshon, J A, Thompson, I D, & Tolhurst, D J.** 1978, The Journal of Physiology,, Vol. 283 .

15. *Stochastic ERK Activation Induced by Noise and Cell-to-Cell Propagation Regulates Cell Density-Dependent Proliferation.* **Aoki, K., Kumagai, Y., Sakurai, A., Komatsu, N., Fujita, Y., Shionyu, C., & Matsuda, M.** 2013, Molecular cell, Vol. 52, pp. 529-540.

16. *Rescaling of Spatio-Temporal Sensing in Eukaryotic Chemotaxis.* **Kamino, K., & Kondo, Y.** 2016, PloS one, Vol. 11.

17. *Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation.* **Marshall, J.C.** 1995, Cell, Vol. 80, pp. 179-185.

18. *Behavioral flexibility as a mechanism for coping with climate change.* **Beever, E.A., Hall, L.E., Varner, J., Loosen, A.E., Dunham, J.B., Gahl, M.K., Smith, F.A., Lawler, J.J.** 2017, Front Ecol Environ, Vol. 15, pp. 299–308.

19. *Phenotypic flexibility and the evolution of organismal design.* **Piersma, T., Drent, J.** 2003, Trends in Ecology & Evolution, pp. 228-233.

20. *A transcriptional program underlying the circannual rhythms of gonadal development in medaka.* **Nakayama, T., Tanikawa, M., Okushi, Y., Itoh, T., Shimmura, T., Maruyama, M., Yamaguchi, T., Matsumiya, A., Shinomiya, A., Guh, Y. J., Chen, J., Naruse, K., Kudoh, H., Kondo, Y., Naoki, H., Aoki, K., Nagano, A. J., & Yoshimura, T.** 2023, Proceedings of the National Academy of Sciences of the United States of America, Vol. 120, p. 52.

21. *メダカに於ける温熱抵抗の季節的變化 [Seasonal Changes in Thermal Resistance in Medaka].* **Matsui, Kizô.** 1940, Zoological magazine, Vol. 52, pp. 385-389.

22. *clusterProfiler: an R package for comparing biological themes among gene clusters.* **Yu, G., Wang, L. G., Han, Y., & He, Q. Y.** 2012, OMICS: A Journal of Integrative Biology, Vol. 16, pp. 284-287.

23. *Capturing hidden regulation based on noise change of gene expression level from single cell RNA-seq in yeast.* **Itoh, T., & Makino, T.** 2021, *Scientific Reports*, Vol. 11.

24. *Cold-induced degradation of core clock proteins implements temperature compensation in the Arabidopsis circadian clock.* **Maeda, A. E., Matsuo, H., Muranaka, T., & Nakamichi, N.** 2024, Science advances, Vol. 10.

25. *Pattern recognition and machine learning.* **Bishop, Christopher M.** New York: Springer, 2006.

26. *Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. Statistics in medicine.* **Pavlou, M., Ambler, G., Seaman, S., De Iorio, M., & Omar, R. Z.** 2016, Statistics in medicine, Vol. 37, pp. 1159-1177.

27. *A stochastic iterative amplitude adjusted Fourier Transform algorithm with improved accuracy.* **Venema, V., Ament, F. & Simmer, C.** 2006, Nonlin. Proc. Geophys., pp. 13, pp. 321-328.

28. *Learning to Segment Medical Images with Scribble-Supervision Alone.* **Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., & Baumgartner, C.F.** 2018, Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer , pp. 236-244.