

Influential Force: from Higgs to the *Ab Initio* Genetic Orbital Method

Tsutomu Mori^{1*}, Takashi Kawamura¹, Daisuke D. Ikeda², Susumu Goyama³,
Hiroshi Haeno⁴, Kazuhiko Ikeda⁵, Katsue Adachi⁶, Yoshihiko Saito⁶,
Tomoyoshi Horisawa⁶, Junzo Suzuki⁷, and Seiichi Takenoshita⁸

¹Department of Human Lifesciences,
Fukushima Medical University School of Nursing, Fukushima, Japan

²Kitajima-cho, Itano-gun, Tokushima, Japan

³Division of Molecular Oncology, Department of Computational Biology and Medical Sciences,
Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

⁴Medical Mathematical Modelling Laboratory, Department of Computational Biology and Medical
Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

⁵Division of Blood Transfusion and Transplantation Immunology,
Fukushima Medical University, Fukushima, Japan

⁶Chi Co., Ltd., Tokyo, Japan

⁷Fukushima Preservative Service Association of Health, Fukushima, Japan

⁸Fukushima Medical University, Fukushima, Japan

*Correspondence: tmori@fmu.ac.jp

Abstract

Nature has a tendency to evolve to a more probable state, and we regard this tendency as a force of probability. All fundamental interactions are transmitted by mediator particles, and their unification has long been attempted. However, no theory has fully unified the interactions into a single simple formula. Here, we show that fundamental interactions are unified by a novel “influential force” driven by probability under the canonical distribution of the transmitted information. This force affects the information distance denoting transmission difficulty, based on a universal gauge symmetry within information coordinate spacetime. We develop statistical mechanics of mutual information and reveal that the influential force exists in both physical and biological systems, offering clues to solve many intractable problems. In the field of physics, the influential force provides a coherent explanation of the spontaneous symmetry breaking of the Higgs field and its relationship with gravity, inflation, quantum entanglement entropy, and the hierarchy problem. In the field of biology, the force endows genes with huge network information, which leads to the development of the *ab initio* genetic orbital method and identification of a novel immune checkpoint, *KYNU*/kynureninase. Our findings demonstrate that the influential force acts between highly divergent beings, thereby shaping the essential properties of nature.

Keywords: influential force, unification theory, Higgs mechanism, *ab initio* genetic orbital method, immune checkpoint

CONTENTS

1. Introduction	4
2. Influential force and physics	8
(a) Influential force based on MI	8
(b) Influential force and field theory	13
(c) Influential force and physical forces	17
(d) Motion generated by the influential force	25
(e) Influential force and quantum oscillators	32
(f) Time evolution of MI in classical systems	39
(g) Time evolution of MI in quantum systems	44
(h) Influential force and quantum entanglement	52
(i) Influential force field and the Higgs field	53
(j) Many-body system	69
3. Influential force and biology	72
(a) Gene information and life	72
(b) Gene information and the influential force	74
(c) Relationship between gene information and fitness	75
(d) Relationship between intergenic MI and fitness	79
(e) Influential force acting between two genes	80
(f) Evolution of the MI of two genes	88
(g) Evolution of a network of two genes	93
(h) Evolution of networks of multiple genes	95
4. <i>Ab initio</i> genetic orbital (GO) method	103
(a) Basic theory of the <i>ab initio</i> GO method	103
(b) Super-high resolution analysis of intergenic interactions	105
(c) Frontier information theory	109
(d) Weighted pathway analysis	115
(e) Application of the <i>ab initio</i> GO method	116
(f) Searching for a novel immune checkpoint in cancer	121
5. Conclusions and Discussion	129

Appendices	139
1. Distances, metrics, and coordinates	139
2. Probability distribution of MI	140
3. MI and natural force potentials	141
4. Probabilistic influential force potential and Morse potential	141
5. Probabilistic influential force potential and dispersion force	142
6. Micromutual information summation theorem	143
7. MI of informatons that accompanies fluctuations	144
8. MI of free particles in an exchange interaction	146
9. Multidimensional information and linear index	149
10. Relationship between effective self-information and fitness	150
11. Relationship between MI and fitness	151
12. Moran process model of evolution	152
13. Influential force acting between genes	154
14. Temporal change in the MI of genes	156
15. Network information of edge weights	157
16. Functional mutual information	157
17. Infocanonical distribution	158
18. Infocanonical distribution of micromutual information	159
19. Frontier information theory	160
20. Equivalence principle of information and probability	161
21. Weighted pathway analysis	168
22. Equation of motion in spacetime with probability as the coordinate	168
23. Equilibrium information distance	169
24. Influential microforce	170
25. Information entropy of the multidimensional random variable	172
26. Temporal change in effective self-information	173
27. Relationship between the rank change in MI and fitness	174
28. Pareto distribution of intergenic MI	176
29. Multivariate central limit theorem	177
30. Integration theory of the Moran process and branching process	178
31. Examples of the <i>ab initio</i> GO calculation	179
Acknowledgments	181
Author contributions	181
Conflict of interest	181
References	181

1. Introduction

Nature has a general tendency to evolve from a less probable to a more probable state [1], and we regard this tendency as a force arising from probability. All fundamental interactions (i.e., the electromagnetic, weak, strong, and gravitational interactions) are thought to be transmitted by mediator particles, and the unification of these interactions has long been attempted [2-6]. In particular, superstring theory has been proposed to unify gravity and the other interactions [7,8]. M-theory was further developed to unify five kinds of superstring theory [9]. However, no theory has fully succeeded in unifying the interactions into a single simple formula [10]. Notably, no supersymmetric particles have been discovered by the Large Hadron Collider [11]. This poses serious questions about the existence of supersymmetry, on which superstring theory is based. Here, we show that all fundamental interactions are unified by a force driven by probability under the canonical distribution of the transmitted information. Furthermore, we introduce an information coordinate system to establish the formalism of the force. The information coordinate uses the information distance that represents the difficulty of information transmission. The unified force, the “influential force,” originates from a universal gauge symmetry within this information coordinate spacetime.

We briefly introduce the information distance between factors having information, where we call these factors informatons. Suppose that informatons X and Y constitute a closed composite system XY . Let $H(X)$ and $H(Y)$ be the information entropies of X and Y , respectively. The mutual information (MI) is $MI(X;Y) = H(X) + H(Y) - H(X,Y)$, where $H(X,Y)$ is the joint entropy of the composite system [12] (Fig. 1a). When the conventional distance between X and Y is x , the information distance I_O is defined by the self-information as $I_O = -\log p(x)$, in which $p(x)$ is the probability density of information transmission by the mediator particles. I_O also expresses the least-action path length for a mediator particle traveling between the informatons. In turn, I_O is subject to change by MI, which represents the amount of transmitted information. Here, I_O involves two kinds of information distance: one of external spacetime and one of internal spacetime. The two information distances constitute I_O in an approximately equivalent manner. We refer to this equivalence as metric symmetry.

The composite system XY is represented by a single wave function. The change in its energy is proportional to that in its entropy when the Helmholtz free energy is constant. In this case, the probability increases in the direction in which the system's energy decreases, and its entropy also decreases. At this time, MI shared by X and Y follows a probability distribution $p = \exp(MI)$. This means $\Delta MI > 0$; that is, there is the tendency for information exchange to develop between the informatons, which we refer to as the genesis of the influential force. Notably, this force decreases the overall entropy of the composite system XY . As a result, the force reduces the system's total energy and becomes equivalent to the energetic force [13,14]. This particular force also decreases the information distance I_O , such that $\Delta I_O = -\Delta MI < 0$, and is therefore an attractive force.

In contrast to the attractive force, by the second law of thermodynamics, a counteracting force arises to recover the entropy of the system. In this case, MI follows another probability distribution, $p = \exp(-2MI)$. This relationship demonstrates an additional tendency in the direction in which $\Delta MI < 0$; that is, opposite the direction of the attractive force. This second force brings about an increase in I_O , such that $\Delta I_O = -\Delta MI > 0$, and thus acts as a repulsive force. Under isothermal conditions, this repulsive force increases the total entropy and energy of the composite system XY and is classified as an entropic force [14-16].

On the basis of the generality of both probability and information, the influential force is applied to diverse interactions and thus explains a wide range of phenomena (Figs. 1b–1e). This is because probability and information are everywhere in the universe [17]; that is, they exist in physical, chemical, biological, engineering, and sociological systems, as well as in human relationships. This paper aims to prove the existence of the influential force systematically. For this purpose, it is desirable to provide evidence for more than one system using an identical formula. We therefore give examples in which the influential force has a considerable effect on either a physical or biological system. Furthermore, we demonstrate that, in either system, we can solve many long-lasting problems with particular importance by considering the influential force.

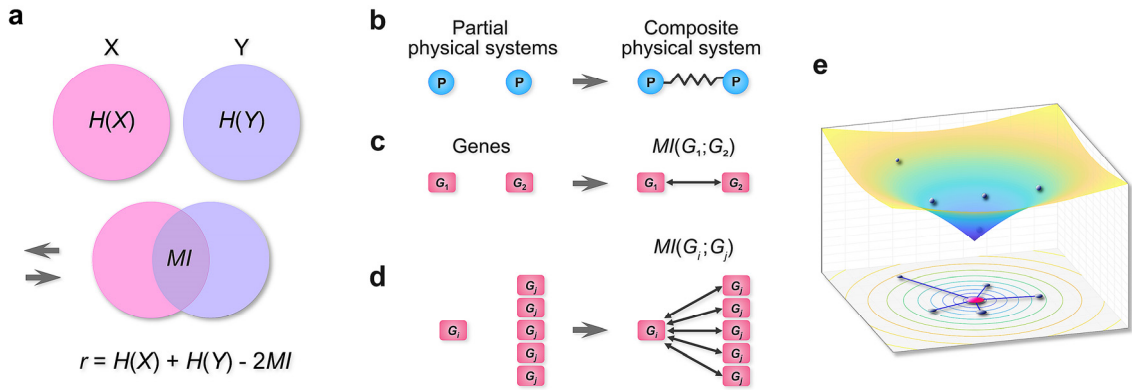


Figure 1 | Informatons and mutual information. **a**, MI . $H(X)$ and $H(Y)$ are the Shannon entropies of informatons X and Y , respectively, shown as discs in the Venn diagram. MI of X and Y corresponds to the intersection. The information metric is $r = H(X) + H(Y) - 2MI$. Total entropy $H(X, Y)$ corresponds to the union. **b**, X and Y are partial physical systems, constituting a composite physical system based on MI . **c**, X and Y are genes G_1 and G_2 . **d**, Gene G_i interacting with other genes G_j through $MI(G_i; G_j)$. **e**, Gene networks. The bottom plane represents the two-dimensional information metric space, in which six genes in **(d)** are located, with G_i at the center and the other genes on the periphery. G_i and the other genes interact owing to the influential force based on $MI(G_i, G_j)$. The surface expresses the influential force potential, its vertex corresponds to G_i , and the other genes are located on the surface such that their projections on the bottom plane correspond to the genes.

In the physical systems described in Chapter 2, we establish an information coordinate system, thereby demonstrating the gauge symmetry of the unified force and explaining its mechanics within the information coordinate spacetime. In the following, we list examples in which we address and demonstrate the involvement of the influential force.

- 1) Generation of the influential force through the relativistic scattering of two particles, which is associated with the creation of MI and potential energy
- 2) Emergence of MI upon mixing two inert gases, which appears to provide stabilization energy that contributes to the formation of the mixed state of the gas
- 3) Creation of the influential force acting between identical free particles through a novel exchange interaction, whose energy may be a candidate for dark matter
- 4) Application of the influential force to the Higgs field to solve many problems facing the Standard Model, which likely provides a unified understanding of the Higgs boson, gravity, and inflation
- 5) Adoption of the information coordinates to deal with the entanglement entropy S_{EE} , which may incorporate the outcome of string theory, offering clues to the gauge–gravity correspondence
- 6) Application of the influential force to many-body systems to predict the repulsive force acting between the bodies, which may quantitatively explain the dark energy of the universe

In 1), the scattering reduces the joint entropy of the composite system, generates MI, and thereby produces potential energy. In 2), the collision of gas molecules induces fluctuations in kinetic energy, which creates the MI of two populations of molecules. In 3), the new exchange interaction appreciably differs from the known interactions. This interaction energy is expected to exist in vast quantities in the universe and is thus a candidate for dark matter. In 4), the standard Higgs potential can be replaced by the influential force potential (i.e., the influential Higgs). The influential Higgs allows the unification of the gravitational and Higgs fields, explains spontaneous symmetry breaking, illustrates the equivalence of inertial and gravitational mass, and provides a clue to the hierarchy problem. In addition, an influential Higgs inflation model fairly reproduces the data of the cosmic microwave background. In 5), S_{EE} may cause the influential force and the gravity between spacetime quanta, thereby creating the Higgs boson as a blackhole dimer. In 6), we explain why dark energy occupies two-thirds (67%) of the mass-energy density of the universe, close to the measurement of 68%.

In the biological systems discussed in Chapter 3, we consider the influential force in terms of population genetics to address the involvement of gene information in biological evolution. We demonstrate that the influential force acts between genes in the direction in which the biological fitness (i.e., the existence probability of the gene information itself) increases. This is done by generating order and by providing homeostasis for organisms. As listed below, we provide examples in which we address and illustrate the involvement of the influential force.

- 7) Generation of the influential force acting between two genes, which allows information exchange between the genes and thus their coordinated functions
- 8) Creation of gene networks based on the influential force acting between genes (Figs. 1d and 1e), which likely produces a large quantity of network information
- 9) Development of the network function of genes with a considerable amount of information, which can create a large fitness advantage and likely promotes the phenotypic evolution of organisms

The above observations also apply to cancer biology, as in the following example.

- 10) Genesis of cancer gene networks, whose abundant information can accelerate the evolution of cancer progenitor cells to cancer cells

In 7), the intergenic attractive force is expected to be proportional to $\exp(MI)$, as it is for physical systems. In 8), gene networks will encode much more information than proteins, which suggests the priority of the network functions of genes. In 9), the substantial fitness advantage of the network function will likely provide an answer to a persistent problem in evolutionary biology; that is, it has been difficult to explain the phenotypic evolution of organisms through the molecular evolution of genes. In 10), the cancer gene network will have a structure similar to that of the normal gene network. Of course, the cancer gene network is harmful to the host organisms.

In Chapter 4, we describe an application of the influential force to medical informatics. As an inference from the above, on the basis of generating MI of genes, the influential force appears to endow genes with advanced network functions. To prove this, we developed an *in silico* strategy that calculates the multidimensional MI of all genes and thereby enables a rapid computation of gene functions. This methodology involves the following advances in theory and technology. The methodology yielded a series of information theoretical theorems and two informatic algorithms that led to the identification of a novel candidate for an immune checkpoint in cancer.

- 11) Development of statistics for the microstates of MI, which allows the faster computation of *MI* and that of Fisher's exact *p*-value for two highly multidimensional random variables
- 12) Demonstration of a theorem, the equivalence principle of information and probability, which states that *MI* for two informatons is equivalent to the *p*-value of Fisher's exact test, thereby combining information theory, probability theory, and statistics
- 13) Development of the *ab initio* genetic orbital (GO) method, which rapidly calculates the network function of unknown genes without performing experiments
- 14) Development of STAIC (a Strategic Tool for Ab Initio Identification of Cancer genes), which facilitates the discovery of cancer genes by incorporating multiple informatics approaches
- 15) Identification of a potential immune checkpoint *KYNU*, encoding a metabolic enzyme kynureninase, as a candidate clinical marker and therapeutic target for cancer

In 11), the algorithm uses the statistics of micro-MI at each microstate. The statistical mechanics of highly multidimensional informatons is necessary for analyzing the relationship between genes as well as that between quanta, by which we can obtain the probabilistic representation of the influential force between the informatons. In 12), while the theorem enables an informatic interpretation of the probabilistic-statistical relationship between informatons, it conversely offers a probabilistic interpretation of MI. In 13), the GO represents the influential force in a multi-gene system, which originates from one gene and targets another gene. The *ab initio* GO method uses the same formula as the *ab initio* molecular orbital (MO) method. In 14), the combination of multiple informatic algorithms makes the identification of cancer genes efficient. In 15), the *ab initio* GO method identifies a potential immune checkpoint *KYNU*, which affects the prognosis of lung and pancreatic adenocarcinoma and is a candidate therapeutic target for cancer. This example demonstrates the applicability of the method to the medical sciences.

This paper contains 31 appendices for information theoretical theorems, supplementary equations, and supplementary figures.

2. Influential force and physics

(a) Influential force based on MI

MI and influential force in thermodynamics

We define an *informaton* as a factor that has information. In this paper, we limit ourselves to the case in which information exerts an influence on the realization probability of events [18]. In physics, the situation in which information combines with energy is such a case [19,20]. In addition, we limit ourselves to considering only isothermal conditions. Let X and Y be random variables and $H(X)$ and $H(Y)$ be their respective Shannon information entropies. In this case, the *mutual information MI* of X and Y is expressed as $MI = H(X) + H(Y) - H(X, Y)$, where $H(X, Y)$ is the joint entropy. The *information metric r* of X and Y is then expressed [21] as

$$r = H(X) + H(Y) - 2MI \quad (1)$$

(Fig. 1). It is noted that r satisfies the axioms of the metric space (Appendix 1). In this case, if we regard $\mathcal{M}(X) = \exp[H(X)/2]$ and $\mathcal{M}(Y) = \exp[H(Y)/2]$ as the information masses, then

$$\mathcal{F} = e^{MI} = \frac{\mathcal{M}(X)\mathcal{M}(Y)}{e^{\frac{1}{2}r}} \quad (2)$$

represents the force acting between informatons X and Y . Here, \mathcal{F} represents an attractive force with magnitude e^{MI} , which decreases the information metric r of the two informatons. In the following, we demonstrate that \mathcal{F} acts as the force.

We first consider the Shannon MI of two physical bodies, X and Y , which are regarded as physical informatons. We let x and y be the states of X and Y , while $p_X(x)$ and $p_Y(y)$ are their respective marginal probabilities. Additionally, we let $p(x, y)$ be the joint probability. The Shannon MI, MI , is then expressed as

$$MI(X;Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p_X(x)p_Y(y)} , \quad (3)$$

which represents the amount of information shared by physical informatons X and Y .

We next discuss the relationship between information and energy in quantum systems. If we assume that energy E of a body in a closed equilibrium system follows a canonical distribution [22], then the realization probability p_ν of energy level E_ν satisfies $p_\nu = \exp(-\beta E_\nu)/Z$, where β is the inverse temperature $1/k_B T$ expressed in terms of the Boltzmann constant k_B and absolute temperature T , and Z is the partition function $\sum_\nu \exp(-\beta E_\nu)$. We here define the **information level** I_ν of this body as the self-information of the energy level E_ν ; that is,

$$I_\nu = -\log p_\nu = \beta E_\nu + \log Z . \quad (4)$$

This indicates that the information and energy at each level are related through the realization probability. Furthermore, this relationship can be expressed at the entropy level; that is, the Shannon information entropy $H(X)$ and thermodynamic entropy $S(X)$ are connected by $k_B H(X) = S(X)$ [19].

We here describe the interaction between the two physical informatons X and Y , from an information theoretical point of view. Concerning the two-body system, the **joint thermodynamic entropy** $S(X, Y)$ and **joint information entropy** $H(X, Y)$ satisfy $S(X, Y) = k_B H(X, Y)$. Hence,

$$S = k_B [H(X) + H(Y) - MI] . \quad (5)$$

Therefore, the internal energy U is expressed as

$$\begin{aligned} U &= F + TS = -k_B T \log Z + k_B T [H(X) + H(Y) - MI] \\ &= k_B T [H(X) + H(Y) - MI - \log Z] = \frac{1}{2} k_B T [H(X) + H(Y) + r - 2 \log Z] , \end{aligned} \quad (6)$$

where F is the Helmholtz free energy. Therefore, under isothermal conditions where the free energy F is constant, the change in the internal energy is

$$\Delta U = -k_B T \Delta MI = \frac{1}{2} k_B T \Delta r . \quad (7)$$

This indicates that an increase in MI (i.e., a decrease in r) reduces the internal energy U and thereby stabilizes the system. Moreover, given that $MI \geq 0$ (where the equality holds if and only if X and Y are independent), r is more likely to decrease, which demonstrates that $\mathcal{F} = e^{MI}$ is an attractive force. Thus, \mathcal{F} facilitates the information transmission between the informatons, thereby affecting the mutual state of existence. We refer to this novel force \mathcal{F} as the **influential force**.

MI in quantum systems

As presented above, we have obtained the connection between ΔU and ΔMI from thermodynamics. We now examine whether this relationship is also valid in the quantum regime. For this purpose, we specify the MI of quantum bodies in comparison with the *quantum mutual information* MI_{qu} [23-26]. We then show that MI and not MI_{qu} represents the energy produced by the quantum interaction.

Let us assume isothermal conditions. Suppose that two quanta, X and Y , do not interact in a ground state that is the thermal state. This initial mutual independence is satisfied, e.g., if the quanta are created at a particular time point in the universe and if their probability density functions are delta functions at the creation. Under these conditions, we set the ground state as the time ($t = 0$) when both quanta have appeared. We further consider that X and Y constitute a composite system XY . Then, by the definition of MI_{qu} , $MI_{qu} = 0$ at $t = 0$. Moreover, when $t > 0$, MI_{qu} of X and Y is obtained as

$$\begin{aligned} MI_{qu} &= tr(\rho_{XY} \log \rho_{XY}) - tr(\rho_{XY} \log \rho_{X_R} \otimes \rho_{Y_R}) \\ &= tr(\rho_{XY} \log \rho_{XY}) - tr(\rho_{X_R} \log \rho_{X_R}) - tr(\rho_{Y_R} \log \rho_{Y_R}) , \end{aligned} \quad (8)$$

where ρ_{X_R} and ρ_{Y_R} are respectively the reduced density matrices of X and Y , which are respectively obtained as the partial traces $tr_Y \rho_{XY}$ and $tr_X \rho_{XY}$ of the joint density matrix ρ_{XY} . MI_{qu} is equal to the quantum relative entropy S_{qu} of ρ_{XY} and $\rho_{X_R} \otimes \rho_{Y_R}$; that is,

$$MI_{qu} = S_{qu}(\rho_{XY} || \rho_{X_R} \otimes \rho_{Y_R}) , \quad (9)$$

which is associated with the change in the free energy ΔF of the composite system XY according to $\Delta F = k_B T MI_{qu}$ [23]. It is noted that ΔF can be ignored as the first-order perturbation of the state [23].

Instead of MI_{qu} , we will specify MI using the von Neumann entropy of non-reduced density matrices. We again suppose that the two subsystems X and Y are originally independent at $t = 0$. Then, by the definition of MI , $MI = 0$ at $t = 0$. Moreover, when $t > 0$, MI of X and Y can be determined as

$$\begin{aligned} MI &= H(X) + H(Y) - H(X, Y) \\ &= -tr(\rho_X \log \rho_X) - tr(\rho_Y \log \rho_Y) + tr(\rho_{XY} \log \rho_{XY}) , \end{aligned} \quad (10)$$

where ρ_X and ρ_Y are the non-reduced density matrices of X and Y .

What is important is that, whereas MI_{qu} is linked to the change in free energy ΔF , MI is directly connected to the change in thermodynamic entropy ΔS ; that is,

$$\begin{aligned} \Delta S &= -k_B tr(\rho_{XY} \log \rho_{XY}) + k_B tr(\rho_X \otimes \rho_Y \log \rho_X \otimes \rho_Y) \\ &= -k_B tr(\rho_{XY} \log \rho_{XY}) + k_B tr(\rho_X \log \rho_X) + k_B tr(\rho_Y \log \rho_Y) = -k_B MI . \end{aligned} \quad (11)$$

This formula highlights the importance of the MI in quantum systems.

On the basis of the above distinction between the two types of MI, we further discuss their relevance to the internal energy U of the composite system; that is, $k_B T MI_{qu}$ and $-k_B MI$ respectively correspond to the free energy ΔF and thermodynamic entropy ΔS , and therefore

$$\Delta U = \Delta F + T\Delta S = k_B T \cdot (MI_{qu} - MI) \quad . \quad (12)$$

Thus, the difference between the two types of MI is the source of energy. Moreover, $\Delta F = k_B T MI_{qu}$ can be ignored as a first-order perturbation of the state that is initially the thermal state [23]. This is because MI_{qu} is determined for a pair of X and Y initially independent of each other. Additionally, because the situations thereafter are close to thermodynamic equilibrium conditions where $\Delta F = 0$,

$$\Delta U = -k_B T MI \quad , \quad (13)$$

which implies Eq. (7). Finally, the information metric r of X and Y is defined as

$$\begin{aligned} r &= 2H(X, Y) - H(X) - H(Y) \\ &= -2tr(\rho_{XY} \log \rho_{XY}) + tr(\rho_X \log \rho_X) + tr(\rho_Y \log \rho_Y) \end{aligned} \quad (14)$$

and decreases as MI increases. This equation allows the application of the information metric r to quantum systems. To summarize, the MI is the source of changes in the internal energy U and the information metric r in the context of quantum and superstring theory. These observations support the validity of the influential force in the quantum regime.

Attractive influential force in physical systems

Now that we have verified the validity of MI in physical systems, we further consider the relationship between MI and the influential force \mathcal{F} . We first assume that the two informatons X and Y together constitute one composite system XY . We now express the magnitude of the force \mathcal{F} as the ratio of the realization probabilities before (p) and after (p') the change in MI:

$$p = \frac{1}{Z} \exp\left(-\frac{U}{k_B T}\right) \quad , \quad p' = \frac{1}{Z} \exp\left(-\frac{U}{k_B T} + \Delta MI\right) \quad . \quad (15)$$

A **probabilistic representation** of the influential force \mathcal{F} is then written as

$$\mathcal{F} = p'/p = e^{\Delta MI} \quad . \quad (16)$$

If MI is initially zero, then $\Delta MI = MI$ and thus $\mathcal{F} = e^{MI}$ as in Eq. (2). Here, the influential force \mathcal{F} represents the relative increase in the realization probability associated with MI . In turn, MI is linked with the decrease in the system energy by $\Delta U = -k_B T MI$. In this situation, because $MI \geq 0$ is always true, $\mathcal{F} \geq 1$ invariably holds. Therefore, the influential force \mathcal{F} facilitates both the information transmission and the energy transfer between the physical informatons and thereby acts as the attractive force that decreases their information metric r . Hence, we use the notation \mathcal{F}_{att} to represent this attractive influential force. Finally, the above feature of \mathcal{F}_{att} is classified as energetic, according to the conventional distinction of energetic and entropic forces [13,14].

Spontaneous probability of MI

In contrast to the preceding case, when the change in energy is negligible, the spontaneous probability $p(MI)$ that the magnitude of MI of the composite system XY becomes MI follows a canonical distribution:

$$p(MI) = e^{-MI}, \quad (17)$$

which we also call the *infocanonical distribution* (Appendices 2 and 17). That is to say, as MI increases, X and Y constrain each other, which reduces the total number of states of the composite system XY . This leads to a decrease in $p(MI)$ according to the principle of equal a priori probabilities. Notably, $p(MI)$ is equal to the inverse of the attractive influential force \mathcal{F}_{att} [Eq. (A6)], implying the occurrence of a repulsive force. Thus, Eq. (17) is compatible with the increasing tendency of $H(X, Y)$; that is, the second law of thermodynamics.

Repulsive influential force in physical systems

Considering the decreasing propensity of MI , we next analyze the properties of the repulsive influential force, which we denote as \mathcal{F}_{rep} . In contrast with the attractive force \mathcal{F}_{att} , the repulsive force \mathcal{F}_{rep} is an entropic force that results from the system's statistical tendency to increase in entropy [13-16]. Here, to characterize \mathcal{F}_{rep} only between two bodies, we assume that the closed composite system XY is informationally isolated from the rest of the universe. The most typical examples consistent with this model are interatomic and intermolecular interactions, where we can ignore interference from other bodies. Thus, for the time being in this paper, we essentially restrict ourselves to assume an isothermal spacetime in which only the given bodies exist. However, we will consider more general cases in the section "Many-body system" at the end of this chapter, in which we will discuss the magnitude of the repulsive forces counteracting the fundamental natural forces.

To examine the repulsive force, we here regard each physical informaton X and Y as one system. Then, for each of them, an increase in MI decreases the number of states and its realization probability. The repulsion therefore arises from both informatons. This repulsion decreases MI and thus restores the joint entropy $S(X, Y) = k_B[H(X) + H(Y) - MI]$ as well as the realization probability. Given that the repulsion is e^{MI} for each X and Y , the total repulsive force \mathcal{F}_{rep} becomes $e^{MI} \times e^{MI} = e^{2MI}$, which is proportional to the square of the attractive force \mathcal{F}_{att} (Appendix 2). Indeed, in physicochemical systems, \mathcal{F}_{rep} generally increases rapidly as the distance decreases. We will later present examples consistent with this model, namely the Morse potential, dispersion force, and Higgs potential.

Basic equations of the influential force

On the basis of the preceding discussion, we construct a composite formula for the influential force \mathcal{F} . If \mathcal{F}_{att} and \mathcal{F}_{rep} act independently in opposite directions, the expected magnitude of \mathcal{F} becomes $\mathcal{F}_{att}/\mathcal{F}_{rep} = e^{-MI}$. This composite probability equals $p(MI)$ above, supporting its overall validity.

However, because these opposing forces arise in a mutually exclusive manner, we rewrite the equation for \mathcal{F} by adding the two forces using k_1 and k_2 as attractive and repulsive constants, respectively, as

$$\begin{aligned}\mathcal{F} &= \mathcal{F}_{att} + \mathcal{F}_{rep} \\ &= \mathcal{M}(X)\mathcal{M}(Y)[k_1e^{-\acute{\alpha}r} - k_2\mathcal{M}(X)\mathcal{M}(Y)e^{-2\acute{\alpha}r}].\end{aligned}\quad (18)$$

We determine the signatures of \mathcal{F}_{att} and \mathcal{F}_{rep} such that the direction of increasing realization probability is positive. Meanwhile, \mathcal{M} is the information mass, which is specified with a parameter $\acute{\alpha}$, the velocity of light c , the mass of the body m , and the momentum P , such that

$$\mathcal{M}(X) = e^{\acute{\alpha}H(X)} = e^{\acute{\alpha}\beta\sqrt{(mc^2)^2 + (cP)^2}}.\quad (19)$$

In some systems, the exponent in \mathcal{F}_{rep} can be represented as $\exp(-\acute{\alpha}_0r)$ using another constant $\acute{\alpha}_0$. In addition, the force \mathcal{F} constitutes the **probabilistic influential force potential** φ as

$$\begin{aligned}\varphi &= -\int_r^\infty \mathcal{F} dr \\ &= \mathcal{M}(X)\mathcal{M}(Y)\left[-\frac{k_1}{\acute{\alpha}}e^{-\acute{\alpha}r} + \frac{k_2}{2\acute{\alpha}}\mathcal{M}(X)\mathcal{M}(Y)e^{-2\acute{\alpha}r}\right].\end{aligned}\quad (20)$$

Moreover, the force generates the **probabilistic influential force field** $\tilde{\mathcal{V}}$ that comprises an attractive force field $\tilde{\mathcal{V}}_{att}$ and repulsive force field $\tilde{\mathcal{V}}_{rep}$:

$$\tilde{\mathcal{V}}_{att} = -\frac{k_1}{\acute{\alpha}}\mathcal{M}(X)e^{-\acute{\alpha}r}, \quad \tilde{\mathcal{V}}_{rep} = \frac{k_2}{2\acute{\alpha}}\mathcal{M}(X)^2e^{-2\acute{\alpha}r}, \quad \tilde{\mathcal{V}} = \tilde{\mathcal{V}}_{att} + \tilde{\mathcal{V}}_{rep},\quad (21)$$

which pertain to the range of influence of the informaton.

The influential force \mathcal{F} has a remarkable feature that it exists in the **information metric spacetime** (r -spacetime). Nonetheless, the above equations for \mathcal{F} are similar to those for known natural forces, suggesting some connection between them. Notably, the influential force represents the universal relationship between information and energy for natural forces. This is because the known physical forces generate the canonical distribution at energy levels under thermal equilibrium conditions.

(b) Influential force and field theory

This section explores the relationship between the influential force \mathcal{F} and the known natural forces. For this purpose, we discuss the source of MI in terms of field theory.

MI and field theory

We calculate the MI of two fields at points P and Q to derive an expression for the influential force \mathcal{F} in the conventional space. We now let

$$\mu_P[\varphi(P)] = \int \mu[\varphi(P), \varphi(Q)] d\varphi(Q), \quad \mu_Q[\varphi(Q)] = \int \mu[\varphi(P), \varphi(Q)] d\varphi(P)\quad (22)$$

be the probability density functions of $\varphi(P)$ and $\varphi(Q)$ that respectively represent the natural force potentials or the field functions at P and Q . MI of these fields is then calculated as

$$MI[\varphi(P); \varphi(Q)] = \int_P \int_Q \mu[\varphi(P), \varphi(Q)] \log \frac{\mu[\varphi(P), \varphi(Q)]}{\mu_P[\varphi(P)] \mu_Q[\varphi(Q)]} d\varphi(P) d\varphi(Q), \quad (23)$$

where the integrand is the density of MI . We can derive the influential force \mathcal{F} acting between the two fields from Eq. (18); that is, the influential force can express all natural forces. Finally, the information metric r of $\varphi(P)$ and $\varphi(Q)$ is calculated from Eq. (1) by setting X and Y to $\varphi(P)$ and $\varphi(Q)$, respectively.

Influential force and mediator particles

The natural forces acting between two informatons are considered to be transmitted by mediator particles. Their field functions satisfy the Klein–Gordon equation, and the static potential is the solution to the equation when the time derivative is set to zero. If the mass of the mediator particle is m and the distance between two informatons is x , then the Yukawa potential $V_{yukawa}(x) = \exp(-mcx/\hbar)/(4\pi x)$ is derived, where \hbar is the Planck constant divided by 2π .

The mediator particles contribute to the sharing of energy and information between two informatons. It is thus conceivable that the particles mediate MI as a form of action. If we let S_a and $-V(x)$ respectively be the action and energy exerted by the mediator particle during time t , then $S_a = -V(x)t$. By analogy between quantum mechanics and statistical mechanics through the Wick rotation [27], it/\hbar in the path integral corresponds to $\beta (= 1/k_B T)$, where i is the imaginary unit. It follows that $\exp(iS_a/\hbar)$ corresponds to $\exp(\beta S_a/t) = \exp[-\beta V(x)]$, which indicates that S_a follows the canonical distribution. This further supports the relationship between S_a and MI because MI is linked to $-\beta V(x)$ [Eq. (A14)] and follows the canonical distribution [Eq. (17)]. Finally, the four-dimensional path length also follows the canonical distribution because it is proportional to S_a .

In the static setting described above, the total energy of the mediator particle equals the potential energy $V(x)$ between the two informatons. Accordingly, $\Delta MI = -\beta \Delta V(x)$ as described by Appendix 3. Thus, S_a/\hbar and $-\beta V(x)$ are equivalent to MI ; that is,

$$\frac{|S_a|}{\hbar} = \beta |V(x)| = MI. \quad (24)$$

Hence, for a single mediator particle that follows the least-action path, the average ratio of the realization probability with (p') versus without (p) the action is

$$\begin{aligned} p'/p &= \exp\left(\frac{S_a}{\hbar}\right) = \exp[-\beta V(x)] \\ &= \begin{cases} \exp(MI) &= \mathcal{F}_{att} & (S_a > 0) \\ \exp(-MI) &= \frac{1}{\sqrt{-\mathcal{F}_{rep}}} & (S_a < 0) \end{cases}, \end{aligned} \quad (25)$$

where we set $k_1 = k_2 = 1$ and $\hat{\alpha} = 0.5$ for simplicity. This formula indicates the relative realization probability of the information transmission by the particle. Again, the signs of \mathcal{F}_{att} and \mathcal{F}_{rep} are determined such that the direction of increasing realization probability is positive.

In terms of quantum mechanics, Eq. (25) demonstrates that the attractive force \mathcal{F}_{att} represents the fold increase in the bidirectional transition amplitude for mediator particles to move between two informatons following the least-action pathway. Given that $\Delta MI = -\Delta r/2$, $\mathcal{F}_{att} = \exp(MI) = \mathcal{M}(X)\mathcal{M}(Y)\exp(-r/2)$, and \mathcal{F}_{att} is thus a Slater-type real wave function in the information metric spacetime (r -spacetime). Meanwhile, the repulsive force $\mathcal{F}_{rep} [= (\mathcal{F}_{att})^2]$ is another Slater-type function, which indicates the enhancement of the two bidirectional transition amplitudes. Combining the two forces, $\mathcal{F} = \mathcal{F}_{att} + \mathcal{F}_{rep}$ provides the relative probability amplitude of composite information transmission at the particle level. Finally, \mathcal{F} indicates the fold increase in the probability of unidirectional information transmission.

Influential force and path integral

In addition to the static setting above, the influential force \mathcal{F} is related to the path integral in the dynamic setting, assuming the canonical distribution of action S_a . In this case, the path length again follows the canonical distribution. The propagator of the mediator particle is then described as

$$K(x, x'; t) = \int_{x'}^x \exp\left(\frac{iS_a}{\hbar}\right) Dx = \int_{x'}^x \exp(-iMI) Dx, \quad (26)$$

where Dx is the path integral in the *conventional coordinate spacetime* (x -spacetime) and K is the unidirectional transition amplitude from the point x' at $t = 0$ to the point x at time t . Furthermore, considering the least-action principle, the attractive and repulsive forces are expressed using the bidirectional propagator $K_{inf}(x, x'; t) := [K(x, x'; t)]^2$ as

$$\mathcal{F}_{att} = |K_{inf}(x, x'; t)| = e^{MI}, \quad \mathcal{F}_{rep} = |K_{inf}(x, x'; t)|^2 = e^{2MI}. \quad (27)$$

Thus, the influential force is formulated using the path integral of information transmission.

We finally consider the path integral in the r -spacetime. If both informatons X and Y have unit information mass $\mathcal{M} = 1$, then the path integral of $\exp(ir/2)$ from the origin O with the coordinate $r = r' = 0$ at $t = 0$ to the point P with $r = r$ at $t = t$ is given by the least-action principle as

$$K(r, r'; t) = \int_{r'}^r \exp\left(\frac{ir}{2}\right) Dr = e^{\frac{(i-1)r}{4}}, \quad (28)$$

where Dr indicates the path integral in the r -spacetime. Again using the bidirectional propagator $K_{inf}(r, r'; t) := [K(r, r'; t)]^2$, the *basic attractive force* and *basic repulsive force* in the r -spacetime are respectively written as

$$\mathcal{F}_{att} = |K_{inf}(r, r'; t)| = e^{-\frac{1}{2}r}, \quad \mathcal{F}_{rep} = |K_{inf}(r, r'; t)|^2 = e^{-r}. \quad (29)$$

The information metric r equals the shortest path length and is expressed as

$$r = -2 \log |K_{inf}(r, r'; t)|. \quad (30)$$

Thus, the influential force is directly linked to the path integral formalism in the r -spacetime.

Scattering and influential force

We next give an example of the above discussion by considering the relativistic scattering of two particles, X and Y . Let x and y be the four-vectors of X and Y , respectively. The scattering amplitude at x and y mediated by scalar mediator particles is then expressed as

$$S_{fi} = -ig^2 e^{-i(p_1-p_3)x} \Delta_F(x-y) e^{-i(p_2-p_4)y} . \quad (31)$$

Here, p_1 and p_3 are respectively the four-momenta of X before and after the scattering, p_2 and p_4 are respectively the four-momenta of Y before and after the scattering, g is the coupling constant of the mediator particle and both of X and Y , and $\Delta_F(x-y)$ represents the Feynman propagator; that is,

$$\Delta_F(x-y) = \frac{1}{(2\pi)^4} \int \frac{e^{-iq(x-y)}}{q^2 - m^2 + i\epsilon} d^4q , \quad (32)$$

where m is the mass of the mediator particle, and q is the change in the four-momentum such that $q = p_3 - p_1 = p_2 - p_4$.

To avoid the divergence problem of the scattering amplitude, we here use the measurement of g . The scattering probability is then expressed as $p(x, y)$, which is the joint probability density function of X and Y :

$$p(x, y) = C |S_{fi}|^2 = C g^4 |\Delta_F(x-y)|^2 , \quad C = \left[\iint g^4 |\Delta_F(x-y)|^2 d^4x d^4y \right]^{-1} , \quad (33)$$

where C is the normalization constant. Meanwhile, the respective existence probabilities of X and Y are given by $p(x)$ and $p(y)$, which are the marginal probability density functions of $p(x, y)$:

$$\begin{aligned} p(x) &= \int p(x, y) d^4y = \int C g^4 |\Delta_F(x-y)|^2 d^4y \\ p(y) &= \int p(x, y) d^4x = \int C g^4 |\Delta_F(x-y)|^2 d^4x \\ p(x) &= p(y) = \text{const} , \end{aligned} \quad (34)$$

where $p(x)$ and $p(y)$ are equal and constant.

The joint information entropy of the composite system XY is given as

$$\begin{aligned} H(X, Y) &= - \iint p(x, y) \log p(x, y) d^4x d^4y \\ &= - \iint C g^4 |\Delta_F(x-y)|^2 \log [C g^4 |\Delta_F(x-y)|^2] d^4x d^4y . \end{aligned} \quad (35)$$

The information entropies of X and Y are written as

$$\begin{aligned} H(X) &= - \int p(x) \log p(x) d^4x , \quad H(Y) = - \int p(y) \log p(y) d^4y \\ H(X) &= H(Y) = \text{const} , \end{aligned} \quad (36)$$

where $H(X)$ and $H(Y)$ are equal and constant. This implies that when the mediator particles' energy is also summed, the energies of X and Y are preserved throughout the scattering. Thus,

$$E_x = k_B T \cdot H(X) = E_y = k_B T \cdot H(Y) = \text{const} , \quad (37)$$

where E_x and E_y are respectively the energies of X and Y .

We next introduce a fluctuation factor ε_{xy} , which is the ratio of the joint probability density to the product of the marginal probability densities. We also adopt an MI density at x and y , MI_d . Then,

$$\varepsilon_{xy} = \frac{p(x,y)}{p(x)p(y)} , \quad MI_d = p(x,y) \log \varepsilon_{xy} . \quad (38)$$

The MI of X and Y is now expressed as

$$\begin{aligned} MI(X;Y) &= \iint MI_d d^4x d^4y \\ &= \iint Cg^4 |\Delta_F(x-y)|^2 \log \frac{Cg^4 |\Delta_F(x-y)|^2}{p(x)p(y)} d^4x d^4y \\ &= -H(X,Y) + H(X) + H(Y) . \end{aligned} \quad (39)$$

Given that $H(X)$ and $H(Y)$ are constant, the MI difference before and after scattering, MI_{fi} , is

$$\begin{aligned} MI_{fi} &= -\Delta H(X,Y) \\ &= \Delta \iint p(x,y) \log p(x,y) d^4x d^4y \\ &= \Delta \iint Cg^4 |\Delta_F(x-y)|^2 \log [Cg^4 |\Delta_F(x-y)|^2] d^4x d^4y > 0 , \end{aligned} \quad (40)$$

which is always positive, thus demonstrating the reduction in the total entropy of the composite system. Moreover, the third line of the above equation provides a statistical description of the action S_F of the Feynman propagator, which is proportional to the Yukawa potential V_{yukawa} . Thus,

$$MI_{fi} = \frac{|S_F|}{\hbar} = \beta V_{yukawa} . \quad (41)$$

Therefore, despite preserving the energies of X and Y , the scattering reduces the joint entropy and energy of the composite system XY . This is accompanied by the generation of potential energy, which marks the advent of the influential force.

(c) Influential force and physical forces

We discuss gauge transformations in an *information coordinate spacetime*, which we will initially define in this section. All the forces in the conventional coordinate spacetime are transformed into a constant force in the information coordinate spacetime. This constant force is the equivalent of the influential force and is invariant under the transformations of all the gauge groups. Hence, the natural forces are uniformly described by the influential force.

One-dimensional local information coordinate

We revisit field theory from an information theoretical point of view to transform the conventional distance to the information distance. Specifically, we consider the characteristics of the spacetime as a mediator of information. In this paper, we newly construct the local information coordinate so that it reflects the *information distance* in the spacetime. We consider that the information distance

represents the difficulty of information transmission. By applying this difficulty to remeasure the distance in the spacetime, we define the local information coordinate. The information is transmitted by the mediator particles, and the coordinate thus reflects the probability density function of the particles, which follows a canonical distribution in terms of both the path length and potential energy.

First, we discuss the one-dimensional space without considering the potential energy explicitly. Let $p(x) = \alpha \exp(-\alpha x)$ be the probability density function of the mediator particle, where $x (\geq 0)$ is the conventional distance from the origin without gravity while α is a parameter of the distribution. α is set to $E/\hbar v$, where E and v are respectively the energy and velocity of the mediator particle in the conventional coordinate spacetime (x -spacetime). αx then corresponds to the path length Et/\hbar of the mediator particle in the x -spacetime. If the mediator particle has mass, then $\alpha \approx mc^2/\hbar v$ is inversely proportional to the reduced Compton wavelength \hbar/mc , which represents the range of the force. Eventually, within the one-dimensional space, the self-information of the existence of the particle is $I(x) = -\log p(x) = \alpha x - \log \alpha$. This formula directly relates the conventional distance to the difficulty of information transmission. We thus regard $|I(x) - I(0)|$ as the information distance between the origin and point x in the absence of a potential.

Second, we consider the potential energy $V(x)$. The probability density function $p'(x)$ of the body under the canonical distribution is expressed as $p'(x) = \exp[-\beta V(x)]/Z$, where $Z = \int_0^\infty \exp[-\beta V(x)] dx$ is the partition function. Here, $\beta V(x) = r/2$ expresses the path length Et/\hbar of the mediator particle in the information metric spacetime (r -spacetime).

Third, we define another potential $V'(x) = V(x) + \alpha x/\beta$ to compose the two probability density functions $p(x)$ and $p'(x)$, and determine a third probability density function $p''(x) = \exp[-\beta V'(x)]/Z'$, where $Z' = \int_0^\infty \exp[-\beta V'(x)] dx$ is the second partition function. In the above definition for $V'(x) = V(x) + \alpha x/\beta$, not only $V(x)$ but also $\alpha x/\beta$, which is the information level required to carry the mediator particle from the origin to the target at distance x , is regarded as a potential. In turn, the particle reaching the target generates a potential that is equal to $\alpha x/\beta$ and appended to the potential $V(x)$.

Finally, we define the one-dimensional local information coordinate $I_o(x)$ as

$$I_o(x) = -\log p''(x) = \alpha x + \beta V(x) + \log Z' . \quad (42)$$

In differential form,

$$dI_o(x) = \alpha dx + \beta dV(x) = \left[\alpha + \beta \frac{dV(x)}{dx} \right] dx . \quad (43)$$

In conclusion, not only the conventional distance x but also the potential $V(x)$ affects the information transmission. Equation (42) further demonstrates that $V(x)$ alters the information distance and thereby deforms the information coordinate spacetime. In addition, $I_o(x)$ approaches αx as x goes to

infinity because the other two terms on the right-hand side of Eq. (42) become negligible. Hence, from a global point of view, the information coordinate $I_O(x)$ becomes nearly proportional to the conventional coordinate x .

Framework and deformation of the information coordinate spacetime

In the static setting described in the preceding section, the potential $V(x)$ in Eq. (42) can be replaced by the information metric r such that $\beta V(x) = r/2$; that is,

$$I_O(x, r) = \alpha x + \frac{1}{2} r, \quad (44)$$

where $\log Z'$ is set to zero, such that $I_O(0, 0) = 0$ at the origin O . Here, we let the interval of x , dx , be small enough such that $p''(x) \leq 1$ and thus $I_O(x, r) \geq 0$. Again, αx and $r/2$ represent the path length Et/\hbar in the conventional spacetime (x -spacetime) and information metric spacetime (r -spacetime), respectively. Equation (44) combines the conventional distance x and information metric r as a unified coordinate $I_O(x, r)$. In this expression, the **information distance** between the origin O and a point P at the conventional coordinate x and information metric r is denoted $I_O(x, r)$. This is an **entropic representation of the distance** between the points O and P , in which the number of states is proportional to $\exp[-I_O(x, r)]$. As described later, this distance provides an information theoretical basis for constructing higher-dimensional coordinate systems centered on the origin.

In contrast to conventional coordinates, $I_O(x, r) = \alpha x + r/2$ has remarkable properties of the information coordinate spacetime. Above all, $I_O(x, r)$ involves the information metric r , which is lacking in the previous coordinate system. This r denotes a difference in the state associated with that of the potential $V(x)$ between the informatoms. Therefore, if the difference in $V(x)$ is high, the second term $r/2$ becomes much larger than the first term αx . This implies that $V(x)$ strongly deforms the information coordinate spacetime, especially at a microscopic scale.

In the case of the electromagnetic force acting between the proton and electron in a hydrogen atom, the distance is the Bohr radius $a_0 = 5.3 \times 10^{-11}$ m and the Coulomb potential V_C is 4.4×10^{-18} J. If this V_C is mediated by a photon with a wavelength of 4×10^{-7} m, then α is 1.6×10^7 m⁻¹, leading to a ratio of $r/2 = \beta V_C$ to $\alpha x = \alpha a_0$ of 1.3×10^6 . This indicates that, at the atomic level, the effect of $r/2$ on $I_O(x, r)$ far exceeds that of αx with the electromagnetic force. Furthermore, the strong interaction is much more powerful at the nuclear level. According to Inoue's quantum chromodynamics calculation [28], the potential of the strong interaction between two nucleons increases to a maximum of 2800 MeV as the separation decreases from 0.5 to 0 fm. Inoue supposed that a pseudo-scalar meson of 469 MeV mediates the strong interaction. In this instance, α is 2.4×10^{13} m⁻¹ while the ratio of $r/2$ to αx is as much as 9.1×10^{12} . The effect of $r/2$ is thus much greater than that of αx in the strong interaction as well. Meanwhile, when x is sufficiently large such as in the interaction between celestial bodies, the effect of αx is larger than that of $r/2$, because

interactions other than those of gravity are negligible in this case. These examples thus demonstrate that the information metric r has an immense effect on the structure of the information coordinate spacetime at the nuclear level (10^{-15} to 10^{-11} m) as well as at the subnuclear level (10^{-35} to 10^{-15} m), whereas the conventional distance is compatible with the information distance in larger systems.

Metric symmetry

Despite the above inequivalence of ax and $r/2$, because $I_0(x, r)$ is a linear function of both x and r , ax and $r/2$ contribute to $I_0(x, r)$ in a qualitatively similar manner. Therefore, with respect to physical equations in the information coordinate spacetime (I_0 -spacetime), this similarity indicates an approximate symmetry between the variables x and r ; that is, these variables become apparently interchangeable if their interdependence can be ignored. More specifically, the related but distinct physical phenomena within either the conventional coordinate spacetime (x -spacetime) or the information metric spacetime (r -spacetime) can be identically formulated by either x or r , respectively. We designate this putative phenomenon the *metric symmetry*. We will describe relevant cases in the following sections. These cases include the Higgs field, where the metric symmetry illustrates the connection between the internal and external spaces, a prerequisite for explaining gravity.

Influential force in one-dimensional information coordinate spacetime

Within *one-dimensional information coordinate spacetime*, we define the *influential force* \mathcal{F}' acting between the origin O and a point P at which the information distance from the origin is I_0 . Given that $I_0(x, r) = ax + r/2$, I_0 corresponds to an extension of the information metric r by the conventional distance x . Hence, substituting I_0 for ar in Eq. (18) yields

$$\begin{aligned}\mathcal{F}' &= \mathcal{F}'_{att} + \mathcal{F}'_{rep} \\ &= \mathcal{M}(X)\mathcal{M}(Y) [k_1 e^{-I_0} - k_2 \mathcal{M}(X)\mathcal{M}(Y) e^{-2I_0}] ,\end{aligned}\quad (45)$$

which is a *probabilistic representation* of the influential force \mathcal{F}' in the one-dimensional information coordinate spacetime. Similar to the case for Eq. (18), the attractive force \mathcal{F}'_{att} and the repulsive force \mathcal{F}'_{rep} are respectively energetic and entropic forces. Here, k_1 and k_2 are respectively the attractive and repulsive constants.

The force \mathcal{F}' generates the motion of a body in this spacetime, where the attractive force \mathcal{F}'_{att} and repulsive force \mathcal{F}'_{rep} determine the direction and magnitude of \mathcal{F}' . Under the conditions $\mathcal{M}(X) = \mathcal{M}(Y) = 1$ and $k_1 = k_2 = 1$, the attractive term and repulsive term are expressed by

$$\mathcal{F}'_{att} = e^{-I_0}, \quad \mathcal{F}'_{rep} = e^{-2I_0}, \quad (46)$$

which are *probabilistic representations* of the *basic attractive* and *basic repulsive forces*, respectively.

Path integral in one-dimensional information coordinate spacetime

I_O represents the difficulty for a mediator particle to go from one spacetime coordinate to another. The propagator in the one-dimensional conventional spacetime (x -spacetime) is given as

$$K(x, x'; t) = \int_{x'}^x \exp\left(\frac{iS_a}{\hbar}\right) Dx = \int_{x'}^x \exp(-iMI) Dx = \int_{x'}^x \exp[i\beta|V(x)|] Dx . \quad (47)$$

Here, we introduce αx , which represents the information exchange between the mediator particle and spacetime. $\Delta\beta V(x)$ is then replaced with $\Delta[\alpha x + \beta V(x)] = \Delta I_O$. We have

$$\frac{|\Delta S_a|}{\hbar} = -\Delta[\alpha x + \beta V(x)] = -\Delta I_O . \quad (48)$$

Hence,

$$K(x, x'; t) = \int_{x'}^x \exp\{i[\alpha x + \beta V(x)]\} Dx = \int_{x'}^x \exp(iI_O) Dx . \quad (49)$$

The information distance I_O equals the shortest path length and is written using the bidirectional propagator K_{inf} as

$$I_O = A \log |K_{inf}(x, x'; t)| , \quad (50)$$

where A is a constant. This equation directly relates I_O to the path integral formalism.

Four-dimensional polar local information coordinates

We now extend our discussion to four-dimensional spacetime. We define the **four-dimensional polar local information coordinate system** based on the one-dimensional information coordinate. We first suppose the four-dimensional conventional polar coordinates $(ct, r_p, \theta_p, \varphi_p)$, where the radial component r_p corresponds to the one-dimensional conventional coordinate x . We next define the four-dimensional polar local information coordinates $(\alpha c I_O(t), I_O(r_p), I_O(\theta_p), I_O(\varphi_p))$ as

$$I_O(ct, r_p, \theta_p, \varphi_p) = (\alpha c I_O(t), I_O(r_p), I_O(\theta_p), I_O(\varphi_p)) . \quad (51)$$

The time coordinate $\alpha c I_O(t)$ is then defined analogously to the one-dimensional case by $\alpha c I_O(t) = \beta V(t, r_p, \theta_p, \varphi_p) + \alpha ct + \log Z(t)$, in which $V(t, r_p, \theta_p, \varphi_p)$ is the potential at the point $(r_p, \theta_p, \varphi_p)$ at time t within the three-dimensional conventional polar coordinate spacetime, and $Z(t) = \int_0^\infty \int_0^\pi \int_0^{2\pi} \exp[-\beta V(t, r_p, \theta_p, \varphi_p)] r_p^2 \sin\theta_p dr_p d\theta_p d\varphi_p$. Similarly, the coordinate $I_O(r_p)$ is defined by $I_O(r_p) = \beta V(t, r_p, \theta_p, \varphi_p) + \alpha r_p + \log Z_r$, where $Z_r = \int_0^\infty \exp[-\beta V(t, r_p, \theta_p, \varphi_p) - \alpha r_p] dr_p$. $I_O(\theta_p)$ and $I_O(\varphi_p)$ are defined in a similar manner.

The difference r'_{mt} between $\alpha c I_O(t + \Delta t)$ and $\alpha c I_O(t)$ for small Δt is then calculated as $r'_{mt} = \alpha c [I_O(t + \Delta t) - I_O(t)] = \alpha c \Delta I_O(t) \simeq \alpha c (\partial I_O(t) / \partial t) \Delta t \simeq (\beta \partial V / \partial t + \alpha c) \Delta t$. Additionally, the difference r'_{mr} between $I_O(r_p + \Delta r_p)$ and $I_O(r_p)$ for small Δr_p is expressed as $r'_{mr} \simeq (\partial I_O(r_p) / \partial r_p) \Delta r_p = (\beta \partial V / \partial r_p + \alpha) \Delta r_p$. Similarly, the differences $r'_{m\theta}$ and $r'_{m\varphi}$ are respectively calculated as $r'_{m\theta} \simeq (\beta \partial V / \partial \theta_p + \alpha r_p) \Delta \theta_p$ and $r'_{m\varphi} \simeq (\beta \partial V / \partial \varphi_p + \alpha r_p \sin\theta_p) \Delta \varphi_p$. We finally define a distance r'_m between the point $(\alpha c I_O(t + \Delta t), I_O(r_p + \Delta r_p), I_O(\theta_p + \Delta \theta_p), I_O(\varphi_p + \Delta \varphi_p))$ and the point $(\alpha c I_O(t), I_O(r_p), I_O(\theta_p), I_O(\varphi_p))$ in the information coordinate spacetime as

$$\begin{aligned}
r_m'^2 &= -r_{mt}'^2 + r_{mr}'^2 + r_{m\theta}'^2 + r_{m\varphi}'^2 \\
&\simeq -\left[\alpha \frac{\partial I_O(t)}{\partial t}\right]^2 (c\Delta t)^2 + \left[\frac{\partial I_O(r_p)}{\partial r_p}\right]^2 (\Delta r_p)^2 + \left[\frac{\partial I_O(\theta_p)}{\partial \theta_p}\right]^2 (\Delta \theta_p)^2 + \left[\frac{\partial I_O(\varphi_p)}{\partial \varphi_p}\right]^2 (\Delta \varphi_p)^2 \\
&= g_{00}(c\Delta t)^2 + g_{11}(\Delta r_p)^2 + g_{22}(\Delta \theta_p)^2 + g_{33}(\Delta \varphi_p)^2,
\end{aligned} \tag{52}$$

where g_{00} , g_{11} , g_{22} , and g_{33} are respectively the coefficients of $(c\Delta t)^2$, $(\Delta r_p)^2$, $(\Delta \theta_p)^2$, and $(\Delta \varphi_p)^2$. These are the diagonal entries of the metric tensor g_{ij} of the information coordinate spacetime. If we denote $(c\Delta t)^2$, $(\Delta r_p)^2$, $(\Delta \theta_p)^2$, and $(\Delta \varphi_p)^2$ by $(\Delta x_0)^2$, $(\Delta x_1)^2$, $(\Delta x_2)^2$, and $(\Delta x_3)^2$ respectively, then $r_m'^2 = \sum g_{ii}(\Delta x_i)^2$. As an example,

$$g_{11} = \left[\frac{\partial I_O(r_p)}{\partial r_p}\right]^2 = \left(\beta \frac{\partial V}{\partial r_p} + \alpha\right)^2 = [-\beta F(r_p) + \alpha]^2, \tag{53}$$

where $F(r_p)$ is the force in the conventional coordinate spacetime (described below). This indicates that the force $F(r_p)$ is linked with the deformation of spacetime.

Relationship between the radial coordinate and information metric

In the above static setting, the potential $V(t, r_p, \theta_p, \varphi_p)$ can be replaced by the information metric r such that $\beta V(t, r_p, \theta_p, \varphi_p) = r/2$. Therefore, $I_O(r_p, r)$ is represented as

$$I_O(r_p, r) = \alpha r_p + \frac{1}{2}r, \tag{54}$$

where $\log Z_r$ is set to zero, such that $I_O(0,0) = 0$ at the origin O . We let the interval of r_p , dr_p , be small enough such that $I_O(r_p, r) \geq 0$. In this expression, the **information distance** between the origin O and a point P at the conventional polar coordinate r_p and information metric r equals $I_O(r_p, r)$. The potential at the point P is $V'(r_p, r) = V(r_p) + \alpha r_p/\beta$. Finally, Eq. (54) predicts that there is also **metric symmetry** between r_p and r in the four-dimensional spacetime.

Influential force in the four-dimensional information coordinate spacetime

Within the four-dimensional polar local information coordinate spacetime, the **probabilistic representation** of the **influential force** \mathcal{F}' is expressed as

$$\begin{aligned}
\mathcal{F}' &= \mathcal{F}'_{att} + \mathcal{F}'_{rep} \\
&= \mathcal{M}(X)\mathcal{M}(Y) [k_1 e^{-l_0} - k_2 \mathcal{M}(X)\mathcal{M}(Y) e^{-2l_0}],
\end{aligned} \tag{55}$$

which is identical to Eq. (45). The attractive force \mathcal{F}'_{att} and repulsive force \mathcal{F}'_{rep} are respectively energetic and entropic forces. Here, k_1 and k_2 are respectively the attractive and repulsive constants.

Finally, under the conditions $\mathcal{M}(X) = \mathcal{M}(Y) = 1$ and $k_1 = k_2 = 1$, the **basic attractive** and **basic repulsive forces** are expressed as

$$\mathcal{F}'_{att} = e^{-l_0}, \quad \mathcal{F}'_{rep} = e^{-2l_0}, \tag{56, 57}$$

which are identical to the expressions in Eq. (46).

Constant attractive force in the information coordinate spacetime

When we employ the information coordinate system, the basic forces within the conventional coordinate system are transformed into a new form. The dimensionless potential $\Phi'(I_O)$ in the information coordinate spacetime is $\Phi'(I_O) = \beta V'(I_O) = I_O$, in terms of $I_O = I_O(r_p, r)$. Therefore, within this particular spacetime, all basic natural forces in the conventional coordinate spacetime are transformed into a universal attractive force $F'(I_O)$:

$$F'(I_O) = -\frac{d\Phi'(I_O)}{dI_O} = -1. \quad (58)$$

This is the constant force that operates everywhere in the I_O -spacetime. Notably, this equation expresses both attractive and repulsive forces when the conventional radial coordinate r_p is considered as a reference. In contrast, when the information coordinate I_O is used as in the above equation, both forces are transformed into an attractive force. Eventually, $F'(I_O)$ is a symmetric central force with a constant magnitude -1 within this polar local information coordinate spacetime.

Gauge invariance in the information coordinate spacetime

We now discuss gauge transformations of the basic natural forces in the four-dimensional polar local information coordinate spacetime (I_O -spacetime). The equation $F'(I_O) = -1$ always holds, and $F'(I_O)$ is thus invariant under any gauge transformation as shown below. Let Λ be a gauge transformation, let r_{p1} be the transformed radial coordinate Λr_p , and let $V_1(r_{p1})$ be the transformed potential $\Lambda V(r_p)$. Then, $I_O(r_p)$ is transformed into

$$I_1(r_{p1}) = \alpha r_{p1} + \beta V_1(r_{p1}) + \log Z'_1 = \beta V'_1(r_{p1}) + \log Z'_1, \quad (59)$$

where $V'_1(r_{p1}) = V_1(r_{p1}) + \alpha r_{p1}/\beta$ and $Z'_1 = \int_0^\infty \exp[-\beta V_1(r_{p1}) - \alpha r_{p1}] dr_{p1}$. The transformed potential $\Phi'(I_1)$ in the I_O -spacetime is thus expressed as $\Phi'(I_1) = \beta V'_1(I_1) = I_1 - \log Z'_1$, and the force derived from it is $F'(I_1) = -d\Phi'(I_1)/dI_1 = F'(I_O)$. Thus, $F'(I_O) = -1$ is invariant under the transformations of all gauge groups, such as $U(1)$, $SU(2)$, $SU(3)$, and $O(3, 1)$. In conclusion, $F'(I_O)$ is a universal gauge force that arises from these general gauge symmetries. These observations highlight an important and unusual aspect of the information coordinate spacetime.

Equation (58) also implies that the potential $\Phi'(I_O)$ is a gauge potential. In general, gauge potentials have degrees of freedom that allow gauge transformations. Indeed, the potential $\Phi'(I_O)$ has degrees of freedom derived from those of $V(x)$, which is a gauge field in the conventional spacetime (x -spacetime). Given that $V(x) = -k_B T \log \mathcal{F}$, the influential force \mathcal{F} in the information metric spacetime (r -spacetime) is associated with $\Phi'(I_O)$ as the source of degrees of freedom. This indicates that there is a close interconnection between $F'(I_O)$ in the I_O -spacetime and \mathcal{F} in the r -spacetime.

Mechanistic force and probabilistic force

In addition to the above relationships, the constant force $F'(I_O) = -1$ can be connected to the probabilistic influential force \mathcal{F}' in the I_O -spacetime. Let us consider the potential difference Φ'_d between the origin O and a point P with radial coordinates of $I_O(0,0) = 0$ and $I_O(r_p, r) = I_O$, respectively. In this case, $\Phi'_d = \Phi'[I_O(r_p, r)] - \Phi'[I_O(0,0)] = I_O(r_p, r) - I_O(0,0) = I_O$. The probability density function of Φ'_d is expressed as $p(\Phi'_d) = e^{-I_O}$. Considering that this expression is identical to that of the basic attractive force $\mathcal{F}'_{att} = e^{-I_O}$ given as Eq. (56), we regard that $F'(I_O)$ is an equivalent of, and is a **mechanistic representation** of, the probabilistic attractive force \mathcal{F}'_{att} .

The attractive influential force \mathcal{F}'_{att} is a Slater-type function that represents the bidirectional transition amplitude between two informatoms along the least-action path. The propagator in the four-dimensional polar local information coordinate spacetime (I_O -spacetime) can be derived from that in the one-dimensional conventional coordinate spacetime (x -spacetime) given as Eq. (47). The path integral of $\exp(iI_O)$ from the origin O with the radial coordinate $I_O(0,0) = I'_O = 0$ at $t = 0$ to the point P with $I_O(r_p, r) = I_O$ at $t = t$ is given by the least-action principle as

$$K(I_O, I'_O; t) = \int_{I'_O}^{I_O} \exp(iI_O) DI_O = e^{\frac{(i-1)I_O}{2}}, \quad (60)$$

where DI_O indicates the path integral in the I_O -spacetime. Again, we introduce the bidirectional propagator $K_{inf}(I_O, I'_O; t) := [K(I_O, I'_O; t)]^2$. The basic attractive force \mathcal{F}'_{att} is then written as

$$\mathcal{F}'_{att} = |K_{inf}(I_O, I'_O; t)| = e^{-I_O}. \quad (61)$$

The information distance I_O equals the shortest path length and is derived as

$$I_O = -\log |K_{inf}(I_O, I'_O; t)|. \quad (62)$$

This equation directly relates I_O to the path integral formalism in the I_O -spacetime, where the path integration is performed over the whole space. Regarding the propagator $K_{inf}(I_O, I'_O; t)$, it is noteworthy that its phase angle, potential, and path length all have the same single value, I_O . Moreover, this coherence is conserved among all four natural forces. In summary, the constant mechanistic force $F'(I_O) = -d\Phi'(I_O)/dI_O = -1$ is derived from this remarkable uniformity.

The main difference between $F'(I_O)$ and conventional natural forces is that the strengths of most of the latter follow an inverse-square law. This law is explained as the dilution effect of point-source radiation into a three-dimensional sphere whose surface area is proportional to the square of its radius. In contrast, I_O and its equivalence $\Phi'(I_O)$ are derived from the shortest path length described above, where the path integral is determined in view of the diffusion effect over the entire space. Moreover, $\Phi'(I_O) = I_O = \alpha r_p + r/2$ is invariant irrespective of the four-dimensional path lengths, r_p and r . The three-dimensional dilution effect thus no longer needs to be considered in obtaining $F'(I_O)$.

Uniform description of natural forces

We again use x instead of r_p to deal with the conventional coordinate spacetime (x -spacetime). From the gauge force $F'(I_O)$ in the information coordinate spacetime (I_O -spacetime), we derive a general force $F(x)$ in the x -spacetime as $F(x) = -\partial V(x)/\partial x = k_B T[\alpha + F'(I_O) \cdot \partial I_O(x)/\partial x]$. In other words, $F(x)$ originates from $V(x)$, representing all the natural force potentials in the x -spacetime. In turn, $V(x)$ is derived from the mechanistic influential force $F'(I_O)$ in the I_O -spacetime. Moreover, the versatility of $F(x)$ is consistent with the universality of the influential force, which has been deduced from Eq. (23). In summary, the observable force $F(x)$ in the x -spacetime represents all four basic forces and is essentially an expression of the mechanistic influential force $F'(I_O)$ in the I_O -spacetime.

On the basis of the above discussions, we can describe all the natural interactions in terms of information transmission. When $V(x)$ is a static potential, $\Delta V(x) = -k_B T \Delta MI$ according to Eq. (A14). Then,

$$F(x) = -\frac{\partial V}{\partial x} = k_B T \frac{\partial MI}{\partial x}. \quad (63)$$

This formula demonstrates that all the natural forces are uniformly described with MI . Together with the universal gauge symmetry explained by the influential force $F'(I_O)$, and with the equivalence between $F'(I_O)$ and $F(x)$, this equation provides a plausible description of a **Unification Theory**. This concept would serve as a basis for expanding our knowledge of the relationship between energy, information, and the universe. The following section focuses on the mechanistic aspect of the influential force, whereby we discuss the manner in how informatons are forced to move under the influential force.

(d) Motion generated by the influential force

This section considers the motion generated by the influential force \mathcal{F}' . \mathcal{F}' in the information coordinate spacetime is determined by $I_O = \alpha x + r/2$, and the motion is thus characterized by I_O and its relationship with x and r . We here describe several equations of motion using different coordinates. We then demonstrate that \mathcal{F}' is a ‘mechanistic force’ in terms of the probability p , and that it plays a pivotal role in various physical systems. We finally discuss the significance of the repulsive force.

Uniform acceleration in the information metric spacetime

We first discuss the relationship between a uniform acceleration in the information metric spacetime (r -spacetime) and motion in the conventional coordinate spacetime (x -spacetime). Let r be the

information metric of two informatons and $r(0)$ be its initial value. Let x be the conventional distance between the bodies in the conventional spacetime and $x(0)$ be its initial value. We here consider the case in which the internal energy U is approximated by the potential energy V . We then have $dV \simeq dU = \frac{1}{2}k_B T dr$. Thus,

$$m \frac{d^2 x}{dt^2} = -\frac{dV}{dx} \simeq -\frac{1}{2}k_B T \frac{dr}{dx} = \text{const} . \quad (64)$$

This equation indicates that r and x have a linear relationship. If we let $\alpha_{inf} (< 0)$ be the acceleration in the r -spacetime, then

$$r(t)-r(0) = \frac{1}{2} \alpha_{inf} t^2, \quad x(t)-x(0) = -\frac{1}{2} \sqrt{-\frac{\alpha_{inf} k_B T}{2m}} t^2, \quad (65)$$

where m is the reduced mass in the x -spacetime. The right-hand equation shows that the motion in the r -spacetime also generates the motion of uniform acceleration in the x -spacetime. This motion resembles that in classical mechanics under constant gravitational and electromagnetic forces.

Equation of motion along the r -axis in the information metric spacetime

We next examine the more general movement in the information metric spacetime (r -spacetime). We derive Newton's equation of motion from Lagrange's equation of motion in this spacetime. Let V be the potential difference between the informatons. The change in the Lagrangian $d\mathcal{L}$ of an informaton is represented as $d\mathcal{L} = dT_K - dV$, where dT_K and dV are respectively changes in the kinetic energy and potential energy. We here consider only static settings, such as the static potential generated by mediator particles. Then, as in the preceding case, dT_K is negligible relative to dV . The change in the information metric becomes $dr = 2\beta(dV + dT_K) = 2\beta dV$. We then have

$$d\mathcal{L} = dT_K - dV = m \frac{dr}{dt} d\left(\frac{dr}{dt}\right) - \frac{dr}{2\beta}, \quad (66)$$

where m is the reduced mass of the informaton. Substituting this expression into Lagrange's equation of motion $d/dt [\partial\mathcal{L}/\partial(dr/dt)] - \partial\mathcal{L}/\partial r = 0$, we derive Newton's equation of motion as

$$m \frac{d^2 r}{dt^2} = -\frac{1}{2\beta} = -\frac{dV}{dr} =: F_r, \quad (67)$$

where F_r is the force in the r -spacetime along the r -axis. This equation demonstrates that F_r in the r -spacetime is constant, as is the case for the mechanistic attractive force $F'(I_O)$ in the information coordinate spacetime (I_O -spacetime) [Eq. (58)]. Conversely, F_r is related to \mathcal{F}_{att} by

$$\frac{d}{dr} \log \mathcal{F}_{att} = \frac{dMI}{dr} = -\frac{1}{2} = \beta F_r. \quad (68)$$

Thus, a body in the r -spacetime moves under the constant force F_r , which is derived from the attractive force \mathcal{F}_{att} . This results in uniform acceleration toward the origin, which leads to uniform acceleration also in the conventional coordinate spacetime (x -spacetime), as mentioned above.

We finally extend Eq. (68) with $\mathcal{F} = \mathcal{F}_{att} + \mathcal{F}_{rep}$. Let r be the information metric of two informatons G_1 and G_2 , and let their information masses be $\mathcal{M}(G_1)$ and $\mathcal{M}(G_2)$, respectively. The macroscopic equation of motion is then described using the information metric r as

$$\begin{aligned} F_r &= \frac{\mathcal{M}(G_1)\mathcal{M}(G_2)}{\mathcal{M}(G_1)+\mathcal{M}(G_2)} \frac{d^2r}{dt^2} = k_B T \frac{d}{dr} \log(\mathcal{F}_{att} + \mathcal{F}_{rep}) = k_B T \frac{d}{dr} \log \mathcal{F} \\ &= k_B T \frac{d}{dr} \log[k_1 e^{-\acute{\alpha}r} - k_2 \mathcal{M}(G_1)\mathcal{M}(G_2) e^{-2\acute{\alpha}r}] . \end{aligned} \quad (69)$$

In this paper, if the antilogarithm is negative, then the logarithm is interpreted as that of the absolute value of the antilogarithm. The above equation expresses the direction and magnitude of F_r along the r -axis, which is derived from the influential force \mathcal{F} .

Equation of motion along the p -axis in the information metric spacetime

In addition to motion along the r -axis, we consider movement in terms of the probability density p . When the system follows the canonical distribution for the potential difference V , the informatons move in the direction in which p increases. We formulate another equation of motion by performing a coordinate transformation from r to p (Appendix 1). As described in Appendix 22, the probability density of MI becoming MI , $p(MI)$, should be balanced with that of the potential difference, $p(V)$; that is, $p = \exp(-MI) = \exp(-\beta V)/Z$. Thus, $dV = -dp/(\beta p)$. We have

$$d\mathcal{L} = dT_K - dV = m \frac{dp}{dt} d\left(\frac{dp}{dt}\right) + \frac{dp}{\beta p} . \quad (70)$$

We now apply Lagrange's equation of motion $d/dt [\partial\mathcal{L}/\partial(dp/dt)] - \partial\mathcal{L}/\partial p = 0$. Given that $p^{-1} = p(MI)^{-1} = \mathcal{F}_{att}$ [Eqs. (A5, A6)], we obtain an equation of motion of Newton's type as

$$m \frac{d^2p}{dt^2} = \frac{1}{\beta p} = k_B T \mathcal{F}_{att} =: F_p , \quad (71)$$

where F_p is the force in the information metric spacetime (r -spacetime) in terms of the probability. While \mathcal{F}_{att} increases the relative probability of information transmission as a probabilistic force, the above equation shows that \mathcal{F}_{att} can also be regarded as a mechanistic force acting along the p -axis. This notion supports the previous discussion in Section (c) that the influential force can be formulated from both probabilistic and mechanistic points of view.

We further extend Eq. (71) with $\mathcal{F} = \mathcal{F}_{att} + \mathcal{F}_{rep}$. We employ the same parameters used for Eq. (69). In this case, the equation of motion is written using the probability density p as

$$\begin{aligned} F_p &= \frac{\mathcal{M}(G_1)\mathcal{M}(G_2)}{\mathcal{M}(G_1)+\mathcal{M}(G_2)} \frac{d^2p}{dt^2} = k_B T \mathcal{F} \\ &= k_B T \mathcal{M}(G_1)\mathcal{M}(G_2) [k_1 e^{-\acute{\alpha}r} - k_2 \mathcal{M}(G_1)\mathcal{M}(G_2) e^{-2\acute{\alpha}r}] . \end{aligned} \quad (72)$$

This equation expresses the direction and magnitude of the force F_p acting along the p -axis, which is

derived from the influential force \mathcal{F} . The two formulas for F_r and F_p complement each other and together indicate the motion of informatons in the r -spacetime for a canonical distribution.

In summary, the influential force \mathcal{F} increases the probability of information transmission between physical informatons, which generates the mechanistic force in the conventional coordinate spacetime (x -spacetime). The force $F(x)$ in the x -spacetime is then related to \mathcal{F} in the r -spacetime by

$$F(x) = k_B T \frac{\partial MI}{\partial x} = k_B T \frac{\partial}{\partial x} \log \mathcal{F} . \quad (73)$$

Thus, as a mechanistic force acting along the p -axis, \mathcal{F} becomes the common source of physical forces in our spacetime. Furthermore, the probability is a general measure, and the influential force \mathcal{F} thus acts in a wide variety of systems other than physical systems, as discussed in later sections.

Equation of motion along the I_0 -axis in the information coordinate spacetime

We next derive equations of motion in the information coordinate spacetime (I_0 -spacetime), where the conventional distance x and the information metric r together constitute a unified coordinate I_0 according to $I_0 = \alpha x + r/2$. With the **probabilistic influential force** $\mathcal{F}' = \mathcal{F}'_{att} + \mathcal{F}'_{rep}$, we write a mechanistic equation of motion of Newton's type as

$$\begin{aligned} F'(I_0) &= \frac{\mathcal{M}(G_1)\mathcal{M}(G_2)}{\mathcal{M}(G_1)+\mathcal{M}(G_2)} \frac{d^2 I_0}{dt^2} = \frac{d}{dI_0} \log(\mathcal{F}'_{att} + \mathcal{F}'_{rep}) = \frac{d}{dI_0} \log \mathcal{F}' \\ &= \frac{d}{dI_0} \log[k_1 e^{-I_0} - k_2 \mathcal{M}(G_1)\mathcal{M}(G_2) e^{-2I_0}] , \end{aligned} \quad (74)$$

which is the same as Eq. (69) for F_r , except that I_0 is used instead of r . This equation is an extension of $F'(I_0) = -d\Phi'(I_0)/dI_0 = -1$ [Eq. (58)] in that it also contains a repulsive term.

We here rewrite the preceding equations using the attractive and repulsive terms. We set $\mathcal{M}(X) = \mathcal{M}(Y) = 1$ and $k_1 = k_2 = 1$ so that the **mechanistic influential force potential** $\Phi'(I_0)$ at the origin is zero. The probabilistic forces \mathcal{F}'_{att} and \mathcal{F}'_{rep} then express the basic attractive force e^{-I_0} and the basic repulsive force e^{-2I_0} , respectively. $\Phi'(I_0)$ is now rewritten as

$$\Phi'(I_0) = k'_1 \Phi'_{att}(I_0) + k'_2 \Phi'_{rep}(I_0) , \quad (75)$$

where $\Phi'_{att}(I_0) = \log \mathcal{F}'_{att} = -I_0$ and $\Phi'_{rep}(I_0) = -\log \mathcal{F}'_{rep} = 2I_0$ are respectively the **basic attractive** and **basic repulsive potentials**. Meanwhile, k'_1 and k'_2 are respectively the attractive and repulsive coefficients. As detailed later in this chapter, k'_2 is statistically determined as

$$k'_2 = \frac{1}{N-1} k'_0 , \quad (76)$$

where N is the total number of bodies in the system, and k'_0 is the coefficient for a two-body system. The **mechanistic influential force** $F'(I_0)$ is also rewritten using the above k'_1 and k'_2 as

$$F'(I_0) = k'_1 F'_{att}(I_0) + k'_2 F'_{rep}(I_0) , \quad (77)$$

where $F'_{att}(I_0)$ and $F'_{rep}(I_0)$ are respectively the **basic attractive** and **basic repulsive forces**.

$F'_{att}(I_0)$ was previously denoted $F'(I_0)$ and is a mechanistic representation of the probabilistic attractive force \mathcal{F}'_{att} . It is a constant attractive force acting toward the origin:

$$F'_{att}(I_0) = \frac{d}{dI_0} \log \mathcal{F}'_{att} = -\frac{d\Phi'_{att}(I_0)}{dI_0} = -1. \quad (78)$$

This is universally invariant upon gauge transformations, which allows the representation of all basic forces. While $F'_{att}(I_0)$ is an energetic attractive force, $F'_{rep}(I_0)$ is an entropic repulsion acting against $F'_{att}(I_0)$ and is a mechanistic representation of the probabilistic repulsive force \mathcal{F}'_{rep} . $F'_{rep}(I_0)$ is a constant repulsive force acting from the origin:

$$F'_{rep}(I_0) = -\frac{d}{dI_0} \log \mathcal{F}'_{rep} = \frac{d\Phi'_{rep}(I_0)}{dI_0} = 2. \quad (79)$$

This is again universally invariant upon gauge transformations, which suggests a certain generality of $F'_{rep}(I_0)$, as will be discussed later.

Equation of motion along the p -axis in the information coordinate spacetime

As in the r -spacetime, another equation of motion is derived in the I_0 -spacetime through a coordinate transformation from I_0 to p (Appendix 1). We consider that MI tends to zero when $I_0 \rightarrow \infty$. Then, the probability density of MI being MI , $p(MI)$, should balance with the probability density of the information distance, $p(I_0)$; that is, $p = \exp(-MI) = \exp(-I_0)$. We thus have

$$\begin{aligned} F'_p(I_0) &= \frac{\mathcal{M}(G_1)\mathcal{M}(G_2)}{\mathcal{M}(G_1)+\mathcal{M}(G_2)} \frac{d^2p}{dt^2} = \mathcal{F}' \\ &= \mathcal{M}(G_1)\mathcal{M}(G_2) [k_1 e^{-I_0} - k_2 \mathcal{M}(G_1)\mathcal{M}(G_2) e^{-2I_0}], \end{aligned} \quad (80)$$

where $F'_p(I_0)$ is the force in the I_0 -spacetime in terms of probability.

$F'(I_0)$ and $F'_p(I_0)$ together indicate the motion of informatons within this specific spacetime. Despite this complementarity, however, $F'_p(I_0) = \mathcal{F}'$ demonstrates that the influential force \mathcal{F}' is originally the mechanistic force $F'_p(I_0)$ acting along the p -axis and drives to increase the probability of the information transmission. Accordingly, \mathcal{F}' is related to the observable force $F(x)$ within the conventional coordinate spacetime (x -spacetime) by

$$F(x) = k_B T \left(\alpha + \frac{\partial}{\partial x} \log \mathcal{F}' \right), \quad (81)$$

which includes the term α , representing the effect of mediator particles. This equation is identical to Eq. (73) when α tends to zero, and these equations together imply the critical role of probability in physics. Given the universality of the probability, this leads to an understanding that all the bodies in the universe move under the influential force $\mathcal{F}' = F'_p(I_0)$ of probability.

In addition to the force \mathcal{F}' , the **probabilistic influential force potential** $\varphi'(I_0)$, which is an integration of \mathcal{F}' , is a mechanistic potential along the p -axis. When the metric symmetry is applied, $\varphi'(I_0)$ takes shape as real potentials in both the r -spacetime and x -spacetime. In summary, the influential force \mathcal{F}' resides in the I_0 -spacetime, where it exhibits its most distinctive properties.

Attractive and repulsive forces in the information coordinate spacetime

The configuration of the mechanistic influential force $F'(I_O)$ comprises both the attractive force $F'_{att}(I_O)$ and repulsive force $F'_{rep}(I_O)$. As described in the next section, $F'(I_O)$ is well compatible with physicochemical interactions, such as the Morse oscillator (MO) potential. In addition to these composite forces, the basic natural forces are assigned to $F'_{att}(I_O)$. However, whether there exists a repulsive force that counteracts these basic natural forces is unknown. We here discuss the need to consider a repulsive force associated with the basic forces. We suggest that this particular $F'_{rep}(I_O)$ can be measured by paying attention to its statistical properties.

Two major problems arise as important consequences of the attractive force $F'_{att}(I_O)$. First, as Einstein anticipated, a universal attractive force should contract the universe and lead to its collapse. This assumption also applies to the I_O -spacetime because the information distance I_O is macroscopically equivalent to the conventional distance x . While Einstein introduced a cosmological constant that counteracts gravity, recent studies have shown that 68% of the mass-energy density of the universe can be attributed to “dark energy” [29], a cosmological repulsive force.

Second, from an information theoretical point of view, attractive motion under the effect of $F'_{att}(I_O)$ will reduce the entropy of a system by $\Delta I_O(r_p, r)$ because of information transmission. We emphasize that the energy transfer is coupled with the sharing of information as $\Delta MI = -\Delta I_O$, which decreases the joint entropy, such that $\Delta S(G_1, G_2) = k_B[H(G_1) + H(G_2) - \Delta MI]$ for two informatons, G_1 and G_2 . Therefore, at least locally, it appears that all the fundamental forces act against the second law of thermodynamics. It is then expected that a repulsive force will be generated to restore the joint entropy. Taken together, these two issues strongly suggest the existence of a universal repulsive force $F'_{rep}(I_O)$ that opposes $F'_{att}(I_O)$.

Despite the indications of $F'_{rep}(I_O)$, however, the existence of $F'_{rep}(I_O)$ is not consistent from system to system. $F'_{rep}(I_O)$ is evident in some physicochemical interactions (e.g., interatomic and intermolecular forces). However, $F'_{rep}(I_O)$ acting against the four basic forces (e.g., antigravity) has not been identified. In clarifying these contradictory observations, it is noted that there is a statistical difference among the systems because $F'_{rep}(I_O)$ reflects a statistical tendency to increase the entropy. We propose that the inconsistency between systems can be explained by considering the number N of information-sharing bodies in each statistical ensemble.

Two-body system and many-body system

In the first category mentioned above, the systems typically comprise only two bodies, as exemplified by the MO and the dispersion force. It could be regarded that these systems are virtually isolated from the rest of the universe in terms of information transmission. $F'_{rep}(I_O)$ results from the statistical

tendency of the whole system, and it thus most typically emerges within such isolated systems with $N = 2$ bodies, in which the two informatons share the repulsive information.

As an example of the two-body system, we discuss the Higgs field later in this chapter. We argue that the Higgs field is an excellent example of the influential force field in that it is expressed by $\varphi'(I_0)$ and comprises both the attractive and repulsive terms. We then show that this field is an equivalent of the gravitational field. $F'_{rep}(I_0)$ acting against $F'_{att}(I_0)$ in this case is thus antigravity. However, this repulsive force has a very short effective range, which has allowed its evasion from investigation.

In the second category, the basic forces act within systems comprising a large number N of bodies (e.g., the universe). The change in entropy $\Delta H(G_1)$ becomes shared by the many bodies in the system. This particular information sharing occurs through a mechanism detailed in the “Many-body system” section. When $H(G_1)$ is decreased by $\Delta I_0(G_1, G_2)$ upon information transmission with G_2 , the information distances between G_1 and the other bodies tend to expand to restore $H(G_1)$. The metric symmetry then works and expands the conventional distances between bodies. The average repulsion at each interaction is then proportional to $1/(N - 1)$.

Description of the composite force $F'(I_0)$

We finally revisit the equation for $F'(I_0)$. Given the uncertainty regarding the consistent existence of the repulsive force, we use coefficients k'_1 and k'_2 to describe the composite force $F'(I_0)$ as

$$F'(I_0) = k'_1 F'_{att}(I_0) + k'_2 F'_{rep}(I_0) , \quad (82)$$

where k'_2 is the statistical factor that expresses the average degree of shared information distributed to each informaton. Specifically, in a two-body system, k'_2 takes its maximum value. Conversely, in the many-body system with a very large number N of bodies, k'_2 becomes $1/(N - 1)$ of the maximum value, which is almost zero. k'_2 is thus determined as

$$k'_2 = \frac{1}{N-1} k'_0 , \quad (83)$$

where k'_0 is the coefficient for the two-body system.

Regarding the interatomic and intermolecular forces acting between two bodies, k'_1 and k'_2 vary with distance, and $F'(I_0)$ thus changes with I_0 . Conversely, concerning the basic forces, k'_1 is constant while $k'_2 \simeq 0$. Regarding the whereabouts of the repulsive force, we will find an answer in the section “Many-body system,” and this answer possibly explains the origin of dark energy. In summary, despite being unfamiliar to most scientists, the repulsive force $F'_{rep}(I_0)$ is present in almost every case, which suggests the universality of the entropic repulsion counteracting the energetic attractive force $F'_{att}(I_0)$.

(e) Influential force and quantum oscillators

The influential force \mathcal{F}' is a superposition of Slater-type functions \mathcal{F}'_{att} and \mathcal{F}'_{rep} in the I_O -spacetime. In this canonical spacetime, the radial coordinate I_O is symmetric with respect to the origin, where the potential $\Phi'(I_O)$ is also symmetric such that $F'_{att}(I_O) = -1$ and $F'_{rep}(I_O) = 2$ are constant. This symmetry in the potential leads one to assume that \mathcal{F}' is associated with quantum oscillators. Moreover, given that \mathcal{F}' is involved in the generation of physical forces, it is natural that \mathcal{F}' is closely linked to the essential constituents of spacetime. In this connection, many of those constituents (i.e., quantum fields [30] and superstrings [8]) are composed of harmonic oscillators. Here, we suggest that \mathcal{F}' likely represents an anharmonic oscillator comprising the superposition of harmonic oscillators.

Accommodating coordinate system

At the subatomic level, the effect of the dimensionless potential $\beta V(x)$ far exceeds that of the conventional distance αx . We then have $I_O = \alpha x + r/2 \simeq r/2$, where $r/2 = \beta V(x)$ in the above static setting. For now, we tentatively assume that \mathcal{F}' in the I_O -spacetime equals \mathcal{F} in the information metric spacetime (r -spacetime). In this situation, on the basis of the ubiquitous nature of harmonic oscillators, we discuss the possibility that \mathcal{F}' has connections with such oscillators. We thus assume that a mass point with mass m performs a harmonic oscillation with angular frequency ω and the origin at the center of the conventional coordinate spacetime (x -spacetime). The oscillator potential $V(x)$ is proportional to the square of the conventional distance x , while the information metric r is proportional to $V(x)$. Then, \mathcal{F} can be linked to $r \propto V(x) \propto x^2$ in any energy state.

We begin by considering the ground state of the harmonic oscillators. The wave function $\phi_0(x)$ in the x -spacetime is then expressed as

$$\phi_0(x) = \left(\frac{2m\omega}{h} \right)^{\frac{1}{4}} \exp\left(-\frac{\pi m\omega x^2}{h} \right). \quad (84)$$

If we set $r = \pi m\omega x^2 / (\hbar)$ and $k_1 \mathcal{M}(G_1) \mathcal{M}(G_2) = (2m\omega/h)^{1/4}$, then $\mathcal{F}_{att} = \phi_0(x)$. Meanwhile, because $\mathcal{F}_{rep} \propto (\mathcal{F}_{att})^2$, \mathcal{F}_{rep} can be expressed as another wave function $\tilde{\phi}_0(x)$ of a harmonic oscillator with angular frequency 2ω :

$$\tilde{\phi}_0(x) = \left(\frac{4m\omega}{h} \right)^{\frac{1}{4}} \exp\left(-\frac{2\pi m\omega x^2}{h} \right). \quad (85)$$

We additionally set $k_2 [\mathcal{M}(G_1) \mathcal{M}(G_2)]^2 = (4m\omega/h)^{1/4}$. We then have $\mathcal{F}_{rep} = -\tilde{\phi}_0(x)$. Thus,

$$\mathcal{F} = \mathcal{F}_{att} + \mathcal{F}_{rep} = \phi_0(x) - \tilde{\phi}_0(x). \quad (86)$$

This suggests a new point of view regarding the properties of the influential force \mathcal{F} in the r -spacetime. That is to say, \mathcal{F} can be expressed as the difference between ground-state wave functions of harmonic oscillators in the x -spacetime. In addition, because the oscillators $\phi_0(x)$ and $\tilde{\phi}_0(x)$ have different angular frequencies, \mathcal{F} represents an anharmonic oscillator in the ground state. Thus, the possible relationship between \mathcal{F} and the harmonic oscillators conforms to our initial assumption.

Oscillators in the information metric spacetime

Despite the consistency with the original model, the influential force \mathcal{F} described until now does not represent oscillators in the higher energy states. Hence, to meet the wider configurations of wave functions in the information metric spacetime (r -spacetime), we extend the format of the influential force \mathcal{F} . Generally, in the n -th excited state, the wave function $\phi_n(x)$ of the harmonic oscillator is expressed as $\phi_n(x) = A_n H_n(\xi) \exp(-\xi^2/2)$, where $\xi = (2\pi m\omega/h)^{1/2}x$ and $A_n = (2m\omega)^{1/4}/[2^{n/2}(n!)^{1/2}]$. $H_n(\xi)$ is the Hermite polynomial defined as $H_n(\xi) = (-1)^n \exp(\xi^2) (d^n/d\xi^n) [\exp(-\xi^2)]$. In this case, ξ is the dimensionless form of conventional coordinate x , and $r = \xi^2/(2\acute{\alpha})$. If we set $\acute{\alpha} = 1/2$ as has been the default in this paper, then

$$r = \xi^2. \quad (87)$$

Therefore, the information metric r can be set to coincide with the square of the dimensionless coordinate ξ , which supports a close relationship between the r -spacetime and conventional coordinate spacetime (x -spacetime). Using this relationship, \mathcal{F}_{att} in energy state n is defined as

$$\mathcal{F}_{att} = A_n H_n(\sqrt{r}) \exp\left(-\frac{r}{2}\right) = \phi_n(x). \quad (88)$$

Meanwhile, \mathcal{F}_{rep} in energy state n is determined as another wave function $\tilde{\phi}_n(x)$ of a harmonic oscillator with angular frequency 2ω as

$$\mathcal{F}_{rep} = \sqrt[4]{2} A_n H_n(\sqrt{2r}) \exp(-r) = \tilde{\phi}_n(x). \quad (89)$$

Together, \mathcal{F} in any state n is represented as

$$\mathcal{F} = \mathcal{F}_{att} + \mathcal{F}_{rep} = \phi_n(x) - \tilde{\phi}_n(x). \quad (90)$$

This is an extended definition of the so-far-described influential force \mathcal{F} in the r -spacetime. Although the definition involves a set of wave functions in excited states, the above arguments for the ground state also hold. Thus, considering the coupling between r and ξ , it is plausible that \mathcal{F} represents an anharmonic oscillator as the superposition of harmonic oscillators in the x -spacetime.

By integrating Eq. (86) with respect to r , the probabilistic influential force potential $\varphi(r)$ is written in a decomposed form as

$$\varphi(r) = \varphi_{att}(r) + \varphi_{rep}(r) = \int [\phi_0(x) - \tilde{\phi}_0(x)] dr, \quad (91)$$

where $\varphi_{att}(r)$ and $\varphi_{rep}(r)$ are respectively the attractive and repulsive terms of $\varphi(r)$. In turn, $\varphi(r)$ is determined by $V(x) = m\omega^2 x^2/2$ and is applied to all energy states despite the extension of the formalism for \mathcal{F} . This further supports the correspondence between \mathcal{F} and the symmetric parabolic potential $V(x)$ in the x -spacetime. In summary, the above findings suggest that \mathcal{F} represents an anharmonic oscillator of information in the r -spacetime, which can be related to the anharmonic oscillator in the x -spacetime via the coordinate transformation between r and x .

Oscillators in the information coordinate spacetime

We return to the discussion on the influential force \mathcal{F}' in the information coordinate spacetime (I_O -spacetime). We here consider the effect of the conventional distance αx within the dimensionless potential $I_O = \alpha x + r/2$. Now, by replacing $r/2$ with I_O , the above equations in the information metric spacetime (r -spacetime) can be transformed into their counterparts in the I_O -spacetime:

$$\mathcal{F}'_{att} = A_n H_n(\sqrt{2I_O}) \exp(-I_O) \simeq \phi_n(x) \quad (92)$$

and

$$\mathcal{F}'_{rep} = \sqrt[4]{2} A_n H_n(2\sqrt{I_O}) \exp(-2I_O) \simeq \tilde{\phi}_n(x) . \quad (93)$$

Together, \mathcal{F}' in energy state n is represented as

$$\mathcal{F}' = \mathcal{F}'_{att} + \mathcal{F}'_{rep} \simeq \phi_n(x) - \tilde{\phi}_n(x) . \quad (94)$$

This is an extended definition of the influential force \mathcal{F}' in the I_O -spacetime. This formula demonstrates that \mathcal{F}' also represents an anharmonic oscillator, which is approximately equal to the superposition of harmonic oscillators in the conventional coordinate spacetime (x -spacetime). Moreover, the probabilistic influential force potential $\varphi'(I_O)$ in the I_O -spacetime is defined as

$$\varphi'(I_O) = \varphi'_{att}(I_O) + \varphi'_{rep}(I_O) = \int [\phi_0(x) - \tilde{\phi}_0(x)] dI_O , \quad (95)$$

where $\varphi'_{att}(I_O)$ and $\varphi'_{rep}(I_O)$ are respectively the attractive and repulsive terms of $\varphi'(I_O)$. Again, $\varphi'(I_O)$ is approximately equal to $\varphi(r)$, which corresponds to an anharmonic oscillator. These formulas demonstrate the close relationship between \mathcal{F}' in the I_O -spacetime and \mathcal{F} in the r -spacetime, which is in turn tightly coupled to the anharmonic oscillator in the x -spacetime.

Despite the preceding argument, however, the transformed formulas in the I_O -spacetime are no longer exactly equivalent to their parental formulas in the r -spacetime. This is because the transformation equation $I_O = \alpha x + r/2$ has the term αx , which represents the action of mediator particles. As far as the subatomic scale is considered, $\alpha x \ll r/2$. Nevertheless, αx is essential to preserve the potential symmetry of $\Phi'_{att}(I_O)$ and $\Phi'_{rep}(I_O)$, which underlies the constant forces $F'_{att}(I_O) = -1$ and $F'_{rep}(I_O) = 2$. The gauge invariance comes from this symmetry, and it is thus conceivable that oscillation occurs under $\Phi'(I_O)$ in the I_O -spacetime but not under $V(x)$ in the x -spacetime. Therefore, it is suggested that the oscillation in the x -spacetime is an extremely close approximation to that in the I_O -spacetime, which signifies the priority of the I_O -spacetime.

In contrast, when the supra atomic scale is considered, $\alpha x \gg r/2$ becomes true as x increases, which makes the contribution of $r/2$ negligible. Then, $I_O = \alpha x$ is substantially accurate, and a vast discrepancy between \mathcal{F}' and \mathcal{F} arises. In this case, however, the metric symmetry becomes apparent. Because $\varphi'(I_O) = \varphi'(\alpha x)$ is a function of the conventional coordinate x , $\varphi'(I_O)$ after all represents an anharmonic oscillator in the x -spacetime. Once again, this implies the priority of the I_O -spacetime, which is corroborated in the following subsections.

Metric symmetry and oscillators

To demonstrate the validity of our model, we provide evidence that supports the concept of metric symmetry. The *probabilistic influential force potential* $\varphi'(I_O)$ in the I_O -spacetime is a function of $I_O(x, r)$:

$$\varphi'(I_O) = \mathcal{M}(X)\mathcal{M}(Y)\left[-k_1e^{-I_O} + \frac{k_2}{2}\mathcal{M}(X)\mathcal{M}(Y)e^{-2I_O}\right], \quad (96)$$

which is suggested to represent an anharmonic oscillator. Here, the first term and second term on the right-hand side respectively correspond to $\varphi'_{att}(I_O)$ and $\varphi'_{rep}(I_O)$. The equation $I_O(x, r) = \alpha x + r/2$ is linear with respect to both r and x , and these coordinates thus contribute to the potential $\varphi'(I_O)$ in a similar fashion. A physical formula concerning the potential $\varphi'(I_O)$ would therefore have a symmetrical appearance with regard to r and x ; that is, these variables are apparently interchangeable. The two examples below show that this metric symmetry is expected to work between the equations that represent related but distinct phenomena. In contrast, for harmonic oscillators satisfying $r \propto x^2$, the identical phenomenon can be described by either r or x but with different formulas, where the metric symmetry does not hold.

We here present examples that illustrate this symmetry, thereby demonstrating the existence of a group of related oscillators in the r -spacetime and x -spacetime. In the first place, the probabilistic influential force potential $\varphi(r)$ in the r -spacetime is described as

$$\varphi(r) = \mathcal{M}(X)\mathcal{M}(Y)\left[-\frac{k_1}{\alpha}e^{-\alpha r} + \frac{k_2}{2\alpha}\mathcal{M}(X)\mathcal{M}(Y)e^{-2\alpha r}\right], \quad (97)$$

which can be written as

$$\varphi(r) = \frac{k_1^2}{2\alpha k_2} \left\{ \exp\left[-2\alpha\left(r - H(X) - H(Y) - \frac{1}{\alpha} \log \frac{k_2}{k_1}\right)\right] - 2\exp\left[-\alpha\left(r - H(X) - H(Y) - \frac{1}{\alpha} \log \frac{k_2}{k_1}\right)\right] \right\} \quad (98)$$

(Fig. 2). We here examine the metric symmetry described above to find the correspondence of $\varphi(r)$. We substitute x for r instead of using the aforementioned relation $r = \pi m \omega x^2 / (\alpha h)$. If we adopt the replacements of $\varphi(r)$ as $r = x$, $k_1^2 / (2\alpha k_2) = D$ and $H(X) + H(Y) + 1/\alpha \log(k_2/k_1) = x_e$, then the MO potential $V_M(x)$ [31] for a diatomic molecule is obtained as

$$V_M(x) = D \{ \exp[-2\alpha(x - x_e)] - 2\exp[-\alpha(x - x_e)] \}, \quad (99)$$

where D is the dissociation energy and x_e is the equilibrium distance (Appendix 4). This implies a connection between the probabilistic influential force potential $\varphi(r)$ in the r -spacetime and the MO potential $V_M(x)$ in the x -spacetime; that is, $k_B T \varphi(r) = V_M(x)$. The observed symmetry suggests that the MO potential and $\varphi(r)$ have the same background; that is, the probabilistic influential force potential $\varphi'(I_O)$ in the I_O -spacetime.

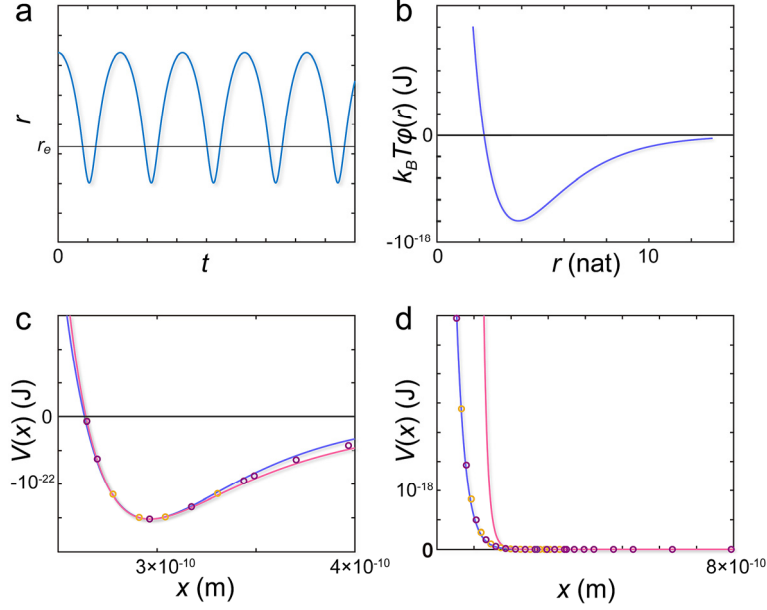


Figure 2 | Influential force potential and gas molecules. **a**, Information metric r of two informatons G_1 and G_2 under the influential force. r represents an anharmonic oscillation around the equilibrium metric r_e . **b**, Influential force potential $\varphi(r)$ conforming to the Morse potential $V(x)$ between the two nuclei in a hydrogen molecule. **c**, **d**, Potential energy between helium atoms for a short (**c**) and a wide (**d**) range of interatomic distances. The influential force potential (blue line) fits the reported potentials (purple [32] and yellow [33] circles) better than the Lennard–Jones potential (red).

In addition to the above relevance to $\varphi(r)$, the MO has universality in that its wave function is related to that of the two-dimensional isotropic harmonic oscillator (2DIHO) [34]. Here, x is used as the coordinate of the conventional distance spacetime, while r' is used as that of the radial component of the canonical coordinates of the 2DIHO. If we set $z = 2\pi(8mD)^{1/2}\exp[-\alpha(x - x_e)]/(\alpha h)$ and $\rho = (2\pi\omega/h)^{1/2}r'$, then $z = \rho^2$. Notably, the wave function of the MO in terms of z coincides with that of the radial part of the 2DIHO in terms of ρ^2 . Given that $x^2 \propto r$, we can remap z and ρ^2 onto the x - and r -coordinates, respectively, which allows us to consider that the MO and the radial part of the 2DIHO reside in the x - and r -spacetime, respectively. It is then regarded that the radial component of the wave equation of the 2DIHO has an anharmonic oscillation in the r -spacetime. This case thus illustrates the second example of the metric symmetry between the oscillators in the r - and x -spacetime. Furthermore, the potential for the radial component of the 2DIHO conforms to the formula for $\varphi(r)$.

In summary, the probabilistic influential force potential $\varphi(r)$ in the r -spacetime constitutes a group of anharmonic oscillators, together with the MO and the radial component of the 2DIHO. This coherence suggests that the group has a background identical to that of $\varphi'(I_O)$ in the I_O -spacetime, which supports metric symmetry and thus the framework of our theory.

Oscillation under the influential force

The probabilistic influential force potential $\varphi'(I_O)$ is an entirely abstract representation of the interaction between any objects having information. On the basis of the universality of information and probability, the action of the influential force \mathcal{F}' is inherently ubiquitous. Therefore, $\varphi'(I_O)$ allows the description of a wide variety of phenomena. Specifically, taking the limit of α to zero, the application scope of Eq. (96) may be broad and not confined to physics. Nevertheless, $\varphi'(I_O)$ has distinct properties of the anharmonic oscillator. Indeed, $\varphi'(I_O)$ represents the MO potential and 2DIHO potential, which are putative examples of metric symmetry in the x -spacetime and r -spacetime, respectively. We here focus on the correspondence between $\varphi'(I_O)$ and its direct transformation, the MO potential.

We assume that two informatons X and Y oscillate under the probabilistic influential force potential $\varphi'(I_O)$. For the oscillation of informatons, we use the reduced information distance newly set as $I_O := (\alpha x + r/2)/2$, which corresponds to the reduced mass-energy of the quantum. Then, $\varphi'(I_O)$ is a minimum when I_O is $I_e = H(X) + H(Y) + \log(k_2/k_1)$ (Fig. 2b). I_e is defined as the equilibrium information distance between the informatons, where the state probability takes a maximum value. Unless the repulsion coefficient k_2 is zero, $I_e > 0$, and the vibration is anharmonic. In this case, $\varphi'(I_O)$ increases rapidly when I_O becomes smaller than I_e , increases moderately when I_O becomes larger than I_e , and becomes zero asymptotically at large I_O . Following deformation into the MO potential using Eq. (94), these characteristics of $\varphi'(I_O)$ are completely reproduced as features of the MO potential. In addition to these rises and falls, $\varphi'_{rep} \propto \varphi'_{att}{}^2$. Therefore, φ'_{rep} increases rapidly as the distance decreases, which indicates a strong repulsive force in the vicinity of the origin. This feature again effectively reproduces the exponential repulsive wall in the MO potential.

We next examine the oscillation equation to decipher the physical reality of $\varphi'(I_O)$. When two informatons X and Y oscillate under the probabilistic influential force potential $\varphi'(I_O)$, the wave equation is written as

$$\left[-\frac{\hbar^2}{2\mu_m} \frac{d^2}{dI_O^2} + \varphi'(I_O) - I_n \right] \chi(I_O) = 0, \quad (100)$$

where μ_m is the reduced mass, $\chi(I_O)$ is the wave function, and I_n is the information level. The solution to this equation is

$$\chi(I_O) = \exp\left(-\frac{z}{2}\right) z^{\frac{b}{2}} L_n^{(b)}(z), \quad (101)$$

where $I_n = -D + \hbar\omega(n + 1/2) - [\hbar\omega(n + 1/2)^2]/4D$, $z = (8\mu_m D)^{1/2} \exp(I_e - I_O)/\hbar$, $b = (-8\mu_m I_n)^{1/2}/\hbar$, $D = k_1^2/(2k_2)$, and $I_e = H(X) + H(Y) + \log(k_2/k_1)$. $L_n^{(b)}(z)$ is the generalized Laguerre polynomial. $\chi(I_O)$ has n nodes, and it takes a maximum value at $I_O = I_e$. These features are typical of anharmonic quantum oscillators and are consistent with those of the wave function of the MO.

The classical equation of motion of the above oscillation is next written as

$$\frac{d^2 I_O}{dt^2} = \frac{2D}{\mu_m} \{-\exp[-(I_O - I_e)] + \exp[-2(I_O - I_e)]\}, \quad (102)$$

where I_O is the reduced information distance between X and Y in the I_O -spacetime while μ_m is the reduced mass. The solution to this equation is

$$I_O = \log \left\{ 1 + \left(1 - \frac{|E|}{D} \right)^{\frac{1}{2}} \cos \left[\left(\frac{2|E|}{\mu_m} \right)^{\frac{1}{2}} t \right] \right\} - \log \left[1 + \left(1 - \frac{|E|}{D} \right)^{\frac{1}{2}} \right] + I_O(0), \quad (103)$$

where E is the total energy expressed as $E = D(\exp\{-2[I_O(0) - I_e]\} - 2\exp\{-[I_O(0) - I_e]\})$, which takes a negative value. This equation presents the classical motion of the anharmonic oscillator. The oscillation is around $I_O = I_e$, which is similar to the case of the MO. When t is sufficiently small, I_O can be approximated by a quadratic function of t as

$$\begin{aligned} I_O &\simeq \left(1 - \frac{|E|}{D} \right)^{\frac{1}{2}} \cos \left[\left(\frac{2|E|}{\mu_m} \right)^{\frac{1}{2}} t \right] - \left(1 - \frac{|E|}{D} \right)^{\frac{1}{2}} + I_O(0) \\ &\simeq - \left(1 - \frac{|E|}{D} \right)^{\frac{1}{2}} \frac{|E|}{\mu_m} t^2 + I_O(0). \end{aligned} \quad (104)$$

This motion can be considered uniform acceleration within a sufficiently narrow time window. More specifically, if the repulsive constant k_2 in Eq. (102) equals zero, then the acceleration $d^2 I_O / dt^2$ is constant, resulting in uniform acceleration without a time limitation.

Despite these remarkable coincidences, there is a difference in the known generation mechanism between $\varphi'(I_O)$ and the MO potential. While $\varphi'(I_O)$ arises from the changes in entropy associated with the information transmission, the Morse potential arises from the electromagnetic force. Although these two mechanisms appear to be different, the influential force provides a basis for the unification of gauge fields, to which the electromagnetic field belongs. This suggests that the MO potential is created by informatic mechanisms behind the electromagnetic interaction. In summary, the above observations demonstrate a substantial unity between $\varphi'(I_O)$ and the MO potential. To our knowledge, this is the first example where information theory has been successfully applied in predicting a quantum interaction.

It will be meaningful to ascertain whether $\varphi'(I_O)$ is a cause of the MO potential. In this case, the information shared by the two atoms is that of the positions and momenta of the nuclei and electrons. In previous work [35], the ionicity of the valence bond was calculated using MI for the states of two hydrogen atoms and that of a hydrogen molecule. However, the methodology was different from our method in that it calculated MI of electron assignments using atomic orbitals.

Probabilistic influential force potential and dispersion force

In addition to the MO potential, the interaction between two helium molecules follows the probabilistic influential force potential $\varphi'(I_O)$ as described below. Helium atoms form a dimer based on the

dispersion force at low temperatures. Hellmann [32] and others [33] calculated the potential energy of helium atoms as a function of the interatomic distance using the *ab initio* molecular orbital method. According to them, the potential takes a minimum value of 11.01 K or 1.520×10^{-22} J at an interatomic distance of 5.60 Bohr ($= 2.964 \times 10^{-10}$ m). We then fit the probabilistic influential force potential $\varphi'(I_0)$ to their calculation results and estimate $\acute{\alpha} = 1.109 \text{ Bohr}^{-1} = 2.096 \times 10^{10} \text{ m}^{-1}$ (Figs. 2c and 2d). The potential curve of the influential force fits their results well. In particular, when the interatomic distance is short, the fit is much better than that of the Lennard–Jones potential (Appendix 5).

According to the above observations, the interaction between the two helium atoms forming a dimer can also be explained by $\varphi'(I_0)$, and the atoms are considered to have an anharmonic oscillation. Thus, the above two examples of the MO potential and helium dimer illustrate the possibility of applying the influential force to other physicochemical bonds. We expect that, in future work, various intramolecular as well as intermolecular interactions will be explained from an information theoretical point of view.

Prediction of motion in non-physical systems from metric symmetry

In addition to motion in the x -spacetime, motion is predicted with an equivalent formula in the r -spacetime from metric symmetry. In such cases, Eqs. (20) and (96) should work through the replacement of x by r . From a general point of view, this concept can be extended to deal with systems other than physical systems. This is done by taking the limit $ax \rightarrow 0$ and obtaining $I_0 = ax + r/2 = r/2$. If we assume a system of two informatons in which the probability of information transmission follows the infocanonical distribution [Eq. (17), Appendix 17] (i.e., $p = e^{-I_0} = e^{-r/2}$), then the influential force $\mathcal{F}_{att} = e^{MI}$ decreases the information metric r . Under these conditions, the above Eqs. (20) and (96) also hold in the r -spacetime when changing the variable and constant as $r = x$ and $k_1^2/(2\acute{\alpha}k_2) = D$. As a result, even if the system is not a physical one, the informatons undergo an anharmonic oscillation in the r -spacetime. In the second half of this paper, we discuss the interaction between informatons from this point of view unless otherwise stated.

(f) Time evolution of MI in classical systems

The preceding sections discussed information transmission related to energy transfer. We concluded that the MI describes all fundamental interactions. However, the basic mechanisms underlying the genesis of MI can be understood only if the statistical mechanics are sufficiently improved, such that we can analyze the extremely high-dimensional MI of informatons. This is because all the constituents of our spacetime (e.g., quantum fields and superstrings) are supposed to have information with a large number of dimensions.

To address the above issue, we begin an exploration to clarify the mechanisms that underlie the emergence of the very multidimensional MI. We adopt two strategies for this purpose. First, we tentatively ignore the mass of the mediator particle and consider the limit as $\alpha x \rightarrow 0$. This strategy is based on the assumption that the information metric r of the informatons is too large and the effect of the mediator particle is thus negligible. The influential force \mathcal{F}' in the I_O -spacetime then becomes equal to \mathcal{F} in the r -spacetime. Second, we consider microstates of informatons and seek a rationale that generates MI at the microscopic level. Given that the working principle of the influential force \mathcal{F} is based on statistical mechanics, \mathcal{F} can also be defined by employing a microscopic method. Adopting this strategy, we demonstrate that MI evolves between informatons through fluctuations in their microstates followed by changes in the joint probability distribution.

Microscopic method

Let X and Y be the information levels of informatons G_X and G_Y , which can be evenly split into m and n levels and discretely distinguished by indices k ($1 \leq k \leq m$) and l ($1 \leq l \leq n$), respectively. This section considers that m and n are sufficiently large. Let $p(X_k)$ and $p(Y_l)$ be the realization probabilities of information levels X_k and Y_l , respectively, and let $I_k = -\log p(X_k)$ and $I_l = -\log p(Y_l)$ be their self-information. Additionally, we let $p(X_k, Y_l)$ be the joint probability of X_k and Y_l and we refer to $I_{kl} = -\log p(X_k, Y_l)$ as the **joint self-information**. Furthermore, let the **micromutual information** MI_{kl} be MI for X and Y shared at the specific microstate (k, l) . We compute MI_{kl} as the MI of the $m \times n$ contingency table with respect to a cell (k, l) (Appendix 6). We finally introduce a fluctuation factor $\epsilon_{kl} = p(X_k, Y_l) / [p(X_k)p(Y_l)]$ as the ratio of the joint probability to the product of marginal probabilities, while its logarithm $\log(\epsilon_{kl})$ is referred to as the pointwise MI [36].

According to the above definitions, we can construct equations that express the microscopic interactions between informatons. We first describe the joint self-information I_{kl} and define the **microinformation metric** r_{kl} by

$$I_{kl} = I_k + I_l - \log(\epsilon_{kl}), \quad r_{kl} = I_k + I_l - 2 \log(\epsilon_{kl}), \quad (105)$$

respectively. I_{kl} and r_{kl} are respectively the microscopic counterparts of the joint entropy $H(X, Y)$ and the information metric r . Accordingly, if we multiply these two formulas by $p(X_k, Y_l)$ and take the sum concerning k and l , then we obtain the macroscopic equations

$$H(X, Y) = H(X) + H(Y) - MI, \quad r = H(X) + H(Y) - 2MI \quad (106)$$

(Appendix 24). Similarly, we can define the influential forces working in microstates. Using the infocanonical distribution (Appendix 17), we obtain the **attractive microforce** \mathcal{F}_{kl_att} and **repulsive microforce** \mathcal{F}_{kl_rep} for each microstate as

$$\mathcal{F}_{kl_att} \simeq \exp\left[\frac{\Delta MI_{kl}}{p(X_k, Y_l)}\right], \quad \mathcal{F}_{kl_rep} \simeq \exp\left[\frac{2\Delta MI_{kl}}{p(X_k, Y_l)}\right]. \quad (107)$$

Then, by multiplying these two expressions by $p(X_k, Y_l)$ and taking the sum with respect to k and l , we get the macroscopic attractive and repulsive forces as

$$\sum_k \sum_l p(X_k, Y_l) \mathcal{F}_{kl_att} \simeq \exp(\Delta MI) = \mathcal{F}_{att}, \quad \sum_k \sum_l p(X_k, Y_l) \mathcal{F}_{kl_rep} \simeq \exp(2\Delta MI) = \mathcal{F}_{rep}, \quad (108)$$

respectively.

In addition to the description with MI , we can calculate the microscopic influential force using the microinformation metric r_{kl} . Let $W(X_k)$ and $W(Y_l)$ be the number of microstates X_k and Y_l , respectively. Then, given that $\exp(I_k) = W(X_k)$ and $\exp(I_l) = W(Y_l)$, the microforces are represented with r_{kl} as

$$\mathcal{F}_{kl_att} \simeq \frac{\sqrt{W(X_k)} \sqrt{W(Y_l)}}{\exp\left(\frac{1}{2} r_{kl}\right)}, \quad \mathcal{F}_{kl_rep} \simeq \frac{W(X_k) W(Y_l)}{\exp(r_{kl})}, \quad (109)$$

respectively. Then, for each microstate, a **composite microforce** \mathcal{F}_{kl} is derived from the attractive and repulsive microforces as $\mathcal{F}_{kl} = \mathcal{F}_{kl_att} + \mathcal{F}_{kl_rep}$; that is,

$$\mathcal{F}_{kl} = [W(X_k) W(Y_l)]^{\acute{\alpha}} \{ k_1 \exp(-\acute{\alpha} r_{kl}) - k_2 [W(X_k) W(Y_l)]^{\acute{\alpha}} \exp(-2\acute{\alpha} r_{kl}) \}. \quad (110)$$

We now write the **equation of motion for each microstate** as

$$\frac{d^2 p_{kl}}{dt^2} = k_B T [W(X_k)^{\acute{\alpha}} + W(Y_l)^{\acute{\alpha}}] \{ k_1 \exp(-\acute{\alpha} r_{kl}) - k_2 [W(X_k) W(Y_l)]^{\acute{\alpha}} \exp(-2\acute{\alpha} r_{kl}) \}, \quad (111)$$

where $p_{kl} = \exp(-\log \varepsilon_{kl}) = 1/\varepsilon_{kl}$. If we multiply both sides of this equation by $p(X_k, Y_l)$ and sum over all k and l , then we obtain the macroscopic equation of motion given as Eq. (72). In summary, the influential force can be defined from both microscopic and macroscopic points of view.

MI generated through fluctuations in classical physicochemical systems

Using the microscopic method above, we propose a fluctuation-induced mechanism in the generation of MI. When two types of inert gas— X and Y —are combined under isothermal conditions, mutual diffusion occurs. An inelastic collision then arises, and the gas mixture begins to behave as if a single gas. At this time, the average kinetic energy is constant for each molecule X and Y . However, it is noted that inelastic collisions reduce the average relative velocity between the two molecules. In this case, the collision heat is transferred to the heat bath, maintaining the isothermal condition. Thus, the emergence of cooperative movement indicates that the collision increases $MI(X; Y)$, thereby reducing the energy associated with the joint entropy $H(X, Y)$ while preserving $H(X)$ and $H(Y)$.

The inelastic collision between gas molecules affects their mutual Brownian motion. Therefore, for the information levels of X and Y , the joint probability after the collision is generally different from the product of the marginal probabilities, which implies the generation of MI. The information shared between the two types of gas molecules is their positions and momenta, as well as their Hamiltonians.

We here assume that the potential energy is negligible, and the information levels of the kinetic energy then follow an exponential distribution, which we call an infocanonical distribution (Appendix 17).

We now return to a general discussion. With regard to the informatons X and Y , we assume that the realization probability of the k -th information level of X , X_k , follows the infocanonical distribution $p(X_k) = [\exp(1/\lambda) - 1]\exp(-k/\lambda)$, whereas that of the l -th level of Y , Y_l , follows the infocanonical distribution $p(Y_l) = [\exp(1/\nu) - 1]\exp(-l/\nu)$, where λ and ν are respectively the mean index values of X_k and Y_l . We suppose that $\varepsilon_{kl} = p(X_k, Y_l)/[p(X_k)p(Y_l)]$ fluctuates close to a value of 1. We also suppose that ε_{kl} is represented by $\varepsilon_{kl} = \gamma_{kl}\delta_{kl}$ using the two factors γ_{kl} and δ_{kl} , which independently fluctuate owing to the properties of X and Y , respectively. In addition, we let γ and δ be the averages of γ_{kl} and δ_{kl} over k and l , respectively, and we assume that both γ and δ follow a normal distribution with a mean of 1 and variance of σ^2 . As defined earlier, we let the micromutual information MI_{kl} be MI for X and Y shared at the specific microstate (k , l) (Appendix 6). Now that m and n are sufficiently large, MI of X and Y is obtained as

$$\begin{aligned} MI &\simeq \sum_{k=1}^m \sum_{l=1}^n MI_{kl} \\ &\simeq (1-\gamma\delta)^2 \left(1 - e^{-\frac{m}{\lambda}}\right) \left(1 - e^{-\frac{n}{\nu}}\right) \end{aligned} \quad (112)$$

(Appendix 7). This demonstrates that MI arises because of fluctuations of the realization probability of each microstate, and its magnitude is determined by the numbers of information levels of X and Y .

In the above example, X and Y are two gas molecules, and X_k and Y_l correspond to their information levels of kinetic energy. The mixing of two gases initiates collisions, where we let τ be the average time between collisions of X and Y . As the molecules repeatedly collide, their information levels are subject to Brownian fluctuations, which are depicted as two-dimensional spreading in the contingency table (Fig. 3). Most changes are within the neighborhood of the original state. These fluctuations result in the spreading of their information levels, which approximately approaches a two-dimensional normal distribution under $m, n \rightarrow \infty$. Conversely, $p(X_k, Y_l)$ at a particular state (k , l) becomes a complex superposition of nearly normal distributions originating from many other states. As a result, ε_{kl} fluctuates as the model above, leading to the change in MI_{kl} at each microstate. Finally, $MI(X; Y)$ increases as the summation of all MI_{kl} .

Along with the increase in $MI(X; Y)$, the average relative kinetic entropy between X and Y decreases, which implies the generation of the coordinated motions of X and Y . Thus, on the basis of the shared information, the two gases come to act together. This discussion leads to the general conclusion that a collision is a form of information exchange between physical informatons. Here, in the above description, probabilistic indices such as $p(X_k, Y_l)$ and ε_{kl} at time t are defined for the pair of X and Y that caused a collision within a period between $t - \tau$ and t .

$Y \backslash X$	1	2	...	k	...	$m-1$	m	
1								$\rho(Y_1)$
2								$\rho(Y_2)$
...								...
l								$\rho(Y_l)$
...								...
$n-1$								$\rho(Y_{n-1})$
n								$\rho(Y_n)$
	$\rho(X_1)$	$\rho(X_2)$...	$\rho(X_k)$...	$\rho(X_{m-1})$	$\rho(X_m)$	1

Figure 3 | Brownian fluctuations in an $m \times n$ contingency table for a pair of molecules X and Y . Initially, X takes the state X_k while Y takes the state Y_l (red rectangle), with a joint probability $p(X_k, Y_l)$ and marginal probabilities $p(X_k)$ and $p(Y_l)$, respectively. Repeated collisions occur with an average interval of τ , which enhances Brownian fluctuations in the states (blue arrows).

Finally, we examine the time course of MI. We assume that the distributions of γ and δ are represented by the diffusion equation with diffusion coefficient D and that both γ and δ follow a normal distribution having a mean of 1 and variance $\sigma^2 \simeq 2Dt$ at time t after initiation. Because m/λ and n/ν are constant under isothermal conditions, there is fluctuation in the joint probability $p(X_k, Y_l)$ without changes in the marginal probabilities $p(X_k)$ and $p(Y_l)$. Thus, $H(X)$ and $H(Y)$ are considered constant throughout the process. The expectation value of ΔMI is then

$$\langle \Delta MI \rangle \simeq (8D^2 t^2 + 8Dt) \left(1 - e^{-\frac{m}{\lambda}}\right) \left(1 - e^{-\frac{n}{\nu}}\right). \quad (113)$$

Of note is that $\langle \Delta MI \rangle$ increases with time irrespective of whether $\gamma\delta$ increases or decreases. The expectation value of the information metric $\langle r \rangle$ of X and Y then decreases spontaneously and monotonously as a quadratic function of t . Thus, when t is sufficiently small, there is motion with nearly uniform acceleration in the information metric spacetime (r -spacetime).

Stabilization of the combined classical system through MI

With respect to the above example of gas mixing, the fluctuations in the kinetic energies generate MI for the two populations of gas molecules. The generated MI is expected to provide a stabilization energy U_{stab} that contributes to the formation of the mixed state of the gases; that is,

$$U_{stab} = -k_B T MI. \quad (114)$$

To the best of our knowledge, this is novel energy related to the statistical mechanics of composite systems, which provides a clue for elucidating the interactions between informatons with many different states, such as superstrings. We propose that this is the basic manner by which the attractive

influential force \mathcal{F}_{att} in the r -spacetime is generated between bodies as

$$\mathcal{F}_{att} = e^{MI}, \quad (115)$$

where the repulsive force is for now not considered. In summary, the microscopic interaction that affects the Brownian fluctuation of informatons is the source of macroscopic MI. For the initial period, this causes movement under an almost constant attractive force \mathcal{F}_{att} in the r -spacetime.

(g) Time evolution of MI in quantum systems

This section explores whether MI is generated through fluctuations also in quantum systems. Here, by observing the state change after the generation of two particles constituting a composite system, we will demonstrate the occurrence of the MI of the particles. According to the discussion thus far, the generated MI can contribute to the shortening of the information metric r between the particles, implying the emergence of the attractive influential force \mathcal{F}_{att} .

Generation of MI in composite systems

Let us assume a composite quantum system XY that comprises two particles, X and Y , under isothermal conditions. Suppose that X and Y appear respectively at times $t_X \leq 0$ and $t_Y = 0$ and that their probability density functions are delta functions at their emergence. They are independent at $t = 0$, which can be regarded as corresponding to the ground state of the composite system. By the definition of MI, $MI(X;Y) = 0$ at $t = 0$.

When $t > 0$, we can calculate MI as below to represent the information exchange in the quantum system. Let \vec{r}_1 and \vec{r}_2 be the position vectors of X and Y , respectively. Let $p(\vec{r}_1, \vec{r}_2, t)$ be the joint probability density function of XY , which represents the coexistence probability of the particles. Let $p(\vec{r}_1, t)$ and $p(\vec{r}_2, t)$ be respectively the marginal probability density functions of X and Y , which represent the existence probability of each particle. We show that if there is mutual interference between X and Y , then $MI(X;Y)$ increases. This $MI(X;Y)$ is described using a time parameter t :

$$\begin{aligned} MI(X;Y) &= \iiint \left[\iiint p(\vec{r}_1, \vec{r}_2, t) \log \frac{p(\vec{r}_1, \vec{r}_2, t)}{p(\vec{r}_1, t)p(\vec{r}_2, t)} d\vec{r}_2 \right] d\vec{r}_1 \\ &= \iiint \left[\iiint p(\vec{r}_1, \vec{r}_2, t) \log p(\vec{r}_1, \vec{r}_2, t) d\vec{r}_2 \right] d\vec{r}_1 \\ &\quad - \iiint \left[\iiint p(\vec{r}_1, \vec{r}_2, t) \log p(\vec{r}_1, t) d\vec{r}_2 \right] d\vec{r}_1 \\ &\quad - \iiint \left[\iiint p(\vec{r}_1, \vec{r}_2, t) \log p(\vec{r}_2, t) d\vec{r}_1 \right] d\vec{r}_2 \\ &= -H(X, Y) - \iiint p(\vec{r}_1, t) \log p(\vec{r}_1, t) d\vec{r}_1 - \iiint p(\vec{r}_2, t) \log p(\vec{r}_2, t) d\vec{r}_2 \\ &= -H(X, Y) + H(X) + H(Y) \quad , \end{aligned} \quad (116)$$

which follows the formula for MI, given by Eq. (10). Taken together, we obtain the MI of the particles X and Y at any time $t \geq 0$.

Influential force in composite systems

We can now define the information metric r of X and Y using $MI(X;Y)$. The change in r satisfies $\Delta r = -2\Delta MI = 2\Delta S_a/\hbar$. We can then express the probability amplitude of information propagating from X to Y as

$$K(\vec{r}_2, \vec{r}_1, t) = \int \exp\left(\frac{iS_a}{\hbar}\right) Dx = \int \exp\left(\frac{ir}{2}\right) Dx. \quad (117)$$

r is therefore proportional not only to the action S_a but also to the path length between X and Y . Additionally, r represents the phase difference between X and Y . Here, we initially assume that $\alpha x = 0$, and the information distance then becomes $I_0 = r/2$. As described earlier, we obtain a constant force F_r in the information metric spacetime (r -spacetime); that is,

$$F_r = -\frac{dV}{dr} = -\frac{1}{2}k_B T, \quad (118)$$

where V is the potential difference between X and Y . This equation is invariant irrespective of V and r , demonstrating the gauge symmetry in the r -spacetime. It shows that F_r is a constant attractive force acting toward the origin. If we define \mathcal{F}_{att} as described in this paper, the above formula indicates that \mathcal{F}_{att} is a probabilistic representation of a gauge force F_r in the r -spacetime. Then,

$$\mathcal{F}_{att} = \exp(\Delta MI) = \exp\left(-\frac{\Delta r}{2}\right) = \exp\left(-\frac{\Delta S_a}{\hbar}\right). \quad (119)$$

This attractive influential force \mathcal{F}_{att} decreases the information metric r of X and Y and expresses the action that diminishes their phase difference.

Diffusion of free quantum particles

We next discuss the MI generated through the exchange interaction between two free quantum particles. The exchange interaction is a quantum mechanical effect that occurs between identical particles. It has no classical analog and is not considered a true force. However, by introducing probability density functions, we can identify the MI of identical free particles, which suggests the generation of the attractive influential force \mathcal{F}_{att} . For the quantitative analysis of MI over time, we will deal with the propagator of the free particles under diffusion processes.

We first mention the similarity of the diffusion equation and Schrödinger equation. The diffusion equation is written as

$$\frac{\partial u(x,t)}{\partial t} = D \frac{\partial^2 u(x,t)}{\partial x^2}, \quad (120)$$

where $u(x, t)$ is the probability density function of the particle at position x and time t . Meanwhile, the Schrödinger equation of a free particle is written as

$$\frac{\partial \psi(x,t)}{\partial t} = \frac{i\hbar}{2m} \frac{\partial^2 \psi(x,t)}{\partial x^2}, \quad (121)$$

where $\psi(x, t)$ is the wave function and m is the mass of the particle.

If we set the diffusion coefficient to $i\hbar/(2m)$, then the diffusion equation coincides with the Schrödinger equation. This coincidence suggests a general analogy between Brownian motion and quantum mechanics [27]. The analogy becomes clearer when we consider the propagator of the free particle; that is,

$$K(x, x'; t) = \left(\frac{m}{2\pi i\hbar t} \right)^{\frac{1}{2}} \exp\left[-\frac{m(x-x')^2}{2i\hbar t} \right]. \quad (122)$$

The propagator of the free particle also satisfies the Schrödinger equation and again follows a diffusion equation with the diffusion coefficient $i\hbar/(2m)$. $p(x, x', t) = |K(x, x'; t)|^2$ is then the probability density function for the particle initially located at x' at $t = 0$ being found at x at time t .

Analytical calculation of the MI of identical free particles

We next examine the MI of the exchange interaction between two free particles. We assume that particles X and Y are simultaneously created in three-dimensional space at $t = 0$ and start diffusing. When they are created, their probability density functions are expressed by delta functions. The position vectors of the two particles $\vec{r}_1 = (x_1, y_1, z_1)$ and $\vec{r}_2 = (x_2, y_2, z_2)$ take the values $(d, 0, 0)$ and $(-d, 0, 0)$ at $t = 0$, respectively. Nearer bounds of the diffusion region approach each other and cross at the origin. We assume that X and Y have the same mass m and that this mass is similar to the mass of an electron m_e . The propagators are then

$$K_1(\vec{r}_1, d; t) = \left(\frac{m}{2\pi i\hbar t} \right)^{\frac{3}{2}} \exp\left\{ -\frac{m}{2i\hbar t} [(x_1-d)^2 + y_1^2 + z_1^2] \right\} \quad (123)$$

and

$$K_2(\vec{r}_2, -d; t) = \left(\frac{m}{2\pi i\hbar t} \right)^{\frac{3}{2}} \exp\left\{ -\frac{m}{2i\hbar t} [(x_2+d)^2 + y_2^2 + z_2^2] \right\}. \quad (124)$$

Using the propagators $K_1(\vec{r}_1, d; t)$ and $K_2(\vec{r}_2, -d; t)$, the existence probability densities are obtained as described below.

To analyze the MI of free particles, we first calculate the joint probability density that represents the coexistence of the particles. We consider the case in which the two particles are identical bosons of spin zero. Meanwhile, when the two particles are identical fermions, similar arguments hold except for the sign (Appendix 8). We here let $\psi(\vec{r}_1, \vec{r}_2, t)$ be the composite wave function of X and Y , and suppose that it is expressed as

$$\psi(\vec{r}_1, \vec{r}_2, t) = A[K_1(\vec{r}_1, d; t)K_2(\vec{r}_2, -d; t) + K_1(\vec{r}_2, d; t)K_2(\vec{r}_1, -d; t)], \quad (125)$$

where A is a normalization constant. This A is calculated by setting the integral of $|\psi(x, y, t)|^2$ to 1 on the integration region, which is defined by $|x_1 - d| < (ht/m)^{1/2}/2$, $|y_1| < (ht/m)^{1/2}/2$, $|z_1| < (ht/m)^{1/2}/2$, $|x_2 + d| < (ht/m)^{1/2}/2$, $|y_2| < (ht/m)^{1/2}/2$, and $|z_2| < (ht/m)^{1/2}/2$.

The normalization constant A is then expressed as

$$A = \frac{1}{\sqrt{2}} \left(\frac{ht}{m} \right)^2 \left[\frac{ht}{m} + \frac{1}{4} \left(\frac{ht}{\pi md} \right)^2 \sin^2 \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left(\frac{8\pi md^2}{ht} \right) \right]^{-\frac{1}{2}}. \quad (126)$$

Using this A , we now obtain the joint probability density function $p(\vec{r}_1, \vec{r}_2, t)$ as

$$p(\vec{r}_1, \vec{r}_2, t) = |\psi(\vec{r}_1, \vec{r}_2, t)|^2 = 2A^2 \left(\frac{m}{ht} \right)^6 \left\{ 1 + \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\}, \quad (127)$$

which represents the joint existence probability of X and Y (Fig. 4d). Although $p(\vec{r}_1, \vec{r}_2, t)$ changes periodically with respect to $x_1 - x_2$, it always takes a maximum value at $x_1 = x_2$. This tendency of colocalization conforms to the known features of the exchange interaction between identical bosons.

In addition to the joint probability density function, the marginal probability density functions are expressed as

$$\begin{aligned} p(\vec{r}_1, t) &= \iiint p(\vec{r}_1, \vec{r}_2, t) d\vec{r}_2 \\ &= 2A^2 \left(\frac{m}{ht} \right)^5 \left\{ \sqrt{\frac{ht}{m}} + \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_1 + d) \right] \right\} \end{aligned} \quad (128)$$

and

$$\begin{aligned} p(\vec{r}_2, t) &= \iiint p(\vec{r}_1, \vec{r}_2, t) d\vec{r}_1 \\ &= 2A^2 \left(\frac{m}{ht} \right)^5 \left\{ \sqrt{\frac{ht}{m}} + \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_2 - d) \right] \right\}, \end{aligned} \quad (129)$$

which respectively express the existence probability of X and Y within each diffusion region. On the right side of each equation, the second term in the braces is the interference term. The interference term causes an oscillation of each probability density, in a time- and position-dependent manner. Figures 4a and 4b show $p(\vec{r}_1, t)$ and $p(\vec{r}_2, t)$, respectively, while Figure 4c depicts their product $p(\vec{r}_1, t)p(\vec{r}_2, t)$.

We next introduce a fluctuation factor $\varepsilon_{\vec{r}_1, \vec{r}_2} = p(\vec{r}_1, \vec{r}_2, t) / [p(\vec{r}_1, t)p(\vec{r}_2, t)]$, which is the ratio of the joint probability density to the product of the marginal probability densities. This $\varepsilon_{\vec{r}_1, \vec{r}_2}$ is the equivalent of ε_{kl} in the preceding section and is expressed as

$$\begin{aligned} \varepsilon_{\vec{r}_1, \vec{r}_2} &= \frac{p(\vec{r}_1, \vec{r}_2, t)}{p(\vec{r}_1, t)p(\vec{r}_2, t)} \\ &= \frac{1}{2A^2} \left(\frac{ht}{m} \right)^4 \left\{ 1 + \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\} \\ &\quad \left\{ \sqrt{\frac{ht}{m}} + \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_1 + d) \right] \right\}^{-1} \\ &\quad \left\{ \sqrt{\frac{ht}{m}} + \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_2 - d) \right] \right\}^{-1}. \end{aligned} \quad (130)$$

This formula indicates that $\varepsilon_{\vec{r}_1, \vec{r}_2}$ fluctuates with the phase difference between $p(\vec{r}_1, \vec{r}_2, t)$ and the product of $p(\vec{r}_1, t)$ and $p(\vec{r}_2, t)$ (Fig. 4e). This situation is similar to that in the preceding section, where the fluctuations in microstates generate MI. Indeed, MI again originates as described below.

In the current model, MI_d , the density of MI at time t , is expressed as

$$\begin{aligned}
MI_d &= p(\vec{r}_1, \vec{r}_2, t) \log \varepsilon_{\vec{r}_1, \vec{r}_2} \\
&= 2A^2 \left(\frac{m}{ht} \right)^6 \left\{ 1 + \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\} \\
&\quad \log \left[\frac{1}{2A^2} \left(\frac{ht}{m} \right)^4 \left\{ 1 + \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\} \right. \\
&\quad \left. \left\{ \sqrt{\frac{ht}{m}} + \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_1 + d) \right] \right\}^{-1} \right. \\
&\quad \left. \left\{ \sqrt{\frac{ht}{m}} + \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_2 - d) \right] \right\}^{-1} \right]. \tag{131}
\end{aligned}$$

This MI_d also fluctuates in a time- and position-dependent manner (Fig. 4f), which leads to a positive value of MI .

Finally, we obtain MI by integrating MI_d with respect to \vec{r}_1 and \vec{r}_2 over the integration region; that is,

$$\begin{aligned}
MI &= \iiint (\iiint MI_d d\vec{r}_2) d\vec{r}_1 \\
&= \left(\frac{ht}{m} \right)^2 \int_{d-\frac{1}{2}\sqrt{\frac{ht}{m}}}^{d+\frac{1}{2}\sqrt{\frac{ht}{m}}} \left(\int_{-d-\frac{1}{2}\sqrt{\frac{ht}{m}}}^{-d+\frac{1}{2}\sqrt{\frac{ht}{m}}} MI_d dx_2 \right) dx_1, \tag{132}
\end{aligned}$$

where the integration ranges increase with time through the diffusion of the bosons. Compared with the model in the prior section, the integration of MI_d is the nearly equivalent operation of taking the sum of micromutual information MI_{kl} .

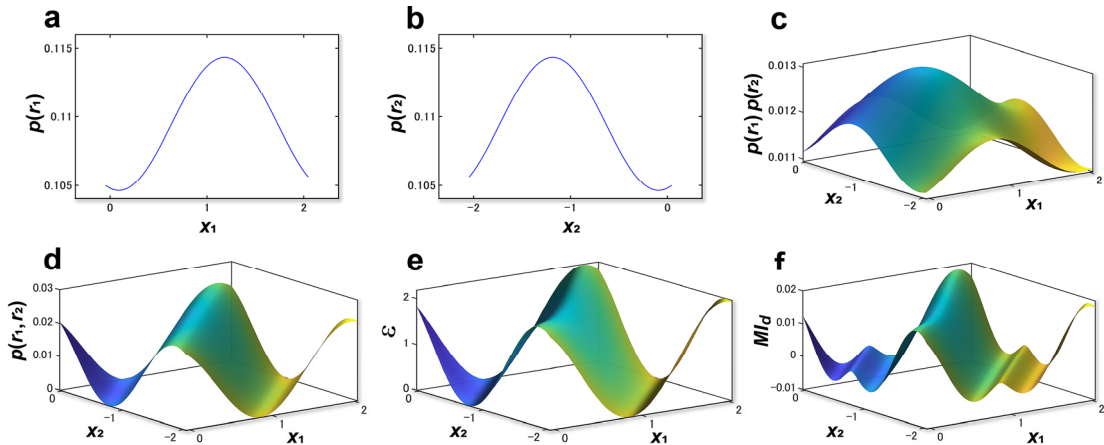


Figure 4 | Generation of MI of identical bosons X and Y in three-dimensional space. **a**, Marginal probability density function of X . **b**, Marginal probability density function of Y . **c**, Product of the marginal probability density functions of X and Y . **d**, Joint probability density function of X and Y . **e**, Fluctuation factor. **f**, Density of MI . Parameters are $d = 1$ m, $t = 6,000$ s, and $m = m_e$.

MI of identical particles

The above observations show that MI arises from the exchange interaction of identical free bosons at any time $t > 0$. The interference term in the composite wave function $\psi(\vec{r}_1, \vec{r}_2, t)$ causes the phase difference between the joint probability density function $p(\vec{r}_1, \vec{r}_2, t)$ and both the marginal probability density functions $p(\vec{r}_1, t)$ and $p(\vec{r}_2, t)$, which generates the fluctuation of $\varepsilon_{\vec{r}_1, \vec{r}_2}$, and MI thus arises.

At the limit $t \rightarrow 0$, the interference terms in the expressions for $p(\vec{r}_1, t)$ and $p(\vec{r}_2, t)$ in Eqs. (128) and (129) become zero. Then, $MI(X;Y) = 0$, because $\varepsilon_{\vec{r}_1, \vec{r}_2} = 1$. For this reason, at the very time that X and Y appear in the space, they are mutually independent. This satisfies the ground state condition discussed in the preceding Section (a) and allows MI to represent the information exchange in a quantum system.

When $t > 0$, because the interference terms are not zero, $MI(X;Y) > 0$. As a result,

$$H(X) + H(Y) > H(X, Y) \quad (133)$$

holds at any time $t > 0$. For the closed composite system under isothermal conditions, the change in free energy is $\Delta F = 0$ [23]. Therefore, the internal energy U decreases owing to the mutual interference between X and Y as

$$\Delta U = -k_B T \Delta MI(X;Y) < 0. \quad (134)$$

Hence, the equilibrium tends to proceed to the right-hand side of Eq. (133), forming the interaction between X and Y . At this time, the information metric r of the particles is shortened, indicating the emergence of the attractive influential force \mathcal{F}_{att} .

In addition to the shortening of r , the phase difference $\theta = r/2$ between X and Y decreases. Moreover, from Eq. (131), when MI_d takes a maximum value, x_1 and x_2 approach one another, and their respective velocities $v_1 \simeq (x_1 + d)/t$ and $v_2 \simeq (x_2 - d)/t$ converge. Additionally, the individual kinetic energies E_{K1} and E_{K2} approach one another. In summary, the exchange interaction reduces the conventional distance x , information metric r , phase difference θ , and the difference in kinetic energies E_K between the identical free bosons.

Unique features of the MI of identical particles

The preceding observations revealed that $MI(X;Y)$ should be relevant to statistical energy. As written above, the internal energy of the closed composite system is reduced as $\Delta U = -k_B T \Delta MI < 0$. Notably, this expression for ΔU is independent of the Hamiltonian \mathcal{H} of the system. Conversely, the previously known exchange interactions depend on the exchange integral $\int \psi^* \mathcal{H} \psi d\tau$, where \mathcal{H} corresponds to the Coulomb potential when electrons are exchanged, τ denotes the coordinates, and $*$ indicates the complex conjugate. Thus, $\Delta U = -k_B T \Delta MI$ expresses a new interaction in that it purely reflects the information exchange between quanta.

On the basis of the differences presented above, the statistical energy of MI has five distinct features. First, the generated $\Delta U = -k_B T \Delta MI$ is overwhelmingly smaller than the exchange integral. For example, the exchange integral between atoms is on the order of 1 eV, which is much larger than the generated $\Delta U = -k_B T \Delta MI$ calculated as 2.59×10^{-5} eV. Therefore, it is suggested that the effect of MI becomes apparent only as a massive statistical effect of exchange interaction among a vast number of identical particles. Second, the time dependence of MI is weak because MI changes at approximately the zeroth power of t . MI thus remains almost constant for a long time. Third, owing to this time stability, the maximum value of MI is virtually independent of the initial distance between X and Y . Even if the two identical particles are far apart in the universe, the maximum value of MI is almost the same. This long-range effect is in stark contrast to the exchange integral that is effective only for very close bodies. Fourth, the interference term in $p(\vec{r}_1, \vec{r}_2, t)$ appears immediately after the emergence of the particles, even if the diffusion of the particles with mass m takes a long time depending on the distance. Therefore, it is likely that this new exchange interaction propagates at nearly the speed of light. This immediate action at a distance is highly characteristic, suggesting that the found exchange interaction is a non-local quantum mechanical effect. Fifth, MI is independent of the masses of the identical particles, indicating its universality within the quantum regime. Thus, these observations imply that $\Delta U = -k_B T \Delta MI$ is appreciably different from the previously known exchange interactions.

Generality of the MI of identical particles

It is expected that there is always MI for identical particles, even if the particles are not generated simultaneously. The only requirement is that their probability density functions are delta functions at the creation. Then, $MI = 0$ at the appearance of the composite system, and $MI > 0$ thereafter. In such cases, our arguments hold for any pair of identical particles. Moreover, this kind of information exchange occurs irrespective of spins.

Our study considered the orbital wave functions of bosons of spin zero, and we concluded the emergence of the attractive force \mathcal{F}_{att} . This conclusion is unaffected even if we consider the spin wave functions. Therefore, regardless of whether the particles are bosons or fermions, it is conceivable that interaction effects among a huge number of distant identical particles will be newly discovered. In the case of bosons, because they tend to have the same position in the phase space, \mathcal{F}_{att} is expected to induce Bose–Einstein condensation. Moreover, considering the occurrence of phase synchronization between bosons of spin zero, it is plausible that this mechanism explains Higgs condensation.

In addition to bosons, the above discussion can be applied to the case of identical free fermions except for the sign (Appendix 8). Thus, the exchange interaction causes MI not only between identical free bosons but also between identical free fermions. The fact that $MI(X; Y) > 0$ even for identical fermions implies that, although the fermions avoid the same position, \mathcal{F}_{att} arises statistically via a mechanism analogous to that for identical bosons.

Stabilization of the composite quantum system through MI

In summarizing the above, a considerable fraction of identical particles exchanges information as MI. Moreover, from the inherent analogy between the Schrödinger and diffusion equations, it is expected that the emergence of MI is a universal phenomenon occurring for identical quantum particles of any combination. In turn, the originating MI is anticipated to provide stabilization energy U_{stab} that statistically contributes to the formation of the composite system; that is,

$$U_{stab} = -k_B T MI. \quad (114)$$

We propose that this creation of MI is the basic manner by which the influential force \mathcal{F}_{att} acting between identical quanta universally arises in the information metric spacetime (r -spacetime) as

$$\mathcal{F}_{att} = e^{MI}, \quad (115)$$

where the repulsive force is for now not considered. \mathcal{F}_{att} decreases the phase and velocity differences between the identical particles, allowing the coordinated movement of the particles. In conclusion, the attractive force \mathcal{F}_{att} in the r -spacetime is acting between virtually all identical quantum bodies based on the newly identified exchange interaction, thereby contributing to the stabilization of the universe.

Exchange interactions in the information coordinate spacetime

We now extend our discussion to the information coordinate spacetime (I_O -spacetime). The new exchange interaction that we found is non-local in the conventional coordinate spacetime (x -spacetime), in contrast to the previously known exchange interactions. Indeed, the above calculated MI does not depend on the conventional distance x . This distance independence has been observed because we have assumed $\alpha x = 0$ in this section. Thus, until now, the information distance has been $I_O = r/2$. However, the universal gauge symmetry in the I_O -spacetime requires the term αx (>0) to formulate $I_O = \alpha x + r/2$. Therefore, we again introduce αx to preserve the gauge symmetry, which characterizes the canonicity of the I_O -spacetime.

Traditionally, the exchange interaction has been regarded as a quantum mechanical "effect" and has not been considered a true force. However, the introduction of αx allows us to deal with a non-local interaction effect within the x -spacetime as a gauge force within the I_O -spacetime. In using αx , we hypothesize that a particular mediator particle carries MI and that αx expresses the information exchange between this imaginary particle and the I_O -spacetime. As written above, the MI of identical particles can be derived from the interference between the wave functions as $MI = \iint MI_d d\vec{r}_1 d\vec{r}_2$. This MI is then transformed into the information distance according to $I_O = \alpha x + r/2 = \alpha x - MI$. Eventually, both MI and αx contribute to the generation of the attractive force as $\mathcal{F}'_{att} = \exp(-\alpha x + MI) = \exp[-(\alpha x + r/2)]$. Finally, we suppose the occurrence of repulsive force \mathcal{F}'_{rep} . The identical particles then undergo an anharmonic oscillation [Eq. (101)]. Thus, applying the influential force leads to a revision of the paradigms for exchange interaction and non-locality.

Hypothetical consideration: a possible origin of dark matter

The information coordinate system uses $I_O = \alpha x + r/2$, which supposes metric symmetry between the conventional distance x and the information metric r . If we consider this symmetry, the effects of MI are predicted to be observable also within the x -spacetime. Specifically, relevant phenomena would become apparent when the energy of MI is converted to mass (i.e., becomes non-relativistic). This will be especially the case for the MI of identical particles with mass.

We conducted a numerical calculation to obtain MI . When the diffusion region of each free particle crosses the origin at $t = 4d^2m/h$, MI is 0.30685 nat, which is equivalent to $U_{stab} = 4.238 \times 10^{-24}$ J at 1 K. We hypothesize that this energy corresponds to the putative mediator particle mentioned above, which would mediate the MI of identical particles. We therefore name this hypothetical boson the *mion*. If the mion has mass, it would be as small as 2.645×10^{-5} eV/c² ($= 4.715 \times 10^{-41}$ kg) and the mion would thus be much less massive than the neutrino. However, the mion might constitute a large fraction of energy in the spacetime because it would exist in vast quantities in the universe. This energy may account for dark matter, which is thought to produce gravity and occupy approximately 85% of the matter and 27% of the mass–energy of the universe. This hypothesis requires further study.

(h) Influential force and quantum entanglement

We here investigate MI arising from the entanglement of two quanta. Although we have considered the generation of MI by the fluctuations of microstates in previous sections, we will demonstrate that MI and the influential force arise also from the entanglement of quantum superposition.

We suppose that the entanglement arises between two initially independent quanta X and Y of the mixed states, and the composite system XY in the pure state arises. Let X_0 and Y_0 be X and Y before the entanglement, respectively, and X_1 and Y_1 be X and Y after the entanglement, respectively. We let $S(X) = -tr(\rho_{X_R} \log \rho_{X_R})$ be the von Neumann entropy for the reduced density matrix ρ_{X_R} of X using the partial trace. Meanwhile, we let $H(X) = -tr(\rho_X \log \rho_X)$ be that for the density matrix ρ_X without using the partial trace. Then, whereas $S(X_0) = S(Y_0) = 0$ in the initial state, $H(X_0) > 0$ and $H(Y_0) > 0$. In the composite system after the entanglement, $S(X_1, Y_1) = H(X_1, Y_1) = 0$. The entanglement entropy S_{EE} satisfies $S_{EE} = S(X_1) = S(Y_1)$.

The change in the free energy ΔF before and after the entanglement is expressed as

$$\Delta F = k_B T \cdot [S(X_0) + S(Y_0) - S(X_1, Y_1)] = k_B T \cdot MI_{qu} = 0. \quad (135)$$

Meanwhile, the change in the thermodynamic entropy ΔS satisfies

$$\Delta S = -k_B \cdot [H(X_0) + H(Y_0) - H(X_1, Y_1)] = -k_B \cdot MI. \quad (136)$$

The MI shared by the two entangled quanta is

$$MI = S(X_1) + S(Y_1) = 2S_{EE}, \quad (137)$$

and the change in the energy ΔU before and after the entanglement is therefore expressed as

$$\Delta U = \Delta F + T\Delta S = k_B T \cdot (MI_{qu} - MI) = -2k_B T \cdot S_{EE} . \quad (138)$$

This demonstrates that the energy

$$\Delta U = -k_B T \cdot S_{EE} \quad (139)$$

is generated per one quantum based on the entanglement entropy. The total energy is preserved because for the total energy U ,

$$\beta U = H(X_0) + H(Y_0) = S(X_1) + S(Y_1) = 2S_{EE} \quad (140)$$

holds before and after the entanglement.

In the above scenario, the interaction energy ΔU of the entanglement becomes large, seeing that the energy of the composite system is $k_B T \cdot S(X_1, Y_1) = k_B T \cdot H(X_1, Y_1) = 0$. This significant energy difference suggests that the entanglement is likely to occur. Indeed, the entanglement exists universally and is thus supposed to occur spontaneously. At this time, because

$$\Delta I_O = -\Delta MI = -2S_{EE} , \quad (141)$$

an attractive influential force arises in the information coordinate spacetime (I_O -spacetime) as

$$\mathcal{F}'_{att} = e^{-I_O} = e^{2S_{EE}} . \quad (142)$$

Applying the metric symmetry, it is implied that the attractive force also emerges in the conventional coordinate spacetime (x -spacetime), which corresponds to gravity in the anti-de Sitter/conformal field theory correspondence. Meanwhile, the repulsive influential force is generated in the direction preserving entropy as

$$\mathcal{F}'_{rep} = e^{-2I_O} = e^{4S_{EE}} . \quad (143)$$

Under the equivalence of energy and entropy represented as $\Delta U = -2k_B T \cdot S_{EE}$, this \mathcal{F}'_{rep} is equivalent to the level repulsion counteracting the attractive potential. Finally, on the basis of the above discussion, an anharmonic oscillation will occur within the I_O -spacetime. Taken together, we have incorporated the quantum gravity of superstring theory into the framework of the influential force. In the next section, we investigate the relationship between the Higgs field and gravity using these findings.

(i) Influential force field and the Higgs field

The influential force uniformly represents the natural forces, which include gravity caused by mass. Its formula reveals that the quantum field generally accompanies the repulsive field. As a candidate of the quantum field with this property, we consider the Higgs field involved in the generation of mass. We start our research by investigating the relationship between the influential force, Higgs field, and gravity. According to the Standard Model, the Higgs field has a non-zero vacuum expectation value arising from spontaneous symmetry breaking and produces mass in proportion to the expectation value. However, the mechanism of symmetry breaking itself remains unknown. Moreover, the relationship between the Higgs mechanism and gravity is poorly understood.

Here, we apply the equation of the influential force to the Higgs field and demonstrate that the repulsive term breaks the gauge symmetry of the influential force field, producing the vacuum expectation value and the Higgs boson. The influential force field can express the gravity and Higgs fields and explains the equivalence of gravitational and inertial mass. The relationship between the influential force and gravity is supported by an inflation model using the influential-force-explained Higgs field (i.e., the influential Higgs), which provides a good fit to the cosmic microwave background data. In addition, the entanglement entropy is linked with the influential Higgs and the Higgs mass. Thus, the influential force unifies the Higgs mechanism, gravity, inflation, and entanglement entropy.

Formula for the influential force field and Higgs field

In the Standard Model, the Higgs field is expressed by a complex scalar field ϕ and codes for the Higgs boson with spin zero. The standard Higgs potential V_H is expressed as

$$V_H = -\mu^2 |\phi|^2 + \lambda |\phi|^4, \quad (144)$$

where μ and λ are arbitrary constants. The equation was established by considering renormalizability based on ϕ^4 theory. In the Standard Model, μ^2 becomes positive at the phase transition of a vacuum, and the Higgs field gives a non-zero vacuum expectation value. Thereby, the electroweak symmetry is broken spontaneously, and the weak bosons and fermions obtain mass. However, the mechanism of the phase transition itself remains unknown. Moreover, the relationship between the Higgs field and the gravity field remains obscure, clearly demonstrating the need for physics beyond the Standard Model. Furthermore, the relationship between gravity, inflation theory, and entanglement entropy is currently under intense investigation, highlighting the importance of clarifying the association between the Higgs field and these gravity-related issues.

First, let us compare the formulas of the standard Higgs potential V_H and the influential force potential $\varphi'(I_O)$. For example, if we replace $|\phi|^2$ in the expression for V_H by e^{-I_O} , then the equation is isomorphic to that of the probabilistic influential force potential,

$$\varphi'(I_O) = -k_1 e^{-I_O} + \frac{k_2}{2} e^{-2I_O}, \quad (145)$$

which is the inverse superposition of the two potentials generated by two vacuum field quanta having a unit information mass, one of which has the ground state energy whereas the other has twice that energy. In comparison, $\varphi'(I_O)$ and V_H are similar in that the second term is opposite the first term and its magnitude is proportional to the square of the first term. Therefore, both formulas consist of an attractive term and a repulsive term, although their order is reversed. Despite the similarities, because $\varphi'(I_O)$ has finite magnitude, unlike the case for V_H , divergence does not occur, and renormalization is unnecessary. This lack of divergence may provide a promising solution to the hierarchy problem. There is thus merit to expressing the Higgs field using the influential force field.

Inspired by the similarities mentioned above, we examine whether the influential force field can represent the Higgs field. Suppose that the Higgs field is a complex scalar field with a modulus I_O and a potential $\varphi'(I_O)$. From the preceding discussion, $\varphi'(I_O)$ resides in the I_O -spacetime, encoding the anharmonic oscillator with the attractive term $\varphi'_{att}(I_O)$ and repulsive term $\varphi'_{rep}(I_O)$. I_O is a coordinate expressing both the position and energy, and $\varphi'(I_O)$ thus represents the energy vibration of the I_O -spacetime itself. Under the above supposition, this oscillator will represent a quantum particle associated with the Higgs field. We will explore the validity of these considerations.

Application of Nambu theory

The universality of the Higgs field suggests that its origin is the quantum of the vacuum field. We first consider the situation where the two quanta of the zero-point oscillation interact under the influential force field. We assume the superconductive state of the vacuum according to Nambu–Higgs theory and investigate a model considering the influential force. Nambu–Higgs theory explains electroweak symmetry breaking by analogy to Bardeen–Cooper–Schrieffer (BCS) theory on superconductivity. Nambu proposed a composite Higgs model in which the Higgs boson is composed of a pair of top and anti-top quarks. However, there is no evidence on the identity of the phonon and the Cooper pair (i.e., the Higgs boson). We revisit this Nambu theory, with a modification that the pair is made up of spacetime quanta in the vacuum.

On the basis of the canonicity of I_O -spacetime, we here consider the interaction between two quantized I_O -spacetimes as a candidate for the origin of the Cooper pair and suppose that the vibration of the background x -spacetime corresponds to the phonon. In this case, if the quantized I_O -spacetime is a fermion, then there will be an interaction with another spacetime fermion with opposite spin via the oscillation of the background spacetime. The complex thus generated can be a boson corresponding to the Cooper pair, namely the Higgs boson, which will undergo Bose–Einstein condensation.

We initially perform a thought experiment where the classical BCS theory is applied to the conventional coordinate spacetime (x -spacetime), and the interaction arises between the spacetime quanta. The Bose–Einstein condensation is supposed to lead to superconductivity in the vacuum and is represented by the BCS wave function

$$|\phi_{\text{BCS}}\rangle = \prod_{k'} (u_{k'} + v_{k'} c_{-k'\downarrow}^\dagger c_{k'\uparrow}^\dagger) |0\rangle, \quad (146)$$

where $c_{k'\uparrow}^\dagger$ and $c_{-k'\downarrow}^\dagger$ are generators of quanta in the vacuum with momentum k' and $-k'$, respectively having up and down spin. Here $u_{k'}$ and $v_{k'}$ are the variational parameters satisfying $u_{k'}^2 + v_{k'}^2 = 1$. In this discussion of the influential Higgs field, the vacuum state $|0\rangle$ corresponds to the Fermi surface, the lowest energy state in the normal conduction state without condensation. Adopting BCS theory, if there is an attractive force due to the interaction of quanta, then $|\phi_{\text{BCS}}\rangle$ has lower energy than $|0\rangle$ and the Cooper pair forms.

Energy gap and influential force field

According to BCS theory, the energy of the attraction between the quanta forming the Cooper pair is the order parameter Δ , which is the attractive pair potential. This energy corresponds to the decrease in energy from the vacuum state (i.e., the Fermi energy E_F). The order parameter Δ_k for momentum k is expressed as

$$\Delta_k = - \sum_{k'} V_{kk'} \langle \phi_{\text{BCS}} | c_{k'\uparrow} c_{-k'\downarrow} | \phi_{\text{BCS}} \rangle, \quad (147)$$

where $V_{kk'}$ is the magnitude of the interaction. The attractive force acts within a small energy range, and we thus consider that $V_{kk'}$ is constant in this range and thus $\Delta_k = \Delta$. Therefore, a new ground state arises near $E_F - \Delta$ owing to the superconductivity. Then, through the level repulsion, the energy gap 2Δ opens in the range $E_F \pm \Delta$.

The level repulsion implies the extension of the information distance between above and below the gap, and the application of information theory is thus valuable for understanding the phenomenon. Let X and Y be two interacting fermions, let $H(X)$ and $H(Y)$ be their respective entropies, and let $H(X, Y)$ be their joint entropy. The magnitude of attraction due to the BCS pair potential is

$$k_B T [H(X) + H(Y) - H(X, Y)] = \Delta, \quad (148)$$

and we can thus regard $\beta\Delta = MI$. The attractive influential force then acts between X and Y , and

$$\mathcal{F}_{\text{att}} = e^{\beta\Delta}. \quad (149)$$

In contrast, the magnitude of the energy gap due to the level repulsion is

$$2k_B T [H(X) + H(Y) - H(X, Y)] = 2\Delta. \quad (150)$$

This implies that a repulsive influential force also acts between X and Y , and

$$\mathcal{F}_{\text{rep}} = e^{2\beta\Delta}. \quad (151)$$

Therefore, when the Cooper pair forms, not only does the attractive force \mathcal{F}_{att} based on Δ act between the quanta but also there is the repulsive force \mathcal{F}_{rep} resulting from 2Δ . The composite force is

$$\mathcal{F} = \mathcal{F}_{\text{att}} + \mathcal{F}_{\text{rep}}. \quad (152)$$

According to Nambu theory, the vacuum expectation value $\langle \phi \rangle_0$ is generated based on the energy gap 2Δ , and the elementary particles are endowed with mass. In other words, the mechanism of generation of the repulsive force \mathcal{F}_{rep} , which acts between the spacetime quanta within the Higgs boson as the Cooper pair, is expected to lead to symmetry breaking. Thus, the above demonstrates that the BCS phenomenon is a typical example of the influential force proposed in this paper.

Oscillation of spacetime

We next explore how the lattice oscillation, the phonon, described by BCS theory is related to the spacetime vibration. We note that while both the attractive force $\mathcal{F}'_{\text{att}}$ and repulsive force $\mathcal{F}'_{\text{rep}}$ take maximum values at $I_0 = 0$, the spacetime oscillation becomes intense beyond the Planck scale. This

concordance at an extremely close proximity suggests that the influential force originates from the spacetime vibration within the world shorter than the Planck length.

To focus on the phonon generated by the spacetime vibration, we introduce asymptotic background-free quantum gravity with BRST conformal symmetry proposed by Hamada [37]. The world beyond the Planck scale has a large quantum fluctuation of gravity and is the background-free world. Hamada constructed conformal field theory with BRST conformal symmetry and quantized this ultra-fine spacetime. The metric field $g_{\mu\nu}$ is decomposed as

$$g_{\mu\nu} = e^{2\phi}(\hat{g}_{\mu\nu} + th_{\mu\nu} + \dots), \quad (153)$$

where ϕ is the conformal factor field, $\hat{g}_{\mu\nu}$ is the background metric, and $h_{\mu\nu}$ is the traceless tensor field. Additionally, t is the dimensionless gravity coupling constant, representing a deviation from the conformal invariance. Because $t \rightarrow 0$ beyond the Planck scale, the background-free dynamics are realized on this scale. The above $g_{\mu\nu}$ is therefore the metric that uniquely holds at any scale. In the following, we will adopt this particular $g_{\mu\nu}$ as the metric of spacetime.

While Hamada's theory succeeded in the quantization of Einstein gravity, Hamada did not mention the Higgs field, which is expected to be associated with gravity. Hence, with BCS theory in mind, we focus on the oscillation of the metric field $g_{\mu\nu}$ to elucidate the characteristics of the phonon causing the interaction between the vacuum fermions. As noted above, both the attractive and repulsive influential forces strengthen as I_0 approaches zero, and we thus deal with the limit $t \rightarrow 0$. When we fix the original metric field $g_{0\mu\nu}$, the metric field relative to $g_{0\mu\nu}$ is expressed as

$$g_{\mu\nu} - g_{0\mu\nu} = e^{2\phi}(\hat{g}_{\mu\nu} - \hat{g}_{0\mu\nu}). \quad (154)$$

Thus, the metric field $g_{\mu\nu}$ oscillates, reflecting vibrations of the background metric $\hat{g}_{\mu\nu}$ and the conformal field ϕ . The Liouville Hamiltonian H^L of this quantized metric $g_{\mu\nu}$ in $R \times S^1$ is then

$$\begin{aligned} H^L &= \hat{p}^2 + \frac{b}{2} + \sum_{n=1}^{\infty} (\alpha_n^{+\dagger} \alpha_n^+ + \alpha_n^{-\dagger} \alpha_n^-) \\ &= H_O + H_Q, \end{aligned} \quad (155)$$

where \hat{p} is the momentum, $b/2$ represents the Casimir effect, α^+ and α^- respectively represent the left-moving and right-moving modes [37], $H_O = \hat{p}^2 + b/2$ and $H_Q = H^L - H_O$.

Here, H^L in the ground state represents a harmonic oscillation around the original metric $g_{0\mu\nu}$, and its angular frequency is the eigenvalue of $2H^L/\hbar$. At this time, H^L involves the repulsive force, which moves $g_{\mu\nu}$ away from $g_{0\mu\nu}$. Meanwhile, Eq. (154) expresses how $g_{\mu\nu}$ and $g_{0\mu\nu}$ are associated with each other via the vibrating background metric $\hat{g}_{\mu\nu}$, revealing a causal relationship. In other words, the probabilistic attractive force also acts between the spacetime quanta $g_{\mu\nu}$ and $g_{0\mu\nu}$. We therefore regard the background vibration as the equivalence of the lattice oscillation of the BCS theory and consider it as the phonon. In this way, we have obtained a clue as to how to connect the vibration of the background spacetime with that of the influential force field.

Cooper pair and Bogoliubov quasiparticle

BCS theory predicts the appearance of the Bogoliubov quasiparticle, which is the excitation mode under Cooper pair condensation. We examine the emergence of the quasiparticle as follows. The quantized metric field $g_{\mu\nu}$ can be regarded as a fermion because α_n^\pm satisfies the anti-commutation relationship. To clarify the interrelationship between the spacetime quanta, we perform an inverse Bogoliubov transformation of H_Q on the right-hand side of Eq. (155). Then, a pair of interacting fermions emerges according to

$$\begin{aligned} H_Q &= \sum_{n=1}^{\infty} (\alpha_n^{+\dagger} \alpha_n^+ + \alpha_n^{-\dagger} \alpha_n^-) \\ &= \sum_k \xi_k (c_{k\uparrow}^\dagger c_{k\uparrow} + c_{k\downarrow}^\dagger c_{k\downarrow}) - \Delta \sum_k (c_{k\uparrow}^\dagger c_{-k\downarrow}^\dagger + c_{-k\downarrow} c_{k\uparrow}) + C_O = H_{BCS}, \end{aligned} \quad (156A)$$

which is just the BCS Hamiltonian. We here let ξ_n be the energy of a spacetime quantum with respect to the Fermi energy E_F and let $E_n = \sqrt{\xi_n^2 + \Delta^2}$ be the energy of the Bogoliubov quasiparticle and make the replacements $\xi_k = \xi_n/E_n$, $\Delta := \Delta/E_n$, $n = k$, $C_O = 2 \sum_{n=1}^{\infty} (1 - \xi_n/E_n)$, $c_{k\uparrow} = \alpha_n^+ \sqrt{(E_n + \xi_n)/2E_n} + \alpha_n^{-\dagger} \sqrt{(E_n - \xi_n)/2E_n}$, and $c_{-k\downarrow} = -\alpha_n^{+\dagger} \sqrt{(E_n - \xi_n)/2E_n} + \alpha_n^- \sqrt{(E_n + \xi_n)/2E_n}$.

Adopting the inverse Bogoliubov transformation, it is shown that an interacting pair of metric fermions is generated below the Planck length. According to Nambu theory, this pair of fermions can form a complex and undergo Bose–Einstein condensation as a boson. This complex is formed by the attractive pair potential Δ (or the attractive order parameter Δ) acting between the spacetime quanta, and the condensate is regarded as the Higgs field. It is important that Δ represents gravity, which allows us to connect the Higgs field with gravity.

Meanwhile, the above transformation implies that the pair of spacetime quanta $g_{\mu\nu}$ and $g_{0\mu\nu}$ can be regarded as one Bogoliubov quasiparticle; that is,

$$H_Q = \sum_{n=1}^{\infty} \sum_{\sigma=\pm} \gamma_n^{\sigma\dagger} \gamma_n^\sigma, \quad (156B)$$

where $\gamma_n^\pm = \alpha_n^\pm$. The excitation energy of the quasiparticle equals the energy gap 2Δ due to the level repulsion, supporting the expectation that the repulsive force also acts between the pair of fermions. The emergence of this particular quasiparticle thus supports the involvement of both the attractive potential Δ and the repulsive potential 2Δ . Accordingly, the above findings suggest the origin of the influential force \mathcal{F}' with attractive and repulsive terms and the association of this force with gravity.

Order parameters and information distance

The magnitude of the order parameter Δ , which represents the attractive force as gravity acting between the spacetime quanta within the Higgs complex, varies depending on the distance between the quanta and is supposed to follow a canonical distribution. Notably, the 1:2 ratio between the

attractive potential Δ and the repulsive potential 2Δ , which characterizes Nambu theory, coincides with the ratio $I_0:2I_0$ between the resource energy of the attractive term $\mathcal{F}'_{att} = k_1 e^{-I_0}$ and the repulsive term $\mathcal{F}'_{rep} = k_2 e^{-2I_0}$ of the influential force. Together with the probability distribution, this particular energy ratio supports the validity of applying the influential force to the Higgs field.

From the above point of view, we now connect the order parameters to the influential force in the I_0 -spacetime. We first express the conventional distance x using the metric. When the metric is $g_{\mu\nu}$, the distance x varies from the original metric $g_{0\mu\nu}$ by

$$dx = \sqrt{\frac{ds^2 - g_{00}(cdt)^2 - g_{22}dy^2 - g_{33}dz^2}{g_{11}}} . \quad (157)$$

If we set $x = dx$ in terms of the origin, then the information distance $I_0 \geq 0$ between the spacetime quanta is expressed as

$$I_0 = \alpha x - \beta\Delta , \quad (158)$$

where $\alpha = E/\hbar v$ is a parameter of the distribution of the particle mediating information. Meanwhile, the Higgs field ϕ is an order parameter representing the breaking of electroweak symmetry and $\phi = 2\Delta$. The relationship with the information distance $I_0 \geq 0$ is described as

$$I_0 = \alpha x + \frac{1}{2}\beta\phi . \quad (159)$$

Thus, we have linked Hamada's quantum gravity to the I_0 -spacetime. If we neglect the spread in the conventional four-dimensional spacetime and set $x = 0$, then $I_0 = \beta\phi/2$ is proportional to ϕ and inversely proportional to the absolute temperature T .

Metric symmetry combining internal spacetime and external spacetime

A theory linking general relativity and quantum mechanics requires a symmetry that combines the internal spacetime and external spacetime. When spontaneous symmetry breaking occurs, the phase coherence of the Higgs field emerges. The internal spacetime representing the spin dynamics of spacetime quanta is then linked to the order-parameter collective mode of the Higgs condensate. Meanwhile, in terms of $I_0 = \alpha x + r/2$, Eqs. (158) and (159) regard the order parameters as the information metric r of the spacetime quanta and associate them with the conventional distance x . Here, the principle of least action will lead to phase coherence, which connects the internal and external spacetime in the context of the Higgs mechanism. Thus, the combination of the internal and external coordinates within the Higgs field illustrates how the metric symmetry plays a role in the unification of natural forces by the influential force.

Order parameters and influential force potential

As we initially envisioned, we consider the complex scalar field with modulus I_0 in the I_0 -spacetime. Let the potential of the field be the probabilistic influential force potential $\varphi'(I_0)$ between two spacetime quanta X and Y . We again use the reduced information distance for oscillation potentials.

Within this polar local information coordinate spacetime, the use of the reduced information distance supposes that one of the two spacetime quanta is fixed at the origin. At this time, the reduced information distance corresponds to the reduced mass-energy of the other oscillating quantum. This is because I_O is a special coordinate expressing both the position and energy. Then,

$$\varphi'(I_O) = \mathcal{M}(X)\mathcal{M}(Y) \left[-k_1 e^{-I_O} + \frac{k_2}{2} \mathcal{M}(X)\mathcal{M}(Y) e^{-2I_O} \right]. \quad (160)$$

Suppose both spacetime quanta have unit information mass $\mathcal{M} = 1$. In that case, $\varphi'(I_O)$ can be expressed as $\varphi'(I_O) = \varphi'_{att}(I_O) + \varphi'_{rep}(I_O)$, where $\varphi'_{att}(I_O) = -k_1 e^{-I_O}$ is the attractive force potential, which has the information distance I_O corresponding to the attractive pair potential Δ , and $\varphi'_{rep}(I_O) = \frac{1}{2} k_2 e^{-2I_O}$ is the repulsive force potential, which has the information distance $2I_O$ corresponding to the repulsive pair potential $\phi = 2\Delta$. The existence probability of the Higgs boson is then determined by $\varphi'(I_O)$. Thus, we have obtained the Higgs potential represented by the influential force potential. Hereafter, we set the unit information mass $\mathcal{M} = 1$ as that corresponding to the vacuum expectation value $\langle \phi \rangle_0 = 246$ GeV. Additionally, we set the magnitude of the attractive potential Δ corresponding to $\langle \phi \rangle_0$ as $\Delta_0 := \langle \phi \rangle_0 / 2 = 123$ GeV.

$\varphi'_{att}(I_O)$ and $\varphi'_{rep}(I_O)$ are respectively the probabilistic influential force potentials corresponding to the spacetime quanta $\phi_0(x)$ and $\tilde{\phi}_0(x)$ in the conventional x -spacetime with the metric $g_{\mu\nu}$. The energy ratio 1:2 between the attractive term and repulsive term of $\varphi'(I_O)$ also coincides with that of $\phi_0(x)$ and $\tilde{\phi}_0(x)$, which are respectively wave functions of the zero-point oscillation and double-frequency oscillation. These characteristics are again consistent with the discussion presented in this paper.

Taken together, the influential force potential $\varphi'(I_O)$ is well compatible with both order parameters Δ and ϕ of Nambu theory. This supports the involvement of the phonon-based influential force \mathcal{F} in the Higgs mechanism. Furthermore, it is conceivable that the exchange interaction between the generated identical Higgs bosons causes another \mathcal{F}' acting between them, thereby producing the Higgs condensate with phase coherence.

Probabilistic influential force potential and energetic potential

The probabilistic influential force potential $\varphi'(I_O)$ is the cumulative probability amplitude of exchanging information between the origin and a point at the information distance I_O . Meanwhile, consistent with the energy ratio characterizing Nambu theory, $\varphi'(I_O)$ represents the potential constructed by probabilistically distributing the energy between the two energy levels of I_O and $2I_O$. This composition is similar to that of the Morse potential V_M that expresses the interatomic interaction of a diatomic molecule. Therefore, $\varphi'(I_O)$ can be converted into the potential V_H' according to

$$V_H' = -hc[\varphi'(I_O)]^2, \quad (161)$$

which has the dimension of energy.

Influential force field and gravity

Peter Higgs himself denied the association of gravity with the Higgs field in the Standard Model. However, we have demonstrated that the influential force field $\varphi'(I_O)$ explains the Higgs field and is associated with gravity on the quantum level. In addition, we here demonstrate the relevance between the influential force field $\varphi'(I_O)$ and the gravity field V_g on the macroscopic level. If we let \mathcal{F}' be the influential force and let $F_g(x)$ be gravity, then

$$\begin{aligned} F_g(x) &= k_B T \left(\alpha + \frac{\partial}{\partial x} \log \mathcal{F}' \right) \\ &= k_B T \left\{ \alpha + [k_1 e^{-I_0} - k_2 \mathcal{M}(G_1) \mathcal{M}(G_2) e^{-2I_0}]^{-1} \right. \\ &\quad \left. [-k_1 e^{-I_0} + 2k_2 \mathcal{M}(G_1) \mathcal{M}(G_2) e^{-2I_0}] \left(\alpha + \beta \frac{\partial V_g}{\partial x} \right) \right\}. \end{aligned} \quad (162)$$

If $k_1 \gg 2k_2 \mathcal{M}(G_1) \mathcal{M}(G_2) e^{-I_0}$, then

$$F_g(x) = -\frac{\partial V_g}{\partial x}. \quad (163)$$

Therefore, as far as the repulsive force can be neglected, the influential force \mathcal{F}' represents conventional gravity. This suggests that the influential force \mathcal{F}' expresses both the Higgs field and gravity in a unified manner. Moreover, the repulsive force is expected to arise at the unobservable ultramicroscopic level, which in turn conforms to the anticipation of inflation theory.

Visualization of the influential Higgs potential

As the influential Higgs potential, we visualize the probabilistic influential force potential $\varphi'(I_O)$ between the pair of spacetime quanta $g_{\mu\nu}$ and $g_{O\mu\nu}$ (Fig. 5). Assuming that the spacetime fermions have energy corresponding to the unit mass $\mathcal{M} = 1$, we consider the potential between them. The potential equals the lowest excitation energy of the Bogoliubov quasiparticle expressed by H_Q in Eq. (156B). According to Nambu theory, its energy is the gap energy $2\Delta_0$, which is the vacuum expectation value $\langle \phi \rangle_0 = 246$ GeV. However, if we consider that the one spacetime fermion is fixed at the origin, then we can use the reduced mass (i.e., $\langle \phi \rangle_0 / 2 = 123$ GeV) for the other. In this case, both the attractive potential Δ_0 and the repulsive potential $2\Delta_0$ are halved to $\Delta_0/2$ and Δ_0 , respectively. In addition, the reduced information distance $I_O = (\alpha x - \beta \Delta) / 2$ is applied again. The application of this distance allows the focusing on one of the two spacetime quanta; that is, a quantum that oscillates around the origin of the information coordinate spacetime (I_O -spacetime).

Then, $\varphi'(I_O)$ in the two-dimensional figure resembles the Morse potential V_M . $\varphi'(I_O)$ takes the minimum $\varphi'(I_e) = -\left(\sqrt{(\pi)} c \mu_H / 2h\right)^{1/2}$ (i.e., $V_H' = -108.8$ GeV) at the equilibrium information distance $I_e = (\log 2) / 4$ (Fig. 5a). The standard Higgs potential V_H also takes a minimum here, where ϕ becomes $\langle \phi \rangle_0$. As I_O becomes smaller than I_e , $\varphi'(I_O)$ increases, which indicates the strong repulsive force near the origin. Meanwhile, as I_O becomes larger than I_e , $\varphi'(I_O)$ increases moderately and $\lim_{I_O \rightarrow \infty} \varphi'(I_O) = 0$.

As shown in Figure 5a, the influential Higgs potential $\varphi'(I_0)$ is flat in the high-energy region. Its shape is not that of a Mexican hat or a wine bottle like that of the standard Higgs potential V_H . The difference from the Morse potential lies in the attractive and repulsive coefficients; thus, the potential at the origin $\varphi'(0)$ (i.e., $V_H' = -101.1$ GeV) is lower than that at the Fermi surface $\varphi' = 0$. This is consistent with the fact that the attractive force is dominant for gravity in the observable universe.

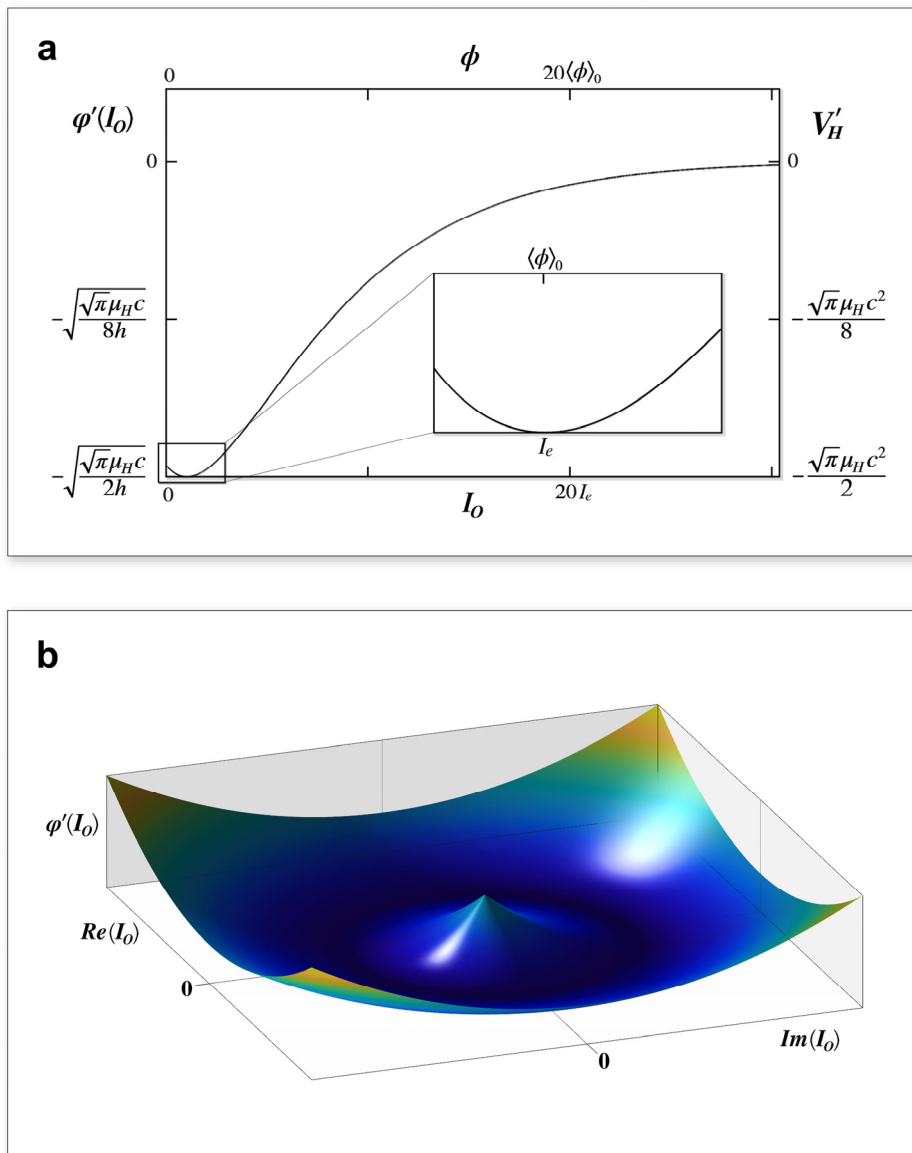


Figure 5. | Influential Higgs potential. a, Probabilistic influential force potential representing the Higgs potential as a function of I_0 . The upper abscissa expresses the corresponding field, while the right ordinate represents the corresponding potential energy. The inset magnifies the neighborhood of the minimal point. **b,** Three-dimensional figure of the potential in the vicinity of the minimal point.

Figure 5b shows a three-dimensional plot of the influential Higgs potential $\varphi'(I_O)$ in the neighborhood of the origin and the minimal points. The standard Higgs potential V_H diverges rapidly as ϕ increases, and its shape is that of a Mexican hat. In contrast, $\varphi'(I_O)$ is finite. Therefore, there is no divergence even if I_O approaches zero or goes to infinity. Because divergence does not occur, renormalization is unnecessary, and we can use the same formula continuously from below the electroweak scale to beyond the Planck scale. This is a large difference from the case of the standard Higgs potential V_H . Simultaneously, this property is critical in terms of the potential that represents gravity. The most significant difficulty in quantizing gravity has been that the gravity potential diverges. In contrast, when expressing gravity by the influential force \mathcal{F}' , the quantum gravity theory will be much easier to establish because the scattering amplitude does not diverge.

From the spacetime quantum to the Higgs sea

The influential Higgs potential $\varphi'(I_O)$ is a Morse-like potential and is a central force that allows us to reduce the two-body problem of the interacting spacetime quanta to a one-body problem. In this case, the mass of the quantum that oscillates at the bottom of the potential is obtained by applying the harmonic approximation and then transforming the result into energy. We thus get a mass of $\langle\phi\rangle_0/2 = 123$ GeV, which can be referred to as the reduced unit mass μ_H . In the influential force potential [Eq. (145)], this μ_H corresponds to the dissociation energy and is close to the Higgs mass of 125 GeV observed by the Large Hadron Collider. We are thus justified in regarding this quantum oscillator as the Higgs boson. Therefore, the spacetime quanta in Hamada theory constitute Higgs bosons. They will undergo Bose–Einstein condensation probably through the exchange interactions that we described in Section (g). This process results in the coherent Higgs condensate, the Higgs sea.

Importance of the equilibrium information distance

The probabilistic influential force \mathcal{F}'_{att} and the repulsive force \mathcal{F}'_{rep} are balanced at the equilibrium information distance I_e . The minimum of $\varphi'(I_O)$ corresponds to -7.7 GeV relative to $\varphi'(0)$. This shallowness of the potential bottom is consistent with the observed small mass of the Higgs boson. Notably, the value of $I_e = (\log 2)/4$ is constant irrespective of the masses of the interacting quanta or their distance in the x -spacetime, and I_e can therefore be regarded as a unit of quantum information (Appendix 23). This particular equilibrium point has a special meaning within the I_O -spacetime as well as within the conventional x -spacetime, as described below.

The distance I_e corresponds to the point where the attractive force and repulsive force are balanced in the I_O -spacetime and thus corresponds to the Schwarzschild radius $r_s = 2Gm/c^2$ in the x -spacetime, where G is the gravity constant and m is the mass of the particle. Moreover, the spacetime quantum located at the origin of the information coordinates is a black hole, and $I_e = (\log 2)/4$ is its

radius in the I_O -spacetime. This black hole has entropy $S_{BH} = I_e$. Moreover, because the attractive pair potential is $\Delta_0/2$, the energy of the black hole is $\mu_{BH} = k_B T \cdot I_e = \mu_H - \Delta_0/2 = \langle \phi \rangle_0/4$.

Although the background freedom holds on the scale of r_s in general, the expectation value of the unique metric field $g_{\mu\nu}$ can be expressed as

$$\langle g_{\mu\nu} - g_{O\mu\nu} \rangle = \langle \hat{g}_{\mu\nu} - \hat{g}_{O\mu\nu} \rangle. \quad (164)$$

By calculating the x -coordinate using the metric as mentioned above, we obtain the expectation value of r_s as $\langle r_s \rangle = \langle \hat{r}_s \rangle$, where $\langle \hat{r}_s \rangle$ is the expectation value of r_s in the background metric. On this basis, $\langle r_s \rangle = 3.24 \times 10^{-52}$ m for the spacetime quantum with a mass of 123 GeV. The parameter of the distribution of the influential force is then $\alpha = I_e / \langle r_s \rangle = 5.35 \times 10^{50} \text{ m}^{-1}$. This α is much larger than $\alpha = 2.4 \times 10^{13} \text{ m}^{-1}$, which is the value calculated adopting lattice quantum chromodynamics [see Section (c)]. The large value of α for gravity demonstrates that the strong repulsive force acts inside the Schwarzschild radius, showing the huge magnitude of the Hamiltonian H_Q of the Bogoliubov quasiparticle generated by the spacetime vibration.

Influential Higgs inflation

The influential force in the I_O -spacetime predicts the strong repulsive force on the ultramicroscopic scale. This coincides with the inflation theory of the early universe. Bezrukov proposed a Higgs inflation model in which the standard Higgs boson is the inflaton [38]. This model assumes a non-minimal coupling between the standard Higgs potential V_H and gravity. Following Bezrukov's Higgs inflation, we consider a model in which the influential Higgs is the inflaton (i.e., influential Higgs inflation). The shape of the curve in Figure 5a highly resembles that of the Higgs inflation potential except for the region inside I_e . However, the influential Higgs potential is $\varphi'(I_O)$, which directly represents the gravity from the very first, and non-minimal coupling with gravity is thus not required. Similarly, $\varphi'(I_O)$ does not have the square of the Ricci scalar, which is included in Starobinsky's R^2 inflation model [39] to allow the coupling of gravity with quantum fields. Contrary to the above two models, the Lagrangian and the action in the influential Higgs inflation model are respectively expressed as

$$\begin{aligned} L_H &= -V'_H = hc[\varphi'(I_O)]^2 \\ S_H &= \int L_H dt = hc \int [\varphi'(I_O)]^2 dt, \end{aligned} \quad (165)$$

which do not include any artificial interaction terms with gravity.

If we suppose that the inflation began at $I_O = 4.97 = 28.7I_e$, then the e-fold is $N = 60$. The slow roll parameters are $\varepsilon = 1.375 \times 10^{-4} \ll 1$ and $|\eta| = 0.0164 \ll 1$, which satisfy the slow-roll conditions. The calculated values of quantum fluctuation are then the spectral index $n_s = 0.966$ and the

tensor-to-scalar ratio $r = 0.0022$, which are compatible with the observations of the Wilkinson Microwave Anisotropy Probe [40] and Planck spacecraft [41]. The good reproduction of the satellite observation data by the calculated values supports the validity of the influential Higgs inflation model, in which the universal force potential $\varphi'(I_0)$ inherently expresses gravity. Notably, the models proposed by both Bezrukov and Starobinsky predicted that $r = 0.003$, which is appreciably higher than our prediction. In the future, it will be desirable to examine observations of the polarized cosmic microwave background made by the next-generation satellite LiteBIRD [42].

Breaking of influential force gauge symmetry

If the inflaton is the influential Higgs, then I_0 decreased as the inflation proceeded. In the early universe at high temperature, spacetime quanta formed the Higgs field based on the gauge symmetry of the attractive term of the mechanistic influential force, which is expressed by $F'_{att}(I_0) = -d\Phi'_{att}(I_0)/dI_0 = -1$. According to the Standard Model, the phase transition of the field occurred at the transition to the low-temperature state. At that time, the relative strength of the repulsive term $|F'_{rep}(I_0)| = d\Phi'_{rep}(I_0)/dI_0 = 2$ to the attractive term $|F'_{att}(I_0)| = 1$ was greatly increased by the decrease in I_0 . As a result, it appears that $F'(I_0) = F'_{att}(I_0)$ spontaneously changed to $F'(I_0) = k_1 F'_{att}(I_0) + k_2 F'_{rep}(I_0)$.

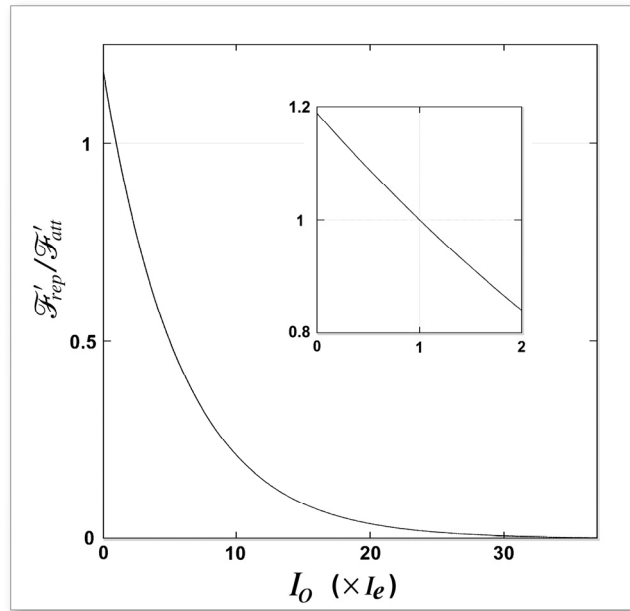


Figure 6. | Ratio between the repulsive influential force and attractive influential force as a function of I_0 . The inset expresses the neighborhood of the equilibrium point I_e .

We now evaluate the weight of the repulsive force. If we calculate the ratio of the repulsive term to the attractive term in the force \mathcal{F}' acting between two spacetime quanta, then $\mathcal{F}'_{rep}/\mathcal{F}'_{att} = \sqrt[4]{2}e^{-I_0}$ (Fig. 6). This ratio is invariant irrespective of either the masses of the quanta or the distance between the quanta in the x -spacetime. I_0 was sufficiently large at the beginning of the inflation, and the repulsive term was negligible. However, with the progression of the inflation, the weight of the repulsive term increased as I_0 decreased. When $I_0 = I_e$, the ratio became 1, and the repulsive force balanced with the attractive force. This means that the strong repulsive force \mathcal{F}'_{rep} acting between the spacetime quanta broke the influential gauge symmetry spontaneously.

Influential force and hierarchy problem

The hierarchy problem asks why the Higgs mass is much smaller than expected from renormalization. To solve this issue, we focus on an example in which a matter particle M with zero mass collides with the Higgs boson H with mass m_H and obtains mass m'_M . The scattering causes an information exchange between the two particles, and MI arises as Eq. (41). The energy difference between before and after the information sharing is expressed as

$$k_B T [H(H) + H(M) - H(H, M)] = -k_B T \cdot \Delta I_0 = k_B T \cdot \Delta MI = -\Delta U. \quad (166)$$

This means that the information distance between the particles decreases by ΔI_0 , and the energy U of the composite system decreases by $-k_B T \cdot \Delta I_0 = k_B T \cdot \Delta MI$. If we now set the conventional distance at the collision as $x = 0$, then the influential force

$$\mathcal{F}'_{att} = e^{\Delta MI} = e^{-\Delta I_0} \quad (167)$$

acts between the particles. The information distance $I_0(M, H)$ is symmetric with respect to the two particles, and the action of the influential force is thus also symmetric. Therefore, the total energy U lost is $(k_B T \cdot \Delta MI)/2$ for each particle. The influential force \mathcal{F}'_{att} decreases the phase difference between the two particles because $\Delta MI = -\Delta I_0 = -\Delta S_a/\hbar$.

If we let T_i ($i = M$ or H) and V_i be the kinetic energy and potential energy of the two particles, respectively, then the Lagrangian L_i before the scattering is

$$L_i = T_i - V_i. \quad (168)$$

After the scattering, however, this L_i changes to

$$L'_i = T_i - \left(V_i + \frac{1}{2} k_B T \cdot \Delta MI \right). \quad (169)$$

This shows that the potential energy increases by $(k_B T \cdot \Delta MI)/2$. Then, according to Nambu theory, this increase in potential implies that the particle M obtains the mass $m'_M = (k_B T \cdot \Delta MI)/2c^2$.

As the matter particle M gains mass, by the symmetry of the influential force, the potential energy of the Higgs boson H increases by the same amount $(k_B T \cdot \Delta MI)/2c^2$. This increase in the potential

indicates that the minimum value of the influential Higgs potential converted to energy decreases from $V'_H(I_e)$ to $V'_H(I_e) - (k_B T \cdot \Delta MI)/2$, and the Higgs mass thus increases. The Higgs mass m'_H after the increase is

$$m'_H = m_H + \frac{1}{2C^2} k_B T \cdot \Delta MI = m_H + m'_M. \quad (170)$$

That is to say, the Higgs mass increases by the mass m'_M , which is the same mass that the matter particle obtains. Notably, the influential force potential $\varphi'(I_0)$ does not diverge; thus, even if m'_M is very large, there is no need to correct the Higgs mass by renormalization. Therefore, applying the influential force \mathcal{F}' can lead to the solution of the hierarchy problem.

Equivalence of gravitational and inertial mass

The equivalence principle states the equivalence of gravitational and inertial mass. Supposing the same scattering as supposed in the previous subsection, we now explain the mechanisms underlying the generation of gravity in proportion to the particle mass. When the mass of Higgs boson is m_H , the Higgs potential $\varphi'(m_H)$ due to the influential force is expressed as

$$\varphi'(m_H) = \sqrt[4]{\pi} \sqrt{\frac{m_H c}{h}} [-2^{\frac{3}{4}} e^{-I_0} + e^{-2I_0}]. \quad (171)$$

If we set

$$r' = r \left(1 + \frac{m'_M}{m_H} \right), \quad I'_0 = \alpha x + \frac{1}{2} r', \quad (172)$$

then after the increase in mass by m'_H ,

$$\varphi'(m_H + m'_M) = \sqrt[4]{\pi} \sqrt{\frac{(m_H + m'_M) c}{h}} [-2^{\frac{3}{4}} e^{-I'_0} + e^{-2I'_0}]. \quad (173)$$

This means that $r = 2\beta V_g(x)$ increases, and the gravity potential thus increases from $V_g(x)$ to $V'_g(x) = V_g(x)(1 + m'_M/m_H)$. Hence, the gravity force $dV_g(x)/dx$ also increases in proportion to the mass obtained by the material particle. If we define the gravitational mass as $m'_{GM} = \{ [V'_g(x)/V_g(x)] - 1 \} m_H$, then it coincides with the inertial mass m'_M :

$$m'_{GM} = m'_M. \quad (174)$$

Thus, the equivalence of gravitational and inertial mass is proved, demonstrating the solution to a central mystery in modern physics once again.

Entanglement entropy and the Higgs field

In this section, we apply BCS theory to spacetime quanta and described how the MI of the spacetime quanta forms the Higgs field. Section (h) showed that $MI = 2S_{EE}$ if the entanglement holds between two spacetime quanta. Meanwhile, the MI explained by BCS theory satisfies $MI = \beta\Delta$. These results suggest a close relationship between S_{EE} and Δ . Here, we derive the relationship between S_{EE} and the Higgs field by considering the entanglement of spacetime quanta.

The electrons that form Cooper pairs in the BCS theory of superconductivity are entangled with each other. The entanglement entropy is proportional to the pairing energy Δ [43,44]. We apply the BCS theory to spacetime quanta, and therefore the entanglement is also supposed to hold between the spacetime quanta by a similar argument. We then ask, even in the case of entanglement between spacetime quanta, does a proportional relationship hold between S_{EE} and Δ ? We here consider the following two points.

- 1) Our model of the Higgs boson is a combination of two spacetime quanta via the phonon, which is the oscillation of the background x -spacetime. The equilibrium information distance $I_e = (\alpha x - \beta \Delta_0)/2 = (\log 2)/4$ is the reduced information distance at which the attractive and repulsive forces are balanced. The use of the reduced information distance corresponds to tracing out one of the two subsystems. This distance I_e corresponds to the Schwarzschild radius within the x -spacetime when we regard the spacetime quantum as a black hole. Here, the black hole has entropy $S_{BH} = I_e = (\log 2)/4$ and energy $\mu_{BH} = k_B T \cdot I_e = \Delta_0/2 = \langle \phi \rangle_0/4$.
- 2) According to the Bekenstein–Hawking formula, the black hole entropy is $A/4$, where A is the area of the event horizon. This entropy is equal to the entanglement entropy S_{EE} according to the Ryu–Takayanagi formula, and A corresponds to the number of Bell pairs (or EPR states). The spacetime quanta in Hamada’s theory are the quantized version of Einstein gravity, and it is thus appropriate to apply the Ryu–Takayanagi formula to the entropy calculation of quantum interactions. If we use this formula, the area A represents the number of one-qubit entanglements per one spacetime quantum. The positive minimum of A is one Bell pair (i.e., $\min(A) = \log 2$), and then $S_{EE} = (\log 2)/4$. This value is equal to $S_{BH} = I_e = (\log 2)/4$; that is, I_e is the entropy of the minimum black hole. The energy of this black hole is $\mu_{BH} = k_B T \cdot S_{EE}$.

It follows from 1) and 2) that at the equilibrium information distance,

$$S_{EE} = I_e = \frac{\log 2}{4} . \quad (175)$$

More generally, there is a proportional relationship between S_{EE} , the pairing energy Δ , and the Higgs field strength ϕ :

$$S_{EE} = \frac{1}{2} \beta \Delta = \frac{1}{4} \beta \phi . \quad (176)$$

In this way, S_{EE} expresses the interaction between spacetime quanta, which explains the Higgs field. Furthermore, $\Delta = 2k_B T \cdot S_{EE}$ expresses the energy ΔU of the gravitational interaction, which satisfies

$$\Delta U = -2k_B T \cdot S_{EE} = -\Delta . \quad (177)$$

Moreover, the relationship with the Higgs mass is expressed as

$$\mu_H = \Delta_0 = 2k_B T \cdot I_e = 2k_B T \cdot S_{EE} , \quad (178)$$

and the source of mass is thus created according to S_{EE} .

As noted above, we have related S_{EE} to the Higgs field as well as the gravitational field, which is an application of the gauge–gravity correspondence. The obtained picture of the Higgs boson is a combination of two black holes separated from each other by their Schwarzschild radii. We will call this arrangement a **black hole dimer**. Given that the dimers are all identical, the above-discovered exchange interaction will induce Bose–Einstein condensation to form the Higgs condensate. However, it is noted that, unlike superstring theory, we deal with a scale much smaller than the Planck scale.

Unification of diverse fields of physics

By replacing the standard Higgs potential V_H with the probabilistic influential force potential $\varphi'(I_O)$, we provided a clue to understanding the principle underlying spontaneous symmetry breaking. We also obtained findings that can lead to the comprehensive unification of the Higgs potential, gravity, inflation, and entanglement entropy. The predicted Higgs mass and the inflation parameters reproduce observations well. Moreover, we presented a plausible explanation for the equivalence of gravitational and inertial mass and provided a possible solution to the hierarchy problem. Thus, applying $\varphi'(I_O)$ has offered explanations for a variety of the issues that exist in modern physics.

(j) Many-body system

This section discusses the influential force in a many-body system, using the information distance $I_O = \alpha x + r/2$. Each physical body shares MI with many other bodies in the universe. In other words, a state change in a single informaton has the possibility of affecting the states of all other informatons, and vice versa. Such information transmission affects I_O of informatons. According to the metric symmetry, this change in I_O affects both the information metric r and conventional distance x . We here discuss the outcome of these events from the information theoretical point of view.

Linear combination of influential orbitals (LCIO)

The attractive influential force ϕ_{ij} that acts between two bodies G_i and G_j is expressed as a function of the information distance $I_O(G_i, G_j)$ between the bodies:

$$\phi_{ij}(I_O) = \mathcal{M}(G_i)\mathcal{M}(G_j)\exp[-I_O(G_i, G_j)] . \quad (179)$$

This equation has the same form as the Slater-type wave function used in the molecular orbital method. We here refer to $\chi_{ij}(I_O) = (\phi_{ij}(I_O), [\phi_{ij}(I_O)]^2)$ as the **influential force orbital**. If we let k_{1ij} be the attractive coefficient, k_{2ij} be the repulsive coefficient, and c_{ij} be the **weight coefficient vector** defined as $c_{ij} := (k_{1ij}, -k_{2ij})$, then the influential force acting between G_i and G_j is expressed as

$$\varphi(G_i, G_j) = c_{ij} \cdot \chi_{ij}(I_O) . \quad (180)$$

The influential force φ of G_i in the many-body system is now expressed by the combination of influential force orbitals as

$$\varphi[\vec{I}_O(G_i)] = \sum_{j \neq i} c_{ij} \cdot \chi_{ij}(I_O) \quad (181)$$

We designate this φ as the **information orbital** of G_i . Equation (181) shows that the information orbital can be represented by the **LCIO** with all other bodies. Furthermore, I_O is the entropic representation of the distance, and the LCIO has a relationship with the information entropy of the many-body system.

Genesis of the repulsive force in a many-body system

We now consider a closed many-body system that comprises N bodies, G_1 to G_N , and examine the outcome of the four fundamental forces in this system. This system can be regarded as a subsystem of the universe. We initially assume that isothermal conditions are maintained and that the entropy of the system $H_{\text{sys}} := H(G_1, \dots, G_N)$ is constant. H_{sys} is then expressed as

$$H_{\text{sys}} = \sum_{i=1}^N H(G_i) - \sum_{i=1}^{N-1} \sum_{j=i+1}^N MI(G_i; G_j) \quad (182)$$

Similar to the discussion of the two-body system, when the mediator particles' energy is also summed, we consider that each body's entropy $H(G_i)$ remains constant irrespective of the change in $MI(G_i, G_j)$. Here, if the attractive force $F'_{att}(I_O)$ acts between the two bodies G_i and G_j , then $MI(G_i, G_j)$ increases while $I_O(G_i, G_j)$ decreases. Because H_{sys} is assumed to be constant, both $\sum MI(G_i, G_k)$ and $\sum MI(G_j, G_k)$ decrease while both $\sum I_O(G_i, G_k)$ and $\sum I_O(G_j, G_k)$ increase, where the summations are taken with respect to k running from 1 to N excepting i and j . This implies that repulsive forces arise both between G_i and G_k and between G_j and G_k in the information coordinate spacetime (I_O -spacetime). Thus, $F'_{att}(I_O)$ for only a single pair of bodies affects the distribution of I_O . Furthermore, from a global point of view, the action of $F'_{att}(I_O)$ is generally regarded as local. Nonetheless, $F'_{att}(I_O)$ affects the global distribution of I_O in the many-body system.

Metric symmetry explaining the repulsive force in a many-body system

To extend beyond the above discussion, we next consider the overall effect of $F'_{att}(I_O)$ that acts between bodies in the system. We now apply the concept of metric symmetry to address this issue. Then, using equations $I_O = \alpha x + r/2$ and $MI = -I_O$, the system entropy H_{sys} is expressed as

$$\begin{aligned} H_{\text{sys}} &= \sum_{i=1}^N H(G_i) + \sum_{i=1}^{N-1} \sum_{j=i+1}^N I_O(G_i, G_j) \\ &= \sum_{i=1}^N H(G_i) + \alpha \sum_{i=1}^{N-1} \sum_{j=i+1}^N x(G_i, G_j) + \frac{1}{2} \sum_{i=1}^{N-1} \sum_{j=i+1}^N r(G_i, G_j) \end{aligned} \quad (183)$$

This equation indicates that H_{sys} equals the total of the sum of each entropy $H(G_i)$ and the sum of each information distance $I_O(G_i, G_j)$. The latter equals the sum of two terms representing the effects of the conventional distance x and the information metric r . Thus, the metric symmetry between x and r also holds in the many-body system.

We here perform a thought experiment in which $F'_{att}(I_O)$ begins to act on bodies before initiating the movement within the conventional coordinate spacetime (x -spacetime). In this case, $\sum_{i=1}^{N-1} \sum_{j=i+1}^N r(G_i, G_j)$ first decreases, indicating that the average information metric $\langle r \rangle$ of the bodies contracts. Then, because H_{sys} is constant, $\sum_{i=1}^{N-1} \sum_{j=i+1}^N x(G_i, G_j)$ tends to increase, owing to the metric symmetry. This implies that the average conventional distance $\langle x \rangle$ between bodies is prone to expand. This observation again shows that the repulsive force $F'_{rep}(I_O)$ arises in the many-body system. After that, according to $F'_{att}(I_O)$, the distance x between bodies begins to decrease locally, which causes an additional decrease in H_{sys} , thereby inducing further repulsion and fluctuation.

We next extend our discussion to the universe by applying the above argument. We consider that the universe is an isolated many-body system, in which we assume that the isothermal conditions prevail. We here set H_{univ} as the total entropy of the universe. Then, $\Delta H_{univ} \geq 0$ always holds. Accordingly, H_{univ} generally increases through the action of $F'_{att}(I_O)$. Therefore, in a thought experiment similar to that described above, $\langle x \rangle$ in the universe tends to expand even more rapidly than that in the aforementioned many-body subsystem. This observation reinforces the production of the repulsive force $F'_{rep}(I_O)$, which is consistent with the currently observed expansion of the universe.

Finally, we present an overview of the counteracting forces; that is, $F'_{att}(I_O)$ and $F'_{rep}(I_O)$. Regarding every two-body interaction, their absolute proportion is $F'_{att}(I_O) : F'_{rep}(I_O) = \Phi'_{att}(I_O) : \Phi'_{rep}(I_O) = 1 : 2$, which holds everywhere in the I_O -spacetime. Thus, $F'_{att}(I_O)$ and $F'_{rep}(I_O)$ may respectively constitute 33% and 67% of the cosmic energy, which likely explains the observation that the total energy of the universe contains 32% ordinary and dark matter and 68% dark energy.

Chapter summary

The influential force unifies all natural forces and exists in diverse physical systems. First, we found that the influential force arises from the fluctuation of microstates. Next, a novel exchange interaction was discovered, which is caused by the MI of identical quanta. We then explained spontaneous symmetry breaking for the Higgs field, accounted for the equivalence of inertial and gravitational mass, and obtained a possible resolution to the hierarchy problem. The influential Higgs effectively reproduced the data of the cosmic microwave background even without a non-minimal coupling with gravity. The Higgs field, gravity, inflation, and entanglement entropy were all explained in a unified manner. Finally, we estimated the magnitude of the dark energy as 67% of the cosmic energy density. The influential force thus offers a starting point for a variety of next-generation physics research.

3. Influential force and biology

(a) Gene information and life

Life is the mode of existence of information, whereas a gene is a representative informaton. The purpose of the presence of genes is to contribute to the *fitness* of organisms; that is, the *existence probability* of the informaton itself. For this purpose, genes encode molecular structures of gene products by which they control materials and energy, thereby regulating organisms' architectures and functions. After 4 billion years of evolution, genes have acquired profound capabilities. Genes now have the ability to construct complex body plans and establish higher functions represented by the nervous system and immune system. The evolution of gene information has played a central role in this remarkable progress.

Enormity of gene information

Genes encode a vast amount of information, and hence the entire picture is difficult to understand. Primarily, gene information is expressed at a molecular level; that is, it determines the state of biomolecules in many ways. These include quantitative change, structural alteration, chemical modification, control of degradation, a shift in localization, complex formation, and chromatin regulation. Moreover, the expression profiles of genes are heterogeneous, being either continuous or discrete. Genes are simultaneously expressed in many different forms, in multiple subcellular components, and in diverse types of cells that compose various organs. With these characteristic expression patterns, the information is represented as the morphological and physiological properties of organisms with sociobiological phenotypes and is subject to natural selection through them [45].

Because of the extreme complexity of gene information, it has been challenging to understand the broad totality of gene information using a unified measure. To address this issue, we employ a statistical approach as follows.

Measure of gene information

We consider gene information as an *extremely high-dimensional random variable*. To deal with the large complexity, we introduce a *linear index* that allows us to treat the information as if it were one-dimensional (Appendix 9). Let the gene G_X be an informaton that expresses at most n different *observable states*. With a linear index k ($1 \leq k \leq n$), we can arrange the states according to their likelihood. For the k -th observable state, the realization probability $p(I_k)$ of the self-information I_k follows an exponential distribution, which we call the infocanonical distribution (Appendix 17):

$$p(I_k) = e^{-I_k}, \quad (184)$$

where the information level of the observable state is I_k . This is the basic equation applied to all self-information. However, because of a substantial number of observable states, the realization probability can be represented by a continuous random variable, as detailed below.

The observable states expressed by G_X depend on the deoxyribonucleic acid (DNA) sequence. Because the average length of human protein-coding genes is approximately 50 kb [46], the total number of states per haploid (n_{tot}) is $n_{tot} = 4^{50,000} = 9.99 \times 10^{30,102}$, which is incomparably larger than the Avogadro number 6.022×10^{23} in the physical system. Although the number of constituent particles in the physical system is enormous, we cannot distinguish the particles from one another. In contrast, although there are few constituent components in the biological system, the permutation becomes the source of information, thus allowing the generation of an immense amount of highly multidimensional information. However, even if G_X takes a set of n ($\leq n_{tot}$) observable states in a multidimensional space, each state can be distinguished by the linear index k ($1 \leq k \leq n$). Furthermore, because n is immensely large, the information levels of the observable states can be divided in an extremely fine manner and sufficiently represented by a one-dimensional continuous random variable X .

On the basis of the preceding discussion, we propose that the realization probability of the information level X of the observable state follows the infocanonical distribution (Appendix 17):

$$p(X) = \frac{1}{I} e^{-\frac{X}{I}}, \quad (185)$$

where I is the expectation value of X that represents the likelihood that X has a more unlikely information level (Fig. 7). When $I = 1$, this equation always holds because it is equivalent to that of self-information [Eq. (184)]. The use of I allows us to contrast a pair of genes with particularly closely related properties, such as a wild type and its mutant. In general, most mutants have characteristics very similar to those of their wild-type counterparts in the context of neutral or nearly neutral theory.

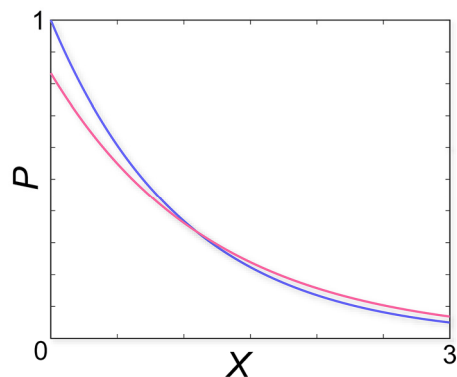


Figure 7 | Probability distribution of the observable information level of genes.

The realization probability of the observable information level X of a gene G follows the infocanonical distribution. Blue and red lines respectively represent the probability density functions for G_{wt} and G_{mut} . I_{wt} and I_{mut} are thus set to 1 and 1.2, respectively.

In such cases, I is a valuable measure with which to compare the relative ability to express the same set of states. We denote the wild-type and mutant values of I by I_{wt} and I_{mut} , respectively, and we set $I_{wt} = 1$. At this time, $\Delta I = I_{mut} - I_{wt}$ represents the average increase in the relative ability accompanying the mutation. We refer to this I as the *effective self-information* and X as the *observable information level*.

Besides the continuous representation, X can be evenly split into n levels and distinguished by the linear index k ($1 \leq k \leq n$). The realization probability $p(X_k)$ of the k -th information level X_k is obtained as an infocanonical distribution (Appendix 17) by integrating $p(X)$ from $(k-1)/n$ to k/n as

$$p(X_k) = (e^{\frac{1}{\lambda}} - 1)e^{-\frac{k}{\lambda}}, \quad (186)$$

where $\lambda = nI$ is the mean index value of X_k . Finally, if the gene G_X evolves to have large λ , then the information entropy $H(G_X)$ is expressed with the same λ as (Appendix 25)

$$H(G_X) = - \sum_{k=1}^n p(X_k) \log p(X_k) \simeq 1 + \log \lambda \quad (187)$$

(b) Gene information and the influential force

In physics, the influential force generates the MI of informatons, thereby affecting the existence probability of the system. In a biological system, genes encode the information that affects the existence probability of the organism. Therefore, similar to the case for physical informatons, genes may produce MI and exert the influential force on the whole body, including other genes. Indeed, many functional interactions exist between genes. Specifically, the complex function of a gene is exerted not only by the gene itself but also by the collaborations with many other genes. Such an ability to achieve an appropriate partnership has evolved in the long history of life. Genes thus encode not only molecular structures but also mutual relationships among themselves.

MI of the gene and body

If a gene G affects other parts of the body B , then the information of G alters the state of B , and vice versa. If we regard the states of both G and B as extremely high-dimensional random variables, then the information entropy of the whole body $H(G, B)$ is

$$H(G, B) = H(G) + H(B) - MI(G; B), \quad (188)$$

which demonstrates that an increase in $MI(G; B)$ reduces the overall entropy of the organism. Therefore, as in the physical system, genes exert an influential force on the body, thereby decreasing the uncertainty, bringing order to the biological system, and ultimately achieving homeostasis in the organism. In this manner, genes increase the fitness; that is, the probability of the existence of a living thing. We propose that this is the most essential function of genes.

We can ignore changes in energy and temperature during the evolution of the probability distribution of $MI(G;B)$. Thus, under natural conditions, $MI(G;B)$ follows an infocanonical distribution (Appendix 2, 17):

$$p = e^{-MI(G;B)} . \quad (189)$$

MI of genes

When a gene G_i functions in an organism, it exchanges information with other genes. As an example, we consider a case in which the products of two genes G_i and G_j function through the formation of a dimer. In this instance, the functional complex cannot form if the two genes are expressed in different tissues at different times. Therefore, selective pressure favors their information sharing regarding spatiotemporal expression, and MI evolves between genes. The increase in $MI(G_i;G_j)$ leads to an ordered state within the biological system, which is advantageous to the organism's survival. The probability of co-expression of two genes increases with time, and the *influential force* is thus regarded as acting between the genes. The important thing is that the influential force manifests itself in the process of evolution.

In the present case, the joint information entropy $H(G_i, G_j)$ of the two-gene system is

$$H(G_i, G_j) = H(G_i) + H(G_j) - MI(G_i; G_j) . \quad (190)$$

We can again ignore changes in energy and temperature during evolution. Therefore, under natural conditions, the magnitude of $MI(G_i; G_j)$ follows an infocanonical distribution (Appendix 2, 17):

$$p = e^{-MI(G_i; G_j)} . \quad (191)$$

If we let G_j be the gene whose $MI(G_i; G_j)$ with G_i is of j -th strength, then $MI(G_i; G_j)$ follows the logarithmic distribution of magnitude rank j :

$$MI(G_i; G_j) = -A \log(j) + B , \quad (192)$$

where A and B are constants (Appendix 2).

(c) Relationship between gene information and fitness

To address the role of gene information, we should consider the evolutionary process of genes in light of information theory. During evolution, base substitutions in DNA occur at the rate of approximately $1E-9$ /year·base; that is, mutations arise at a frequency of approximately $5E-5$ /year on average per gene of length 50 kb (or once in approximately 20,000 years). With these mutations, the gene information changes stochastically both at the individual and population levels. We here consider this process from a statistical and probabilistic point of view, applying the analogy of the ergodic hypothesis in statistical mechanics.

Gene information and fitness

We consider a large population of haploid organisms and initially discuss a single gene. We assume that the population is under sufficiently stable conditions. Let X be the continuous random variable for the observable information level of a gene G . Then, X follows an infocanonical distribution at the individual level [Eq. (185), Appendix 17]. Meanwhile, at the population level, the population frequency of X is supposed to follow a distribution according to its realization probability in each organism. This is an analogy to the ergodic hypothesis. In this case, X is expected to follow an infocanonical distribution within the population. Such individual variance arises stochastically, such as through differences in the genetic background and developmental environment. Moreover, at the individual level, gene information makes unlikely (preferred) physiological states more likely to occur. From a macroscopic point of view, gene information increases the frequency of individuals with unlikely (desirable) states. These gene activities are represented by the effective self-information I .

We now discuss the relationship between gene information and fitness. For the meantime, we ignore the time-lapse and adverse effects of mutations. We initially consider cases in which mutations are advantageous for survival unless stated otherwise. We assume that a wild-type G_{wt} of gene G can mutate into a mutant G_{mut} , thereby increasing the relative fitness from $r_{wt} = 1$ to $r_{mut} = 1 + s$, where s is the *selection coefficient*. Given that G_{wt} and G_{mut} generally have very closely related properties, they are assumed to take virtually the same set of observable states. If we let I_{wt} and I_{mut} be the effective self-information I of G_{wt} and G_{mut} , respectively, the change in I through the mutation is $\Delta I = I_{mut} - I_{wt}$. If we assume that all the mutations contribute to positive fitness and if $0 < \Delta I \ll 1$, then

$$s = \Delta I, \quad r_{mut} = 1 + \Delta I \simeq e^{\Delta I} \quad (193)$$

(Appendix 10). These formulas demonstrate that the gene information is linked to fitness; that is, to the probability of the existence of the gene itself after one generation. A change in gene information thus contributes to the evolutionary process.

Major problem in evolutionary biology

We next discuss the evolution of protein-coding genes, considering deleterious mutations. Selection coefficient s is a factor that determines the probability of the ultimate fixation of G_{mut} in the population [47]. When the product $N_e s$ of the effective population size N_e and s is greater than 1, the fixation probability increases owing to positive selection. However, $N_e s$ takes a negative value far more frequently than a positive value. Moreover, in negative cases, most mutations are excluded from the population because they are deleterious. Hence, the mutants with $N_e s < 0$ are not fixed, except where $|N_e s|$ is sufficiently small ($|N_e s| < 1$) and thus *nearly neutral*. For these reasons, a single amino acid substitution arises at a rate of only $5E-6$ /year to $2E-7$ /year per 100 amino acid residues. Moreover, even when mutations occur, most substitutions are conservative, such that there are no appreciable

changes in the structures and functions of proteins [45]. These observations indicate that protein evolution is highly conservative.

In contrast to the preceding discussion, the evolution of phenotypes is, in some cases, much more rapid than that of proteins. The first example is the *Cambrian explosion*, which is the most important evolutionary event in the history of life on Earth. Approximately 542 million years ago in the Cambrian period, complex animals suddenly appeared in the fossil record [48]. However, despite conspicuous morphological evolution, extensive gene evolution did not occur during this period. Conversely, large-scale gene evolution took place as many as 300 million years before the Cambrian explosion [49].

The second example is the evolution of the human nervous system, which was accompanied by the establishment of a highly organized neuronal network. The human genus has evolved for approximately 2 million years. Nevertheless, *Homo sapiens* with excellent brain function emerged only 250,000 years ago. Furthermore, the remarkable cultural progress called the *great leap forwards* began merely 50,000 years ago [50].

For the above examples, it is difficult to fully explain phenotypic evolution based on the traditional theory that depends on the mutations in a protein sequence. Thus, there may be another mechanism underlying evolution.

Temporal dynamics of exonic gene information

To solve the serious problem presented so far, it is necessary to focus on the amount of information that a gene can express. Hence, from the perspective of information theory, we consider temporal changes in the fitness associated with mutations in coding sequences. We assume that the information expressed from a gene G whose total length of the exon is L bases takes at most $N (\leq 4^L)$ observable states as multidimensional random variable X (Appendix 25). Suppose X is the information level of X that can be evenly split into n levels and distinguished by a linear index k ($1 \leq k \leq n$). Let $p(X_k)$ be the realization probability of the k -th information level X_k . Then, X_k follows the infocanonical distribution $p(X_k) = [\exp(1/\lambda) - 1] \exp(-k/\lambda)$, where λ is the mean index value of X_k [Eq. (186)].

Moreover, we assume a situation in which stochastic mutations cause fluctuations in each occurrence probability of gene information. Let the realization probability $p(X_k)$ of each level X_k change γ_k -fold at time t , owing to mutations in the entire region of gene G , including non-coding sequences. We assume that γ , the average value of γ_k , is expressed by the diffusion equation with diffusion coefficient D and follows a normal distribution with a mean of 1 and variance of $\sigma^2 \simeq 2Dt$. The absolute expectation value of ΔI is then expressed as

$$\langle |\Delta I| \rangle \simeq 2 \sqrt{\frac{Dt}{\pi}}, \quad (194)$$

which increases in proportion to the square root of t (Appendix 26).

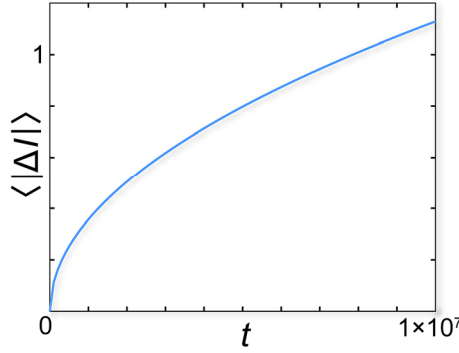


Figure 8 | Temporal change in the effective self-information of a gene. The mean of $|\Delta I|$ at time t increases as a square-root function of t . $D = 1\text{E-}7$.

Figure 8 depicts a simulation of the temporal profile of $\langle |\Delta I| \rangle$ corresponding to Eq. (194) (Appendix 26). It is indicated that, even if the biological effect of each mutation is subtle, the information changes accumulate with time and can generate more information and thus greater fitness. We propose that this is the fundamental mechanism by which mutations in a single gene lead to novel phenotypes. However, considering the difficulty of fully explaining phenotypic evolution by mutations in exonic sequences, it is probable that $\langle |\Delta I| \rangle$ does not necessarily increase efficiently. We will thus explore other mechanisms that facilitate the evolution of organisms.

Temporal dynamics of non-exonic gene information

We next examine whether $|\Delta I|$ becomes sufficient to account for the evolution of phenotypes when we also consider mutations in the non-coding regions. We consider a medium-sized gene of 50 kb [46] that contains 1.5 kb exons. When the non-coding region is considered in addition to the exons, the mutable sequence is expected to become 30 times longer. It is known that the mutation rate in non-coding regions is approximately 3 times the rate of non-synonymous substitutions in exons [51]. We assume that the magnitude of the fitness generated by a mutation in the non-coding region is s_0 times that generated by a non-synonymous substitution in an exon. The expected ratio of the change in fitness due to the non-coding region s_{nc} ($:= \Delta I_{nc}$) to that due to the non-synonymous substitution in the exon s_{ex} ($:= \Delta I_{ex}$) is

$$\frac{s_{nc}}{s_{ex}} = \frac{\Delta I_{nc}}{\Delta I_{ex}} = 90 \times s_0. \quad (195)$$

In Eq. (195), however, it is known that the non-coding sequence contributes much less than the exon sequence to the fitness. If we estimate s_0 as $1\text{E-}4$ or $1\text{E-}3$, then ΔI_{nc} is 0.9% or 9% of ΔI_{ex} , respectively. These values are much less than those of the fitness associated with protein evolution. In conclusion, even if we consider the non-coding regions as well as the coding regions, it is difficult to fully explain phenotypic evolution only by changes in a single gene.

(d) Relationship between intergenic MI and fitness

This section considers a two-gene system instead of the single-gene system that can hardly explain the evolutionary process of organisms. Specifically, we focus on the interaction between genes from the perspective of information theory. We consider the MI and attractive influential force \mathcal{F}_{att} acting between genes as in the physical system. Here, again adopting the analogy of the ergodic hypothesis, we suggest that \mathcal{F}_{att} likely contributes to the fitness of organisms.

Two-gene system without MI

Regarding the organismal population described in the preceding section, we consider two genes, $G1$ and $G2$, whose linkage can be neglected. The wild types $G1_{wt}$ and $G2_{wt}$ of the genes can mutate into $G1_{mut}$ and $G2_{mut}$, respectively. Let the relative fitness of the wild-type combination be $r_0 = 1$. Let the increase in the fitness due to the single mutation $G1_{wt} \rightarrow G1_{mut}$ be s_1 with the relative fitness $r_1 = 1 + s_1$. Let the rise in the fitness due to the mutation $G2_{wt} \rightarrow G2_{mut}$ be s_2 with the relative fitness $r_2 = 1 + s_2$. The relative fitness of the double mutant is then expressed as $r_3 = 1 + s_1 + s_2$ (Table 1, where we set the epistatic coefficient $\varepsilon_{two} = 0$) [47]. However, from the aforementioned observations of a single gene mutation, $s_1 < 0$, $s_2 < 0$, and thus $r_3 < 1$ in most cases. Therefore, it is unlikely that changes in each gene readily lead to phenotypic evolution.

Two-gene system with MI

We next consider a scenario in which, between two mutants, a new interaction is formed, accompanied by an increase in fitness. Let $\varepsilon_{two} (>0)$ be the *epistatic coefficient* of the mutants, which is the increase in fitness due to the interaction of the mutants. The relative fitness of the double mutant is then expressed as $r_3 = 1 + s_1 + s_2 + \varepsilon_{two}$ (Table 1) [52]. Let MI_{wt} be *MI* of the two wild-types $G1_{wt}$ and $G2_{wt}$, and MI_{mut} be *MI* of the two mutants $G1_{mut}$ and $G2_{mut}$. The increase in mean *MI* due to the double mutation is then $\Delta MI = MI_{mut} - MI_{wt}$. This shift in *MI* reflects changes in gene relationships that accompany all possible heritable changes in the genome. The changes include structural alterations in DNA that range from both gene regions to the entire genome; structural changes in the chromatin; epigenetic alterations; changes in the spatiotemporal expression; quantitative and qualitative changes; modifications and degradations; shifts in localization; and the complex formation of gene products, such as proteins and RNA (including miRNA and lncRNA).

We initially consider the case that the change in intergenic *MI* provides a fitness advantage. We assume that, in each organism, the spontaneous probability $p(MI)$ that the magnitude of *MI* becomes MI follows an infocanonical distribution $p(MI) = e^{-MI}$ [Eq. (17), Appendix 17] (Fig. 9a, blue line). The population frequency of *MI* is supposed to follow a distribution according to the realization probability in each organism. Again, this is an analogy to the ergodic hypothesis. As a result, *MI* is expected to follow an infocanonical distribution within the population of organisms.

Table 1. Relative fitness of each genotype

Type	Genotype	Fitness
0	$(G 1_{wt}, G 2_{wt})$	$r_0 = 1$
1	$(G 1_{mut}, G 2_{wt})$	$r_1 = 1 + s_1$
2	$(G 1_{wt}, G 2_{mut})$	$r_2 = 1 + s_2$
3	$(G 1_{mut}, G 2_{mut})$	$r_3 = 1 + s_1 + s_2 + \varepsilon_{two}$

We assume that $\Delta MI > 0$ and that all new interactions between the mutants contribute to the positive fitness. We then have

$$\varepsilon_{two} = \Delta MI \quad (196)$$

(Appendix 11). Keeping in mind that most mutations are nearly neutral, if we consider $s_1 \simeq 0$ and $s_2 \simeq 0$, we have $r_3 \simeq 1 + \varepsilon_{two}$. Moreover, if $0 < \Delta MI \ll 1$, then

$$r_3 \simeq 1 + \Delta MI \simeq e^{\Delta MI} . \quad (197)$$

The above discussion of a haploid population also applies to a diploid population as long as the change in fitness is sufficiently small [45]. The positive fitness values for most mutations are extremely small, and this requirement is thus satisfied in most cases.

When compared with a quantum system, MI_{wt} and MI_{mut} correspond to MI of physical bodies when the information metric r is large and small, respectively. Meanwhile, the fitness corresponds to the realization probability of each state. These correspondences respectively coincide mathematically: $MI_{mut} - MI_{wt} = \Delta MI = -\Delta r/2$ and $r_3 \simeq e^{\Delta MI} = e^{-\Delta r/2} = e^{-\beta \Delta U}$. These equations support the following notion that the attractive force \mathcal{F}_{att} acts similarly in the biological and quantum systems.

(e) Influential force acting between two genes

According to the discussion presented thus far, $\varepsilon_{two} = \Delta MI$ is expected to generate an attractive force \mathcal{F}_{att} between genes, facilitating the biological evolution. The present section considers this possibility by adopting a probabilistic model and simulations.

Simultaneous mutation model

Suppose that the mutations of two genes arise simultaneously. In that case, the ultimate fixation probability of the double mutant is calculated by substituting $r_3 = 1 + s_1 + s_2 + \varepsilon_{two}$ into the solution to the Kolmogorov backward equation [53] obtained by Kimura [47]. Let N_e be the effective population size, $s_0 = 0$ be the selection coefficient of the neutral gene, and u_0 be the ultimate fixation probability

of the neutral gene. When ΔMI is zero, if we let $s = s_1 + s_2$ be the selection coefficient of the double mutant and u be its fixation probability, then the relative fixation probability u/u_0 is

$$u/u_0 = \frac{N_e s}{1 - e^{-N_e s}} . \quad (198)$$

In contrast, when $\varepsilon_{two} = \Delta MI (> 0)$ is generated, if we let u' be the fixation probability, then its relative fixation probability u'/u_0 is

$$u'/u_0 = \frac{N_e (s + \varepsilon_{two})}{1 - e^{-N_e (s + \varepsilon_{two})}} . \quad (199)$$

The graph of u'/u_0 is a translation of ε_{two} towards the left of the graph of u/u_0 (Fig. 9b). In the latter graph, the fixation probability increases with ε_{two} . When $s > 0$, the effect of ε_{two} is subsidiary. When $s \approx 0$, however, if $\varepsilon_{two} > 1/N_e$, then $u'/u_0 \approx N_e \cdot \varepsilon_{two} > 1$. This indicates that $\varepsilon_{two} = \Delta MI$ determines the fixation of the nearly neutral mutants. More remarkably, in the region $s < 0$, which covers most cases, ε_{two} has a large multiplication effect on the fixation probability (Fig. 9c); that is, compensatory effects for deleterious mutations are observed depending on ε_{two} . Thus, under an ideal condition where the two mutations arise simultaneously, $\varepsilon_{two} = \Delta MI$ is critical for the fixation.

Moran process model

We now discuss the more general case in which the two mutations occur independently. To explore the effects of $\varepsilon_{two} = \Delta MI$, we examine the temporal changes in genotype frequencies adopting the Moran process model [54]. Using Kolmogorov forward equations [53], we explore how the combination of mutant genes spreads and becomes fixed in a haploid population of N individuals. Whereas a randomly chosen individual proliferates in proportion to the fitness r at time t , another randomly chosen individual dies; thus, the population remains constant. Moreover, mutations arise with probabilities u_1 – u_4 defined as follows. We assume backward mutations do not occur. Let u_1 and u_2 be the probabilities that $G1_{wt}$ and $G2_{wt}$ mutate into $G1_{mut}$ and $G2_{mut}$ at time t , respectively. Let u_3 be the probability that $G1_{wt} \rightarrow G1_{mut}$ arises under the condition of $G2_{wt} \rightarrow G2_{mut}$ and let u_4 be the probability that $G2_{wt} \rightarrow G2_{mut}$ occurs under the condition of $G1_{wt} \rightarrow G1_{mut}$. We refer to individuals whose genotypes ($G1, G2$) are (wt, wt), (mut, wt), (wt, mut), and (mut, mut) as types 0, 1, 2, and 3, respectively, and let r_0, r_1, r_2 , and r_3 be their relative fitness values.

Under the above conditions, the probabilities X_0 – X_3 that the entire population is occupied by individuals of types 0–3, respectively, are expressed by Kolmogorov forward equations [55]:

$$\begin{aligned} \frac{dX_0}{dt} &= -[a(t) + b(t) + c(t)]X_0, & \frac{dX_1}{dt} &= a(t)X_0 - d(t)X_1, \\ \frac{dX_2}{dt} &= b(t)X_0 - e(t)X_2, & \frac{dX_3}{dt} &= c(t)X_0 + d(t)X_1 + e(t)X_2, \end{aligned} \quad (200)$$

where $a(t)$, $b(t)$, $d(t)$, and $e(t)$ are respectively the transition rates of type 0→1, type 0→2, type 1→3, and type 2→3 at time t , and $c(t)$ is the tunneling rate at which the transition type 0→3 occurs directly through the two mutations without remaining in middle stages [56]. If $a(t)$ to $e(t)$ are constant, then the solutions to these equations are (Appendix 12)

$$\begin{aligned} X_0(t) &= \exp[-(a+b+c)t], & X_1(t) &= \frac{a}{d-a-b-c} \{\exp[-(a+b+c)t] - \exp(-dt)\}, \\ X_2(t) &= \frac{b}{e-a-b-c} \{\exp[-(a+b+c)t] - \exp(-et)\}, & X_3(t) &= 1 - X_0(t) - X_1(t) - X_2(t). \end{aligned} \quad (201)$$

Next, after defining the influential force acting between genes, we will examine the effect of the force on these genotype frequencies.

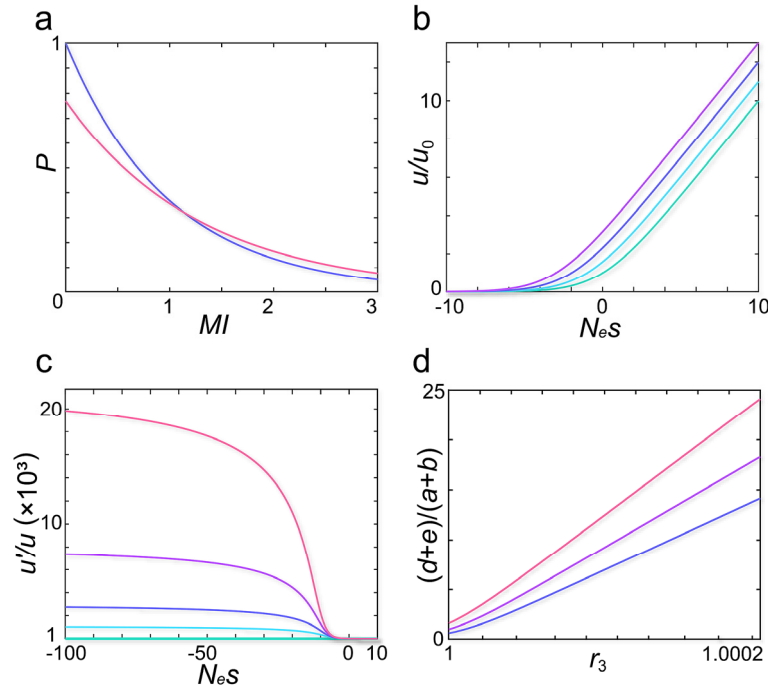


Figure 9 | Fitness and MI. **a**, Distribution of MI for $G1$ and $G2$. The curves represent the probability density functions of MI in the wild-type (blue line) and mutant populations (red). **b**, Relative fixation probability of mutant to neutral genes. N_e is the effective population size set to $1E5$, and s ($= s_1 + s_2$) is the selective coefficient. u_0 and u are the ultimate fixation probabilities of neutral and mutant genes, respectively. The epistatic fitness ε_{two} is set to 0 (green line), $1E-5$ (light blue), $2E-5$ (blue), or $3E-5$ (purple). **c**, Ratios of the relative fixation probabilities. Green, light blue, blue, purple, and red lines represent the ratios in the cases that $\varepsilon_{two} = 0, 7E-5, 8E-5, 9E-5,$ and $1E-4$ to that of $\varepsilon_{two} = 0$, respectively. **d**, Relationship between $(d + e)/(a + b)$ and $r_3 \approx \mathcal{F}_{att}$ when $s_1 \approx 0$ and $s_2 \approx 0$ in the Moran process. r_1 and r_2 are $(1-5E-6)$ (red line), 1 (purple), or $(1+5E-6)$ (blue).

Influential force acting between genes

To address the involvement of the intergenic MI in evolution, we determine the attractive force \mathcal{F}_{att} in terms of population genetics. We here define **MI inherent** (MI_{inh}) as the true value of $MI(G1;G2)$ and **MI observed** (MI_{obs}) as the apparent value calculated from genotype frequencies. We hereafter formulate the influential force acting between genes, evaluate its relationship with MI_{inh} and MI_{obs} , and show that the attractive force \mathcal{F}_{att} decreases the information metric r of $G1$ and $G2$.

As previously described in this section, we assume a new interaction between the mutated states of genes. ΔMI_{inh} for two genes is then equal to ΔMI for their mutants. To define the influential force acting between genes, we again employ the probabilistic definition in Eq. (16), which has been used for the force acting between physical bodies. Considering that the fitness represents the **existence probability** of a gene after one generation, the attractive force acting between genes is viewed as the ratio of the fitness value with the epistatic fitness ε_{two} to that without it. Instead of physics, however, we here apply population genetics that deals with the allele frequency in a given population. The attractive force \mathcal{F}_{att} acting between two genes is therefore defined as the ratio of the genotype frequency of double mutants with (p') and without (p) ε_{two} in the next generation, where $p' + p = 1$ and the initial ratio of frequency is set to 1:1.

In focusing on the effects of ε_{two} , we consider the case that both mutations $G1_{mut}$ and $G2_{mut}$ are neutral; that is, $s_1 = s_2 = 0$. The selection coefficient of the double mutant is then $s_1 + s_2 + \varepsilon_{two} = \varepsilon_{two}$. Unless the selection is strong, $dp'/dt = \varepsilon_{two}p'p$ is sufficiently accurate [45], where t is the generation number. The solution to this differential equation is $p'_t/p_t = p'_0/p_0 \exp(\varepsilon_{two} \cdot t)$, where p'_0/p_0 is the initial ratio of frequency and takes a value of 1 at $t = 0$. We finally obtain \mathcal{F}_{att} by setting $t = 1$ as

$$\mathcal{F}_{att} = e^{\varepsilon_{two}} = e^{\Delta MI_{inh}}. \quad (202)$$

If we regard that the wild types are ground states, and if we let the initial $MI = MI_{inh}(G1_{wt}; G2_{wt})$ be zero, then the **attractive force acting between genes** is formulated as $\mathcal{F}_{att} = e^{MI}$. This formula is identical to that acting between physical bodies irrespective of s_1 and s_2 [Eq. (A72), Appendix 13].

Moran process model, influential force, and MI inherent (MI_{inh})

We apply the Moran model to evaluate the effects of the influential force on evolutionary processes. As t increases, owing to the population's finite size, dynamic properties of the biological informants dissociate from those of physical informants. The characteristics of the force are therefore best assessed by taking the limit $t \rightarrow 0$. Thus, $\mathcal{F}_{att} = \exp(MI_{inh})$ is expressed as the attractive microforce (Appendix 24) acting between $G1_{mut}$ and $G2_{mut}$ at $t \rightarrow 0$:

$$\mathcal{F}_{att} = e^{\Delta MI_{inh}} \simeq \lim_{t \rightarrow 0} \frac{X_3}{(X_1+X_3)(X_2+X_3)} \simeq \frac{ad+be}{2ab} . \quad (203)$$

If we further assume $a \simeq b$, then \mathcal{F}_{att} can be described as

$$\mathcal{F}_{att} \simeq \frac{d+e}{a+b} \quad (204)$$

(Appendix 13). This formula indicates that the attractive force \mathcal{F}_{att} accelerates the transition from the single to double mutant and thus facilitates the biological interaction between $G1$ and $G2$.

As shown in Figure 9d, we perform a simulation to verify this equation. Using the Moran model, transition rates $a-e$ are calculated, and $(d+e)/(a+b)$ is plotted against r_3 . If $0 < \varepsilon_{two} \ll 1$ and $s_1, s_2 \simeq 0$, then $(d+e)/(a+b)$ has a nearly linear relationship with $r_3 \simeq 1 + \varepsilon_{two} \simeq \exp(\Delta MI_{inh}) = \mathcal{F}_{att}$. This observation supports Eq. (204) and implies that the Moran process model is a good representation of the influential force \mathcal{F}_{att} acting between genes.

Moran process model, influential force, and MI observed (MI_{obs})

Whereas MI_{inh} is not observable, MI_{obs} is an observable value calculated from the genotype frequencies. Using the probabilities X_0-X_3 , MI_{obs} is expressed as

$$\begin{aligned} MI_{obs} = & X_0 \log \frac{X_0}{(X_0+X_2)(X_0+X_1)} + X_1 \log \frac{X_1}{(X_1+X_3)(X_0+X_1)} \\ & + X_2 \log \frac{X_2}{(X_0+X_2)(X_2+X_3)} + X_3 \log \frac{X_3}{(X_1+X_3)(X_2+X_3)} . \end{aligned} \quad (205)$$

As detailed by Appendix 13(3), we derive temporal changes in MI_{obs} applying $X_0(t)-X_3(t)$ described in Eq. (201). If we set $X_0(0) = 1$ and $X_1(0) = X_2(0) = X_3(0) = 0$, the second derivative of MI_{obs} at $t = 0$ is expressed as

$$\begin{aligned} \left. \frac{d^2 MI_{obs}}{dt^2} \right|_{t=0} & \simeq 2(ad+be) \log \mathcal{F}_{att} + 2(2ab-ad-be) \\ & = 2(ad+be) \Delta MI_{inh} + 2(2ab-ad-be) . \end{aligned} \quad (206)$$

This equation shows the relationship between MI_{inh} and MI_{obs} . Furthermore, we refer to \mathcal{F}_{obs} as an **apparent attractive force** observed from the genotype frequencies in a finite-size population. At $t = 0$, this \mathcal{F}_{obs} is derived as the product of the reduced mass of informatons and the second derivative of the information metric $r(G_1, G_2) = H(G_1) + H(G_2) - 2MI_{obs}$. We here assume that $H(G_1)$ and $H(G_2)$ are large and virtually constant throughout the process. If $0 < \Delta MI_{inh} \ll 1$, then \mathcal{F}_{obs} is expressed as

$$\begin{aligned} \mathcal{F}_{obs} & \simeq \frac{4}{W(G_1)^{-\hat{\alpha}} + W(G_2)^{-\hat{\alpha}}} [(ad+be) \Delta MI_{inh} + 2ab-ad-be] \\ & = A_M \log \mathcal{F}_{att} + B_M \simeq A_M (\mathcal{F}_{att} - 1) , \end{aligned} \quad (207)$$

where $A_M = 4(ad+be)/[W(G_1)^{-\hat{\alpha}} + W(G_2)^{-\hat{\alpha}}]$ and $B_M = 4(2ab-ad-be)/[W(G_1)^{-\hat{\alpha}} + W(G_2)^{-\hat{\alpha}}]$.

Equation (207) shows that the true attractive force \mathcal{F}_{att} working at the gene level and the apparent attractive force \mathcal{F}_{obs} observed at the population level are nearly linear. Thus, \mathcal{F}_{att} acts between genes both microscopically and macroscopically. It accelerates the formation of the double mutant, reduces the information metric r of genes, and thereby enhances their mutual relationships.

Simulations

The above discussion suggests that the influential force \mathcal{F}_{att} generates gene interactions via the effect of $\varepsilon_{two} = \Delta MI$. We next examine the effect of ε_{two} by simulating temporal changes in the genotype frequencies using Eq. (201). We use the following parameters in the simulations [45]. The population size N is set to $1E5$, which is generally used for the population genetics of mammals. The base substitution rate is $1E-9/\text{year}\cdot\text{base}$, and the mutation rates u_1-u_4 (Fig. 10a) are thus set to $1E-6$ per generation per gene with total exons of 1 kb. One generation takes 1 year. We regard the time when type 3 (i.e., the double mutant) accounts for 99.5% of the population as the fixation time of type 3.

• Case 1: $s_1, s_2 \geq 0$

The solution for a pair of neutral mutations (i.e., $s_1 = s_2 = 0$) is shown in Figure 10b. The figure shows how mutants $G1_{mut}$ and $G2_{mut}$ arise and are finally fixed in the population. When $\varepsilon_{two} = 0$, type 3 becomes fixed after 5.43 million years. However, when a cooperative interaction arises with $\varepsilon_{two} = 1E-4$, type 3 is fixed 2.16 times earlier; that is, at 2.51 million years (Fig. 10c). We next show the cases in which $s_1 = s_2 = 1E-4$; that is, both mutations are strongly advantageous. When $\varepsilon_{two} = 0$ (Fig. 10d) and $\varepsilon_{two} = 1E-4$ (Fig. 10e), the time to fixation is largely shortened to 589,000 and 368,000 years, respectively. Although less effective than in neutral mutants, a positive effect of ε_{two} on the fixation is again evident. These results indicate that ε_{two} accelerates the fixation of both neutral and advantageous mutants. Finally, because $\mathcal{F}_{att} \approx (d + e)/(a + b)$ is equal to 1.00, 9.24, 1.00, and 1.98 in Figures 10b, c, d, and e, respectively, an increase in ε_{two} is considered to accelerate the fixation of the double mutant through \mathcal{F}_{att} in Figures c and e.

• Case 2: $s_1, s_2 < 0$

In contrast to Case 1, in which both mutations are nondeleterious (i.e., $s_1, s_2 \geq 0$), most mutations are deleterious (i.e., $s_1, s_2 < 0$), categorized as Case 2, and are subject to negative selection. Therefore, the solution in the region $s_1 < 0$ and $s_2 < 0$ is essential to understanding what allows the phenotypic evolution of organisms. As presented below, even in this adverse situation, advantageous information can be fixed owing to the intergenic MI.

When a single gene mutation is deleterious (i.e., $s < 0$), the mutation is regarded as nearly neutral [45] only if it satisfies $|N_e s| < 1$. Therefore, unless the effective population size N_e is sufficiently small, fixation hardly occurs because of negative selection pressure [45]. When N_e is $1E5$, the maximal $|s|$

satisfying $|N_e s| < 1$ is as little as $1E-5$; thus, only a limited range of mutant genes can survive. Likewise, when we consider $s = s_1 + s_2$ in a two-gene system, only $|N_e(s_1 + s_2)| < 1$ is nearly neutral. In contrast to this limitation associated with near neutrality, however, it is worth remarking that the *tunneling effect* becomes more pronounced in populations with larger N_e [56]. Thus, tunneling might overcome the inherent limitation of the theory of near neutrality.

We calculate the tunneling rate c in a two-gene system (Appendix 12) and assess its contribution to the fixation process (Fig. 10f–h). With fixed settings $\varepsilon_{two} = 0.10$ and $s_1 = s_2$, we search for a range in which tunneling occurs. Under the same conditions as above ($N = 1E5$, $u_1 - u_4 = 1E-6/\text{year}$), we find that tunneling occurs over a wide range; that is, $-10,000 < N_e(s_1 + s_2) < -1$. In this case, the tunneling

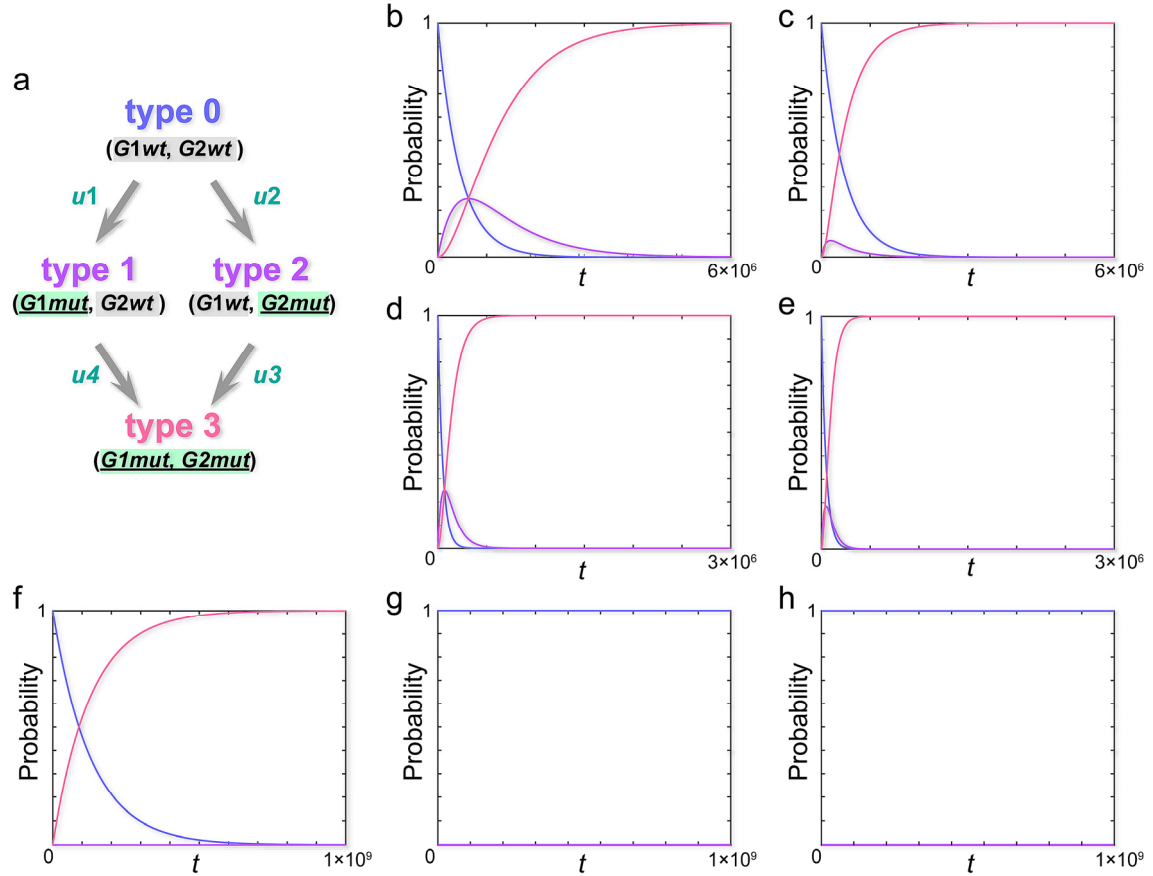


Figure 10 | Moran process. **a**, Mutation rates in the Moran process. $u_1 - u_4$ are the probabilities of mutations between types 0–3. **b–e**, Effect of ε_{two} on the fixation probability of a double mutant when $s_1 = s_2 \geq 0$. Blue, purple, and red lines represent X_0 , $X_1 (= X_2)$, and X_3 , the probabilities that the entire population is occupied by types 0–3 at time t , respectively. The parameters are $N = 1E5$ and $u_1 - u_4 = 1E-6/\text{year}$. **(b)** $\varepsilon_{two} = 0$ and $s_1 = s_2 = 0$. **(c)** $\varepsilon_{two} = 1E-4$ and $s_1 = s_2 = 0$. **(d)** $\varepsilon_{two} = 0$ and $s_1 = s_2 = 1E-4$. **(e)** $\varepsilon_{two} = 1E-4$ and $s_1 = s_2 = 1E-4$. **f–h**, Effect of ε_{two} on the probability when $s_1 = s_2 = -0.049$. **(f)** $\varepsilon_{two} = 0.10$ and a tunneling rate $c = 7.76E-9$. **(g)** $\varepsilon_{two} = 0.10$ and $c = 0$. **(h)** $\varepsilon_{two} = 0$.

effect increases the prevalence of type 3, which leads to its ultimate fixation. When we examine the population change with settings $s_1 = s_2 = -0.049$, which is close to the lower bound of the range, the transition from type 0 to type 3 arises without increasing the prevalence of types 1 and 2 (Fig. 10f). This indicates the emergence of the tunneling phenomenon, and type 3 is finally fixed at 0.7 billion years. In contrast, type 3 does not appear when we set either $c = 0$ (Fig. 10g) or $\varepsilon_{two} = 0$ (Fig. 10h). These results indicate that type 3 becomes fixed depending on both ε_{two} and the tunneling effect.

In addition to the range in which tunneling is possible, when $-1 \leq N_e(s_1 + s_2) < 0$, type 3 is also fixed depending on ε_{two} , although there is no tunneling event in this range. Taken together, $\varepsilon_{two} = 0.10$ contributes to the fixation under the union of the two ranges; that is, $-10,000 < N_e(s_1 + s_2) < 0$, which amounts to up to 94.9% of the joint distribution of $s_1 < 0$ and $s_2 < 0$ [57]. In summary, advantageous information can be fixed according to both MI and the tunneling phenomenon in most cases of $s_1, s_2 < 0$, similar to the situation for $s_1, s_2 \geq 0$.

• Conclusions

Comparing the results of Kolmogorov forward and backward equations, the conditions required for the ultimate fixation of the double mutant coincide with each other; that is,

$$s_1 + s_2 + \varepsilon_{two} > 0 . \quad (208)$$

Accordingly, even when $s_1 + s_2 < 0$, if the epistatic fitness $\varepsilon_{two} = \Delta MI$ is sufficiently large, then there is compensation for the harm of deleterious mutations, resulting in an increase in the fixation probability. Considering that most mutations in protein sequences are at least slightly deleterious (i.e., $s_1 < 0$ and $s_2 < 0$) [57], a large part of the fixed advantageous information can be attributed to ΔMI . Thus, as a result of the influential force \mathcal{F}_{att} , the intergenic MI can function as a determinant of the evolution of the two-gene system.

In contrast to the potential roles of $\Delta MI = \varepsilon_{two}$ in the fixation, however, the time required for fixation varies dramatically depending on s_1 and s_2 . Fixation takes a short time (e.g., 1E5–1E7 years) when $0 < s_1 + s_2$ but a very long time (e.g., 1E8–1E9 years) when $-\varepsilon_{two} < s_1 + s_2 < 0$. For the latter condition, the fixation depends on the tunneling rate, which slows the fixation process and may ultimately reduce the actual chance of fixation. Nevertheless, we should remember that this adverse condition accounts for most cases. Therefore, additional compensatory mechanisms need to be considered in explaining the phenotypic evolution properly. Such additional mechanisms will be further discussed in later sections of this chapter. In summary, in both analytical and simulative approaches, $\mathcal{F}_{att} = \exp(\Delta MI)$ was demonstrated to be involved in the evolutionary process of the two-gene system. The next section elucidates the causal mechanisms of the MI of genes and considers the relationship between biological and physical systems.

(f) Evolution of the MI of two genes

Gene information is extremely high-dimensional with respect to the number of features. Accordingly, the degree of freedom that can affect intergenic MI is very high. For example, the MI of two genes can be increased successively by the detailed adjustment of a broad spectrum of conditions, such as optimizing protein sequences and expression conditions, adding transport signals, and integrating miRNA regulations.

Intergenic MI caused by fluctuation

We again apply the fluctuation model to explain the evolution of the MI of two genes. In the evolutionary process, stochastic fluctuations occur for each gene informaton owing to mutational events. Intergenic MI can therefore be generated by mutations of both genes in the nucleotide sequences, including untranslated regions. We consider two genes G_X and G_Y , and let L_X and L_Y bases be their respective lengths on chromosomes. We assume that the information expressed from G_X and G_Y takes at most $M (\leq 4^{L_X})$ and $N (\leq 4^{L_Y})$ observable states as multidimensional random variables X and Y , with information levels X and Y , respectively. Suppose that X and Y can be evenly split into m and n levels that are distinguished by linear indices k ($1 \leq k \leq m$) and l ($1 \leq l \leq n$), respectively. Additionally, we assume that the realization probability of the k -th information level of X , X_k , follows the infocanonical distribution $p(X_k) = [\exp(1/\lambda) - 1] \exp(-k/\lambda)$ while that of the l -th information level of Y , Y_l , follows $p(Y_l) = [\exp(1/\nu) - 1] \exp(-l/\nu)$, where λ and ν are respectively the mean index values of X_k and Y_l .

As in the case of physics, we employ a **fluctuation factor** ε_{kl} that represents the relationship between the joint probability and marginal probabilities; that is, $\varepsilon_{kl} = p(X_k, Y_l) / [p(X_k)p(Y_l)]$. We again suppose that ε_{kl} fluctuates close to a value of 1, and we express ε_{kl} as $\varepsilon_{kl} = \gamma_{kl}\delta_{kl}$ using the two factors γ_{kl} and δ_{kl} , which independently fluctuate owing to changes in X and Y , respectively. The introduction of γ_{kl} and δ_{kl} is justified in that the mutations of each gene generate epistasis. We further let γ and δ be the averages of γ_{kl} and δ_{kl} over k and l , respectively, and we assume that both γ and δ follow a normal distribution with a mean of 1 and variance of σ^2 . We here think that both m and n are sufficiently large. Then, MI (in terms of MI_{inh}) of G_X and G_Y is expressed as

$$\begin{aligned} MI &\simeq \sum_{k=1}^m \sum_{l=1}^n MI_{kl} \\ &\simeq (1-\gamma\delta)^2 \left(1 - e^{-\frac{m}{\lambda}}\right) \left(1 - e^{-\frac{n}{\nu}}\right). \end{aligned} \quad (209)$$

This is the identical formula as for physical informatons [Eq. (112), Appendix 7]. Because m/λ and n/ν hardly change according to the nearly neutral theory, ε_{kl} is supposed to fluctuate without changes in the marginal probabilities $p(X_k)$ and $p(Y_l)$. Thus, MI is also generated between genes because of fluctuations in their microstates caused by mutations.

Temporal profiles of intergenic MI

We examine temporal changes in the MI of genes. In evolution, mutational events arise stochastically in the gene regions of G_X and G_Y , including untranslated sequences. These mutations lead to both quantitative and qualitative alterations in the gene information, and we will therefore extend the previous considerations for physical systems. As described in Section (f) of Chapter 2, we examined the effects of fluctuations in the stochastic factor ε_{kl} and obtained Eq. (113). We here consider a situation where mutations affect the probability distribution of the information levels of the genes.

We make the following assumptions in analyzing the effects of mutations on $MI(G_X; G_Y)$. First, as previously mentioned, we assume that a diffusion equation describes the distributions of γ and δ with diffusion coefficient D and that both γ and δ follow a normal distribution having a mean of 1 and variance $\sigma^2 \approx 2Dt$ at time t . Second, we assume that the total numbers m and n of the information levels in G_X and G_Y increase by Δm and Δn per unit time, respectively. Third, we assume that both Δm and Δn follow a Poisson distribution having a mean μ per unit time. Finally, we assume that λ and ν , the expectation values of k and l , respectively, increase at a rate κ , accompanying the evolution of each gene. Then, whereas γ and δ follow the diffusion equation, Δm , Δn , $\Delta\lambda$, and $\Delta\nu$ are expected to be proportional to t . Under these conditions, the expectation value of MI in terms of MI_{inh} is

$$\begin{aligned} \langle \Delta MI \rangle \simeq (4D^2t^2 + 4Dt) & \left\{ \left(1 - e^{-\frac{m}{\lambda}}\right) \left(1 - e^{-\frac{n}{\nu}}\right) \left[\left(\frac{1}{\lambda} + \frac{1}{\nu}\right) \mu t + 2 \right] \right. \\ & \left. - \left[\frac{m}{\lambda^2} e^{-\frac{m}{\lambda}} \left(1 - e^{-\frac{n}{\nu}}\right) + \frac{n}{\nu^2} e^{-\frac{n}{\nu}} \left(1 - e^{-\frac{m}{\lambda}}\right) \right] \kappa t \right\} \end{aligned} \quad (210)$$

(Appendix 14). Note that, whereas $\langle \Delta MI \rangle$ is the cubic function of t , $\langle |\Delta I| \rangle$ is the square root function of t [Eq. (194)]. Hence, $\langle \Delta MI \rangle$ increases with time and can be greater than $\langle |\Delta I| \rangle$. If we consider that the same values of $\langle |\Delta I| \rangle$ and $\langle \Delta MI \rangle$ contribute equally to the fitness, then it is possible that $\langle \Delta MI \rangle$ contributes more than $\langle |\Delta I| \rangle$ to the positive fitness (Fig. 11).

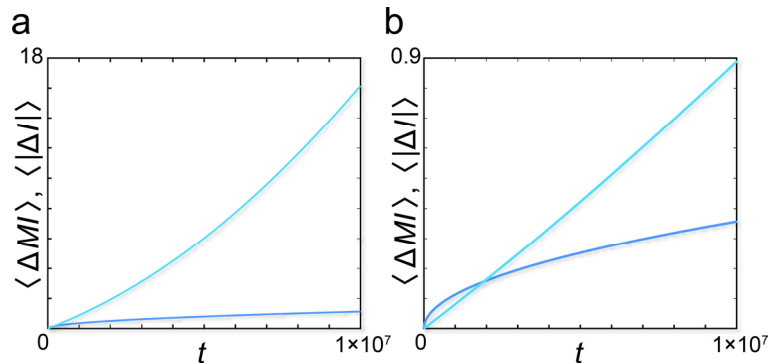


Figure 11 | Evolution of intergenic MI. **a, b,** Temporal change in means of ΔMI (light blue) and effective self-information $|\Delta I|$ (blue). The diffusion coefficient D is $1E-7$ (**a**) or $1E-8$ (**b**).

According to Eq. (187), the sum of information of the genes is $H(G_X) + H(G_Y) \approx 2 + \log(\lambda + \Delta\lambda) + \log(v + \Delta v)$, and the mean satisfies $\langle H(G_X) + H(G_Y) \rangle \approx 2 + \log(\lambda + \kappa t) + \log(v + \kappa t)$. Here, the expectation value of the information metric is $\langle r \rangle = \langle H(G_X) + H(G_Y) \rangle - 2\langle \Delta MI \rangle$. Because $\langle H(G_X) + H(G_Y) \rangle$ increases as the sum of the logarithmic functions of t while $\langle \Delta MI \rangle$ is the cubic function of t , $\langle r \rangle$ decreases with time. The average change rate of the information metric $\langle r \rangle$ is thus expressed as

$$\begin{aligned} \frac{d\langle r \rangle}{dt} \simeq & \kappa \left(\frac{1}{\lambda + \kappa t} + \frac{1}{v + \kappa t} \right) \\ & - 2 \left(1 - e^{-\frac{m}{\lambda}} \right) \left(1 - e^{-\frac{n}{v}} \right) \left\{ 12D^2 \left(\frac{1}{\lambda} + \frac{1}{v} \right) \mu t^2 + \left[4D \left(\frac{1}{\lambda} + \frac{1}{v} \right) \mu + 16D^2 \right] t + 8D \right\} \\ & + 2 \left[\frac{m}{\lambda^2} e^{-\frac{m}{\lambda}} \left(1 - e^{-\frac{n}{v}} \right) + \frac{n}{v^2} e^{-\frac{n}{v}} \left(1 - e^{-\frac{m}{\lambda}} \right) \right] \kappa (12D^2 t^2 + 8Dt) . \end{aligned} \quad (211)$$

As long as $d\langle r \rangle/dt < 0$, $\langle r \rangle$ decreases monotonously. This condition is satisfied by the parameters used in the following simulations. In contrast, $\kappa = 0$ and $\mu = 0$ in physical systems under isothermal conditions. Therefore, the above condition is always satisfied, and $\langle r \rangle$ decreases monotonously for physical informatons.

Simulations

To confirm the analytical observations described above, we perform numerical simulations as presented in the following part of this subsection. As expected, $\langle \Delta MI \rangle$ considerably increases while $\langle r \rangle$ monotonously decreases with time t , for the parameters given below.

Adopting the Moran process, we examine the effects of the increase in ΔMI [Eq. (210)] on the evolution of the two-gene system comprising $G1$ and $G2$. According to the literature [57], 1% of newly generated mutations are supposed to be advantageous. With this in mind, we here assume that the epistatic fitness can be approximated as a linear function of $\langle \Delta MI \rangle$ such that $\varepsilon_{t_{wo}} = 0.01 \times \langle \Delta MI \rangle$. According to this model, $\varepsilon_{t_{wo}}$ increases with ΔMI over time. In contrast, the preceding simulations employed a static model in which an advantageous fitness $\varepsilon_{t_{wo}} = \Delta MI$ was set to a constant value.

We use time-dependent parameters in the following simulations. We now regard the transition rates $a(t) - e(t)$ in Eq. (200) as functions of time according to Eq. (A66) in Appendix 12 and calculate the frequency of each genotype using the general solution given by Eq. (A64). We let the population size $N = 1E5$, numbers of gene information levels $m = n = 1000$, the expectation values of k and l be $\lambda = v = 100$, and $\kappa = 1E-7$ (/year) be the change rates of λ and v . We set the mutation rates as $u_1 - u_4 = 1E-6$ (hit/year). The population is then homogeneous with respect to the genotype most of the time [56]. Moreover, we estimate the change rate of the number of information levels and the diffusion coefficient as 1/10 of $u_1 - u_4$ and thus set $\mu = 1E-7$ and $D = 1E-7$.

• **Case 1: $s_1, s_2 = 0$**

We first assume that both mutations are neutral; that is, $s_1 = s_2 = 0$. Fixation then occurs depending on the increasing value of $\varepsilon_{two} = 0.01 \times \langle \Delta MI \rangle$. When the tunneling rate is artificially set as $c = 0$, type 3 (i.e., the double mutant) fixes at 2.40 million years without tunneling. In contrast, when c is naturally allowed to become positive, tunneling occurs with a maximal c value of $1.69\text{E-}5$. As a result, type 3 fixes at 0.56 million years, in a process that is 4.31 times faster than that without tunneling (Fig. 12a). Compared with the slopes in the cases using the fixed value of ΔMI [Fig. 10b and 10c in Section (g)], the slopes of these fixation curves are steeper, reflecting the continuous increase in ΔMI .

We here calculate the percentage occupancy of $\langle \Delta MI \rangle$ within the fixed total information comprising $\langle \Delta MI \rangle$ and $\langle |\Delta I| \rangle$. The occupancies of $\langle \Delta MI \rangle$ at the time of fixation with and without tunneling are 46.9% and 68.3%, respectively. In terms of the information metric of the two genes, $\langle r \rangle$ shortens by 0.940 and 4.77 nat with and without tunneling, respectively. These observations indicate that the increasing ΔMI and the tunneling effect cooperatively promote the fixation of neutral mutants and contribute to the fixation of advantageous information. As a result, the two genes become closer in the information metric spacetime (r -spacetime).

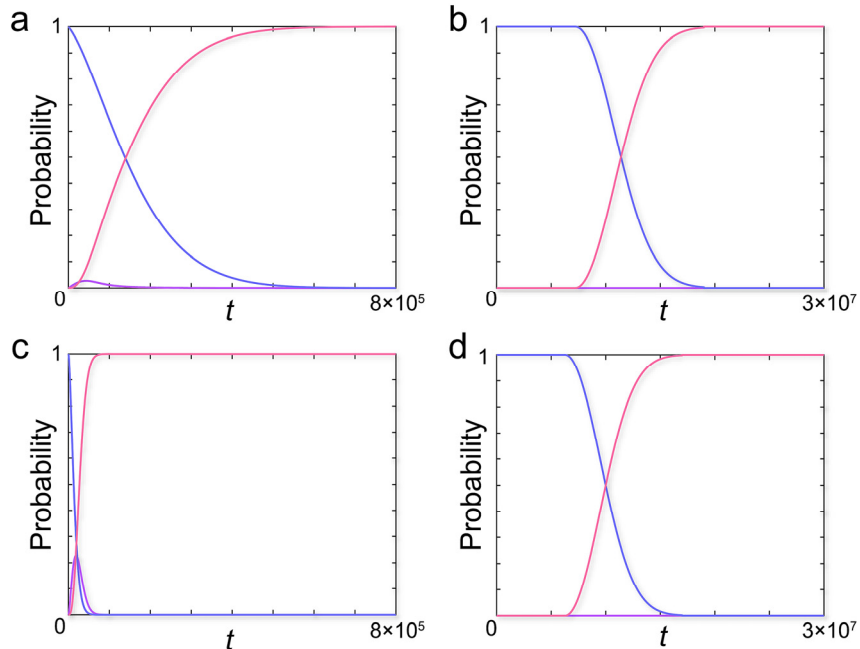


Figure 12 | Effect of time-dependent ε_{two} on the fixation probability of a double mutant. The parameters are $N = 1\text{E}5$ and $u_1-u_4 = 1\text{E-}6/\text{year}$. ε_{two} changes as $\varepsilon_{two} = 0.01 \times \langle \Delta MI \rangle$ through Eq. (210). Blue, purple, and red lines respectively represent X_0 , $X_1 (= X_2)$, and X_3 . **a**, $s_1 = s_2 = 0$. **b**, $s_1 = s_2 = -0.05$. **c**, $s_1 = s_2 = 0.01 \times \langle |\Delta I| \rangle$. **d**, $s_1 = s_2 = -0.05 + 0.01 \times \langle |\Delta I| \rangle$.

• **Case 2: $s_1, s_2 = -0.05$**

We next consider the case in which the fitness of each mutant takes a constant negative value, similar to the case in Fig. 10f. With settings $s_1 = s_2 = -0.05$ [i.e., $N_e(s_1 + s_2) = -10,000$], $\varepsilon_{\text{two}} = 0.01 \times \langle \Delta MI \rangle$ increases by more than 0.10 at 7.21 million years, at which time the total fitness $s = s_1 + s_2 + \varepsilon_{\text{two}}$ becomes positive. Type 3 is finally fixed at 18.6 million years owing to the tunneling effect, with the maximal c value being $9.83\text{E-}07$ (Fig. 12b).

The above results show that an increased $\varepsilon_{\text{two}} = 0.01 \times \langle \Delta MI \rangle$ overcomes the deleterious effects of the harmful mutations, even though $N_e(s_1 + s_2) = -10,000$ is much less than the lower limit of the near neutrality; that is, $N_e(s_1 + s_2) = -1$. Notably, the information occupancy of $\langle \Delta MI \rangle$ at the time of fixation is as high as 93.4%. Thus, most of the fixed information corresponds to $\langle \Delta MI \rangle$ but not to $\langle |\Delta I| \rangle$ of each gene. Taking these results together with those of Case 1, while almost all mutations fall in the range of $-10,000 < N_e(s_1 + s_2) < 0$, a fraction of the fixed advantageous information is attributed to ΔMI in this range. Moreover, at fixation, $\langle r \rangle$ decreases 86.4 nat, which is more than 91 times the value in Case 1 with tunneling. These findings demonstrate that $\langle \Delta MI \rangle$ largely contributes to the attraction of the two genes in the information metric spacetime (r -spacetime).

Case 3: $s_1, s_2 = 0.01 \times \langle |\Delta I| \rangle$

We next examine the case in which $\langle |\Delta I| \rangle$ as well as $\langle \Delta MI \rangle$ increases from an initial value of zero [Eq. (194)]. When we set $s_1 = s_2 = 0.01 \times \langle |\Delta I| \rangle$, the fixation takes 70,020 years without tunneling (Fig. 12c), which is much faster than that in Case 1 (i.e., $s_1 = s_2 = 0$). This is because s_1 and s_2 are positive under these conditions (Fig. 10d showing the control). Although small, $\langle r \rangle$ decreases by 0.113 nat at the fixation. At the same time, $\langle \Delta MI \rangle$ accounts for 23.0% of all fixed information, which is the smallest value among the four cases tested.

• **Case 4: $s_1, s_2 = -0.05 + 0.01 \times \langle |\Delta I| \rangle$**

We finally explore the case in which s_1 and s_2 increase from a negative initial value (Fig. 12d). When we set $s_1 = s_2 = -0.05 + 0.01 \times \langle |\Delta I| \rangle$, the fixation occurs at 16.1 million years, which is a shorter time than in Case 2 (i.e., $s_1 = s_2 = -0.05$). This is because r_1 and r_2 continue to increase under these conditions. In Case 4, however, the occupancy of $\langle \Delta MI \rangle$ reaches 92.3% at the fixation, which is almost the same value as in Case 2. Likewise, $\langle r \rangle$ decreases 68.2 nat, which is the second-largest value, following Case 2. These results indicate that, even though $\langle |\Delta I| \rangle$ is increasing, the increase in $\langle \Delta MI \rangle$ is much more rapid, and $\langle \Delta MI \rangle$ is thus the main factor driving the fixation process. In contrast to the fixed value of ΔMI (Fig. 10f), type 3 does not emerge gradually in the population, and the slope of the fixation curve is steeper. It may be that the mutants categorized in Case 4 spread abruptly in the population.

• Conclusions

At the onset of mutations, most mutations are at least slightly deleterious. Afterward, the gene information is subject to fluctuation, increasing $\langle |\Delta I| \rangle$ and $\langle \Delta MI \rangle$. It is thus probable that most real situations are similar to Case 4 (Fig. 12d), in which $\varepsilon_{two} = \Delta MI$ plays a major role in the evolutionary process. Also in the other three cases, a monotonous increase and a monotonous decrease were observed for $\langle \Delta MI \rangle$ and $\langle r \rangle$, respectively, although with various magnitudes.

In summary, the above simulations demonstrate that the intergenic interactions promote evolution. Conversely, a considerable amount of information acquired during evolution is ascribed to the intergenic MI. Finally, $\langle r \rangle$ continuously decreases in all cases examined, which substantiates the involvement of the attractive force \mathcal{F}_{att} in the evolution of the two-gene system. Conversely, in a multigene network, ΔMI of a particular gene with many other genes sums up to a large amount. Thus, as discussed in the next section, ΔMI of the whole network affects the fitness.

Similarity to physical systems

In addition to the dynamic simulations presented thus far, we consider the case in which protein-coding genes have reached sufficient stable conditions. Many genes that constitute present organisms have already evolved highly, and an amino acid substitution thus takes as long as approximately $1E7$ years per 100 amino acids. If we assume that the numbers of information levels, m and n , and the mean index values of information levels, λ and ν , do not change within this long stable period, then both $H(G_X)$ and $H(G_Y)$ are constant. In this case, the expectation value of the MI of G_X and G_Y increases with time, as expressed in Eq. (210), while that of the information metric r decreases. The influential force thus acts between genes, just as the influential force acts between physical informatoms, and is considered the driving force of biological evolution.

(g) Evolution of a network of two genes

Attractive and repulsive forces

In this section, we will consider the two-gene system. Mutations accumulate over time in the regions of two genes, affecting information and fitness. The genes are then subject to successive cycles of mutation and selection, resulting in either fixation or extinction and ultimately leading to the evolution of each gene. Along with these events, MI for two genes gradually increases while the information metric r decreases, finally facilitating the functional interplay of the two genes. Gene interactions have now evolved fairly well in complex organisms, demonstrating that the attractive force \mathcal{F}_{att} has worked between genes for a long time.

Despite the roles played by \mathcal{F}_{att} , Eqs. (17) and (18) predict that the repulsive force \mathcal{F}_{rep} also acts between genes. From a physiological point of view, an optimum metric r_e should exist for the survival of organisms. For example, if r is too small, then the interference between two genes becomes excessive. In this case, the activity of each gene becomes confined, which would not be beneficial for survival. Selective pressure is therefore expected to act in the direction in which the metric r increases. This indicates the emergence of the repulsive force between genes. Conversely, the observation of a repulsive force implies that the relationship between genes is strong.

Influential force orbital

Through evolution, intergenic interactions evolved, which are detailed, strong, multifaceted, attractive, or repulsive. These interactions can be assigned in the information metric spacetime (r -spacetime). The attractive force acting between genes G_i and G_j is expressed using the Slater-type function of the information metric r as

$$\phi_{ij}(r) = \mathcal{M}(G_i)\mathcal{M}(G_j)\exp[-\alpha_{ij}r(G_i, G_j)] , \quad (212)$$

where α_{ij} is the influential force parameter. This is the same as Eq. (179) for physical systems, except that r is used instead of the information distance $I_o = \alpha x + r/2$ under the limit $\alpha x \rightarrow 0$. When we use the attractive coefficient k_{1ij} and repulsive coefficient k_{2ij} , set the **influential force orbital** to $\chi_{ij}(\mathbf{r}) = (\phi_{ij}(\mathbf{r}), [\phi_{ij}(\mathbf{r})]^2)$, and set the **coefficient vector** to $c_{ij} := (k_{1ij}, -k_{2ij})$, the influential force acting between G_i and G_j is expressed as

$$\varphi(G_i, G_j) = c_{ij} \cdot \chi_{ij}(r) . \quad (213)$$

Again, this is essentially an equal representation for physical informatons written as Eq. (180).

Graph with two nodes

To focus on the informatics aspects of the influential force acting between G_i and G_j , we regard the intergenic interaction as a graph with two nodes and an edge of weight $MI(G_i; G_j)$. We call this interaction a **two-gene network** N_{ij} . We now define $H(N_{ij})$ as the information that characterizes this particular two-gene interaction. We express $H(N_{ij})$ using $MI(G_i; G_j)$ and the signature information $H_{sgn}(G_i, G_j)$ that indicates the direction of the edge between genes:

$$H(N_{ij}) = [MI(G_i; G_j), H_{sgn}(G_i, G_j)] , \quad (214)$$

which represents the union of the two pieces of information. However, inversion in the direction of gene interaction would considerably affect the fitness of the organism. It is therefore natural to suppose that such an inversion occurs infrequently, and thus, $\Delta H(N_{ij}) = \Delta MI(G_i; G_j)$ most of the time.

Instead of the fluctuation model used in the preceding sections, we here use the equation of motion in the information metric spacetime [Eq. (103)]. If t is sufficiently small, then

$$MI(G_1;G_2) \simeq \frac{1}{2} \alpha |E| \sqrt{1 - \frac{2\alpha |E| k_2}{k_1^2 [\mathcal{M}(G_1)^{-1} + \mathcal{M}(G_2)^{-1}]} } t^2 \quad (215)$$

This solution to the equation of motion is similar to MI (in terms of MI_{inh}) calculated using the fluctuation model in that they are quadratic functions of t [Eq. (210)]. From the discussion in the previous section, the advantageous information can be acquired as $\Delta H(N_{ij})$ through the evolution of gene interactions. Again, it is important that $\Delta H(N_{ij}) = \Delta MI(G_i;G_j)$ is larger than $\Delta H(G_1) + \Delta H(G_2)$, which represents the amount of information obtained from the evolution of two individual genes.

(h) Evolution of networks of multiple genes

This section considers a multigene system that is constructed as a one-to-many network with G_i at the center. An extensive network, between G_i at the center and the other genes G_j on the periphery, then evolves with time. In humans, G_j involves approximately 20,000 protein-coding genes. Even a weak relationship is amplified in evolution [58], and there are thus detailed interactions among virtually all genes. Moreover, associations with non-protein-coding genes are constructed similarly to those with protein-coding genes.

LCIO as a genetic orbital

In the multi-gene system around G_i , the influential force of G_i is expressed in a multidimensional information metric spacetime (Fig. 13). This is done by combining the influential force orbitals as

$$\varphi[\vec{r}(G_i)] = \sum_{j \neq i} c_{ij} \chi_{ij}(r) \quad (216)$$

which is the sum of the influential forces that G_i exerts on all other genes. In a biological sense, this represents the extended functions of G_i . We call this φ the **genetic orbital** (GO) of G_i . It is expressed as the **LCIO** between G_i and all other genes, as given by Eq. (181). This formula is analogous to the molecular orbital as the linear combination of atomic orbitals (LCAO) [59]. However, unlike the case for the LCAO method, we must consider the dimensions of the information metric spacetime to illustrate the multi-gene system. This is because genes have diverse structural properties and widely distinct functions. Genes are thus separated from each other by the information metric r in the multidimensional spacetime.

We calculate the dimension M of the information metric spacetime as follows. The number of relative coordinates of N genes in the M -dimensional space is $(N-1)M$. These coordinates must satisfy the $N(N-1)/2$ equations for the information metric r between each pair of genes. Therefore, $(N-1)M \geq N(N-1)/2$ and $M = \lceil (N+1)/2 \rceil$, where N is the total number of genes and $\lceil x \rceil$ is the Gaussian notation for the largest integer that does not exceed x . Thus, when we consider $N = 2E4$, we have $M = 1E4$.

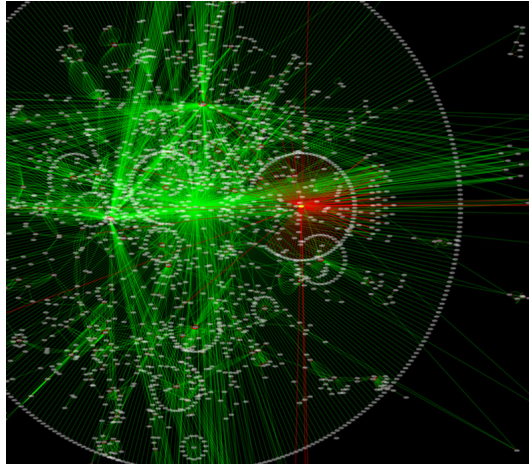


Figure 13 | Intricate network of genes. This schema represents complex network interactions between multiple genes (provided by Dr. Keiichiro Ono, <http://www.cytoscape.org>). On the basis of LCIO, each gene exerts highly evolved network functions through its effect on many other genes.

Molecular information and network information

To decipher the informatic implications of the LCIO, we consider a gene network N_i of the gene G_i . N_i is constructed for the central node G_i and peripheral nodes G_j as the summation of each two-gene network N_{ij} . N_{ij} is characterized by the two-gene network information $H(N_{ij})$, and $H(N_i)$ is thus expressed as the superposition of each $H(N_{ij})$ in a multidimensional spacetime as

$$H(N_i) = \sum_{j \neq i} H(N_{ij}) \quad (217)$$

This is an equivalent derivative of the LCIO in the multidimensional information metric spacetime. However, the current paradigm of gene information is that it encodes the structure of gene products (e.g., proteins, miRNAs, and lncRNAs). Therefore, the total information of gene G_i , $H(G_i)$, is the union of molecular information $H(M_i)$ expressed by the **molecular structure** of the gene product and the network information $H(N_i)$ represented by the **network structure** of the LCIO:

$$H(G_i) = H(M_i, N_i) \quad (218)$$

Biological implications of network information

Equation (218) states that each gene encodes not only the molecular structure but also the network structure that represents mutual relationships with other genes. The network based on the LCIO assembles all the gene nodes around G_i using the edges of weight $MI(G_i; G_j)$ (Fig. 1e) and thereby endows G_i with the following new functions. First, G_i affects the other $N - 1$ genes ($N \geq 2E4$), making

it possible to use them appropriately as tools with which to realize the physiological functions of G_i itself. Next, the $N - 1$ genes concertedly exert their influential force on G_i , allowing the fine tuning of the G_i functions according to the intracellular and extracellular conditions. This also provides a molecular basis underlying the information processing machinery required to establish complex multicellular organisms. Such highly sophisticated functions are possible only because the gene network has a very high density; that is, $H(N_i)$ is immense.

Quantity of network information

With G_i at the center, $H(N_i)$ is the information of the union of the entire two-gene network N_{ij} . However, we must keep in mind that $H(N_i)$ is not the simple sum of each $H(N_{ij})$ but also depends on the network structure in the multidimensional space. To determine $H(N_i)$, we consider the projection of the network structure to three-dimensional space. Here, we must take account of the differences among individual genes, in contrast to the case of physical systems in which the distinction among bodies is not necessary. According to Bianconi [60], $H(N_i)$ is calculated by applying statistical mechanics to the network ensemble. While $H(N_i)$ takes the maximum value $\max H(N_i)$ when all edges are independent, we suppose that $H(N_i)$ is basically equal to $\max H(N_i)$ because genes are distinct from each other. $H(N_i)$ is then the sum of each information of edge weights (Σ^W), directions (Σ^{sgn}), and ranks (Σ^{rank}):

$$H(N_i) = \Sigma^W + \Sigma^{sgn} + \Sigma^{rank} . \quad (219)$$

To calculate $H(N_i)$, we deal with a one-to-many network with G_i at the center. We then have

$$\begin{aligned} \Sigma^W &= \log \left(\frac{N-1+S}{N-1} \right), \quad \text{where } S = \sum_{j \neq i} MI(G_i; G_j) \\ \Sigma^{sgn} &= \frac{1}{N} \log 2^{\frac{N(N-1)}{2}}, \quad \Sigma^{rank} = \log[(N-1)!] . \end{aligned} \quad (220)$$

As described below, we investigate how $H(N_i) = \Sigma^W + \Sigma^{sgn} + \Sigma^{rank}$ changes with time. First, however, we assume that Σ^{sgn} derived from the edge directions does not change. This is because an inversion of interaction direction will cause a significant fitness effect and is supposed to be rare.

Σ^W : network information originating from edge weights

Σ^W originates from all edge weights around G_i . The increase in Σ^W is expressed by the change in $S = \sum_{j \neq i} MI(G_i; G_j)$, from S to $S + (N - 1)\Delta MI$, where ΔMI is the average increase in edge weights. When each $MI(G_i; G_j)$ increases with time, $\Delta \Sigma^W$ as the total sum of $\Delta MI(G_i; G_j)$ is approximated as

$$\Delta \Sigma^W \simeq \left[(N-1) \log \left(\frac{N-1}{S} + 1 \right) \right] \Delta MI \quad (221)$$

(Appendix 15). As an example, we now consider a case in which $N = 2.0E4$, $S = 200$ nat, and each $MI(G_i;G_j)$ increases by $\Delta MI = 1.0E-4$ or $1.0E-3$ on average. This value of S is a typical value estimated from the *ab initio* GO analysis of *CLSTN3* mentioned later. Σ^W then increases respectively as $\Delta\Sigma^W = 9.23$ and 92.3 in the entire network; both these values are 92,300 times ΔMI . If we assume that the increases in MI act additively on the fitness and the summed fitness value [= $\varepsilon_w(N_i)$] increases by 1% of $\Delta\Sigma^W$, then $\varepsilon_w(N_i)$ respectively becomes 0.0923 and 0.923, which are 923 times ΔMI . Such large fitness values are generated by the sum of epistatic effects of each $\varepsilon_{two} = \Delta MI(G_i;G_j)$.

With respect to the fixation of a new mutant of a single gene G_i , $\varepsilon_w(N_i) = 0.01 \times \Delta\Sigma^W$ is expected to have a large compensation effect, even if this mutant has a deleterious effect. To examine the effect of $\Delta\Sigma^W$, we here assume the same conditions as those for Fig. 10, except that we consider only one of the two genes as G_i . The time to fixation of a neutral mutant then becomes as short as 629 years (or 15,700 years for humans whose generation takes 25 years) (Table 2). This is in clear contrast to 4.80E6 years in the case without this effect, which indicates that the fixation is remarkably accelerated by an increase in $\varepsilon_w(N_i)$. Even when the mutation is strongly deleterious such that $s_1 = -0.05$ (i.e., $N_e s = -5000$), the fixation time with $\varepsilon_w(N_i)$ is only 1310 years (32,700 years for humans).

In accordance with the acceleration of fixation, the relative fixation probability of the neutral mutant with $\varepsilon_w(N_i)$ is 9230 times that without $\varepsilon_w(N_i)$. Even if the mutation is strongly deleterious as above ($s_1 = -0.05$), the massive effect of $\varepsilon_w(N_i)$ largely precludes harm. It increases the fixation probability to a level 4230 times that of the control (i.e., $s_1 = \varepsilon_w(N_i) = 0$). Finally, it is convincing that $\varepsilon_w(N_i)$ occupies much of the advantageous information that G_i acquires through evolution, because $\varepsilon_w(N_i)$ is much larger than $\varepsilon_{two} = \Delta MI$, let alone $|s_1| = |\Delta I|$, a selection coefficient of the G_i mutant itself. In conclusion, Σ^W greatly affects the evolution of a gene in terms of probability and rapidity.

Table 2. Effects of epistasis on fixation

Gene fitness	Epistatic fitness	Total fitness	Time to fixation (year)	Relative fixation probability
0	0	0	4.80 E6	1
s_1^*	0	-0.05	not fixed	0
0	ε_{two}^{**}	0.0001	5.30 E5	10
s_1	ε_{two}	-0.0499	not fixed	0
0	$\varepsilon_w(N_i)^{***}$	0.0923	629	9,230
s_1	$\varepsilon_w(N_i)$	0.0423	1,310	4,230

* $s_1 = -0.05$, ** $\varepsilon_{two} = 0.0001$, *** $\varepsilon_w(N_i) = 0.0923$, $u_1 = 1E-6$, $N = 1E5$.

Σ^{rank} : network information generated by gene rank

Σ^{rank} does not change over time in contrast to Σ^W , which increases with time. This is because Σ^{rank} arises from the permutation of the magnitude rank of $MI(G_i; G_j)$; that is, $\Sigma^{rank} = \log[(N-1)!]$, which is constant over time. However, for the following reasons, if there are changes in the rank order of $MI(G_i; G_j)$, there will be qualitative changes in the network functions of G_i , which is located at the center of the network. Given that different peripheral genes G_j have distinct structures and functions, the information shared with the central gene G_i varies from G_j to G_j . As a result, a rank order arises spontaneously in the magnitude of $MI(G_i; G_j)$. It is notable that $MI(G_i; G_j)$ indicates the functional proximity of G_i and G_j (Appendix 16). Thus, the rank order of $MI(G_i; G_j)$ determines what gene G_j preferentially interacts with G_i , thereby qualitatively defining the network functions of G_i .

We now investigate how network functions undergo qualitative evolution through changes in the rank order of $MI(G_i; G_j)$. For this purpose, we calculate the increase in the fitness associated with such events. We consider a case in which the weight of a particular edge around G_i increases more rapidly than the weights of other edges, and the rank of its weight thereby rises among all edges. The edge weights $MI(G_i; G_j)$ follow an infocanonical distribution [Eq. (17)] (Appendices 2 and 17). Hence, we set $MI(G_i; G_j) = -A \times \log(j) + B$ (e.g., $A = 1.032$ and $B = 14.05$ in *CLSTN3* described later) for the j -th largest MI. Then, when $MI(G_i; G_j)$ increases by ΔMI , the average rise λ_r in its rank is expressed as

$$\lambda_r \simeq j \frac{\Delta MI}{A} . \quad (222)$$

We here consider a situation in which the increase in MI per unit time increases from $\Delta MI_0 (\simeq 0)$ to ΔMI and $|\Delta MI| \ll 1$. If we initially assume that all newly generated information is associated with the positive fitness, then the increase in the fitness value [=: $\varepsilon_{rank}(N_i)$] is calculated as

$$\varepsilon_{rank}(N_i) \simeq \frac{j}{A} (\Delta MI - \Delta MI_0) \simeq \frac{j}{A} \Delta MI \quad (223)$$

(Appendix 27). This implies that the edge rank, in addition to the edge weight, affects the fitness. When the change in MI per unit time increases from $\Delta MI_0 \simeq 0$ to $\Delta MI = 1.0E-4$, the average increase in the rank is calculated as $\lambda_r = 0.969$. If we assume that a 1% fraction of ΔMI is reflected in Eq. (223), then the increase in the fitness is $\varepsilon_{rank}(N_i) = 0.00969$. Moreover, considering the contributions of both genes G_i and G_j , the expectation value of the increase is $\varepsilon_{rank} = \varepsilon_{rank}(N_i) + \varepsilon_{rank}(N_j) = 2\varepsilon_{rank}(N_i) = 0.0194$, which is 194 times that for $\varepsilon_{two} = \Delta MI$.

The above increase in $\varepsilon_{rank}(N_i)$ means the rise in importance of the role that G_i plays through the interaction with G_j , implying a qualitative change in the network function of G_i . Given that selective pressure favors intergenic interactions that are advantageous to the organism, a mutant with this new preferred network function spreads in the population rapidly. Thus, changes appear in the network functions of genes, leading to the qualitative evolution of the genes.

Evolution driven by network information

The preceding observations revealed that the network information $H(N_i)$ involving Σ^W and Σ^{rank} greatly affects the evolution of gene G_i . While $\varepsilon_w(N_i)$ affects both the probability and rapidity of the fixation process of new genes, $\varepsilon_{rank}(N_i)$ influences functional aspects of the gene network. Their multiplier coefficients are respectively 923 and 194 times those for $\varepsilon_{two} = \Delta MI$. These effects of $H(N_i)$ on fitness are so strong that the biological effects of the mutations in non-coding DNA regions may also become apparent. If this is the case, the mutations in the long non-coding regions will increase the effective rate of information change by orders of magnitude. Taken together, even though many mutations in protein sequences are deleterious, changes in network information $H(N_i)$ provide opportunities for creating advantageous information to more than compensate for the harm, which may facilitate the phenotypic evolution of organisms.

Particularity of network information

In evolution, both the rank orders and direction of edges $MI(G_i; G_j)$ are generated quite spontaneously. This is because there is originally no preference in the direction of mutation. However, despite this natural spontaneity, the generated network information $H(N_i)$ is large enough to provide a basis for the vast variations and profoundness of gene functions. As discussed earlier, this is a situation analogous to the generation of MI in physical systems, where massive information is produced with the multi-body system. Nevertheless, in contrast to the physical system made up of homogeneous particles, there is remarkable heterogeneity in the gene nodes of the network. Accordingly, evolution allows the development of a highly complex network of genes, which is a prerequisite for the establishment of sophisticated phenotypes. These properties highlight the particularity of the biological information and illustrate the need for an in-depth inquiry into the network architecture. Thus, we next examine the qualitative and quantitative properties of the gene network structure.

Rapid evolution implied by the Pareto distribution of $MI(G_i; G_j)$

From the above consideration, the probability distribution of $MI(G_i; G_j)$ is supposed to have an intimate relationship with the properties of the gene G_i . In some cases, mutations in G_i will affect the distribution of $MI(G_i; G_j)$. In particular, if G_i has a strong influential force \mathcal{F} and is evolving rapidly, then mutations in G_i can affect most of the relationships with the other genes. For example, if such G_i acquires new functions through evolution, then there will be changes in the probability distribution of the observable information level X of G_i . This will cause shifts in the distribution of micromutual information MI_{kl} of G_i and G_j , which eventually affects the probability distribution of $MI(G_i; G_j)$. Such scenarios apply when a new phosphorylation site arises in the encoded protein and in another case where a mutation occurs in the protein interaction domain.

We presume that genes initially have an infocanonical distribution in $MI(G_i; G_j)$ under sufficiently stable evolution (Appendices 2 and 17). However, in the cases above, we propose that $MI(G_i; G_j)$ undergoes a transition from the original infocanonical distribution and approaches the Pareto distribution (Appendix 28). It is thus expected that the genes following the Pareto distribution are rapidly evolving and that their influential forces are strong. Indeed, the results of *ab initio* GO analysis in the next chapter fairly meet our expectations; that is, while most genes have an infocanonical distribution of $MI(G_i; G_j)$, most genes in the immune system have a Pareto distribution (to be published elsewhere). It is known that genes of the immune system have a potent biological activity and a high evolution rate, which conforms to the above predictions. We present examples in the next chapter.

Large amount of network information

As noted above, the network information $H(N_i)$ constitutes a large part of the information that a gene G_i acquires through evolution. G_i is then supposed to express most of its information as $H(N_i)$ but not as the molecular information $H(M_i)$. We here compare the maximum possible magnitude between $H(M_i)$ and $H(N_i)$. For $H(M_i)$, we consider a protein of standard molecular weight with 500 amino acids. The maximum number of states $W(M_i)$ expressed by the permutation of the amino acids is then $20^{500} = 3.27 \times 10^{650}$, which is equivalent to $H(M_i) = 2161$ bit.

Meanwhile, for $H(N_i)$, we consider Σ^{rank} and Σ^{sgn} as components of $H(N_i) = \Sigma^{rank} + \Sigma^W + \Sigma^{sgn}$, and then $H(N_i) > \Sigma^{rank} + \Sigma^{sgn}$. If we set $N = 20,000$, then the maximum number of states of the network is $W(N_i) > \exp(\Sigma^{rank}) \times \exp(\Sigma^{sgn}) = (8.77 \times 10^{84,580}) \times (2^{9999.5}) = 1.24 \times 10^{87,591}$, where $\Sigma^{rank} = 281$ kbit and $\Sigma^{sgn} = 10$ kbit. A comparison of the two types of information gives $W(N_i)/W(M_i) > 3.78 \times 10^{86,940}$, implying more than a 289-kbit difference between $H(N_i)$ and $H(M_i)$. This large difference indicates that even if we ignore Σ^W , $H(N_i)$ is by far larger than $H(M_i)$. It is thus suggested that the major part of the gene information is encoded as the network structure and not as the molecular structure.

In the *ab initio* GO method described in the next chapter, we choose the related genes G_j whose ranks of $MI(G_i; G_j)$ are within the top 10% and then calculate the relevant pathways of G_i and direction of each edge. As detailed there, the functions of G_i deduced from this calculation provide good results, which supports the importance of $H(N_i)$. However, to reduce the calculation burden, we use the combined information as a substitute for $\exp(\Sigma^{rank})$ because the combination reflects the gene rank order. At this time, $W(N_i) > C(20,000, 8000) \times (2^{8,000}) = 5.07 \times 10^{8251}$ (27.4 kbit), which is still far larger than $W(M_i) = 3.27 \times 10^{650}$, despite the reduction in the calculation cost.

Non-coding regions and network information

At the end of this chapter, we consider the effect of non-coding chromosomal regions on network information. It is known that the complexity of higher organisms correlates with the relative amount of non-coding DNA sequences in the genome [61]. This suggests that mutations in the non-coding

regions have led to the evolution of $H(N_i)$, making it possible to express complex phenotypes. When mutations arise in the non-coding regions, they do not change the structural properties of the gene product but can affect the spatiotemporal regulation of gene expression.

The changes induced in gene expression profiles then cause fluctuations in the microstate probability of intergenic MI, resulting in the alteration of each $MI(G_i; G_j)$ and the overall network structure; that is, all of Σ^{ank} , Σ^W , and Σ^{sgn} may change. Because $H(N_i) \gg H(M_i)$, the evolution of the gene network function, $\Delta H(N_i)$, greatly exceeds that of the molecular function remaining at $\Delta H(M_i) = 0$. As a result, $H(N_i)$ gains considerable variability in both network architecture and activity and there is thus an opportunity for creating the complexity of higher organisms. Together with an inherently large $H(N_i)$, the above inference may offer a further plausible explanation for why the phenotypic evolution is occasionally much more rapid than the gene evolution.

Chapter summary

This chapter showed that the influential force \mathcal{F} acts between genes and is expressed by the same mathematical formula as is the influential force in the physical system. The force arises from $MI(G_i; G_j)$ by gene interaction and represents the relative fitness. In a multi-gene system, the force provides power to construct the LCIO, which has an immense magnitude of information encoding the precise structure of the gene network. Unlike physical systems, the gene nodes in the network have highly divergent properties, such that the network functions $H(N_i)$ of each gene G_i vary enormously.

The findings obtained in this chapter demonstrate that the network function constitutes the main body of the gene information. This leads to the striking conclusion that

$$H(N_i) \gg H(M_i) , \quad (224)$$

which indicates that the gene functions based on the network structure are far more essential than those based on the molecular structure. Especially in higher organisms, genes primarily encode the intergenic MI presumably based on information in untranslated regions, thereby determining the network structure. This concept rewrites the traditional paradigm based on the Watson–Click model, which claims that genes encode the amino acid sequence based on exon sequences, thereby determining protein structure.

In turn, the above conclusion will drastically change our perception of both biology and medicine. We should be aware of the priority of $H(N_i)$ when investigating gene properties and their relevance to pathophysiology. Focusing on the LCIO-based network structure will allow the determination of the gene function of interest without experimentation. This method could also be applied to genes that do not encode proteins. In the next chapter, we will present examples of how modern medicine can be advanced by analyzing the LCIO-based gene network information.

4. *Ab initio* GO method

(a) Basic theory of the *ab initio* GO method

Medical science has made much progress owing to the availability of massive information about the human genome. However, there remain many diseases without known etiology or effective therapies, indicating that our knowledge about the genome is still far from complete. Medicine can be advanced by the functional characterization of all genes, which will broaden the perspective of pathophysiology and help identify novel disease-associated genes, diagnostic markers, and molecular targets for therapy. For advancement in this direction, we employ a new strategy to reveal the network functions of genes, which we have proposed to be encoded by the LCIO.

Here, to advance medicine and demonstrate that genes perform functions based on the LCIO, we developed an informatics program named the *ab initio* GO method. Using The Cancer Genome Atlas (TCGA), this method calculates the function of a gene of interest within minutes. Additionally, we developed another program, STAIC (a Strategic Tool for *Ab Initio* Identification of Cancer Genes), to facilitate the discovery of cancer genes. Combining these two methods, we identified a *potential immune checkpoint, KYNU/kynureninase*. While traditional techniques of molecular biology have never elucidated this function, our informatics demonstrated the possible involvement of *KYNU* in cancer immunity. Thus, with rapidity and accuracy, the *ab initio* GO method could accelerate the functional analysis of the human genome and advance medical science.

LCIO in cancer cells

In Chapter 3, we discussed the organismal level of gene network evolution. The formation of intergenic MI is accompanied by the establishment of the LCIO, which codes for the structures and functions of gene networks. Meanwhile, genes are also subject to evolution by natural selection in carcinogenesis, like in gene evolution in organisms. Therefore, if the development of the LCIO constitutes a key feature of the gene evolution at the organismal level, then the LCIO would also develop through the evolutionary process of genes in cancer. Furthermore, given that the LCIO substantially contributes to gene functions in normal cells, it would also play a pivotal role in the gene activities in cancer cells.

Of note, cancer cells usually use the gene functions of their normal cell counterparts [62]. Therefore, conversely, we can deduce many genes' actions in normal cells from those in cancer cells. Indeed, this strategy has been widely applied to experimentally elucidate the functions of many genes in normal cells. However, although it has been a standard procedure, the experimental approach has

two significant drawbacks. First, as expressed by $H(N_i) \gg H(M_i)$ [Eq. (224)], molecular functions constitute only a small part of the total gene activity. Hence, whereas ordinary ‘wet’ experiments clarify the molecular functions of genes, the obtained knowledge may be insufficient to understand the whole picture of the physiological role of the genes. Second, the experimental process takes a long time to uncover the properties of a gene of interest; that is, it usually takes more than 10 years and therefore requires multidirectional, large-scale investments.

To evade the two drawbacks mentioned above, we developed an informatics algorithm that rapidly calculates gene functions without performing experiments. In principle, this algorithm deciphers the LCIO in cancer cells, whereby we can deduce the network functions of a particular gene in normal cells. This procedure is designated as the *ab initio* GO method because its rationale is analogous to that of the *ab initio* molecular orbital method. Our method computes the intergenic MI at a super-high resolution and thereby predicts a wide range of gene properties, ranging from the molecular function to the pathophysiological relevance.

Generation and amplification of intergenic MI

In the process of cancer evolution, various changes (e.g., mutations and epimutations) arise in numerous genes. As a result, intergenic MI is generated globally. Within the many gene alterations, there are recurrent patterns characteristic to each gene, each case, and each type of tumor. This tendency is also true for the co-occurrence of gene alterations, which reflects the intergenic MI. Therefore, by analyzing the relationships between various changes in many genes, in many cases for multiple kinds of tumor, we can explore MI shared between virtually all the genes with all other genes. In addition, much of the intergenic MI in cancer cells is related to that in normal cells because, as noted above, the functions of the corresponding genes are similar. We can thus deduce much of the intergenic MI in normal cells.

In addition to the global generation of intergenic MI, the *ab initio* GO method uses the property that the evolutionary process of genes in cancer is analogous to that in organisms. From the beginning, all the genetic information (e.g., genes and intergenic MI) in organisms has evolved through effects on the fitness of the organisms. Similarly, in cancer evolution, the alteration of genetic information [e.g., the (epi)mutations and altered intergenic MI] also affects the fitness of the generating cancer cells. The cells with changed fitness are subject to natural selection, thereby eliciting the fluctuation in the frequency of the altered information within the cell population.

(b) Super-high resolution analysis of intergenic interactions

Detection of intergenic MI using a contingency table

As the prevailing carcinogenesis model, we again adopt the Moran process, which explains genotype amplification. First, we consider the case of single-gene evolution at an organismal level. Even if the fitness advantage is as small as $s = 1E-8$ to $1E-7$, this effect is amplified through generations, resulting in a dramatic increase in its genotype frequency. Second, in the case of single-gene evolution in cancer, the fitness effect is again unambiguously appreciable. Finally, when two genes G_i and G_j are subject to alterations in cancer cells, $MI_{inh}(G_i, G_j)$ arises because of the fluctuation in information for both genes [Eq. (210)]. An influential force then arises between the genes and affects the evolutionary fitness of the cells: $\mathcal{F}_{att} = \exp[MI_{inh}(G_i, G_j)] \approx 1 + \varepsilon_{two}$ [Eq. (202)]. As discussed in Chapter 3, this epistatic fitness ε_{two} increases the frequency of the double mutant.

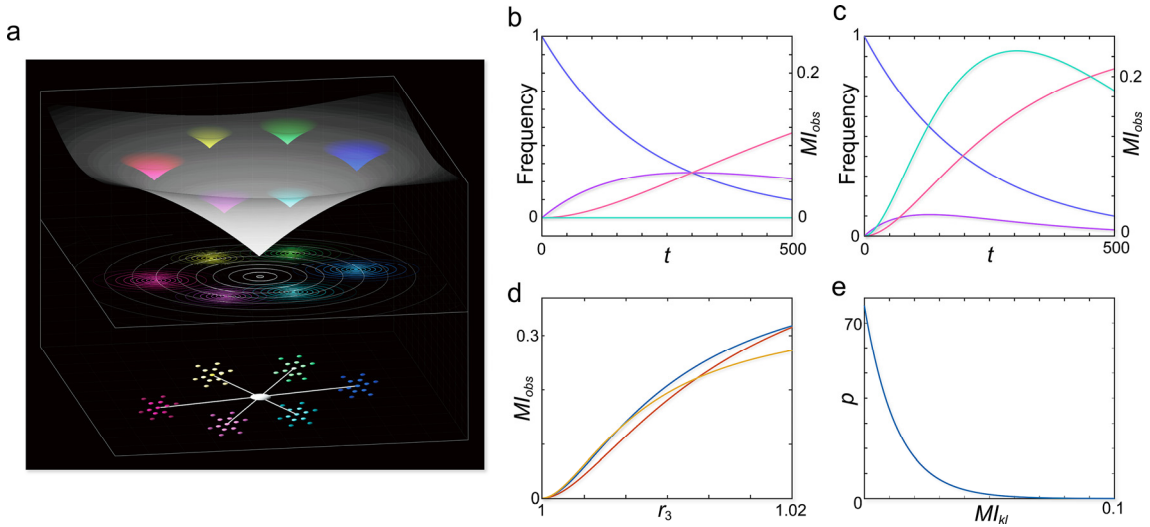


Figure 14 | Theory of the *ab initio* GO method. **a**, Network of genes around G_i . The bottom plane represents the gene network around G_i in the information metric space. Six clusters represent a group of genes with related functions. The large surface expresses the influential force potential, and its vertex corresponds to G_i . The clusters are represented as six small surfaces on the large surface, and their vertices express the centers of information mass of the clusters. **b**, **c**, Change in genotype frequencies in the Moran process. Blue, purple, and red lines represent X_0 , $X_1(= X_2)$, and X_3 , the frequencies of types 0–3 in cancer samples at time t , respectively, and the green line represents MI_{obs} . $N = 1000$, $u_1-u_4 = 1E-3$, $r_0-r_2 = 1$ and $r_3 = 1$ (**b**) or 1.01 (**c**). MI_{obs} remains zero when $r_3 = 1$. **d**, Effect of r_3 on MI_{obs} in the Moran process. MI_{obs} increases as r_3 increases. t is 200 (red line), 300 (blue) or 400 (yellow). **e**, Infocanonical distribution of micromutual information MI_{kl} .

Table 3. Contingency tables of genotype frequency

Table 3A

	A	A'	
B	X_0	X_1	X_0+X_1
B'	X_2	X_3	X_2+X_3
	X_0+X_2	X_1+X_3	1

Table 3B

	A	A'	
B	NX_0	NX_1	$N(X_0+X_1)$
B'	NX_2	NX_3	$N(X_2+X_3)$
	$N(X_0+X_2)$	$N(X_1+X_3)$	N

In a multigene system, $MI_{inh}(G_i;G_j)$ as the epistatic fitness value differs for each combination of G_i and G_j (Fig. 14a). The differences in the fitness effect are then amplified through the cycles of cell division, which results in a shift in the allele frequency distribution. This phenomenon can be presented in a contingency table as a deviation in the frequency of genotype combinations (Fig. 14b, c and Table 3), which is measured by the increase in MI_{obs} . Conversely, by detecting the frequency deviation as MI_{obs} in cancer, we get a clue for estimating intergenic interactions as MI_{inh} in normal cells [Eq. (206)].

For example, we consider two genes, A and B, and let them take two different states, A / A' and B / B', respectively. Table 3A gives the relative frequency of the combination of gene states. Here, X_0 , X_1 , X_2 , and X_3 are respectively the proportions of combinations AB, A'B, AB', and A'B'. Table 3B gives the observable number of cases. Using this frequency distribution, we can calculate the one-dimensional MI_{obs} for A and B as

$$\begin{aligned}
 MI_{obs} = & X_0 \log \frac{X_0}{(X_0+X_2)(X_0+X_1)} + X_1 \log \frac{X_1}{(X_1+X_3)(X_0+X_1)} \\
 & + X_2 \log \frac{X_2}{(X_0+X_2)(X_2+X_3)} + X_3 \log \frac{X_3}{(X_1+X_3)(X_2+X_3)} . \quad (225)
 \end{aligned}$$

Furthermore, in the latter part of this section, we describe a procedure for calculating extremely high-dimensional MI_{obs} , which is an advanced methodology based on a series of novel mathematical theorems (Appendices 18–20).

Similarity of networks between normal and cancer cells

Before proceeding, we establish two conjectures about the LCIO-based gene network. Both assumptions derive from empirical observations that cancer gene networks are, on average, similar to the gene networks of normal cell counterparts. The first conjecture is that the $MI(G_i;G_j)$ distribution around each gene G_i is fixed as either an infocanonical or Pareto distribution and that the distribution parameters do not notably change upon transformation from normal cells to cancer cells. It is known that few gene alterations are uniformly common to many cases in various types of cancer. Therefore, when assessed using many datasets with more cancer data, the distribution of $MI(G_i;G_j)$ would become

stable on average and approach that in normal cells. The second conjecture is that, upon transformation, there is little change in the rank order of the magnitude of $MI(G_i;G_j)$ around G_i . This supposition is made because the realization probability of changes in the rank order follows a Poisson distribution and is thus expected to be sufficiently small on average (Appendix 27).

If the above two conjectures are correct, then both $MI_{can}(G_i;G_j)$ in cancer cells and $MI_{norm}(G_i;G_j)$ in normal cells follow either an infocanonical or Pareto distribution, and

$$MI_{norm}(G_i;G_j) \simeq \frac{\langle MI_{norm}(G_i;G_j) \rangle}{\langle MI_{can}(G_i;G_j) \rangle} MI_{can}(G_i;G_j), \quad (226)$$

where $\langle MI_{norm}(G_i;G_j) \rangle$ and $\langle MI_{can}(G_i;G_j) \rangle$ are respectively the mean values of $MI(G_i;G_j)$ in normal cells and cancer cells. Therefore, if we successfully analyze the LCIO in the cancer cells, we can better understand the LCIO in the normal cells. Furthermore, even if the data in the TCGA are heterogeneous in terms of the sampling conditions, the numbers of cases and types of cancer are sufficiently large that the precision of the estimation of each $MI(G_i;G_j)$ would be guaranteed by the multivariate central limit theorem (Appendix 29). In conclusion, when large-scale data of many kinds of cancer are available, the above conjectures will ultimately become true on average, which would provide a wealth of information about the LCIO in normal cells. Therefore, we will next focus on the development of strategies that allow the management of massive data.

Evolutionary microscope

To obtain the highest possible resolution of the structures of the LCIO-based gene networks, we exploit the dramatic amplification effect of MI_{inh} through evolution. MI_{obs} is a multivariate function of fitness (r_1-r_3), time t , mutation rates (u_1-u_4), and population size N [Eqs. (201) and (205)]. If $u_1 \leq u_4$ and $u_2 \leq u_3$, then MI_{obs} mainly depends on r_3 , the relative fitness of the double mutant. MI_{obs} is a monotonously increasing function of $r_3 \simeq \mathcal{F}_{att} = \exp(MI_{inh})$, even in carcinogenesis (Fig. 14d, Appendix 13). Below are the examples that illustrate the dominant effect of r_3 on MI_{obs} .

• Case 1: increases in mutation rates

As an example of the strong dependence of MI_{obs} on r_3 , we here discuss a single type of cancer, and set $N = 1000$, $u_1-u_4 = 1E-3$, $r_0-r_2 = 1.0$, and $\max(t) = 200$. Initially, we examine a two-gene system in which both rates of reciprocal mutations (i.e., u_3 and u_4) increase by 1%. The transition rates d and e then increase by 1.7% from $2.31E-3$ to $2.35E-3$. However, MI_{obs} is as small as $7.83E-4$ nat, and the obtained p -value is only $1E-0.34$ even at $t = 200$. The increases in mutation rates thus do not appreciably enhance the intergenic MI .

• Case 2: increase in fitness

We next consider the increase in the fitness of the double mutant instead of increases in the mutation rates. If r_3 increases by 1% from 1.00 to 1.01, then $d = e = 1.20\text{E-}2$, which increases the transition rates by 419%. Moreover, when $r_3 = 1.01$, MI_{obs} increases with time and reaches 0.193 nat at $t = 200$, which is 246 times higher than that in Case 1 (Fig. 14d). In addition, the p -value becomes $1\text{E-}84$, which is the 246th power of that in Case 1. These observations indicate that a 1% difference in the fitness values can be discriminated at extremely high sensitivity and significance levels. Thus, the intergenic interactions can be detected sensitively on the basis of the amplification of the fitness effect that arises from evolution. We refer to this amplification effect as the *evolutionary microscope*.

Meta-analysis via a novel theorem

In the past, conventional information theory did not consider the application of meta-analysis to the calculation of MI. The inability of meta-analysis constituted a limiting factor for the precision and accuracy of the MI analysis. To overcome this issue, we have developed a novel theorem named the “equivalence principle of information and probability”, detailed in the next section and Appendix 20. This theorem asserts that MI can be calculated from the p -value of Fisher’s exact test. Therefore, MI can be computed from a meta-analysis of Fisher’s exact test using many large-scale data, allowing the quantitative estimation of MI at an extremely high significance level.

• Case 3: conducting a meta-analysis

To conduct a meta-analysis [63], we next explore 30 types of cancer dataset with the same sample size simultaneously, instead of considering only one type of cancer dataset. When r_3 increases by 1% from 1.00 to 1.01 as in Case 2, MI_{obs} remains the same, 0.193 nat. This invariance shows that the sample size does not affect MI , which conforms to its well-known quantitative nature. However, the p -value becomes $1\text{E-}2520$, which is as much as the 30th power of that in Case 2. Such a drastic decrease in the p -value demonstrates that if we perform a **meta-analysis** using many types of cancer dataset, we can improve the accuracy of the evolutionary microscope even more.

When we combine the two principles mentioned above, namely the evolutionary microscope and the meta-analysis, the magnitude of Δr_3 that can be discriminated at a significance level of $p = 0.05$ is $5.17\text{E-}6$, which is remarkably small. Even when we assume $\Delta r_3 = 0.01 \times MI_{inh}$, ΔMI_{inh} that can be distinguished is as small as $5.17\text{E-}4$. Moreover, the meta-analysis has the merit of assessing the LCIO without being disturbed by the noise that accompanies the individual properties of cancer.

Integration of the Moran process and branching process

The initiation of cancer divides tumor development into two phases: the pre-initiation phase and the post-initiation phase [64,65]. The Moran process presupposes a constant cell number and is used to model the pre-initiation phase. By contrast, the branching process is chosen to model the post-initiation phase in which the cell number increases [64,66]. In the branching process, the cell with fitness $1 + s$ divides into two daughter cells with probability $(1 + s)/2$ and dies with probability $(1 - s)/2$. It is assumed that a mutation occurs in the mitosis with the probability u . It is also assumed that the fitness increases as the cell has more mutations.

Whereas the Moran model of the pre-initiation phase underlies the LCIO-based network model described in this paper, the data in the TCGA derives from the cancer samples obtained in the post-initiation phase. Therefore, in estimating the LCIO, we should consider the possibility of the deformation of the network structure, which might accompany the phase transition. However, because the increase in the fitness due to the mutations in the post-initiation phase is sufficiently large, the genotype frequencies used for the Moran model remain virtually constant (Appendix 30). In other words, the fitness values of gene alterations established in the pre-initiation phase are so small that there is no interference with the large fitness values arising after the initiation. Therefore, the Moran model can also be applied to the genetic data even after cancer initiation. Thus, we can use the TCGA as a source with which to analyze the LCIO-based gene network.

(c) Frontier information theory

Immensely high-dimensional intergenic MI

The gene information is an extremely high-dimensional random variable. Therefore, MI of genes becomes an *immensely high-dimensional random variable*. This remarkable complexity poses a serious problem when we wish to estimate MI of genes because we cannot perfectly measure the probability of all combinations of the information states of the genes. To resolve this issue, we developed a calculation strategy that generally allows us to compute the highly multi-dimensional MI of two random variables.

The computing methodology employed is based on two principles detailed later. In short, 1) the micromutual information MI_{kl} follows an infocanonical distribution when the number of dimensions is sufficiently large, and 2) the information exchange mainly occurs in a particular state where MI_{kl} takes the maximum value. In addition, a prerequisite for the calculation scheme is that the specific condition noted in 2) is experimentally observable. This precondition assumes that the particular state producing the highest MI_{kl} value occurs at a considerable frequency and can thus be readily measured.

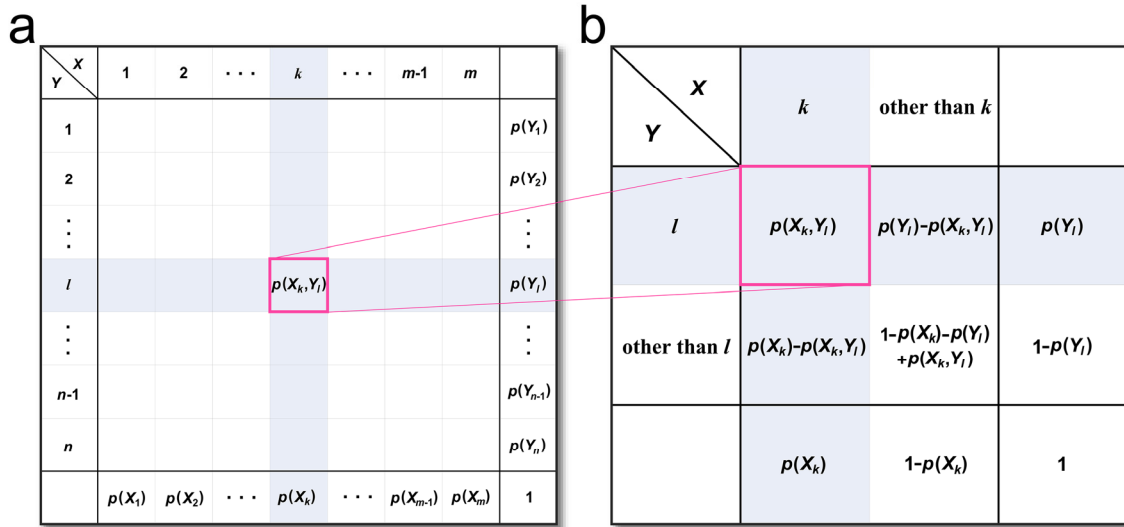


Figure 15 | Transformation from an $m \times n$ contingency table into a 2×2 contingency table. a, $m \times n$ contingency table with respect to two random variables X and Y . b, 2×2 contingency table concerning the cell (k, l) in a.

In addition, the measurement method is expected to allow a wide range of observations that provides the most valuable data related to gene interactions. If these prerequisites are satisfied, we can adopt the highest MI_{kl} obtained as the representative value of the intergenic relationship.

The TCGA is a comprehensive gene database for multiple types of cancer. It provides a broad collection of cancer information (e.g., mutations, expression levels of mRNA and proteins, DNA copy numbers, DNA methylation, and clinical outcomes) for many kinds of tumor. The TCGA contains a wide variety of data, and it is thus expected to provide sufficient information with which to analyze the high-dimensional MI of genes and be an ideal source for the computation. We therefore designed the *ab initio* GO method to incorporate multiple types of data from the TCGA as different dimensions of information. In this way, the *ab initio* GO method integrates large-scale, multi-dimensional genomic data and calculates MI as a unified indicator of gene interaction.

Lowering the number of dimensions using a linear indexing method

We consider two genes G_X and G_Y , which take multi-dimensional states X and Y , respectively. To the collected data representing these variables, we apply a linear indexing technique (Appendix 9). Adopting this technique, we can transform X and Y into one-dimensional variables labelled with the linear indices k and l , respectively. If k and l take m and n different positive integers, respectively, we obtain an $m \times n$ large contingency table (Fig. 15a). We are now able to measure the intergenic MI according to

$$MI = \sum_{k=1}^m \sum_{l=1}^n p(X_k, Y_l) \log \frac{p(X_k, Y_l)}{p(X_k)p(Y_l)} . \quad (227)$$

Although this equation follows the exact calculation method for MI , it has drawbacks. MI obtained using Eq. (227) is derived from the relative frequency of each cell (k, l) in the table, and it thus does not reflect the sample size and is irrespective of the significance level. This lack of significance level can lead to an increase in the computation time without yielding meaningful data. Therefore, to consider each cell's significance level and avoid valueless computation, we devised a novel analysis protocol that uses the micromutual information MI_{kl} .

Micromutual information MI_{kl}

To apply the statistical approach, the $m \times n$ large contingency table is transformed into a 2×2 small contingency table (Fig. 15b). For the cell (k, l) of the large contingency table, the micromutual information MI_{kl} is defined as MI of the corresponding 2×2 contingency table. When the indices k' and l' in the small contingency table take values of 1 or 2, respectively, MI_{kl} is obtained as

$$MI_{kl} = \sum_{k'=1}^2 \sum_{l'=1}^2 p(X_{k'}, Y_{l'}) \log \frac{p(X_{k'}, Y_{l'})}{p(X_{k'})p(Y_{l'})} , \quad (228)$$

which indicates MI of X and Y shared at the specific microstate (k, l) .

Managing massive information using MI_{kl}

We next discuss the methodology of using MI_{kl} obtained above and the small 2×2 contingency table to manage much information initially included in the large $m \times n$ contingency table. We have developed a series of mathematical theorems that use MI_{kl} and MI as statistical variables. These theorems include 1) the MI_{kl} summation theorem, 2) the infocanonical distribution of MI_{kl} , 3) the frontier information theory, and 4) the equivalence principle of information and probability (Appendices 6, 18, 19, and 20).

Filtering using the odds ratio (OR)

Our goal is to examine the network information $H(N_i)$ of a gene G_i in a multi-gene system. We first characterize the multidimensional interaction between genes G_i and G_j in a dataset h . The respective states of the two genes (including both normal and aberrant states) were classified into m and n types. Thus, the $m \times n$ large contingency table is prepared (Fig. 15a). We next calculate the OR for each cell in the table and select only cells that exhibit a strong association between the two genes. At the same time, we distinguish the interaction in each cell between being cooperative and mutually exclusive [67]. To calculate the OR, we use the 2×2 small contingency table (Fig. 15b) and judge each interaction

as follows. If the OR is more than 2, then the interaction is strong and cooperative; if the OR is less than 0.5, then the interaction is strong and mutually exclusive.

We note here that the formation of mutually exclusive interaction is less frequent and less important than that of cooperative interaction. We suppose that the mutually exclusive interaction is the result of a highly acute evolution. While such a rapid alteration in the gene relationship is often observed in cancer, it rarely occurs in organismal evolution. Thus, in the following part of this paper, we mainly consider cells with an OR greater than 2 in the contingency table.

Equivalence principle of information and probability

With an $m \times n$ contingency table, we introduce a novel point of view on the relationship between the two genes G_i and G_j as informatons. On the one hand, we must evaluate the significance (p -value) of intergenic interaction, which varies greatly between datasets based on sample size differences. On the other hand, we require a quantitative measure of information transmission that is robust against variations in the sample size. However, to satisfy these contradictory demands, neither current information theory nor current probability theory is sufficient. In particular, because our analysis uses multiple large-scale genomic data, p -value instability is a major issue.

Here, we describe a theorem that connects information theory and probability. It claims the equivalence between MI and the p -value, which provides a dualistic (i.e., informational/probabilistic) point of view of the interaction between the two informatons. The realization probability of MI of two random variables is calculated using the following two equations. We refer to them collectively as the ***equivalence principle of information and probability***. This principle holds in the fields of both probability theory and statistics as follows.

- 1) In probability theory, the following formula holds as described by Eq. (17):

$$p = e^{-MI}, \quad MI = -\log p, \quad (229)$$

where the p -value does not depend on the sample size (Appendix 2). In Chapter 2, we used this formula to deal with the realization probability of MI of physical informatons.

- 2) In statistics, the following formula holds asymptotically in the limit of large N :

$$P_F = e^{-N \cdot MI}, \quad MI = -\frac{1}{N} \log P_F, \quad (230)$$

where P_F is Fisher's exact probability (Appendix 20). P_F depends on the sample size N and represents the statistical significance of MI . In this Chapter 4, we will apply this formula to the analysis of intergenic interactions using large-scale data.

Equations (229) and (230) together imply the infocanonical distribution of MI (Appendix 17) and constitute the basis for the statistical mechanics of information exchange between informatons, and they apply to diverse disciplines such as particle physics and genome medicine.

Features of the equivalence principle

In developing the *ab initio* GO method, we emphasize five essential points of the above theorem, especially regarding the second equation.

- 1) For a single informaton, the relationship between the self-information I and the probability p is expressed as $p = e^{-I}$ and $I = -\log p$ [12]. However, there has been no such corresponding paradigm regarding the information exchange between informatons.
- 2) While the theorem provides an informatics approach to analyze the probabilistic-statistical events, it conversely offers a way to characterize the probabilistic-statistical properties of informational phenomena.
- 3) The theorem provides the information-theoretical interpretation of Fisher's exact probability. Accordingly, it has the potential to make a paradigm shift in our understanding of so-called ***statistical significance***.
- 4) While MI is a robust measure that is not biased by the sample size, the theorem allows the evaluation of the statistical significance (p -value) of MI that is sensitive to the sample size.
- 5) On the basis of the p -value calculated for MI , we can apply meta-analysis to integrate multiple datasets, thereby improving the precision and accuracy of the estimation of MI , especially when using large-scale data.

Frontier mutual information MI_{front}

We revisit the $m \times n$ contingency table for G_i and G_j in the dataset h . When m and n are sufficiently large, MI_{kl} in the cell (k, l) follows an infocanonical distribution (Appendices 17 and 18). That is, if we set $x = MI_{kl}$ and $\lambda = mn/MI_{obs}$, then the probability density function of x is

$$p(x) = \lambda e^{-\lambda x}, \quad \lambda = \frac{mn}{MI_{obs}}, \quad (231)$$

where MI_{obs} is $MI_{obs}(G_i; G_j)$ calculated from the $m \times n$ contingency table. This means that the two genes interact predominantly in a limited number of cells in the table with large MI_{kl} values (Fig. 14e). Furthermore, when m and n become sufficiently large, the number of conditions that cause the substantive information exchange converges to only one. This conclusion is consistent with a common biological observation that genes (or gene products) interact with each other in their specific molecular forms, under particular spatiotemporal conditions, within restricted types of cells.

The infocanonical distribution of MI_{kl} noted above is generally applied to the information exchange between informatons. In this connection, MI for informatons relates to the physical forces, as described in Chapter 2. Accordingly, it is not surprising in physicochemical systems that information exchange with energy transfer commonly occurs under a limited set of conditions, where the most efficient transfer takes place (Fig. 14e). Such a specific condition for each informaton

corresponds to a state that has the nearly highest level of energy/information. Indeed, *frontier orbital theory* claims that organic compounds undergo electrophilic reactions with the atoms whose electron density of the highest occupied frontier orbital (HOMO) is large [68]. Of note, the HOMO is the *frontier orbital* with the highest energy level, and it therefore has the highest information level.

From the above considerations, in the $m \times n$ contingency table for G_i and G_j in the dataset h , we designate a cell with the highest MI_{kl} value as a **frontier cell**, where the two genes exchange information predominantly. Additionally, we define the **frontier mutual information (MI_{front})** as the maximum MI_{kl} manifested in the frontier cell; that is, $MI_{front} := \max(MI_{kl})$. We here assume that m and n are sufficiently large, and MI_{kl} thus follows an infocanonical distribution (Appendix 18). MI_{obs} can then be estimated using MI_{front} (Appendix 19) as

$$MI_{obs} \simeq \frac{mn}{\log(mn-1)} MI_{front} \quad (232)$$

Next, for a cell (k, l) in the original $m \times n$ large contingency table, we set $P_{F_{kl}}$ as the P_F -value of the transformed 2×2 small contingency table. We additionally define the **frontier probability (P_{front})** as the minimum of $P_{F_{kl}}$; that is, $P_{front} := \min(P_{F_{kl}})$. We can now apply the equivalence principle of information and probability [Eq. (230)]. If N is sufficiently large, then we can directly calculate MI_{front} from P_{front} according to

$$MI_{front} = - \lim_{N \rightarrow \infty} \frac{1}{N} \log P_{front} \quad (233)$$

Furthermore, we get the 95% confidence interval of MI_{obs} according to

$$\frac{mn}{\log(mn-1) + 2.970} MI_{front} \leq MI_{obs} \leq \frac{mn}{\log(mn-1) - 2.970} MI_{front} \quad (234)$$

If we assume, for example, that both G_i and G_j express 20 discrete states for each of 50 different properties, then $m = n = 20^{50}$. The relative error of MI_{obs} is within 0.99% in this instance because $2.970/\log(mn-1)$ is 0.0099. In conclusion, owing to there being a great number of states, the intergenic interaction $MI_{obs}(G_i; G_j)$ can be computed from MI_{front} with a fairly small error. This procedure has the advantage that we can reduce the calculation time by omitting the cells without statistical significance. In addition, the P_F -value for G_i and G_j is readily estimated from MI_{front} and $P_F \simeq e^{-N \cdot MI_{obs}}$.

Signature information H_{sgn}

In addition to $MI_{obs}(G_i; G_j)$, we can obtain the **signature information $H_{sgn}(G_i, G_j)$** in the dataset h . This is done by comparing the direction of the mRNA expression between the two genes in the frontier cell. Depending on the direction, H_{sgn} is defined as taking a value of 1 or -1 , indicating the same or opposite orientation. The frontier cell represents a gene interaction with a significance value of P_{front} , and H_{sgn} is therefore also expected to have a significance that reflects P_{front} . In addition, because we focus on cooperative interactions, $H_{sgn} = 1$ and -1 respectively indicate a gene direction that the coexpression and reciprocal expression of the genes G_i and G_j promotes their collaborative functioning.

Data aggregation of MI_{obs} and H_{sgn}

To accomplish the computation of the interaction between G_i and G_j , we aggregate $MI_{obs}(G_i;G_j,h)$ and $H_{sgn}(G_i,G_j,h)$ from multiple datasets, where the index h denotes the h -th dataset. First, we repeat the above $MI_{front}(h)$ calculations across various datasets for distinct types of tumor. We then obtain $P_{overall}$ by summarizing each $P_{front}(h)$ based on a meta-analysis. After the calculation of $P_{overall}$, we can transform $P_{overall}$ into MI ($=: MI_{overall}$) based on the equivalence principle [Eq. (230)] because the principle also holds after meta-analysis. Thus,

$$MI_{overall} = - \lim_{N_{overall} \rightarrow \infty} \frac{1}{N_{overall}} \log P_{overall}, \quad (235)$$

where $N_{overall}$ is the total number of samples across all datasets examined. Second, the overall value of $H_{sgn}(G_i, G_j)$ ($=: H_{sgn_overall}$) is calculated as the signature of the total sum of each $H_{sgn}(G_i, G_j, h)$; that is,

$$H_{sgn_overall} = \text{sgn} \left[\sum_h H_{sgn}(h) \right]. \quad (236)$$

The two-gene network N_{ij} is now determined from the network information $H(N_{ij})$ using Eq. (214) as

$$H(N_{ij}) = [MI_{overall}(G_i;G_j), H_{sgn_overall}(G_i, G_j)]. \quad (237)$$

Finally, the network information $H(N_i)$ of the gene G_i is obtained by iterating the above computations for all other genes G_j . Then, as described in Chapter 3, $H(N_i)$ is expressed as the superposition of each $H(N_{ij})$ in multidimensional space:

$$H(N_i) = \sum_{j \neq i} H(N_{ij}). \quad (238)$$

Again, this is an equivalent derivative of the LCIO in the multidimensional information metric spacetime.

Consequently, using the evolutionary microscope and conducting a meta-analysis of frontier information, we can reveal the LCIO in cancer cells. Owing to the vast volume of data in the TCGA, the network structure calculated in cancer cells is thought to reflect the network structure in normal cells. We thus adopt the obtained results as values consistent with those of the normal gene network.

(d) Weighted pathway analysis

The LCIO predicts that G_i performs network functions based on $H(N_i)$. In other words, the network functions of a gene are determined by what kind of genes it affects. This can be validated using the weighted pathway analysis described herein (Appendix 21). The analysis adopts the principle that genes with related functions share large MI with each other (Appendix 16). Therefore, using $H(N_i)$ obtained in the preceding section, we can elucidate the network functions statistically. We here suppose that G_i affects the functions of other genes G_j . We now set $F_j(G_i) = N_{ij} = [MI(G_i;G_j), H_{sgn}(G_i,G_j)]$ as the

j -th network function of G_i . The total network function of G_i is then equal to the network information $H(N_i)$ that is superimposed on the multidimensional space as

$$H(N_i) = \sum_{j \neq i} F_j(G_i) = \sum_{j \neq i} [MI(G_i; G_j), H_{sgn}(G_i, G_j)] \quad (239)$$

By considering the projection of $H(N_i)$ to three-dimensional space, we can compute Σ^W , Σ^{sgn} , and Σ^{rank} in Eq. (220), which provides a clue with which to decipher the network function of G_i .

More specifically, we can elucidate the network functions of G_i through weighted pathway analysis. This analysis determines the significance level at which G_i is involved in a specified pathway. As detailed in Appendix 21, we know the network functions of G_i with P_{WE} expressed by

$$P_{WE} \simeq \sum_m \frac{M!}{m!(M-m)!} \cdot \frac{(N-M-1)!}{(n-m)!(N-M-n+m-1)!} \cdot \frac{n!(N-1-n)!}{(N-1)!} \\ \times \frac{1}{\sqrt{2\pi}} \int_{\sqrt{m}(\frac{w}{m}-\mu)/\sigma}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \quad (240)$$

Practically speaking, we adopt Ingenuity Pathway Analysis (IPA), although it cannot evaluate edge weights. We perform Canonical Pathways analysis using a set of G_j with $MI_{overall}(G_i; G_j)$ ranks above 2000 or 8000. In addition, we use $H_{sgn_overall}$ to conduct Diseases and Functions analysis, which computes the magnitude and direction of G_i functions and the relationships of G_i with pathophysiological conditions. Furthermore, other informatics programs of IPA, such as Upstream Analysis, Regulator Effects Analysis, and Causal Network Analysis, can also be performed. Thus, through the application of IPA, we can readily interpret a wealth of information regarding the functional properties of the LCIO.

(e) Application of the *ab initio* GO method

Genes have evolved for 4 billion years. They now possess profound functions, which are far beyond human understanding. Genes are involved in many diseases, and a technique that clarifies their broad functions rapidly and accurately will lead to substantial advances in medical sciences. The *ab initio* GO method allows the rapid functional analysis of genes of interest, which took more than 10 years by conducting *wet* experiments. Because this method uses a large amount of network information of the LCIO, we can obtain a wide range of knowledge about the features of gene function and its pathophysiological relevance. This section examines the characteristics of gene networks by adopting the *ab initio* GO analysis. We subsequently provide examples of practical interpretation regarding the properties of the gene network function.

Network properties of the LCIO

From a network theoretical point of view, the LCIO presents a novel type of architecture for gene networks. The LCIO-based network model determines the properties of both nodes and edges. As for the nodes, the Shannon information entropy $H(G_i)$ of a gene G_i follows an exponential distribution by applying the maximum entropy principle [69,70]. This is analogous to physical systems, where the thermodynamic entropy S follows an exponential distribution under isothermal conditions. For the edges, the LCIO employs MI as edge weights, and MI is also expected to follow an exponential distribution, which we call the infocanonical distribution [Eq. (191)] (Appendices 2 and 17).

In contrast to the LCIO, the traditional theory states that many biological networks are *scale free* [71]. In this model, degree k of nodes follows a power law as $P(k) \sim k^{-\gamma}$; that is, the Pareto distribution. Importantly, this model neglects edge weights and assumes that the existence of the edges between nodes follows the all-or-none principle. However, because genes have been under selective pressure for a very long time, there is no doubt that mutual interactions have occurred for virtually all pairs of genes. Furthermore, it is natural that a physiological level of variation has emerged among the edge weights. Hence, we propose that the LCIO revises the classical model of gene networks. The validity of the LCIO-based network will be confirmed using examples later.

Distribution of MI obtained using the *ab initio* GO method

We performed an *ab initio* GO computation to get an overview of gene networks. We initially calculated the quantity distribution of gene information $H(G_i)$. $H(G_i)$ is approximated from the network information $H(N_i)$ using Eq. (224), and we computed $MI(G_i;G_j)$ for all genes using Eq. (235).

We found that, for most cases, $MI(G_i;G_j)$ of a single gene G_i with other genes G_j follows the infocanonical distribution (Appendices 2 and 17). In this case, $MI(G_i;G_j)$ follows the logarithmic distribution of magnitude rank j [Eq. (192)] (Fig. 16). By contrast, $MI(G_i;G_j)$ for rapidly evolving genes with strong influential forces tended to follow the Pareto distribution. Indeed, most of the immune genes exhibited the Pareto distribution. This particular phenomenon is consistent with the general notion that immune genes evolve rapidly to cope with the expeditious evolution of pathogens.

In addition, we noted that each gene tended to interact with other genes having the same kind of MI distribution as its own. (Results are to be published elsewhere.) Thus, genes are divided by the MI distribution into two groups: an infocanonical type and a Pareto type. Furthermore, it was found that genes with closely related functions share much MI with each other (Appendix 16). Thus, genes constitute a biological pathway by sharing MI with other genes that express interrelated information. Therefore, it is conceivable that strong influential forces acting between genes form a pathway, thereby producing greater biological fitness.

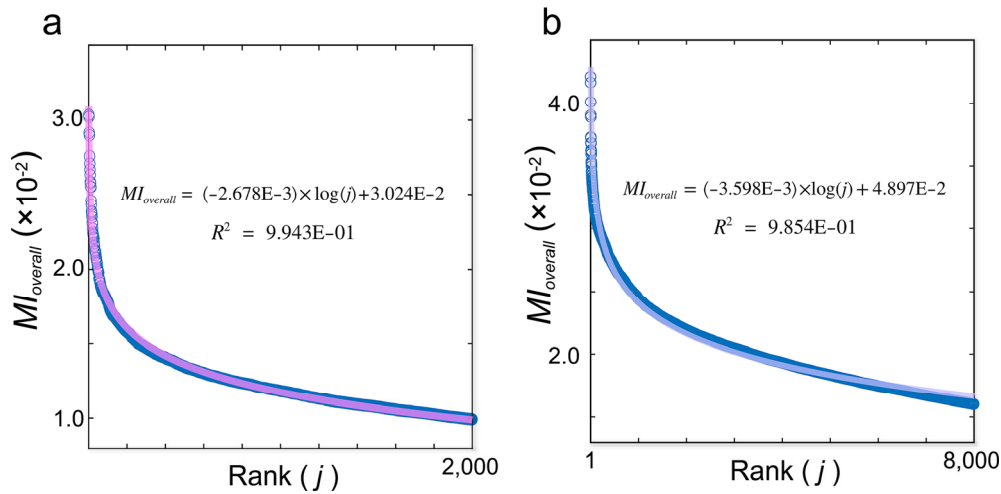


Figure 16 | Rank distribution of MI . $MI_{overall}$ of $G_i = ASXL1$ (a) or $CLSTN3$ (b) with other genes G_j is represented by the blue circles, where j is the rank of the magnitude of $MI_{overall}$ with G_i . The regression is shown as the pink line in (a) and the blue line in (b).

In an earlier study [62], the authors described a method for constructing a gene network by calculating MI only for regulators and targets, assuming that the mRNA expression of genes follows a normal distribution. Our methods are different from previous methods because we calculate MI for all genes without assuming any distribution.

ASXL1

Additional sex combs like 1 (ASXL1) is a member of the *ASXL* family and is involved in epigenetic regulation. Somatic mutations of the *ASXL1* gene are frequently detected in hematopoietic neoplasms, including those of myelodysplastic syndromes and acute myeloid leukemia. *ASXL1* interacts with a deubiquitinase BAP1 to form polycomb-repressive deubiquitinase, which removes histone H2A lysine 119 ubiquitylation (H2AK119Ub) [72]. It has been shown that the disease-associated *ASXL1* mutations form hyperactive deubiquitinase complex with BAP1 to upregulate several target genes, including HOX genes [73,74]. Additionally, a recent report showed that the mutant *ASXL1*/BAP1 complex induces deubiquitination of phosphorylated AKT, thereby activating the AKT/mTOR pathway in hematopoietic stem cells. The mutant *ASXL1*-induced overactivation of AKT/mTOR signaling provokes mitochondrial dysregulation and promotes premature aging of hematopoietic stem cells [75]. Thus, *ASXL1* regulates the ubiquitination of histones (e.g., H2AK119ub) and non-histone proteins (e.g., phosphorylated AKT) to maintain the hematopoietic system.

We here applied the *ab initio* GO analysis to *ASXL1*. The levels of mRNA expression and DNA methylation were used as computation dimensions unless otherwise stated. First, we made a list of

relevant genes and calculated the intensities and directions of their actions using the *ab initio* GO method. We used the highly related genes whose *MI* shared with *ASXL1* was ranked within the top 2000. As shown in Fig. 16a, $MI_{overall}$ values were distributed logarithmically with rank j at a high coefficient of determination ($R^2 = 0.9943$), showing the infocanonical distribution of *MI* [Eq. (191)] and thereby supporting the precision of our prediction (Appendix 2). $MI_{overall}$ of *AKT2* with *ASXL1* was ranked in the top 0.052% (10th) of the 19,195 genes, demonstrating the involvement of *ASXL1* in the AKT pathway, consistent with the experimental observation.

We next estimated the pathways and functions of *ASXL1* using the Canonical Pathways analysis and Diseases and Functions analysis of IPA, respectively. In the Canonical Pathways analysis, we found that the Protein Ubiquitination Pathway and mTOR Signaling ranked second (goodness-of-fit, $p = 2.04E-5$) and third ($p = 3.11E-5$), respectively (Fig. 17a). In addition, the first-ranked was EIF2 Signaling ($p = 6.81E-8$), downstream of the ATK/mTOR pathway involved in cell survival. In the Diseases and Functions analysis, activation was deduced for Cell survival, Cell transformation, Leukopoiesis, and Ubiquitination, with z -scores 6.203 ($p = 1.46E-10$), 2.920 ($p = 2.44E-8$), 2.421 ($p = 1.45E-6$), and 2.254 ($p = 1.22E-5$), respectively (Fig. 17b). Finally, when mRNA expression levels, mutation, and DNA copy number variation were used as dimensions, Oxidative Phosphorylation and Mitochondrial Dysfunction ranked first ($p = 2.76E-28$) and second ($p = 5.83E-24$) respectively in the Canonical Pathways analysis. Thus, the pathways and functions identified by the *ab initio* GO method agree well with the known roles of *ASXL1*.

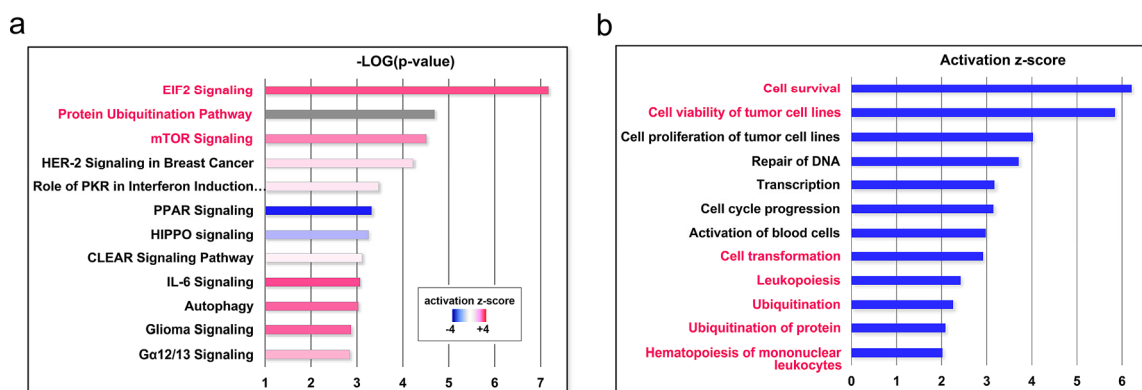


Figure 17 | *Ab initio* GO analysis for *ASXL1*. Functions of *ASXL1* calculated using the *ab initio* GO method and IPA. Items relevant to the text description are shown in red. **a**, Canonical Pathways analysis. The magnitudes of z -scores are shown by the color scale bar in the inset. A grey bar means that the z -score was not determined. **b**, Diseases and Functions analysis. The lengths of bars express the goodness-of-fit in (a) and activation z -scores in (b).

CLSTN3

CLSTN3 is involved in synaptogenesis and synaptic plasticity. The calyntenin family links between vesicles and kinesin motors for axonal transport [76,77] and is implicated in memory and learning [78] as well as the secretory pathway [79,80]. We calculated the functions of *CLSTN3* using the *ab initio* GO method in the same way as for *ASXL1*. We used the highly related genes whose *MI* with *CLSTN3* was ranked within the top 2000. *MI_{overall}* followed the infocanonical distribution with a high coefficient of determination ($R^2 = 0.985$), supporting computation accuracy (Fig. 16b, Appendix 2).

When the selected genes were subjected to the Canonical Pathways analysis in IPA (Fig. 18a), the Synaptogenesis Signaling Pathway ranked top, exactly indicating the primary function of *CLSTN3*. In the Diseases and Functions analysis (Fig. 18b), the Microtubule dynamics and Transport of vesicles had high activation *z*-scores, 6.353 ($p = 1.64E-22$) and 2.764 ($p = 5.63E-7$) respectively. These were illustrated by the findings that *MI* of *CLSTN3* and either kinesin 3B or 3C ranked in the top 0.59% (117th) or 0.11% (21st), respectively. The kinesin-mediated transport was thus correctly estimated [81]. In this connection, *CASY-1*, the homolog of *CLSTN3* in *Caenorhabditis elegans* [82], transports the insulin receptor by complexing with kinesin moving on microtubules [83].

Additionally, *CLSTN3* was deduced to augment Cognition and Learning, higher physiological activities of the nervous system. Furthermore, Synaptic transmission as well as both Long-term depression and Long-term potentiation, which are the basic mechanisms of memory, were elucidated. All of these activities were detected with *z*-scores greater than 2.0 ($p < 3.0E-7$). Thus, the *ab initio* GO method successfully predicted both molecular and physiological functions of *CLSTN3*. We present other examples of *ab initio* GO calculation in Appendix 31.

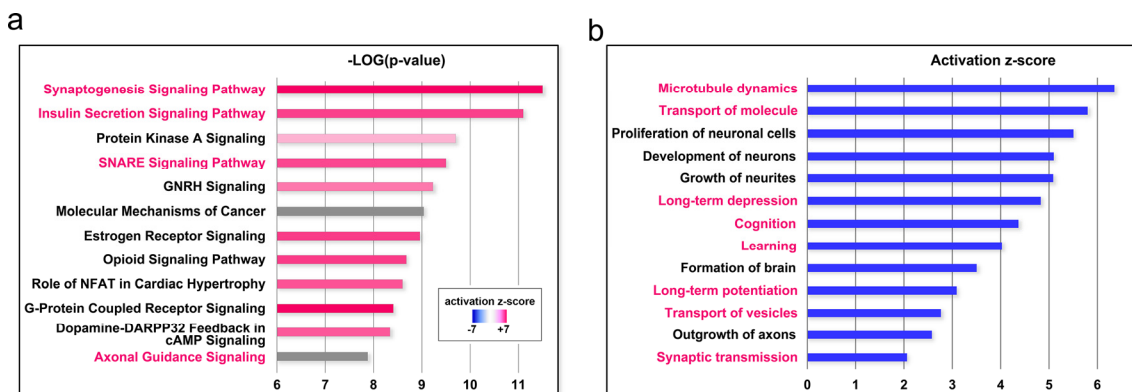


Figure 18 | *Ab initio* GO analysis for *CLSTN3*. Functions of *CLSTN3* calculated using the *ab initio* GO method using IPA. Items relevant to the text description are shown in red. **a**, Canonical Pathways analysis. The magnitudes of *z*-scores are shown by the color scale bar in the inset. A grey bar means that the *z*-score was not determined. **b**, Diseases and Functions analysis. The lengths of bars express the goodness-of-fit in (a) and activation *z*-scores in (b).

(f) Searching for a novel immune checkpoint in cancer

Progress in modern medicine has revealed that genetic changes in cancer are highly divergent among patients, broader than previously thought, and range over the whole genome. Therefore, the extraction of essential genetic alterations from enormous data is a critical issue in current oncology. However, existing informatics tools have poor functionality in suggesting which genes are the most intimately related to cancer pathophysiology, although they enable us to browse genes one by one on databases. Furthermore, the registered gene information in databases depends only on the knowledge obtained from experimental observations in the past, which limits the range of information to less than that needed for rapid advancement.

STAIC—a strategic tool for cancer genome analysis

We developed STAIC to discover cancer genes more efficiently and overcome the drawbacks of conventional informatics programs. STAIC is a program that includes data of the 33 kinds of tumor in the TCGA and allows the efficient identification of genes associated with cancer pathophysiology (initiation, progression, and prognosis) by considering simultaneously four types of information on the mRNA expression profile, DNA copy number, mutation, and clinical course of all the genes. Specifically, to better identify the ‘driver’ genes involved in cancer initiation, STAIC uses the COPA algorithm [84] in analyzing the mRNA expression and DNA copy number profiles. Moreover, STAIC offers a convenient command to immediately start the *ab initio* GO analysis of the discovered genes. These features allow a rapid identification of candidate genes followed by the successive, quick but detailed prediction of their functional and pathophysiological properties.

After developing STAIC, we used it to investigate datasets for 19 kinds of tumor with sufficient information. We detected 90.6% of previously reported tumor genes (92.3% for epithelial and 84.5% for non-epithelial tumors). Moreover, we found many unpublished genes having a strong relationship with cancer. For example, 30 and 38 genes had a mutation frequency exceeding 20% in lung adenocarcinoma (LUAD) and pancreatic adenocarcinoma (PAAD). For LUAD and PAAD, very few of these genes have been reported in the literature on their role in cancer. This lack of literature is noteworthy if we consider the enormous past investments spent in identifying oncogenes.

Finally, we performed the *ab initio* GO analysis of the discovered genes and found that their biofunctions were compatible with the cancer pathology. (Data will be published elsewhere.) In this way, the STAIC analysis followed by the *ab initio* GO method facilitates the rapid identification and characterization of novel cancer genes that have potential applications as biomarkers and/or therapeutic targets.

Immune checkpoint in cancer

Immune checkpoints are gatekeeper mechanisms that maintain self-tolerance and moderate immune responses [85]. In cancer, immune checkpoints are frequently activated to evade antitumor immunity. In 2018, James P. Allison and Tasuku Honjo won the Nobel Prize in Physiology or Medicine as the founders of cancer immune therapy that aims at restoring the antitumor reaction by targeting the checkpoint molecules *CTLA-4* and *PD-1* (gene symbol, *PDCDI*), respectively. Following their success, the search for other immune checkpoints has become a hotspot of research in oncology. Thus, we search for a novel immune checkpoint gene by combining STAIC and the *ab initio* GO method.

Searching for a novel immune checkpoint in LUAD

LUAD is a representative immune-driven tumor. It involves widespread gene alterations causing various cellular dysfunctions, and it is thus possible that mutations of unknown immune genes may underlie the abrogation of anti-cancer immunity. Therefore, it is an ideal target of the investigation to identify a novel immune checkpoint. Hence, we performed a search procedure using the LUAD dataset of the TCGA by employing a two-step process; that is, a STAIC-based enrichment of candidate genes coupled with an *ab initio* GO prediction of their functions.

The impairment of anti-cancer immunity is not only an early event in carcinogenesis but also associated with cancer development at a later stage, leading to a poor outcome [86]. In contrast, the number of genes showing such dual nature is not high among the classical oncogenes and anti-oncogenes. Therefore, by taking advantage of an enrichment function of STAIC, we extracted genes with deep involvement in both carcinogenesis and malignancy. We thus began a STAIC analysis for searching for genes that fulfill the following criteria.

- 1) The genes have the highest COPA scores for mRNA expression levels; that is, they are ranked within the top 10% of all genes. The COPA is an excellent algorithm for finding ‘driver’ genes, mainly involved in cancer initiation [84].
- 2) The mutation frequency is more than 3%. This criterion enriches the cancer-associated genes.
- 3) The *p*-value of the log-rank test is sufficiently small. This criterion ensures that the gene is involved in cancer progression.
- 4) The proportion of patients with a combination of genetic alterations (out of 47 types) that gives the lowest *p*-value in the log-rank test reaches 15%. This criterion is set assuming that specific conserved changes in relatively limited genes underlie the frequent immune dysfunction of LUAD.

After the enrichment procedure, we performed the *ab initio* GO analysis of the candidate genes that gave significant *p*-values in the log-rank test. We then found *KYNU*, which exhibited the smallest *p*-value related to severe prognosis (Fig. 19). Finally, an in-depth IPA analysis concluded that *KYNU* is a novel immune regulator involved in the pathophysiology of LUAD.

Genes	RANK (EXP)	mRNA (%)	CNV (%)	MUT (%)	Log-Rank (p-value)	-1 -LOG[P]	Select Condition (%)	-LOG[P] x (%)	Description	Map Location	Chromosome
KYNU	11500	5.61	57.98	3.68	9.34e-05	4.03	21.38	86.17	kyreninase	2q22.2	2
AKAP12	4970	2.42	91.99	3.33	3.32e-04	3.48	29.77	103.55	A-kinase anchoring protein 12	6q25.1	6
COL7A1	29760	14.03	96.73	8.76	3.58e-04	3.43	29.93	99.26	collagen type VII alpha 1 chain	3p21.31	3
CD109	21810	10.64	89.53	4.90	4.69e-04	3.33	30.40	101.20	CD109 molecule	6p13	6
TRIM29	9320	4.55	38.24	4.20	1.10e-03	2.96	19.08	56.47	tripartite motif containing 29	11q23.3	11
MKI67	16100	8.33	64.52	8.76	1.67e-03	2.78	67.09	186.40	marker of proliferation Ki-67	10q26.2	10
FGA	340	0.17	90.71	3.85	2.44e-03	2.61	28.72	75.04	fibronogen alpha chain	4q31.3	4
COL11A1	8980	4.19	91.00	21.72	3.33e-03	2.48	61.43	152.18	collagen type XI alpha 1 chain	1q21.1	1
CDC42BPA	16880	8.23	4.90	3.33	8.91e-03	2.05	84.70	173.83	CDC42 binding protein kinase alpha	14q23.1	14
CPS1	830	0.40	62.27	13.31	0.020	1.70	31.03	52.65	carbamoyl-phosphate synthase 1	2q34	2
SERPINF3	14230	6.94	82.23	5.08	0.021	1.69	18.66	31.44	serpin family B member 3	18q21.33	18
ITGA2	9860	4.81	15.76	4.55	0.023	1.63	83.23	136.02	integrin subunit alpha 2	5q11.2	5
CEHFP	27940	13.19	4.90	9.11	0.041	1.38	74.63	103.82	centromere protein F	1q41	1
NES	28180	13.75	1.53	7.36	0.045	1.34	51.36	69.04	nestin	1q23.1	1
DSC2	29290	14.29	49.57	4.38	0.049	1.31	51.57	67.53	desmocollin 2	18q12.1	18
NR4A2	23970	11.69	58.76	3.88	0.054	1.27	26.42	33.49	nuclear receptor subfamily 4 group A member 2	2q24.1	2
ATAD2	30740	14.99	6.59	3.89	0.055	1.26	83.69	104.06	ATPase family, AAA domain containing 2	9q24.13	9
MYO5B	30130	14.70	87.13	3.88	0.063	1.20	94.76	113.57	myosin VB	18q21.1	18
MUC16	13280	6.48	95.66	45.36	0.068	1.17	50.10	58.82	mucin 16, cell surface associated	19p13.2	19
MAP1B	18460	8.03	17.98	8.41	0.071	1.15	44.65	51.23	microtubule associated protein 1B	5q13.2	5
CAPN6	28620	13.96	31.66	5.08	0.080	1.10	17.19	18.83	calpain 6	Xq23	X
COL12A1	4310	2.10	86.89	10.88	0.082	1.09	66.25	72.12	collagen type XII alpha 1 chain	8q13-q14.1	8
MIA3	21400	10.44	4.90	3.15	0.093	1.03	95.44	99.60	MIA family member 3, ER export factor	1q41	1
MUC4	4570	2.23	1.31	8.41	0.095	1.02	74.00	75.51	mucin 4, cell surface associated	3q29	3
PCDH1	23800	11.61	20.15	4.20	0.108	0.97	96.44	93.09	protocadherin 1	5q31.3	5
IGF1R	24530	11.97	32.18	3.33	0.113	0.95	75.47	71.37	insulin like growth factor 1 receptor	17q29.3	17
ABCC5	6590	3.21	22.38	3.15	0.116	0.93	93.29	87.23	ATP binding cassette subfamily C member 3	17q21.33	17
HK2	15540	7.58	56.01	3.33	0.117	0.93	89.52	83.48	hexokinase 2	2p12	2
DST	10020	4.89	17.01	20.14	0.130	0.89	78.83	69.87	dystonin	6p12.1	6
CST1	11340	6.53	7.57	3.33	0.132	0.88	23.77	25.21	costalin	20p11.21	20
PDE4DIP	21930	10.70	0.01	11.91	0.154	0.81	97.27	78.92	phosphodiesterase 4D interacting protein	1q21.2	1
FNDC1	27160	13.25	86.26	6.48	0.159	0.80	61.43	49.02	fibronectin type III domain containing 1	6q25.3	6
FS	17110	8.35	2.98	8.06	0.159	0.80	16.03	14.37	copulation factor V	1q24.2	1
MAP2	23720	11.97	56.61	8.58	0.164	0.79	45.27	38.73	microtubule associated protein 2	2q34	2
LAMA3	5490	2.68	46.79	6.13	0.180	0.75	78.62	58.82	laminin subunit alpha 3	18q11.2	18
TJPI	28220	12.79	82.73	3.33	0.188	0.73	97.27	71.07	tight junction protein 1	15q13.1	15
AHNAK2	8920	3.23	51.14	19.09	0.189	0.73	94.07	61.11	AHNAK nucleoprotein 2	14q23.33	14
ANTXR1	6980	3.40	56.01	4.03	0.191	0.72	92.24	66.41	anthrax toxin receptor 1	2p13.3	2
CAD	26340	12.36	63.80	5.60	0.191	0.72	81.34	58.43	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase...	2q23.3	2
LAMA5	6170	3.01	8.38	7.18	0.195	0.71	92.45	65.33	laminin subunit alpha 5	20q13.33	20
COL6A3	1040	0.51	60.96	15.06	0.201	0.70	15.51	10.81	collagen type VI alpha 3 chain	2q37.3	2
PLCE	26670	19.69	64.50	4.90	0.202	0.69	83.88	88.26	phosphoinositide 3-kinase class II delta	10q22.3	10

Figure 19 | STAIC analysis for LUAD in the TCGA. Gene enrichment was performed by COPA (mRNA) \leq top 15%, mutation \geq 3%, and affected patient population \geq 15%. The enriched genes were sorted in decreasing order of the minus log(p-value) of the log-rank test. *KYNU* was ranked highest of all genes.

KYNU role in LUAD

The LCIO predicts that the network function of a gene is of primary importance [Eq. (224)]. For this reason, we cannot obtain a complete understanding of the gene function when we rely only on conventional molecular biological techniques. Our STAIC analysis revealed that many unreported genes showed prominent associations with tumors. Therefore, there may remain undiscovered immune checkpoints in cancer. In this regard, it is noted that controversy exists regarding the importance of *PD-1* (*PDCDI*) and its ligands, *PD-L1* (*CD274*) and *PD-L2* (*PDCDILG2*), in LUAD [87-90]. We hypothesized that the known immune checkpoints are insufficient to fully explain the immune dysfunction in LUAD. Thus, to broaden our perspective, we applied our informatics algorithms and identified *KYNU*. When we conducted the *ab initio* GO method with IPA, the Disease and Functions analysis indicated that *KYNU* is intimately associated with immunity (Fig. 20a).

To improve the sensitivity of detecting gene involvement in the disease outcome, STAIC was implemented with automatic optimization of the search conditions. This function facilitated the discovery of many unpublished genes associated with LUAD. Despite this high sensitivity, however, the optimized STAIC produced starkly contrasting results between *KYNU* and the *PD-1* ligands as follows. The patients with altered *KYNU* [higher mRNA ($>$ mean+2SD) and/or DNA amplification] showed a marked association with adverse prognosis in an optimized log-rank test ($p = 9.34E-5$) (Fig.

20b). This index indicated a much worse prognosis than in the cases of altered *IDO1* ($p = 0.00268$, Fig. 20c), altered *PD-L1* ($p = 0.148$, Fig. 20d) or altered *PD-L2* ($p = 0.111$) [where the latter two groups had higher mRNA ($> \text{mean}+2\text{SD}$) and/or DNA amplification, and/or DNA gain]. Moreover, the population with altered *KYNU* (21.38%) was much greater than that with altered *PD-L1* (2.31%) or *PD-L2* (0.63%). Similarly, a more significant p -value was observed with *KYNU* than *PD-L1* or *PD-L2* in PAAD and lung squamous cell carcinoma (LUSC) under the optimized search conditions. These findings suggest that *KYNU* is prominently involved in cancer progression as a major pathogenic factor.

In addition to the later stages of LUAD, *KYNU* is likely involved in the earlier stages. This is evidenced by the COPA analysis on STAIC. *KYNU* ranked in the upper 5.61% of all genes when we calculated the COPA score under the condition of mRNA overexpression at the first percentile of the patient population. This observation suggests that the overexpression of *KYNU* is a critical upstream event in carcinogenesis, which is consistent with its putative immune function. In contrast, the COPA scores of *PD-L1* and *PD-L2* were ranked in the upper 47.79% and 60.13%, respectively. These ranks are almost the average for all genes, and it is thus difficult to conclude that these *PD-I* ligands are cancer drivers. Thus, *KYNU* is associated with LUAD pathology more deeply than the *PD-I* pathway.

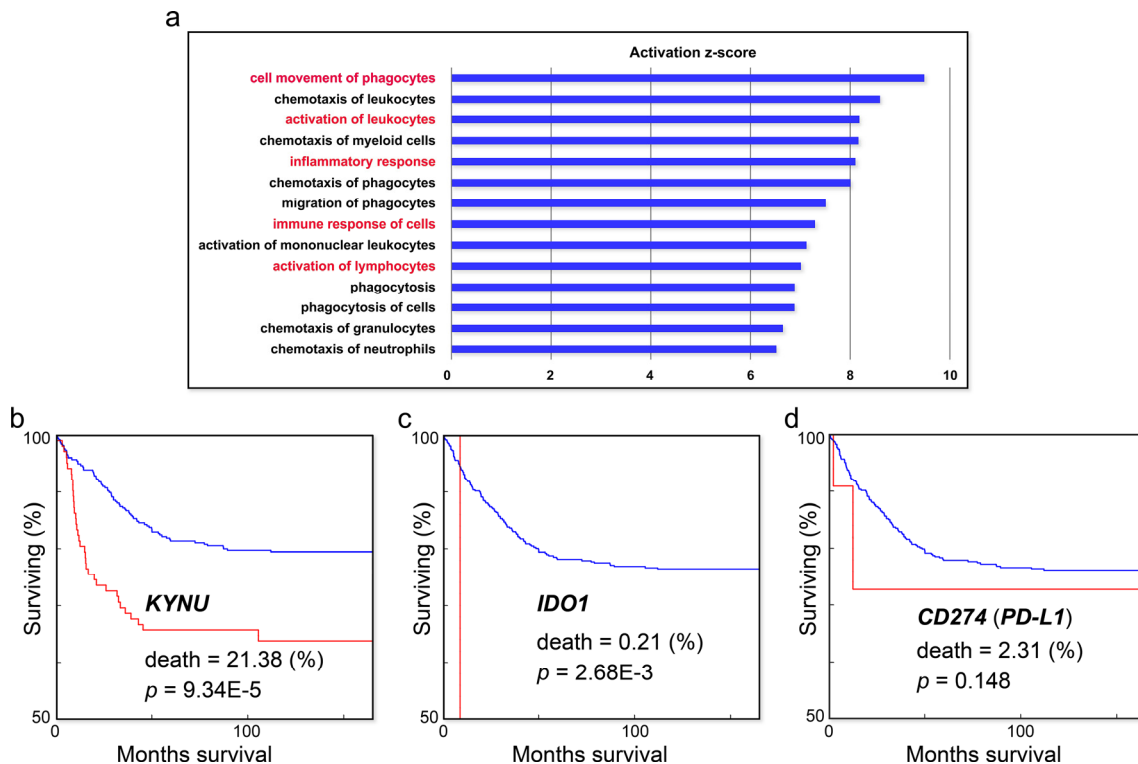


Figure 20 | Search for an immune checkpoint using the *ab initio* GO method. a, Functions of *KYNU* calculated using the *ab initio* GO method and the Diseases and Functions analysis of IPA. **b**, **c**, **d**, Prognosis of patients of LUAD with overexpression or mutations of *KYNU* (**b**), *IDO1* (**c**), and *CD274* (**d**) from the TCGA ($n = 477$). The red and blue lines represent the survival with overexpression or mutations and normal expression, respectively. Ratios of alteration-related deaths to controls and p -values from log-rank tests are indicated.

***KYNU* role in cancers other than LUAD**

Among previous studies on *KYNU* and cancer, some studies supported an oncogenic role [91-96] whereas a few showed a tumor suppressor role [97]. Such controversy occasionally arises for genes involved in cancer. We performed the optimized log-rank test on STAIC. Figure 21 illustrates the examples of Kaplan–Meier curves with significance ($p < 0.05$). Panels (a) to (e) indicate that the overexpression of *KYNU* leads to a worse prognosis, while panel (f) shows that its low expression predicts a better outcome. Especially in PAAD (Fig. 21a), the overexpression of *KYNU* was more strongly associated with unfortunate outcomes than in LUAD.

In summary, the aberrant increased expression of *KYNU* is often accompanied by poorer outcomes in various cancers. Notably, in LUAD and PAAD, *KYNU* was altered in expression levels at a much higher frequency than the *PD-1* pathway, indicating that this gene is deeply involved in the pathogenesis of these two types of adenocarcinoma.

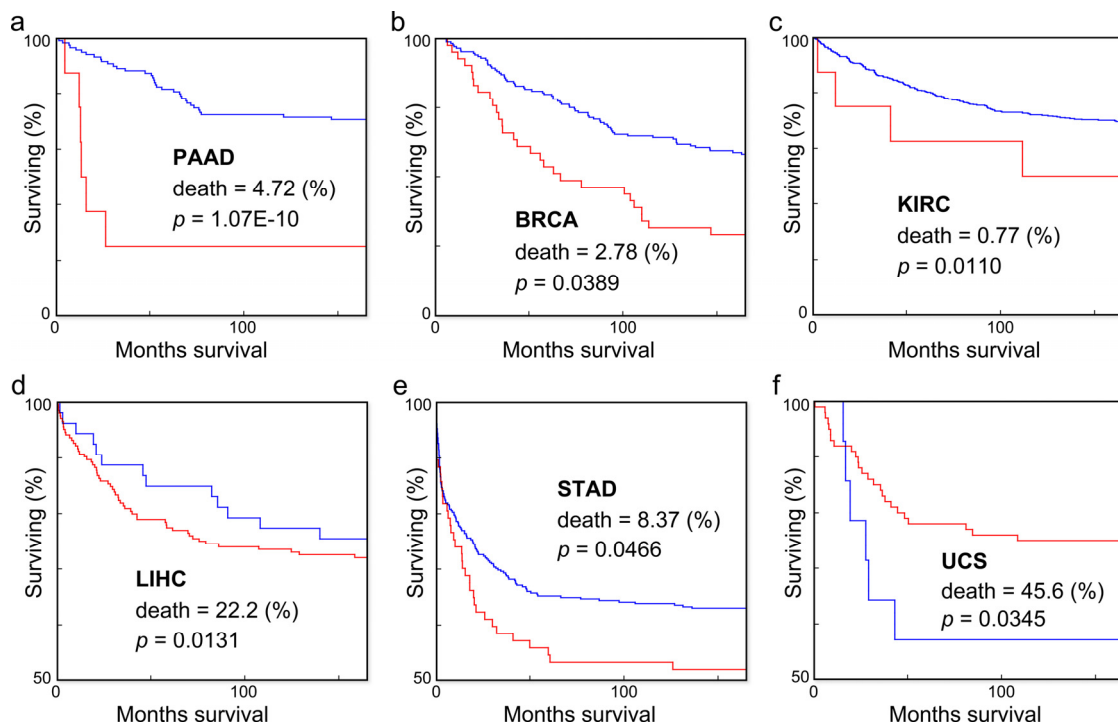


Figure 21 | Prognosis of patients of various cancer types with altered *KYNU* expression from the TCGA. The red and blue lines represent the survival with altered expression or mutations and normal expression, respectively. **a**, PAAD with amplification or mRNA expression > mean + 2SD ($n = 127$). **b**, Breast invasive carcinoma with mutation or mRNA expression > mean + 1SD ($n = 1043$). **c**, Kidney renal clear cell carcinoma with amplification or mRNA expression > mean + 2SD ($n = 521$). **d**, Liver hepatocellular carcinoma with amplification, mutation, or mRNA expression > mean + 0.5SD ($n = 257$). **e**, Stomach adenocarcinoma with mRNA expression > mean + 1SD ($n = 442$). **f**, Uterine carcinosarcoma with homozygous deletion or mRNA expression < mean - 1SD ($n = 57$). Ratios of alteration-related deaths to controls, and p -values from log-rank tests are indicated.

Characterizing the immune function of *KYNU*

KYNU encodes kynureninase, a metabolic enzyme on the kynurenine pathway [98]. This pathway catalyzes the degradation of tryptophan, which leads to the production of nicotinamide adenine dinucleotide (NAD⁺). The known molecular function of kynureninase is the cleavage of 3-hydroxykynurenine to 3-hydroxyanthranilic acid (3-HAA). No paper has shown that *KYNU* functions as an immune checkpoint until now, which may be related to the difficulty of experimental verification. However, when we applied the *ab initio* GO method with IPA, the Canonical Pathways analysis further indicated that *KYNU* is prominently associated with immunity (Fig. 22).

The most significantly affected pathways include Th1/Th2 activation, cytokine signaling, T cell differentiation, NK cell signaling, autoimmunity, and inflammatory pathways. The goodness-of-fit $-\text{Log}(p\text{-value})$ scores for these pathways were extremely high, exceeding 20; that is, $p < 1\text{E-}20$. Notably, the activation z -scores for these pathways were strongly positive (3.280–7.042), except that the *PD-1* pathway was significantly negative (−3.395). This indicates that, whereas the *KYNU* gene is co-regulated with most of the immune regulatory pathways, only the *PD-1* pathway displays a mutually exclusive alteration pattern. Such mutual exclusivity is a strong indication that *KYNU* has the redundant/same role as the *PD-1* pathway. Moreover, the Disease and Functions analysis showed that *KYNU* is intensely associated with lung cancer ($p = 4.40\text{E-}54$, $z = 2.581$). Thus, the overexpressed *KYNU* and the increased production of 3-HAA may be responsible for the immune dysfunction, which leads to the immune evasion of the cancer cells, resulting in a worse prognosis.

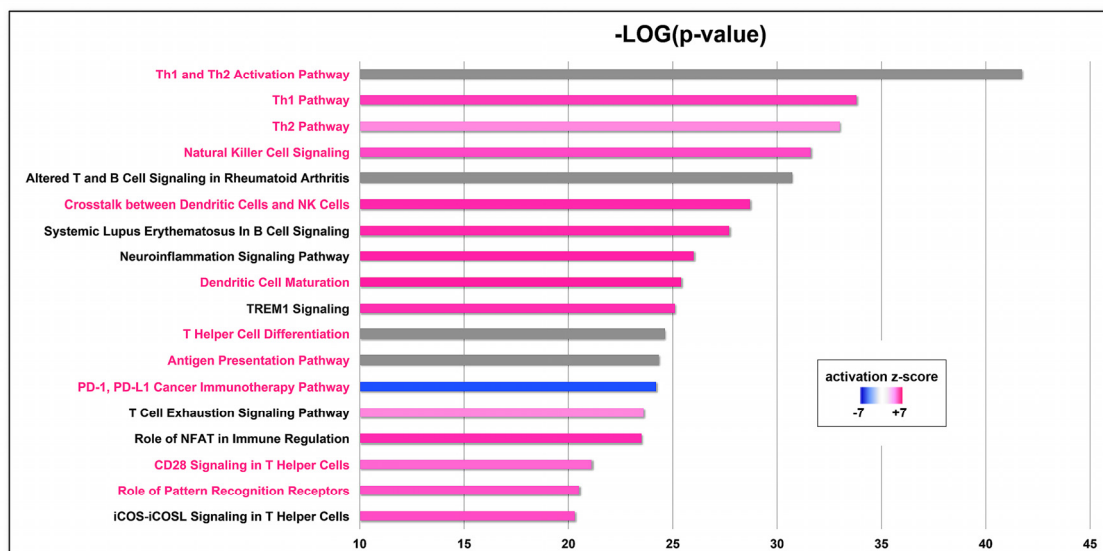


Figure 22 | *Ab initio* GO analysis for *KYNU* using the Canonical Pathways analysis of IPA. Items relevant to the text description are shown in red. The lengths of bars express the goodness of fit to each pathway. Red and blue bars represent positive and negative activation, respectively. The magnitudes of z -scores are shown by the color scale bar in the inset. Grey bars: z -scores not determined.

In addition to the overall findings of IPA, the *ab initio* GO method identified the genes that have the highest *MI* with *KYNU*. The gene with the most *MI* was *TIM-3 (HAVCR2)* [99]. This gene is an immune checkpoint, mainly expressed on T cells, and suppresses immunity [100]. To compare network functions between *KYNU* and *TIM-3*, we searched for genes exhibiting the most potent interaction with these genes. In comparing the 100 genes showing the most *MI* with *KYNU* and the 100 genes showing that with *TIM-3*, as many as 72 genes were in common. Given this strong commonality, it is conceivable that there is a critical relationship between the gene functions of *KYNU* and *TIM-3*. Thus, these findings further demonstrate that *KYNU* is substantially involved in anti-cancer immunity in coordination with *TIM-3*.

In contrast to the pathogenic role of *KYNU*, STAIC revealed that *HAAO* had a positive effect on the prognosis of cancers. These cancers include PAAD, kidney renal clear cell carcinoma (KIRC), and colon adenocarcinoma (COAD). *HAAO* encodes the downstream enzyme of kynureninase (i.e., 3-hydroxyanthranilate 3,4-dioxygenase), which oxidizes 3-HAA to 2-amino-3-carboxymuconic semialdehyde. The use of the *ab initio* GO method demonstrated that *HAAO* is also involved in immunity. These observations collectively imply that *HAAO* antagonizes the pathological effect of *KYNU* by washing out 3-HAA, which leads to the restoration of the anti-cancer immunity. In conclusion, many observations support that *KYNU* and 3-HAA respectively promote cancer development as an immune checkpoint gene and molecule, both of which might be used as tumor markers. Consistent with our predictions, 3-HAA is considered to suppress immunity [101].

In line with the informatics observations above, the immune regulatory function has been recently ascribed to the kynurenine pathway. *IDO1* encodes the first and rate-limiting step enzyme of this pathway, namely indoleamine 2,3-dioxygenase 1, which catalyzes the O₂-dependent oxidation of L-tryptophan to N-formylkynurenine. *IDO1* and related genes (*IDO2* and *TDO*) have been implicated in immune surveillance and tumor progression [102]. Clinical tests have been conducted to evaluate the effects of anti-cancer therapeutics targeting *IDO1* [103]. However, pivotal phase-3 trials (NCT02752074, NCT03386838, and NCT03417037) had negative results and were halted. Although surprising, these failures are consistent with our informatics findings. In comparison with *KYNU*, STAIC showed that *IDO1* and isozymes were altered much less frequently and/or with much lower significance in tumors such as those of LUAD (Fig. 20c), PAAD, and LUSC. These observations demonstrate that, at least when we focus on the immune checkpoint function in cancer, *KYNU* is an essential molecule on the kynurenine pathway.

***KYNU* as a potential tumor marker and a therapeutic target**

The gene product of *KYNU* is a metabolic enzyme, kynureninase, and it would thus be ideal as a tumor marker and therapeutic target. Inhibitors of *KYNU*, such 2-amino-5-methyl-S-phenyl cysteine S,S-dioxide and 2-amino-4-[3'-hydroxyphenyl]-4-hydroxybutanoic acid, have been developed [104].

However, they are chiefly used to treat neuropsychiatric diseases, and there have been no reports that they are applied to cancer treatment. By optimizing the chemical structure of inhibitors, it might be possible to develop their applications for cancer. Especially, pancreatic cancer has a very poor prognosis, and the indication for anti-*PD-1* antibody (pembrolizumab) is limited. In this respect, the involvement of *KYNU* in the prognosis of pancreatic cancer is evident (Fig. 21a), and the number of affected patients is not small. Thus, the development of *KYNU* inhibitors may lead to remarkable advances in the treatment of pancreatic cancer.

***KYNU* gene network supporting the immune function**

From the *ab initio* GO calculation, the distribution of *MI* of the *KYNU*-neighboring gene was shown to follow a Pareto distribution. This is consistent with the empirical findings that immune genes generally follow a Pareto distribution. That is, if we set $G_i = \textit{KYNU}$, then *MI* with another gene G_j of the j -th ranked interaction is $MI(G_i;G_j) = 33.7j^{-0.245}$ ($R^2 = 0.9559$), which is consistent with the above description. Thus, the structural features of the gene network again support the involvement of *KYNU* in immune function. The Pareto distribution demonstrates that the interaction of *KYNU* with the functionally related genes is more potent than those in the infocanonical distribution (Appendices 17 and 18), which is expected under natural conditions.

On the basis of the characteristic network structure, we suppose that the immune network of *KYNU* has strengthened the gene association in the vicinity within the network at a rapid evolutionary velocity for the fierce battle against pathogenic microorganisms. An analytical model that may explain this evolutionary process is presented in Appendix 28. In the context of distinguishing between self and non-self, *KYNU*'s immune checkpoint function may be associated with the evolution of the immune system against pathogenic microorganisms.

Chapter summary

By developing two informatics programs, the *ab initio* GO method and STAIC, we found *KYNU* as a potential immune checkpoint in cancers such as LUAD and PAAD. The *ab initio* GO method reveals the pathological role of genes, even when the *wet* experiment is difficult to perform. In the near future, the validity and availability of the *ab initio* GO method could be determined through the multidisciplinary delineation of the *KYNU* function.

We hope that our informatics programs will lead to the identification of new disease-related genes other than *KYNU* and the development of treatments and clinical laboratory tests for them. The target disease genes will include those challenging to experiment with, such as non-coding RNAs and genes of the nervous system and immune system. Finally, the medical advances thus brought about will confirm the correctness of the *ab initio* GO method, supporting the assumption that gene networks are formed as the LCIO based on the intergenic influential force in evolution.

5. Conclusions and Discussion

This paper discovered the influential force, which is a force based on probability, and elucidated its characteristics as a unifying force. We showed through many examples that the influential force exists in both physical and biological systems. In the study, we developed statistical mechanics that deals with MI in microstates and proceeded with analysis based on the statistical mechanics. As a result, for physical systems, we obtained a formula that describes the four natural forces in a unified manner. We further explained many critical issues in modern physics, including spontaneous symmetry breaking of the Higgs field and its relationship with entanglement entropy. Meanwhile, for biological systems, we demonstrated that the influential force endows genes with much network information, which led to the development of the *ab initio* GO method and the identification of a novel immune checkpoint, *KYNU*. In this way, we found answers and clues for many previously intractable problems. The existence of the influential force has become clear from the abundant findings described in this paper.

Statistical mechanics and probability theory of MI

The realization probability of the MI of two random variables is calculated from the equivalence principle of information and probability. This principle holds in the fields of both probability theory and statistics as follows.

1) In probability theory, the following formula holds:

$$p = e^{-MI}, \quad MI = -\log p. \quad (241)$$

This corresponds to the realization probability $p = \exp(-I)$ of the self-information I of one random variable. In this case, the p -value does not depend on the sample size. We used this formula to derive the state probabilities of physical systems.

2) In statistics, the following formula holds asymptotically in the limit of large N :

$$P_F = e^{-N \cdot MI}, \quad MI = -\frac{1}{N} \log P_F, \quad (242)$$

where P_F is Fisher's exact probability. P_F depends on the sample size N and represents the statistical significance of MI . In this paper, we applied this formula to the analysis of intergenic interactions using large-scale data.

$p = \exp(-MI)$ demonstrates that it is natural for informatons not to share much information, and there is a spontaneous tendency to reduce MI . This shows the trend of increasing the joint entropy of two informatons, which follows the second law of thermodynamics. However, in most cases, MI takes a certain positive value. It is therefore presumed that a force exists and contributes to the generation of MI , thereby antagonizing its decreasing tendency. Hence, we verified the presence of the force in various systems.

Spontaneity of MI

In all the cases examined, MI occurs spontaneously by the mechanisms described in (1) and (2) below.

(1) MI due to microstate fluctuations

The influential force is based on the statistical mechanics and probability theory of MI. Traditionally, MI analysis methods for physical systems have been limited. However, we developed a method of calculating high-dimensional MI between informatons with multi-dimensional information. Whether random variables are treated discretely or continuously, the MI of two informatons X and Y is generated from the fluctuations in their microstates.

As for discrete random variables, we investigated classical physical systems and genetic systems. In these examples, MI is calculated as the sum of micromutual information MI_{kl} :

$$MI = \lim_{m,n \rightarrow \infty} \sum_{k=1}^m \sum_{l=1}^n MI_{kl} \quad (243)$$

Here, the condition that $MI > 0$ is that the fluctuation factor $\varepsilon_{kl} = p(X_k, Y_l)/[p(X_k)p(Y_l)]$ and $MI_{kl} = p(X_k, Y_l)\log\varepsilon_{kl}$ fluctuate in the microstates. We considered the case of mixing two kinds of inert gas as a classical system. At this time, molecular collisions due to mutual diffusion cause fluctuations in the microstates' probabilities, such as those of position, momentum, and energy. MI between molecules then arises, and the gas molecules begin coordinated movements. Meanwhile, for the genetic system, we discussed the process of gene mutation in evolution. In this case, fluctuations arise in the microstate probabilities of the information expressed from two genes, and MI is generated.

As for continuous random variables, we dealt with the probability density functions of quantum systems. If \vec{x} and \vec{y} are respectively the position vectors of X and Y , then MI in the quantum system can be calculated by integrating the MI density, MI_d , corresponding to MI_{kl} as

$$MI = \iint MI_d d\vec{x} d\vec{y} \quad (244)$$

Here, the condition that $MI > 0$ is first that the fluctuation factor $\varepsilon_{xy} = p(x, y)/[p(x)p(y)]$ corresponding to ε_{kl} and $MI_d = p(x, y)\log\varepsilon_{xy}$ fluctuate in the microstates. In addition, the subsystems X and Y must be independent of each other in the initial state. In this paper, two examples were given as situations that satisfy these conditions. First, in relativistic scattering, two particles are mutually independent before the scattering. Second, in the exchange interaction of two identical quanta, if one quantum appears in the vacuum as a delta function at a certain point in time, the two quanta are independent at the time of appearance. Under these mutually independent conditions, $MI(X;Y) = 0$, and thereafter, MI becomes positive owing to the fluctuations of the microstates.

In all cases examined, for either discrete or continuous systems, the microstate fluctuations occur naturally (and in many instances randomly), and MI is thus supposed to arise spontaneously.

(2) MI due to quantum entanglement

In addition to the above examples, we took up the entanglement of quantum superposition. We consider that two quanta are mutually independent in the initial state, and we assume that the entanglement produces a composite system in the pure state. If the entanglement entropy is S_{EE} , then the interaction energy ΔU per one quantum is described by MI as

$$\Delta U = -\frac{1}{2}k_B T \cdot MI = -k_B T \cdot S_{EE} . \quad (245)$$

The large negative value of arising ΔU indicates that the entanglement is likely to occur, and indeed, the entanglement exists universally. Summarizing (1) and (2) above, it is concluded that MI occurs spontaneously through either microstate fluctuations or quantum entanglement.

Information distance and the influential force

To formulate the force between informatons, we introduced the information distance I_o and a coordinate system based on the distance. I_o represents the difficulty of information transmission between two informatons and is defined as

$$I_o(x,r) = \alpha x + \frac{1}{2}r , \quad (246)$$

where x is the conventional distance and r is the information metric. Then, αx is the dimensionless path length of the particle mediating the force between quanta in the external spacetime. Meanwhile, r is a measure that quantitatively expresses the difference in the internal state between informatons and satisfies the metric axiom. In physical systems, $r/2$ expresses not only the potential difference and phase difference between the quanta but also the dimensionless path length in the information metric spacetime. In cases other than physical systems, the information distance can be applied by taking the limit of $\alpha x \rightarrow 0$ and setting $I_o = r/2$.

I_o is associated with the MI of informatons as

$$I_o(x) = \frac{1}{2} [H(X,Y) - MI] + \alpha x . \quad (247)$$

We here remark on the energy change ΔU of the closed composite system XY . $\Delta U = k_B T \cdot \Delta H(X,Y)$ is likely to take a negative value, and the joint entropy $H(X,Y)$ tends to decrease. Therefore, when x is constant, the information distance I_o tends to decrease as MI increases. In other words, the spontaneous increase in MI can be regarded as a state in which an attractive force acts between informatons. Meanwhile, when applying the metric symmetry, αx becomes equivalent to $r/2$ and x tends to become shorter, reducing I_o and energy. This again indicates the action of the attractive force.

The influential force includes a repulsive force in addition to the attractive force. The cause of the repulsion is the recovery of entropy, and the magnitude of the repulsive force is proportional to the square of the attractive force. This repulsive force is the same phenomenon as level repulsion under the condition that the energy and entropy of quanta can be regarded as equivalent.

Metric symmetry and the unifying force

I_O is the least-action path length within the information coordinate spacetime (I_O -spacetime). On the basis of additivity [8], we define S_{uni} as the sum of the actions of all natural forces. Then,

$$\frac{S_{uni}}{\hbar} = I_O = -MI, \quad (248)$$

where $\delta S_{uni} = 0$. The unifying force is derived from this unified information distance I_O .

αx and $r/2$ are the information distances in the external and internal spacetimes, respectively. We introduced metric symmetry, which considers the approximate equivalence of these distances. The metric symmetry is a probabilistic symmetry based on the canonical distribution of the three information distances (i.e., I_O , αx , and $r/2$). As a result of its introduction, the universal gauge symmetry is derived, which allows a unified expression of the four natural forces; that is, $F'_{att}(I_O) = -1$ and $F'_{rep}(I_O) = 2$. The influential force is then expressed either mechanistically or probabilistically as

$$F'(I_O) = k'_1 F'_{att}(I_O) + k'_2 F'_{rep}(I_O), \quad (249)$$

$$\mathcal{F}' = \mathcal{F}'_{att} + \mathcal{F}'_{rep} = \mathcal{M}(X)\mathcal{M}(Y) [k_1 e^{-I_O} - k_2 \mathcal{M}(X)\mathcal{M}(Y) e^{-2I_O}]. \quad (250)$$

In turn, the unifying force in the conventional coordinate spacetime is expressed as

$$F(x) = k_B T \left(\alpha + \frac{\partial}{\partial x} \log \mathcal{F}' \right). \quad (251)$$

At present, no supersymmetric particles have been discovered. Meanwhile, metric symmetry, which interrelates the external and internal spacetimes, is a promising alternative to supersymmetry.

Influential force and thermodynamics

As mentioned above, there is a natural tendency for MI to increase, whether in a physical system or genetic system. It is noted that the joint entropy $H(X,Y)$ of the composite system XY becomes less than that of the initial state entropy $H(X)+H(Y)$, where X and Y are independent. Thus, although the increase in MI does not contradict the increase in entropy for the entire universe, it reduces the overall entropy of the composite system, which is a local system.

We here propose the spontaneous increasing tendency of MI as a novel law of thermodynamics. In other words, \mathcal{F}'_{att} arises naturally between informatons, and the entropy of the composite system thereby decreases spontaneously. At first glance, this is the opposite of the conventional wisdom that systems tend to progress toward increasing entropy. However, this novel rule was valid in all the cases examined. Especially in the case of physical systems, the decrease in entropy of the composite system is accompanied by the reduction of its energy, which highlights the tight coupling between entropy and energy. MI is a kind of entropy, and this new rule will therefore be understandable if we consider that the law of increasing entropy also applies to the MI itself.

Canonicity of the information coordinate system

We have verified from the above perspectives that the influential force explains all the fundamental interactions. Conventionally, the three fundamental forces other than gravity (i.e., the electromagnetic, weak, and strong forces) have been described by scattering theory. However, by applying information theory, we have shown that in relativistic scattering, the potential V is generated through the exchange of information between quanta. At this time, $I_O = \alpha x + r/2$ is shortened by $\Delta r/2 = \beta \Delta V$. Meanwhile, in the case of gravity, I_O for spacetime quanta is expressed using the entanglement entropy S_{EE} as $I_O = \alpha x - 2S_{EE}$. In this case, the gravitational potential V_g is generated by an information exchange through the entanglement of the spacetime quanta, and I_O is shortened by $\Delta r/2 = -2\Delta S_{EE}$. Taken together, the gauge interactions and the quantum entanglement are equivalent information exchanges within the information coordinate spacetime (I_O -spacetime), with both leading to I_O shortening by $\Delta r/2$.

It is noted that all the fundamental forces distort the I_O -spacetime by $\Delta r/2$, just as gravity distorts the conventional coordinate spacetime (x -spacetime). This is a remarkable commonality found for every unified natural force, which supports the canonicity of the information coordinate system. We refer to this general phenomenon as the *relativization of relativity*. This generality also implies that the information coordinate system has successfully incorporated the outcome of superstring theory and suggests that the gauge–gravity correspondence can be resolved within the I_O -spacetime.

Along with the canonicity described above, the ability of the influential force to unify all the natural forces is brought about by adopting the information coordinates. In contrast, when using the conventional coordinate system, it is impossible to express all the natural forces with one mathematical formula, even on the basis of superstring theory. The fact that the unified gauge symmetry of the influential force depends on the information distance I_O suggests that future physics should employ the information coordinate system.

Exploration of physics based on the influential force

Next, by considering the influential force, we addressed multiple issues in modern physics.

(1) Novel exchange interaction as a dark-matter candidate

The universal gauge symmetry of the influential force requires αx (> 0) to formulate $I_O = \alpha x + r/2$. The use of αx allows us to deal with a non-local interaction within the x -spacetime as a gauge force within the I_O -spacetime. We identified a novel exchange interaction between identical particles. This interaction is non-local in the x -spacetime and generates MI as well as the influential force between the particles within the I_O -spacetime. Of note, the energy being represented by αx suggests that the influential force also acts within the x -spacetime. We proposed that a hypothetical boson named the mion mediates this new exchange interaction, and we estimated its energy. Although the calculated

energy is much lower than that of the known exchange interactions, it is almost independent of the distance within the x -spacetime. This is consistent with the non-locality of the new exchange interaction and sharply contrasts with known exchange interactions that work over a very short range. Given the immense possible number of new exchange interactions, the mion is considered a novel candidate for dark matter.

(2) Elucidation of the identity of the Higgs field

The existence of the repulsive term is a distinctive feature of the influential force and leads to solutions for several major problems of the Standard Model. First, we identified a pair of spacetime quanta that interact under quantum gravity according to Hamada's theory. These are the spacetime fermions with opposite spins. By applying the modified Nambu BCS theory and considering the vibration of the background metric field as a phonon, it was shown that an attractive pair potential Δ is generated between the spacetime quanta. In the I_0 -spacetime, this quantum pair is a composite particle corresponding to the Cooper pair; that is, the Higgs boson. Here, Δ acting between quanta is gravity, which can be regarded as the influential force derived from the information exchange between the quanta. At the same time, there is an energy gap due to level repulsion, which gives the vacuum expectation value $\langle\phi\rangle_0 = 2\Delta_0 (= 246 \text{ GeV})$. This level repulsion is equivalent to the repulsive term of the influential force, indicating that the gauge symmetry of the influential force field is spontaneously broken. The Higgs mass of $\Delta_0 = 123 \text{ GeV}$ is then generated; this value is close to the observed value of 125 GeV.

(3) On the equivalence of inertial and gravitational mass

The findings described above show the equivalence of the influential Higgs field and the gravitational field. On this basis, the equivalence of inertial mass and gravitational mass was also explained. When the matter particles interact with the Higgs field to gain mass, MI of the matter particles and Higgs particles is generated. This MI increases the mass of the Higgs boson by the mass obtained by the matter particle. The increase in the Higgs mass means that the influential force potential and the gravitational potential become steeper, indicating that gravitational mass is created. From this discussion, we explained the rationale for why the inertial and gravitational masses become equal.

(4) On the hierarchy problem

The hierarchy problem is that the Higgs boson is so much lighter than the Planck mass. We solved this problem by again considering the influential Higgs. As in (3), the Higgs mass increases by the same magnitude as the mass acquired by the matter particles, which results from the deepened bottom of the influential Higgs potential. For all natural forces, the influential force potential $\varphi'(I_0)$ represents the cumulative probability amplitude of exchanging information between the origin and a point at the information distance I_0 . Therefore, $\varphi'(I_0)$ can be converted into potential energy by

$$V'_H = -hc[\varphi'(I_O)]^2, \quad (252)$$

which applies to the Higgs potential. It is noted that this probability-based potential does not diverge. Therefore, it is unnecessary to apply renormalization to the Higgs mass, and the hierarchy problem is thus avoided. Similarly, it is worth noting that the gravitational potential expressed by the influential force potential $\varphi'(I_O)$ also does not diverge. This implies that the influential force \mathcal{F}' can faithfully describe quantum gravity.

(5) Influential Higgs inflation

Next, the influential Higgs inflation was discussed to clarify the importance of the repulsive force to the cosmic structure. The influential Higgs potential $\varphi'(I_O)$ does not diverge and becomes flat as I_O goes to infinity. This is in sharp contrast to the case of the standard Higgs potential V_H , which diverges rapidly and is shaped like a Mexican hat. It is noted that because the influential Higgs potential expresses gravity from the beginning, it does not require an artificial R^2 term or a non-minimal coupling to gravity. Nevertheless, the influential Higgs inflation model reproduces measurements of the cosmic microwave background well ($n_s = 0.966$, $r = 0.0022$). This suggests that the Higgs field, gravitational field, and inflaton field are originally the same, all being derived from the influential force field. Considering this scenario, spontaneous symmetry breaking is thought to have occurred owing to the strengthening of the repulsive force in the process of inflation. It is possible that the influential force played a major role in cosmogenesis.

(6) Influential Higgs and quantum entanglement

Cosmology based on the influential force field has succeeded in incorporating the outcome of superstring theory. It is known that the pair potential Δ between electrons, which corresponds to the attractive force of BCS theory, is proportional to the quantum entanglement entropy S_{EE} . Here, it was shown that a proportional relationship $S_{EE} = \beta\Delta/2$ holds even in the case of the influential Higgs. The probabilistic influential force potential $\varphi'(I_O)$ of the information coordinate system is minimized at the equilibrium information distance $I_e = (\log 2)/4$. This I_e is the information distance corresponding to the Schwarzschild radius when the spacetime quantum of the origin is regarded as a black hole. According to the Ryu–Takayanagi formula, $S_{EE} = I_e = (\log 2)/4$ is the minimum number of entanglements that one spacetime quantum can take and equals the minimum surface area and minimum gravity of a black hole. Thus, the spacetime quantum existing at the origin is a black hole whose radius is $I_e = S_{EE}$. In turn, the above spacetime quantum pair is a Higgs boson and is a black hole dimer, separated by I_e .

(7) Repulsive influential force as a dark-energy candidate

All Higgs bosons present in a vacuum are identical, and our novel exchange interaction probably forms the Higgs condensate. The phase coherence created at this time implies the shortening of I_O between

spacetime quanta by $\Delta I_O = -\Delta MI$; that is, the distortion of the I_O -spacetime. A gravitational force field is thus induced. Again, because there is a large number of information exchanges between the identical Higgs bosons, the energy of the new exchange interaction may be the origin of dark matter. Meanwhile, by considering the influential force of many-body systems, it was shown that a cosmic repulsive force, which is equivalent to dark energy, is generated. Whereas the magnitude of the attractive force is $\Delta (= k_B T \cdot MI)$, that of the repulsive force is 2Δ . Therefore, dark energy is two-thirds (67%) of the cosmic energy. This proportion is close to the measurement of 68%.

In summary, spontaneous symmetry breaking is explained by considering the influential force field. Furthermore, the Higgs field, gravity, inflation, and entanglement entropy are explained in a unified manner, which provides clues for clarifying the essential structure of a universe that has dark matter and dark energy.

Exploration of biology based on the influential force

(1) Gene information offered by the influential force

The influential force acting between genes functions in the same way as that acting between objects in a physical system. The intergenic MI is spontaneously generated by mutation-induced fluctuations of MI_{kl} . By combining information theory and population genetics, we proved that the intergenic MI is equal to the epistatic fitness, ε_{two} . As an effect of this intergenic MI, the influential force \mathcal{F}_{att} affects the existence probability of gene information according to

$$\mathcal{F}_{att} = e^{\varepsilon_{two}} = e^{\Delta MI_{inh}}. \quad (253)$$

Using the Moran process, we revealed how the MI of genes becomes fixed in the population. Even though most individual mutations are harmful, if the intergenic MI compensates for the harm and is sufficiently favorable to the organism's survival, then MI is fixed. On the basis of this MI of two genes created by the influential force, all the other genes gather around a gene G_i to form a highly multi-dimensional network represented by the LCIO; that is, $\varphi[\vec{r}(G_i)] = \sum c_{ij} \cdot \chi_{ij}$.

The LCIO imparts a substantial amount of highly diverse network information to genes, far exceeding the amount of molecular information encoded by the structure of gene products. This is because gene nodes have widely distinct properties, and permutations generate immense information. Thus, most of the information expressed by genes is encoded as the network structure; that is,

$$H(N_i) \gg H(M_i). \quad (254)$$

The substantial amount of network information can explain the relationship between gene evolution and phenotypic evolution, which has traditionally been a major problem in evolutionary biology. Therefore, to clarify the function of a gene, it is essentially more important to know the network function of the LCIO than to know the molecular function of the gene product.

(2) *Ab initio* GO method

We developed the *ab initio* GO method to advance medicine and demonstrate that genes perform functions based on the LCIO. This method uses large-scale genomic data of the TCGA, computes the intergenic MI at a super-high resolution, and predicts a wide range of gene properties within minutes. In developing the *ab initio* GO method, we discovered a series of mathematical theorems that use MI_{kl} and MI as statistical variables. These theorems include 1) the MI_{kl} summation theorem, 2) the infocanonical distribution of MI_{kl} , 3) frontier information theory, and 4) the equivalence principle of information and probability.

The *ab initio* GO method uses the above principles, integrates large-scale, multi-dimensional genomic data, and calculates MI as a unified indicator of gene interaction. The network information $H(N_i)$ of the gene G_i is obtained by comprehensive calculation and superposition of all two-gene network information $H(N_{ij}) = [MI_{overall}(G_i;G_j), H_{sgn_overall}(G_i,G_j)]$ in multi-dimensional space as

$$H(N_i) = \sum_{j \neq i} H(N_{ij}) \quad (255)$$

Finally, applying IPA allows the interpretation of a wealth of information regarding the wide range of functional properties of the LCIO, from molecular functions to pathophysiological relevance.

(3) *Identification of a potential immune checkpoint, KYNU/kynureninase*

The *ab initio* GO method allows us to quickly decipher gene functions, even when the experimental elucidation is difficult. In addition, we developed another program, STAIC, to facilitate cancer gene discovery. Combining the two strategies, we identified a *potential immune checkpoint, KYNU*. Our analysis showed that the overexpressed *KYNU* could lead to the abrogation of the immune checkpoint, resulting in a prominently worse prognosis in several types of cancers. In particular, the involvement of *KYNU* in clinical outcomes is highly evident in LUAD and PAAD. *KYNU* would thus be ideal for these representative refractory cancers as a tumor marker and therapeutic target. The development of *KYNU* inhibitors is highly awaited to advance cancer treatment.

From the success of the *ab initio* GO method, it was demonstrated that an influential force-based gene network is indeed formed as the LCIO. As a future application of informatics to medicine with LCIO in mind, we propose promoting the functional genome project by applying comprehensive network analysis technology. This project will provide a great deal of novel valuable information about diseases of unknown etiology. Hopefully, various molecular targeted therapies such as anticancer drugs targeting *KYNU/kynureninase* will be developed.

Concluding remarks

The preceding findings demonstrate that the influential force explains the laws of physical and biological systems in a unified manner. Although genes and physical objects appear absolutely different, they have commonality as informatons. Notably, between the informatons within each system, an equally expressed influential force acts, resulting in the formation of many-body systems.

Despite the commonality in each system, they are different in terms of whether the components of the constructed many-body systems are homogeneous or heterogeneous. As a result, exchange interactions occur in the physical system, the accumulated energy of which is large enough to affect the entire structure of the universe. Meanwhile, in the multi-gene system, giving each gene substantial network information enabled the creation of organisms with complex information processing machinery that has survived a long geological age. Although the representations of the two many-body systems are absolutely different, they are common in that a vast amount of MI is generated. We can understand the emergence of these enormous amounts of MI in the unified context of the new law of thermodynamics that we have proposed; that is, MI arises between informatons spontaneously. Even with different features and shapes, the generation of two massive volumes of MI by the influential force demonstrates a surprising commonality between the totally different beings of the universe and life.

Finally, from the universality of information and probability, we can presume that the influential force also affects relationships between other informatons such as humans. It is conceivable that the preceding arguments for the interaction between informatons will also apply to human relationships. This is because we have successfully described interactions between genes, which are sufficiently complex as those between humans. In this case, the influential force acting on the mutual relationship between humans will appear in two forms: the attractive and repulsive forces. Through the action of two opposing forces acting between individuals, the influential force provides power to construct the network of human beings, thus building society and bringing about its time evolution; that is, the history of humanity.

In finishing this manuscript, we await further studies that will seek to validate the conjectures presented in this paper.

Appendices

1. Distances, metrics, and coordinates

This paper uses several kinds of coordinate: the information metric r , conventional distance x , information distance I_O (either original or reduced), and probability p .

(1) Information metric, r

The information metric r is also known as the variation of information [10]. This r is a true metric because it satisfies the axioms of the metric space; that is, triangle inequality, non-negativity, indiscernibility, and symmetry. r is a metric of the distance between informatons. We associate r with the potential $V(x)$ and partition function Z' by $r = 2\beta V(x) + 2 \log Z'$ [see Chapter 2, Section (c)].

(2) Conventional distance, x

The conventional distance x is a physical distance in the conventional coordinate spacetime without gravity, measured in terms of the velocity of light in a vacuum.

(3) Information distance, I_O

(Original) information distance

The information distance I_O represents the difficulty of information transmission between two informatons. If the probability density of information transmission from informaton X to Y is p , then their information distance is expressed as

$$I_O = -\log p. \quad (\text{A1})$$

By setting $Z' = 0$, I_O is written using the conventional distance x and information metric r as

$$I_O(x, r) = \alpha x + \frac{1}{2} r, \quad (\text{A2})$$

where α is a distribution parameter and αx and $r/2$ are the lengths of least action paths in the conventional spacetime (x -spacetime) and information metric spacetime (r -spacetime), respectively.

Reduced information distance

For the oscillation of informatons, we use the reduced information distance $I_O := (\alpha x + r/2)/2$, which corresponds to the reduced mass-energy of the quantum.

(4) Probability, p

Suppose a one-dimensional spacetime with probability density p as a coordinate, where p satisfies the axioms of the metric space. If the probability density $p = p(X)$ of the random variable X follows a canonical distribution, then p is a strictly monotonically decreasing function of X , and X and p are injective. Here, X applies, for example, to the information distance I_O , information metric r , potential V , and MI. In particular, I_O and p are bijective and thus reversible.

2. Probability distribution of MI

(1) Realization probability of MI

We consider that two systems X and Y interact and share MI, where the system is under isothermal conditions. Let W_X and W_Y be the numbers of states of X and Y , respectively, and W_{MI} be the number of states of MI . Here, the number of states W_{tot} when no information is shared is expressed as

$$W_{tot} = W_X \times W_Y . \quad (\text{A3})$$

When the information is shared, the number of states W of the composite system XY is expressed as

$$W = \frac{W_X}{W_{MI}} \times W_{MI} \times \frac{W_Y}{W_{MI}} . \quad (\text{A4})$$

Then, the spontaneous probability $p(MI)$ that the magnitude of MI becomes MI is

$$p(MI) = \frac{W}{W_{tot}} = \frac{1}{W_{MI}} = e^{-MI} . \quad (\text{A5})$$

We call this the infocanonical distribution of MI , which represents the "equivalence principle of information and probability" in the field of probability theory (Appendices 17 and 20). Finally, $p(MI)$ is equal to the inverse of the attractive influential force; that is,

$$p(MI) = \mathcal{F}_{att}^{-1} . \quad (\text{A6})$$

(2) Repulsive force generated by MI

We next consider each of systems X and Y mentioned above. We assume that in the initial state, X and Y share information, where the magnitude of the MI is MI_0 . First, we consider only X . The initial number of states $W_X(MI_0)$ of X is then expressed as

$$W_X(MI_0) = \frac{W_X}{W_{MI_0}} . \quad (\text{A7})$$

When the MI decreases by ΔMI , the number of states of X is

$$W_X(MI_0 - \Delta MI) = \frac{W_X}{W_{MI_0 - \Delta MI}} = W_X(MI_0) \times W_{\Delta MI} . \quad (\text{A8})$$

The MI tends to decrease because $W_X(MI_0 - \Delta MI) > W_X(MI_0)$. Let $p_X(MI_0)$ be the realization probability of X when the magnitude of MI is MI_0 . The relative probability of a decrease in MI is then expressed as

$$\frac{p_X(MI_0 - \Delta MI)}{p_X(MI_0)} = \frac{W_X(MI_0 - \Delta MI)}{W_X(MI_0)} = W_{\Delta MI} = e^{\Delta MI} . \quad (\text{A9})$$

Likewise, MI tends to decrease also for Y , and its relative probability is $e^{\Delta MI}$ equally. Thus, a repulsive force \mathcal{F}_{rep} acts between X and Y , and its magnitude is the product of both changes in the probability:

$$\mathcal{F}_{rep} = e^{\Delta MI} \times e^{\Delta MI} = e^{2\Delta MI} . \quad (\text{A10})$$

If the initial MI is zero, then $\mathcal{F}_{rep} = e^{2MI}$. This is equal to the square of the attractive force $\mathcal{F}_{att} = e^{MI}$.

(3) Rank distribution of MI

Let G_j be the gene whose MI shared with G_i ranks the j -th largest. $MI(G_i;G_j)$ is then expressed as

$$MI(G_i;G_j) = -A \log(j) + B, \quad (\text{A11})$$

where A and B are constants. The rank j is then expressed as

$$j = \exp\left[\frac{B - MI(G_i;G_j)}{A}\right]. \quad (\text{A12})$$

3. MI and natural force potentials

We consider a system of two informatons X and Y . Let $H(X)$ and $H(Y)$ be the information entropy of X and Y , respectively. Let also $S(X,Y)$ be their joint thermodynamic entropy, and $\Delta V(x)$ be the potential difference between X and Y . We here examine a situation where $H(X)$ and $H(Y)$ are constant when the mediator particles' entropy is summed [Eq. (36)]. The internal energy U of the composite system XY is then expressed as

$$U = F + TS(X,Y) = -k_B T \log Z + k_B T [H(X) + H(Y) - MI], \quad (\text{A13})$$

where F is the Helmholtz free energy. $\Delta F = -k_B T \Delta \log Z = 0$ as we assume the thermodynamic equilibrium condition. Moreover, if we suppose a static setting, such as the static potential generated by mediator particles, then $\Delta V(x)$ is equal to ΔU . Thus,

$$\Delta U = \Delta V(x) = -k_B T \Delta MI. \quad (\text{A14})$$

4. Probabilistic influential force potential and Morse potential

We numerically compare the probabilistic influential force potential $\varphi(r)$ and the Morse potential $V_M(x)$. In the 1Σ state of the hydrogen molecule, $D = 7.964 \times 10^{-19}$ J and $\alpha = 1.85 \times 10^{10}$ m⁻¹, the reduced mass is half that of the proton $\mu = 8.365 \times 10^{-28}$ kg, and there are 19 energy levels. Using these parameters, the vibration entropy is calculated as $S = 1.21 \times 10^{-30}$ J/K.

In addition, because the standard molar entropy of the hydrogen molecule in the standard state is 130.684 J/mol·K and that of the hydrogen atom is 114.713 J/mol·K, the MI shared by the two atoms in the hydrogen molecule is calculated as $(2 \times 114.713 - 130.684) / (k_B \times 6.02 \times 10^{23}) = 11.886$ nat, and the information metric for the nuclei is $114.713 \times 2 / (k_B \times 6.02 \times 10^{23}) - 2 \times 11.886 = 3.844$ nat. This is the equilibrium information metric r_e for the two atoms. We assume that this corresponds to the equilibrium conventional distance $x_e = 8.9 \times 10^{-11}$ m of $V_M(x)$. Then, based on the metric symmetry, 1 nat can be converted to 2.32×10^{-11} m.

When we represent the internuclear distance x in terms of meters, the Morse potential [31] is expressed as

$$V_M(x) = 7.964 \times 10^{-19} \{ \exp[-3.7 \times 10^{10}(x - 8.9 \times 10^{-11})] - 2 \exp[-1.85 \times 10^{10}(x - 8.9 \times 10^{-11})] \}, \quad (\text{A15})$$

Using the relationship that $\acute{\alpha} = 1.85 \times 10^{10} \text{ m}^{-1}$ corresponds to 0.429 nat^{-1} , we can represent r in terms of nat. Then, as an alternative representation of the Morse potential, we obtain the description of the probabilistic influential force potential $\varphi(r)$ between the hydrogen nuclei as

$$k_B T \varphi(r) = 7.964 \times 10^{-19} \{ \exp[-0.858(r - 3.84)] - 2 \exp[-0.429(r - 3.84)] \}, \quad (\text{A16})$$

which follows $V_M(x) = k_B T \varphi(r)$. If this formula is valid, then we can expect the number of information levels of the anharmonic oscillation to be 19 and that the vibration entropy is $S = 1.21 \times 10^{-30} \text{ J/K}$. These constants are consistent with those of the Morse potential. The above calculations thus suggest that the Morse potential $V_M(x)$ may numerically conform to the probabilistic influential force potential $\varphi(r)$.

5. Probabilistic influential force potential and dispersion force

We explore whether the dispersion force acting between the helium atoms conforms to the probabilistic influential force potential $\varphi(r)$. When we performed a fitting analysis, the following function $V_{He}(x)$ provided a good fit to the potentials between two monoatomic helium molecules calculated by Hellmann [32] and Hurly and Mehl [33]:

$$V_{He}(x) = 1.520 \times 10^{-22} \{ \exp[-4.192 \times 10^{10}(x - 2.964 \times 10^{-10})] - 2 \exp[-2.096 \times 10^{10}(x - 2.964 \times 10^{-10})] \}, \quad (\text{A17})$$

where x is expressed in terms of meters. If we accept this formula as a representation of the dispersion force acting between the helium molecules, because $\acute{\alpha} = 2.096 \times 10^{10} \text{ m}^{-1}$ is converted to 0.4836 nat^{-1} , the probabilistic influential force potential is expressed as

$$k_B T \varphi(r) = 1.520 \times 10^{-22} \{ \exp[-0.9726(r - 12.78)] - 2 \exp[-0.4863(r - 12.78)] \}, \quad (\text{A18})$$

where r is expressed in terms of nat.

Comparing the above two formulas with those of the hydrogen molecule in the previous section, the equilibrium information metric r_e is 3.33 times longer, whereas the dissociation energy D is one-5238th as large. The interaction in the helium dimer is based on the dispersion force; thus, the obtained formulas reflect that the dispersion force is much weaker than the covalent bond. As described above, it is feasible that the dispersion force acting between monoatomic molecules also conforms to the probabilistic influential force potential $\varphi(r)$.

6. Micromutual information summation theorem

We describe a theorem that allows the calculation of highly multidimensional MI. Let MI be the MI of the $m \times n$ contingency table with respect to two random variables X and Y . Let MI_{kl} be the MI, defined as the MI of the 2×2 contingency table with respect to a cell (k, l) . Additionally, let $p(X_k, Y_l)$ be the relative frequency of the cell (k, l) , and let $p(X_k)$ and $p(Y_l)$ be the marginal relative frequencies of the k -th column and l -th row, respectively. We here examine the equivalence of MI and the sum of MI_{kl} over all cells by taking their difference as

$$\begin{aligned}
 MI - \sum_{k=1}^m \sum_{l=1}^n MI_{kl} &= \sum_{k=1}^m \sum_{l=1}^n \left\{ -[p(X_k) - p(X_k, Y_l)] \log \left[1 - \frac{p(X_k, Y_l)}{p(X_k)} \right] - [p(Y_l) - p(X_k, Y_l)] \log \left[1 - \frac{p(X_k, Y_l)}{p(Y_l)} \right] \right. \\
 &\quad - [1 - p(X_k) - p(Y_l) + p(X_k, Y_l)] \log [1 - p(X_k) - p(Y_l) + p(X_k, Y_l)] \\
 &\quad \left. + [1 - p(X_k)] \log [1 - p(X_k)] + [1 - p(Y_l)] \log [1 - p(Y_l)] \right\} .
 \end{aligned} \tag{A19}$$

We now divide each cell of the original contingency table into h equal parts for both directions of rows and columns. If we take the sum of MI_{kl} over all the divided cells, we obtain

$$\begin{aligned}
 MI - \sum_{k=1}^m \sum_{l=1}^n MI_{kl} &= \sum_{k=1}^m \sum_{l=1}^n \left\{ -[hp(X_k) - p(X_k, Y_l)] \log \left[1 - \frac{p(X_k, Y_l)}{hp(X_k)} \right] - [hp(Y_l) - p(X_k, Y_l)] \log \left[1 - \frac{p(X_k, Y_l)}{hp(Y_l)} \right] \right. \\
 &\quad - [h^2 - hp(X_k) - hp(Y_l) + p(X_k, Y_l)] \log \left[1 - \frac{1}{h} p(X_k) - \frac{1}{h} p(Y_l) + \frac{1}{h^2} p(X_k, Y_l) \right] \\
 &\quad \left. + [h^2 - hp(X_k)] \log \left[1 - \frac{1}{h} p(X_k) \right] + [h^2 - hp(Y_l)] \log \left[1 - \frac{1}{h} p(Y_l) \right] \right\} .
 \end{aligned} \tag{A20}$$

We define $\Delta(h)$ as the left-hand side of Eq. (A20). Using L'Hospital's theorem, we calculate the limit of each term on the right-hand side as h tends to infinity. The limits of the first and second terms are then both $p(X_k, Y_l)$. Meanwhile, the third term approaches

$$\frac{h}{2} [p(X_k) + p(Y_l)] - p(X_k, Y_l) - \frac{1}{2} [p(X_k) + p(Y_l)]^2 \tag{A21}$$

and the fourth and fifth terms approach

$$\frac{h}{2} p(X_k) - \frac{1}{2} [p(X_k)]^2, \quad \frac{h}{2} p(Y_l) - \frac{1}{2} [p(Y_l)]^2, \tag{A22}$$

respectively. If we sum these limits and consider that the sum of $p(X_k, Y_l)$, that of $p(X_k)$, and that of $p(Y_l)$ are each equal to 1, then we can prove that the limit of the right-hand side is zero. Therefore, $\lim_{h \rightarrow \infty} \Delta(h) = 0$. Thus, when m and n are sufficiently large,

$$MI \simeq \sum_{k=1}^m \sum_{l=1}^n MI_{kl} . \tag{A23}$$

In conclusion, the highly multidimensional MI can be calculated by taking the total sum of the micromutual information MI_{kl} .

7. MI of informatons that accompanies fluctuations

We calculate the MI of two highly multidimensional informatons, G_X and G_Y , under the condition that a fluctuation occurs in the realization probability of each information. Let the information expressed by G_X and G_Y take at most M and N observable states as multidimensional random variables X and Y , respectively. Under $M, N \rightarrow \infty$, the information levels of these states can be represented by one-dimensional continuous random variables X and Y , respectively. In this case, the realization probabilities of X and Y respectively follow the infocanonical distributions (Appendix 17) as

$$p(X) = \frac{1}{I_X} \exp\left(-\frac{X}{I_X}\right), \quad p(Y) = \frac{1}{I_Y} \exp\left(-\frac{Y}{I_Y}\right), \quad (\text{A24})$$

where I_X and I_Y are the expectation values (i.e., the effective self-information) of X and Y . We also assume X and Y can be evenly split into m and n levels and are discretely distinguished by linear indices k ($1 \leq k \leq m$) and l ($1 \leq l \leq n$), respectively. Then, the realization probability of the k -th information level of X , X_k , follows the infocanonical distribution as

$$p(X_k) = \int_{k-1}^k p(X) dX = [\exp(1/\lambda) - 1] \exp(-k/\lambda), \quad (\text{A25})$$

while that of the l -th information level of Y , Y_l , follows

$$p(Y_l) = \int_{l-1}^l p(Y) dY = [\exp(1/\nu) - 1] \exp(-l/\nu). \quad (\text{A26})$$

where λ and ν are the mean index values of X_k and Y_l , respectively.

We assume that owing to the random fluctuations, the realization probabilities of each level X_k and Y_l change stochastically. We use a fluctuation factor $\varepsilon_{kl} = p(X_k, Y_l) / [p(X_k)p(Y_l)]$, and suppose that ε_{kl} fluctuates close to a value of 1 and that ε_{kl} is expressed by $\varepsilon_{kl} = \gamma_{kl}\delta_{kl}$ using the two factors γ_{kl} and δ_{kl} , which independently fluctuate owing to changes in X and Y , respectively. We let γ and δ be the averages of γ_{kl} and δ_{kl} over k and l , respectively, and additionally assume that both γ and δ follow a normal distribution having a mean of 1 and variance σ^2 .

In this case, the micromutual information MI_{kl} is calculated as

$$MI_{kl} \simeq \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{1-k}{\lambda}} e^{-\frac{1-l}{\nu}} \left[1 + \frac{\left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{-\frac{n+1-l}{\nu}}}{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{-\frac{n+1-l}{\nu}}} + \frac{\left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{-\frac{m+1-k}{\lambda}}}{1 - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{-\frac{m+1-k}{\lambda}}} \right] + \frac{\left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{1-k}{\lambda}} e^{-\frac{1-l}{\nu}}}{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{-\frac{n+1-l}{\nu}} - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{-\frac{m+1-k}{\lambda}} + \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{1-k}{\lambda}} e^{-\frac{1-l}{\nu}}} \right] (\gamma\delta)^2$$

$$\begin{aligned}
& - \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{\frac{1-k}{\lambda}} e^{\frac{1-l}{\nu}} \\
& \left[1 + \frac{1 + \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}}}{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}}} + \frac{1 + \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}}}{1 - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}}} \right. \\
& \left. - \frac{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}} - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}} - \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{\frac{1-k}{\lambda}} e^{\frac{1-l}{\nu}}}{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}} - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}} + \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{\frac{1-k}{\lambda}} e^{\frac{1-l}{\nu}}} \right] \gamma \delta \\
& + \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{\frac{1-k}{\lambda}} e^{\frac{1-l}{\nu}} \\
& \left[\frac{1}{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}}} + \frac{1}{1 - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}}} \right. \\
& \left. - \frac{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}} - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}}}{1 - \left(1 - e^{-\frac{1}{\lambda}}\right) e^{-\frac{m}{\lambda}} e^{\frac{n+1-l}{\nu}} - \left(1 - e^{-\frac{1}{\nu}}\right) e^{-\frac{n}{\nu}} e^{\frac{m+1-k}{\lambda}} + \left(1 - e^{-\frac{1}{\lambda}}\right) \left(1 - e^{-\frac{1}{\nu}}\right) e^{\frac{1-k}{\lambda}} e^{\frac{1-l}{\nu}}} \right], \tag{A27}
\end{aligned}$$

which is simplified as

$$\begin{aligned}
MI_{kl} &= \left(e^{\frac{k-m-1}{\lambda}} - e^{-\frac{m}{\lambda}} + e^{-\frac{m+1}{\lambda}} \right)^{-1} \left(e^{\frac{l-n-1}{\nu}} - e^{-\frac{n}{\nu}} + e^{-\frac{n+1}{\nu}} \right)^{-1} \\
& \quad e^{-\frac{n+1}{\lambda}} e^{-\frac{m+1}{\nu}} \left(e^{\frac{1}{\lambda}} - 1 \right) \left(e^{\frac{1}{\nu}} - 1 \right) (1 - \gamma \delta)^2 \\
& \simeq e^{-\left(\frac{k}{\lambda} + \frac{l}{\nu}\right)} \left(e^{\frac{1}{\lambda}} - 1 \right) \left(e^{\frac{1}{\nu}} - 1 \right) (1 - \gamma \delta)^2. \tag{A28}
\end{aligned}$$

Hence, if m and n are sufficiently large, then

$$\begin{aligned}
MI &\simeq \sum_{k=1}^m \sum_{l=1}^n MI_{kl} \simeq \left(e^{\frac{1}{\lambda}} - 1 \right) \left(e^{\frac{1}{\nu}} - 1 \right) (1 - \gamma \delta)^2 \sum_{k=1}^m e^{-\frac{k}{\lambda}} \sum_{l=1}^n e^{-\frac{l}{\nu}} \\
&= (1 - \gamma \delta)^2 \left(1 - e^{-\frac{m}{\lambda}} \right) \left(1 - e^{-\frac{n}{\nu}} \right). \tag{A29}
\end{aligned}$$

This equation applies to the various high-dimensional informatons whose states fluctuate randomly.

8. MI of free particles in an exchange interaction

(1) MI of free particles in a one-dimensional spacetime

We describe the MI of free quantum free particles that undergo diffusion processes. A case for identical bosons is discussed in the text (Chapter 2, Section (g)). Appendix 8 considers two identical free fermions X and Y with the same mass m . Their position coordinates x and y take values of d and $-d$ at $t = 0$, respectively. Their propagators are then expressed as

$$K_1(x, d; t) = \left(\frac{m}{2\pi i \hbar t} \right)^{\frac{1}{2}} \exp\left[-\frac{m(x-d)^2}{2i\hbar t} \right] \quad \text{and} \quad (\text{A30})$$

$$K_2(y, -d; t) = \left(\frac{m}{2\pi i \hbar t} \right)^{\frac{1}{2}} \exp\left[-\frac{m(y+d)^2}{2i\hbar t} \right]. \quad (\text{A31})$$

The probability density functions for X and Y are calculated as below under the limited conditions $|x - d| < (ht/m)^{1/2}/2$ and $|y + d| < (ht/m)^{1/2}/2$, respectively, where h is the Planck constant.

Let $\psi(x, y, t)$ be the composite wave function of X and Y , which we express as

$$\psi(x, y, t) = A[K_1(x, d; t)K_2(y, -d; t) - K_1(y, d; t)K_2(x, -d; t)], \quad (\text{A32})$$

where A is the normalization constant. When $(ht/m)^{1/2}/1 < 2d$ (i.e., $t < 4d^2m/h$), if $x < d - (ht/m)^{1/2}/2$ and $y > -d + (ht/m)^{1/2}/2$, then $\psi(x, y, t) = 0$. Meanwhile, when $t > 4d^2m/h$, if $x > d - (ht/m)^{1/2}/2$ and $y < -d + (ht/m)^{1/2}/2$, then $\psi(x, y, t)$ can take non-zero values. The constant A is calculated by setting the integral of $|\psi(x, y, t)|^2$ on the rectangle $|x - d| < (ht/m)^{1/2}/2$ and $|y + d| < (ht/m)^{1/2}/2$ to 1.

A is calculated as

$$A = \frac{\hbar t}{\sqrt{2}m} \left[\int_{d-\frac{1}{2}\sqrt{\frac{\hbar t}{m}}}^{d+\frac{1}{2}\sqrt{\frac{\hbar t}{m}}} \left(\int_{-d-\frac{1}{2}\sqrt{\frac{\hbar t}{m}}}^{-d+\frac{1}{2}\sqrt{\frac{\hbar t}{m}}} \left\{ 1 - \cos\left[\frac{4\pi m d}{\hbar t}(x-y) \right] \right\} dy \right) dx \right]^{-\frac{1}{2}}. \quad (\text{A33})$$

We therefore obtain the joint probability density function $p(x, y, t)$ as

$$p(x, y, t) = |\psi(x, y, t)|^2 = 2A^2 \left(\frac{m}{\hbar t} \right)^2 \left\{ 1 - \cos\left[\frac{4\pi m d}{\hbar t}(x-y) \right] \right\}. \quad (\text{A34})$$

The marginal probability density functions are expressed as

$$\begin{aligned} p(x, d, t) &= \int_{-d-\frac{1}{2}\sqrt{\frac{\hbar t}{m}}}^{-d+\frac{1}{2}\sqrt{\frac{\hbar t}{m}}} |\psi(x, y, t)|^2 dy \\ &= 2A^2 \left(\frac{m}{\hbar t} \right)^2 \left\{ \sqrt{\frac{\hbar t}{m}} - \frac{\hbar t}{2\pi m d} \sin\left(2\pi d \sqrt{\frac{m}{\hbar t}} \right) \cos\left[\frac{4\pi m d}{\hbar t}(x+d) \right] \right\} \end{aligned} \quad (\text{A35})$$

and

$$\begin{aligned}
p(y, -d, t) &= \int_{d-\frac{1}{2}\sqrt{\frac{\hbar t}{m}}}^{d+\frac{1}{2}\sqrt{\frac{\hbar t}{m}}} |\psi(x, y, t)|^2 dx \\
&= 2A^2 \left(\frac{m}{\hbar t} \right)^2 \left\{ \sqrt{\frac{\hbar t}{m}} - \frac{\hbar t}{2\pi m d} \sin\left(2\pi d \sqrt{\frac{m}{\hbar t}}\right) \cos\left[\frac{4\pi m d}{\hbar t}(y-d)\right] \right\}. \quad (\text{A36})
\end{aligned}$$

Thus, ε_{xy} , the ratio of the joint probability density to the product of the marginal probability densities, is expressed as

$$\begin{aligned}
\varepsilon_{xy} &= \frac{p(x, y, t)}{p(x, d, t)p(y, -d, t)} \\
&= \frac{1}{2A^2} \left(\frac{\hbar t}{m} \right)^2 \left\{ 1 - \cos\left[\frac{4\pi m d}{\hbar t}(x-y)\right] \right\} \\
&\quad \left\{ \sqrt{\frac{\hbar t}{m}} - \frac{\hbar t}{2\pi m d} \sin\left(2\pi d \sqrt{\frac{m}{\hbar t}}\right) \cos\left[\frac{4\pi m d}{\hbar t}(x+d)\right] \right\}^{-1} \\
&\quad \left\{ \sqrt{\frac{\hbar t}{m}} - \frac{\hbar t}{2\pi m d} \sin\left(2\pi d \sqrt{\frac{m}{\hbar t}}\right) \cos\left[\frac{4\pi m d}{\hbar t}(y-d)\right] \right\}^{-1}. \quad (\text{A37})
\end{aligned}$$

Therefore, MI_d , the density of MI for X and Y , is

$$\begin{aligned}
MI_d &= p(x, y, t) \log \varepsilon_{xy} \\
&= 2A^2 \left(\frac{m}{\hbar t} \right)^2 \left\{ 1 - \cos\left[\frac{4\pi m d}{\hbar t}(x-y)\right] \right\} \\
&\quad \log \left[\frac{1}{2A^2} \left(\frac{\hbar t}{m} \right)^2 \left\{ 1 - \cos\left[\frac{4\pi m d}{\hbar t}(x-y)\right] \right\} \right. \\
&\quad \left. \left\{ \sqrt{\frac{\hbar t}{m}} - \frac{\hbar t}{2\pi m d} \sin\left(2\pi d \sqrt{\frac{m}{\hbar t}}\right) \cos\left[\frac{4\pi m d}{\hbar t}(x+d)\right] \right\}^{-1} \right. \\
&\quad \left. \left\{ \sqrt{\frac{\hbar t}{m}} - \frac{\hbar t}{2\pi m d} \sin\left(2\pi d \sqrt{\frac{m}{\hbar t}}\right) \cos\left[\frac{4\pi m d}{\hbar t}(y-d)\right] \right\}^{-1} \right]. \quad (\text{A38})
\end{aligned}$$

We calculate MI by integrating MI_d as

$$MI = \int_{d-\frac{1}{2}\sqrt{\frac{\hbar t}{m}}}^{d+\frac{1}{2}\sqrt{\frac{\hbar t}{m}}} \left(\int_{-d-\frac{1}{2}\sqrt{\frac{\hbar t}{m}}}^{-d+\frac{1}{2}\sqrt{\frac{\hbar t}{m}}} MI_d dy \right) dx. \quad (\text{A39})$$

We thus obtain a positive value of MI for identical fermions in the one-dimensional spacetime.

(2) MI of free particles in a three-dimensional spacetime

Similar arguments as above hold for the three-dimensional case. The position vectors of two particles $r_1 = (x_1, y_1, z_1)$ and $r_2 = (x_2, y_2, z_2)$ take the values $(d, 0, 0)$ and $(-d, 0, 0)$ at $t = 0$, respectively. Then, their propagators are

$$K_1(\vec{r}_1, d; t) = \left(\frac{m}{2\pi i \hbar t} \right)^{\frac{3}{2}} \exp\left\{ -\frac{m}{2i\hbar t} [(x_1-d)^2 + y_1^2 + z_1^2] \right\} \quad \text{and} \quad (\text{A40})$$

$$K_2(\vec{r}_2, -d; t) = \left(\frac{m}{2\pi i \hbar t} \right)^{\frac{3}{2}} \exp\left\{ -\frac{m}{2i\hbar t} [(x_2+d)^2 + y_2^2 + z_2^2] \right\}, \quad (\text{A41})$$

respectively. The composite wave function is expressed as

$$\psi(\vec{r}_1, \vec{r}_2, t) = A[K_1(\vec{r}_1, d; t)K_2(\vec{r}_2, -d; t) - K_1(\vec{r}_2, d; t)K_2(\vec{r}_1, -d; t)], \quad (\text{A42})$$

where A is the three-dimensional normalization constant. A is calculated as

$$A = \frac{1}{\sqrt{2}} \left(\frac{ht}{m} \right)^2 \left[\frac{ht}{m} - \frac{1}{4} \left(\frac{ht}{\pi md} \right)^2 \sin^2 \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left(\frac{8\pi md^2}{ht} \right) \right]^{-\frac{1}{2}}. \quad (\text{A43})$$

We obtain the joint probability density function as

$$p(\vec{r}_1, \vec{r}_2, t) = |\psi(\vec{r}_1, \vec{r}_2, t)|^2 = 2A^2 \left(\frac{m}{ht} \right)^6 \left\{ 1 - \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\}. \quad (\text{A44})$$

The ratio of the joint probability density to the product of the marginal probability densities is expressed as

$$\begin{aligned} \varepsilon_{\vec{r}_1, \vec{r}_2} &= \frac{p(\vec{r}_1, \vec{r}_2, t)}{p(\vec{r}_1, t)p(\vec{r}_2, t)} \\ &= \frac{1}{2A^2} \left(\frac{ht}{m} \right)^4 \left\{ 1 - \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\} \\ &\quad \left\{ \sqrt{\frac{ht}{m}} - \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_1 + d) \right] \right\}^{-1} \\ &\quad \left\{ \sqrt{\frac{ht}{m}} - \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_2 - d) \right] \right\}^{-1}. \end{aligned} \quad (\text{A45})$$

Therefore, MI_d is expressed as

$$\begin{aligned} MI_d &= p(\vec{r}_1, \vec{r}_2, t) \log \varepsilon_{\vec{r}_1, \vec{r}_2} \\ &= 2A^2 \left(\frac{m}{ht} \right)^6 \left\{ 1 - \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\} \\ &\quad \log \left[\frac{1}{2A^2} \left(\frac{ht}{m} \right)^4 \left\{ 1 - \cos \left[\frac{4\pi md}{ht} (x_1 - x_2) \right] \right\} \right. \\ &\quad \left. \left\{ \sqrt{\frac{ht}{m}} - \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_1 + d) \right] \right\}^{-1} \right. \\ &\quad \left. \left\{ \sqrt{\frac{ht}{m}} - \frac{ht}{2\pi md} \sin \left(2\pi d \sqrt{\frac{m}{ht}} \right) \cos \left[\frac{4\pi md}{ht} (x_2 - d) \right] \right\}^{-1} \right], \end{aligned} \quad (\text{A46})$$

where A is the three-dimensional normalization constant. By integrating MI_d with respect to r_1 and r_2 , we obtain MI as

$$\begin{aligned} MI &= \iiint (\iiint MI_d d\vec{r}_2) d\vec{r}_1 \\ &= \left(\frac{ht}{m} \right)^2 \int_{d-\frac{1}{2}\sqrt{\frac{ht}{m}}}^{d+\frac{1}{2}\sqrt{\frac{ht}{m}}} \left(\int_{-d-\frac{1}{2}\sqrt{\frac{ht}{m}}}^{-d+\frac{1}{2}\sqrt{\frac{ht}{m}}} MI_d dx_2 \right) dx_1, \end{aligned} \quad (\text{A47})$$

We thus obtain a positive value of MI for identical fermions in the three-dimensional spacetime similarly to the one-dimensional case.

9. Multidimensional information and linear index

We here describe a method that allows the management of highly multidimensional information. While we use genes as examples here, the procedure below is also applied to other multidimensional informatons. The information expressed by a gene G is detailed and ranges over many categories. We regard this information as the n -dimensional random variable X and describe it as the n -dimensional row vector

$$X = (X_1, X_2, \dots, X_n), \quad (\text{A48})$$

where the i -th entry X_i represents information categories, such as the mRNA expression, translation rate, chemical modification, and intracellular localization. Meanwhile, X_i is the m -dimensional column vector whose j -th entry is X_{ij} , and m depends on i . X_i is then expressed as

$$X_i = (X_{i1}, X_{i2}, \dots, X_{im})^T, \quad (\text{A49})$$

where X_{ij} represents either the discrete values that X_i can take or the ranges of divided data if X_i takes a continuous value.

The information expressed by the gene G is written as the combination of entries of each category X_i . To the combinations, we assign the linear index k' arranged lexicographically. Each combination can then be described as the row vector $X_{k'}$ whose entries are the values of j for each i . As an example, if the combination of (i, j) with the k' -th linear index is $(i, j) = (1, 2), (2, 5), (3, 4), \dots$, then $X_{k'}$ is expressed as

$$X_{k'} = (2, 5, 4, \dots), \quad (\text{A50})$$

where the total number N_{tot} of $X_{k'}$ is expressed as

$$N_{tot} = \prod_{i=1}^n m_i. \quad (\text{A51})$$

Given that the information expressed by the gene is large m and n are large finite numbers, and N_{tot} thus becomes an overwhelming number.

Finally, using another linear index k , we rearrange $X_{k'}$ into another row vector X_k according to the order of its realization probability. The realization probability $p(X_k)$ of the k -th observable state X_k then follows the infocanonical distribution $p(X_k) = e^{-I_k}$, where $I_k = -\log p(X_k)$ is the information level of X_k (Appendix 17). Here, even though the information level I_k takes discrete values, because N_{tot} is immensely large, the level can be divided extremely finely and sufficiently expressed by a one-dimensional continuous random variable X_c . The realization probability $p(X_c)$ of the information level X_c is then expressed by a continuous distribution as

$$p(X_c) = \frac{1}{I} e^{-\frac{X_c}{I}}. \quad (\text{A52})$$

In this equation, I is the effective self-information, which is the expectation value of X_c that represents the likelihood that X_c takes a more unlikely information level.

10. Relationship between effective self-information and fitness

In this section, we describe how gene information affects biological fitness. According to Section (a) of Chapter 3, we let X be the observable information level of a gene G . Then, at an individual level, X is related to its realization probability $p(X)$ according to

$$p(X) = \frac{1}{I} e^{-\frac{X}{I}}, \quad (\text{A53})$$

where I is the effective self-information, which is the expectation value of X . In a sufficiently large and stable population, the realization frequency of X follows the same infocanonical distribution. We now assume that the population comprises a subpopulation with the wild-type gene G_{wt} and a subpopulation with the mutant gene G_{mut} . Let I_{wt} and I_{mut} be the effective self-information of G_{wt} and that of G_{mut} , respectively. The difference ΔI in the effective self-information is then $\Delta I = I_{mut} - I_{wt}$.

In each subpopulation, the frequency of individuals whose X takes a value x is expressed as

$$P_{wt}(x) = \frac{1}{I_{wt}} \exp\left(-\frac{1}{I_{wt}}x\right), \quad P_{mut}(x) = \frac{1}{I_{mut}} \exp\left(-\frac{1}{I_{mut}}x\right). \quad (\text{A54})$$

The relative fitness $r(x)$ of the individual with G_{mut} to that with G_{wt} is then expressed as

$$r(x) = \frac{I_{mut} P_{mut}}{I_{wt} P_{wt}} = \exp\left[\left(\frac{1}{I_{wt}} - \frac{1}{I_{mut}}\right)x\right]. \quad (\text{A55})$$

Therefore, the relative fitness r_{mut} of the mutant subpopulation to the wild-type subpopulation is

$$\begin{aligned} r_{mut} &= \overline{r(x)} = \int_0^{\infty} r(x) P_{wt}(x) dx \\ &= \int_0^{\infty} \exp\left[\left(\frac{1}{I_{wt}} - \frac{1}{I_{mut}}\right)x\right] \cdot \frac{1}{I_{wt}} \exp\left(-\frac{1}{I_{wt}}x\right) dx \\ &= \frac{1}{I_{wt}} \int_0^{\infty} \exp\left(-\frac{1}{I_{mut}}x\right) dx \\ &= \frac{I_{mut}}{I_{wt}} = 1 + \frac{\Delta I}{I_{wt}}. \end{aligned} \quad (\text{A56})$$

When we set $I_{wt} = 1$, if $\Delta I \approx 0$ and all the mutations contribute to the positive fitness, then

$$s = \Delta I, \quad r_{mut} = 1 + \Delta I \simeq e^{\Delta I}, \quad (\text{A57})$$

where s is the selection coefficient. In this formula, $\exp(\Delta I)$ is the ratio of the genotype frequency of individuals with a fitness of $1 + \Delta I$ to that with a fitness of 1 after one generation [45]. Thus, the increase in the fitness of the mutant, s , is equal to the rise in the effective self-information, ΔI .

11. Relationship between MI and fitness

We describe how the MI of genes affects biological fitness. If the change in energy is negligible, the spontaneous probability $p(MI)$ that the magnitude of MI becomes MI is

$$p(MI) = e^{-MI} . \quad (\text{A58})$$

On this basis, we consider that MI of two genes $G1$ and $G2$ in a large haploid population follows an infocanonical distribution. We assume that the population comprises two subpopulations; that is, double wild-types ($G1_{wt}$ and $G2_{wt}$) and double mutants ($G1_{mut}$ and $G2_{mut}$). Let MI_{wt} be the population mean MI of $G1_{wt}$ and $G2_{wt}$ in the wild-type subpopulation and MI_{mut} be that between $G1_{mut}$ and $G2_{mut}$ in the mutant subpopulation. The increase in MI caused by the double mutation is then described as $\Delta MI = MI_{mut} - MI_{wt}$.

In each subpopulation, the frequency of the individual whose $MI(G1;G2)$ is x is

$$P_{wt}(x) = \frac{1}{MI_{wt}} \exp\left(-\frac{1}{MI_{wt}}x\right) , \quad P_{mut}(x) = \frac{1}{MI_{mut}} \exp\left(-\frac{1}{MI_{mut}}x\right) , \quad (\text{A59})$$

respectively. Then, the relative fitness $r(x)$ of the double mutant to the double wild-type is

$$r(x) = \frac{MI_{mut} P_{mut}}{MI_{wt} P_{wt}} = \exp\left[\left(\frac{1}{MI_{wt}} - \frac{1}{MI_{mut}}\right)x\right] . \quad (\text{A60})$$

The relative fitness of the double-mutant subpopulation to the double wild-type subpopulation is

$$\begin{aligned} \overline{r(x)} &= \int_0^{\infty} r(x) P_{wt}(x) dx \\ &= \int_0^{\infty} \exp\left[\left(\frac{1}{MI_{wt}} - \frac{1}{MI_{mut}}\right)x\right] \cdot \frac{1}{MI_{wt}} \exp\left(-\frac{1}{MI_{wt}}x\right) dx \\ &= \frac{1}{MI_{wt}} \int_0^{\infty} \exp\left(-\frac{1}{MI_{mut}}x\right) dx \\ &= \frac{MI_{mut}}{MI_{wt}} = 1 + \frac{\Delta MI}{MI_{wt}} . \end{aligned} \quad (\text{A61})$$

When we set $MI_{wt} = 1$, if the values of the relative fitness of each single mutant are nearly equal to 1 and all the new interactions between mutants contribute to the positive fitness, then

$$\varepsilon_{two} = \Delta MI \simeq r_3 - 1 , \quad (\text{A62})$$

where ε_{two} is the epistatic coefficient, and r_3 is the relative fitness of the double mutant. Moreover, if $0 < \Delta MI \ll 1$, then

$$r_3 \simeq 1 + \Delta MI \simeq e^{\Delta MI} . \quad (\text{A63})$$

Here, $\exp(\Delta MI)$ is the ratio of the genotype frequency of individuals with fitness of $1 + \Delta MI$ to that with fitness of 1 after one generation [45], which equals the attractive force \mathcal{F}_{att} [Eq. (A71)].

12. Moran process model of evolution

(1) General solutions to the Kolmogorov forward equation

The general solutions to Eq. (200) under initial conditions $X_0(0) = 1$ and $X_1(0) = X_2(0) = X_3(0) = 0$ are calculated as

$$\begin{aligned}
 X_0(t) &= \exp\left\{-\int_0^t [a(s)+b(s)+c(s)] ds\right\} \\
 X_1(t) &= \frac{\int_0^t a(\tau) \exp\left\{\int_0^\tau [d(s)-a(s)-b(s)-c(s)] ds\right\} d\tau}{\exp\left[\int_0^t d(s) ds\right]} \\
 X_2(t) &= \frac{\int_0^t b(\tau) \exp\left\{\int_0^\tau [e(s)-a(s)-b(s)-c(s)] ds\right\} d\tau}{\exp\left[\int_0^t e(s) ds\right]} \\
 X_3(t) &= \int_0^t c(\tau) \exp\left\{-\int_0^\tau [a(s)+b(s)+c(s)] ds\right\} d\tau \\
 &\quad + \int_0^t d(\tau) \frac{\int_0^\tau a(s) \exp\left\{\int_0^s [d(\delta)-a(\delta)-b(\delta)-c(\delta)] d\delta\right\} ds}{\exp\left[\int_0^\tau d(s) ds\right]} d\tau \\
 &\quad + \int_0^t e(\tau) \frac{\int_0^\tau b(s) \exp\left\{\int_0^s [e(\delta)-a(\delta)-b(\delta)-c(\delta)] d\delta\right\} ds}{\exp\left[\int_0^\tau e(s) ds\right]} d\tau .
 \end{aligned} \tag{A64}$$

(2) Numerical solutions to the Kolmogorov forward equation

If $a(t)$ to $e(t)$ are constant, then the numerical solutions to Eq. (200) under initial conditions $X_0(0) = 1$ and $X_1(0) = X_2(0) = X_3(0) = 0$ are calculated as

$$\begin{aligned}
 X_0(t) &= \exp[-(a+b+c)t], \quad X_1(t) = \frac{a}{d-a-b-c} \{\exp[-(a+b+c)t] - \exp(-dt)\}, \\
 X_2(t) &= \frac{b}{e-a-b-c} \{\exp[-(a+b+c)t] - \exp(-et)\}, \quad X_3(t) = 1 - X_0(t) - X_1(t) - X_2(t).
 \end{aligned} \tag{A65}$$

Here, a , b , d , and e are expressed as

$$\begin{aligned}
 a &= Nu_1\rho_1, \quad \rho_1 = \frac{1-r_0(1-u_1)/(r_1+r_0u_1)}{1-[r_0(1-u_1)/(r_1+r_0u_1)]^N} \\
 b &= Nu_2\rho_2, \quad \rho_2 = \frac{1-r_0(1-u_2)/(r_2+r_0u_2)}{1-[r_0(1-u_2)/(r_2+r_0u_2)]^N} \\
 d &= Nu_4\rho_4, \quad \rho_4 = \frac{1-r_1(1-u_4)/(r_3+r_1u_4)}{1-[r_1(1-u_4)/(r_3+r_1u_4)]^N} \\
 e &= Nu_3\rho_5, \quad \rho_5 = \frac{1-r_2(1-u_3)/(r_3+r_2u_3)}{1-[r_2(1-u_3)/(r_3+r_2u_3)]^N},
 \end{aligned} \tag{A66}$$

where ρ_1 – ρ_5 are the probabilities that the organisms after the mutations are ultimately fixed.

(3) Tunneling rate

In Eq. (A65), c is called the tunneling rate. If it is independent of time, then

$$c = Nu_1(1-V_1-\rho_1) + Nu_2(1-V'_1-\rho_2), \quad (\text{A67})$$

and

$$\rho_3 = \frac{1-r_0(1-u_1)/(r_3+r_1u_4)}{1-[r_0(1-u_1)/(r_3+r_1u_4)]^N}, \quad \rho_3' = \frac{1-r_0(1-u_2)/(r_3+r_2u_3)}{1-[r_0(1-u_2)/(r_3+r_2u_3)]^N}. \quad (\text{A68})$$

Moreover, V_1 is the solution to a system of equations of $(N-1)$ variables:

$$\begin{aligned} & \left[-r_1u_4\rho_3-(r_1-r_1u_4+r_0)\frac{N-1}{r_1(1-u_4)+r_0(N-1)} \right] V_1 + \frac{r_1(1-u_4)(N-1)}{r_1(1-u_4)+r_0(N-1)} V_2 \\ & = -\frac{r_0(N-1)}{r_1(1-u_4)+r_0(N-1)}, \\ & \frac{r_0(N-i)}{ir_1(1-u_4)+r_0(N-i)} V_{i-1} + \left[-r_1u_4\rho_3-(r_1-r_1u_4+r_0)\frac{N-i}{ir_1(1-u_4)+r_0(N-i)} \right] V_i \\ & + \frac{r_1(1-u_4)(N-i)}{ir_1(1-u_4)+r_0(N-i)} V_{i+1} = 0 \quad (2 \leq i \leq N-2), \\ & \frac{r_0}{(N-1)r_1(1-u_4)+r_0} V_{N-2} + \left[-r_1u_4\rho_3-(r_1-r_1u_4+r_0)\frac{1}{(N-1)r_1(1-u_4)+r_0} \right] V_{N-1} = 0. \end{aligned} \quad (\text{A69})$$

Similarly, V'_1 is the solution to a system of equations of $(N-1)$ variables:

$$\begin{aligned} & \left[-r_2u_3\rho_3'-(r_2-r_2u_3+r_0)\frac{N-1}{r_2(1-u_3)+r_0(N-1)} \right] V'_1 + \frac{r_2(1-u_3)(N-1)}{r_2(1-u_3)+r_0(N-1)} V'_2 \\ & = -\frac{r_0(N-1)}{r_2(1-u_3)+r_0(N-1)}, \\ & \frac{r_0(N-i)}{ir_2(1-u_3)+r_0(N-i)} V'_{i-1} + \left[-r_2u_3\rho_3'-(r_2-r_2u_3+r_0)\frac{N-i}{ir_2(1-u_3)+r_0(N-i)} \right] V'_i \\ & + \frac{r_2(1-u_3)(N-i)}{ir_2(1-u_3)+r_0(N-i)} V'_{i+1} = 0 \quad (2 \leq i \leq N-2), \\ & \frac{r_0}{(N-1)r_2(1-u_3)+r_0} V'_{N-2} + \left[-r_2u_3\rho_3'-(r_2-r_2u_3+r_0)\frac{1}{(N-1)r_2(1-u_3)+r_0} \right] V'_{N-1} = 0. \end{aligned} \quad (\text{A70})$$

Finally, the tunneling rate can be calculated using Eq. (A67).

13. Influential force acting between genes

Using the Moran process model, we examine the influential force behind the evolutionary process. Under the assumptions described in Section (d) of Chapter 3, we consider a situation in which gene interactions newly arise between two mutants, thereby increasing the fitness. We here define MI_{inh} as the inherent MI of two genes, which represents the MI of their mutants.

(1) Influential force acting between genes

We define an attractive force \mathcal{F}_{att} acting between genes $G1$ and $G2$ as the ratio of the fitness of double mutants with and without the epistatic coefficient ε_{two} . In terms of population genetics, the intergenic \mathcal{F}_{att} is the ratio of the genotype frequency of double mutants with (p') and without (p) ε_{two} in the next generation, where $p' + p = 1$ and the initial ratio is 1:1. According to the differential equation of Kimura [45], $p'_t/p_t = p'_0/p_0 \exp(\varepsilon_{two} \cdot t)$, where t is the generation number and p'_0/p_0 is the initial ratio of the frequency at $t = 0$, which takes a value of 1. As a result, \mathcal{F}_{att} is obtained by setting $t = 1$ as

$$\mathcal{F}_{att} = e^{\varepsilon_{two}} = e^{\Delta MI_{inh}} . \quad (A71)$$

If we let the initial $MI_{inh} = MI(G1;G2)$ be zero, the intergenic attractive force is $\mathcal{F}_{att} = e^{MI}$. Moreover, the intergenic \mathcal{F}_{att} is independent of the selection coefficient (i.e., s_1 and s_2) for each gene because

$$\frac{p'_1}{p_1} = \frac{e^{s_1 + s_2 + \varepsilon_{two}}}{e^{s_1 + s_2}} = e^{\varepsilon_{two}} = \mathcal{F}_{att} . \quad (A72)$$

Thus, irrespective of s_1 and s_2 , the intergenic \mathcal{F}_{att} is identical to that acting between physical bodies.

(2) Influential force and MI_{inh}

We describe the influential force \mathcal{F}_{att} acting between $G1$ and $G2$ in the Moran model. We construct a 2×2 contingency table of genotype frequencies (Table A1). Using this table, we examine \mathcal{F}_{att} by dividing it into the microforces \mathcal{F}_{kl_att} for each cell (k, l) (Appendix 24). As depicted in the table, we assign columns and rows for the combination of $G1_{wt}/G1_{mut}$ and $G2_{wt}/G2_{mut}$. The cells in the table then give the genotype frequencies X_0, X_1, X_2 , and X_3 .

We assume that a new interaction arises between $G1_{mut}$ and $G2_{mut}$. The attractive force \mathcal{F}_{att} at $t \rightarrow 0$ is thus expressed by a microforce \mathcal{F}_{22_att} in cell (2, 2) of the table [Eqs. (107)] as

$$\mathcal{F}_{att} \simeq \mathcal{F}_{22_att} \simeq \frac{X_3}{(X_1 + X_3)(X_2 + X_3)} . \quad (A73)$$

Additionally, the solutions in Eq. (A65) are approximated to the second order of t :

$$\begin{aligned} X_1(t) &\simeq at - \frac{1}{2} a(a+b+d)t^2 \\ X_2(t) &\simeq bt - \frac{1}{2} b(a+b+e)t^2 \\ X_3(t) &\simeq \frac{1}{2} (ad+be)t^2 . \end{aligned} \quad (A74)$$

Table A1.
Contingency table for gene frequencies

	$G1_{wt}$	$G1_{mt}$	
$G2_{wt}$	X_0	X_1	X_0+X_1
$G2_{mt}$	X_2	X_3	X_2+X_3
	X_0+X_2	X_1+X_3	1

By substituting Eq. (A74) into the right-hand side of Eq. (A73) and by taking the limit with respect to $t \rightarrow 0$, we obtain

$$\mathcal{F}_{att} = e^{\Delta MI_{inh}} \simeq \lim_{t \rightarrow 0} \frac{X_3}{(X_1+X_3)(X_2+X_3)} \simeq \frac{ad+be}{2ab} . \quad (A75)$$

Moreover, if $a \simeq b$, then the attractive force is nearly equal to the ratio of the transition rates; that is,

$$\mathcal{F}_{att} \simeq \frac{d+e}{a+b} . \quad (A76)$$

The influential force \mathcal{F}_{att} thus facilitates the transition from each single mutant to the double mutant.

(3) Influential force and MI_{obs}

We investigate the change in MI observed (MI_{obs}) by applying the probability model. Using the genotype frequencies X_0 - X_3 indicated above, MI_{obs} can be calculated as

$$\begin{aligned} MI_{obs} = & X_0 \log \frac{X_0}{(X_0+X_2)(X_0+X_1)} + X_1 \log \frac{X_1}{(X_1+X_3)(X_0+X_1)} \\ & + X_2 \log \frac{X_2}{(X_0+X_2)(X_2+X_3)} + X_3 \log \frac{X_3}{(X_1+X_3)(X_2+X_3)} . \end{aligned} \quad (A77)$$

If we set the initial conditions to $X_0(0) = X_{i0}$, $X_1(0) = X_{i1}$, $X_2(0) = X_{i2}$, and $X_3(0) = X_{i3}$, then

$$\begin{aligned} MI_{obs} = & \log \left(-\frac{\sigma_4 - \sigma_5}{\sigma_2(\sigma_6 + \sigma_5 + \sigma_1 - 1)} \right) (\sigma_4 - \sigma_5) - \log \left(-\frac{\sigma_4 + \sigma_6 + \sigma_1 - 1}{(\sigma_4 + \sigma_7 + \sigma_1 - 1)(\sigma_6 + \sigma_5 + \sigma_1 - 1)} \right) (\sigma_4 + \sigma_6 + \sigma_1 - 1) \\ & + \log \left(-\frac{\sigma_6 - \sigma_7}{\sigma_3(\sigma_4 + \sigma_7 + \sigma_1 - 1)} \right) (\sigma_6 - \sigma_7) + X_{i0} \log \left(\frac{X_{i0} \sigma_8}{\sigma_2 \sigma_3} \right) \sigma_8 , \end{aligned} \quad (A78)$$

where

$$\begin{aligned} \sigma_1 &= \frac{X_{i0} \sigma_8 (ac - ad + bc - be - cd - ce + de + c^2)}{(a+b+c-d)(a+b+c-e)} \\ \sigma_2 &= X_{i0} \sigma_8 + \sigma_4 - \sigma_5 & \sigma_3 &= X_{i0} \sigma_8 + \sigma_6 - \sigma_7 \\ \sigma_4 &= \frac{e^{-dt} (X_{i0} a + X_{i1} a + X_{i1} b + X_{i1} c - X_{i1} d)}{a+b+c-d} & \sigma_5 &= \frac{X_{i0} a \sigma_8}{a+b+c-d} \\ \sigma_6 &= \frac{e^{-et} (X_{i2} a + X_{i0} b + X_{i2} b + X_{i2} c - X_{i2} e)}{a+b+c-e} & \sigma_7 &= \frac{X_{i0} b \sigma_8}{a+b+c-e} \\ \sigma_8 &= e^{-(a+b+c)} . \end{aligned}$$

Next, to focus on the initial state, we set $X_0(0) = 1$ and $X_1(0) = X_2(0) = X_3(0) = 0$. By taking the Maclaurin series, the half of the second derivative of MI_{obs} at $t = 0$ is expressed as

$$\begin{aligned} \frac{1}{2} \frac{d^2 MI_{obs}}{dt^2} \Big|_{t=0} &= 2ab - ad - be + (ad + be) \log\left(\frac{ad + be}{2ab}\right) \\ &\simeq 2ab - ad - be + (ad + be) \log \mathcal{F}_{att} \\ &= (ad + be) \Delta MI_{inh} + 2ab - ad - be \quad . \end{aligned} \quad (\text{A79})$$

The apparent attractive force \mathcal{F}_{obs} acting between G_1 and G_2 observed at the population level is calculated according to Eq. (69) using the information metric $r(G_1, G_2) = H(G_1) + H(G_2) - 2MI_{obs}$. The gene information is enormous, and we thus assume that $H(G_1)$ and $H(G_2)$ are constant. We then have

$$\begin{aligned} \mathcal{F}_{obs} &= \frac{\mathcal{M}(G_1)\mathcal{M}(G_2)}{\mathcal{M}(G_1) + \mathcal{M}(G_2)} \frac{d^2 r}{dt^2} \\ &\simeq \frac{4}{W(G_1)^{-\alpha} + W(G_2)^{-\alpha}} [(ad + be) \Delta MI_{inh} + 2ab - ad - be] \\ &= A_M \log \mathcal{F}_{att} + B_M \simeq A_M (\mathcal{F}_{att} - 1) \quad , \end{aligned} \quad (\text{A80})$$

where $A_M = 4(ad + be) / [W(G_1)^{-\alpha} + W(G_2)^{-\alpha}]$ and $B_M = 4(2ab - ad - be) / [W(G_1)^{-\alpha} + W(G_2)^{-\alpha}]$. This equation relates the apparent attractive force \mathcal{F}_{obs} to the true attractive force \mathcal{F}_{att} .

14. Temporal change in the MI of genes

MI of genes in terms of MI_{inh} is expressed by Eq. (A29) in Appendix 7,

$$MI \simeq (1 - \gamma\delta)^2 (1 - e^{-\frac{m}{\lambda}}) (1 - e^{-\frac{n}{\nu}}) \quad . \quad (\text{A81})$$

Let ΔMI be the change in MI when $\gamma\delta$, m , and n change from here. When the initial value of $\gamma\delta$ is 1, $\Delta(\gamma\delta) = \gamma\delta - 1$. It then follows that

$$\begin{aligned} \Delta MI &\simeq \frac{\partial(MI)}{\partial(\gamma\delta)} \Delta(\gamma\delta) + \frac{\partial(MI)}{\partial m} \Delta m + \frac{\partial(MI)}{\partial n} \Delta n + \frac{\partial(MI)}{\partial \lambda} \Delta \lambda + \frac{\partial(MI)}{\partial \nu} \Delta \nu \\ &\simeq (1 - \gamma\delta)^2 (1 - e^{-\frac{m}{\lambda}}) (1 - e^{-\frac{n}{\nu}}) \left(2 + \frac{\Delta m}{\lambda} + \frac{\Delta n}{\nu}\right) \\ &\quad - (1 - \gamma\delta)^2 \left[\frac{m}{\lambda^2} e^{-\frac{m}{\lambda}} (1 - e^{-\frac{n}{\nu}}) \Delta \lambda + \frac{n}{\nu^2} e^{-\frac{n}{\nu}} (1 - e^{-\frac{m}{\lambda}}) \Delta \nu \right] \quad . \end{aligned} \quad (\text{A82})$$

If we assume that $\Delta \lambda = \Delta \nu = \kappa t$, then the expectation value of ΔMI is calculated as

$$\begin{aligned} \langle \Delta MI \rangle &\simeq (4D^2 t^2 + 4Dt) \left\{ (1 - e^{-\frac{m}{\lambda}}) (1 - e^{-\frac{n}{\nu}}) \left[\left(\frac{1}{\lambda} + \frac{1}{\nu} \right) \mu t + 2 \right] \right. \\ &\quad \left. - \left[\frac{m}{\lambda^2} e^{-\frac{m}{\lambda}} (1 - e^{-\frac{n}{\nu}}) + \frac{n}{\nu^2} e^{-\frac{n}{\nu}} (1 - e^{-\frac{m}{\lambda}}) \right] \kappa t \right\} \quad . \end{aligned} \quad (\text{A83})$$

Thus, $\langle \Delta MI \rangle$ increases spontaneously and monotonically for sufficiently small κ .

15. Network information of edge weights

We calculate the Σ^W component of the network information, which comprises edge weights. As the edge weight of the gene network, we use MI of two genes. Σ^W in Eq. (219) is approximated using Stirling's formula as

$$\begin{aligned}\Sigma^W &= \log \frac{(N-1+S)!}{(N-1)!S!} \\ &\simeq (N-1+S) \log(N-1+S) - (N-1) \log(N-1) - S \log S \\ &= (N-1) \log\left(1 + \frac{S}{N-1}\right) + S \log\left(\frac{N-1}{S} + 1\right),\end{aligned}\quad (\text{A84})$$

where $S = \sum_{j \neq i} MI(G_i; G_j)$. When MI changes by ΔMI on average, the change in Σ^W is expressed as

$$\begin{aligned}\Delta \Sigma^W &\simeq \left(\frac{d\Sigma^W}{dS}\right) \Delta S \\ &\simeq \left[(N-1) \log\left(\frac{N-1}{S} + 1\right)\right] \Delta MI.\end{aligned}\quad (\text{A85})$$

Thus, the change in Σ^W is approximately proportional to the average change in MI .

16. Functional mutual information

We consider the MI that represents the functional relationship between genes, which we refer to as **functional mutual information (fMI)**. We assume that two genes X and Y express information that is categorized into one of n functions (X_1, \dots, X_n) and (Y_1, \dots, Y_n), respectively. We additionally assume that each function X_i and Y_j takes m states (X_{i1}, \dots, X_{im}) and (Y_{j1}, \dots, Y_{jm}), respectively. Let p_{ijkl} be the joint probability of X_{ik} and Y_{jl} in the contingency table of X_i and Y_j . The marginal probabilities are then expressed as $p(X_{ik}) = \sum_l p_{ijkl}$ and $p(Y_{jl}) = \sum_k p_{ijkl}$. Under these settings, we define the functional mutual information $fMI(X_i; Y_j)$ of the functions X_i and Y_j as

$$fMI(X_i; Y_j) = \sum_k \sum_l p_{ijkl} \log \left[\frac{p_{ijkl}}{p(X_{ik})p(Y_{jl})} \right]. \quad (\text{A86})$$

By contrast, the total $MI(X; Y)$ of X and Y is described as

$$\begin{aligned}MI(X; Y) &= \sum_i \sum_j \sum_k \sum_l \frac{p_{ijkl}}{n^2} \log \left\{ \frac{p_{ijkl}}{n^2} \div \left[\sum_i \sum_k \left(\frac{p_{ijkl}}{n^2} \right) \sum_j \sum_l \left(\frac{p_{ijkl}}{n^2} \right) \right] \right\} \\ &= \frac{1}{n^2} \sum_i \sum_j \sum_k \sum_l p_{ijkl} \log \frac{p_{ijkl}}{p(X_{ik})p(Y_{jl})} \\ &= \frac{1}{n^2} \sum_i \sum_j fMI(X_i; Y_j).\end{aligned}\quad (\text{A87})$$

Thus, the total $MI(X; Y)$ is the average of $fMI(X_i; Y_j)$, which implies that $MI(X; Y)$ represents the functional proximity of the genes X and Y .

17. Infocanonical distribution

(1) Infocanonical distribution of self-information

Suppose a closed system A, which interacts and exchanges information with the surroundings. Let λ be a state of A, let $P(\lambda)$ be its realization probability, and let $I(\lambda)$ be its self-information. Then,

$$p(\lambda) = \exp[-I(\lambda)] \quad (\text{A88})$$

holds anytime, irrespective of energy and temperature. Next, we suppose a closed system comprising two systems A and B. The systems interact with each other and exchange information. Let λ and ν be the states of A and B, respectively, let $P(\lambda, \nu)$ be their joint probability, and let $I(\lambda, \nu)$ be their joint self-information. Then,

$$p(\lambda, \nu) = \exp[-I(\lambda, \nu)] \quad (\text{A89})$$

holds anytime, irrespective of energy and temperature.

In addition to the self-information, if we let X be the information level of a multidimensional random variable $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$, then the realization probability of X is expressed as

$$p(X) = \frac{1}{I} e^{-\frac{X}{I}}, \quad (\text{A90})$$

where I is the expectation value of X ; that is, the effective self-information [Chapter 3, Section (a)].

In contrast, the thermodynamic canonical distribution for the relationship between the system and the heat bath is described as follows [22]. Let μ be a state of a closed system, $E(\mu)$ be its energy level, and let $P(\mu)$ be its realization probability. Then,

$$p(\mu) \propto \exp[-\beta E(\mu)]. \quad (\text{A91})$$

holds, where β is the inverse temperature $1/k_B T$.

In Eqs. (A88) and (A89), the energy level $E(\mu)$ of the thermodynamic canonical distribution (A91) is replaced by the information levels $I(\lambda)$ and $I(\lambda, \nu)$, respectively. Therefore, we refer to Eqs. (A88) and (A89) as the infocanonical distributions of one system and two systems, respectively. Moreover, we call Eq. (A90) the infocanonical distribution of the multidimensional random variable.

Whereas the thermodynamic canonical distribution holds only under the isothermal condition, Eqs. (A88) and (A89) hold under any conditions irrespective of temperature and energy.

(2) Infocanonical distribution of MI

Besides the descriptions above, the "equivalence principle of information and probability" [Eqs. (A5) and (A104) in Appendices 2 and 20, respectively] implies that MI follows the infocanonical distribution. This infocanonical distribution of MI holds under the isothermal condition, contrary to the case of self-information.

18. Infocanonical distribution of micromutual information

Let MI be the MI of the $m \times n$ contingency table with respect to two discrete random variables X and Y . It is defined as

$$MI = \sum_{k=1}^m \sum_{l=1}^n p(X_k, Y_l) \log \frac{p(X_k, Y_l)}{p(X_k)p(Y_l)}, \quad (\text{A92})$$

where k and l run from 1 to m and 1 to n , respectively. Here, $\sum_k p(X_k, Y_l) = \sum_k p(X_k) = \sum_l p(Y_l) = 1$. Let MI_{kl} be the micromutual information, which is defined as the MI of the 2×2 contingency table with respect to each cell (k, l) :

$$\begin{aligned} MI_{kl} &= p(X_k, Y_l) \log \frac{p(X_k, Y_l)}{p(X_k)p(Y_l)} \\ &+ [p(X_k) - p(X_k, Y_l)] \log \frac{p(X_k) - p(X_k, Y_l)}{p(X_k)[1 - p(Y_l)]} \\ &+ [p(Y_l) - p(X_k, Y_l)] \log \frac{p(Y_l) - p(X_k, Y_l)}{[1 - p(X_k)]p(Y_l)} \\ &+ [1 - p(X_k) - p(Y_l) + p(X_k, Y_l)] \log \frac{1 - p(X_k) - p(Y_l) + p(X_k, Y_l)}{[1 - p(X_k)][1 - p(Y_l)]}. \end{aligned} \quad (\text{A93})$$

When m and n become sufficiently large, by the micromutual information summation theorem,

$$MI \simeq \sum_{k=1}^m \sum_{l=1}^n MI_{kl}. \quad (\text{A94})$$

Each summand of the right-hand side of Eq. (A92) is the same as the first term on the right-hand side of Eq. (A93). Hence, in Eq. (A93), the total sum of the second, third, and fourth terms on the right-hand side with respect to k and l converges to zero. This implies that, as m and n approach infinity, the expectation value for each MI_{kl} of the cell (k, l) becomes more independent of the micromutual information of the other cells.

As m and n increase, the distributions of X and Y become approximated by the uniform distribution, because X and Y respectively take the values X_k and Y_l in a random manner. Meanwhile, the relative frequencies of the k -th column and l -th row respectively approach $1/m$ and $1/n$ according to the law of large numbers and become less dependent on the relative frequencies of the other columns and rows.

Taken together, when m and n are sufficiently large, from a microscopic point of view, we can apply Eq. (17) and the discussion in Appendix 2(1) by setting X and Y as X_k and Y_l , respectively. MI_{kl} thus follows an infocanonical distribution. Meanwhile, from a macroscopic point of view, the mean value of MI_{kl} is calculated as

$$\langle MI_{kl} \rangle = \frac{1}{mn} MI. \quad (\text{A95})$$

In conclusion, the probability density function $p(x)$ of $x = MI_{kl}$ is expressed as

$$p(x) = \lambda e^{-\lambda x}, \quad \lambda = \frac{1}{\langle MI_{kl} \rangle}. \quad (\text{A96})$$

The micromutual information MI_{kl} thus follows an infocanonical distribution with a mean of MI/mn .

19. Frontier information theory

The micromutual information MI_{kl} follows an infocanonical distribution according to Eq. (A96). Therefore, if we let j be the rank of the magnitude of MI_{kl} , say, x , then,

$$j \simeq (mn-1)[\exp(-\lambda x) - \exp(-\lambda x_1)] + 1 . \quad (\text{A97})$$

where x_1 is the maximum MI_{kl} , which we refer to as the frontier mutual information MI_{front} .

Hence, x is approximately expressed as a logarithmic function of j as

$$x \simeq -\frac{1}{\lambda} \log\{[j-1+(mn-1)\exp(-\lambda x_1)]/(mn-1)\} , \quad (\text{A98})$$

and

$$(mn-1)\exp(-\lambda x_1) - 1 \simeq 0 . \quad (\text{A99})$$

It follows that the frontier mutual information MI_{front} is related to MI by

$$\begin{aligned} MI_{front} &:= \max(MI_{kl}) \\ &\simeq \frac{MI}{mn} \log(mn-1) \simeq \langle MI_{kl} \rangle \log(mn-1) . \end{aligned} \quad (\text{A100})$$

Thus, MI_{front} is almost proportional to both the entire MI and the average MI_{kl} . Conversely, the entire MI is estimated from the MI_{front} according to

$$MI \simeq \frac{mn}{\log(mn-1)} MI_{front} . \quad (\text{A101})$$

In addition to the above derivation, we can estimate the confidence intervals of MI_{front} and MI . The distribution function of MI_{front} is expressed as

$$F[MI_{front}] = [1 - \exp(-\lambda MI_{front})]^{mn} , \quad (\text{A102})$$

where λ is the mean of MI_{front} . Its mean $\langle MI_{front} \rangle$ is approximated as $(MI/mn)\log(mn-1)$ according to Eq. (A101) and its standard deviation $\sigma[MI_{front}]$ is approximately equal to $(\pi/6^{0.5})(MI/mn)$. Therefore, the 95% confidence interval of MI_{front} is expressed as $[\langle MI_{front} \rangle - 2.316 \sigma[MI_{front}], \langle MI_{front} \rangle + 2.316 \sigma[MI_{front}]]$, which is $[(MI/mn)[\log(mn-1) - 2.970], (MI/mn)[\log(mn-1) + 2.970]]$. We can therefore estimate the 95% confidence interval of MI from MI_{front} as

$$\frac{mn}{\log(mn-1) + 2.970} MI_{front} \leq MI \leq \frac{mn}{\log(mn-1) - 2.970} MI_{front} . \quad (\text{A103})$$

Thus, using MI_{front} , we can predict MI with a 95% confidence interval. By increasing m and n , it is possible to narrow the confidence interval.

20. Equivalence principle of information and probability

In the field of probability theory, the realization probability of the MI of two random variables follows the "equivalence principle of information and probability"; that is,

$$p = e^{-MI}, \quad MI = -\log p. \quad (\text{A104})$$

This principle also holds in the field of statistics. Here, we prove that MI is asymptotically equal to the logarithm of Fisher's p -value divided by the sample size. Our theorem in statistics is as follows.

Theorem

Let P_F be the p -value of Fisher's exact test in a 2×2 contingency table. Let MI be the MI of the two variables in the same table and N be the sample size. Then,

$$MI = -\lim_{N \rightarrow \infty} \frac{1}{N} \log P_F. \quad (\text{A105})$$

Proof

We consider a 2×2 contingency table for two random variables A and B, which take two values A and A' and B and B', respectively. Table 3A [reshown, see Chapter 4, Section (b)] gives the relative frequency of the combination of the variables. Hence, $X_0, X_1, X_2,$ and X_3 are the proportions of AB, A'B, AB', and A'B', respectively. Table 3B gives the frequency itself obtained from the relative frequency by multiplying it by N . Then, MI is defined as

$$\begin{aligned} MI &= X_0 \log \frac{X_0}{(X_0+X_2)(X_0+X_1)} + X_1 \log \frac{X_1}{(X_1+X_3)(X_0+X_1)} \\ &\quad + X_2 \log \frac{X_2}{(X_0+X_2)(X_2+X_3)} + X_3 \log \frac{X_3}{(X_1+X_3)(X_2+X_3)} \\ &= X_0 \log X_0 - X_0 \log(X_0+X_2) - X_0 \log(X_0+X_1) + X_1 \log X_1 - X_1 \log(X_1+X_3) \\ &\quad - X_1 \log(X_0+X_1) + X_2 \log X_2 - X_2 \log(X_0+X_2) - X_2 \log(X_2+X_3) + X_3 \log X_3 \\ &\quad - X_3 \log(X_1+X_3) - X_3 \log(X_2+X_3) \\ &= X_0 \log X_0 + X_1 \log X_1 + X_2 \log X_2 + X_3 \log X_3 - (X_0+X_1) \log(X_0+X_1) \\ &\quad - (X_0+X_2) \log(X_0+X_2) - (X_1+X_3) \log(X_1+X_3) - (X_2+X_3) \log(X_2+X_3) \quad . \quad (\text{A106}) \end{aligned}$$

Table 3. Contingency tables of genotype frequency (reshown)

Table 3A

	A	A'	
B	X_0	X_1	X_0+X_1
B'	X_2	X_3	X_2+X_3
	X_0+X_2	X_1+X_3	1

Table 3B

	A	A'	
B	NX_0	NX_1	$N(X_0+X_1)$
B'	NX_2	NX_3	$N(X_2+X_3)$
	$N(X_0+X_2)$	$N(X_1+X_3)$	N

In the following, we consider P_F , the p -value of Fisher's exact test. We divide P_F into two terms, the main term P_H and the sum of the remaining terms $P_{F_rem} = P_F - P_H$. First, P_H is the probability that the observed result is obtained. Second, P_{F_rem} represents the probabilities that more impossible results occur. As shown below, P_{F_rem} asymptotically vanishes, compared with the main term.

To begin with, the main term, P_H , is calculated using the hypergeometric distribution:

$$P_H = \frac{\binom{N(X_0+X_1)}{NX_0} \binom{N(X_2+X_3)}{NX_2}}{\binom{N}{N(X_0+X_2)}} = \frac{[N(X_0+X_1)!]}{[(NX_0)!(NX_1)!]} \times \frac{[N(X_2+X_3)!]}{[(NX_2)!(NX_3)!]} \div \frac{N!}{[N(X_0+X_2)!][N(X_1+X_3)!]} \quad (\text{A107})$$

Taking the logarithm yields

$$-\log P_H = \log(NX_0)! + \log(NX_1)! + \log(NX_2)! + \log(NX_3)! + \log N! \\ - \log[N(X_0+X_1)!] - \log[N(X_0+X_2)!] - \log[N(X_1+X_3)!] \\ - \log[N(X_2+X_3)!] \quad (\text{A108})$$

When n is large, Stirling's formula, $\log n! \simeq n \log n - n$, and $X_0 + X_1 + X_2 + X_3 = 1$, yields

$$-\log P_H \simeq NX_0 \log NX_0 - NX_0 + NX_1 \log NX_1 - NX_1 + NX_2 \log NX_2 - NX_2 \\ + NX_3 \log NX_3 - NX_3 + N \log N - N - N(X_0+X_1) \log N(X_0+X_1) \\ + N(X_0+X_1) - N(X_0+X_2) \log N(X_0+X_2) + N(X_0+X_2) \\ - N(X_1+X_3) \log N(X_1+X_3) + N(X_1+X_3) - N(X_2+X_3) \log N(X_2+X_3) \\ + N(X_2+X_3) \\ = NX_0 \log N + NX_0 \log X_0 + NX_1 \log N + NX_1 \log X_1 + NX_2 \log N + NX_2 \log X_2 \\ + NX_3 \log N + NX_3 \log X_3 + N \log N - N(X_0+X_1) \log N \\ - N(X_0+X_1) \log(X_0+X_1) - N(X_0+X_2) \log N - N(X_0+X_2) \log(X_0+X_2) \\ - N(X_1+X_3) \log N - N(X_1+X_3) \log(X_1+X_3) - N(X_2+X_3) \log N \\ - N(X_2+X_3) \log(X_2+X_3) \\ = NX_0 \log X_0 + NX_1 \log X_1 + NX_2 \log X_2 + NX_3 \log X_3 \\ - N(X_0+X_1) \log(X_0+X_1) - N(X_0+X_2) \log(X_0+X_2) \\ - N(X_1+X_3) \log(X_1+X_3) - N(X_2+X_3) \log(X_2+X_3) \\ = N \times MI \quad (\text{A109})$$

Hence,

$$MI \simeq -\frac{1}{N} \log P_H \quad (\text{A110})$$

Therefore, if N is large, then MI is approximately equal to the logarithm of P_H divided by N .

We next evaluate the error of this equation. The error comprises two parts. The first part is the sum of the remaining terms $P_{F_rem} = P_F - P_H$, and the second part is the error of Stirling's formula applied to the main term P_H . Let ER_1 be $-\log P_H - (-\log P_F)$, which is an equivalent of P_{F_rem} , and will be evaluated below.

First part of the error

We first estimate the first part of the error P_{F_rem} . We can assume that $X_0 + X_2 < \min\{X_1 + X_3, X_0 + X_1, X_2 + X_3\}$ and $X_1X_2 < X_0X_3$ without loss of generality. It follows that $X_2 < X_1$, $X_0 < X_3$, and $X_2 < X_3$. The sum of the remaining terms P_{F_rem} is then the probability that the frequency of AB is greater than NX_0 when the marginal frequency is fixed.

We let $p(X)$ be the probability that the frequency of AB is X , which follows the hypergeometric distribution. The main term is then equal to $p(NX_0)$, and

$$p(NX_0+1) = p(NX_0) \frac{NX_1NX_2}{(NX_0+1)(NX_3+1)} < \frac{p(NX_0)}{OR}, \quad (\text{A111})$$

where OR is the odds ratio $(X_0X_3) \div (X_1X_2) > 1$. Similarly,

$$p(NX_0+2) = p(NX_0+1) \frac{(NX_1-1)(NX_2-1)}{(NX_0+2)(NX_3+2)} < \frac{p(NX_0)}{OR^2}, \quad (\text{A112})$$

and

$$p(NX_0+NX_2) < \frac{p(NX_0)}{OR^{NX_2}}. \quad (\text{A113})$$

Therefore, the sum of the remaining terms is

$$\begin{aligned} & p(NX_0+1) + p(NX_0+2) + \dots + p(NX_0+NX_2) \\ & < p(NX_0) \left(\frac{1}{OR} + \frac{1}{OR^2} + \dots + \frac{1}{OR^{NX_2}} \right) \\ & = p(NX_0) \times \left(1 - \frac{1}{OR^{NX_2}} \right) \div (OR-1) \\ & < \frac{p(NX_0)}{OR-1}. \end{aligned} \quad (\text{A114})$$

P_H , the p -value of the hypergeometric distribution, is equal to $p(NX_0)$. Then,

$$P_F - P_H < \frac{P_H}{OR-1}. \quad (\text{A115})$$

Hence,

$$P_F < P_H \frac{OR}{OR-1}. \quad (\text{A116})$$

Thus, $ER_1 = -\log P_H - (-\log P_F)$ is evaluated as

$$0 < -\log P_H - (-\log P_F) < \log \frac{OR}{OR-1}. \quad (\text{A117})$$

Second part of the error

We next estimate the second part of the error $ER_2 = -\log P_H - NMI$. Stirling's formula is expressed as

$$\Gamma(N+1) = N! = (N+1)^{N+\frac{1}{2}} \exp[-(N+1)] \sqrt{2\pi} \exp\left[\sum_{n=1}^{\infty} \frac{(-1)^{n-1} B_{2n}}{2n(2n-1)(N+1)^{2n-1}}\right], \quad (\text{A118})$$

where B_{2n} denotes Bernoulli numbers. Taking the logarithm and substituting $B_2 = 1/6$ and $B_4 = 1/30$, we obtain

$$\begin{aligned} & \log N! - N \log N + N \\ &= N \log \frac{N+1}{N} + \frac{1}{2} \log(N+1) - 1 + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} - \frac{1}{360(N+1)^3} + \dots \end{aligned} \quad (\text{A119})$$

Using (A119) and neglecting $\frac{1}{360(N+1)^3}$, we derive the difference (error) between the logarithm of P_H and MI as

$$\begin{aligned} & -\log P_H - N \times MI \\ &= \frac{1}{2} \log(N+1)(NX_0+1)(NX_1+1)(NX_2+1)(NX_3+1) \\ & - \frac{1}{2} \log[N(X_0+X_1)+1][N(X_0+X_2)+1][N(X_1+X_3)+1][N(X_2+X_3)+1] \\ & + N \log \frac{N+1}{N} + NX_0 \log \frac{NX_0+1}{NX_0} + NX_1 \log \frac{NX_1+1}{NX_1} + NX_2 \log \frac{NX_2+1}{NX_2} + NX_3 \log \frac{NX_3+1}{NX_3} \\ & - N(X_0+X_1) \log \frac{N(X_0+X_1)+1}{N(X_0+X_1)} - N(X_0+X_2) \log \frac{N(X_0+X_2)+1}{N(X_0+X_2)} \\ & - N(X_1+X_3) \log \frac{N(X_1+X_3)+1}{N(X_1+X_3)} - N(X_2+X_3) \log \frac{N(X_2+X_3)+1}{N(X_2+X_3)} \\ & - 1 + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} + \frac{1}{12(NX_0+1)} + \frac{1}{12(NX_1+1)} + \frac{1}{12(NX_2+1)} + \frac{1}{12(NX_3+1)} \\ & - \frac{1}{12(NX_0+NX_1+1)} - \frac{1}{12(NX_0+NX_2+1)} - \frac{1}{12(NX_1+NX_3+1)} - \frac{1}{12(NX_2+NX_3+1)}. \end{aligned} \quad (\text{A120})$$

Here, $N \log \frac{N+1}{N} = N \log\left(1 + \frac{1}{N}\right) < N \times \frac{1}{N} = 1$,

$$NX_0 \log \frac{NX_0+1}{NX_0} \times \frac{N(X_0+X_1)}{N(X_0+X_1)+1} \times \frac{N(X_0+X_2)}{N(X_0+X_2)+1} < NX_0 \log\left(1 + \frac{1}{NX_0}\right) < 1, \quad (\text{A121})$$

and similar inequalities hold. Moreover,

$$\frac{1}{12(NX_0+1)} + \frac{1}{12(NX_1+1)} + \frac{1}{12(NX_2+1)} + \frac{1}{12(NX_3+1)} < \frac{1}{24} + 3 \times \frac{1}{12} = \frac{7}{24}. \quad (\text{A122})$$

Thus,

$$\begin{aligned}
& -\log P_H - N \times MI \\
& < \frac{1}{2} \log(N+1) + 5 - 1 + \frac{1}{2} \log(2\pi) + \frac{1}{24} + \frac{7}{24} \\
& < \frac{1}{2} \log(N+1) + 5.253 \quad .
\end{aligned} \tag{A123}$$

We next evaluate the lower bound of this error. Using

$$N \log\left(1 + \frac{1}{N}\right) > N\left(\frac{1}{N} - \frac{1}{2N^2}\right) = 1 - \frac{1}{2N} \tag{A124}$$

and similar inequalities, we obtain

$$\begin{aligned}
& -\log P_H - N \times MI \\
& > \frac{1}{2} \log \frac{2(N+1)}{(N+1)^4} + 1 - \frac{1}{2N} + 1 - \frac{1}{2NX_0} + 1 - \frac{1}{2NX_1} \\
& \quad + 1 - \frac{1}{2NX_2} + 1 - \frac{1}{2NX_3} - 5 + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} \\
& > -\frac{3}{2} \log(N+1) + \frac{1}{2} \log 2 - \frac{1}{2N} - 2 \\
& \quad + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} \\
& > -\frac{3}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - 0.735 \quad .
\end{aligned} \tag{A125}$$

Thus,

$$-\frac{3}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - 0.735 < -\log P_H - N \times MI < \frac{1}{2} \log(N+1) + 5.253 \quad . \tag{A126}$$

Hence, we have evaluated the second part of the error ER_2 .

Combining the two parts of the error

Combining ER_1 and ER_2 , we can evaluate the total error as

$$\begin{aligned}
& -\log \frac{OR}{OR-1} - \frac{3}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - 0.735 < -\log P_F - N \times MI \\
& < \frac{1}{2} \log(N+1) + 5.253 \quad .
\end{aligned} \tag{A127}$$

Therefore,

$$\begin{aligned}
& N \times MI - \log \frac{OR}{OR-1} - \frac{3}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - 0.735 \\
& < -\log P_F \\
& < N \times MI + \frac{1}{2} \log(N+1) + 5.253 \quad .
\end{aligned} \tag{A128}$$

That is,

$$\begin{aligned}
& MI + \frac{1}{N} \left[-\log \frac{OR}{OR-1} - \frac{3}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - 0.735 \right] \\
& < -\frac{1}{N} \log P_F \\
& < MI + \frac{1}{2N} \log(N+1) + \frac{5.253}{N} .
\end{aligned} \tag{A129}$$

We thus have

$$MI = -\lim_{N \rightarrow \infty} \frac{1}{N} \log P_F . \tag{A130}$$

q. e. d.

Corollary

Let P_F be the p -value of Fisher's exact test in an $m \times n$ contingency table. Let MI be the MI between the two variables in the same table and N be the sample size. Then,

$$MI = -\lim_{N \rightarrow \infty} \frac{1}{N} \log P_F . \tag{A131}$$

Proof

We consider an $m \times n$ contingency table for two random variables A and B . A takes integer values from 1 to m , and B takes integer values from 1 to n . Let X_{ij} be the relative frequency that A takes i and B takes j . Then, $-\log P_H - NMI$ is calculated and evaluated similarly to Eqs. (A120) to (A123) as

$$\begin{aligned}
& -\log P_H - NMI \\
& = \frac{1}{2} \log \left[(N+1) \prod_{i=1}^m \prod_{j=1}^n (NX_{ij}+1) \right] - \frac{1}{2} \log \left[\prod_{j=1}^n \left(N \sum_{i=1}^m X_{ij} + 1 \right) \right] \left[\prod_{i=1}^m \left(N \sum_{j=1}^n X_{ij} + 1 \right) \right] \\
& + N \log \frac{N+1}{N} + \sum_{i=1}^m \sum_{j=1}^n NX_{ij} \log \frac{NX_{ij}+1}{NX_{ij}} \\
& - N \sum_{j=1}^n \left(\sum_{i=1}^m X_{ij} \right) \log \frac{N \sum_{i=1}^m X_{ij} + 1}{N \sum_{i=1}^m X_{ij}} - N \sum_{i=1}^m \left(\sum_{j=1}^n X_{ij} \right) \log \frac{N \sum_{j=1}^n X_{ij} + 1}{N \sum_{j=1}^n X_{ij}} \\
& - 1 + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} + \sum_{i=1}^m \sum_{j=1}^n \frac{1}{12(NX_{ij}+1)} - \sum_{j=1}^n \frac{1}{12 \left(N \sum_{i=1}^m X_{ij} + 1 \right)} - \sum_{i=1}^m \frac{1}{12 \left(N \sum_{j=1}^n X_{ij} + 1 \right)} \\
& < \frac{1}{2} \log(N+1) + 1 + mn - 1 + \frac{1}{2} \log(2\pi) + \frac{1}{24} + \frac{1}{24} + \frac{1}{12} (mn-1) \\
& < \frac{1}{2} \log(N+1) + \frac{13}{12} mn + 0.919 .
\end{aligned} \tag{A132}$$

Meanwhile, $-\log P_H - NMI$ is evaluated similarly to (A124) and (A125) as

$$\begin{aligned}
& -\log P_H - NMI \\
& > \frac{1}{2} \log \frac{2(N+1)}{(N+1)^{mn}} + 1 - \frac{1}{2N} + \sum_{i=1}^m \sum_{j=1}^n \left(1 - \frac{1}{2NX_{ij}}\right) - mn - 1 + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} \\
& > -\frac{mn-1}{2} \log(N+1) + \frac{1}{2} \log 2 - \frac{1}{2N} - \frac{mn}{2} + \frac{1}{2} \log(2\pi) + \frac{1}{12(N+1)} \\
& > -\frac{mn-1}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - \frac{mn}{2} + 1.265 \quad .
\end{aligned} \tag{A133}$$

Finally, we evaluate the difference between P_F and P_H . Because the total number of contingency tables is ${}_{mn}H_N = (N + mn - 1)! / [(mn - 1)! N!]$, P_F is evaluated as

$$P_F < {}_{mn}H_N P_H < (N+m-1)^{mn-1} P_H \quad . \tag{A134}$$

Hence,

$$-\log P_H - (mn-1) \log(N+mn-1) - NMI < -\log P_F - NMI \quad . \tag{A135}$$

Combining the above inequalities, we obtain

$$-\frac{mn-1}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - \frac{mn}{2} + 1.265 - (mn-1) \log(N+mn-1) < -\log P_F - NMI \quad . \tag{A136}$$

Meanwhile,

$$-\log P_F - NMI < -\log P_H - NMI < \frac{1}{2} \log(N+1) + \frac{13}{12} mn + 0.919 \quad . \tag{A137}$$

Therefore,

$$\begin{aligned}
NMI - \frac{mn-1}{2} \log(N+1) + \frac{1}{12(N+1)} - \frac{1}{2N} - \frac{mn}{2} + 1.265 - (mn-1) \log(N+mn-1) \\
& < -\log P_F < \\
NMI + \frac{1}{2} \log(N+1) + \frac{13}{12} mn + 0.919 \quad .
\end{aligned} \tag{A138}$$

Hence,

$$MI = -\lim_{N \rightarrow \infty} \frac{1}{N} \log P_F \quad . \tag{A139}$$

q. e. d.

We have thus proved Eq. (A105) also in the case of an $m \times n$ contingency table. MI of multi-dimensional random variables approaches the logarithm of the p -value of Fisher's exact test divided by the sample size N .

21. Weighted pathway analysis

We calculate how much a gene G_i with unknown functions is involved in a biological pathway A . We first consider the pathway $A = \{x_1, x_2, \dots, x_n\}$ that comprises n genes. Let $B = \{x_1, x_2, \dots, x_m\} \subset A$ be the set of genes in A whose MI with G_i ranks above M . Moreover, let W_j be the total sum of $MI(G_i; G_k)$ of G_i and each element $G_k \in B$. Then, probability P_{WE} that $W_j > W$ under the null hypothesis that G_i is not involved in A is expressed as

$$P_{WE} \simeq \sum_m \frac{M!}{m!(M-m)!} \cdot \frac{(N-M-1)!}{(n-m)!(N-M-n+m-1)!} \cdot \frac{n!(N-1-n)!}{(N-1)!} \\ \times \frac{1}{\sqrt{2\pi}} \int_{\sqrt{m}(\frac{W}{m}-\mu)/\sigma}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \quad . \quad (\text{A140})$$

Here, N is the total number of genes, μ is the mean of $MI(G_i; G_j)$ of G_i and G_j ($j \neq i$), and σ is the standard deviation of $MI(G_i; G_j)$. If we define weighted information I_{WE} as

$$I_{WE} = -\log(P_{WE}) \quad , \quad (\text{A141})$$

then I_{WE} represents how much G_i is involved in A .

22. Equation of motion in spacetime with probability as the coordinate

We describe an equation of motion using the probability p as the coordinate. We assume a composite system XY comprising two informatons X and Y , which have the potential difference V and MI MI . On the one hand, when this system follows the canonical distribution for the potential difference, the state probability is $p(V) = \exp(-\beta V)/Z$, where Z is the partition function. This equation means that the probability $p(V)$ increases in the direction in which the potential difference V decreases and the information metric r between X and Y reduces.

Meanwhile, the probability of MI being MI is $p(MI) = \exp(-MI)$ by Eq. (17). This equation is based on a principle that the probability is proportional to the number of states of the composite system XY , which conforms to the second law of thermodynamics. Increasing MI decreases the total entropy by $H(X, Y) = H(X) + H(Y) - MI$, and $p(MI)$ thus increases in the direction in which MI decreases, and the system entropy $H(X, Y)$ increases. In this case, because the internal energy U is proportional to $H(X, Y)$ under the thermal equilibrium condition, the energy U increases. We here assume a static condition where $V \simeq \Delta U$, and the information metric r also increases. Clearly, this is opposite to the above observed direction for the information metric r .

The realization probability p of the system can be elucidated by considering a balance between the above two opposing forces. Hence, by applying an equality $p = p(V) = p(MI)$, we get

$$p = \frac{1}{Z} e^{-\beta V} = e^{-MI}. \quad (\text{A142})$$

Therefore, $dV = -dp/(\beta p)$, where dV is the change in the potential difference. The change in the Lagrangian $d\mathcal{L}$ of the system is then expressed as $d\mathcal{L} = dT_K - dV$, where dT_K is the change in the kinetic energy; that is,

$$d\mathcal{L} = dT_K - dV = m \frac{dp}{dt} d\left(\frac{dp}{dt}\right) + \frac{dp}{\beta p}. \quad (\text{A143})$$

We now apply Lagrange's equation of motion $d/dt [\partial\mathcal{L}/\partial(dp/dt)] - \partial\mathcal{L}/\partial p = 0$. Given that $p^{-1} = p(MI)^{-1} = \mathcal{F}_{att}$, we can obtain an equation of motion of Newton's type as

$$m \frac{d^2 p}{dt^2} = \frac{1}{\beta p} = k_B T \mathcal{F}_{att} =: F_p. \quad (\text{A144})$$

where F_p is the force of probability, which acts to reduce the information metric r between X and Y . In summary, \mathcal{F}_{att} is a mechanistic force in this spacetime, which increases the relative probability of information transmission and the system's state probability.

23. Equilibrium information distance

We define I_e as the equilibrium information distance between the informatons, where the state probability takes a maximum value. We demonstrate that $I_e = (\log 2)/4$ irrespective of the informatons. The probabilistic influential force potential $\varphi'(I_O)$ in the I_O -spacetime is a function of $I_O(x, r)$:

$$\varphi'(I_O) = \mathcal{M}(X)\mathcal{M}(Y) \left[-k_1 e^{-I_O} + \frac{k_2}{2} \mathcal{M}(X)\mathcal{M}(Y) e^{-2I_O} \right]. \quad (\text{A145})$$

Then, $\varphi'(I_O)$ takes the minimum when $I_O = I_e$, which is calculated as

$$\begin{aligned} I_e &= \log \left[\left(\frac{k_2}{k_1} \right) \mathcal{M}(X)\mathcal{M}(Y) \right] \\ &= \acute{\alpha} H(X) + \acute{\alpha} H(Y) + \log \left(\frac{k_2}{k_1} \right). \end{aligned} \quad (\text{A146})$$

If we use $k_1 \mathcal{M}(X)\mathcal{M}(Y) = (2m\omega/h)^{1/4}$ and $k_2 [\mathcal{M}(X)\mathcal{M}(Y)]^2 = (4m\omega/h)^{1/4}$ according to Eq. (84) and Eq. (85), then we obtain

$$I_e = \acute{\alpha} H(X) + \acute{\alpha} H(Y) + \log \frac{\sqrt[4]{2}}{\mathcal{M}(X)\mathcal{M}(Y)} = \frac{\log 2}{4}. \quad (\text{A147})$$

Thus, we have demonstrated that $I_e = (\log 2)/4$ is constant irrespective of the mass of the interacting quantum or the separation of the quantum in x -spacetime.

24. Influential microforce

(1) Information of the microstate

In addition to the definitions of self-information $I_k = -\log p(X_k)$ and a fluctuation factor $\varepsilon_{kl} = p(X_k, Y_l)/[p(X_k)p(Y_l)]$, we respectively define joint self-information I_{kl} and microinformation distance r_{kl} as

$$I_{kl} = I_k + I_l - \log(\varepsilon_{kl}), \quad r_{kl} = I_k + I_l - 2 \log(\varepsilon_{kl}) .$$

If we multiply these formulas by $P(X_k, Y_l)$ and sum with respect to k and l , then we obtain

$$\begin{aligned} H(X, Y) &= \sum_k \sum_l P(X_k, Y_l) I_{kl} = \sum_k \left[I_k \sum_l P(X_k, Y_l) \right] + \sum_l \left[I_l \sum_k P(X_k, Y_l) \right] - MI \\ &= \sum_k [I_k P(X_k)] + \sum_l [I_l P(Y_l)] - MI , \end{aligned} \quad (\text{A148})$$

which implies that

$$H(X, Y) = H(X) + H(Y) - MI, \quad r = H(X) + H(Y) - 2MI , \quad (\text{A149})$$

where r is the information metric, which is equal to $\sum_k \sum_l P(X_k, Y_l) r_{kl}$.

(2) Influential force of the microstate

We next derive Eqs. (107) and (108), which represent the relationship among the influential microforce, micromutual information and macroscopic influential force. As we mentioned in Appendix 6, when $m \rightarrow \infty$ and $n \rightarrow \infty$,

$$\sum_k \sum_l MI_{kl} = MI = \sum_k \sum_l P(X_k, Y_l) \log \gamma_k \delta_l , \quad (\text{A150})$$

where $\gamma_k \delta_l$ is equal to ε_{kl} . Therefore, when m and n are sufficiently large,

$$MI_{kl} \simeq P(X_k, Y_l) \log \gamma_k \delta_l . \quad (\text{A151})$$

Hence,

$$\log \gamma_k \delta_l \simeq \frac{MI_{kl}}{P(X_k, Y_l)} . \quad (\text{A152})$$

Let $P(X_k, Y_l)$ and $P'(X_k, Y_l)$ be the joint probabilities before and after the change in $\gamma_k \delta_l$. Then, $P(X_k, Y_l) = \exp[-(I_k + I_l)]$ and $P'(X_k, Y_l) = \exp\{-(I_k + I_l - \Delta \log(\gamma_k \delta_l))\}$. By these probabilities, the attractive microforce \mathcal{F}_{kl_att} is defined as

$$\mathcal{F}_{kl_att} = \frac{P'(X_k, Y_l)}{P(X_k, Y_l)} = \exp[\Delta \log \gamma_k \delta_l] \simeq \exp\left[\frac{\Delta MI_{kl}}{P(X_k, Y_l)}\right]. \quad (\text{A153})$$

From this formula, we can derive the macroscopic attractive force as

$$\begin{aligned} \sum_k \sum_l P(X_k, Y_l) \mathcal{F}_{kl_att} &\simeq \sum_k \sum_l P(X_k, Y_l) \exp\left[\frac{\Delta MI_{kl}}{P(X_k, Y_l)}\right] \\ &\simeq \sum_k \sum_l P(X_k, Y_l) \left[1 + \frac{\Delta MI_{kl}}{P(X_k, Y_l)}\right] = 1 + \sum_k \sum_l \Delta MI_{kl} \\ &\simeq 1 + \Delta MI \simeq \exp(\Delta MI) = \bar{\mathcal{F}}_{att}. \end{aligned} \quad (\text{A154})$$

Meanwhile, the repulsive microforce \mathcal{F}_{kl_rep} is the square of \mathcal{F}_{kl_att} , similar to Eq. (A10) in Appendix

2. Therefore, similar formulas can also be derived with respect to \mathcal{F}_{kl_rep} .

The composite microforce $\mathcal{F}_{kl} = \mathcal{F}_{kl_att} + \mathcal{F}_{kl_rep}$ can be regarded as the microforce for the microstate $X = X_k$ and $Y = Y_l$.

(3) Equation of motion of the microstate

If we use the probability as the coordinate in the information metric spacetime, then the equation of motion for the microstate is

$$\frac{d^2 p_{kl}}{dt^2} = k_B T [W(X_k)^{\acute{\alpha}} + W(Y_l)^{\acute{\alpha}}] \{k_1 \exp(-\acute{\alpha} r_{kl}) - k_2 [W(X_k) W(Y_l)]^{\acute{\alpha}} \exp(-2\acute{\alpha} r_{kl})\}. \quad (\text{A155})$$

If we multiply both sides of Eq. (A155) by $P(X_k, Y_l)$ and sum with respect to k and l , then we obtain

$$\frac{d^2 p}{dt^2} = k_B T [\mathcal{M}(G_1) + \mathcal{M}(G_2)] [k_1 e^{-\acute{\alpha} r} - k_2 \mathcal{M}(G_1) \mathcal{M}(G_2) e^{-2\acute{\alpha} r}]. \quad (\text{A156})$$

Indeed,

$$\sum_k \sum_l P(X_k, Y_l) [W(X_k)^{\acute{\alpha}} + W(Y_l)^{\acute{\alpha}}] \exp(-\acute{\alpha} r_{kl}) \simeq [W(X_k)_{avg}^{\acute{\alpha}} + W(Y_l)_{avg}^{\acute{\alpha}}] \exp(-\acute{\alpha} r), \quad (\text{A157})$$

where

$$\begin{aligned} W(X_k)_{avg}^{\acute{\alpha}} &= \sum_k P(X_k) W(X_k)^{\acute{\alpha}} = \sum_k P(X_k) \exp(\acute{\alpha} I_k) \\ &\simeq \exp[\acute{\alpha} H(G_1)] = W(G_1)^{\acute{\alpha}} = \mathcal{M}(G_1). \end{aligned} \quad (\text{A158})$$

In this way, the influential force can be explained by the influential microforce. We derived the equation of motion with respect to the probability from the microscopic probability.

25. Information entropy of the multidimensional random variable

We assume that the information expressed by gene G_X , whose length on the chromosome equals L bases, takes at most $N (\leq 4^L)$ observable states, represented by a multidimensional random variable $\mathbf{X} = (X_1, X_2, \dots, X_m)^T$. If we let X be the information level of \mathbf{X} , then the realization probability of X follows the infocanonical distribution

$$p(X) = \frac{1}{I} e^{-\frac{X}{I}}, \quad (\text{A159})$$

where I is the expectation value of X ; that is, the effective self-information. We also assume that the information levels are evenly split and discretely distinguished using a linear index k ($1 \leq k \leq n$). The realization probability of the k -th information level X_k is then described as

$$p(X_k) = \int_{(k-1)/n}^{k/n} p(X) dX = \left[\exp\left(\frac{1}{\lambda}\right) - 1 \right] \exp\left(-\frac{k}{\lambda}\right), \quad (\text{A160})$$

where λ is the mean index value $\lambda = nI$.

The information entropy $H(X)$ is then calculated as

$$\begin{aligned} H(X) &= - \sum_{k=1}^n p(X_k) \log p(X_k) = - \sum_{k=1}^n e^{-\frac{k}{\lambda}} (e^{\frac{1}{\lambda}} - 1) \left[-\frac{k}{\lambda} + \log(e^{\frac{1}{\lambda}} - 1) \right] \\ &= (e^{\frac{1}{\lambda}} - 1) \sum_{k=1}^n \frac{k}{\lambda} e^{-\frac{k}{\lambda}} - (e^{\frac{1}{\lambda}} - 1) \log(e^{\frac{1}{\lambda}} - 1) \sum_{k=1}^n e^{-\frac{k}{\lambda}} \\ &\simeq (e^{\frac{1}{\lambda}} - 1) \int_0^n \frac{x}{\lambda} e^{-\frac{x}{\lambda}} dx - (e^{\frac{1}{\lambda}} - 1) \log(e^{\frac{1}{\lambda}} - 1) \frac{e^{-\frac{1}{\lambda}} (e^{-\frac{n}{\lambda}} - 1)}{e^{-\frac{1}{\lambda}} - 1} \\ &= \lambda (e^{\frac{1}{\lambda}} - 1) \left[1 - \left(\frac{n}{\lambda} + 1 \right) e^{-\frac{n}{\lambda}} \right] - (1 - e^{-\frac{n}{\lambda}}) \log(e^{\frac{1}{\lambda}} - 1). \end{aligned} \quad (\text{A161})$$

We suppose a situation in which the evolution of the gene G_X results in an appreciable increase in λ .

If λ goes to infinity, then $H(X)$ is approximated as

$$\begin{aligned} H(X) &\simeq \lambda \cdot \frac{1}{\lambda} \left[1 - \left(\frac{n}{\lambda} + 1 \right) e^{-\frac{n}{\lambda}} \right] - (1 - e^{-\frac{n}{\lambda}}) \log\left(\frac{1}{\lambda}\right) \\ &= 1 - \left(\frac{n}{\lambda} + 1 \right) e^{-\frac{n}{\lambda}} + (1 - e^{-\frac{n}{\lambda}}) \log \lambda. \end{aligned} \quad (\text{A162})$$

If n becomes sufficiently large in comparison to λ , $e^{-\frac{n}{\lambda}} \simeq 0$. Hence,

$$H(X) \simeq 1 + \log \lambda. \quad (\text{A163})$$

Thus, the information entropy $H(X)$ of the multidimensional random variable \mathbf{X} is expressed by the logarithm of the mean index value λ of the discrete information level X_k .

26. Temporal change in effective self-information

As described in Appendix 9, we let X be the observable information level of a gene G_X . Then, at an individual organism level, X is related to its realization probability $p(X)$ by

$$p(X) = \frac{1}{I} e^{-\frac{X}{I}}, \quad (\text{A164})$$

where I is the effective self-information; that is, the expectation value of X . In a sufficiently large population of organisms, the individual frequency follows the same infocanonical distribution. We assume that the whole population comprises two subpopulations, one with the wild-type gene G_{wt} and the other with the mutant gene G_{mut} . Let I_{wt} and I_{mut} be the effective self-information of G_{wt} and G_{mut} , respectively. The difference ΔI in the effective self-information is then described as $\Delta I = I_{mut} - I_{wt}$.

In each subpopulation, the realization probability of the individual whose information level of G_X is x is expressed as

$$P_{wt}(x) = \frac{1}{I_{wt}} \exp\left(-\frac{1}{I_{wt}} x\right), \quad P_{mut}(x) = \frac{1}{I_{mut}} \exp\left(-\frac{1}{I_{mut}} x\right). \quad (\text{A165})$$

We now assume that the information level x is evenly and discretely distinguished by a linear index k ($1 \leq k \leq n$). In this case, the realization probability of the k -th information level X_k follows the infocanonical distribution

$$p(X_k) = (e^{\frac{1}{I}} - 1) e^{-\frac{k}{I}}, \quad (\text{A166})$$

which is the integration of $p(X)$ from $k-1$ to k . Let ΔI be the change in the effective self-information I when the realization probability of each state X_k becomes γ_k -fold owing to the mutation of the gene G_X . If we let γ be the average of γ_k , then ΔI is expressed as

$$\begin{aligned} \Delta I &= -\log[\gamma p(X_k)] + \log p(X_k) \\ &= -\log \gamma \simeq 1 - \gamma. \end{aligned} \quad (\text{A167})$$

We assume that the value of $1-\gamma$ satisfies the diffusion equation and follows a normal distribution with mean zero and variance σ^2 over time. The absolute expectation value of $1-\gamma$ is then expressed as

$$\begin{aligned} E(|1-\gamma|) &= \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} |1-\gamma| \exp\left[-\frac{(1-\gamma)^2}{2\sigma^2}\right] d\gamma \\ &= \frac{2}{\sqrt{2\pi} \sigma} \int_0^{\infty} x \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sqrt{\frac{2}{\pi}} \sigma. \end{aligned} \quad (\text{A168})$$

Here, the expectation value of $|\Delta I|$ at time t is

$$\langle |\Delta I| \rangle \simeq 2 \sqrt{\frac{Dt}{\pi}}, \quad (\text{A169})$$

where D is the diffusion coefficient and $\sigma^2 \simeq 2Dt$. Thus, the absolute expectation value of ΔI increases in proportion to the square root of t .

27. Relationship between the rank change in MI and fitness

We consider the network N_i that comprises N genes with the gene G_i at the center. The nodes are all the genes, and the edges exist between G_i and the other genes, the weights of which are MI . Let G_j be the gene whose MI with G_i is the j -th largest among $N - 1$ genes around G_i . Because MI follows the infocanonical distribution from Eq. (17), we set $MI(G_i; G_j) = -A \times \log(j + B) + C$.

Because

$$j = \exp\left\{\frac{1}{A}[C - MI(G_i; G_j)]\right\} - B, \quad (\text{A170})$$

the rank of G_j when $MI(G_i; G_j)$ increases in ΔMI is calculated as

$$\begin{aligned} j + \Delta j &= \exp\left\{\frac{1}{A}[C - MI(G_i; G_j) - \Delta MI]\right\} - B \\ &= (j + B) \exp\left(-\frac{\Delta MI}{A}\right) - B, \end{aligned}$$

$$\Delta j = (j + B) \left[\exp\left(-\frac{\Delta MI}{A}\right) - 1 \right]. \quad (\text{A171})$$

The expectation value λ_r of the rise in the rank of G_j is expressed as

$$\lambda_r = -\Delta j \simeq (j + B) \frac{\Delta MI}{A}. \quad (\text{A172})$$

In particular, when $B = 0$ as in Fig. 16,

$$\lambda_r \simeq j \frac{\Delta MI}{A}. \quad (\text{A173})$$

If we represent the permutation by the change in rank as the product of transpositions with the minimal number, then the number of transpositions follows a Poisson distribution with mean λ_r . Here, we consider the case that the change in MI per unit time increases from $\Delta MI_0 (\simeq 0)$ of the wild type to ΔMI_1 of the mutant, and λ_r thereby increases from λ_0 to λ_1 . Then, the probabilities $P_0(n)$ and $P_1(n)$ that the transpositions of ranks occur n times are

$$P_0(n) = \frac{1}{n!} \lambda_0^n e^{-\lambda_0}, \quad P_1(n) = \frac{1}{n!} \lambda_1^n e^{-\lambda_1}, \quad (\text{A174})$$

respectively. We here assume that all the increases in MI contribute to positive fitness. The relative fitness $r(n)$ of the individual with the mutant gene to that with the wild-type gene is then expressed as

$$r(n) = \frac{e^{\lambda_1} P_1(n)}{e^{\lambda_0} P_0(n)} = \left(\frac{\lambda_1}{\lambda_0} \right)^n. \quad (\text{A175})$$

Moreover, the relative fitness $r_{net}(N_i)$ of the mutant subpopulation to the wild-type subpopulation is then calculated as

$$\begin{aligned} r_{net}(N_i) &= \sum_{n=0}^{\infty} r(n)P_0(n) = \sum_{n=0}^{\infty} \left(\frac{\lambda_1}{\lambda_0}\right)^n \frac{1}{n!} \lambda_0^n e^{-\lambda_0} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \lambda_1^n e^{-\lambda_0} = e^{\lambda_1 - \lambda_0} . \end{aligned} \quad (\text{A176})$$

Therefore, the increase in the fitness $\varepsilon_{net}(N_i)$ when $|\Delta MI_0| \ll 1$ and $|\Delta MI_1| \ll 1$ is expressed as

$$\begin{aligned} \varepsilon_{net}(N_i) &\simeq e^{\lambda_1 - \lambda_0} - 1 \simeq \exp\left[\frac{j}{A}(\Delta MI_1 - \Delta MI_0)\right] - 1 \\ &\simeq \frac{j}{A}(\Delta MI_1 - \Delta MI_0) . \end{aligned} \quad (\text{A177})$$

In particular, when $\Delta MI_0 \simeq 0$,

$$\varepsilon_{net}(N_i) \simeq \frac{j}{A} \Delta MI_1 . \quad (\text{A178})$$

Moreover, if we calculate the average $(\lambda_r)_{avg}$ of the rise in the rank of all the other $N-1$ genes, then

$$\begin{aligned} (\lambda_r)_{avg} &= (-\Delta j)_{avg} = \frac{1}{N-1} \sum_{j=1}^{N-1} \left\{ (j+B) \left[1 - \exp\left(-\frac{\Delta MI}{A}\right) \right] \right\} \\ &= \left(\frac{N}{2} + B \right) \left[1 - \exp\left(-\frac{\Delta MI}{A}\right) \right] . \end{aligned} \quad (\text{A179})$$

In particular, when $B = 0$ as in Fig.16,

$$\lambda_r = \frac{N}{2} \left[1 - \exp\left(-\frac{\Delta MI}{A}\right) \right] . \quad (\text{A180})$$

When $\Delta MI_0 \simeq 0$ and $|\Delta MI_1| \ll 1$, the expectation value of the increase in fitness over the $N-1$ genes is

$$\varepsilon_{net}(N_i) \simeq \frac{N}{2A} \langle \Delta MI_1 - \Delta MI_0 \rangle \simeq \frac{N}{2A} \langle \Delta MI_1 \rangle . \quad (\text{A181})$$

Thus, the increase in the fitness $\varepsilon_{net}(N_i)$ is approximately proportional to the average increase in MI per unit time and the number of genes N .

28. Pareto distribution of intergenic MI

As described in Section (h) of Chapter 3, we consider the case that mutations of a gene G_i appreciably affect the distribution of $MI(G_i;G_j)$ with other genes G_j . Then, in the course of evolution, $MI(G_i;G_j)$ can transition from the original infocanonical distribution and approach the Pareto distribution. Here, we provide an analytical explanation of this phenomenon as follows.

Let us assume an initial condition in which $MI(G_i;G_j)$ follows the infocanonical distribution with a probability density function $p_1(x) = (1/\lambda_1)\exp(-x/\lambda_1)$ for $x \geq 0$, where $x = MI(G_i;G_j)$ and λ_1 is the initial expectation value of x .

Next, we further assume that a mutation occurs in G_i that strengthens the relationship with a restricted number of G_j that have greater association with G_i . To represent this situation, we introduce another probability density function $p_2(x) = (1/\lambda_2)\exp[-(x-n)/\lambda_2]$ for $x \geq n > 0$, where $\lambda_2 (> \lambda_1)$ is the secondary expectation value of $x - n$. We here suppose that the affected interaction is under the influence of both the initial and second relationships with expectation values of λ_1 and λ_2 , respectively. Hence, we consider a composite probability density function $p(x)$ as $p(x) = p_1(x)$ for $0 \leq x \leq n$ and $p(x) = p_1(x)p_2(x)$ for $x \geq n$. Then,

$$p_1(x)p_2(x) = \frac{1}{\lambda_1\lambda_2} \exp\left[-\frac{\lambda_1 + \lambda_2}{\lambda_1\lambda_2} \left(x - \frac{n\lambda_1}{\lambda_1 + \lambda_2}\right)\right]. \quad (\text{A182})$$

The integration of $p(x)$ from zero to infinity is $1 + \left(\frac{1}{\lambda_1 + \lambda_2} - 1\right) \exp\left(-\frac{n}{\lambda_1}\right)$, which approaches a value of 1 as n tends to infinity. The interactions of G_i with the strongly interacting genes become much stronger because λ_2 is greater than λ_1 .

Generally, the exponential distribution $f_{(0, b, b/a)}(x) = (a/b)\exp[-a(x-b)/b]$ is approximated by the Pareto distribution $f_{(a, b)}(x) = ab^a/(x^{a+1})$. Therefore, if we set $a/b = (\lambda_1 + \lambda_2)/(\lambda_1\lambda_2)$ and $b = n\lambda_1/(\lambda_1 + \lambda_2)$, then $p_1(x)p_2(x)$ is approximated by the Pareto distribution $p_3(x)$ as

$$p_3(x) = \frac{n}{\lambda_2} \cdot \frac{\left[\frac{n\lambda_1}{\lambda_1 + \lambda_2}\right]^{\frac{n}{\lambda_2}}}{x^{\frac{n}{\lambda_2} + 1}}. \quad (\text{A183})$$

The Pareto distribution is defined in the domain $x \geq b > 0$, and x thus has a positive lower bound. This requirement is fulfilled by the above conditions, in which the probability distribution function $p_2(x)$ is

defined in the domain $x \geq n$. Hence, the infocanonical distribution $p_1(x)$ is transformed into the Pareto distribution $p_3(x)$ for $x \geq n$, where G_i has much MI with other genes G_j .

Finally, if we define a new composite probability density function $p_4(x)$ as $p_4(x) = p_1(x)$ for $0 \leq x \leq n$ and $p_4(x) = p_3(x)$ for $x \geq n$, then $p_4(x)$ approaches the Pareto distribution as a whole. If similar changes occur repeatedly and frequently in the course of evolution, then the original infocanonical distribution approaches the Pareto distribution more and more.

Hence, this example demonstrates the change of the infocanonical distribution into the Pareto distribution with time.

29. Multivariate central limit theorem

For the cancer tissues included in the TCGA, the time after onset and fitness are different in each case and vary greatly. However, if the number of cases is sufficiently large, the multivariate central limit theorem guarantees that the data fit the theory of the Moran process through the use of mean values. The theorem is stated as follows.

Let X_1, X_2, \dots, X_n be d -dimensional random variables that are independent and follow identical distributions. If $X_1 = (X_1^1, X_1^2, \dots, X_1^d)$, $E(X_1^i) = 0$ ($1 \leq i \leq d$), and we let Σ be a d -dimensional square matrix with entries $E(X_1^i X_1^j)$ ($1 \leq i, j \leq d$), then the distribution of $(X_1 + X_2 + \dots + X_n)/n^{0.5}$ converges to the normal distribution $N(0, \Sigma)$ as n tends to infinity. The probability density function p of $N(0, \Sigma)$ is expressed as

$$p = (2\pi)^{-\frac{d}{2}} (\det\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{X\Sigma^{-1}X^t}{2}\right) \quad (\text{A184})$$

for $X = (X^1, X^2, \dots, X^d)$, where X^t is the transposed column vector of X .

Therefore, if the number of cases is sufficiently large, like in the case of the TCGA, then the averages of the time, fitness, and relative frequencies of genotypes for all the cases approach the true theoretical averages. Additionally, the differences between them (i.e., error) diminish in inverse proportion to the square root of the case number.

30. Integration theory of the Moran process and branching process

We consider a precancerous population of N cells that at first divides according to the Moran process, described by Eq. (200). Upon initiation of carcinogenesis, the heterogeneous cell population proceeds to the branching process. At this time point, the numbers of cells of types 0 to 3 are nearly equal to NX_0 to NX_3 , respectively.

In the next branching process, the cell of type 3 divides with probability $r_3/2$ and dies with probability $1 - r_3/2$. When it divides, a daughter cell undergoes another mutation with probability u , which increases the fitness by s . At time t after the transition into the branching process, the expected number of the cells of type 3 without another mutation is

$$NX_3 \left[\frac{r_3(2-u)}{2} \right]^t, \quad (\text{A185})$$

and that with another mutation is

$$NX_3 r_3 u (2-u)^{t-1} \frac{(r_3+s)^t - r_3^t}{2^t s} \quad (\text{A186})$$

by elementary calculations. Similarly, the expected number of the cells of type 1 without another mutation is

$$NX_1 \left[\frac{r_1(2-u)}{2} \right]^t, \quad (\text{A187})$$

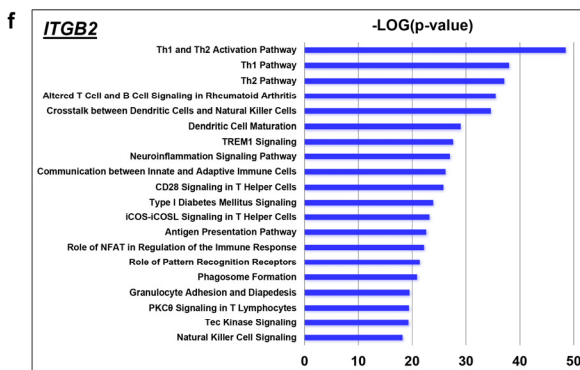
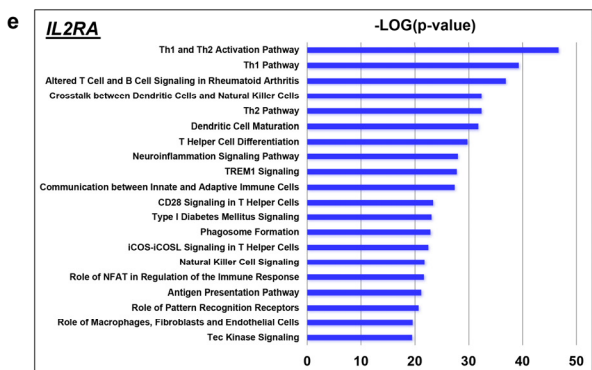
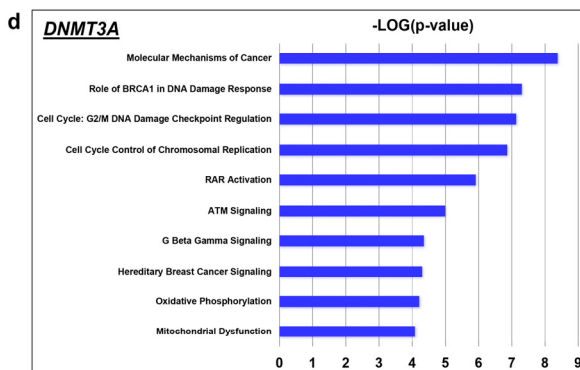
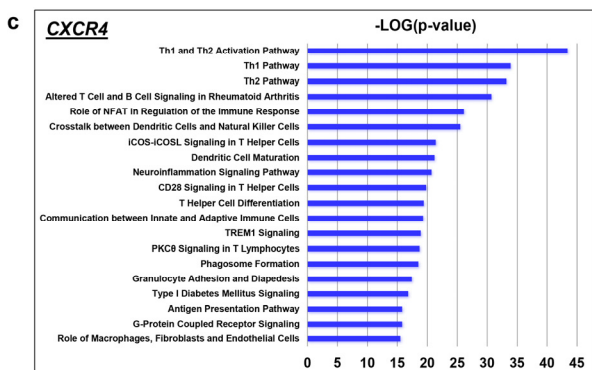
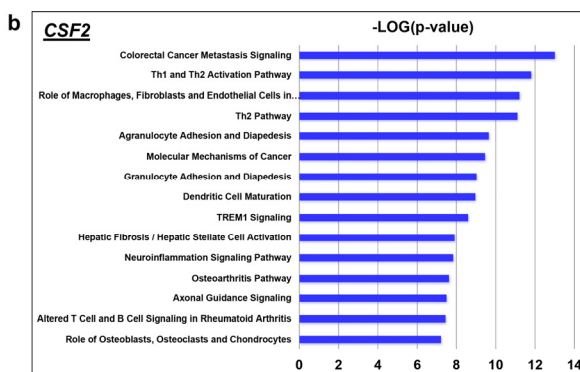
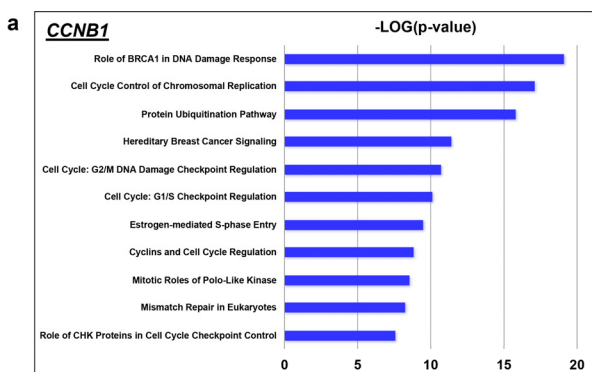
and that with another mutation is

$$NX_1 r_1 u (2-u)^{t-1} \frac{(r_1+s)^t - r_1^t}{2^t s}. \quad (\text{A188})$$

Therefore, if r_1 is nearly equal to r_3 , and s is sufficiently larger than $r_1 - 1$ and $r_3 - 1$, then the ratio of the number of the type-3 cells to the number of type-1 cells with another mutation is approximately equal to X_3/X_1 . Similar arguments hold for the case that the cells undergo plural mutations in the branching process. Thus, the initial ratio of the number of cells in the Moran process remains almost constant after progression to the branching process.

31. Examples of the *ab initio* GO calculation

We present other examples of the *ab initio* GO calculation in Fig. A1. With extremely high sensitivity, the calculation is readily applicable to the analysis of intermolecular interactions, pathway analysis, prospect of relations to diseases, and the specification of disease markers and therapeutic targets. The results in Fig. A1 demonstrate that our method successfully predicted functions of genes that act in various tissues that range from neuron to immune cells. Moreover, the analyzed genes are of various categories, such as enzymes, signal transduction factors, and epigenetics.



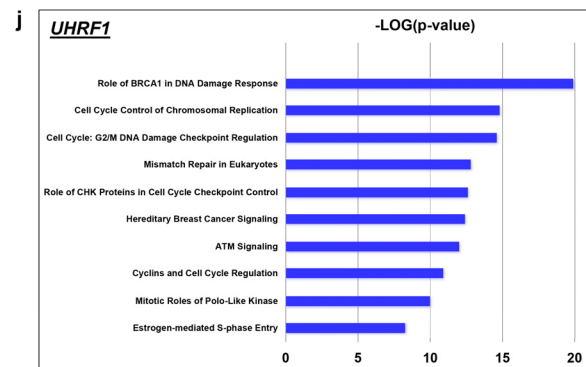
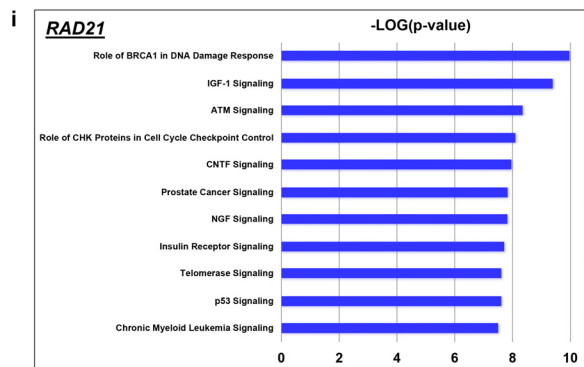
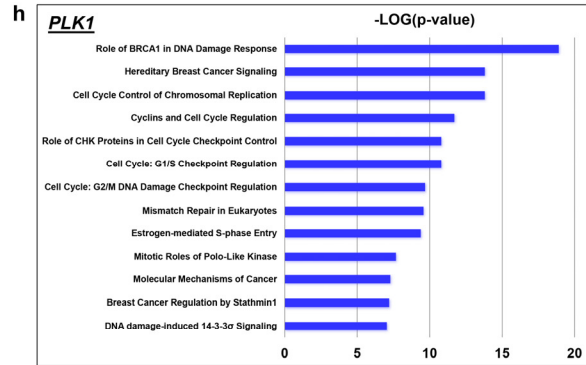
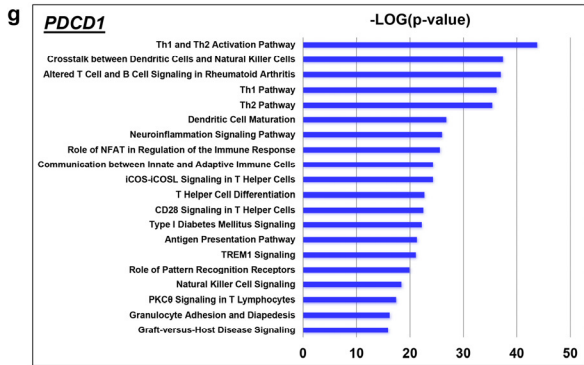


Figure A1 | Functions of 10 genes calculated using the *ab initio* GO method and the Canonical Pathways analysis of IPA. The bars express the p -values for the null hypotheses that the genes are not involved in the pathways. The pathways whose p -values are less than $1E-7$ are represented for up to 25 pathways for each gene. **a**, *CCNB1*. **b**, *CSF2*. **c**, *CXCR4*. **d**, *DNMT3A*. **e**, *IL2RA*. **f**, *ITGB2*. **g**, *PDCD1*. **h**, *PLK1*. **i**, *RAD21*. **j**, *UHRF1*.

Acknowledgments

This paper was notarized by a notary appointed by the Minister of Justice in Japan on April 28, 2022. After that, a paper appeared that demonstrated a mechanism of *NRF2* tumoral immunosuppression through upregulation of *KYNU* in LUAD [105]. Our work has a considerably extended scope than their findings by applying a comprehensive informatics approach, thereby supporting that *KYNU* is an immune checkpoint. In a planned revision of this manuscript, we will provide more robust evidence that *KYNU* has a primary role in cancer.

We thank H. Chiba and T. Nakayama for informatic programming. We acknowledge T. Okada, Y. Matsuoka, H. Yoshida, and I. Wada for valuable comments. We are grateful to Y. Ono and M. Sato for meaningful discussions. We are appreciative of N. Sato for drawing figures. We thank M. Garcia and G. Pennycook from Edanz (<https://jp.edanz.com/ac>) for correcting the manuscript.

Author contributions

TM conceived the study and constructed the theory. TK assisted in the construction of the theory. TM, TK, KA, YS, and TH designed the algorithm and performed the numerical computation and mathematical calculation. TM, TK, SG, and DDI wrote the manuscript. TM and DDI discussed the biological application of the theory. DDI developed the cloud platform for data sharing and informatics calculation. SG validated the results of the calculation through experiments. HH constructed the differential equations for the Moran process. KI verified the results of the analysis from a medical point of view. JS and ST supervised the research. Percentage contributions are TM: 95%, and TK, DDI, SG, HH, KI, JS, ST: 5%. All authors read and approved the final manuscript.

Conflict of interest

The authors declare no potential conflict of interests.

References

1. Van Wylen, G. J., Sonntag, R. E. & Borgnakke, C. *Fundamentals of classical thermodynamics*. 4th edn, (Wiley, 1994).
2. Maxwell, J. C. VIII. A dynamical theory of the electromagnetic field. *Philosophical transactions of the Royal Society of London*, 459-512 (1865).
3. Salam, A. & Ward, J. C. Weak and electromagnetic interactions. *Il Nuovo Cimento* **11**, 568-577, doi:10.1007/bf02726525 (1959).

4. Weinberg, S. A Model of Leptons. *Physical Review Letters* **19**, 1264-1266, doi:10.1103/PhysRevLett.19.1264 (1967).
5. Georgi, H. & Glashow, S. Unity of All Elementary Particle Forces. *Physical Review Letters* **32**, 438-441 (1974).
6. Buras, A. J., Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. Aspects of the grand unification of strong, weak and electromagnetic interactions. *Nuclear Physics B* **135**, 66-92, doi:10.1016/0550-3213(78)90214-6 (1978).
7. Schwarz, J. H. Superstring theory. *Physics Reports* **89**, 223-322, doi:10.1016/0370-1573(82)90087-4 (1982).
8. Kaku, M. *Introduction to superstrings and M-theory*. (Springer Science & Business Media, 2012).
9. Witten, E. String theory dynamics in various dimensions. *Nuclear Physics B* **443**, 85-126 (1995).
10. AccessScience, E. Unification theories and a theory of everything. doi:10.1036/1097-8542.Br0814141 (2014).
11. Collaboration, T. A. Search for squarks and gluinos in final states with jets and missing transverse momentum using 139 fb⁻¹ of $\sqrt{s}=13$ TeV pp collision data with the ATLAS detector. *arXiv* (2021).
12. Shannon, C. E. A Mathematical Theory of Communication. *Bell System Technical Journal* **27**, 379-423, doi:10.1002/j.1538-7305.1948.tb01338.x (1948).
13. Müller, I. *A history of thermodynamics: the doctrine of energy and entropy*. (Springer Science & Business Media, 2007).
14. Satz, H. Quarkonium binding and entropic force. *The European Physical Journal C* **75**, doi:10.1140/epjc/s10052-015-3424-7 (2015).
15. Neumann, R. M. Entropic approach to Brownian movement. *American Journal of Physics* **48**, 354-357 (1980).
16. Verlinde, E. On the Origin of Gravity and the Laws of Newton. *Journal of High Energy Physics* **2011**, 1-27 (2011).
17. Bekenstein, J. D. Information in the holographic universe. *Sci Am* **289**, 58-65, doi:10.1038/scientificamerican0803-58 (2003).
18. Gaudenzi, R. Entropy? Exercices de Style. *Entropy* **21**, 742, doi:10.3390/e21080742 (2019).
19. Parrondo, J. M., Horowitz, J. M. & Sagawa, T. Thermodynamics of information. *Nature physics* **11**, 131-139 (2015).
20. Toyabe, S., Sagawa, T., Ueda, M., Muneyuki, E. & Sano, M. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nature physics* **6**, 988-992 (2010).
21. Cover, T. M. *Elements of information theory*. (John Wiley & Sons, 1999).
22. Gibbs, J. W. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics*. (Yale University Press, 1914).

23. Blanco, D. D., Casini, H., Hung, L.-Y. & Myers, R. C. Relative entropy and holography. *Journal of High Energy Physics* **2013**, 1-65 (2013).
24. Lin, J., Marcolli, M., Ooguri, H. & Stoica, B. Locality of Gravitational Systems from Entanglement of Conformal Field Theories. *Phys Rev Lett* **114**, 221601, doi:10.1103/PhysRevLett.114.221601 (2015).
25. Lashkari, N., Lin, J., Ooguri, H., Stoica, B. & Van Raamsdonk, M. Gravitational positive energy theorems from information inequalities. *Progress of Theoretical and Experimental Physics* **2016** (2016).
26. Witten, E. A mini-introduction to information theory. *La Rivista del Nuovo Cimento* **43**, 187-227 (2020).
27. Gaveau, B., Jacobson, T., Kac, M. & Schulman, L. Relativistic extension of the analogy between quantum mechanics and Brownian motion. *Physical Review Letters* **53**, 419 (1984).
28. Inoue, T. & Collaboration, H. Q. in *AIP Conference Proceedings*. 020002 (AIP Publishing LLC).
29. Ade, P. A. *et al.* Planck 2013 results. I. Overview of products and scientific results. *Astronomy & Astrophysics* **571**, A1 (2014).
30. Weinberg, S. *The quantum theory of fields*. Vol. 2 (Cambridge university press, 1995).
31. Morse, P. M. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical review* **34**, 57 (1929).
32. Hellmann, R., Bich, E. & Vogel, E. Ab initio potential energy curve for the helium atom pair and thermophysical properties of dilute helium gas. I. Helium–helium interatomic potential. *Molecular Physics* **105**, 3013-3023 (2007).
33. Hurly, J. J. & Mehl, J. B. 4He thermophysical properties: new ab initio calculations. *Journal of research of the National Institute of Standards and Technology* **112**, 75 (2007).
34. Granados, V. & Aquino, N. The correspondence between the states of the two-dimensional isotropic harmonic oscillator and the Morse potential. *Journal of Molecular Structure: THEOCHEM* **493**, 37-41 (1999).
35. Nalewajski, R. F. in *Frontiers of quantum chemistry* 315-351 (Springer, 2018).
36. Church, K. & Hanks, P. Word association norms, mutual information, and lexicography. *Computational linguistics* **16**, 22-29 (1990).
37. Hamada, K.-j. BRST Conformal Symmetry as A Background-Free Nature of Quantum Gravity. *arXiv preprint arXiv:1707.06351* (2017).
38. Bezrukov, F. & Shaposhnikov, M. The Standard Model Higgs boson as the inflaton. *Physics Letters B* **659**, 703-706 (2008).
39. Starobinsky, A. A. A new type of isotropic cosmological models without singularity. *Physics Letters B* **91**, 99-102 (1980).

40. Hinshaw, G. *et al.* Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series* **208**, 19 (2013).
41. Aghanim, N. *et al.* Planck 2018 results-VI. Cosmological parameters. *Astronomy & Astrophysics* **641**, A6 (2020).
42. Hazumi, M. *et al.* Litebird: A satellite for the studies of b-mode polarization and inflation from cosmic background radiation detection. *Journal of Low Temperature Physics* **194**, 443-452 (2019).
43. Puspup, X. M., Villegas, K. H. & Paraan, F. N. Entanglement spectrum and number fluctuations in the spin-partitioned BCS ground state. *Physical Review B* **90**, 155123 (2014).
44. Di Tullio, M., Gigena, N. & Rossignoli, R. Fermionic entanglement in superconducting systems. *Physical Review A* **97**, 062109 (2018).
45. Kimura, M. *The neutral theory of molecular evolution.* (Cambridge University Press, 1983).
46. Strachan, T. & Read, A. Human molecular genetics. Garland science. *Edition, Kapitel 13*, 418 (2011).
47. Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713-719, doi:10.1093/genetics/47.6.713 (1962).
48. Maloof, A. C. *et al.* The earliest Cambrian record of animals and ocean geochemical change. *Geological Society of America Bulletin* **122**, 1731-1774 (2010).
49. Nikoh, N. *et al.* An estimate of divergence time of Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by aldolase and triose phosphate isomerase clocks. *Journal of molecular evolution* **45**, 97-106 (1997).
50. Diamond, J. *The third chimpanzee for young people: On the evolution and future of the human animal.* (Seven Stories Press, 2014).
51. Jukes, T. H. Neutral changes during divergent evolution of hemoglobins. *J Mol Evol* **11**, 267-269, doi:10.1007/BF01734488 (1978).
52. Ohta, T. Effect of initial linkage disequilibrium and epistasis on fixation probability in a small population, with two segregating loci. *Theor Appl Genet* **38**, 243-248, doi:10.1007/BF01245624 (1968).
53. Kolmogoroff, A. ber die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen* **104**, 415-458 (1931).
54. Moran, P. A. P. The statistical processes of evolutionary theory. *The statistical processes of evolutionary theory.* (1962).
55. Haeno, H., Maruvka, Y. E., Iwasa, Y. & Michor, F. Stochastic Tunneling of Two Mutations in a Population of Cancer Cells. *PLoS One* **8**, e65724, doi:10.1371/journal.pone.0065724 (2013).
56. Komarova, N. L., Sengupta, A. & Nowak, M. A. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *J Theor Biol* **223**, 433-450, doi:10.1016/s0022-5193(03)00120-6 (2003).

57. Akashi, H., Osada, N. & Ohta, T. Weak selection and protein evolution. *Genetics* **192**, 15-31, doi:10.1534/genetics.112.140178 (2012).
58. Schopf, J. W. *Life's origin: the beginnings of biological evolution*. (Univ of California Press, 2002).
59. Huheey, J. E., Keiter, E. A., Keiter, R. L. & Medhi, O. K. *Inorganic chemistry: principles of structure and reactivity*. (Pearson Education India, 2006).
60. Bianconi, G. Entropy of network ensembles. *Phys Rev E Stat Nonlin Soft Matter Phys* **79**, 036114, doi:10.1103/PhysRevE.79.036114 (2009).
61. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non - protein - coding DNA and eukaryotic complexity. *Bioessays* **29**, 288-299 (2007).
62. Califano, A. & Alvarez, M. J. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer* **17**, 116-130, doi:10.1038/nrc.2016.124 (2017).
63. Rosenthal, R. *Meta-Analytic Procedures for Social Science Research* Sage Publications: Beverly Hills, 1984, 148 pp. *Educational Researcher* **15**, 18-20 (1986).
64. Altrock, P. M., Liu, L. L. & Michor, F. The mathematics of cancer: integrating quantitative models. *Nat Rev Cancer* **15**, 730-745, doi:10.1038/nrc4029 (2015).
65. Foo, J. *et al.* An Evolutionary Approach for Identifying Driver Mutations in Colorectal Cancer. *PLoS Comput Biol* **11**, e1004350, doi:10.1371/journal.pcbi.1004350 (2015).
66. Bozic, I. *et al.* Accumulation of driver and passenger mutations during tumor progression. *Proc Natl Acad Sci U S A* **107**, 18545-18550, doi:10.1073/pnas.1010978107 (2010).
67. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, p11-p11 (2013).
68. Fukui, K., Yonezawa, T. & Shingu, H. A molecular orbital theory of reactivity in aromatic hydrocarbons. *The Journal of Chemical Physics* **20**, 722-725 (1952).
69. Jaynes, E. T. Information theory and statistical mechanics. *Physical review* **106**, 620 (1957).
70. Jaynes, E. T. Information theory and statistical mechanics. II. *Physical review* **108**, 171 (1957).
71. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**, 101-113, doi:10.1038/nrg1272 (2004).
72. Goyama, S. & Kitamura, T. Epigenetics in normal and malignant hematopoiesis: An overview and update 2017. *Cancer Sci* **108**, 553-562, doi:10.1111/cas.13168 (2017).
73. Balasubramani, A. *et al.* Cancer-associated ASXL1 mutations may act as gain-of-function mutations of the ASXL1-BAP1 complex. *Nat Commun* **6**, 7307, doi:10.1038/ncomms8307 (2015).
74. Asada, S. *et al.* Mutant ASXL1 cooperates with BAP1 to promote myeloid leukaemogenesis. *Nature Communications* **9**, doi:10.1038/s41467-018-05085-9 (2018).
75. Fujino, T. *et al.* Mutant ASXL1 induces age-related expansion of phenotypic hematopoietic stem cells through activation of Akt/mTOR pathway. *Nature Communications* **12**, doi:10.1038/s41467-021-22053-y (2021).

76. de Ramon Francas, G., Alther, T. & Stoeckli, E. T. Calsyntenins Are Expressed in a Dynamic and Partially Overlapping Manner during Neural Development. *Front Neuroanat* **11**, 76, doi:10.3389/fnana.2017.00076 (2017).
77. Kim, H. *et al.* Calsyntenin-3 interacts with both α - and β -neurexins in the regulation of excitatory synaptic innervation in specific Schaffer collateral pathways. *Journal of Biological Chemistry* **295**, 9244-9262 (2020).
78. Lu, Z. *et al.* Calsyntenin-3 molecular architecture and interaction with neurexin Ialpha. *J Biol Chem* **289**, 34530-34542, doi:10.1074/jbc.M114.606806 (2014).
79. Rindler, M. J. *et al.* Calsyntenins are secretory granule proteins in anterior pituitary gland and pancreatic islet alpha cells. *J Histochem Cytochem* **56**, 381-388, doi:10.1369/jhc.7A7351.2007 (2008).
80. Witzke, K. E. *et al.* Quantitative Secretome analysis of activated Jurkat cells using click chemistry-based enrichment of secreted glycoproteins. *Journal of Proteome Research* **16**, 137-146 (2017).
81. Gummy, L. F. *et al.* The kinesin-2 family member KIF3C regulates microtubule dynamics and is required for axon growth and regeneration. *Journal of Neuroscience* **33**, 11329-11345 (2013).
82. Ikeda, D. D. *et al.* CASY-1, an ortholog of calsyntenins/alcadeins, is essential for learning in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences* **105**, 5260-5265 (2008).
83. Ohno, H. *et al.* Role of synaptic phosphatidylinositol 3-kinase in a behavioral learning response in *C. elegans*. *Science* **345**, 313-317, doi:10.1126/science.1250709 (2014).
84. MacDonald, J. W. & Ghosh, D. COPA—cancer outlier profile analysis. *Bioinformatics* **22**, 2950-2951 (2006).
85. Marin-Acevedo, J. A. *et al.* Next generation of immune checkpoint therapy in cancer: new developments and challenges. *J Hematol Oncol* **11**, 39, doi:10.1186/s13045-018-0582-8 (2018).
86. Hiam-Galvez, K. J., Allen, B. M. & Spitzer, M. H. Systemic immunity in cancer. *Nat Rev Cancer* **21**, 345-359, doi:10.1038/s41568-021-00347-z (2021).
87. He, J., Hu, Y., Hu, M. & Li, B. Development of PD-1/PD-L1 Pathway in Tumor Immune Microenvironment and Treatment for Non-Small Cell Lung Cancer. *Sci Rep* **5**, 13110, doi:10.1038/srep13110 (2015).
88. Breimer, L. H., Nousios, P., Olsson, L. & Brunnstrom, H. Immune checkpoint inhibitors of the PD-1/PD-L1-axis in non-small cell lung cancer: promise, controversies and ambiguities in the novel treatment paradigm. *Scand J Clin Lab Invest* **80**, 360-369, doi:10.1080/00365513.2020.1742369 (2020).
89. Zhang, B. *et al.* Predictive effect of PD-L1 expression for immune checkpoint inhibitor (PD-1/PD-L1 inhibitors) treatment for non-small cell lung cancer: A meta-analysis. *Int Immunopharmacol* **80**, 106214, doi:10.1016/j.intimp.2020.106214 (2020).

90. Lu, S. *et al.* Comparison of Biomarker Modalities for Predicting Response to PD-1/PD-L1 Checkpoint Blockade: A Systematic Review and Meta-analysis. *JAMA Oncol* **5**, 1195-1204, doi:10.1001/jamaoncol.2019.1549 (2019).
91. Al - Mansoob, M., Gupta, I., Stefan Rusyniak, R. & Ouhtit, A. KYNU, a novel potential target that underpins CD44 - promoted breast tumour cell invasion. *Journal of Cellular and Molecular Medicine* **25**, 2309-2314 (2021).
92. Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005-1009 (2009).
93. D'Amato, N. C. *et al.* A TDO2-AhR signaling axis facilitates anoikis resistance and metastasis in triple-negative breast cancer. *Cancer research* **75**, 4651-4664 (2015).
94. Liu, Q. *et al.* Comprehensive Analysis of the Expression and Prognosis for TDO2 in Breast Cancer. *Molecular Therapy-Oncolytics* **17**, 153-168 (2020).
95. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518-524 (2005).
96. Rose, D. P. The influence of sex, age and breast cancer on tryptophan metabolism. *Clin Chim Acta* **18**, 221-225, doi:10.1016/0009-8981(67)90161-1 (1967).
97. Liu, Y. *et al.* A novel role of kynureninase in the growth control of breast cancer cells and its relationships with breast cancer. *J Cell Mol Med* **23**, 6700-6707, doi:10.1111/jcmm.14547 (2019).
98. Schwarcz, R. The kynurenine pathway of tryptophan degradation as a drug target. *Current opinion in pharmacology* **4**, 12-17 (2004).
99. Monney, L. *et al.* Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* **415**, 536-541, doi:10.1038/415536a (2002).
100. Jin, H. T. *et al.* Cooperation of Tim-3 and PD-1 in CD8 T-cell exhaustion during chronic viral infection. *Proc Natl Acad Sci U S A* **107**, 14733-14738, doi:10.1073/pnas.1009731107 (2010).
101. Darlington, L. G. *et al.* On the Biological Importance of the 3-hydroxyanthranilic Acid: Anthranilic Acid Ratio. *Int J Tryptophan Res* **3**, 51-59, doi:10.4137/ijtr.s4282 (2010).
102. Zhai, L. *et al.* Molecular Pathways: Targeting IDO1 and Other Tryptophan Dioxygenases for Cancer Immunotherapy. *Clin Cancer Res* **21**, 5427-5433, doi:10.1158/1078-0432.CCR-15-0420 (2015).
103. Liu, M. *et al.* Targeting the IDO1 pathway in cancer: from bench to bedside. *J Hematol Oncol* **11**, 100, doi:10.1186/s13045-018-0644-y (2018).
104. Jacobs, K. R., Castellano-Gonzalez, G., Guillemin, G. J. & Lovejoy, D. B. Major Developments in the Design of Inhibitors along the Kynurenine Pathway. *Curr Med Chem* **24**, 2471-2495, doi:10.2174/0929867324666170502123114 (2017).
105. Fahrman, J. F. *et al.* Mutational Activation of the NRF2 Pathway Upregulates Kynureninase Resulting in Tumor Immunosuppression and Poor Outcome in Lung Adenocarcinoma. *Cancers (Basel)* **14**, 2543, doi:10.3390/cancers14102543 (2022).