ISSN: 3066-8115
Research Article

Science Set Journal of Medical and Clinical Case Studies

Supermetabolic Pathways (or Supermetabolons) as Novel Biomarkers for Drug Discovery and Personalized Medicine

Sungchul Ji

Professor Emeritus of Theoretical Cell Biology, Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, N.J

***Corresponding author:** Sungchul Ji, Ph. D., Professor Emeritus of Theoretical Cell Biology, Department of Pharmacology and Toxicology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, N.J.

Submitted: 21 September 2023 **Accepted:** 25 September 2023 **Published:** 30 September 2023

https://doi.org/10.63620/MKSSJMCCS.2023.1019

Citation: *Ji, S. (2023). Supermetabolic Pathways (or Supermetabolons) as Novel Biomarkers for Drug Discovery and Personalized Medicine. Sci Set J of Med Cli Case Stu 2(3), 01-06.*

Abstract

Recent developments in cell biology suggest that living cells use three categories of structures similar to words, sentences, and texts in human language. Functionally, enzymes correspond to words, metabolic pathways (or metabolons more briefly) correspond to sentences, and systems of metabolons (called supermetabolons) correspond to texts or computer programs.

Supermetabolons that can be identified using the computer algorithm known as the Planck- Shannon classifier applied to mRNA expression profiles have been found to serve as novel biomarkers for either beneficial or toxic effects of the anti-cancer drug doxorubicin in breast cancer patients. These findings serve as the empirical and theoretical basis for formulating the new strategies for drug discovery and personalized pharmacotherapy referred to as the transcriptome-based Planck-Shannon classifier method.

Introduction

The concept of metabolic pathways (or metabolons) is well known in biomedical sciences. But the concept of 'supermetabolic pathways' (or 'supermetabolons') defined as systems of metabolic pathways (or systems of metabolons) is rarely discussed, most likely because there has been no mathematical method to identify them, until 2020 when the Planck- Shannon classifier (PSC) was proposed [1].

The first clear evidence for the existence of supermetabolons was provided by the mRNA data measured from breast cancer patients by Perou and his group at Stanford in 2000 [2]. They measured the transcriptome (i.e., the totality of RNA data) of 20 breast cancer patients before and after treating them with the anticancer drug, doxorubicin, for 16 weeks. They found that, of the 20 patients treated, 5 survived on average 10 months, while another group of 5 patients survived on average 70 months [2]. Out of more than two hundred metabolic pathways in the human transcriptome, my students at Rutgers and I analyzed the mRNA data belonging to the 10 metabolic pathways (listed in the legend to Figure 2 below) using the Planck-Shannon classifier [1]. We found that mRNA expression patterns (measured in breast cancer tissues of individual patients before drug treatment) were different between short and long surviving patients. This indicates that the information about how long a patient will survive after doxorubicin treatment appears to be encoded in the dynamic gene expression patterns (i.e., mRNA expression profiles) measured in breast cancer tissue of each patient before drug treatment. For example, we found that Pathways 3, 4 and 5 were activated or expressed only in long survivors, whereas Pathway 7 was activated in short survivors only. These results establish the concept of "supermetabolic pathways" (SMPs) or supermetabolons (SMs) acting as biomarkers for long and short survivors (discussed further in Section 5 below).

One of the main objectives of this manuscript is to propose a drug discovery strategy based on supermetabolic pathways or supermetabolons rather than on conventional metabolic pathways, individual enzymes, or receptors. The approach based on conventional metabolic pathways, enzymes, or receptors has been found to be disappointingly inefficient. For example, the cost of developing a new drug via the conventional method has been estimated to be $1.7 billion and that it takes 12-16 years to complete a drug development process from the compound discovery stage to marketing, and the overall attrition rate for developing a drug has been estimated to be 10,000:1 [3, 4]. If the drug discovery strategy based on supermetabolons is successfully implemented, it is predicted that the cost of developing a drug would be reduced by a factor of at least 100, i.e., to $170 million, and the development time may be shortened to 3-5 years.

Planckian Distribution Equation (PDE)

The Planck-Shannon classifier (PSC) is an algorithm derived from the Planckian Distribution Equation (PDE) which, in turn, is derived from the blackbody radiation equation discovered by M. Planck (1858-1947) in 1900 [1, 5]. PSC can map a set of three or more long-tailed histograms (LTH's) into one or more categories of functions or properties, each category constituting one of the points one a regression line on the Planck-Shannon plot (see Panels c and d in Figure 2). The Planck-Shannon plot consists of a 2-dimensional graph with the Shannon entropy, H, plotted on the x-axis and the Planckian information of the second kind, IPS, plotted on the y-axis. Both H and IPS care computed from the Planckian Distribution Equation (PDE) that fits the data represented by the LTH under analysis [1].

PDE was derived from the blackbody radiation equation (BRE) (see Equation (1) in Figure 1) by replacing the universal constants and temperature with free parameters A, B, and C (see Equation (2) in Figure 1) [1, 6, 7]. The derivation of PDE from BRE was motivated by the similarity between the shape of the blackbody radiation spectra (see Panels a and c in Figure 1) and that of the single-molecule enzyme turnover histogram (see Panels c and the blue curve in panel d in Figure 1)

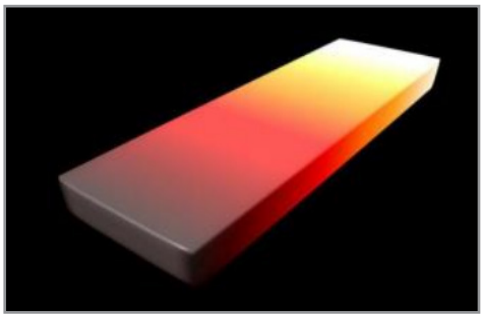
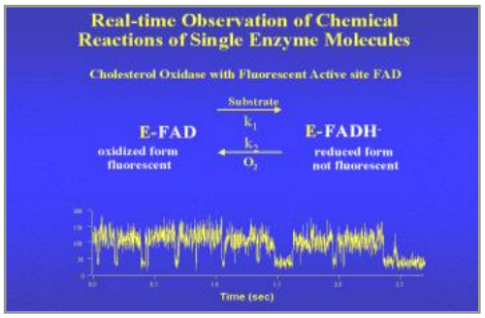
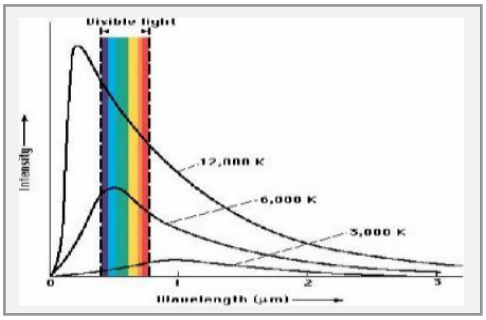
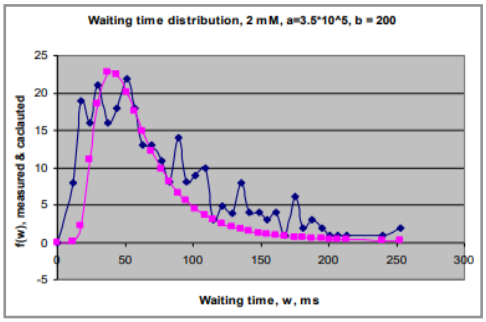
Observation	a) 	b) 
Data	c) 	d) 
Mathematical Equation	**e) Planck radiation equation (PRE):** $u(\lambda, T) = (2hc^2/\lambda^5)/(e^{hc/\lambda kT} - 1)$ **(1)** u = intensity of radiation; h = Planck constant; c = speed of light; λ = wavelength; k = Boltzmann constant; T = temperature	**f) Planckian Distribution Equation (PDE):** $y = (A/(x + B)^5)/(e^{C/(x+B)} - 1)$ **(2)** y = frequency x = bin number A, B, and C = free parameters

Figure 1: The isomorphism between blackbody radiation (left column) and single-molecule enzyme catalysis (right column).

(a) Blackbody radiation [5].

(b) Single-molecule enzyme catalysis [8]. The measurement of the turnover of a cholesterol oxidase (COx) molecule in the presence of cholesterol (0.20 mM) and oxygen (0.25 mM). The prosthetic group, FAD, of COx is fluorescent when in its oxidized state (which is referred to as the "on" state) and non-fluorescent when in its reduced state (which is referred to as the "off" state). Reproduced from http://www.nigms.nih.gov/News/Reports/single_molecules.htm.

(c) Blackbody radiation spectra [5].

(d) The histogram of the COx's "on" and "off" times. Reproduced from http://www.nigms.nih.gov/News/Reports/single_molecules.htm.

(e) The blackbody radiation equation (BRE) that fits blackbody radiation spectra discovered by M. Planck in 1900 [5]. The Planckian Distribution Equation (PDE) derived from BRE in 2008 [6, 7].

Planck-Shannon Classifier

The deviation of Planckian Distribution Equation (PDE) from a symmetric curve such as the Gaussian distribution function can be used as a measure of non-randomness and hence of order and information [8-11]. There are two ways of quantifying the information content of PDE:

(i) The Planckian Information of the First Kind (IPF) defined as the binary logarithm of the ratio of the area under the curve (AUC) of PDE to that of Gaussian-like equation (GLE), Eq. (3):

$$y = Ae^{-(x-\mu)^{\wedge}2/(2\sigma^{\wedge}2)} \quad \textbf{(3)}$$

where A is a free parameter, μ is the mean and σ is the standard deviation.

(ii) The Planckian Information of the Second Kind (I_{PS}) defined as the negative binary logarithm of the skewedness of the long-tailed histogram [10, 11].

$$\mathbf{IPS = - \log2 (|\mu - mode|/\sigma)} \quad \mathbf{(4)}$$

It is often more convenient and more reproducible to calculate IPS from PDE than to calculate I_{PF}.

The Shannon entropy associated with PDE can be calculated based on Equations (5) and (6):

$$\mathbf{H = -\Sigma\, p_i \log_2 p_i} \quad \mathbf{(5)}$$

where p_i is the probability of observing the i^{th} data bin calculated as

$$\mathbf{p_i = y_i / \Sigma y_i} \quad \mathbf{(6)}$$

where y_i is the frequency of the i^{th} data bin and the index i runs from 1 to n, the total number of the data bins.

Each long-tailed histogram (LTH) associated with some function (e.g., a metabolic pathway) can be plotted as a point in the Planck-Shannon plane (also called Planck-Shannon plot, graph, or space) (see Panels c and d, Figure 2).

The Planck-Shannon Plot (PSP)

Once a long-tailed histogram (LTH) is fitted into PDE, two numbers, Planckian information of the second kind, IPS, and Shannon entropy, H, can be computed from the resulting PDE as described in Section 3. With H and IPS computed, the Planck-Shannon plots can be constructed as shown in Panels c and d in Figure 2. The Planck-Shannon classifier (PSC) is an algorithm that maps three or more LTH's into one or more categories with characteristic properties or functions, each category constituting a part of a linearly correlated line on the Planck-Shannon plane (see Panel c in Figure 2). When a set of points forms a linear regression line in the Planck-Shannon plot (PSP), such a set represents a 'superstructure' or 'supermetabolic pathway' with a function beyond the sum of the functions of the individual LTH's [1, 9, 10, 11].

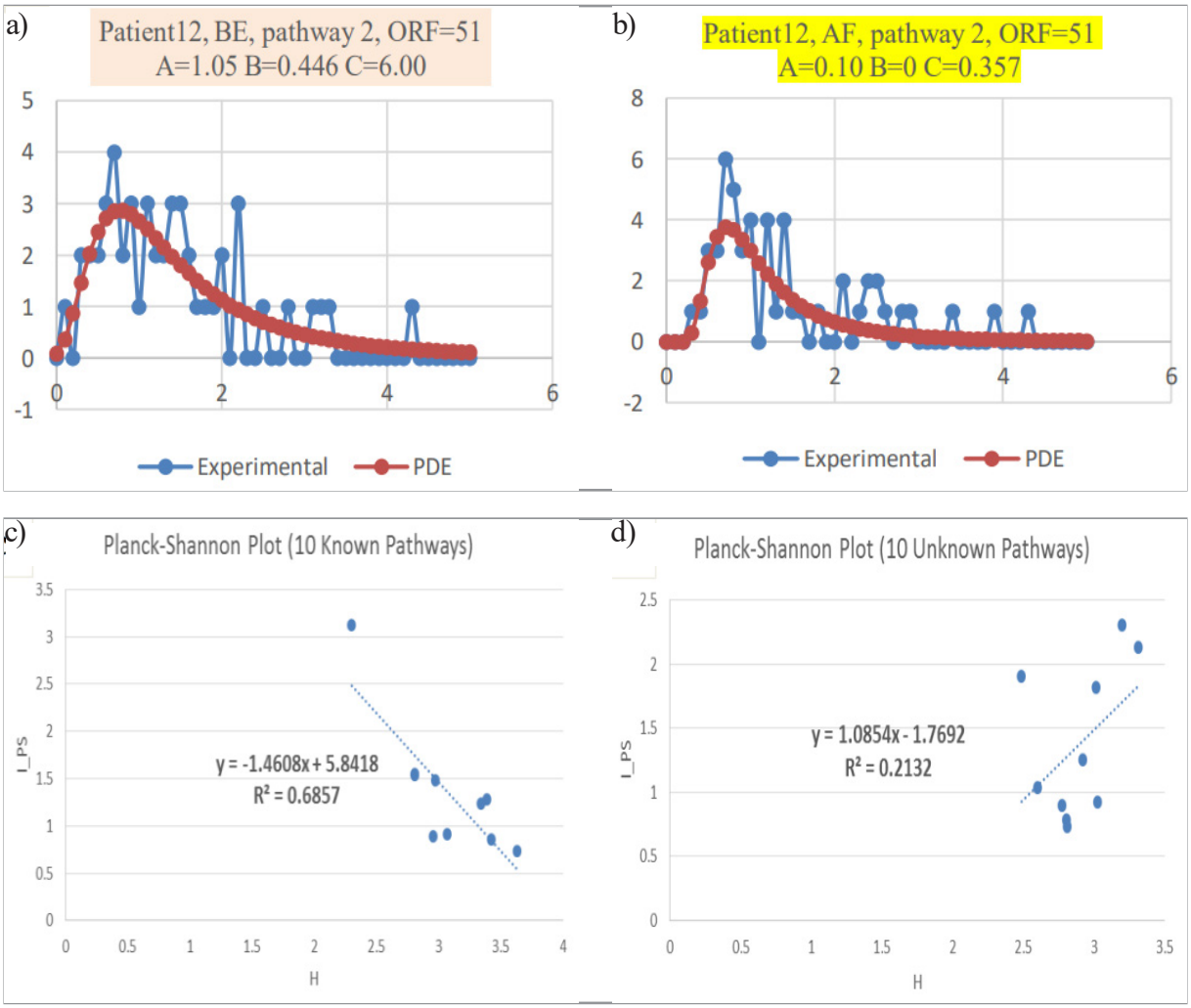


Figure 2: Transforming long-tailed histograms (LTHs) to the Planck-Shannon plot mediated by the histogram-fitting PDE.

a) The histogram and associated PDE of the mRNA levels encoded by cell wall biosynthesis pathway with 51 ORF's (open reading frames) of a breast cancer patient before being treated with doxorubicin [2]. Blue curve = original data; red curve = simulated by PDE.

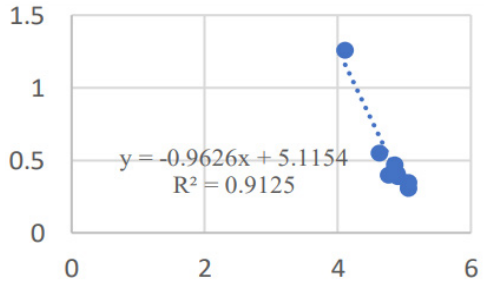
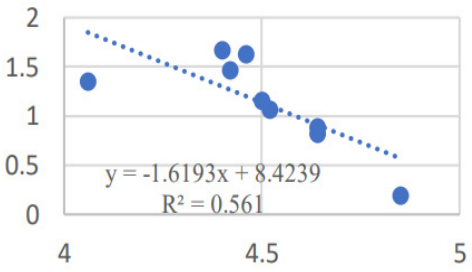
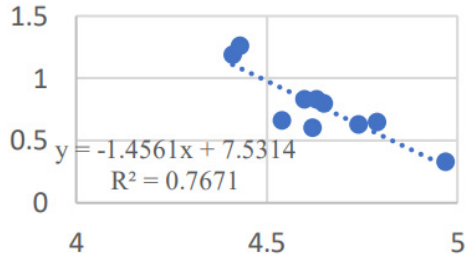
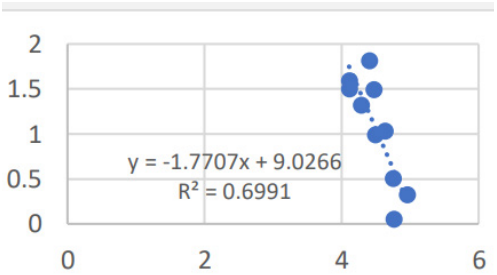
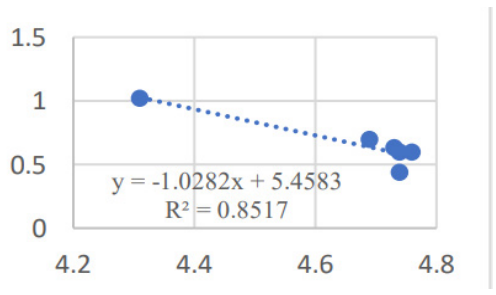
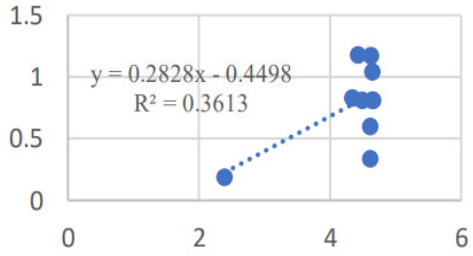
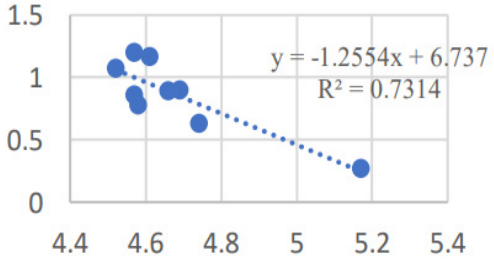
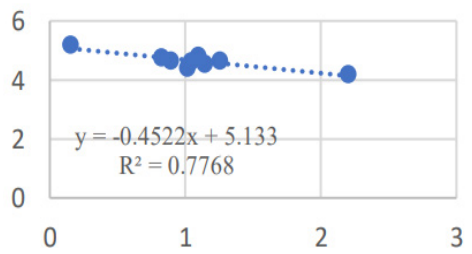
b) The same patient as in (a) except that the mRNA levels of the patient were measured 16 weeks after doxorubicin treatment [2]. Blue curve = original data; red curve = simulated by PDE. It is important to note that plots (a) and (b) cannot be distinguished by visual inspection but the associated PDE's are quantitatively different as shown by the numerical values of their free parameters A, B, and C.

c) The Planck-Shannon plot of 10 metabolic pathways analyzed: 1 = cell cycle, 72 ORFs; 2= cell wall biogenesis, 53 ORFs; 3 = chromatin structures, 44 ORFs; 4 = cytoskeletons, 71 ORFs; 5 = DNA repair, 32 ORFs; 6 = rRNA processing, 37 ORFs; 7 = nuclear protein targeting, 43 ORFs; 8 = protein synthesis, 156 ORFs; 9 = transport, 129 ORFs; and 10 = transcription, 175 ORFs. These data sets all fitted PDE, using the Solver software available in Excel, thereby generating 10 pairs of the IPS and H values, which were plotted as 10 points as shown in Panel c which generated a regression line with a R2 value of 0.7066, thus forming a supermetabolic pathway or a supermetabolon.

d) The Planck-Shannon plot of 10 sets of mRNA levels similar in sizes to the 10 sets used in (c) but with unknown functions. The fact that the R2 value is only 0.021 indicates that this set of 10 points have no correlation among them and hence cannot form a supermetabolic pathway or a supermetabolon, more briefly.

The Drug-Sensitive and Patient-Specific (DSPS) Supermetabolons

When the 10 metabolons were plotted on the Planck-Shannon plane, one or more metabolons had to be removed in order to obtain a good linear regression line with R2 values greater than 0.6 - 0.7. For example, one metabolon was removed from 10 in Patients 2 and 15 (see Figure 3), while 3 metabolons were removed in Patient 14. Therefore, in Patients 2 and 15, the supermetabolons consists of 9 metabolons, while, in Patient 14, the supermetabolon consists of 7metabolons. In Patients 5 and 11, no metabolon was removed, hence their supermetabolons consist of 10 metabolons. Treating with doxorubicin destroyed supermetabolons in Patients 11 and 14 (i.e., 40% of the total patients) but did not affect supermetabolic pathways in the other 3 patients. This phenomenon will be referred to as the "drug-sensitivity and patient-specificity" (DSPS) of supermetabolons.

	Before Drug Treatment	After Drug Treatment
Patient 2 (30; 9)		
Patient 2 (30; 9)		
Patient 11 (66; 10)		
Patient 15 (17; 9)		

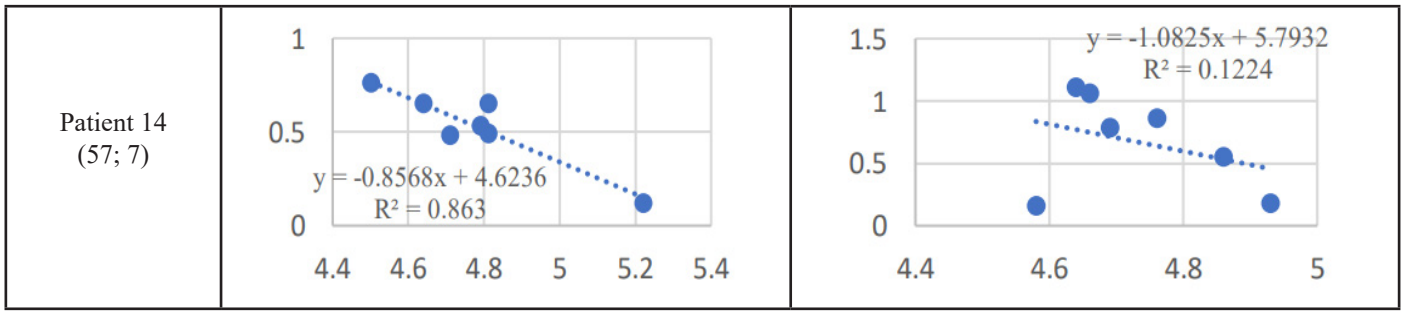


Figure 3: Typical examples of the Planck-Shannon plots. The Planck-Shannon plots of 7-10 metabolic pathways (for each histogram) before and after treating with doxorubicin. Patient # (survival months; number of pathways in the plot). The x-axis encodes the Shannon entropy, H, and the y-axis encodes the Planckian information of the second kind, I_{PS}, [1, 9, 10, 11]. Doxorubicin destroyed the super-metabolic pathways in Patients 11 and 14.

Not all supermetabolons are drug-sensitive (see Patients, 5 and 11). The drug sensitivity of supermetabolons will be found essential for the application of PSC to drug discovery research (as described in Section 7, while the patient-specificity of supermetabolic pathways will be essential for the application of PSC to personalized drug therapy as described in Section 8).

Longevity-Determining Supermetabolons

When 20 breast cancer patients were treated with doxorubicin for 16 weeks, 5 patients survived on average 10 months and another 5 patients survived on average 70 months [2]. The transcripts of 10 metabolons showed different patterns of expression in short and long survivors as analyzed by the Planck-Shannon classifier (PSC) applied to the mRNA data measured from the breast cancer tissues from these patients before doxorubicin treatment. PSC transforms an mRNA data set measured from a given metabolon into a point in the e Planck-Shannon plot as already mentioned. A pair of H and IPS was calculated from the Planck Distribution Equation (PDE) [1] that best fits the histogram generated from the RNA data of a given metabolic pathway. Once a given set of 30 or more data points in a histogram is found to fit PDE, IPS, and H are computed from it as described above. We found that Pathways 3, 4, and 5 formed a supermetabolon (operationally defined as a set of points on the Planck-Shannon plot that lies on a regression line) among the 5 long survivors, whereas only metabolon 7 was commonly expressed in 5 short survivors.

Thus we can predict that any breast cancer patients whose breast tissues mRNA levels analyzed before drug treatment show the activation of metabolons 3, 4, or 5 will most likely survive for about 70 months. This method can also be employed for drug discovery by searching for drug candidates that mimic the behavior of doxorubicin on the Planck-Shannon plot [6-9].

Drug Discovery

The Planck-Shannon classifier as applied to transcriptomic data can be applied to the following 5 levels:

I. **Discovery** of New Drugs
II. Discovery of **natural agents** (to be called bioceuticals) that can replace FDA- approved drugs to avoid their long-term toxicity,
III. **Repurposing** FDA-approved drugs for treating diseases other than those for which drugs were originally approved.
IV. **Personalized pharmacotherapy,** i.e., finding the drug out of many available drugs that can most efficaciously treat the disease of a given patient,
V. **Promotion**, to increase the therapeutic ratio of discarded drug candidates to the level sufficient to pass the threshold for FDA approval.

Due to space limitation, only (i) and (iv) are described in detail below.

The key idea behind the drug discovery strategy based on the Planck-Shannon classifier is that any drug candidates that generate the same supermetabolons on a set of cancer cell cultures that are identical or similar to those produced by any FDA-approved anticancer drugs can be predicted to be effective anticancer drugs at the clinical level. This new strategy involves the following key steps:

I. Select a drug for treating disease X tested on M cell cultures, i.e., select D(X, M).
II. Select M (about 20) cell cultures derived from tissues carrying disease X, i.e., select C(X, M). Prepare M cell cultures so that they will exhibit a distribution of anti-cancer drug efficacies just as the 20 breast cancer cells employed by Perou et al. [2].
III. Select a set of N (about 20) metabolic pathways, i.e., MP(N), where N can run from 1 to 10.
IV. Determine the supermetabolons (SM) characteristic of disease X using N metabolons and M (abut 20) cell cultures characteristic of disease X, i.e., find SM(X, M, N) as described above.
V. Find potential drug for X, i.e., PD(X), that produces the same Supermetabolon as D(X, M, N), i.e., find PD(X, M, N). It is predicted that PD(X, M, N) will be found to be as effective as D(X, M, N), when tested clinically.

Personalized Medicine

The objective of personalized pharmacotherapy is to find the best among the Y FDA-approved drugs for treating disease X that is most therapeutically efficacious for patient Z, i.e., to find BD(X, M, N, Z), BD standing for 'best drug'.

I. Select a set of M cell cultures derived from tissues carrying disease X.
II. Select Y FDA-approved drugs for disease X that produce Y toxic side effects on M cell cultures when examined with N metabolic pathways, i.e., determine toxic SM(X, M, N, Y, Z), where Y is predicted to range from 0 (zero toxicity) to 10 (most toxic).

III. Calculate the therapeutic indexes (i.e., Toxic dose/Effective dose) of the Y FDFA- approved drugs based on the information embodied in SM(X, M, N, Y, Z) and select the drug with the largest therapeutic index for Patient Z.

Conclusions

The drug discovery strategy and the new approach to pharmacotherapy proposed in this paper is based on (i) the cell language theory and (ii) the Planck-Shannon classifier developed in the past couple of decades at Rutgers University (see Table 1 below). The cell language theory predicts the existence of systems of metabolic pathways (called supermetabolons) that are responsible for cellular computing or cellular reasoning underlying human health or diseases and the Planck-Shannon classifier can identify supermetabolons based on the transcriptome of patients. Our preliminary data accumulated during the period 2010-2022 indicate that the transcriptome-based Planck-Shannon classifier method of drug discovery advocated here, when successfully implemented, could reduce the cost of drug development from current $1.2 billion down to $100 million or less per drug and the development time from the current 12-16 years to 3-5 years [10-13].

Table 1: The supermetabolon* approach to drug discovery and personalized pharmacotherapy.

Organization Level	Components	Linguistic Analogy	Characterization Method	Disease due to malfunctioning components
I	enzymes	words (to denote)	chemistry	Level I disease
II	metabolons	sentence (to decide)	biochemistry	Level II disease
III	supermetabolons	texts (to compute or think)	mathematics and computing (e.g., PlanckShannon classifier)	Level III disease

References

1. Ji, S. (2020). The Planck-Shannon plot: A quantitative method for identifying 'superstructures' in cell biology and consciousness study. Cosmos & History 16: 142-164.
2. Perou, C. M., Sørlie, T., Eisen, M. B., de Rijn, M. V., & Jeffrey, S. F. (2000). Molecular portraits of human breast Tumors. Nature 406: 747-752.
3. Bain & Company (2007) Cost of Developing a New Drug Increases to About 1.7 Billion, http://www.loyaltyrules.com/bainweb/publications/publications_detail.asp?id=14245&menu_url=publications_results.asp.
4. Chengalvala, M. V., Chennathukuzhi, V. M., Johnston, D. S., Stevis, P. E., & Kopf, G. S. (2007). Gene Expression Profiling and its Practice in Drug Development. Current Genomics 8: 262-270.
5. Ji, S. (2015). Planckian distributions in molecular machines, living cells, and brains: The wave-particle duality in biomedical sciences. In: Proceedings of the International Conference on Biology and Biomedical Engineering, Vienna, March 15-17: 115-137.
6. Ji, S. (2018). The Cell Language Theory: Connecting Mind and Matter. World Scientific Publishing New Jersey 333-339.
7. Peter Lu H, Luying Xun, Sunney Xie X (1998) Single-Molecule Enzymatic Dynamics. Science 282: 1877-1882.
8. Ji, S.(2018). Mathematical (Quantitative) and Cell Linguistic (Qualitative) Evidence for Hypermetabolic Pathways as Potential Drug Targets. J Mol Genet Med 12: 343.
9. Ji, S. (2015). Planckian Information (Ip): A New Measure of Order in Atoms, Enzymes, Cells, Brains, Human Societies, and the Cosmos. World Scientific 579-589.
10. Ji, S. (2017). Planckian Information (IP): A Measure of the Order in Complex Systems. World Scientific 53-97.
11. Srere PA (1987) Complexes of Sequential Metabolic Enzymes. Ann. Rev Biochem 56: 89-124.
12. Ji, S. (2012). Molecular Theory of the Living Cell: Concepts, Molecular Mechanisms, and Biomedical Applications. Springer, New York.
13. Ji, S. (2023). Planck-Shannon Classifier can Distinguish between Systems and Supersystems: Applications to Drug Discovery and Personalized Medicine. Proceedings of Systemics, Cybernetics and Informatics-2023, September 12-15, 2023, Orlando, Florida.