学术展望

ACADEMIC PERSPECTIVES

美国华裔教授学者协会学术期刊

USA CHINESE SCHOLARS ASSOCIATION

Los Angeles, California, USA, September 12, 2023

ACADEMIC PERSPECTIVES

USA Chinese Scholars Association

Volume 16, 2023

ISBN 1936-1246

Table of Contents

Dirac’s Large Number Hypothesis: An Ongoing Quest for Correlations Between the Infinitesimal and the Infinite

Jonathan H. Jiang, Ethan Liu, and Yanbei Chen………………................. 1

借鉴OC43 引发1889 年人类冠状病毒病（COVID）大流行史：探讨COVID-19 的中枢神经损伤机制

周炳亮，曹虹，黄胜和 ……………………………………………….............. 15

Virginia Woolf’s Between the Acts: A Trauma-Theory Case Study

James J. Lu ………………………………….. 27

癌症的靶向治疗和精准医学

刘西平 …………………………………………………………….................. 40

Comparisons of increased NOergic molecules and neuropeptides between stimuli-evoked axon reflex and stimuli-induced meridian phenomena/tracers along linear pathways resembling meridians: Toward mechanisms of meridian system

Sheng-Xing Ma …………………………………….. 49

Waveguide Effects of a Multi-Layered Fault Zone –Retrospective of the 1992 M7.4 Landers earthquake

Yong-Gang Li ………………………….. 67

S100蛋白家族在中枢神经系统疾病中的研究进展

张沐捷，王希萌， 梁嘉欣，田晓燕，陈宝珠，彭亮.. 82

Environmental Commentary: Paradox in human versus nature

Yue Rong…………………………………………... 94

中国教育：成就、问题与展望

陈岳云 …………………………………………….. 96

Dirac's Large Number Hypothesis: An Ongoing Quest for Correlations Between the Infinitesimal and the Infinite

Jonathan H. Jiang[1], Ethan Liu[2], and Yanbei Chen[3]
[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91099
[2]Conestoga High School, Berwyn, PA 19312
[3]Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125

*Correspondence: Jonathan.H.Jiang@jpl.nasa.gov

Keywords: Large Number Hypothesis, Infinitesimal, Infinite, Gravitational Constant, Continuous, Mass Creation, Anthropic Principle, Variability of Constants, Cosmological Parameters, Theoretical Physics, Universe Models, Einstein's Theory of General Relativity, Cosmic Time, Quantum Physics.

Abstract

Dirac's Large Number Hypothesis (LNH), proposed in 1937, has captivated the scientific community with its exploration of profound correlations between cosmic and atomic scales. This hypothesis delves into the intricate interplay between the infinitesimally small and the vastly large, offering potential insights into the fundamental nature of the universe. As we continue to uncover the mysteries of the cosmos, the LNH oscillates on the precipice of scientific acceptance—neither fully validated nor entirely dismissed. In this paper, we embark on a comprehensive journey through Dirac's LNH, shedding light on its theoretical underpinnings, its implications for universe models, and its resonance with the Anthropic Principle. We delve into the possibilities of variable gravitational constants and continuous mass creation, inviting further exploration into the intricacies of our cosmic symphony. By embracing the ongoing quest for understanding, we endeavor to unravel the profound harmonies that connect the infinitesimal with the infinite, contributing to the symphony of knowledge in theoretical physics and cosmology.

1. Introduction

The indefatigable march of scientific advancement is deeply intertwined with our inherent drive to unravel the fundamental principles that govern our universe. Monumental strides in both physics and mathematics over the preceding centuries have equipped us to wrestle with questions that span the quantum to the cosmic scale, bridging the realms of the infinitesimal and the infinite (Greene, 1999). These leaps in understanding, from Albert Einstein's revolutionary theory of relativity (Einstein, 1915) to Charles-Augustin de Coulomb's foundational work on electrostatics (Coulomb, 1785), have progressively drawn us closer to deciphering the complex cosmic tapestry that constitutes our universe.

Paul Adrien Maurice Dirac, a luminary of physics, proposed a captivating idea—cosmological parameters, colloquially referred to as atomic constants (Dirac, 1937), turned out to related to atomic constants by powers of $\sim 10^{40}$ (Table 1). These parameters appear to underpin the fundamental structure of our cosmos, yet the processes through which they sculpt our universe remain tantalizingly elusive.

In the dawn of the 20th century, several trailblazing physicists set out on a quest to understand the correlations between these cosmological parameters and the nature of our universe. Hermann Weyl was among these pioneers, suggesting that atomic constants seemed "coincidentally" proportionate to their macroscopic counterparts, hence generating ratios close to 40 orders of magnitude (Weyl, 1918).

Table 1: Examples of Large Numbers

Example of large numbers, computed as the radio of	Physical Constants	Large Number
The Electrostatic Force to the Gravitational Force between a Proton and an Electron	$(e^2/G)(m_e m_p)$	~10^{40}
The Planck Mass to the Proton Mass	m_P/m_p	~10^{19}
The Intensity of Electromagnetic - Gravitational Interaction between Elementary Particles	a/G	~10^{40}
The Mass of a Typical Star to the Electron Mass	M_*/m_e	~10^{60}
The Radius of the Observable Universe to the Radius of an Electron	R_U/r_e	~10^{40}
The Hubble Radius to the Planck Length	R_H/l_P	~10^{60}
The Mass of the Universe to the Proton Mass	M_U/m_p	~10^{80}
The Mass of the Universe to the Planck Mass	M_U/m_P	~10^{60}
The Planck Mass Density to the Observed Density of the Universe	ρ_P/ρ_{obs}	~10^{120}
The Planck Energy to the CMBR Temperature	E_P/T_{CMBR}	~10^{30}

In 1919, Weyl postulated that the ratio of the radius of the observable universe to the classical electron radius would yield a Large Number (Weyl, 1919). His computations unveiled a ratio of roughly 10^{42}. This seminal discovery catalyzed a series of similar revelations, cumulating in the unveiling of Eddington's number later that year (Eddington, 1919). However, it was Dirac's insightful work in 1937 that substantially deepened this line of thought (Dirac, 1937). His computations exposed that the ratio of the electrostatic to the gravitational force was around 10^{39}, and the ratio of the mass of the universe to the proton mass approximated a staggering 10^{78}.

These calculations were grounded in the observational data and constants accessible at the time (Dirac, 1937). Over the decades, these values have undergone refinement (Barrow, 2002), yet the scale of these numbers continues to captivate researchers. This paper plunges into the enduring allure and persistent debates surrounding Dirac's Large Number Hypothesis (LNH), as our pursuit of the universe's most profound mysteries remains undiminished.

2. A Captivating Voyage Through the Labyrinth of Improbabilities

The staggering magnitude of these numbers found in computations aroused the intrigue of numerous physicists, not least of which was Dirac (Table 1). Rarely do we stumble upon such prodigious quantities in observations of natural phenomena. It is this captivating pattern that marks the advent of the Large Number Hypothesis (LNH) (Dirac, 1937).

Dirac posited that these remarkable ratios could not be attributed merely to statistical aberration or pure coincidence. If that were the case, it would be highly unlikely for these ratios to remain constant over the astronomical timescales synonymous with cosmic evolution (Dicke, 1961). From this insight, Dirac made a pivotal progression in his hypothesis—the idea that these vast numbers might fluctuate in conjunction with the temporal progression of the universe.

Whether these fluctuations would lead to subtle or significant changes was an open question, yet it served as a cornerstone for the formulation of Dirac's LNH.

The aspiration of Dirac's hypothesis was threefold. The first objective was to decipher the enigma behind the tremendous magnitudes of these large numbers. What led these numbers to such enormity, and could their size potentially illuminate a fundamental facet of our cosmos (Barrow, 2002)? The second goal was to shed light on the implications of these seemingly serendipitous large numbers. Might they reveal an uncharted correlation between the microscopic and macroscopic scales of the universe (Carr, 2005)?

Arguably, the most ambitious objective was Dirac's third—to construct a model of the universe unhindered by anthropocentric constraints. He sought to envision a cosmic structure fundamentally in sync with natural laws and principles, unfettered by the limitations of human comprehension or technological capabilities (Dirac, 1937). Dirac's intent was to curate a cosmological model that reflects the unadulterated nature of the universe, thereby paving the way for new avenues in our quest to decode the cosmic enigmas.

Yet, despite the persuasive premises and the profound implications of Dirac's LNH, it remains a topic of ongoing discourse and investigation within the physics community. The forthcoming sections will delve into the hypothesis's subsequent evolution, the main points of contention, and its prospective ramifications on our understanding of the universe.

3. Unraveling Consequences of LNH

The implications of Dirac's LNH might, at first glance, appear rather opaque. However, when we scrutinize them, they reveal profound implications for both physics and mathematics. Accepting the LNH's premise suggests that as the universe ages, the physical constants that define our observable universe adjust to reflect Dirac's conceptualization of the 'epoch', positioned at a magnitude of $(10^{39})^n$ for some integer n. This notion accommodates the striking observation that all Large Numbers are of the order of 39 or 40 (Barrow, 2002).

Yet, this premise engenders an intriguing conundrum. If these 'constants' are engaged in a temporally dynamic dance, does the term 'constant' still apply? Rather, they transmogrify into time-dependent variables, thereby opening a Pandora's box of fundamental questions concerning the very nature of physical constants and, by extension, the universe itself (Uzan, 2003).

Decades later, in 1974, Dirac refined his insights on the LNH, rigorously examining two preconditions and their ensuing implications requisite for the hypothesis to withstand scrutiny (Dirac, 1974). The first condition revolves around the tenable models of the universe. The universe's magnitude, an integral parameter framing our understanding of the cosmic expanse we inhabit, must align intimately with the 'epoch', precluding a fixed value for the universe's size. As a result, any cosmological model fixating a specific constant as a cosmological parameter would conflict with the stringent criteria of LNH. Thus, only models of a non-static universe might fit into the LNH paradigm.

The second condition presents a formidable challenge. Dirac's LNH necessitates that all cosmological parameters gracefully morph into time-dependent variables. It doesn't require a significant stretch of imagination to deduce that adjusting even one atomic constant within the established physics framework would trigger far-reaching repercussions. Among the myriad of atomic constants, Dirac elected to investigate the gravitational constant (G), the bedrock of Einstein's general relativity, exploring its potential variability and thus adding another layer of complexity to the issue (Sandvik, Barrow, Magueijo, 2002).

With these two conditions at hand, the LNH guides us down a fascinating path hinting at the perpetual genesis of matter within the universe (Dirac, 1974). This spontaneous emergence of matter might manifest through two plausible mechanisms: "additive creation" and "multiplicative creation". Additive creation hypothesizes that matter arises uniformly throughout the universe, even in the ostensibly desolate intergalactic spaces. In contrast, multiplicative creation proposes that matter emerges where matter already exists, proceeding in proportion to the current atomic ensemble (Canuto, Hsieh, Adams, 1977).

While the scope of this paper precludes a deep dive into these mechanisms' nuances, a panoramic view is provided, and further exploration of each element is recommended through Saibal Ray's comprehensive 2019 review on Dirac's LNH (Ray, 2019).

Dirac's foundational 1947 publication paved the way for ongoing research on LNH (Dirac, 1947). These studies typically navigate one of three distinct terrains:

• Cosmological Model Considerations

• Gravitational Constant Variability

• Continuous Mass Creation

By shedding light on the research conducted through these unique lenses, we aim to provide a refreshed perspective on the current state of affairs and the potential future for Dirac's LNH within the esteemed domain of theoretical physics.

4. Cosmological Models under the Lens of Dirac's Large Number Hypothesis

As articulated earlier, the parameters underlying Dirac's LNH impose certain boundaries on the parameters of cosmological models. Specifically, any model governed by a static atomic constant is categorically dismissed under this hypothesis. Although these restrictions may appear to limit the range of potential models, they ensure that the surviving models align with empirical observations. Fundamentally, the hypothesis disallows models proposing a universe that expands to a maximum size and subsequently contracts, as such models would entail a cosmological constant that remains independent of the universe's age (Dirac, 1974).

One significant paradigm affected by this stipulation is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, more commonly recognized as the standard cosmological model. The FLRW metric describes the geometry of the universe and its expansion dynamics through a time-dependent scale factor ($a(t)$) (Friedmann, 1922; Lemaître, 1927; Robertson, 1933; Walker, 1937). However, according to Dirac's LNH, the conventional manifestation of the FLRW model cannot be endorsed (Dirac, 1974).

The FLRW metric can be expressed as:

$$ds^2 = -\, dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \qquad (1)$$

Here s is the distance between two infinitesimally close events in spacetime; t is the time coordinate, which is the time measured by an observer moving along with the expansion of the universe; a is the scale factor; r is the comoving radial coordinate describing the positions of objects in an expanding (or contracting) universe; k is the curvature constant, which can take values of -1, 0, or +1; θ and ϕ are the polar and azimuthal angles in spherical coordinates, respectively.

The evolution of $a(t)$ is governed by the content of the universe, which contains bright matter, dark matter, radiation, and dark energy. Note that dark matter and dark energy are contents whose nature are not well known, but needed to be introduced in order to explain the observed expansion of the universe, as well as the motions of stars in galaxies.

It is worth mentioning that if the scale-covariant theory, which posits that the laws of physics could evolve with time, is validated, it might necessitate a reassessment of the FLRW model's compatibility with Dirac's LNH (Canuto et al., 1977). The scale-covariant theory introduces a time-dependent gravitational constant ($G(t)$) that scales with the temperature ($T(t)$) of the universe, allowing for the evolution of physical laws (Canuto et al., 1977):

$$G(t) = G_0(T(t)/T_0)^n \tag{2}$$

where G_0 is the present value of the gravitational constant, T_0 is the present value of the temperature of the universe, and n is an exponent that determines how $G(t)$ scales with $T(t)$. By introducing a higher degree of flexibility into the FLRW model, the scale-covariant theory could potentially reconcile it with Dirac's hypothesis.

While the FLRW model has historically enjoyed wide acceptance within the scientific community, recent years have seen an upswing in interest towards the Dirac-Milne Universe model (Benoit-Lévy and Chardin, 2012). The Dirac-Milne Universe model provides an alternative theoretical framework for describing the evolution of the universe. In this model, the scale factor evolves as a power law with time:

$$a(t) = a_0(t/t_0)^{2/3} \tag{3}$$

where a_0 is the present-day scale factor and t_0 is the present age of the universe (Benoit-Lévy and Chardin, 2012).

This model has attracted attention due to its congruence with current observational data, as well as its potential insights into dark energy and dark matter, two of the most elusive aspects of our universe. The Dirac-Milne Universe model offers a theoretical framework that allows for the exploration of these enigmatic phenomena, regardless of whether Dirac's LNH eventually secures wider acceptance. Consequently, the Dirac-Milne Universe model could emerge as a cornerstone in advancing our understanding of dark energy, dark matter, and their respective roles in the orchestration of the cosmic order.

Equations (1), (2), and (3) along with the FLRW and Dirac-Milne Universe models discussed above provide mathematical representations and theoretical frameworks that are relevant to the analysis of universe models in the context of Dirac's LNH.

5. Dance of Constants: Variability in the Gravitational Constant

The exhilarating interplay between the macrocosm and the microcosm first sparked interest with Weyl (1917), who proposed a compelling correlation between the estimated radius of the universe (R_U) and the proposed radius of a particle (where rest energy equals the gravitational energy of an electron) r_H. Their ratio, juxtaposed with the classical radius of an electron r_e, lay in the staggering realm of 10^{42}, i.e.

$$\frac{R_U}{r_e} \approx \frac{r_H}{r_e} \approx 10^{42}. \tag{4}$$

Eddington (1931) built on this notion, discerning a similar proportionality between quantum and astronomical realms by comparing the force from electromagnetic interaction with the gravitational interaction of charged particles. This comparison gave rise to a value approximating the square root of N (the total number of charged particles in the universe), i.e.

$$\frac{e^2}{4\pi\epsilon_0 G m_e^2} \approx 4 \times 10^{42} \approx \sqrt{N}. \qquad (5)$$

where e is the elementary charge of an electron, ϵ_0 is the permittivity of free space or vacuum, and m_e is the mass of an electron, a fundamental particle.

Dirac, extending this discourse, examined the ratio between the electrical and gravitational force exerted between a proton and an electron, settling on a strikingly similar magnitude, i.e.

$$\frac{e^2}{4\pi\epsilon_0 G m_e} \times \approx 10^{40} \qquad (6)$$

Furthermore, he proposed a comparable ratio between the age of the universe and the atomic unit of time,

$$\frac{4\pi\epsilon_0 m_e c^3}{e^2} \approx 10^{40} \qquad (7)$$

Here c is the speed of light in a vacuum. From this, Dirac audaciously proposed that this dimensionless constant, pervasive across both macro and micro scales, should not remain invariant across the temporal span of the universe's age. He proposed that the Gravitational constant, G, should evolve as the inverse of time, rendering G a time-dependent entity, i.e.

$$G \propto \frac{1}{t} \qquad (8)$$

This bold proposition triggered a wave of intellectual pursuit in the realm of varying G, establishing the cornerstone of what is now the heart of the LNH.

This innovative approach called into question the fundamental assumptions of Einstein's General Relativity, which posits a constant G and employs relative time for the curving of four-dimensional space-time, disregarding the age of the universe (Einstein, 1915). Dirac's provocative LNH ignited a global scientific interest, suggesting a potential paradigm shift in our understanding of the universe, where the fundamental law of mass/energy conservation might be subject to revision. Yet, despite the profound implications, tangible evidence supporting G's variability remains elusive due to the enormous temporal scale of the universe.

Theories such as the Scale-covariant theory (Canuto et al., 1977) and the Hoyle-Narlikar theory (Hoyle and Narlikar, 1964) have introduced the idea of a universal gauge function wherein G can be expressed as a time-dependent entity. Despite this, the Hoyle-Narlikar theory, largely grounded in a steady-state model, has been largely dismissed in light of more recent cosmic microwave background radiation observations (Wright, 2010).

Numerous studies by Gaztañaga et al. (2002), Nordtvedt (1995), Sahoo et al. (2018), Singh (2007), and Berman (2009) have explored the tantalizing possibility of variations in G. While the changes might be infinitesimally small, these studies suggest they are plausible without undermining the foundational architecture of Einstein's gauge functions.

According to Dirac's LNH, a profound paradigm shift is required: atomic constants, such as Einstein's G, must be inversely tied to the age of the universe or cosmic time. This time-bound variability in G, and therefore the changing distribution of mass in the universe's history, implies the existence of an elusive mass component yet to be discovered. This mysterious mass could manifest as the enigmatic dark matter or as a perpetually created mass.

Dark matter, an exciting frontier in scientific exploration, remains elusive despite a myriad of studies devoted to deciphering its cryptic properties. No research has so far established any temporal correlation with dark matter, making it a challenging endeavor to attribute variations in Einstein's G to the influences of dark matter.

Conversely, Dirac turned his attention towards the intriguing concept of continuous mass creation, an idea that has inspired a multitude of inquiries probing its potential implications. The ensuing section will delve deeper into the captivating narrative woven by this proposition.

6. Symphony of Existence: Continuous Mass Creation

Dirac's LNH unveils an opulent narrative of cosmic proportions, introducing an audacious concept of continuous mass creation. This proposition challenges the traditional portrayal of a universe with a finite quantity of matter, crafting instead a cosmic panorama perpetually infused with emergent matter. This daring concept instigates a revolutionary pivot in our understanding of cosmic evolution, provoking reflections on the origin, nature, and fate of matter in the universe (Davies, 1982).

We previously referred to the potential implications of LNH: if atomic constants such as G were to demonstrate temporal fluctuations, a transformative mechanism must be at work. A compelling candidate for this is the concept of perpetual mass creation. This framework reimagines the universe not as a static construct but as a dynamic theatre where matter is unceasingly born either uniformly throughout space ("additive creation") or localized in areas already rich in matter ("multiplicative creation"), with the former involving the appearance of new matter at a constant rate, and the latter involving a rate that is proportionate to the existing density of matter. This relentless act of creation is postulated to be proportionate to the existing quantity and types of atoms (Hoyle, 1960).

This concept of continuous mass creation has started to reverberate within the scientific community, with a burgeoning body of research scrutinizing its implications and possible harmonization with LNH (Canuto, 1979). Exploration of this phenomenon is crucial not only for delving deeper into the mutable nature of the universe and its fundamental constants, but also for understanding the potential role of dark matter and dark energy in this cosmic ballet (Peebles, 2001).

As an area of scientific inquiry, continuous mass creation has opened intriguing avenues into its potential viability and congruence with observational data. Numerous investigations have attempted to decode the engines propelling mass creation, such as the potential birth of matter from the vacuum state or the transformation of dark energy into matter. For example, Zel'dovich (1971) proposed that the vacuum fluctuations could provide the birthplace for particles, governed by:

$$m = \hbar\omega \tag{9}$$

where $\hbar$ is the reduced Planck constant and ω is the angular frequency of the fluctuation. This equation underlines the potential transformation of vacuum energy into matter, a key idea in continuous mass creation.

On the other hand, Davies (1974) worked on the hypothesis that the creation rate of particles $\dot{N}$ is directly related to the scale factor ($a(t)$) and its derivatives, governed by:

$$\dot{N} = \alpha\frac{\ddot{a}}{a} + \beta\left(\frac{\dot{a}}{a}\right)^2 \tag{10}$$

Here, α and β are constants, and the dot represents time derivatives. Equation (10) is interesting because it suggests that the rate of mass creation is intrinsically linked to the dynamics of the expanding universe, fitting well with Dirac's LNH.

The concept of "additive creation" and "multiplicative creation" could be symbolically represented as:

- Additive creation: $\dot{\rho} = \Gamma$ (11)
- Multiplicative creation: $\dot{\rho} = \Gamma_\rho$ (12)

Here, ρ represents the energy density, and Γ signifies the creation rate. In the case of additive creation, new matter appears at a constant rate Γ, while for multiplicative creation, the rate Γ is proportionate to the existing density ρ.

These inquiries add mathematical rigor to the discussion, laying out possible pathways through which continuous mass creation could occur, either from the vacuum state or from the transformation of existing forms of energy such as dark energy into matter. They also illuminate how such processes could engender mass at a rate consistent with the observed variations in G and other cosmological parameters, as indicated by Dirac's LNH (Davies, 1974).

Continuous mass creation, although a contested and fervently debated topic, presents an enticing and promising frontier in scientific exploration. It tantalizes us with the potential to redefine our understanding of the universe's tapestry, its history, and its future trajectory (Hawking, 2001). As our empirical repositories expand and theoretical models advance in complexity, it is conceivable that the concept of continuous mass creation will gain traction, fortifying its central position in our cosmic comprehension. As we proceed, this hypothesis may disclose a universe that is not merely a stage for the dance of existence, but an active player in the choreography (Rees, 1997).

7. Cosmic Serenade: The Anthropic Principle

The enigma of large numbers permeates the fabric of our universe, embedding a resonance that echoes through the cosmos. Dicke (1961) proposed a theoretical framework, suggesting that some constants could be deduced from established theories, while others, such as the Hubble constant (H), could be anticipated by associating a time scale that aligns with the age of the universe's stars. In fact, all numbers in Table 1 can be reasoned from fundamental laws of physics if we assume two starting points:

I. The age of stars (deductible from atomic and gravitational constants) is comparable to the current age of the universe.

II. The first row of Table 1, which posits that the gravitational interaction is 10^{40} times weaker than electromagnetic interactions, and the fact that the fine structure constant ≈1/137 is not far from the order of unity.

Instead of looking for an inherent reason for this time scale, one can, after the fact, apply the so-called Anthropic Principle. The Anthropic principle asserts that the astounding fine-tuning of physical constants exists only to facilitate an ordered universe and our consequential existence.

This assertion was built on the recognition that certain fundamental physical constants seem to be astoundingly calibrated, thereby facilitating an ordered universe and our consequential existence. The precision of these constants, though measurable, continues to baffle us, underscoring the elusive nature of the cosmic orchestration.

Building on this foundation, Carter (1974) distinguished the Anthropic Principle into two categories: the Weak Anthropic Principle (WAP) and the Strong Anthropic Principle (SAP). The WAP is represented by the conditional probability expression,

$$P\big((O|L)\big) \approx 1, \tag{13}$$

where P denotes the conditional probability. Specifically, equation (13) denotes the conditional probability of O given L, where O signifies the existence of observers and L indicates life-permitting conditions. This implies that, given the life-permitting conditions of the universe (L), the probability of the existence of observers (O) is almost certain (≈ 1). This underpins the idea that our cosmic location, including the epoch we inhabit, is privileged to coincide with our existence as conscious observers. Put differently, it is not surprising to find ourselves in a part of the universe hospitable to life because, otherwise, we wouldn't exist to observe the universe. This underscores that our existence hinges on the specific conditions of our universe that permit life. The WAP aims to elucidate why the current age of the universe aligns with the age of stars—referred to as the starting point I above. It necessitates a reassessment of our cosmic self-perception and counters the Copernican principle by asserting that we, as observers, do not occupy an unprivileged, random location in the universe.

On the other hand, SAP, expressed as

$$P\big((L|O)\big) = 1, \tag{14}$$

posits that the universe and its foundational parameters are configured to permit the emergence of observers. This rekindles Descartes' philosophical assertion, *cogito ergo mundus talis est*—we think, therefore the world is such.

The SAP instead suggests that the atomic and gravitational constants are fine-tuned in order for us to exist, like a cosmic instrument orchestrated to produce a resonant harmony, in which we are a part of. In other words, they are designed to explain point II above. For instance, consider the triple-alpha process in stellar nucleosynthesis, an astoundingly unlikely event that permits the formation of carbon, a building block for life, from primordial helium. The finely-tuned parameters of this process are expressed as

$$\sigma = \frac{S(E)}{E} exp\left(-\frac{2\pi\eta}{\hbar}\right) \tag{15}$$

where σ is the cross-section for the reaction, *S(E)* is the astrophysical *S*-factor, *E* is the energy, η is the Sommerfeld parameter, and $\hbar$ is the reduced Planck constant. The value of *S(E)* at the specific energy E accounts for the effects of the Coulomb barrier and the quantum mechanical tunneling that particles must undergo to interact at low energies in stellar environments. It effectively measures how the rate of a nuclear reaction varies with the energy of the reacting particles — a process that is crucially dependent on the structure of the carbon nucleus (Barrow, 2002).

The crucial aspect of equation (15) is $S(E)$, which is dependent on the structure of the carbon nucleus, and the value of $S(E)$ required for the triple-alpha process is remarkably fine-tuned, meaning it must fall within a very narrow range to allow the formation of carbon. If this were not the case, carbon, a fundamental building block for life, would not form, and life as we know it would not exist. This is an example of fine-tuning in the universe, which the SAP suggests is necessary for the existence of observers like us.

The Anthropic Principle speculates that even minute changes in G could result in discordant notes, disrupting the symphony of conditions necessary for life. As Weinberg (1987) demonstrated, the cosmological constant (Λ), which determines the large-scale structure of the universe in Einstein's equations of general relativity, is astonishingly fine-tuned for life. Any significant deviation in its value could result in a universe hostile to the emergence of complex structures like galaxies, and hence life as we know it. Weinberg derived an anthropic upper bound on the cosmological constant as $\Lambda \approx 10^{-120}$ Planck units, a prediction later confirmed by cosmological observations.

Since Carter's initial formulation, the Anthropic Principle has undergone significant refinements. Bostrom (2022) delved deeper into the concept of selection effects, suggesting that our observations of the universe are not randomly sampled but are influenced by our existence as observers. In the framework of the "many-worlds" interpretation of quantum mechanics, the Anthropic Principle proposes that our observations are determined by the specific "branch" or trajectory of the universe in which we reside.

Taking these ideas further, Hawking (1988) speculated about the existence of an infinite number of parallel universes within a multiverse framework. In this perspective, each universe may harbor different physical laws and fundamental constants, with intelligent observers arising only in those universes that fortuitously possess life-permitting conditions. This expansive concept broadens the scope of the Anthropic Principle beyond the realm of humanity, encompassing any potential observer, regardless of their form or species. By doing so, the principle acknowledges the possibility of non-human intelligent life both within our universe and in others.

The enigmatic melody of the Anthropic Principle beckons researchers to delve deeper into its philosophical and scientific nuances, guiding our quest for a deeper understanding of our place within the grand symphony of the cosmos. Future research directions include investigating the fine-tuning of fundamental constants, exploring the multiverse hypothesis, understanding anthropic selection effects, examining the connection between the Anthropic Principle and the foundations of quantum mechanics, studying the influence of anthropic constraints on cosmological evolution, and delving into the philosophical implications of this principle. This ongoing exploration fuels our curiosity and propels us to uncover the harmonies and intricacies of the celestial concert that is our universe.

8. Conclusion and Discussion

Dirac's Large Number Hypothesis (LNH) has taken us on a remarkable journey through various physical theories and their interrelationships, expanding our understanding of the universe and our place within it. This hypothesis has sparked conceptualizations of universe models, variations in gravitational constants, and continuous mass creation, while also stimulating discussions on the philosophical implications of the Anthropic Principle and our comprehension of cosmological constants.

The LNH, despite its abstract nature, holds deep implications for our understanding of the physical universe. It challenges our conventional wisdom regarding the nature of physical constants and the structure of the universe, pushing us to explore new frontiers in theoretical physics. While it may conflict with established models like the Friedmann-Lemaitre-Robertson-Walker metric, it harmonizes beautifully with more recent compositions such as the scale covariant theory of gravity, offering fresh perspectives on enigmatic phenomena like dark energy and dark matter.

The LNH's suggestion of the variability of the gravitational constant over cosmic time has sparked a vibrant area of research. If such variability is indeed possible, it could reshape our understanding of Einstein's theory of general relativity and prompt us to reconsider the concept of "constants."

Continuous mass creation, another movement in the LNH symphony, holds the potential to illuminate unresolved mysteries of the universe. The ongoing creation of matter, whether through additive or multiplicative processes, offers new avenues to comprehend the nature and distribution of matter in our cosmic concert hall.

Meanwhile, the Anthropic Principle presents us with philosophical enigmas about our existence and the character of the universe. It challenges us to rethink the role of "observers" and ignites the tantalizing prospect of other forms of intelligent life performing in parallel concert halls of the cosmos.

However, the journey of understanding the large number hypothesis is far from over. Future research directions include further exploration of the variations in fundamental constants, deeper investigations into the multiverse hypothesis, understanding the underlying mechanisms of continuous mass creation, examining the anthropic selection effects on cosmological evolution, and delving into the philosophical implications of the Anthropic Principle. These research endeavors will contribute to solving the intriguing question of the large number hypothesis and enhance our understanding of the symphony of the universe.

In the grand concert of scientific inquiry, it is crucial to acknowledge that our journey is ongoing, and no single theory or principle can fully capture the complexity of the symphony of our universe. Dirac's Large Number Hypothesis, the Anthropic Principle, and the associated discussions provide powerful instruments that guide us toward a deeper understanding of our perpetually evolving universe. As our cosmic performance continues, we anticipate further revelations and insights that will enrich our understanding of the universe and our place within it.

Acknowledgement: This work was supported by the Burke Institute for Theoretical Physics, California Institute of Technology (Caltech). The authors also acknowledge the support by the Caltech Jet Propulsion Laboratory, sponsored by NASA.

References

1. Abbott, B. P. et al. (LIGO Scientific Collaboration and Virgo Collaboration) (2016). "Observation of Gravitational Waves from a Binary Black Hole Merger". Physical Review Letters, 116(6), 061102. doi:10.1103/PhysRevLett.116.061102
2. Barrow, J. D. (2002). The Constants of Nature: From Alpha to Omega—The Numbers That Encode the Deepest Secrets of the Universe. Pantheon Books.
3. Benoit-Lévy, A., & Chardin, G. (2012). Introducing the Dirac-Milne Universe. Astronomy & Astrophysics, 537, A78. doi:10.1051/0004-6361/201117492
4. Berman, M. S. (2009). Realization of Einstein's Machian program by the variable mass theory of gravitation. Nuovo Cimento B, 74(2), 173-179. doi:10.1393/ncb/i2008-10420-2
5. Bostrom, N. (2002). Anthropic Bias: Observation Selection Effects in Science and Philosophy. Routledge.

6. Bostrom, N. (2022). Anthropic Bias: Observation Selection Effects in Science and Philosophy. Routledge, London.
7. Canuto, V., Adams, P. J., Hsieh, S.-H., & Tsiang, E. (1977). Scale-covariant theory of gravitation and astrophysical applications. Physical Review D, 16(6), 1643-1663. doi:10.1103/PhysRevD.16.1643
8. Canuto, V., Hsieh, S.H., & Adams, P.J. (1977). Scale-Covariant Theory of Gravitation and Astrophysical Applications. Physical Review Letters, 39(10), 429. doi:10.1103/PhysRevLett.39.429
9. Canuto, V. M., Adams, P. J., Hsieh, S.-H., & Tsiang, E. (1979). Scale-Covariant Theory of Gravitation and Astrophysical Applications. Physical Review D, 19(6), 1547–1562. doi:10.1103/PhysRevD.19.1547
10. Carr, B. J. (2005). Universe or Multiverse?. Cambridge University Press.
11. Carroll, S. (2010). From Eternity to Here: The Quest for the Ultimate Theory of Time. Dutton.
12. Carter, B. (1974). Large Number Coincidences and the Anthropic Principle in Cosmology. Proceedings of the IAU Symposium, 63, 291-298. doi:10.1017/S0074180900235638
13. Carter, B., McCarea, W. (1983). The Anthropic Principle and its Implications for Biological Evolution. Phil. Trans. R. Soc. Lon. A, 310, 347-363. doi:10.1098/rsta.1983.0096
14. Collins, C. B., Hawking, S. W. (1973). "Why is the universe isotropic?". Astrophysical Journal, 180, 317-334. doi:10.1086/151965
15. Coulomb, C. A. (1785). Premier Mémoire sur l'Électricité et le Magnétisme. Histoire de l'Académie Royale des Sciences, 569-577.
16. Davies, P. C. W. (1982). The Accidental Universe. Cambridge University Press.
17. Davies, P. C. W. (1974). The Physics of Time Asymmetry. University of California Press.
18. Dicke, R. H. (1961). Dirac's cosmology and the large numbers hypothesis. Nature, 192(4808), 440-441. doi:10.1038/192440a0
19. Dirac, P. A. M. (1937). The Cosmological Constants. Nature, 139, 323. doi:10.1038/139323a0
20. Dirac, P. A. M. (1947). The Cosmological Constants. Nature, 159, 198. doi:10.1038/159198a0
21. Dirac, P. A. M. (1974). Cosmological Models and the Large Numbers Hypothesis. Proceedings of the Royal Society of London A, 338, 439–446. doi:10.1098/rspa.1974.0095
22. Dirac, P. A. M. (1974). Cosmologies with Varying Light-Speed. Proc. R. Soc. Lond. A, 338, 439–446. doi:10.1098/rspa.1974.0095
23. Eddington, A. S. (1920). A determination of the deflection of light by the Sun's gravitational field, from observations made at the total eclipse of May 29, 1919. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 220, 291–333. doi:10.1098/rsta.1921.0009
24. Eddington, A. (1931). Preliminary Note on the Masses of the Electron, the Proton, and the Universe. Proceedings of the Cambridge Philosophical Society, 27(1), 15–19. doi:10.1017/S0305004100009269
25. Eddington, A. (1931). On the value of the cosmical constant. Proceedings of the Royal Society of London. Series A, 133, 60-72. doi:10.1098/rspa.1931.0111
26. Eddington, A. (1919). Report on the Relativity Theory of Gravitation. Fleetway Press, London.
27. Einstein, A. (1915). The Field Equations of Gravitation. Sitzungsberichte der Preussischen Akademie der Wissenschaften, 844-847.
28. Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. Annalen der Physik, 354(7), 769-822. doi:10.1002/andp.19163540702.
29. Friedmann, A. (1922). Über die Krümmung des Raumes. Zeitschrift für Physik, 10(1), 377-386.

30. Gaztañaga, E., Sánchez, N., & Sarkar, S. (2002). A general test of the Copernican Principle. Monthly Notices of the Royal Astronomical Society, 334(1), L13-L16. doi:10.1046/j.1365-8711.2002.05384.x
31. Greene, B. (1999). The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory. W. W. Norton & Company.
32. Hawking, S. W. (1988). A Brief History of Time. New York: Bantam Books.
33. Hawking, S. (2001). The Universe in a Nutshell. Bantam Books.
34. Hoyle, F. (1960). The Universe: Some Recent Developments in Astronomy. Cambridge University Press.
35. Hoyle, F. (1954). On Nuclear Reactions Occurring in Very Hot Stars. Astrophysical Journal Supplement Series, 1, 121.
36. Hoyle, F., & Narlikar, J. V. (1964). Action at a Distance in Physics and Cosmology. San Francisco: Freeman.
37. Hubble, E. P. (1929). A relation between distance and radial velocity among extra-galactic nebulae. Proceedings of the National Academy of Sciences, 15(3), 168-173. doi:10.1073/pnas.15.3.168.
38. Lemaître, G. (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extragalactiques. Annales de la Société Scientifique de Bruxelles, A47, 49-56.
39. Nordtvedt, K. (1995). Testing Relativity with Laser Ranging to the Moon. Physical Review, 170(5), 1186-1187. doi:10.1103/PhysRev.170.1186
40. Peebles, P. J. E. (2001). Principles of Physical Cosmology. Princeton University Press.
41. Ray, S. (2019). Large Number Hypothesis: A Review. General Relativity and Gravitation, 51(1), 1-26. doi:10.1007/s10714-018-2476-2.
42. Rees, M. (1997). Before the Beginning: Our Universe and Others. Addison-Wesley.
43. Robertson, H. P. (1933). Relativistic Cosmology. Reviews of Modern Physics, 5(1), 62-90.
44. Sahoo, B. K., Rath, P. K., & Biswal, S. (2018). Variations of Newton's Gravitational Constant and Einstein's Cosmological Constant. Modern Physics Letters A, 33(38), 1850223. doi:10.1142/S021773231850223X
45. Saunders, S., Barrett, J., Kent, A., & Wallace, D. (2010). Many Worlds?: Everett, Quantum Theory, and Reality. Oxford University Press.
46. Singh, B. P. (2007). Is the Gravitational Constant, G, really constant? arXiv:0706.1238v1 [gr-qc].
47. Uzan, J-P. (2003). The Fundamental Constants and Their Variation: Observational Status and Theoretical Motivations. Reviews of Modern Physics, 75(2), 403–455. doi:10.1103/RevModPhys.75.403.
48. Walker, A. G. (1937). On Milne's Theory of World-Structure. Proceedings of the London Mathematical Society, 42(1), 90-127.
49. Weinberg, S. (1987). Anthropic Bound on the Cosmological Constant. Physical Review Letters, 59, 2607-2610. doi:10.1103/PhysRevLett.59.2607
50. Weyl, H. (1917). Zur Gravitationstheorie. Annalen der Physik (in German), 359(18), 117–145. doi:10.1002/andp.19173591804
51. Weyl, H. (1917). On the Quantum Mechanics of the Electromagnetic Field. Physikalische Zeitschrift, 18, 722-732.
52. Weyl, H. (1918). Gravitation and Electricity. Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 465-480.
53. Weyl, H. (1919). Eine Neue Erweiterung der Relativitätstheorie. Annalen der Physik, 362(23), 1018-1024.
54. Wright, E. L. (2010). Errors in the Steady State and Quasi-SS Models. arXiv:1001.2011 [astro-ph].

55. Zel'dovich, Y. B. (1971). Cosmological Constant and Elementary Particles. Journal of Experimental and Theoretical Physics Letters, 6, 345–347.

借鉴 OC43 引发 1889 年人类冠状病毒病（COVID）大流行史：探讨 COVID-19 的中枢神经损伤机制

周炳亮 [1]，曹虹 [1]，黄胜和 [2]

[1] 南方医科大学公共卫生学院广东省热带病研究重点实验室微生物学系 [2] 美国南加州大学

COVID-19 的大流行已经结束，但它对人类生活的各方面产生了不可逆转的巨大影响。COVID-19 冠状病毒（SARS-CoV-2）除了导致机体全身和呼吸道症状外，人类 CoV（HCoV）肺炎患者还会出现中枢神经系统症状，包括头痛、意识障碍和感觉异常[1]。随着 SARS-CoV-2 的持续进化，尽管它对神经系统的感染和损伤的发病机制是未知的，但病毒的发展趋势是衍化为普通感冒病原微生物。至今为止，已知的人类冠状病毒（HCoV）有 7 种，包括 HCoV-NL63、HCoV-229E、HCoV-HKU1、HCoV-OC43、严重急性呼吸综合征冠状病毒（SARS-CoV）、中东呼吸综合征冠状病毒（MERS-CoV）和 SARS-CoV-2。研究表明 HCoV-OC43 可能在一百多年前出现并形成大范围流行[2]，当时误认为流感（史称"俄罗斯流感）直到 2005 年才被感染组学（Infectomics）研究证明其病毒为 HCoV-OC43[3]。HCoV-OC43 在人类生态系统中进行了相当长时间的演变和进化，它最终衍化成为历史上引发普通感冒的 4 种冠状病毒之一。HCoV-OC43 与 SARS-CoV-2 同种属，除去部分的结构差异，二者有很高的相似性，并表现了相似的流行趋势[4]。本文资深作者曾在本协会 2020 年 8 月 1 日的学术研讨会上曾预言 COV-19 病毒可衍化为第 5 个引发人类普通感冒的 HCoV。这次 COV-19 流行趋势证明这个预测是准确的。因此，借鉴 OC43 引发 1889 年人类冠状病毒病（COVID）大流行史，是探讨 SARS-CoV-2 乃至冠状病毒对人类神经系统损伤发展和进化的新思路。

HCoV-OC43 的流行病史

将时间线回溯到 19 世纪的一次大流行，第一次被详细描述的 COVID 大流行，是众所众知的 1889 年 12 月-1893 年 6 月的“俄罗斯流感”。一篇经典文章记

录了美国印第安纳州发生的三年半（3.5）疫情。首波疫情发生在 1889 年 12 月至 1890 年 7 月，接着陆续发生了第二波（1890 年 12 月－1891 年 7 月）、第三波（1891 年 10 月-1892 年 7 月）及第四波（1892 年 12 月-1893 年 5 月）疫情[5]。这场大流行的历史经验告诉我们，一种流行病可以在仅有 15 亿人口、联系比现代少得多（铁路和轮船是最快的交通工具）的世界中迅速传播。此次大流行被普遍误认为由 H3N8 病毒引起[6]。然而根据存档的医疗记录和进化模型，“俄罗斯流感”大流行的根源和病原体可能是冠状病毒[7]。利用分子时钟（Molecular Clock）的技术，Leen Vijgen 等人构建了 HCoV-OC43 株和几个 BCOV 分离株的 S 基因系统发育树，使用不同的分子时钟运算，HCoV-OC43 的最近共同祖先和不同的 BCoV 分离株定位于 1890 年左右，并证明病毒是从牛向人类传播[3]。另一个证据是，“俄罗斯流感”对中枢神经系统的症状比其他流感爆发期间更为明显，而 HCoV-OC43 具有神经侵袭性是已被证明的[8]。因此，在 1890 年左右 BCOV 可能跨越了物种屏障，与“俄罗斯流感”大流行对人体交叉感染，导致一种新型人类冠状病毒（HCoV-OC43）的出现[9, 10]。这意味着 HCoV-OC43 在人类生态系统中可能已经进行了漫长的演变和进化。

Baginsky, A 等人编写的专著[4]对“俄罗斯流感”的流行发展有详细的记载，整体而言，不同地区几乎都发生了三次流行浪潮（图 1）。在伦敦，从 1890 年 1 月中旬开始的 4 周内，感染死亡人数在两周内达到峰值，1891 年 5 月至 6 月的 8 周期间出现了第二个高峰，死亡人数甚至更高，之后 1892 年 1 月至 2 月的 5 周出现了第三个死亡率高峰，英国的其他疫区也存在时间趋势相似的三次浪潮[11]；与此同时，距离英国相对遥远的美国也观察到了三个明确的死亡浪潮，如前所述的美国印第安纳州[5]；另一方面，例如哥本哈根等地区发生的浪潮不同于英国和美国，第一次浪潮从 1889 年 12 月持续到 1890 年 7 月，接着 1890 年 10 月发生了一次小规模的疫情爆发，1893 年 5 月观察到了最后一次疫情浪潮。

相比之下，SARS-CoV-2 流行的发展趋势更加清晰。2019 年末，病毒在中国武汉首次爆发，尽管中国迅速采取了防控措施，但到 2020 年 2 月下旬，SARS-CoV-2 已经蔓延到中国的多个地区和世界各地[12]。在接下来的 3 年时间，SARS-CoV-2 遍布全世界。从病毒的进化发展看，SARS-CoV-2 在 3 年间也经历了 3 个典型的阶段（图 1）：在最初的近 8 个月时间里，病毒似乎表现出的进化有限，

全球病毒数量相对较少，传播是有限的，SARS-CoV-2 以相对缓和的速度向全世界蔓延，造成多地疫情局部暴发[13]；直到 2020 年 7 月左右，第一个被发现的关键突变 D614G 占领全球，拉开了变异的序幕[14]。在接下来一年的时间里，确定了四个指定的关注变种，即 α 变异体（最初在英国发现）、β 变异体（最初在南非发现）、γ 变异体（最初在巴西发现）和 δ 变异体（最初在印度发现）[15]，这标志着大流行进入了一个新阶段。这些变体可增加病毒传播速率和/或增加再感染风险，并降低中和单克隆抗体和疫苗接种提供的保护[16]。至此全球疫情持续扩散，全面爆发；2021 年 11 月向世界卫生组织报告的 Omicron（B.1.1.529）变异株[17]，它的毒力减弱，但也是当时突变最多、传染性最强的变异株，并且对流行的免疫疗法或疫苗具有相对抗药性[18, 19]，这种变异最近构成了高度的全球公共卫生威胁，标志着大流行新阶段的开始，直到 2022 年 Omicron 亚系的连续传播在全球 SARS-CoV-2 感染上占据主导地位。

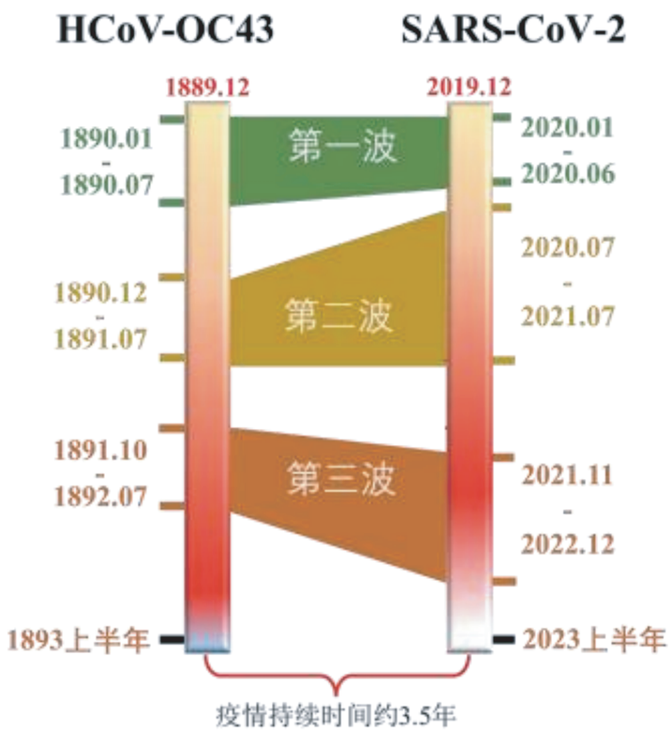
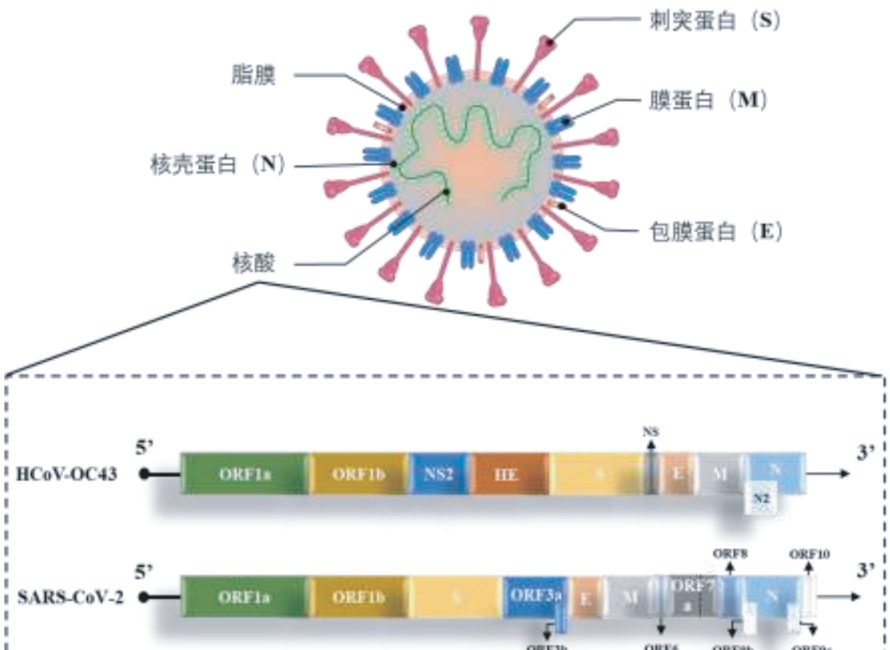


图 1 HCoV-OC43 和 SARS-CoV-2 的疫情时间轨迹对比

HCoV-OC43 于 1967 年首次被发现并鉴定为 β 冠状病毒，它是在人群中较为常见的一种，通常会导致普通感冒[20]。在结构上（图 2），冠状病毒的单链 RNA 基因组被包裹在螺旋核衣壳中[3]，基因组的初始部分编码 16 种非结构蛋白，而

剩余部分编码 4 种必需结构蛋白，即小包膜蛋白（E）、刺突糖蛋白（S）、基质（M）蛋白和核衣壳（N）蛋白[21]。进行基因对比显示 SARS-CoV 全基因组序列与 HCoV-OC43 的同源性为 53.1%，提示其与 HCoV-OC43 有显著的亲缘关系[22]。HCoV-OC43 感染诱导的免疫甚至被证实可以对 SARS-CoV-2 提供短暂的保护[23]。较高的同源性使 HCoV-OC43 能够作为研究 SARS-CoV-2 进化和发展趋势的良好替代品[2]。与其他病毒相比，冠状病毒的起源相对简单，但其宿主主要分布在哺乳动物中[24]。根据 LI C 等人的研究，蝙蝠是冠状病毒的自然宿主，大量家禽和某些野生动物共存在当地乡村[25]。这个结果比其他现象更应该引起我们的关注，因为人类也哺乳动物，而且是高级哺乳动物，人类是否有一天会因为与哺乳动物的过度接触而感染一种新型病毒是未知的[26]。这涉及到了另一个科学问题，人类在与大自然共存的过程中，是否需要更加关注人类与野生动物等的距离，防止未知的病原体由此传播至人类群体，即测定并限制一个平衡的生态健康距离（Eco-health distance）[9]。



图二 HCoV-OC43 和 SARS-CoV-2 的病毒结构和基因序列

HCoV-OC43 的神经侵袭性

肺是人类的重要器官，血供丰富，是气体交换所必需的，而大脑是人体内耗氧量最大的器官，因此，肺部疾病很容易影响大脑[27]。临床研究表明，慢性阻塞

性肺疾病（COPD）通过增加脑白质病变而与脑血管疾病相关，COPD 患者的生活质量和预后在很大程度上取决于是否存在包括卒中和认知障碍在内的共病[28]。人类呼吸道合胞病毒（hRSV）、流感病毒（IV）、冠状病毒（CoV）和人类偏肺病毒（hMPV）等呼吸道病毒一旦进入肺部，就可以扩散到全身并最终影响中枢神经系统[29]。1980 年，从两名多发性硬化症患者尸检获得的脑组织中分离出了 HCoV-OC43[30]；2004 年，在一名儿童患者的鼻咽和脑脊液样本中检测到 HCoV-OC43，这名儿童表现出急性播散性脑脊髓炎，这是一种低发病率的中枢神经系统疾病导致的脱髓鞘[31]。20 多年前，HCoV-OC43 就被证实能够导致星形细胞瘤细胞系、神经胶质瘤和少突胶质细胞的急性和持续感染，并且伴随着感染性病毒粒子的产生和病毒 S1 基因组 RNA 的分子变异[32]。后续研究建立小鼠吸入感染 HCoV-OC43 模型，结果表明它具有感染神经细胞和从外周到达中枢神经系统的内在能力[33]。从早期动物模型和人类研究中积累的证据表明，该病毒具有嗜神经潜能。

研究表明，人体吸入 HCoV-OC43 后，可能通过鼻腔-嗅球、梨状皮质-脑干的途径传播至神经系统，其中神经突触在传播上起重要作用[34]。随后 HCoV-OC43 在大脑中复制导致的病理通常包括免疫介导的成分[35]，感染人星形细胞系 U-373MG 后可导致 IL-6、TNF-α 和 MCP-1 mRNA 表达上调，提示其可能导致中枢神经系统病变相关的炎症分子的产生[36]。HCoV-OC43 持续感染人类神经细胞导致病毒 S 蛋白发生突变[37]。而通过感染具有 S 糖蛋白两点突变的 HCoV-OC43，观察到鼠类和人类神经元细胞的死亡明显增加[38]，甚至某些突变导致受感染小鼠出现后肢麻痹性疾病。数据表明病毒 S 蛋白的突变（在人类神经细胞培养物中病毒持续存在期间可重复获得）导致病毒诱导的神经病理学发生彻底改变，其特征是弛缓性麻痹和脱髓鞘[37]。然而即使有研究为 HCoV-OC43 在中枢神经系统的时空传播提供了新的见解，但当前对 HCoV-OC43 在体内的动态感染过程和机制仍有待深入研究[39]。

临床资料显示，高达 73%的 COVID-19 住院患者累及神经系统症状，主要是头痛、肌痛和意识障碍等类似颅内感染的症状，大多表现为非特异性脑病[40, 41]。事实上，有些人甚至在出现神经系统症状后才出现 COVID-19 相关症状[42]。北京地坛医院在疫情初期报道一例 SARS-CoV-2 攻击中枢神经系统引起的病毒性脑

炎病例[43]。当前 SARS-CoV-2 对神经系统损伤的研究普遍认为，病毒通过将 S 蛋白结合到哺乳动物宿主细胞上的血管紧张素转换酶（ACE）-2 受体上而进入细胞，并通过丝氨酸蛋白酶跨膜蛋白酶丝氨酸 2（TMPRSS2）进行刺突蛋白启动[44]。ACE-2 和 TMPRSS2 均存在于小胶质细胞中，为 CNS 渗透提供途径。SARS-CoV-2 的神经侵袭可通过多种潜在机制发生（图 3-A、B），包括跨突触转移、嗅觉神经进入、血管内皮感染和白细胞迁移穿过血脑屏障（BBB）[45]。可见，HCoV-OC43 和 SARS-CoV-2 在神经侵袭途径上有相似性，神经损伤在疾病的进展中可能导致严重的后果。

肺-脑轴在相关神经损伤上的作用

2020 年疫情早期报告了一类患有 COVID-19 的 3 个月以下婴儿，表现为非局灶性神经现象（轴性低眼压、嗜睡），并伴有发热，但所有患者的没有呼吸道症状，脑脊液均为正常，SARS-CoV-2 检测均为阴性[46]。显然，SARS-CoV-2 可能在不直接侵袭神经系统的情况下进而引发的神经系统损伤。例如病毒感染伴随的缺氧性、感染性、代谢性和血管损伤，可能会导致这些区域严重的神经元损伤和神经炎症[47]。另一方面，肺部与口咽部存在丰富的微生物群[48]，肺部炎症可由获得性微生物区系调节，通过各种机制影响局部神经和免疫微环境[49]。这种与严重的神经系统疾病、功能障碍、肺部感染和损伤相关的途径被称为肺-脑轴，它被认为是双向的[50]，它们的相互影响由神经元、炎症、免疫和神经内分泌信号通路组成的复杂网络所介导[51]。其已知或预测的潜在机制有五种；微生物的直接易位，肺部微生物对全身免疫、神经、下丘脑-垂体-肾上腺轴和代谢变化的影响[48]。

在非直接侵犯神经系统的途径中，肺部微生物对全身性体液因子的影响是通过促炎介质的刺激来实现的，某些抗菌肽可介导体液免疫，进而影响中枢神经系统[52]。COVID-19 相关脑病是一个典型的例子，其严重性与细胞因子风暴强度直接相关[53]。而 HCoV-OC43 的核衣壳蛋白能引起 NF-κB 活化的增强，这种相互作用是 RNA 病毒中免疫逃避的潜在机制[54]，炎症细胞因子在下呼吸道上皮或其他组织中干扰素诱导的跨膜可能为 HCoV-OC43 侵袭这些组织铺平道路[55]。并且干扰素-γ 被证明通过增加 MHC I 类分子的表达来增加 HCoV-OC43 对神经细胞的感染[56]。然而，HCoV-OC43 感染下呼吸道后对神经系统的间接损伤机制并没有

相关研究。

不同于 HCoV-OC43，SARS-CoV-2 在间接影响神经系统的机制上有不少的研究。SARS-CoV-2 感染可以通过炎症、凝血和血栓形成以及缺氧对血脑屏障产生间接影响（图 3-C、D）[57]。首先，SARS-CoV-2 感染引起全身炎症反应，导致促炎细胞因子、趋化因子、炎性蛋白、补体升高，细胞因子风暴还在全身产生连锁反应[58]，这些炎症因子可引起血脑屏障破坏，从而导致神经损伤[57]。并且临床结果的总结支持 SARS-CoV-2 感染增加了免疫激活、补体介导的血栓性微血管病的风险[59]。另外，临床数据表明，SARS-CoV-2 感染肺部导致呼吸困难、呼吸急促，进而可引发低氧血症[60]。大量的体外和体内研究表明，缺氧导致血脑屏障破坏，这可能是随后的中枢神经系统疾病的触发因素之一[61]。另一方面，SARS-CoV-2 感染使机体免疫力下降，损害细胞和肺基础结构，随后，改变的条件使细菌能够增加粘附和侵袭，一些感染 SARS-CoV-2 的患者表现出由免疫系统失调引起的与细菌共感染相关的生物标志物和炎性细胞因子水平升高[62]。然而，呼吸系统微生态的改变进一步加重神经系统损伤的可能机制有待更多而研究。

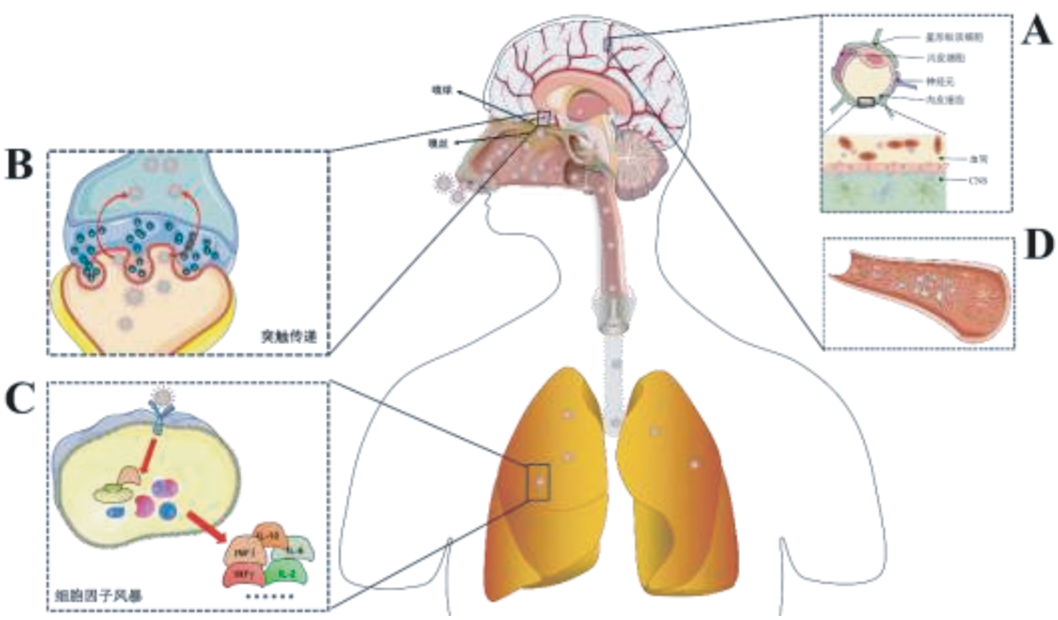


图 3 冠状病毒对神经系统的直接和间接损伤。

A：病毒感染机体后通过血液循环到达脑血管后，穿过血脑屏障直接侵袭神经系统；B：鼻腔吸入病毒，通过嗅球达到嗅神经，再通过突触直接感染神经细胞；C：病毒感染肺部细胞，导致机体发生细胞因子风暴，炎症因子间接影响其他组织器官；D：病毒感染后导致缺氧，血液粘稠度增加，炎症因子增加，导致血栓形成。

SARS-CoV-2 神经损伤的未来

冠状病毒的突变能力及其高重组和/或趋同进化频率是其从动物引入人类并长期共存的主要驱动力[63]。HCoV-OC43的进化潜力使点突变和重组事件在人类群体发挥持续作用[64]。即使HCoV-OC43被发现超过100年，但在儿童呼吸道感染中依旧有它的身影，临床表现可从发烧到呼吸窘迫[65]。2014年11月至2016年11月，对中国大陆13家医院住院的社区获得性肺炎（CAP）儿童进行的一项前瞻性研究中，检测出了一种新的HCoV-OC43重组基因型[66]。一项2016年的儿童热性惊厥病原体分析的研究指出，HCoV-OC43是在小于12个月的儿童中最常见的致病原[67]。种种证据表明，HCoV-OC43没有消散在历史长河中，它还在持续进化并影响人类的健康。特别是，S基因的遗传不稳定性可能在人类群体中的HCoV-OC43基因型持续性发挥重要作用[68]，甚至在加重病毒对机体神经细胞的损伤[38]。在可预见的未来，HCoV-OC 43基因型还在演变，这种基因型转变可能是HCoV-OC 43维持其流行的适应机制[69]。

虽然SARS-CoV-2出现至今时间较短，但即使在急性期病情较轻的非住院COVID-19患者中，症状也可能需要数月才能消退，最持久的症状包括疲劳、嗅觉和味觉丧失以及关节痛等[70]。有临床证据表明SARS-CoV-2感染的神经系统表现可引起儿童重症，却很少有研究评估SARS-CoV-2感染对长期神经发育领域的影响[71]。COVID-19的长期健康后遗症使大量前COVID-19患者在多个健康领域存在严重问题[72]。这是当前被称为“长COVID”的新兴健康问题，它包括长期以来对神经系统造成的影响[73]。参考HCoV-OC43的演变和进化，COVID-19将继续存在，科学家们应该继续关注冠状病毒的未来突变和相关的健康危害[74]。显然，在SARS-CoV-2已经进入与人类长期共存的阶段，借鉴HCoV-OC43历史探究它在长期感染人体神经系统的病理机制和进化历程，不失为预测SARS-CoV-2乃至未来新发冠状病毒感染对机体神经系统的病理发生和发展，为及时做出应对提供参考。

参考文献

[1] MAO L, JIN H, WANG M, et al. Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China[J]. JAMA Neurol, 2020,77(6): 683-690.

[2] KIM M I, LEE C. Human Coronavirus OC43 as a Low-Risk Model to Study COVID-19[J]. Viruses, 2023,15(2).

[3] VIJGEN L, KEYAERTS E, MOËS E, et al. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event[J]. J Virol, 2005,79(3): 1595-1604.

[4] Leyden, E., and Guttmann, S. Die Influenza-Epidemie 1889/90[M].Verlag J.F. Bergmann, Wiesbaden, 1892.

[5] EWING E T. La Grippe or Russian influenza: Mortality statistics during the 1890 Epidemic in Indiana[J]. Influenza Other Respir Viruses, 2019,13(3): 279-287.

[6] VALLERON A J, CORI A, VALTAT S, et al. Transmissibility and geographic spread of the 1889 influenza pandemic[J]. Proc Natl Acad Sci U S A, 2010,107(19): 8778-8781.

[7] DALLMEIER K. Palaeoserology - teeth put into ancient plagues and pandemics[J]. Microb Biotechnol, 2022,15(7): 1940-1942.

[8] ARBOUR N, DAY R, NEWCOMBE J, et al. Neuroinvasion by human respiratory coronaviruses[J]. J Virol, 2000,74(19): 8913-8921.

[9] Victor S Huang. Mirroring History of the “Russian Flu”: Infectomic Decoding of COVID-19 [M]. Smashwords.

[10] SHAW B, GATHERER D. Candidate historical events for the emergence of Human Coronavirus OC43: A critical reassessment of the molecular evidence[J]. PLoS One, 2023,18(5): e285481.

[11] HONIGSBAUM M. The 'Russian' influenza in the UK: lessons learned, opportunities missed[J]. Vaccine, 2011,29 Suppl 2: B11-B15.

[12] KOELLE K, MARTIN M A, ANTIA R, et al. The changing epidemiology of SARS-CoV-2[J]. Science, 2022,375(6585): 1116-1121.

[13] MARKOV P V, GHAFARI M, BEER M, et al. The evolution of SARS-CoV-2[J]. Nat Rev Microbiol, 2023,21(6): 361-379.

[14] WANG C, ZHENG Y, NIU Z, et al. The virological impacts of SARS-CoV-2 D614G mutation[J]. J Mol Cell Biol, 2021,13(10): 712-720.

[15] MOHAMMADI M, SHAYESTEHPOUR M, MIRZAEI H. The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines[J]. Braz J Infect Dis, 2021,25(4): 101606.

[16] TAO K, TZOU P L, NOUHIN J, et al. The biological and clinical significance of emerging SARS-CoV-2 variants[J]. Nat Rev Genet, 2021,22(12): 757-773.

[17] LAMBROU A S, SHIRK P, STEELE M K, et al. Genomic Surveillance for SARS-CoV-2 Variants: Predominance of the Delta (B.1.617.2) and Omicron (B.1.1.529) Variants - United States, June 2021-January 2022[J]. MMWR Morb Mortal Wkly Rep, 2022,71(6): 206-211.

[18] CHEN J, WANG R, GILBY N B, et al. Omicron Variant (B.1.1.529): Infectivity, Vaccine Breakthrough, and Antibody Resistance[J]. J Chem Inf Model, 2022,62(2): 412-422.

[19] TORJESEN I. Covid-19: Omicron may be more transmissible than other variants and partly resistant to existing vaccines, scientists fear[J]. BMJ, 2021,375: n2943.

[20] LAU S K, LEE P, TSANG A K, et al. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination[J]. J Virol, 2011,85(21): 11325-11337.

[21] LEAO J C, GUSMAO T, ZARZAR A M, et al. Coronaviridae-Old friends, new enemy![J]. Oral Dis, 2022,28 Suppl 1(Suppl 1): 858-866.

[22] ST-JEAN J R, JACOMY H, DESFORGES M, et al. Human respiratory coronavirus OC43: genetic stability and neuroinvasion[J]. J Virol, 2004,78(16): 8824-8834.

[23] LAVELL A, SIKKENS J J, EDRIDGE A, et al. Recent infection with HCoV-OC43 may be associated with protection against SARS-CoV-2 infection[J]. iScience, 2022,25(10): 105105.

[24] MULABBI E N, TWEYONGYERE R, BYARUGABA D K. The history of the emergence and transmission of human coronaviruses[J]. Onderstepoort J Vet Res, 2021,88(1): e1-e8.

[25] LI W, SHI Z, YU M, et al. Bats are natural reservoirs of SARS-like coronaviruses[J]. Science, 2005,310(5748): 676-679.

[26] TANG G, LIU Z, CHEN D. Human coronaviruses: Origin, host and receptor[J]. J Clin Virol, 2022,155: 105246.

[27] LI C, CHEN W, LIN F, et al. Functional Two-Way Crosstalk Between Brain and Lung: The Brain-Lung Axis[J]. Cell Mol Neurobiol, 2023,43(3): 991-1003.

[28] LAHOUSSE L, TIEMEIER H, IKRAM M A, et al. Chronic obstructive pulmonary disease and cerebrovascular disease: A comprehensive review[J]. Respir Med, 2015,109(11): 1371-1380.

[29] BOHMWALD K, GÁLVEZ N, RÍOS M, et al. Neurologic Alterations Due to Respiratory Virus Infections[J]. Front Cell Neurosci, 2018,12: 386.

[30] BURKS J S, DEVALD B L, JANKOVSKY L D, et al. Two coronaviruses isolated from central nervous system tissue of two multiple sclerosis patients[J]. Science, 1980,209(4459): 933-934.

[31] YEH E A, COLLINS A, COHEN M E, et al. Detection of coronavirus in the central nervous system of a child with acute disseminated encephalomyelitis[J]. Pediatrics, 2004,113(1 Pt 1): e73-e76.

[32] ARBOUR N, CÔTÉ G, LACHANCE C, et al. Acute and persistent infection of human neural cell lines by human coronavirus OC43[J]. J Virol, 1999,73(4): 3338-3350.

[33] ST-JEAN J R, JACOMY H, DESFORGES M, et al. Human respiratory coronavirus OC43: genetic stability and neuroinvasion[J]. J Virol, 2004,78(16): 8824-8834.

[34] DUBÉ M, Le COUPANEC A, WONG A, et al. Axonal Transport Enables Neuron-to-Neuron Propagation of Human Coronavirus OC43[J]. J Virol, 2018,92(17).

[35] BUTLER N, PEWE L, TRANDEM K, et al. HCoV-OC43-induced encephalitis is in part immune-mediated[J]. Adv Exp Med Biol, 2006,581: 531-534.

[36] EDWARDS J A, DENIS F, TALBOT P J. Activation of glial cells by human coronavirus OC43 infection[J]. J Neuroimmunol, 2000,108(1-2): 73-81.

[37] JACOMY H, ST-JEAN J R, BRISON E, et al. Mutations in the spike glycoprotein of human coronavirus OC43 modulate disease in BALB/c mice from encephalitis to flaccid paralysis and demyelination[J]. J Neurovirol, 2010,16(4): 279-293.

[38] MEESSEN-PINARD M, Le COUPANEC A, DESFORGES M, et al. Pivotal Role of Receptor-Interacting Protein Kinase 1 and Mixed Lineage Kinase Domain-Like in Neuronal Cell Death Induced by the Human Neuroinvasive Coronavirus OC43[J]. J Virol, 2017,91(1).

[39] NIU J, SHEN L, HUANG B, et al. Non-invasive bioluminescence imaging of HCoV-OC43 infection and therapy in the central nervous system of live mice[J]. Antiviral Res, 2020,173: 104646.

[40] GIACOMELLI A, PEZZATI L, CONTI F, et al. Self-reported Olfactory and Taste Disorders in Patients With Severe Acute Respiratory Coronavirus 2 Infection: A Cross-sectional Study[J]. Clin Infect Dis, 2020,71(15): 889-890.

[41] MAURY A, LYOUBI A, PEIFFER-SMADJA N, et al. Neurological manifestations associated with SARS-CoV-2 and other coronaviruses: A narrative review for clinicians[J]. Rev Neurol (Paris), 2021,177(1-2): 51-64.

[42] MAO L, JIN H, WANG M, et al. Neurologic Manifestations of Hospitalized Patients With Coronavirus Disease 2019 in Wuhan, China[J]. JAMA Neurol, 2020,77(6): 683-690.

[43] WU Y, XU X, CHEN Z, et al. Nervous system involvement after infection with COVID-19 and other coronaviruses[J]. Brain Behav Immun, 2020,87: 18-22.

[44] HOFFMANN M, KLEINE-WEBER H, SCHROEDER S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor[J]. Cell, 2020,181(2): 271-280.

[45] HINGORANI K S, BHADOLA S, CERVANTES-ARSLANIAN A M. COVID-19 and the brain[J]. Trends Cardiovasc Med, 2022,32(6): 323-330.

[46] NATHAN N, PREVOST B, CORVOL H. Atypical presentation of COVID-19 in young infants[J]. Lancet, 2020,395(10235): 1481.

[47] GENTILE F, BOCCI T, COPPOLA S, et al. Putative Role of the Lung-Brain Axis in the Pathogenesis of COVID-19-Associated Respiratory Failure: A Systematic Review[J]. Biomedicines, 2022,10(3).

[48] BELL J S, SPENCER J I, YATES R L, et al. Invited Review: From nose to gut - the role of the microbiome in neurological disease[J]. Neuropathol Appl Neurobiol, 2019,45(3): 195-215.

[49] CAIXIA L, YANG X, YURONG T, et al. Involvement of epigenetic modification in epithelial immune responses during respiratory syncytial virus infection[J]. Microb Pathog, 2019,130: 186-189.

[50] NUZZO D, PICONE P. Potential neurological effects of severe COVID-19 infection[J]. Neurosci Res, 2020,158: 1-5.

[51] STEVENS R D, PUYBASSET L. The brain-lung-brain axis[J]. Intensive Care Med, 2011,37(7): 1054-1056.

[52] BAJINKA O, SIMBILYABO L, TAN Y, et al. Lung-brain axis[J]. Crit Rev Microbiol, 2022,48(3): 257-269.

[53] PENSATO U, MUCCIOLI L, CANI I, et al. Brain dysfunction in COVID-19 and CAR-T therapy: cytokine storm-associated encephalopathy[J]. Ann Clin Transl Neurol, 2021,8(4): 968-979.

[54] LAI F W, STEPHENSON K B, MAHONY J, et al. Human coronavirus OC43 nucleocapsid protein binds microRNA 9 and potentiates NF-κB activation[J]. J Virol, 2014,88(1): 54-65.

[55] ZHAO X, GUO F, LIU F, et al. Interferon induction of IFITM proteins promotes infection by human coronavirus OC43[J]. Proc Natl Acad Sci U S A, 2014,111(18): 6756-6761.

[56] COLLINS A R. Interferon gamma potentiates human coronavirus OC43 infection of neuronal cells by modulation of HLA class I expression[J]. Immunol Invest, 1995,24(6): 977-986.

[57] ERICKSON M A, RHEA E M, KNOPP R C, et al. Interactions of SARS-CoV-2 with the

Blood-Brain Barrier[J]. Int J Mol Sci, 2021,22(5).
[58] TAY M Z, POH C M, RÉNIA L, et al. The trinity of COVID-19: immunity, inflammation and intervention[J]. Nat Rev Immunol, 2020,20(6): 363-374.
[59] MERRILL J T, ERKAN D, WINAKUR J, et al. Emerging evidence of a COVID-19 thrombotic syndrome has treatment implications[J]. Nat Rev Rheumatol, 2020,16(10): 581-589.
[60] GALLO M B, AGHAGOLI G, LAVINE K, et al. Predictors of COVID-19 severity: A literature review[J]. Rev Med Virol, 2021,31(1): 1-10.
[61] YANG Y, ROSENBERG G A. Blood-brain barrier breakdown in acute and chronic cerebrovascular disease[J]. Stroke, 2011,42(11): 3323-3328.
[62] MIRZAEI R, GOODARZI P, ASADI M, et al. Bacterial co-infections with SARS-CoV-2[J]. IUBMB Life, 2020,72(10): 2097-2111.
[63] LAM T T, JIA N, ZHANG Y W, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins[J]. Nature, 2020,583(7815): 282-285.
[64] KIN N, MISZCZAK F, LIN W, et al. Genomic Analysis of 15 Human Coronaviruses OC43 (HCoV-OC43s) Circulating in France from 2001 to 2013 Reveals a High Intra-Specific Diversity with New Recombinant Genotypes[J]. Viruses, 2015,7(5): 2358-2377.
[65] OTTOGALLI M E, RODRÍGUEZ P E, FRUTOS M C, et al. Circulation of human coronaviruses OC43 and 229E in Córdoba, Argentina[J]. Arch Virol, 2021,166(3): 929-933.
[66] ZHU Y, LI C, CHEN L, et al. A novel human coronavirus OC43 genotype detected in mainland China[J]. Emerg Microbes Infect, 2018,7(1): 173.
[67] CARMAN K B, CALIK M, KARAL Y, et al. Viral etiological causes of febrile seizures for respiratory pathogens (EFES Study)[J]. Hum Vaccin Immunother, 2019,15(2): 496-502.
[68] REN L, ZHANG Y, LI J, et al. Genetic drift of human coronavirus OC43 spike gene during adaptive evolution[J]. Sci Rep, 2015,5: 11451.
[69] ZHANG Y, LI J, XIAO Y, et al. Genotype shift in human coronavirus OC43 and emergence of a novel genotype by natural recombination[J]. J Infect, 2015,70(6): 641-650.
[70] PETERSEN M S, KRISTIANSEN M F, HANUSSON K D, et al. Long COVID in the Faroe Islands: A Longitudinal Study Among Nonhospitalized Patients[J]. Clin Infect Dis, 2021,73(11): e4058-e4063.
[71] SINGER T G, EVANKOVICH K D, FISHER K, et al. Coronavirus Infections in the Nervous System of Children: A Scoping Review Making the Case for Long-Term Neurodevelopmental Surveillance[J]. Pediatr Neurol, 2021,117: 47-63.
[72] van den BORST B, PETERS J B, BRINK M, et al. Comprehensive Health Assessment 3 Months After Recovery From Acute Coronavirus Disease 2019 (COVID-19)[J]. Clin Infect Dis, 2021,73(5): e1089-e1098.
[73] NATH A. Long-Haul COVID[J]. Neurology, 2020,95(13): 559-560.
[74] ROKNUZZAMAN A, SARKER R, NAZMUNNAHAR, et al. The possibility and challenge evaluation about the declaration of end of the pandemic phase of Covid-19[J]. Ann Med Surg (Lond), 2023,85(5): 2237-2239.

Virginia Woolf's *Between the Acts*: A Trauma-Theory Case Study

James J. Lu

Reading Virginia Woolf's *Between the Acts* for the first time, readers most likely will find themselves puzzled by the novel's utterly chaotic scenes in a virtually plotless narrative. That was exactly what happened to me when I first read the book on my own, along with *To the Lighthouse*, *Mrs. Dalloway* and *The Waves* which were assigned as required texts in an English course I took as a graduate student at Duke University thirty some years ago. Back then, I conveniently and in hindsight presumptuously attributed the descriptive chaos in the novel to the author's innovative display of the stream-of-consciousness, as also employed by James Joyce and others. That approach didn't really help, for I remained baffled. I did not enjoy the book.

In this paper, I intend to reflect on that distant experience, and to demonstrate now a far better understanding gained over years of teaching and research in related subjects. I want to point at the severe damage caused by Germany's 1940 bombing of London not only to Virginia Woolf's homes but to her personally, psychologically. A direct link can be placed between the writer's traumatic experience during WWII and her symbolic representation of fragmented images in *Between the Act*. Taking advantage of some persuasive trauma theories now surging in neurosciences and in literary studies, I wish to highlight the functionality of writing as a potential therapy for artists such as Virginia Woolf to survive and self-express, as long or as temporarily as possible. Finally, I argue that while suicide in general is problematic, Virginia Woolf's suicide deserves reconsideration and reassessment.

Literary historians and biographers have noted that from her childhood, depression and suicidal ideation had haunted Virginia periodically. She suffered sexual abuse at a young age and, according to Emma Woolf, her great niece, between 1913-15 the novelist "made several suicide attempts, including trying to jump from a window…" (Newsweek, http://www.newsweek.com/015/02/27/joyful-gossipy-and-absurd-private-life-virginia-woolf-306438.html). Weak, miserable and perhaps deplorable she might have appeared, Virginia Woolf managed to survive, and later developed into one of the most famous writers of 20th century world literature. What eventually took her life in March, 1941 was apparently the worst nightmare of her life: World War II. Nazi Germany's horrifying destruction of London and other cities throughout Europe, especially of the novelist's own homes, profoundly affected, disturbed, and eventually doomed her.

Virginia Woolf's numerous diaries served as a factual documentation of the war and its impact on her mentally. On Thursday, 5 September 1940, she wrote, "Hot, hot, hot. Record heat wave, record summer if we kept records this summer. At 2.30 a plane zooms; then minutes later air raid sounds…." ("The Bombing of London" 47). Two days later, she recorded, "An air raid in progress. Planes zooming" ("The Bombing of London" 48). On September 10, she continued, "Back from half a day in London…. The house about thirty yards from ours struck at one this morning by a bomb. Completely ruined" ("The Bombing of London" 48). A week later, on September 18, Virginia learned that her own homes at two separate locations in London were severely damaged: "`We have need of all our courage' are the words that came to the surface this morning; upon hearing that all our windows are broken, ceilings down, & most of our china smashed at Meck. Sq. The bomb exploded. Why did we ever leave Tavistock?" ("A Writer's Diary," 338). Under Germans' frequent bombing of London, Virginia found the situation nerve-racking: "When the bombs make our windows rattle I always jump, I don't like sitting of an evening and thinking the drone which is weaving its web above me is about to drop" (LVI, 439).

Apart from diaries, Virginia Woolf also wrote essays. For example, in an essay titled "Thoughts on Peace in an Air Raid," published on 21 October 1940, Woolf describes air raids almost photographically. She also contemplates the possibility of future peace:

> The drone of the planes is now like a sawing of a branch overhead. Round and round it goes, sawing and sawing at a branch directly above the house. Another sound begins sawing its way in the brain… A bomb drops. All the windows rattle.
>
> There are to be no more guns, no army, no navy, no air force in the future. No more young men will be trained to fight with arms. (

No one seems to know for sure exactly when the author began to write her final novel *Between the Acts*. We may speculate that the book evolved into the present form during the last few months of 1940 and the beginning months of 1941 as the war was raging on. During that period of time Virginia struggled with stress and anxiety which took a heavy toll on her. She suffered from a draining depression. In the meantime, from diaries to essays and then to her final novel, she continued to write. Not only did she write to express her feelings, to document the war, but she also, in my opinion, were using writing as a means to search for the continued meaning of life under horrible duress.

Cathy Caruth, author of *Unclaimed Experience: Trauma, Narrative and History; Trauma: Explorations in Memory;* and *The Future of Trauma Theory: Contemporary Literary, Cultural Criticism*, and *Listening to Trauma: Conversations with Leaders in the Theory and Treatment of Catastrophic Experience,* is arguably the best-known trauma theorist in America today. Caruth "pioneered a psychoanalytic poststructural approach that suggests trauma is an unsolvable problem of the unconscious that illuminates the inherent contradictions of experience and language" (Balaev, https://link.springer.com/ chapter/10.1057/9781137365941_1). In *Trauma. Explorations in Memory*, she notes, "For the survivor of trauma, then, the truth of the event may reside not only in its brutal facts, but also in the way that their occurrence defies simple comprehension. The flashback or traumatic reenactment conveys… both the truth of an event, and the truth of its incomprehensibility" (*Trauma: Explorations in Memory*, 153) ed. Cathy Caruth). Literature, for Caruth, "opens a window on traumatic experience because it teaches readers to listen to what can be told only in indirect and surprising ways" (https://www.amazon.com/Trauma-Explorations-Memory-Cathy-Caruth/dp/ 080185007X). "An important aspect of Caruth' s approach to trauma," fellow trauma theorist Ruth Leys points out, "is her suggestion that if language nevertheless succeeds in testifying to the traumatic horror, it does so only when the referential function of words begins to break down, with the result that, as Walter Benn Michaels has put it, what is transmitted is "not the normalizing knowledge of the horror but the horror itself." ("Trauma and the Turn to Affect"; https://www.questia.com/ library/journal/1P3-2849429731/trauma-and-the-turn-to-affect).

The massive destruction caused by World War II, for Virginia Woolf, was an unbearably traumatic event. In *Between the Acts,* the author exhibited a tendency to break up the normal sequence of events, or the conventional syntax of sentences, by filling the narrative with interruptions and leaving gaps between language and empirical experience, which inevitably results in the novel's enigma or incomprehensibility.

Between the Acts is set in the summer of 1939, roughly a year in actuality before the Germans started bombing London. From the very beginning, the author hints at elements of a war. On the first page, mentioned are “graves in the churchyard,” and a bit later the narrator points to an aeroplane propeller spinning “faster, faster, faster…and [f]lying, rushing through the ambient, incandescent, summer....” (15). Much of the story centers on the annual Pageant for which every character prepares for and participates in. Described as “a bit of everything”: “Act; dance; sing….” (58), the Pageant turns out to be a chaotic and discordant play, full of sound and fury.

Shortly after Virginia Woolf committed suicide, on October 5, 1941, a critic named Hudson Strode wrote the following commentary: "`Between the Acts’ is one of Mrs. Woolf’s most seemingly simple books: the plot well integrated, the fancy under deft control. There is even a new calmness, a new clarity. She is not lost in webs of speculation, thin-drawn to incommunicability as in `The Waves’…." (). I do not agree with that observation. Most readers, as I did many years ago, would find the plot of *Between the Acts* random or unpredictable at best. Aside from the novel’s linguistic indeterminacy and ambiguous referentiality, the sequences of events are precariously arranged; they are hard to follow. Interestingly, within the novel’s narrative, author Virginia Woolf provides an answer to the question, “Does the plot matter?” From Page 90 to Page 91, a character named Isabella wonders about the confusion in the Pageant’s plot of which she “could make nothing” out (90): “Did the plot matter? She shifted and looked over her right shoulder...There was no need to puzzle out the plot…. Don’t bother about the plot: the plot’s nothing” (90-91).

The author uses a variety of indicative words to suggest frequent disruptions of normality during the Pageant. The word “interruption,” like other nouns such as “death,” “silence,” “blast,” “scraps and fragments,” or adjectives “empty,” “degenerate,” “dispersed,” “gone,” “forgotten,” appears repeatedly in the narration. Striking, too, are a number of three-word refrains: for instance, “Tick, tick, tick, went the machine in the bushes” (174); “Tick, tick, tick the machine reiterated” (175); “Tick, tick, tick the machine continued” (82); then “Chuff, chuff, chuff the machine ticked” (81); “Chuff, chuff, chuff went the machine” (82); and “The chuff, chuff, chuff of the machine in the bushes…” (157). The monotonous noises somehow resemble the sounds of machine guns. Coupled with words such as “blast,” it is not far fetching to interpret Virginia’s choices of all the words above mentioned as indirect references to the raging war and its impact.

Daniel Ferrer, in his book *Virginia Woolf and the Madness of Language,* analyzes how “Woolf's madness (not as a thematics, but as a condition of texts) [is] related to the essential structure of language, as acts to cover over fundamental voids” (qtd. Geoffrey Bennington and Rachel Bowlby, 16). For Woolf, “Writing [is] only made possible by a process of making the void and denying madness and death, but it is also, contradictorily, a reiteration of a void, a taking up of madness, a repetition of death” (Ferrer, 99).

Virginia Woolf’s madness at destructive forces finds expression in her occasionally mad language. In *Between the Acts*, at one point the author lets one character utter, with anger and distain: “those damned Germans” (151). She even uses the word “maddening” when describing metaphorically how everyone in the audience who stares at the Pageant’s stage looks like a prisoner: “They were all caught and caged; prisoners; watching a spectacle. Nothing happened. The tick of the machine was maddening” (176). Void and death literally scatter everywhere, pervading the story. For example, on page 120, it reads, “The stage was empty”; on page 181, “Civilization…in ruins”; and on page 189, “The voice died away” (189); and on page 147, “All gone. Following the wind. He is gone; she is gone….” Then, as the Pageant continues, readers can hear one character growling, “O that our human pain could here have an ending” (180).

Symbolically, of course, all the above descriptions could be interpreted as related to war and its devastating consequences.

Longer passages implying void, death, sorrow and despair thread through the entire narrative. From Page 36 to Page 37, "Empty, empty, empty; silent, silent, silent. The room was a shell, singing of what was before time was; a vase stood in the heart of the house, alabaster, smooth, cold, holding the still, distilled essence of emptiness, silence." Later, "…the shower fell, sudden, profuse. No one had seen the cloud coming. There it was, black, swollen, on top of them. Down it poured like all the people in the world weeping. Tears, Tears, Tears" (180). And, as a final example, on Page 188, "Look at ourselves, ladies and gentlemen! Then at the wall; and ask how's this wall, the great wall, which we call, perhaps miscall, civilization, to be built by (here the mirrors flicked and flashed) orts, scraps and fragments like ourselves?"

In "Trauma and Recovery in Virginia Woolf's *Mrs. Dalloway*," Karen DeMeester proposes that Virginia Woolf's portrayal of Septimus Smith "illustrates not only the psychological injuries suffered by victims of severe trauma such as war but also the need for them to give meaning to their suffering in order to recover from the trauma" (https://muse.jhu.edu/article/21307/summary). If Woolf has intended in *Mrs. Dalloway* to illustrate "not only the psychological injuries" of a war victim "but also the need…to give meaning to…suffering in order to recover from the trauma," I argue that in *Between the Acts*, although no character is portrayed clearly as a victim of war, we should feel empathy for Virginia Woolf as a victim of World War II, a victim who, through writing, gave profound meaning to suffering not only in order for herself to survive, but to inform the readers of the real suffering and its meaning.

Between the Acts can be interpreted as of "Destruction Art," a term Kristine Stiles has coined to "identify the presentational works that situate the body in the center of the question of destruction and survival" (30). "In destruction art, artists present the `imagery of extinction' localized in the body, the object, which is offered as a destructible material and/or an agent of that destruction" (31). Stiles, too, recognizes writing's therapeutic function: "…the destruction art is not only about the presentation of destruction. It is also one of the means by which a world of consciousness is being formed that may contribute to the construction of an altered sense of self, which is necessary to insure human survival" (31). *Between the Acts* differs from Virginia Woolf's other works such as her diaries and essays written during the final year of her life, in terms of journalistic accuracy or documentary values. The novel, nonetheless, reflects truthfully the author's disturbed state of mind, her strong creative impulses, her sharp perception of reality, and her pessimistic view of the world's future. That Virginia Woolf produced such a novel, as it is, signifies that through a conscious act of writing she managed to draw enough aspiration and inspiration to sustain her life till she could no longer.

We know that on March 27, 1941, Virginia Woolf wrote to her publisher, John Lehmann, about the manuscript of *Between the Acts*, describing it as "too silly and trivial" to bring out in its present form, and promising to revise it for publication later in the year. The next day, after writing another note, a suicide note, to her husband Leonard Woolf, she filled her pockets with stones and drowned herself in a river near her home. Her suicide gave rise to some controversy. Sybil Oldfield, editor of *Afterwords: Letters on the Death of Virginia Woolf,* observes that upon Virginia's death there were critics attacking her "for showing cowardice in the face of the enemy and for setting a bad example to the general population" (https://www.goodreads.com/book/show/108651).

In defense of Virginia Woolf, I propose that it is important to understand her suicide holistically. Before drowning herself, Virginia had attempted suicide more than once, but World War II was the worst traumatic experience in her life. If previously the primary cause for her attempted suicide was largely personal or domestic, now during World War II, with Nazi Germany's constant bombing and rumors of a German invasion, she must have felt a strong sense of loss not merely of personal safety or meaning of existence, but also that of the entire humanity and civilization. It is important to note that it was at the height of Hitler's victories during the War that Virginia Woolf committed suicide. When witnessing evil forces launching a large-scale war with massive destruction of towns, cities and human lives, it is natural for anyone, especially one with chronical depression, to find herself or himself disturbed, traumatized, fearful, powerless, helpless, and hopeless. In Virginia Woolf's case, the fear of death in her and among her fellow countrymen at the time was real. That is perhaps why author Virginia Woolf depicts a character in *Between the Acts* groaning, "The doom of sudden death hanging over us" (114). Although literature and writing provided her relief, Virginia Woolf's depression, understandably, worsened. She revealed in her suicide note to Leonard Woolf: "I feel certain I am going mad again. I feel we can't go through another of those terrible times. And I shan't recover this time. I begin to hear voices, and I can't concentrate…" (https://www.brainpickings.org/2014/03/28/virginia-woolf-suicide-letter/). Indeed, according to an Associated Press report dated April 19, 1941: "Her husband testified that Mrs. Woolf had been depressed for a considerable length of time" (http://hcox7.blogspot.com/2013/11/). Apparently novelist Virginia Woolf was suffering not only mentally or only bodily but rather both mentally and bodily. For a trauma patient, when physical as well as mental suffering reaches a certain point, "death instinct" could rise and intensify. Those who accused Virginia Woolf of cowardice unfortunately failed to take into account her physical and mental illness; therefore, their attack was not fair in the first place.

More importantly, Virginia Woolf's stated reason for suicide should earn her respect if not admiration. In her suicide note, she expresses, in a calm and warm tone, her appreciation of the loving care from her husband over the years. She makes it clear that it is her desire not to continue burdening her beloved ones that she would choose to end her life. The entire suicide note reads as follows:

> Dearest,
>
> I feel certain I am going mad again. I feel we can't go through another of those terrible times. And I shan't recover this time. I begin to hear voices, and I can't concentrate. So I am doing what seems the best thing to do. You have given me the greatest possible happiness. You have been in every way all that anyone could be. I don't think two people could have been happier till this terrible disease came. I can't fight any longer. I know that I am spoiling your life, that without me you could work. And you will I know. You see I can't even write this properly. I can't read. What I want to say is I owe all the happiness of my life to you. You have been entirely patient with me and incredibly good. I want to say that — everybody knows it. If anybody could have saved me it would have been you. Everything has gone from me but the certainty of your goodness. I can't go on spoiling your life any longer. (https://www.brainpickings.org/2014/03/28/virginia-woolf-suicide-letter/)

Finally, while in general suicide means a self-inflicted pain and loss of life, which Christians view as offensive, it has been used, throughout human history, but especially in recent centuries, as a means for political or social protest. As Michael Biggs observes in an article "How Repertoires Evolve: The Diffusion of Suicide Protest in the Twentieth Century," increasingly we see instances of individuals "killing oneself, without harming others, for a collective cause" (http://mobilizationjournal.org/doi/abs/10.17813/maiq.18.4.njnu779530x55082?journalCode=maiq). Therefore, it is possible and reasonable to view Virginia Woolf's suicide as a silent cry, a courageous protest, and a serious indictment of the crimes against humanity committed by the powerful yet evil and destructive forces -- the German Nazis.

Conclusion: Virginia Woolf's final novel *Between the Acts* can be read as a swan song, her final creative performance as a novelist. With a virtually plotless plot and a narrative fragmented by abrupt and discordant verbal expressions, this novel would continue to baffle first-time readers. To better understand the text, it is imperative to examine the broad socio-historical context within which the text was produced, and to analyze the author's state of mind when she was writing it. Relevant trauma theories can help readers penetrate the novel's enigma or incomprehensibility. Virginia Woolf's traumatic experience of World War II, specifically, of the brutal bombings by German Nazis which partially destroyed her own homes in London, led her to a mental breakdown. By engaging herself in writing, the novelist managed to survive as long as she could. In that sense, *Between the Acts* becomes a transformative artwork through which the artist was able to find a way to metaphorically re-register her high-intensity trauma, to reach and push up her fragmented memories of the horrific events now metamorized in the images of broken glasses, empty stage-scenes, and explosive noises – all of which in turn works to effect a process of neuropsychological therapy that functions to sustain Virginia Woolf, at least for the time being. As to her suicide, clearly it had little or nothing to do with cowardice. Her desire and act to sacrifice herself so that her husband could finally be lifted off the heavy burden she knew she had been, deserves respect. Along with her diaries, essays and novel *Between the Acts*, Virginia Woolf's final act of suicide, if viewed historically and politically, can be reasonably interpreted as a solemn indictment, a passive protest against the Nazis for their threat to the entire humanity and civilization.

Works Cited

Balaev, Michelle. *Literary Trauma Theory Reconsidered*. https://link.springer.com/chapter/10.1057/9781137365941_1)

Biggs, Michael. "How Repertoires Evolve: The Diffusion of Suicide Protest in the Twentieth Century." In *Mobilization: An International Quarterly*: December 2013, Vol. 18, No. 4, pp. 407-428).

Caruth, Cathy. *Trauma: Explorations in Memory*.

Ferrer, Daniel. *Virginia Woolf and the Madness of Language.* Trans. Geoffrey Bennington and Rachel Bowlby. London and New York: Routledge Publishers, 1990.

Newsweek, "The Joyful, Gossipy and Absurd Private Life of Virginia Woolf" http://www.newsweek.com/ 2015/02/27/ joyful-gossipy-and-absurd-private-life-virginia-woolf-306438.html

Oldfield, Sybil. "Afterwards." *Letters on the Death of Virginia Woolf* https://www.goodreads.com/ book/show/108651.

Strode, Hudson. "The Genius of Virginia Woolf"

Woolf, Virginia. *Between the Acts*. San Diego, New York, London: A Harvest/HBJ Book, 1941.

"The Bombing of London." in *Life Writing*. Ed. Winifred Bryan Horner. Princeton, New Jersey: A Blair Press Book, 1997.

-----------------. *A Writer's Diary: Being Extracts from the Diary of Virginia Woolf*. Mariner Books;

2003.

癌症的靶向治疗和精准医学

刘西平, PhD
NantWorks/ConjuChem, Culver City, CA 90230, USA

随着现代医学科技的迅猛发展，癌症治疗正经历着革命性的变革。作为一种严重危害人类健康的疾病，癌症一直以来都是医学界和社会的关注焦点。然而，传统的癌症治疗方法，如化疗、放疗等，存在一定的副作用，且对不同患者的疗效有较大差异。为了更有效地应对这一挑战，癌症的靶向治疗和精准医学逐渐崭露头角，为癌症患者带来了新的希望。

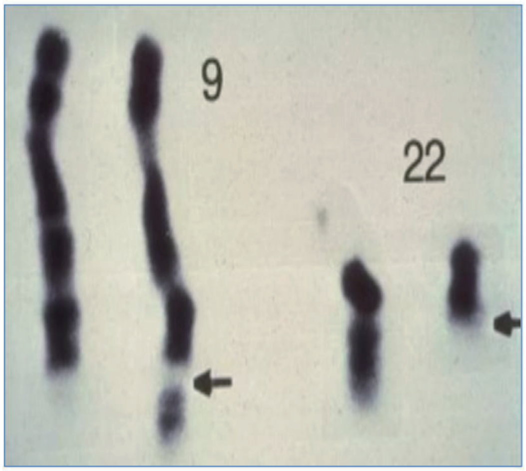
靶向治疗作为一种精确干预癌症生长和发展的策略，通过针对癌症细胞特定的分子靶点，精准地抑制其异常信号通路，从而实现对癌症的有针对性攻击。这种治疗方法不仅可以减少对正常细胞的损害，还能够最大限度地提高治疗效果，为患者带来更优质的生存质量。通过深入研究患者个体的遗传信息、生物标志物和分子特征，精准医学能够为每位患者量身定制最适合的治疗方案，从而实现更加精准、更为有效的治疗效果。

本文将侧重介绍靶向治疗和精准医学在癌症领域的应用和进展。通过相关的科学原理、经典案例以及最新研究，展示在改变癌症治疗、提升患者生活质量方面所取得的突破。同时，还将探讨目前癌症治疗所面临的挑战和未来的发展趋势，为读者了解癌症治疗提供一些参考。

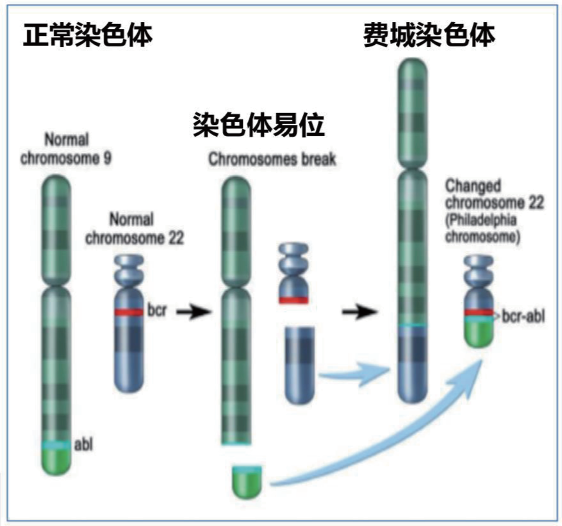
一、慢性髓性白血病：癌症靶向治疗的典范

慢性髓性白血病（又称慢性粒细胞白血病，CML）是一种以髓系增生为主的造血干细胞恶性肿瘤。超过 90%的 CML 患者携带费城染色体，这一遗传特征成为 CML 的重要标志。费城染色体最早是由科学家彼得·克鲁格（Peter Nowell）和大卫·亨利（David Hungerford）于 1960 年代初发现的。当时，克鲁格和亨利在费城的 Wistar 研究所工作，他们利用细胞染色体显微镜技术研究了 CML 患者的细胞。他们观察到在 CML 患者的染色体中，第 9 号染色体与第 22 号染色体发生易位，形成了一种超短的 22 号染色体，被称为费城染色体。进一步研究表明，费城染色体上的某些基因融合产生了一种异常的融合蛋白，即 BCR-Abl 融合蛋白，这一蛋白在 CML 的发展中起着关键作用（图 1）。BCR-Abl 融合蛋白具有异常的酪氨酸激酶活性，促使血液细胞异常增殖和分化，导致 CML 的病理变化。这一发现揭示了 CML 的分子遗传学基础，为后来针对 BCR-Abl 融合蛋白的靶向治疗提供了重要线索，同时也为癌症的分子诊断和治疗开辟了新途径，成为肿瘤学领域的重要突破[1-4]。

图1. 慢性粒细胞白血病的分子生物学



1973年，芝加哥大学的**珍妮特·罗利**（Janet Rowley）教授在彼得的发现基础上，又前进了一层。她的团队发现，费城染色体之所以短，是因为发生了染色体的易位：CML患者的9号染色体与22号染色体发生了一部分的交换，让22号染色体短上那么一截。她敏锐地指出，這个特殊的易位背后，一定存在著某种特殊的致癌机理。



1983年，美国国立癌症研究所（NCI）的学者们发现，由于两条染色体之间发生的交错易位，9号染色体上的Abl基因，恰好与22号染色体上的BCR基因连到了一塊，产生了一条BCR-Abl融合基因。這条融合基因编码了一种奇特的酪氨酸激酶。

CML 经历了放疗、化疗、免疫治疗、骨髓移植、靶向治疗等一系列治疗措施，疗效逐渐提高。放疗、化疗主要用于缓解症状，但难以改变 CML 的自然进程。到了 20 世纪 90 年代末，酪氨酸激酶抑制剂伊马替尼（Imatinib）的成功问世，成为 CML 治疗的重大里程碑，使 CML 治疗进入了靶向治疗时代。在 1998 年伊马替尼开始进行临床试验，I 期研究结果显示，该药物不仅耐受良好，而且表现出显著疗效。在 54 名接受治疗的患者中，有 53 名达到了血液学完全缓解。更令人振奋的是，治疗一年半后，患者的无进展生存率仍高达 89.2%。美国食品药品监督管理局（FDA）于 II 期临床试验后迅速批准伊马替尼上市，全球首款小分子靶向治疗药物正式亮相。这款药物的商品名，即我们所熟知的格列卫（Gleevec）。

然而，部分 CML 患者在治疗过程中出现伊马替尼耐药性，主要是 BCR-Abl 激酶区发生了突变，使得伊马替尼无法结合或结合力下降。为克服伊马替尼耐药性，第二代和第三代 BCR-Abl 酪氨酸激酶抑制剂应运而生，例如达沙替尼（Dasatinib）（商品名：施达赛 Sprycel），适用于携带 ABL 突变型 Y253H，E255K/V，F359V/C/I 的患者治疗；尼洛替尼（Nilotinib）（商品名：达希纳 Tasigna），适用于携带 ABL 突变型 V299L，T315A，F317L/V/I/C 的患者治疗；普纳替尼（Ponatinib）（商品名：Iclusig），适用于携带 ABL 突变型 T315I 的患者治疗。在面对伊马替尼耐药性的 CML 患者时，选择适当的第二代和第

三代 BCR-Abl 酪氨酸激酶抑制剂需要综合考虑患者的临床情况、突变谱、药物特点等因素。对患者进行基因测序，特别是检测 BCR-Abl 激酶区的突变谱，可以帮助确定哪些突变可能导致耐药性。不同的突变可能对不同的抑制剂表现出不同的敏感性，因此了解患者的突变情况可以指导药物选择。有幸的是，第二代和第三代 BCR-Abl 酪氨酸激酶抑制剂能够有效地抑制绝大多数携带 BCR-Abl 激酶区突变的酪氨酸激酶活性[5-7]。

正好在伊马替尼开始临床试验前 10 年，即 1988 年，笔者参与了关于中国慢性粒细胞白血病特异性 BCR 断裂点的研究。这项研究首次在中国 CML 患者中发现了特异性的 BCR 断裂点，表明中国患者的 BCR-Abl 融合蛋白存在区域差异。通过揭示这些特异的 BCR 断裂点，为深入了解中国患者的分子特征以及 CML 患者的靶向治疗和相关药物研究打下了基础[8-9]。

二、PD-1/PD-L1 抑制剂：癌症靶向治疗的突破

随着肿瘤生物学及相关学科的发展，人们逐渐认识到细胞癌变的一个重要原因是细胞信号转导通路的紊乱，从而导致细胞无限增殖。这一认识推动了抗肿瘤药物研发的重心从传统的细胞毒性药物转向能够对肿瘤细胞内异常信号系统的靶点发挥作用的药物。2018 年，诺贝尔生理学或医学奖颁发给了美国安德森癌症中心免疫学系主任詹姆斯·P·艾利森（James P. Allison）教授和日本京都大学免疫学教授本庶佑（Tasuku Honjo）教授，以表彰他们在人类肿瘤免疫治疗方面的卓越贡献。艾利森教授率先提出了免疫检查点（immune checkpoint）的概念，并在小鼠实验中证实杀伤性 T 淋巴细胞相关抗原 4（CTLA-4）抗体能够增强免疫系统对抑制肿瘤的反应。与此同时，本庶佑教授首次发现程序性死亡受体-1（PD-1）在激活 T 淋巴细胞时的诱导作用，并在随后的研究中揭示了 PD-1 作为免疫反应负调节因子的重要性。这些突破性发现为癌症治疗带来了崭新的理念[10-11]。

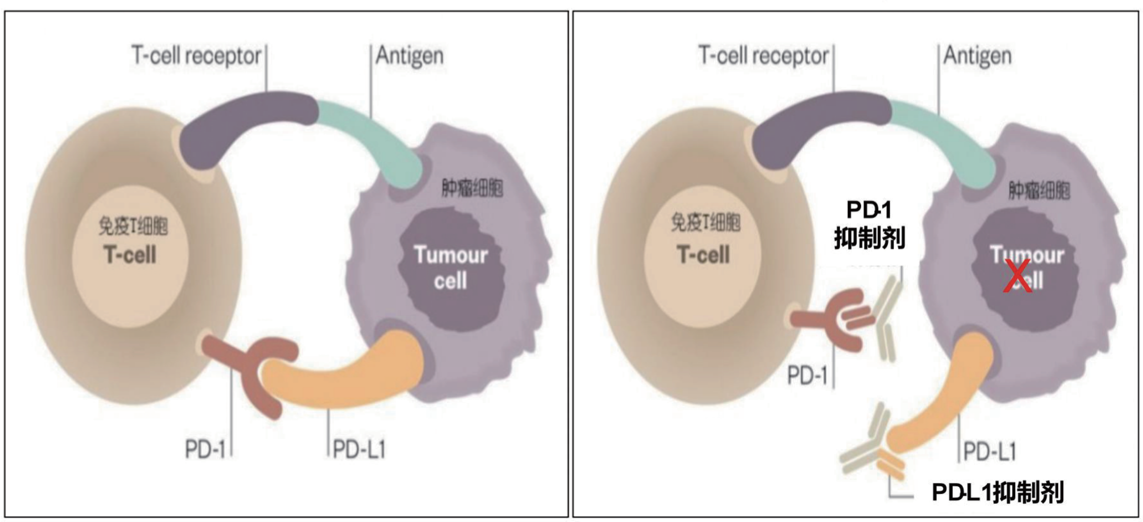
PD-1/PD-L1 抑制剂是一种创新的癌症靶向治疗方法，其基本原理是通过阻断免疫检查点分子 PD-1（程序性死亡受体-1）和其配体 PD-L1（程序性死亡配体-1）之间的相互作用，从而激活患者自身的免疫系统，增强其攻击和清除肿瘤细胞的能力（图 2）。

自从 2011 年 CTLA-4 抗体伊匹木单抗（Ipilimumab，商品名：Yervoy）首次被批准用于治疗转移性黑色素瘤以来，以 PD-1 通路阻断为中心的 7 种免疫检查点抑制剂（Immune Checkpoint Inhibitors，ICIs）已获 FDA 批准，用于治疗 18 种不同类型的癌症。其中有 4 种 PD-1 抑制剂： 1、帕博利珠单抗（Pembrolizumab，商品名：Keytruda）已获批用于多种癌症的治疗，包括黑色素瘤、非小细胞肺癌、肾细胞癌、头颈部鳞状细胞癌、霍奇金淋巴瘤等。2015 年，美国前总统卡特被诊断患有晚期黑色素瘤，在经过放疗和帕博利珠单抗注射液联合治疗 6 个月后，肿瘤消失了。帕博利珠单抗注射液是第一个在一线治疗晚期非小细胞肺癌（NSCLC）中显示优于化疗的 PD-1 抗体药物；2、纳武利尤单抗

（Nivolumab，商品名：Opdivo）已批准用于多种癌症的治疗，包括黑色素瘤、非小细胞肺癌、肾细胞癌、结直肠癌、胃癌、头颈部鳞状细胞癌、霍奇金淋巴瘤等；3、西米普利单抗（Cemiplimab，商品名：Libtayo）是一种用于特定皮肤癌类型治疗的抗 PD-1 单克隆抗体，如普通型和重症鳞状细胞癌；4、多斯塔利单抗（Dostarlimab，商品名：Jemperli）适用于错配修复缺陷复发 (dMMR) 或晚期实体瘤成人患者。2022 年，《新英格兰医学杂志》报道了一项针对局部晚期错配修复缺陷直肠癌患者的研究结果，12 名患者完成了多斯塔利单抗的治疗，并经历了至少 6 个月的随访。所有患者均达到临床完全缓解，核磁共振成像和活检均未发现肿瘤。截至报告发布时，没有患者接受过放化疗或手术，且在随访期间（范围为 6 至 25 个月），没有出现恶化或复发的病例。与此同时，还有 3 个 PD-L1 药物获得 FDA 批准，用于多种癌症治疗，如阿替利珠单抗注射液（Atezolizumab，商品名：Tecentriq）、度伐利尤单抗注射液（Durvalumab，商品名：Infinzi）和阿维单抗（Avelumab，商品名：Bavencio）。尽管它们都是免疫检查点抑制剂，但 PD-1 和 PD-L1 抑制剂之间存在区别。由于 PD-1 的配体还包括 PD-L2，因此，PD-L1 抑制剂仅阻止 PD-1 与 PD-L1 的结合，而 PD-1 抑制剂同时阻止 PD-1 与 PD-L1 以及 PD-L2 的结合，这可能导致它们的作用效果不同。从理论上讲，PD-1 抑制剂更容易引发免疫相关的不良事件，而 PD-L1 抑制剂作用于肿瘤细胞的 PD-L1，可能对正常组织细胞的损伤较小。然而，在临床实际应用中，发现 PD-1 抑制剂在一些患者中的疗效优于 PD-L1 抑制剂[12-14]。

图2. PD-1/PD-L1 抗体如何发挥抗癌作用的?

肿瘤细胞如何逃避免疫系统 **抗PD1/PD1L药物作用原理**



狡猾的癌细胞会表达PD-L1蛋白，并与T细胞上的PD-1受体结合，令T细胞不会对癌细胞作出攻击。

抗PD-1药物原理乃阻断PD-1/PD-L1信号通路，刺激T细胞重新“认识”癌细胞并将癌细胞清除。

尽管 PD-1/PD-L1 抑制剂在许多患者中显示出良好的疗效，但并非所有患者都能够获得相同的益处。治疗效果受多种因素影响，因此在临床实践中需要综合考虑患者的病情、免疫状态等因素。总之，PD-1/PD-L1 抑制剂代表了肿瘤免疫治疗领域的一项重要进展。它们不仅改变了许多癌症患者的生存前景，还为未来的肿瘤治疗开辟了新的道路。

三、T 细胞疗法：癌症靶向治疗的创新

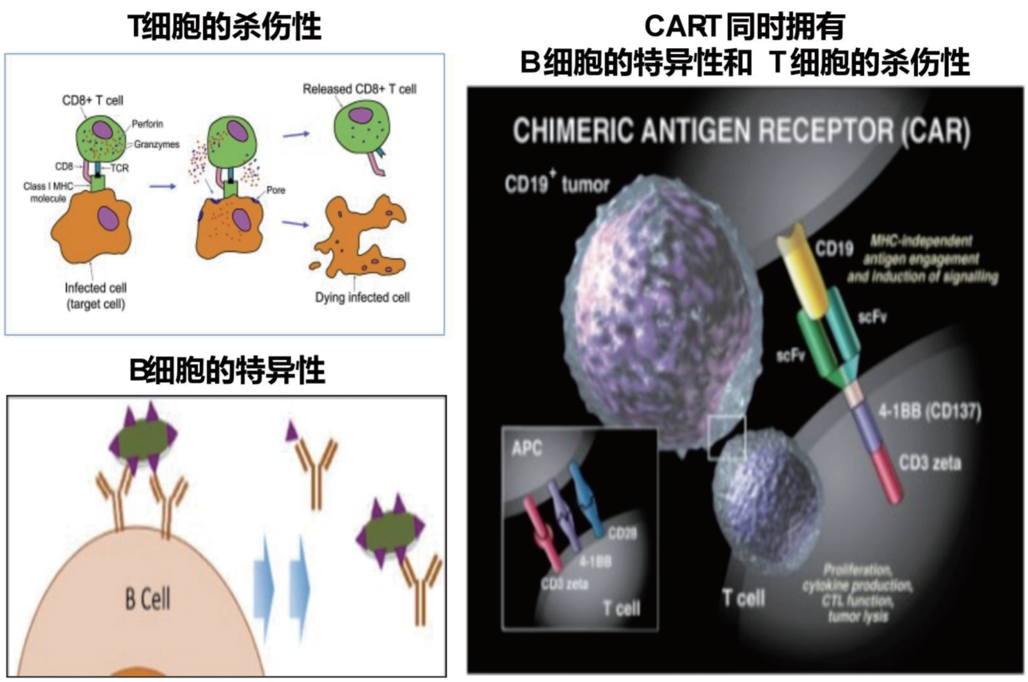
当人体免疫系统检测到体内存在异常细胞，如癌细胞时，杀伤性 T 细胞（Cytotoxic T Lymphocytes, CTLs）成为重要的免疫细胞，发挥着抗癌作用。CTL 通过识别癌细胞表面的特异性抗原来被激活，进而释放细胞毒素，如穿孔蛋白酶和葡萄糖苷酸酶。此外，它们通过与癌细胞表面的肿瘤坏死因子受体超家族成员 6（Fas）结合来诱导癌细胞凋亡。部分 CTL 还会转化为记忆性 CTL，在体内保留对癌细胞的识别能力。这有助于在未来再次遭遇癌细胞时快速作出反应并消灭它们，从而有助于抑制癌症的复发。

人体免疫系统中的杀伤性 T 细胞可以发挥抗癌作用[15]，但癌症细胞也会采取多种策略来逃避免疫系统的检测和攻击，从而限制了免疫系统的效力。其中包括：免疫逃避，即癌细胞通过减少特异性抗原的表达，改变抗原表达模式，或通过其他机制来减低免疫系统对其的识别和攻击。此外，癌细胞还可以通过分泌抑制性分子来影响免疫细胞的功能，从而抑制免疫应答的强度和效力。此外，免疫编辑现象也可能出现，即免疫系统在癌症发展的早期阶段识别和攻击癌细胞，但随着时间的推移，癌细胞可能会逐渐进化，以逃避免疫系统的攻击，导致免疫系统无法有效地清除癌症。为了克服这些挑战，科学家们开发出了一种癌症治疗性疫苗，旨在增强免疫系统对癌细胞的攻击能力。癌症治疗性疫苗的设计可以针对特定的抗原，也可以利用整个癌细胞的抗原组合，从而在患者体内引发持久的免疫应答，有助于抑制癌症的生长和扩散。与预防性疫苗不同，治疗性癌症疫苗是在患者已经患有癌症时使用的，它们的目标是增强机体对癌细胞的免疫反应，从而达到治疗的效果[16-17]。

然而一些晚期癌症患者免疫系统功能受损，难以通过免疫刺激方法获得足够的治疗效果。近年来，一种新兴的免疫疗法-嵌合抗原受体 T 细胞（CAR-T）免疫疗法，为该领域带来了新的希望。CAR-T 细胞疗法基于改造患者的 T 细胞，使其具备更强大的癌细胞识别和攻击能力。这一过程包括采集患者的 T 细胞，通过基因工程技术在其表面引入嵌合抗原受体（CAR），使其能够特异性地识别肿瘤细胞上的抗原，最后将改造后的 CAR-T 细胞重新输注入患者体内。一旦进入患者体内并识别到肿瘤抗原，CAR-T 细胞会被激活并向肿瘤细胞发动攻击，实现对癌症的靶向治疗（图 3）。CAR-T 治疗的优势在于其高度个体化和针对性，因为 CAR-T 细胞是从患者自身的免疫系统中取得并改造的。它在一些晚期和难治性血液恶性肿瘤，如急性淋巴细胞白血病（ALL）、非霍奇金淋巴瘤（NHL）和多

发性骨髓瘤（MM）的治疗中取得了令人瞩目的治疗效果。著名的案例包括美国宾夕法尼亚州 6 岁女孩 Emily Whitehead 和英国 1 岁女孩 Layla Richards，她们在接受 CAR-T 细胞治疗后，白血病得到了控制。随后，FDA 也批准了多种 CAR-T 细胞疗法，如用于治疗难治/复发性 B-ALL 儿童或年轻患者的 CD19 CAR-T 产品（商品名为 kymriah）和用于治疗患有大 B 细胞淋巴瘤的成年患者的 CAR-T 细胞疗法（商品名为 Yescarta）。尽管 CAR-T 细胞疗法在癌症治疗中表现出巨大潜力，但仍然面临着一些挑战，如治疗后的细胞耐受性、副作用控制和生产成本等。随着技术的进步，这些问题有望得到逐步解决，为 CAR-T 细胞疗法的广泛应用创造更有利的环境。

图3. 嵌合抗原受体T细胞（CART)

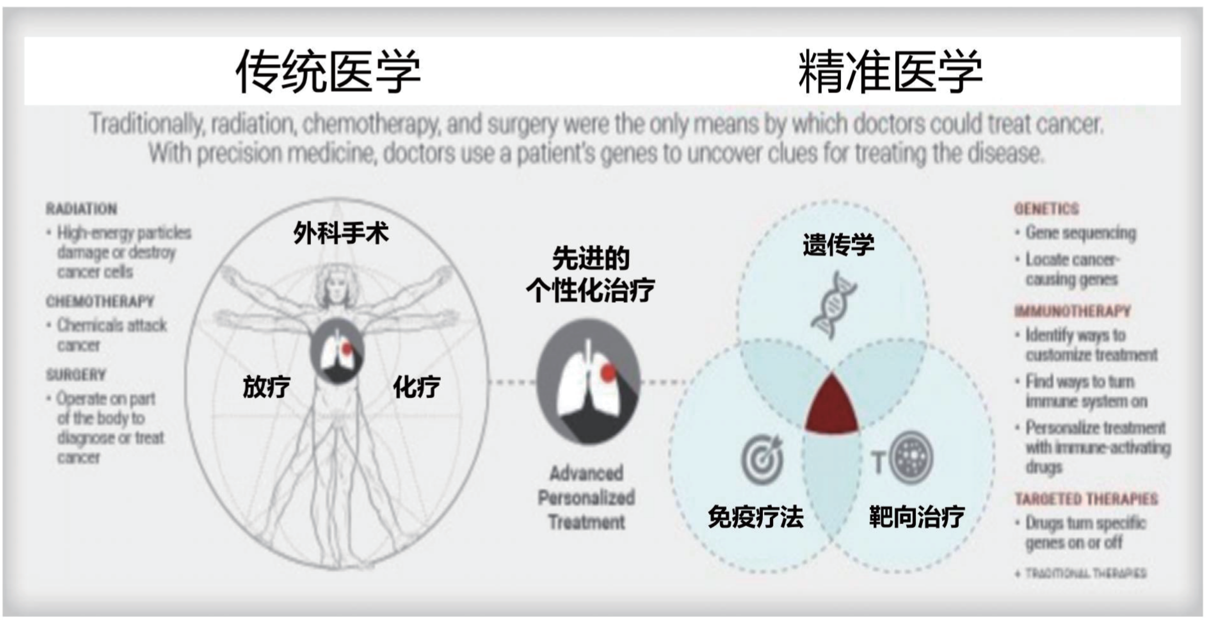


笔者在 2000 年前后参加了早期 T 细胞受体克隆和功能分析以及癌症治疗性疫苗的研究。通过使用免疫转基因小鼠，成功地克隆了能够特异性识别肿瘤蛋白 p53 的 T 细胞受体（TCR）。将这种 T 细胞受体转染到人的突变 T 细胞株中后，转染的 T 细胞能够特异性识别表达肿瘤蛋白 p53 的乳腺癌、胰腺癌和肉瘤细胞。这些研究结果首次证明了特异性 p53 T 细胞受体在异源细胞中的表达，并使其获得了对表达突变 p53 的肿瘤细胞的识别能力。采用这种特异性 T 细胞受体可望对抗人类癌症提供了一个潜在的治疗策略[18-19]。

四、精准医学: 临床诊断和靶向治疗的完美结合

精准医学，也称为个体化医学或定制医学，是一种基于个体的遗传信息、生物标志物和其他生物学特征来指导疾病预防、诊断和治疗的做法。在肿瘤学中，精准医学旨在根据患者肿瘤的个体特征，制定更加个性化和针对性的治疗方案，以提高治疗效果，并减少不必要的治疗和副作用。癌症靶向药物和癌症治疗性疫苗都是精准医学的重要组成部分，这些治疗都是针对肿瘤细胞中特定的异常信号通路或靶点，从而抑制肿瘤细胞的生长和扩散。癌症靶向药物和癌症治疗性疫苗的研发依赖于对肿瘤细胞分子机制的深入理解。通过对肿瘤细胞基因突变、蛋白表达和信号通路的研究，科学家们可以发现一些在肿瘤发生和发展中起关键作用的分子靶点。利用基因测序、蛋白质组学等手段检测患者肿瘤的特定变异，为他们选择最适合的靶向治疗 （图 4）。

图4. 精准医学与肿瘤靶向治疗



今天的癌症治疗应该不再采用一刀切的方法，应该基于构成癌症的基因， 通过检查患者肿瘤的DNA, 识别导致癌症生长的致癌基因。 有了这些信息，就能够找到针对该基因的药物，特异性攻击癌细胞。

当涉及精准医学在癌症治疗中的应用时，以下是一些示例，展示了癌症靶向药物如何根据患者的个体情况进行定制化治疗：1、HER2 阳性乳腺癌治疗： HER2 阳性乳腺癌患者通常表达过多的 HER2 受体。药物如赫赛汀（Herceptin）就是一种靶向 HER2 的抗体药物，它可以与癌细胞表面的 HER2 受体结合，抑制癌细胞生长和扩散。2、EGFR 突变非小细胞肺癌治疗： 某些非小细胞肺癌患者的肿瘤中可能存在 EGFR 基因突变。药物吉非替尼（Gefitinib）和厄洛替尼（Erlotinib）是靶向 EGFR 突变的激酶抑制剂，可以抑制突变的

EGFR 信号通路，从而阻止肿瘤生长。这些例子突显了精准医学在癌症治疗中的应用。通过了解患者的分子特征和肿瘤生物学，医生可以选择最适合的靶向药物或制定最合适的治疗性疫苗方案。这种个体化的治疗方法可以提高治疗效果，同时减少不必要的副作用，为癌症患者提供更好的治疗选择。然而，值得注意的是，并非所有肿瘤都有适用的靶向治疗选择。肿瘤的多样性以及个体之间的分子特征差异，使得精准医学在肿瘤治疗中还面临着许多挑战，如寻找更多有效的靶向药物、开发更先进的检测技术以及解决耐药性等问题。这些挑战将继续激励科学家和医疗专业人员不断探索，以更好地应对癌症挑战并提高患者的生存率和生活质量[20]。

五、参考文献

1. Nowell PC. A minute chromosome in human chronic granulocytic leukemia. Science. 1960; 132:497–501
2. Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 1973, 243 (5405), 290–293
3. Nora Heisterkamp, John R. Stephenson, John Groffen, Pamela F. et al. Localization of the c-abl oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. Nature 1983, 306, 239– 242.
4. Ben-Neriah, Y.; Daley, G. Q.; Mes-Masson, A.-M.; Witte, O. N.; Baltimore, D. The chronic myelogenous leukemia-specific P210 protein is the product of the bcr/abl hybrid gene. Science 1986, 233 (4760), 212– 214
5. Druker, B. J.; Tamura, S.; Buchdunger, E.; Ohno, S.; et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nature medicine 1996, 2 (5), 561– 566
6. Druker, B. J.; Guilhot, F.; O’Brien, S. G.; Gathmann, I.; et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N. Engl. J. Med. 2006, 355 (23), 2408–2417
7. Saglio, G.; Kim, D.-W.; Issaragrisil, S.; Le Coutre, P.; et al. Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. N. Engl. J. Med. 2010, 362 (24), 2251– 2259.
8. Liu X and Yuan TY. Oncogene c-abl and chronic myelocytic leukemia. Overseas Medicine Section of Molecular Biology 1987; 9:176-179.
9. Liu X, Yuan TY and Zhu DE. Sublocation of Ph breakpoints on chromosome 22 in Chinese patients suffering from CML. Journal of Hematology 1988, 9:326-330.
10. Leach D.R., Krummel M.F., Allison J.P. Enhancement of Antitumor Immunity by CTLA-4 Blockade. Science. 1996, 271:1734–1736.
11. Iwai Y., Ishida M., Tanaka Y., Okazaki T., Honjo T., Minato N. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. Proc. Natl. Acad. Sci. USA. 2002, 99:12293–12297.
12. Chamoto K., Hatae R., Honjo T. Current issues and perspectives in PD-1 blockade cancer immunotherapy. Int. J. Clin. Oncol. 2020, 25:790–800.

13. Jii Bum Lee, Hye Ryun Kim, and Sang-Jun Ha. Immune Checkpoint Inhibitors in 10 Years: Contribution of Basic Research and Clinical Application in Cancer Immunotherapy. Immune Netw. 2022 Feb;22(1):e2.
14. Andrea Cercek, Melissa Lumish, Jenna Sinopoli, et al. PD-1 Blockade in Mismatch Repair–Deficient, Locally Advanced Rectal Cancer. N Engl J Med 2022, 386:2363-76.
15. Vijay Shankaran, Hiroaki Ikeda, Allen T. Bruce, J. Michael White, et al. IFNγ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature 2001, 410:1107–1111.
16. Mansi Saxena, Sjoerd H. van der Burg, Cornelis J. M. Melief & Nina Bhardwaj. Therapeutic cancer vaccines. Nature Reviews Cancer 2021, 21:360–378
17. Weber J, Boswell W, Smith J, Liu X et al. Phase I trial of intranodal injection of a Melan-A/MART-1 DNA plasmid vaccine in patients with stage IV melanoma. J Immunother. 2008 31:215-23
18. McCarty TM, Liu X, Sun GY, Peralta EA, Diamond DJ and Ellenhorn JDI. Targeting p53 for adoptive T cell immunotherapy. Cancer Research 1998, 58:2601-2605.
19. Liu X, Peralta EA, Ellenhorn JDI and Diamond DJ. Targeting of human p53-overexpressing tumor cells by an HLA A*0201-restricted murine T-cell receptor expressed in Jurkat T cells. Cancer Research 2000, 60:693-701.
20. Williams SC (2015) News feature: capturing cancer's complexity. Proc Natl Acad Sci USA 112(15):4509–4511

Review and Perspective

Comparisons of increased NOergic molecules and neuropeptides between stimuli-evoked axon reflex and stimuli-induced meridian phenomena/tracers along linear pathways resembling meridians: Toward mechanisms of meridian system

Sheng-Xing Ma

Lundquist Institute for Biomedical Innovation at Harbor-University of California at Los Angeles (UCLA) Medical Center and Department of Obstetrics and Gynecology, David Geffen School of Medicine at UCLA and Harbor-UCLA Medical Center, Torrance, CA 90502

Running Head: NO mediated axon reflex in meridian phenomena

The number of text pages of the whole manuscript: 13

Correspondence: Sheng-Xing Ma, M.D., Ph.D.

Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center

1124 W. Carson Street, E6-room 219

Torrance, CA 90502

Tel: (310) 974-9573; E-mail: sxma@lundquist.org

Abstract: Many studies have demonstrated that low electrical resistance of acupuncture points (acupoints)/meridians, the propagated sensation along channels (PSC), and tracers along a linear pathway are the most commonly identified meridian phenomenon. Anatomical studies have found that acupoints exist higher number of nerve fibers, blood vessels, and sweat glands. Nitric oxide (NO) level is elevated in the acupoints/meridians in rats associated with an enhanced expression of NO synthase. NO increases skin norepinephrine (NE) synthesis/release, which contribute to low electrical resistance of acupoints/meridians. NO release is increased predominantly over acupoints by acupuncture or electrical stimulation. These stimuli have been used to elicit axonal reflexes, which are characterized by local release of NO and neuropeptides. Consistently, all experiments of the tracers along a linear pathway have been conducted by injection of a radiotracer or tracer dyes in solution into an acupoit, which produce much strong stimuli and axon reflex. The stimuli-evoked NOergic biomolecules and neuropeptides mediated axon reflex participate in the processes of the PSC, and increase local blood flow, which move radioactive substances or tracer dyes in the cutaneous tissue under linear pathways resembling meridians. We concluded that the stimuli-induced NO signaling molecules and neuropeptides mediated axon reflexes contribute to meridian phenomena including low electrical resistance of acupoints/meridians, the PSC, and tracers along linear pathways resembling meridians.

KEYWORDS propagated sensation along the meridian, tracers, axon reflex, nitric oxide, neuropeptides, biophysical approaches, acupuncture stimuli, acupuncture points, meridian system

One of the 125 major exploration and discovery questions is that “Is there a scientific basis to the Meridian System in traditional Chinese medicine?” as the international frontier, global common needs and gathering foresight released in 2021 by Shanghai Jiao Tong University and "Science" magazine (Shanghai Jiao Tong University and Science, 2021). The Meridian System (channels and collaterals, jingluo) according the description of Traditional Chinese medicine (TCM) is an essential pathway system that unify all parts of the body and deal with physiological regulation and pathological changes of the human body. Acupuncture points (acupoints) are located along the 14 main meridian pathways, and acupoints/meridians have been used in many unconventional medical practices such as acupuncture, electrical acupuncture (EA), transcutaneous electrical nerve stimulation (TENS), moxibustion (heating acupoints), Qi Gong and meditation (Chan, 1984; Wang, 2010; Ma, 2021). It is described that the meridian systems flow “meridian qi” evoking dynamic processes such as energy exchange, movement, and connection from the surface of the body to the internal organs, as well as communication with the universe, environment and our body as a whole. Various diseases are associated with blocked meridians, and acupuncture stimulates acupoints on the body, releasing and easing the “meridian qi” for treating disorders. There has been a widespread and increasing interest in scientific examinations exploring the Meridian System and the use of acupuncture for treatment of disorders over the world. However, the most logical and direct approach to identify a specialized anatomic entity of acupoints and meridians are still lack (Shanghai Jiao Tong University and Science, 2021; Chan, 1984, Wang, 2010; Ma, 2021). The chemicals, structure, functions and mechanisms of the meridians are unclear.

Over the last few decades there has been an increasing evidence from biophysical, physiological, and biochemical approaches implicates some of specificity for the acupoints in the body around the world. There are various studies from biophysical approaches dealing with specificity for the acupoints and/or median lines in the body (Ma, 2021 and 2023, Wang, 2010; Zhu and Hao, 1989; Yang and Han, 2015). Although no well-defined anatomic entity is available, an abundance of information has now accumulated several meridian phenomenon. It is generally accepted that the most commonly observed and internationally identified meridian phenomenon are: 1) the characteristics of low electrical resistance and high electric conductance of acupoints/meridians; 2) the propagated sensation along channels; and 3) radioactive or fluorescent tracers along a linear path resembling acupoints and meridians (Ma, 2021 and 2023, Wang, 2010; Zhu and Hao, 1989; Yang and Han, 2015). In this paper, the evidence and understanding of the biophysical approaches of acupoints/meridians have been summarized with an emphasis on recent development of stimuli-evoked nitric oxide (NO) release through the axon reflex and neuropeptides mediating meridian (Jingluo) phenomenon.

NOergic signaling molecules and neuropeptides contribute to low electrical resistance properties of acupoints

The acupoints have been studies by various biophysical, anatomic, and biochemical methods, and viewed as points of distinct electrical characteristics. These experimental results provide preliminary findings using biophysical approaches studying electrical resistance and shed some lights into the results of NOergic signaling molecules and neuropeptides involved in skin electrical conductance, which have been summarized in a recent review paper (Ma, 2021). Studies from several research groups have confirmed that most acupoints in both humans and animals correspond to high electrical conductance and low skin resistance points on the body surface along the meridians (Roll 1975; Fraden 1979; Zhu, 1988; Ahn et al., 2008; Ma, 2021). Prior studies have established that skin electrical resistance depends upon the activity of the sympathetic nervous systems and that stimulation of sympathetic pathways lowers the skin resistance levels (Kunimoto, 1995; Chen and Ma, 2005; Chen et al., 2006). Recent evidence has demonstrated that L-arginine-derived NO in the skin contributes to electrical conductance of acupoints through the increases in NE production (Chen and Ma, 2005; Chen et al., 2006). The experimental results in rats have shown that L-arginine-derived NO synthesis and noradrenergic transmission modify skin electric conductance, which contributes to low resistance characteristics of acupoints and meridians (Chen and Ma, 2005). NE synthesis/release was consistently increased in three meridians' acupoints with high skin electric currents. Enhanced 3H-NE synthesis/release in acupoints were facilitated by presence of an exogenous NO donor and inhibited by an inhibitor of NO synthesis (Chen et al., 2006). These results suggesting that peripheral NE synthesis/release from the sympathetic nervous system in acupoints is mediated by NO, and that the NO-NE system plays an important role in the low resistance characteristics of acupoints.

In spite of experiments documenting a biophysical basis of skin electrical conductance and the increasing studies of skin low resistance acupoints, systematic research with an improved equipment and carefully research design to avoid the influencing factors (skin dryness, skin thickness, size of the sensing electrode, pressure applied on the electrode, interelectrode distance, room temperature, and humidity) are requested to be conducted for a definite answer in this field (Pearson et al., 2007; Ahn et al., 2008; Ma, 2021). Other substances, including substance P (SP) and calcitonin gene-related peptide

(CGRP), induce vasodilation and plasma extravasation to increase skin hydration, resulting in the development of high conductance at acupoints, and the high conductance at acupoints, especially vascular leakage following local release of SP and CGRP during neurogenic inflammation (Kim et al., 2017; Fan et al., 2018). Double immunostaining of transient receptor potential vanilloid type-1 (TRPV1) receptor and neuronal nitric oxide synthase (nNOS) revealed co-localization of TRPV1 and nNOS in both subepidermal nerve fibers and in dermal connective tissue cells (Ibrahim et al., 2011). A high expression of TRPV1 endowed with nNOS in subepidermal nerve fibers exist in the acupoints and the expression is increased by EA (Ji et al., 2013). These results are consistent with the studies showed that acupuncture significantly increased components of the TRPV1-related signaling pathway in mice (Chen et al., 2018), and modified acupuncture-induced reflex excitatory cardiovascular responses in humans (Guo et al., 2018). The experimental evidences suggest that the higher expression of TRPV1 in the subepidermal nerve fibers and its upregulation after EA stimulation may play a key role in mediating the transduction of EA signals to the CNS, and its expression in the subepidermal connective tissue cells may play a role in conducting the local effect of the EA. Although several independent groups have reported various biomolecules, NO, NE, TRPV1, CGRP and SP, which can be classified as NOergic signaling molecules (NO-NE/TRPV1) and neuropeptides (CGRP and SP) involved in the skin low resistance at acupoints, the signal transduction pathways and mechanisms of multiple molecules in the skin electrical conductance and low resistance at acupoints are still unclear. Systematic research with an improved equipment and carefully research design to avoid the influencing factors are requested for a definite answer in this field. However, the results from anatomical and biochemical studies consistently show that acupoints exist higher levels of nervous components, and NOergic signaling molecules (NO-NE/TRPV1) and neuropeptides involved in the skin low resistance at acupoints.

Roles of NOergic molecules and neuropeptides mediated axon reflex in propagated sensation along the meridian

The stimuli-evoked NOergic molecules mediated axon reflex and their roles in propagated sensation along the channels/meridian (PSC) have been reviewed in a recent paper (Ma, 2023). Over the last few decades, PSC and the latent PSC (LPSC) over the body surface have been internationally identified in many human subjects, which are consistent with the courses of the classical meridians (Zhu and Hao, 1989; Xu, 2013; Ma, 2023). It appears that PSC is one of the most commonly observed and internationally identified meridian phenomenon over the body surface. The phenomenon occurs most frequently following various stimulation to an acupoint using acupuncture (Xu, 2013, Zhu and Hao, 1989; Razavy et al., 2017; Chen, 2006), EA or TENS (Xu, 2013; Zhu and Hao, 1989; Razavy et al;, 2017; Chen, 2006), acupressure with qigong or meditation (Xu, 2013; Zhu and Hao, 1989), and moxibustion including heat and laser (Xu, 2013; Zhu and Hao, 1989). It is reported that low electrical resistance, PSC, and LPSC are present not only in the acupoints, but also over the entire lines (about 1.0 mm in width) of the meridians, which are described in traditional charts (Ma, 2023; Zhu and Hao, 1989; Xu, 2013). Although there has been increasing evidence reporting that PSC and LPSC along a linear path exist in the acupoints and/or medians in human studies, the most logical and direct approaches to identify biochemical and morphological evidences related to specialized anatomic entities of acupoints and meridians are still unavailable, especially limited by technology used or studied on humans.

Previous anatomical studies have demonstrated that acupoints/meridians exist higher number of nerve fibers/trunks, blood vessels, hair follicles, and sweat glands (Wang et al., 2010; Zhu and Hao, 1989; Monteiro-Riviere et al., 1986). Animal studies showed that NO level and NO synthase expression are consistently increased in the subcutaneous tissue of acupoints and meridian regions compared to non-meridian areas (Ma, 2003). NO level is elevated by EA in the acupoints/meridians in rats associated with an enhanced expression of NO synthase endowed with TRPV1 (Ibrahim et al., 2011; Ji et al., 2013). Dermal microdialysis in humans reported that NO-cGMP releases are increased by EA in the subcutaneous tissue of acupoints (Lim and Ma, 2009). Consistently, several human studies demonstrated that TENS, MA, and electrical heat produce an elevation of NO release predominantly over acupoints (Ma et al., 2015; 2017; Kimura et al., 2013). The axon reflex can be evoked by a wide range of physiologic stimuli such as heat, cold, mechanical and electrical stimuli (Steinhoff et al., 2003; Hassan et al., 1986). The response to local skin stimulus is characterized by its biphasic nature - the first peak is due to C-fiber nociceptor-mediated axon reflex (Schmelz et al., 2000; Magerl and Treede, 1996), which results in vasodilatation through the local release of CGRP (Schmelz et al., 2000; Wallengren et al., 1987) and neuropeptide Y (NPY) (Hodges et al., 2008; Kellogg et al., 1999). The second increase in skin blood flow, characterized by the plateau, has been demonstrated to be largely dependent (about 70%) on NO (Magerl and Treede, 1996), which are characterized by local release of NO and calcitonin gene related peptide mediated by activation of TRPV1 (Ma et al., 2015; 2017; Kimura et al., 2013). Interestedly, the same stimuli, both heat, mechanical and electrical stimuli-evoked axon reflexes, and electrical heat, TENS, and EA-evoked PSC/LPSC, induce NO signaling molecules which involve in the biochemical-physiological processes of axon reflexes and PSC/LPSC, leading to the suggestion that the similar driver induced NO-mediated axon reflexes participate in the somatosensory signal transduction of PSC/LPSC, the important phenomena of meridians induced by the stimuli. Although more sophisticated approaches are requested to address biochemical-physiological interactions in the processes in this field, the results from animal and human studies consistently suggest that stimuli-evoked NO signaling molecules and neuropeptides (CGRP, NPY, SP) mediated axon reflexes contribute to the development and processes of PSC/LPSC, which are evoked by mechanical, electrical, or thermal stimuli.

Studies of tracers along a linear path resembling acupoints and meridians

Some investigators have attempted to explain meridians on the basis of established scientific principles, by postulating biological systems other than those of conventional anatomy and physiology. However, there is little or no scientific evidence currently to support either describing meridians/acupoints as distinct anatomical entities or the existence of alternative biological systems to explain the systems (Ma, 2021 and 2023, Wang, 2010; Zhu and Hao, 1989). Several groups of the investigators have identified the slow-migrating pathways of radioactive tracers along a linear path resembling acupoints and meridians (Kovacs et al., 1992; 1993; 1996; 2000; Wu et al., 1990; 1994; Chen et al, 1994). Early studies (Tiberiu et al., 1981) reported that meridians can be made visible by the shape of radioactive tracer pathway following injection of a radiotracer solution into an acupoit. Migrating pathways of radioactive tracers following injections of isotope (9mTc 300-500 □ci, about 30-70 □l solution) into acupoints are in straight correspondence with classical course of meridians but are quite different of lymphathic vessels in patients and healthy subjects (Meng et al., 1987). The appearance of radioactive paths have been also confirmed after the injection of a radiotracer (Technetium 99m in a volume of 0.05 ml) at acupoints in healthy humans by Lazorthes et al. (1990). A linear radioactive pathway which traveled at approximately 2.5 cm/minute over 11 cm distance hypodermically was induced by (99mTc in a volume of 0.125 ml) sodium pertechnetate injections but not produced by other three nuclear tracers, ^{201}Tl-thallous chloride in a volume of 0.3 ml, (^{131}I) iodine-sodium in a volume of 0.05 ml, and (^{99m}Tc in a volume of 0.125 ml) rhenium sulfide injected into both low-electrical resistance acupoints and adjacent controls on the legs of adult dogs (Kovacs et al., 1992). Direct injections of the dyes into blood vessels led to a much more instantaneous appearance and subsequent disappearance of the trajectory in 10 seconds. However, (^{99m}Tc) rhenium sulfate, which has a predilection for lymphatics, was detected at axillary lymph nodes after 90 minutes in both acupoints and controls, did not have any linear pathway. For the reasons, the authors concluded that veins or lymphatic vessels could not have been responsible for these trajectories (Kovacs et al., 1992; 1993; 1996). Linear and rapid migration of ^{99m}Tc-pertechnetate was not altered or prevented by incisions that did not intersect the migration pathway, but was not observed when the incision was left uncovered, filled with petroleum jelly, or with a solid silicone sheet (Kovacs et al., 2000). Therefore, these investigators concluded that capillary electrophoresis is suggested to explain the hypodermic diffusion of inert particles through specific and constant linear pathways (Kovacs et al., 1992; 1993; 1996; Kovacs et al., 2000).

In addition, previous studies showed that fluorescein injected into acupoints produced similar visualization of interstitial channels along meridians in rats and mini-pigs (Zhang et al., 2008; 2015; Xiong et al., 2020; Li et al., 2021). Several studies reported that injection of tracer dyes at acupoints led to linear migration of the dye along a path aligning with the meridians (Li et al., 2012; Zhang et al., 2008;2015; Li et al., 2021). Low hydraulic resistance channels were found very close to low impedance points along meridians, which were imaged through injection of Alcian blue and isotopic tracing (Zhang et al., 2008; 2015). Animal studies reported that visualization of interstitial channels along meridians with average length of the migration tracks was 5.13 cm following injections of 0.2, 0.4, 0.6, 0.8, and 1 mL of sodium fluorescein into acupoints in mini pigs (Xiong et al., 2020). Similar results were also observed in rats (Gu et al., 2020). Other experiments have identified “perivascular-like-spaces” along the meridians histologically (Ma et al., 2008), which are not directly attributable to lymphatics or blood vessels (Kovacs et al., 1992; 1993; 1996; Ma et al., 2003). Following injecting a small amount of paramagnetic contrast or gadopentetate acid dimeglumine into acupoints, subcutaneous pathways have been visualized in humans, which have a characteristic “puncture-resistant” appearance, indicating a non-tubular structure, not consistent with lymphatic or blood vessels (Li et al., 2008). Fluorescein sodium is found in the pathways of hind limbs and segments of the small intestines after injecting 0.2∼1 mL of fluorescein sodium in acupoint (KI3) of healthy rabbits (Li et al., 2012). Consistently, recent research reported that injection of 0.1 mL of fluorescent dyes into the acupoint PC6 of pericardium meridian, 15 (79%) of fluorescein sodium injection (but not indocyanine green) were associated with slow diffusion of the dye proximally along a path matching closely with the Pericardium meridian in humans (Li et al., 2021). The observed migratory speed, path trajectory, and selective association with Technetium-99 (99mTcO4, but not other nuclear dyes) led authors to speculate that paths were attributed to a yet undescribed channel with low hydraulic resistance through the extracellular matrix

(Zhang et al., 2008; 2015; Xiong et al., 2020; Gu et al., 2020), and not to vein or lymphatics (Kovacs et al., 1992; 1993; 1996).

Stimuli-evoked NOergic molecules and neuropeptides mediated axon reflex contribute to tracers along a linear path resembling acupoints and meridians

In order to examine the radioactive or fluorescent tracer pathway along a linear path resembling acupoints and meridians, all experiments have been conducted by injection of a radiotracer solution or tracer dyes in a volume of either 0.05, 0.1, 0.2, 0.4, 0.6, 0.8 or 1 mL into an acupoit (Kovacs et al., 1992; 1993; 1996; Xiong et al., 2020; Li et al., 2021). Needle injection is a mechanical stimulus and the injected solution is an additional extension stimulus, both of them produce stimuli-evoked axonal reflexes (Izumi et al.; 1991; Sharara et al., 1995). Acupoint injection of solution or drugs has a better effect than traditional acupuncture achieving the purpose of treatment by much stronger stimulating the acupoints (Huang et al., 2019; Sha et al., 2016; Park et al., 2017). Intra-dermal injection of hypertonic saline produces an axon reflex flare response involved neural and vascular mechanisms, and the axon reflex vasodilatation measured by a blood flow value which is inhibited by lignocaine comparably to its effects on histamine and substance P (Izumi et al.; 1991; Sharara et al., 1995). Axonal reflexes have been defined as a response elicited by mechanical, electrical, thermal, or pharmacological stimuli (Steinhoff et al., 2003; Hassan et al., 1986), and are characterized by local release of NO (Magerl and Treede, 1996) and neuropeptides (CGRP mediated by activation of TRPV1 and NPY) (Schmelz et al., 2000; Wallengren et al., 1987; Hodges et al., 2008; Kellogg et al., 1999). Moreover, our results show that local NO release is consistently induced by MA, EA, TENS, and electrical heat (Ma et al., 2015;2017; Kimura et al., 2013). These stimuli, MA, EA, TENS, and electrical heat, have been demonstrated to evoke axon reflexes in various studies (Steinhoff et al., 2003; Hassan et al., 1986). Peripheral axon reflexes play an essential role in the regulation of skin blood flow, and activation of axon reflexes cause vasodilatation and increase local blood flow (Schmelz et al., 2000; Magerl and Treede, 1996; Hodges et al., 2008; Kellogg et al., 1999). It appears that injection of a radiotracer solution or tracer dyes in a volume of solution cause much stronger stimuli than acupuncture stimulation at local acupoints, which produced stimuli-evoked axon reflexes and elevations of NO and neuropeptides levels. The stimuli-evoked axon reflexes and NO signaling molecules/neuropeptides allows for a flare of blood flow, leading to moving radioactive substances or tracer dyes in the skin and subcutaneous tissue under axon reflex pathways. Therefore, the radioactive or fluorescent tracer pathway along a linear path represent a flare of blood flow following stimuli-evoked axon reflexes.

Regarding the radioactive or fluorescent tracer pathway along a linear path resembling acupoints and meridians, it has been concluded “As for the meridians themselves, in 1987, scientists who analyzed the movement of radioisotopes in the body concluded that meridians are actually spaces within the body, not structures. More recently, a 2020 publication in the journal research reached the same conclusion (Shanghai Jiao Tong University and Science, 2021).” The mechanisms of the linear and rapid migration of 99mTc-pertechnetate have been examined by various methods including skin incisions, filled with petroleum jelly, or with a solid silicone sheet (Kovacs et al., 2000), the authors concluded that veins or lymphatic vessels could not have been responsible for linear and rapid migration of 99mTc-pertechnetate (Kovacs et al., 1992; 1993; 1996). The conclusion is consistent with recent fluorescent dye tracer, which showed that both ultrasound imaging and vein locating device did not reveal any corresponding vessels (arterial or venous) or lymphatics at the visualized tracer pathway but demonstrated correlations with intermuscular fascia (Li et al., 2021). Kovacs et al. (2000) indicated that capillary electrophoresis is suggested to explain the hypodermic diffusion of inert particles through specific and constant linear pathways. Our analyses agree with this suggestion and further point out that the stimuli-evoked axon reflex and NO signaling molecules/ neuropeptides produce a flare of blood flow, leading to moving radioactive substances or tracer dyes in the skin and subcutaneous tissue under a linear path resembling acupoints and meridians. NO level and NO synthase expression are consistently increased in the subcutaneous tissue of acupoints and meridian regions compared to non-meridian areas in rats (Ma, 2003) associated with an enhanced expression of NO synthase endowed with TRPV1 (Ibrahim et al., 2011; Ji et al., 2013). Several studies have demonstrated that MA, EA, TENS, and electrical heat-induced elevations of NO release over skin regions with a high level at acupoints (Ma et al., 2015; 2017; Kimura et al., 2013). Axon reflexes-mediated NO and CGRP releases are also induced by the same stimuli (Magerl and Treede, 1996; Schmelz et al., 2000; Wallengren et al., 1987) and acupoint injection of solution (Izumi et al.; 1991; Sharara et al., 1995). Cutaneous small nerve fibers encompass unmyelinated C-fibers and thinly myelinated Aδ-fibers, which innervate dermal vessels (vasomotor fibers), sweat glands (sudomotor fibers), and hair follicles (pilomotor fibers) participate axon reflex (Hijazi et al., 2020; (Steinhoff et al., 2003; Hassan et al., 1986)). Anatomical studies have demonstrated that acupoints/meridians exist higher number of nerve fibers/trunks, blood vessels, hair follicles, and sweat glands (Wang et al., 2010; Zhu and Hao, 1989; Monteiro-Riviere et al., 1986). It appears that the higher number of small nerve fibers and blood vessels with rich NO synthase and neuropeptides, and sweat glands contribute to higher NO related molecules/neuropeptides and axon reflex-evoked blood flow, leading to moving radioactive substances or tracer dyes in the skin and subcutaneous tissue under a linear path resembling acupoints and meridians. Therefore, these analyses support the conclusion "The movement of radioisotopes in the body concluded that meridians are actually spaces within the body, but do not agree "not structures" (Shanghai Jiao Tong University and Science, 2021)". This review suggests that the higher number of small nerve fibers and blood vessels with rich NO synthase, NE, TRPV1, and neuropeptides, and sweat glands in the acupoints/meridians serve as structures, which contribute to biophysical characteristics of low electrical resistance and PSC as well as tracers along a linear pathways resembling meridians.

Conclusion

Over the last few decades, there are various studies from biophysical approaches dealing the phenomena of meridians. An abundance of studies from different international groups have demonstrated that acupoints possess characteristics of low electrical resistance and high electric conductance, the propagated sensation along channels (PSC), and tracers along a linear pathways resembling meridians are the most commonly observed and identified meridian phenomenon. In many texts, although the meridian phenomenon to be genuine evidences for the functional existence of acupoints/meridians, there is a lack of objective and systematic experimental studies to support biochemical and morphological basis of the meridians. Skin electrical resistance depends upon the activity of the sympathetic nervous systems. Recent evidence shows that NO concentrations are enhanced in skin acupoints/meridians, and L-arginine-derived NO synthesis and noradrenergic transmission modify skin electric conductance, which contributes to low resistance characteristics of acupoints and meridians (Chen and Ma, 2005). 3H-NE synthesis/release was enhanced in acupoints and facilitated by presence of an exogenous NO donor and inhibited by an inhibitor of NO synthesis (Chen et al., 2006). NOergic signaling molecules (NO-NE/TRPV1) and neuropeptides play important roles in mediating the skin conductance responses to electrical stimulation and contribute to low electrical resistance and high electric conductance of acupoints/meridians.

The PSC and the latent PSC (LPSC) over the body surface, which are consistent with the courses of the classical meridians, have been internationally elicited by TENS, EA or electrical heat stimulation of acupoints in many human subjects (Ma, 2023, Zhu and Hao, 1989; Xu, 2013). A systematic review found that 63,228 subjects produced an incidence of 12-24% PSC during pulse electrical stimulation of the acupoints. Animal studies showed that NO level is elevated by EA in the acupoints/meridians in rats associated with an enhanced expression of NO synthase endowed with TRPV1. Dermal microdialysis in humans reported that NO-cGMP releases are increased by EA in the subcutaneous tissue of acupoints. Several studies demonstrated that TENS, MA, and electrical heat produce an elevation of NO release predominantly over acupoints. Consistently, these stimuli, TENS, electrical heat, and EA have been used to elicit axonal reflexes, which are characterized by local release of NO and calcitonin gene related peptide mediated by activation of TRPV1. NO signaling molecules and neuropeptides (CGRP, SP, and NPY) involve in the stimuli-evoked axon reflexes, and the same driver elicited biomolecules-mediated axon reflex may also participate in the somatosensory signal transduction and processes of PSC/LPSC.

Moreover, all experiments of the tracers along a linear pathways resembling meridians have been conducted by injection of a radiotracer solution or tracer dyes in a volume of solution into an acupoit. Acupoint solution injections produce much stronger mechanical stimuli, which cause axon reflex. Although more sophisticated approaches are requested to address biochemical-physiological interactions in the processes in this fields, the results from animal and human studies consistently suggest that the stimuli-evoked axon reflex and NOergic biomolecules increase local blood flow, which move radioactive substances or tracer dyes in the skin and subcutaneous tissue under a linear path resembling acupoints and meridians, the important phenomena of meridians induced by the stimuli.

Together, the stimuli-induced NO signaling molecules and axon reflexes participate in the biochemical-physiological processes of all meridian phenomenon, low electrical resistance of acupoints/meridians, the PSC/LPSC, and tracers along a linear pathways resembling meridians, leading to the suggestion that NO signaling molecules are important meridian biomolecules and contributes to structure, functions, and pathways of the meridian system. Anatomical studies have demonstrated that acupoints/meridians exist higher number of nerve fibers/trunks, blood vessels, hair follicles, and sweat glands. The higher number of small nerve fibers, blood vessels with rich endothelia NO synthase, TRPV1 and neuropeptides, and sweat glands in the cutaneous tissues serve as structures for NO signaling molecules and neuropeptides mediated axon reflexes resulting in low electrical resistance, PSC, and tracers along a linear pathways resembling meridians. The increased interest in the meridian phenomenon has led to an open-minded attitude towards understanding this system, which is fundamental important to establish the valid aspects of scientific basis of Chinese medicine mechanisms and therapies.

ACKNOWLEDGMENTS

This project was made possible by NIH Grant (AT002478, AT004620, and AT004504) from the National Center for Complementary and Integrative Medicine (to SXM).

Conflict of Interest

The author has stated explicitly that there are no conflicts of interest in connection with this article.

References

Shanghai Jiao Tong University, Science. (2021). 125 Questions: Exploration and discovery | Science | AAAS. https://www.science.org/content/resource/125-questions-exploration-and-discovery

Chan, S.H. (1984). What is being stimulated in acupuncture: evaluation of the existence of a specific substrate? Neurosci Biobehav Rev 8, 25-33.

Wang, G.J., Ayati, M.H., Zhang, W.B. (2010). Meridian studies in China: a systematic review. J Acupunct Meridian Stud 3, 1-9.

Ma, S.X. (2021). Low electrical resistance properties of acupoints: Roles of NOergic signaling molecules and neuropeptides in skin electrical conductance. Chin J Integr Med 27, 563-569.

Ma, S.X. (2023). Stimuli-evoked NOergic molecules through the dorsal medulla-thalamic pathways may mediate propagated sensation along the meridian. Int J Mol Sci (in press).

Zhu, Z.X., Hao, J.K. (1989). Acupuncture Meridian Biophysics-Scientific Verification of the First Great Invention of China, Beijing Press, Beijing.

Yang, M.N., Han J. X. (2015). Review and analysis on the meridian research of China over the past sixty years. Chin J Integr Med 21, 394-400.

Voll, R. (1975). Twenty years of electroacupuncture diagnosis in Germany. A progress report. Am J Acup 3: 7-17.

Fraden, J. (1979). Active acupuncture point impedance and potential measurements. Am J Acup 7: 137-144.

Zhu, Z.X. (1988). The advances and prospect in physiological and biophysical approaches of acupuncture meridian system. Acup Res 2: 81-89.

Ahn, A.C., Colbert, A.P., Anderson, B.J., Martinsen, O.G., Hammerschlag, R., Cina, S., Wayne, P.M., Langevin, H.M. (2008). Electrical properties of acupuncture points and meridians: a systematic review. Bioelectromagnetics 29:245–256.

Kunimoto, M., Kirno, K., Elam, M., Karlsson, T., Wallin, B.G. (1992). Non-linearity of skin resistance response to intraneural electrical stimulation of sudomotor nerves. Acta Physiol Scand 146: 385-392.

Chen, J.X., Ma, S.X. (2005): Effects of nitric oxide and noradrenergic function on skin electric resistance of acupoints and meridians. J Altern Complement Med 11:423-431.

Chen, J.X., Ibe, B.O., Ma, S.X. (2006). Nitric oxide on modulation of norepinephrine production in acupuncture points. Life Sci 79:2157-2164.

Pearson, S., Colbert, A.P., Mcnames J, Baumgartner, M., Hammerschlag, R. (2007). Electrical skin impedance at acupuncture points. J Altern Complement Med 13 409–418.

Kim DH, Ryu Y, Hahm DH, Sohn BY, Shim I, Kwon OS, et al: Acupuncture points can be identified as cutaneous neurogenic inflammatory spots. Sci Rep 2017; 7:15214.

Fan, Y., Kim, D-H, Ryu, Y., Chang, S., Lee, B.H., Yang, C.H., Kim, H.Y. (2018). Neuropeptides SP and CGRP underlie the electrical properties of acupoints. Front Neurosci doi: 10.3389/fnins.2018.00907.

Ibrahim, T.S., Chen, M.L., Ma, S.X. (2011). TRPV1 expression in acupuncture points: Response to electroacupuncture stimulation. J Chem Neuroanat 41:129-136.

Ji, B., Hu, J., Ma, S.X. (2013). Effects of electroacupuncture Zusanli (ST36) on food intake and expression of POMC and TRPV1 through afferents-medulla pathway in Obese Prone Rats. Peptides 39:188-194.

Chen, H.C., Chen, M.Y., Hsieh, C.L., Wu S.Y., Hsu H.C., Lin Y.W. (2018). TRPV1 is a responding channel for acupuncture manipulation in mice peripheral and central nerve system. Cell Physiol Biochem 49:1813-1824.

Guo, Z.L., Fu, L.W., Su, H.F., Tjen-A-Looi S.C., Longhurst J.C. (2018). Role of TRPV1 in acupuncture modulation of reflex excitatory cardiovascular responses. Am J Physiol Regul Integr Comp Physiol 2018,314:R655-R666.

Xu, J., Zheng, S., Pan, X., Zhu, X., Hu, X. (2013). The existence of propagated sensation along the meridian proved by neuroelectrophysiology. Neural Regen Res 8, 2633-2640.

Razavy, S., Gadau, M., Zhang, S.P., Wang, F.C., Bangrazi, S., Berle, C., Harahap, M., Li, T., Li, W.H., Zaslawski, C. (2017). Investigation of the phenomenon of propagated sensation along the channels in the upper limb following administration of acupuncture and mock laser.

J Acupunct Meridian Stud. 10, 307-316.

Chen, E. (2006). The meridian theory. J Accord Integr Med 2, 1-6.

Monteiro-Riviere, N.A., Hwang, Y.C., Stromberg, M.W. (1981). Light microscopic morphology of low resistance skin points in the guinea pig. Am J Chin Med 9, 155-163.

Ma, S.X. (2003). Enhanced nitric oxide concentrations and expression of nitric oxide synthase in acupuncture points/meridians. J Altern Complement Med 9, 207-215.

Lim, N., Ma, S.X. (2009). Responses of nitric oxide-cGMP releases in acupuncture point to electroacupuncture in human skin in vivo using dermal microdialysis. Microcirculation 16, 434-443.

Ma, S.X., Mayer, E., Lee, P., Li, X.Y., Gao, E.Z. (2015). Transcutaneous electrical nerve stimulation increased nitric oxide-cyclic cGMP release biocaptured over skin surface of the pericardium meridian and acupuncture points in humans. Acup & Electro-Therapeutics Res INT J 40, 73-86.

Ma, S.X., Lee, P., Anderson, T.L., Li, X.Y., Jiang, I. (2017). Response of local nitric oxide release to manual acupuncture and electrical heat in humans: Effects of reinforcement methods. Evid Based Complement Alternat Med. 2017, 1-8.

Kimura, K., Takeuchi, H., Yuri, K., Wakayama, I. (2013). Effects of nitric oxide synthase inhibition on cutaneous vasodilation in response to acupuncture stimulation in humans. Acupunct Med 31, 74-80.

Steinhoff, M., Ständer, S., Seeliger, S., Ansel, J.C., Schmelz, M., Luger, T. (2003). Modern aspects of cutaneous neurogenic inflammation. Arch Dermatol 139, 1479–1488.

Hassan, A.A., Rayman, G., Tooke, J.E (1986). Effect of indirect heating on the postural control of skin blood flow in the human foot. Clin Sci 70, 577–582.

Schmelz, M., Michael, K., Weidner, C., Schmidt, R., Torebjörk, H.E., Handwerker, H.O. (2000). Which nerve fibers mediate the axon reflex flare in human skin? Neuroreport 11, 645-648.

Wallengren, J., Håkanson, R. (1987): Effects of substance P, neurokinin A and calcitonin gene-related peptide in human skin and their involvement in sensory nerve-mediated responses. Eur J Pharmacol 143, 267-273.

Hodges, G.J., Kosiba, W.A., Zhao, K., Johnson, J.M. (2008). The involvement of norepinephrine, neuropeptide Y, and nitric oxide in the cutaneous vasodilator response to local heating in humans. J Appl Physiol 105, 233-240.

Kellogg, D.L. Jr, Liu, Y., Kosiba, I.F., Odonnell, D. (1999). Role of nitric oxide in the vascular effects of local warming of the skin in humans. J Appl Physiol 86, 1185-1190.

Magerl, W., Treede, R.D. (1996). Heat-evoked vasodilatation in human hairy skin: axon reflexes due to low-level activity of nociceptive afferents. J Physiol 497, 837-848.

Tiberiu, R., Gheorge, G., Popescu, S.I. (1981). Do meridians of acupuncture exist? A radioactive tracer study of the bladder meridian. Am J Acupunct 9, 251-256.

Meng, J.B., Gao, H.H., Wang, P., Tian, J.H., Liu, Y.L. (1987). Primary approach to visualize the courses of channels by use of isotope. Acu Res 1:77-81.

Lazorthes, Y., Esquerré, J.P., Simon, J, Guiraud, G., Guiraud, R. (1990). Acupuncture meridians and radiotracers. Pain 40:109-112.

Kovacs, F.M., Gotzens, V., García, A., García, F., Mufraggi, N., Prandi, D., Setoain, J., Román, F.S. (1992). Experimental study on radioactive pathways of hypodermically injected technetium-99m, J Nucl Med 33, 403-407.

Kovacs F.M., Gotzens V., García A., García, F., Mufraggi, N., Prandi, D., Setoain, J., Román, F.S. (1993). Kinetics of hypodermically injected technetium-99m and correlation with cutaneous structures: an experimental study in dogs, Eur J Nucl Med 20, 585-590.

Kovacs F.M., Gotzens V., García A., García, F., Mufraggi, N., Prandi, D., Setoain, J., Román, F.S. (1996). Radioactive pathways of hypodermically injected technetium-99m, Am J Chin Med 24,101-102.

Wu, C.C., Jong, S.B., Lin, C.C., Chen, M.F., Chen, J.R., Chung, C (1990). Subcutaneous injection of 99mTc pertechnetate at acupuncture points K-3 and B-60. Radioisotopes 39, 261-263.

Wu, C.C., Chen, M.F., Lin, C.C (1994). Absorption of subcutaneous injection of Tc-99m pertechnetate via acupuncture points and non-acupuncture points. Am J Chin Med 22, 111-118.

Chen, M.F., Wu, C.C., Jong, S.B., Lin, C.C (1993). Differences in acupuncture point SP-10 and non-acupuncture point following subcutaneous injection of Tc-99m pertechnetate. Am J Chin Med 21, 221-229.

Kovacs, F.M., Garcia, A., Mufraggi, N., García, F., Pavía, J., Prandi D., Gotzens, V., Giralt, I., Piera, C., Setoain, J. (2000). Migration pathways of hypodermically injected technetium-99m in dogs, Eur Radiol 10, 1019-1025.

Zhang, W.B., Tian, Y.Y., Li, H., Tian, J.H., Luo, M.F., Xu, F. L., Wang, G.J., Huang, T., Xu Y. H., Wang, R. H. (2008). A discovery of low hydraulic resistance channel along meridians, J Acupunct Meridian Stud, 20-28.

Zhang, W.B., Wang, G.J., Fuxe, K (2015). Classic and modern meridian studies: A review of low hydraulic resistance channels along meridians and their relevance for therapeutic effects in traditional Chinese medicine, Evid Based Complement Alternat Med DOI: 10.1155/2015/410979

Xiong, F., Song, X., Jia, S. (2020). Preliminary observation of the migration of sodium fluorescein along meridians in the limbs of mini-pigs. Scientia Sinica Vitae 50, doi: 10.1360/SSV-2020-0144.

Gu, X., Wang, Y. P., Wang, G. J, Song X.J., Jia S.Y., Li H.Y., Ye F.Y., Xiong F., Wei Bo Zhang W.B. (2020). In vivo display of low hydraulic resistance channels along conception vessel in rats by photofluorography (in Chinese). Acup Res 45: 227–232.

Ma, W., Tong, H., Xu, W., Hu, J., Liu, N., Li, H.Y., Cao, L.X. (2003) Perivascular spaces: possible anatomical substrate for the meridian. J Altern Complement Med 9: 851–859.

Li, H.Y., Yang, J.F., Chen, M., Xu, L., Wang, W.C., Wang, F., Tong, J.B., Wang, C.Y. (2008) Visualized regional hypodermic migration channels of interstitial fluid in human beings – Are these ancient meridians? J Altern Complement Med 6: 621–628.

Li, H.Y., Chen, M., Yang, J.F., Yang, C.Q., Xu, L., Wang, F., Tong, J.B., Lv, Y, Suonan, C. (2012). Fluid flow along venous adventitia in rabbits: is it a potential drainage system complementary to vascular circulations? PLoS One DOI: 10.1371/journal.pone.0041395

Li, T., Tang, B.Q., Zhang, W.B., Zhao, M.Y., Hu, Q.C., Ahn, A. (2021). In-vivo visualization of pericardium meridian with fluorescent dyes. Evid Based Complement Alternat Med 2021:1-8.

Huang, R., Xu, S.H., Li, D.P., Xie, X,W. (2019). Acupoint injection treatment for primary osteoporosis: A systematic review and meta-analysis of randomized controlled trials protocol. Medicine (Baltimore) 98, e16735.

Sha, T., Gao, L.L., Zhang, C.H., Zheng, J.G., Z H Meng, Z.H. (2016). An update on acupuncture point injection. QJM 109:639–641.

Park, K.M., Cho, T.H. (2017). Therapeutic effect of acupuncture point injection with placental extract in knee osteoarthritis. J Integr Med 15:135–41.

Izumi, H., Karita, K. (1991). Axon reflex vasodilatation in human skin measured by a laser Doppler technique. Jpn J Physiol 41:693-702.

Sharara, A.M., Higham, M.A., Iredale, M.J., Ind, P.W. (1995). Intradermal actions of hypertonic saline involve neural and vascular mechanisms. Br J Clin Pharmacol 40:98-100.

Hijazi, M.M., Buchmann, S.J., Sedghi, A., Illigens, B.M., Reichmann, H., Schackert, G., Siepmann, T., Hijazi, M.M., et al (2020). Assessment of cutaneous axon-reflex responses to evaluate functional integrity of autonomic small nerve fibers. Neurol Sci 41:1685-1696.

Waveguide Effects of a Multi-Layered Fault Zone – Retrospective of the 1992 M7.4 Landers earthquake

Yong-Gang Li
Department of Earth Sciences, University of Southern California (Email: ygli@usc.edu)

Abstract

We recently carried out numerical investigations of trapping efficiency for a depth-dependent layered low-velocity waveguide (LVWG) in aid to further study the internal damage structure of actual fault zones at seismogenic depth. The fault-zone trapped waves (FZTWs) produced by this layered LVWG show multiple amplitude peaks in the long-duration post-S coda, consistent with the FZTWs recorded in our previous experiments conducted at rupture zones of the 1992 M7.4 Landers and the 1999 M7.1 Hector Mine (HM) earthquakes in Southern California. Although the early portion of FZTWs with larger amplitude peaks is most likely produced within the upper layers of the LVWG having the slower velocities, the late-arriving FZTWs with smaller amplitudes could come from the deeper portion of the LVWG having lower velocity reductions. These numerical tests would be helpful in interpretation of observed FZTWs to depict the depth extension of fault damage zones, which is a crucial issue in earthquake physics and strong ground motion prediction. As three decades retrospective of the 1992 Landers earthquake, we review our previous investigations on the co-mainshock damage and post-seismic heal progressions associated with the Landers and HM rupture zones using FZTWs recorded in the field to show the depth-dependent structure of fault damage zones. Then, we exhibit 3-D finite-difference synthetic FZTWs from a layered LVWG model for comparison with observations. The more detailed insight of earthquake-born fault zones at seismogenic depth would help achieve the better evaluation of the source effect, dynamic rupture, earthquake clustering/triggering as well as predicting the behavior and hazard of future earthquakes to reduce the loss in the seismic disaster.

1. Introduction

Structurally, crustal faults are often marked by zones of lowered velocity with width of a few hundred meters to a few km (Mooney and Ginzburg, 1986; Michelini and McEvilly, 1991; Cochran et al., 2009; Li et al., 1997a, b; 1998b; 2000; 2001; 2012a, b). These low-velocity zones are thought to be caused by intense fracturing during earthquakes, brecciation, liquid-saturation and possibly high pore-fluid pressure nears the fault (Nur, 1972; Sibson, 1977). The strength of the low-velocity anomalies might vary over the earthquake cycle (Vidale et al., 1994; Marone, 1998; Li et al., 1998a; 2006; 2007). As a result of the seismic velocity reduction, the fault zone forms a natural LVWG to trap and focus seismic energy as normal modes when a source located in or close to the fault zone. The FZTWs were first time discovered at active faults near Oroville and Parkfield, California and numerically synthesized in 1980s (Li et al., 1990; Li and Leary, 1990). Because FZTWs arise from coherent interference of multiple reflections at the boundaries between the low-velocity fault zone and the high-velocity surrounding rock, their amplitudes, frequencies, and dispersive waveforms strongly depend on the fault geometry and physical properties (Li and Vidale, 1996; Li, 2021). These waves enable us to insight the internal structure of fault zones at seismogenic depths with a higher resolution than before.

We have used FZTWs generated by explosions and aftershocks, and recorded at dense linear seismic arrays deployed across and along the rupture zones to investigate the subsurface geometry and physical properties of fault zones ruptured in the 1992 M7.4 Landers and 1999 Hector Mine

(HM) M7.1 earthquakes. We depict the depth-dependent LVWG with a width of a few hundred meters on the rupture zones that extend from the ground surface to 7-8 km depth, within which seismic velocities are reduced by 25-50% from wall-rock velocities. This LVWG is interpreted as being a remnant of cohesive damage zone acting as a spatial smoothing operator on fault slip where inelastic deformation occurs around the propagating crack tip during dynamic rupture in earthquake. The data recorded in our repeated seismic experiments at the Landers and HM rupture zones revealed the post-seismic fault heal progress accompanied with the seismic wave velocity increase by a few per centages within the rupture zones in several years after the mainshocks. The healing rates are depth dependent and with the greater rate in the earlier stage. The healing process on the Landers rupture zone was interrupted by the 1999 M7.1 HM earthquake which added damage as a temporal reversal of the healing on Landers faults by strong shaking and permanent strain. Our investigation at Landers and HM rupture zones is important for several reasons. First, FZTWs may offer information to find if the pattern of fault traces observed on the ground surface persists to seismogenic depth. The 3-D geometry and material properties of the fault damage zone will influence rupture propagation and may offer a basis for such elusive source parameters as the "asperities", "barriers" and upper bound frequency “fmax” of seismic waves that can be radiated from a fault zone proposed by seismologists for the strong ground motion prediction. Second, since the characteristics of FZTWs are determined directly by the mechanical properties of the fault zone, they may offer the most effective tool for monitoring the temporal change in fault zone properties for the purpose of earthquake prediction.

The depth extension of damage zone on an active fault zone is still a crucial and challenging issue in the view point of fracture mechanics and the field seismic profiling. Based on lithology, the actual fault zones are usually embedded in the layered crust composed by different rocks at different ages so that the fault zones are layered too. Our recent numerical tests for the FZTWs produced by a depth-dependent multi-layered LVWG show multiple amplitude peaks in the post-S long-duration coda, consistent with the observed FZTWs at Landers and HM rupture zones, suggesting the layered damage zone structure on realistic faults.

2. Co-Mainshock Damage Zones in the Landers and Hector Mine Earthquakes

2.1. Landers Rupture Zones Viewed by FZTWs

The M7.4 Landers earthquake on June 26, 1992 produced surface breaks of total ~70 km in the length with the maximum right-lateral slip of 7-m in Mojave Desert, Southern California (Fig.1a). Immediately after the mainshock, a mobile seismic array of 9 seismometers was used at 11 sites along the rupture zone from the Johnson Valley fault (JVF) to Camp Rock fault (CRF) to acquire FZTWs generated by aftershocks (Li et al., 1994a, b). It was the first time to deploy a linear seismic array across the surface rupture of a major earthquake to record FZTWs generated by aftershocks.

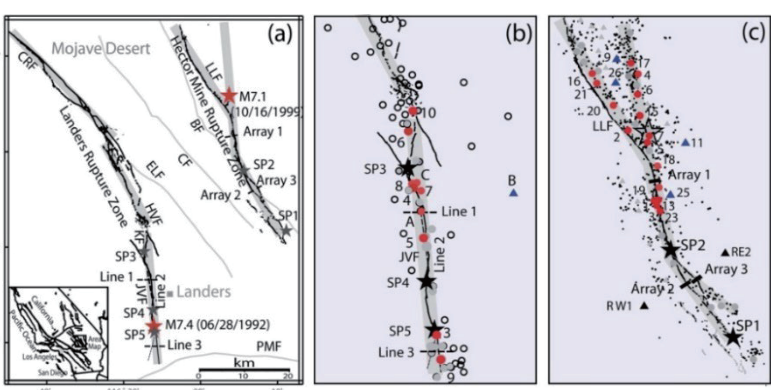


Fig. 1 Maps show locations of seismic arrays (short dashed lines with number) deployed at rupture zones (grey belts) and active faults (black lines) of the 1992 M7.4 Landers and 1999 M7.1 HM earthquakes (red stars) in Southern California (in inset), explosions (black stars with SP number) detonated within the ruptures. Dots: aftershocks recorded in experiments. Red circles with number denote events generating prominent FZTWs

FZTWs at 2-7 Hz with large amplitudes and long-duration wavetrains following the S-arrivals were recorded as both stations and sources located within the rupture zones. Later, denser seismic arrays of more than thirty seismometers (Lines 1 to 3 in Fig.1a) were installed across and along the JVF to record FZTWs that had been used to characterize the internal damage structure and segmentation of the Landers rupture zone (Li et al., 1994a, b; 1999, 2000). In our repeated experiments at Landers, the data recorded by cross-fault arrays for explosions detonated within the rupture zones were used to monitor the post-seismic fault healing with time (Li et al. 1998a; Li and Vidale, 2001; Vidale and Li, 2003).

Fig.2a exhibits fault-parallel component seismograms recorded at cross-fault array Line 1 for 3 aftershocks (events A, B, and C in Fig.1b). FZTWs with large amplitudes and long-duration wavetrains following S-arrivals at stations between W2 and E6 located close to the main fault trace for events A and C occurring on the JVF, but much shorter post-S wavetrains for event B occurring at 4-km depth and ~5 km away from the fault. Stations located out of the rupture zone registered only the brief S-wave for three events. The duration time of FZTWs recorded for aftershocks A and C increases with increasing travel distance within the rupture zone. Empirically, the ratio of the FZTW duration time to the time difference between the P and S arrivals $(t_C-t_S)/(t_S-t_P)$ (called PSSP ratio) is higher than ~1.2 for the FZTWs well generated and recorded (Li et al., 2014, 2016, 2019). Here, t_P and t_S refer to the P- and S-arrival times, and t_C-t_S refers to the post-S-duration time, in which the amplitudes of FZTWs are greater than twice the background noise level. PSSP ratios are 1.4-1.6 for on-fault aftershocks A and C, but 1.0 for the off-fault aftershock C. These observations depict a 200-250-m wide LVWG on the JVF, consistent with the width of the shear zone spanning all surface breaks on both sides of the JVF trace mapped by Johnson et al. (1997). The rupture zone is not symmetry with the main fault trace closer to the west edge of the zone. The FZTWs from aftershock C at 10-km depth show the longer post-S duration than those from aftershock A at 6.7-km depth, suggesting that the LVWG on the JVF like extends to the depth of at least ~6.7 km. Multiple amplitude peaks of FZTWs appear in post-S long durations. We interpret that the early portion of FZTWs with the largest amplitude peak are produced within the top layer of fault damage zone having the lowest velocities, and the later portion of FZTWs with smaller amplitude peaks might come from the deeper layers of fault damage zone having the smaller velocity reduction.

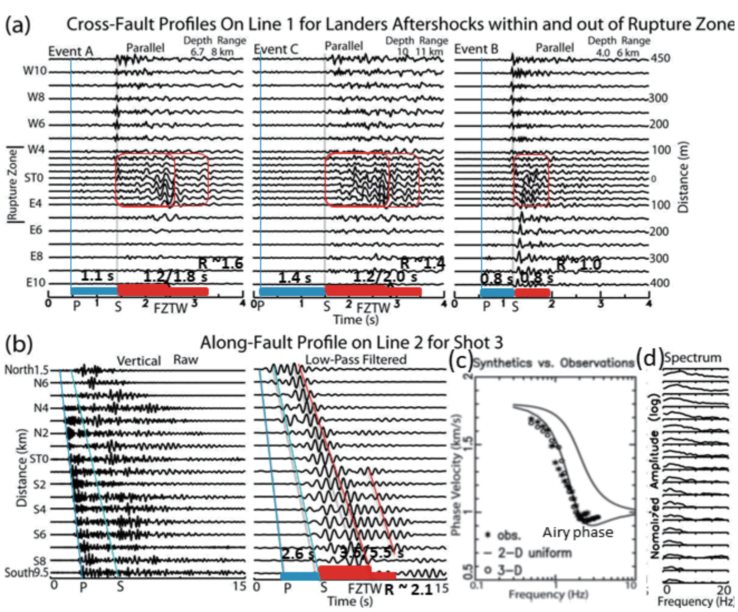


Fig. 2 (a) Seismograms recorded at cross-fault Line 1 show FZTWs with multiple amplitude peaks (in red boxes) in much longer post-S coda duration (marked by red bars) at stations within ~200-m-wide rupture zone for on-fault events A and C than the off-fault event B. P-to-S arrival times, post-S-duration of FZTWs and PSSP ratio are shown for each event. **(b)** FZTWs recorded at along-fault Line 2 for shot SP3. **(c)** Measured group velocities (stars) of FZTWs recorded at Line 2 for shot SP3 compared with synthetic dispersion curve of fundamental mode of FZTWs with Airy phase travel at the lowest group velocity. **(d)** Q derived from the decay of spectral amplitudes of FZTWs vs. station distance.

Fig.2b illustrates vertical-component seismograms recorded at 8-km-long Line 2 consisting of 16 stations along the JVF for an explosion SP3 detonated within the rupture zone. The duration of post-S FZTWs increases with distance between the shot and station, as expected for FZTWs traveling within a continuous LVWG along the JVF. FZTWs at lower frequencies travel faster than at higher frequencies, showing dispersion feature. Fig. 2c shows group velocities of FZTWs measured from multiple band-pass filtered seismograms from SP3, agreeable with synthetic dispersion curves using a depth-dependent fault zone model in Fig.3b. The PSSP ratio is 2.1 for the explosion-generating FZTWs, suggesting the strong trapping efficiency of the shallow portion of LVWG which is formed by severely damaged fault rocks. The coda-normalized spectra of FZTWs show a decrease in amplitude with distance along the rupture zone (Fig.1d), from which the Q value was measured to be ~20 at 2 Hz within the shallow Landers rupture zone.

Fig.3a exhibits group velocities derived from multiple band-pass filtered seismograms for 3 explosions and 7 aftershocks located at different depths within the Landers rupture zone, showing the depth-dependent velocity structure within the fault damage zone (Li et al., 1999, 2000). A depth-dependent fault zone structure is expected because the increasing pressure with depth will strongly affect the crack density, fluid pressure, amount of fluids, development of fault gouge and healing rate of the damage zone (Sibson, 1977; Byerlee, 1990; Scholz, 1990; Marone, 1998). For all these reasons, a realistic fault zone structure is not uniform, but varies with depth and usually layered with variations of the rock formation and fracture magnitude in stratigraphic crust.

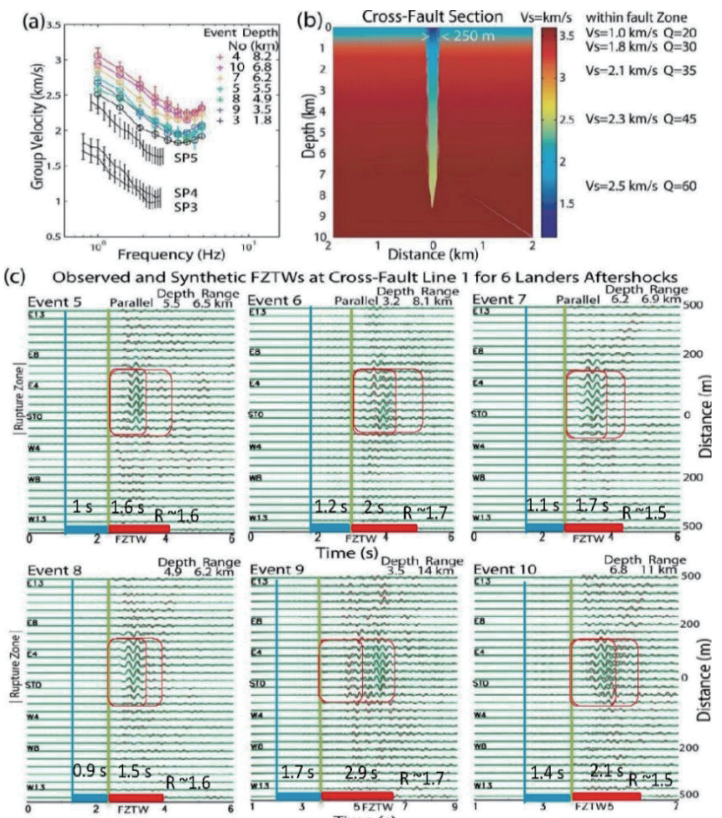


Fig. 3 (a) Group velocities of FZTWs measured from multiple band-pass filtered seismograms recorded at five stations for 3 explosions and 7 aftershocks indicate the depth-dependent velocity structure within the Landers rupture zone. Circles are the mean values at 5 stations within the rupture zone; error bars are standard deviations. The group velocities range from 1.9 km/s at 4 Hz to 2.6 km/s at 1 Hz for shallow events, but range from 2.3 km/s at 4 Hz to 3.1 km/s at 1 Hz for deep events. **(b)** The best-fit model. Shear-velocities within the LVWG reduced by 25-50% from wall-rock velocities between the ground surface and ~8-km depth. Q ~20-60. **(c)** Parallel-fault component of 3-D finite-difference synthetic seismograms (green) at Line 1 for six aftershocks (events 5-10 in Fig.1b), using the model in (b), best fit to observed FZTWs (red). Seismograms have been low-pass (<5 Hz) filtered. FZTWs (in red boxes) show multiple amplitude peaks in long wavetrains (marked by red bars). PSSP ratios (R) for these events are plotted.

Based on the measured group velocities, Q values and the width of the LVWG on the JVF revealed by FZTWs, we constructed a depth-dependent multi-layered model for the Landers rupture zone on the JVF (Fig.3b). Model parameters are obtained from simulation of explosion-generating FZTWs for the top 1-2 km of fault damage zone first, and then from aftershocks for its deeper structure. In this model, the JVF rupture zone is characterized by a low velocity and low Q waveguide, 250-m wide at the surface and tapers to 50 m at ~8 km depth. This model was used to simulate FZTWs generated by explosions and aftershocks within the rupture zone at different depths and epicentral distances from cross-fault Line 1. Fig.3c shows synthetic seismograms recorded at Line 1 for 6 aftershocks (events 5 to 10 in Fig.1b), using a 3-D finite-difference code (Vidale, 1989; Graves, 1996). A double-couple source is located within the LVWG. Although there is a trade-off among model parameters, the error in them can be constrained when independent measurements of group velocities of dispersive FZTWs and Q values from amplitude attenuation with distance.

2.2. Hector Mine Rupture Zones Viewed with FZTWs

The HM earthquake occurred on October 16, 1999, only 25 km northeast of the epicenter of the 1992 M7.3 Landers earthquake (Fig.1). The rupture zone at the surface is ~75 m wide exposed at the ground surface. The faulting with the maximum right-lateral strike slip of 5 m on the Lavic Lake fault (LLF) in the Bullion Mountains. The complex faulting pattern with bifurcation appeared in the northern and southern portions of the rupture zone. Immediately after the HM mainshock, we deployed several dense seismic arrays across rupture zones (Fig.1a, c) and recorded prominent FZTWs at 4–7 Hz for aftershocks within the rupture zone (Li, 2000, Li et al., 2002, 2003a).

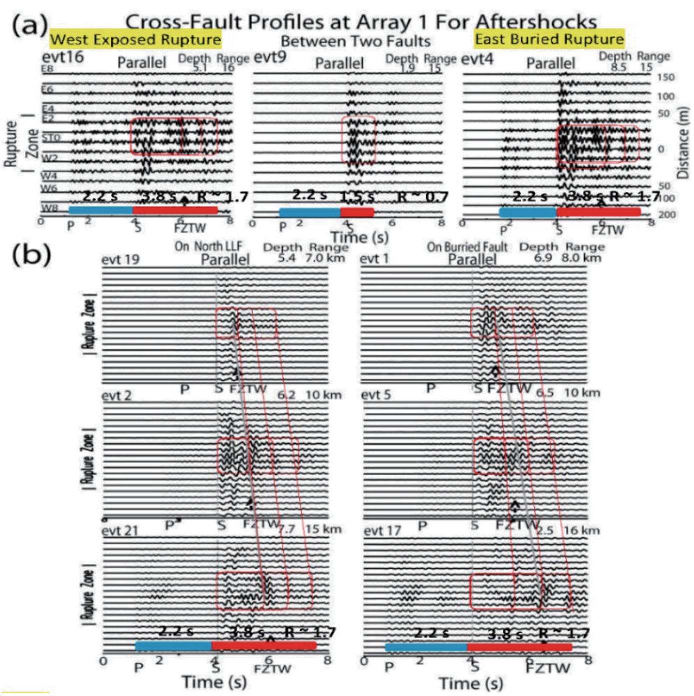


Fig. 4 (a) Seismograms recorded at Array 1 for aftershocks 16, 4, and 9 (locations shown in Fig. 1c) occurring on the north LLF exposed at the surface, the more northly-trending buried fault, and between them. FZTWs (in red boxes) with multiple amplitude peaks in long posdt-S wavetrains at stations within rupture zones for events 4 and 16, but short wavetrain for event 9. PSSP ratios (R) are ~1.7 for on-fault events, but ~0.7 for the off-fault event, indicating that LVWGs exist along the exposed rupture zone on the north LLF and the buried rupture zone. **(b)** Seismograms recorded at Array 1 for six aftershocks (events 1, 2, 5, 17, 19, and 21) occurring on the north LLF and the buried fault. Post-S durations of FZTWs (in red boxes) increase with hypocentral distance, showing the continuous LVWGs along the north LLF and the buried fault at depth. Red lines aligning multiple amplitude peaks of FZTWs denote the traveltime move-out with distance.

For example, Fig.4a exhibits fault-parallel component seismograms recorded at Array 1 for 3 aftershocks located at ~15 km north of the array. The FZTWs show multiple amplitude peaks and the PSSA ratios for the HM rupture zones are similar as those at Landers rupture zones, suggesting the fault rock damage magnitudes at Lander and HM are likely the same. Fig.4b shows FZTWs recorded at Array 1 for 6 aftershocks occurring along the bifurcated rupture zones along the north LLF and the buried fault at different hypocentral distances. Post-S durations of FZTWs increase with the distance between the events and Array1 along the north LLF and the buried fault are nearly same, suggesting the similar LVWG existing on the two northern rupture branches at depths (Li et al., 2002, 2003a). This is in accord with the complex pattern of aftershock distribution

(Hauksson et al., 2002), and the rupture model of inversion of ground motion, telemetry, geodetic data and surface breaks (Ji et al., 2002), as well as the model of dynamic rupture using a finite-element code (Oglesby et al., 2003). The slip on the northern buried rupture segment can jump to the LLF segment. The moment release on both of rupture branches is approximately equal.

We used a 3-D finite-difference code to simulate FZTWs recorded at the HM rupture zone (Fig.5). The good fit of synthetics to observations shows the depth-dependent multi-layered velocity structural model applicable for the HM rupture zone. The LVWG on the HM rupture zone is interpreted as being a remanent of process (damage) zone during dynamic rupture in the 1999 M7.1 earthquake. The reductions of velocities and Q within the fault damage zone are similar to those at the Landers, suggesting that fault rocks have been damaged to the same magnitude in two earthquakes. The LVWG width (~100 m) on the HM rupture is about a half of that (~200 m) on the Landers rupture, consisting with the scale of process zone size to rupture length as predicted in published dynamic rupture models (e.g., Scholz, 1990).

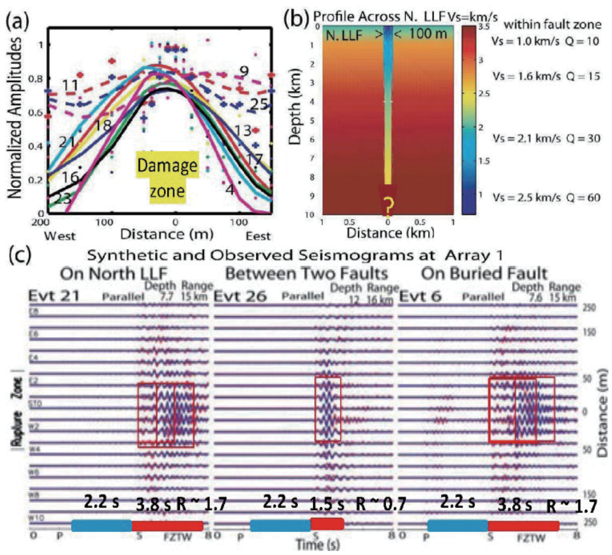


Fig. 5 (a) Computed amplitude spectral ratios of FZTW to P-wave at 15 stations of Array 1 across the LLF for 10 aftershocks in Fig.1c versus distance from the main fault trace, showing the maximum amplitude ratios at 4-6 Hz appearing at stations close to the fault trace for aftershocks occurring within the rupture zone; the amplitude ratio decreases rapidly with the distance from the fault trace. Solid (dashed) curves are for events showing (without) FZTWs. **(b)** Depth-dependent layered shear velocity model across the HM rupture zone, within which velocities are reduced by ~20-50% from wall-rock velocities, and Q is ~10-60. **(c)** Synthetic (red) and observed (blue) seismograms at Array 1 for 3 aftershocks on the north LLF, buried fault, and between them. FZTWs with multiple amplitude peaks in long-duration wavetrains shown for aftershocks 21 and 6.

3. Post-Seismic Fault Heal after Landers and Hector Mine Earthquakes

3.1. Fault Healing on Landers Rupture Zones

Studies of repeated earthquakes along a fault (e. g. Vidale et al., 1994) have shown that the ruptured fault zone could regain its strength following a large earthquake. This trend is consistent with state- and rate-dependent healing models (Dieterich, 1978; Marone, 1998). Rupture models that involve variations in fault-zone fluid pressure over the earthquake cycle have also been proposed (Sibson, 1997; Blanpied et al., 1998). Moreover, structural and rheological fault variations, as well as spatial and temporal variations in strength and stress, will affect the earthquake rupture (e.g., Wesson and Ellsworth, 1973; Das and Aki, 1977; Rice, 1980; Aki, 1984; Vidale et al., 1994; Beroza et al., 1995).

In this section, we review the post-seismic recovery in fault strength with damaged rock healing on the JVF from repeated experiments using explosions at the Landers rupture zone from 1994 to 2001 (Li et al., 1998a, Li and Vidale, 2001, Vidale and Li, 2003). Fig.6 (left) exhibits similar seismograms recorded at Line 1 for shot SP4 in five repeated experiments conducted at Landers between 1994 and 2001. However, waveform auto- and cross-correlations show that traveltimes of P, S, and FZTW advanced between 1994 and 1996, further decreased between 1996 and 1998, implying the heal of Landers damaged zone with the greater healing rate in the earlier stage. This healing trend interrupted by the 1999 M7.1 HM earthquake in 1999 between 1998 and 2000, and then recovered again between 2000 and 2001. Fig.6 (middle and right) displays similar waveforms of P, S, and FZTW in seismic profiles recorded at Line 1 for SP3 in repeated experiments. The measured decreases in traveltime would be simply to interpret. For example, the traveltime of the P wave was ~1.5 s, so the P wave velocity increased by 2.6% between 1994 and 1998. Traveltimes of the S and FZTW were 3.1 s and 4.2 s, so the shear velocities increased by ~2% between 1994 and 1998. However, wave velocities decreased by ~0.4% between 1998 and 2000, and then increased again by ~0.5% between 2000 and 2001.

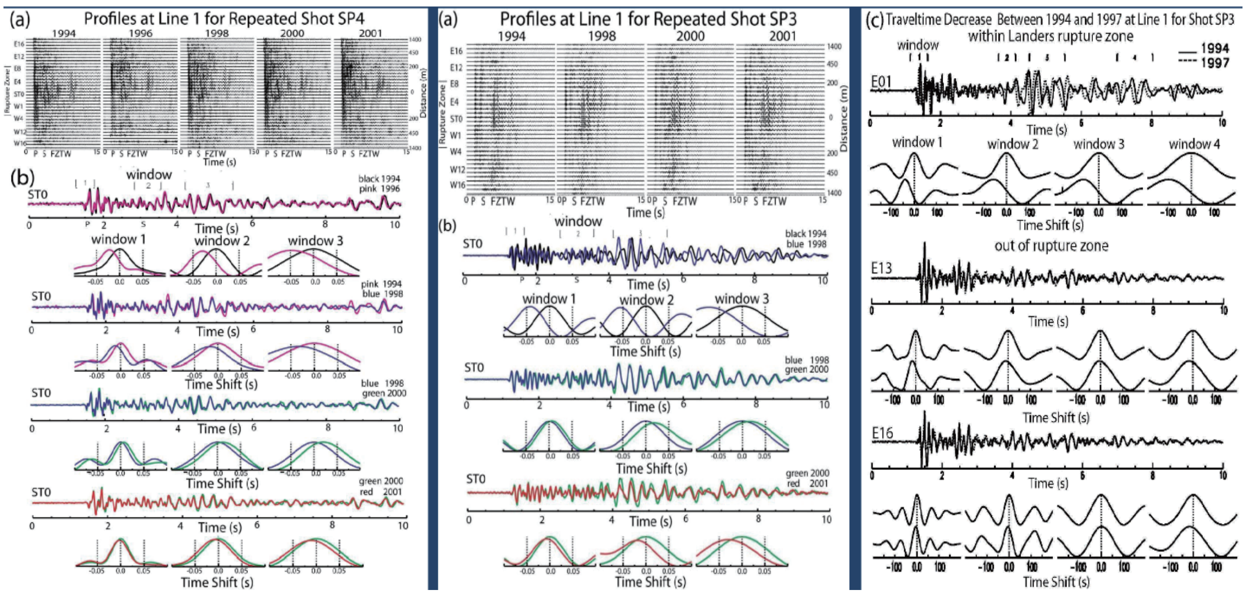


Fig. 6 left: Similar seismic profiles recorded at Line 1 across the JVF for shot SP4 detonated within the Landers rupture zone in five repeated experiments in October of 1994, 1996, 1998, 2000, and 2001. Seismograms recorded at station ST0, and auto- and cross-correlations (in blue and red) in 3 time-windows between a pair of recordings in repeated experiments show that traveltimes of P, S, and FZTW were decreased by 26, 35, and 50 ms between 1994 and 1996, and further decreased by 15, 25, and 35 ms between 1996 and 1998, but increased by 7, 12, and 16 ms between 1998 and 2000, and then decreased again by 7, 14, and 20 ms between 2000 and 2001. **middle:** Same as in left but for shot SP3. Waveform cross-correlations between repeated recordings at station ST0 show that traveltimes of these waves advanced by 40, 65 and 82 ms between 1994 and 1998. **right:** Waveform cross-correlations in time windows (1)-(4) at 3 stations of Line 1 for repeat SP3 show that travetimes arriving at station E01 within the rupture zone advanced by 50, 70, 85 and 105 ms for P, S, and FZTWs, but much smaller changes (<10 ms) at E13 and E16 outside the rupture zone between 1994 and 1997.

Fig.7a shows the decrease in traveltime of P, S, and FZTWs measured from cross-correlations of seismograms recorded at all stations of Line 1 and Line 3 for shots SP3, SP4, and SP5 between 1994 and 1996, and 1996 and 1998. Fig.7b illuminates the shear velocity increases around the JVF between 1994 and 1998. This process may be interpreted as reductive dilatancy (Nur, 1972). Estimates of the change in velocity due to the change in the density of cracks can be calculated using equations in which the elastic constants of fractured rock are functions of the crack density (O'Connell and Budiansky, 1974). The apparent crack density is defined by $e = N(a^3)/V$, where a is the radius of the flat penny-shaped crack and N is the number of cracks in a volume V. Assume randomly oriented cracks to be partially-water filled and Poisson's ratio to be 0.33. Using Vp = 2.0 km/s and Vs = 1.0 km/s for the fault rock at shallow depth, the calculated apparent crack density within the Landers rupture zone was decreased by 0.03 between 1994 and 1998, which caused ~4% increase in shear rigidity of the fault-zone rock. Fig.7c shows the depth dependence of the inferred fault healing with the greater heal at the shallower depth. We also note that the healing rate is greater near the stepover between the JVF and HVF at shallow depth probably related to the stress-controlled fluid redistribution around the fault stepover (Peltzer et al., 1996).

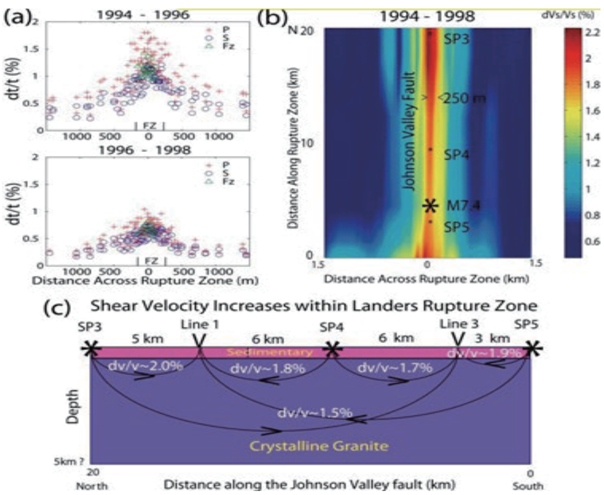


Fig.7 (a) Measred traveltime decreases of P, S and FZTWs in repeated experiments at Landers between 1994 and 1998. On average, the traveltimes of S and FZTWs within the rupture zone decreased by ~1.2% between 1994 and 1996, and decreased further by ~0.7% between 1996 and 1998, but smaller changes in the surrounding rocks. The greater velocity increase within the Landers rupture zone indicates that the fault has been healing by strengthening of rocks after the mainshock, most likely due to the closure of cracks that opened during the 1992 mainshock. **(b)** Map view shows the greater shear velocity increase within rupture zone on the JVF in 1994-1998. **(c)** S-velocity increases within rupture zones along the JVF in 1994-2001 show the larger healing rate at shallower depth.

Fig.8 summarizes measurements of cumulated and annual velocity changes within and out of the Landers rupture zone between 1994 and 2001 indicate that the healing rate is not constant but decreasing with time, consistent with the existing healing model (Dieterich, 1978; Marone, 1998). P and S velocities decreased by ~0.4% between 1998 and 2000, and then increased again by about ~0.5% from 2000 to 2001. The decrease in velocity between 1998 and 2000 might be caused by dynamic stresses during the HM earthquake on the JVF (Fialko et al., 2002), because concentrated deformation may cause damage, low-strength fault zones are sensitive to damage, and the wave amplification is caused by their low impedance. The ratio of traveltime changes for P to S waves ($\Delta t_P/\Delta t_S$) decreased from 0.75 to 0.55 between 1994 and 2001, indicating that cracks within the rupture zone were partially water-saturated and became wetter with time after the mainshock (Garbin and Knopoff, 1975; Li and Vidale, 2001; Vidale and Li, 2003). From this evidence, it appears that some partially saturated cracks which had opened during the mainshock closed soon thereafter. Closure of cracks increases the frictional strength of the fault zone, as well as its stiffness. This is consistent with our tentative interpretation of a well-developed LVWG on Landers rupture zone as being at least partially created during the 1992 M7.4 mainshock, although it represents a worn zone that has accumulated over geologic time.

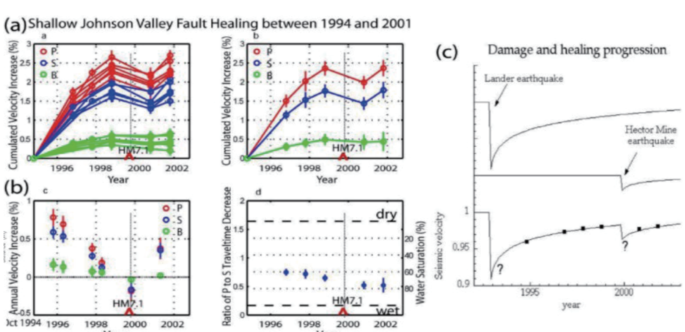


Fig. 8 (a) Measured annual increases in velocity within rupture zone show ~0.8% for P waves and ~0.6% for S and FZTWs in the early stage, but ~0.3% and ~0.2% of them in the late stage between 1994 and 1998. **(b)** The annual increase in velocity, and the ratio of P to S traveltime changes with time. **(c)** Conceptual model shows velocity versus time owing to damage from Landers rupture and HM shaking compared with observations. Healing as the logarithm of time.

3.2 Fault Healing on Hector Mine Rupture Zones

We carried out repeated seismic experiments using explosion sources at rupture zones in October of 2000 and November of 2001 to monitor seismic velocity changes with time after the 1999 M7.1 HM earthquake (Li et al., 2003b). Similar waveforms of P, S and FZTWs were recorded in repeated experiments. In summary, waveform cross-correlations of the recordings show the percentage decrease in travel time of P, S, and FZTWs at all stations on Arrays 1, 2, and 3 for shots SP1 and SP2 between 2000 and 2001 (Fig.9a). The changes in arrival times were larger at stations within the rupture zone than outside the zone, and larger on the LLF than on the BF. Accordingly, the velocity increased by ~0.7-1.4% for the P wave and ~0.5-1.0% for the S wave within the rupture zone between 2000 and 2001, but by ~0.45% for P waves and ~0.35% for S waves in surrounding rocks. These observations indicate that the rock rigidity in fault damage zone increased with time, due to the closure of cracks that opened during the mainshock. The mean ratio of decrease in travel time for P to S waves is 0.725 within rupture zone between 2000 and 2001, but is 0.79 for surrounding rocks, suggesting that cracks within the fault zone are more wet.

However, the post-seismic velocity increases vary on multiple rupture segments (Fig.9b), and are also depth dependent (Fig.19c). These variations show that the multiple fault segments may regain strength at different rates due to the heterogeneity in rock type, slip and stress distribution as well as rheology. We tentatively interpret that the rupture zone on the middle LLF at shallow depth in a region with larger slip and soft sedimentary rock would inflict greater damage due to elastic response and strong shaking, and thus greater healing occurred. Fig.9d exhibits the annual healing rate with velocity increases on the HM rupture zone on the LLF between 1 and 2 years after the 1999 mainshock and for larger intervals on the Landers rupture zone on the JVF after the 1992 mainshock, showing that a fault may regain strength more rapidly in the earlier stage of the interseismic period, but may take a long time to fully recover the strength for the next earthquake.

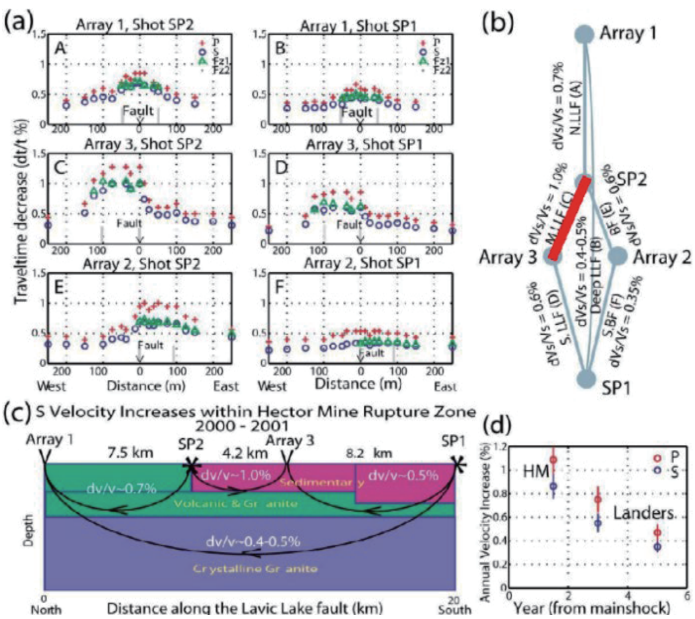


Fig.9 (a) Measured traveltime decreases in percentage for P, S, and FZTWs recorded at Arrays 1, 2, and 3 for shots SP1 and SP2 detonated in 2000 and 2001. Six shot-array pairs are labeled A to F. **(b)** Shear velocity increases within multiple rupture segments (A to F) in 2000-2001. **(c)** The depth section shows shear velocity increases within rupture zones along the LLF in 2000-2001. **(d)** Percentage velocity increase per year within HM and Landers rupture zones. At HM, the velocities increased by ~1.1% for P waves and ~0.85% for S waves for the LLF between 1 and 2 years after the 1999 mainshock. At Landers, the velocities measured on the JVF increased by ~0.77% for the P wave and ~0.58% for the S waves per year between 2 and 4 years, and further increased by ~0.47% for the P wave and 0.35% for the S wave per year between 4 and 6 years after the 1992 mainshock.

4. Synthetic FZTWs Produced by a Layered LVGW

Multiple amplitude peaks of FZTWs in long duration post-S coda recorded at rupture zones of the 1992 Landers M7.4 and 1999 Hector Mine M7.1 earthquakes might correspond multiple layers within the fault damage zone. We interpret that the early portion of FZTWs with the maximum amplitude is mainly produced within the top layer of the fault damage zone having the lowest seismic velocity, and the late portion of FZTWs with minor amplitude peaks likely produced within the deeper layers having smaller velocity reductions from wall-rock velocities. These minor amplitude peaks in the late portion of FZTWs could be worth in evaluation of the depth extension of the fault damage zone.

We recently used a 3-D finite-difference codes (Vidale, 1989; Graves, 1996) to test the trapping efficiency of a depth-dependent layered low-velocity waveguide (LVWG) for FZTWs (Li, 2022a, b). The model is composed by four layers (L1 to L4 from top to bottom), each layer is 2.5 km thick, including a 200-m-wide LVWG shown in the upper left of Fig.10. Seismic velocities within the LVWG are reduced by 50% from wall-rock velocities. Q values are 30-60 within and 150-300 out of the LVWG. A double-couple source is located within or 2-km away from the LVWG at different depths and distance from the receiver array with the station spacing of 50 m, which is perpendicularly across the LVWG at the free surface. Fig.10 (lower left) shows the diagram of ray-paths from a source located within the LVWG at 10-km depth beneath the center of the receiver array. Green lines (labeled by Fd1 to Fd4) denote ray-paths of direct waves traveling within each layer of the LVWG between the source and receiver. Black lines (labeled by Fr1 to Fr4) denote ray-paths of refracted waves at a critical angle traveling within each layer and along the boundary between the LVWG and surrounding rock. Red lines denote ray-paths of reflected waves to produce FZTWs by coherent interference within each layer of the LVWG.

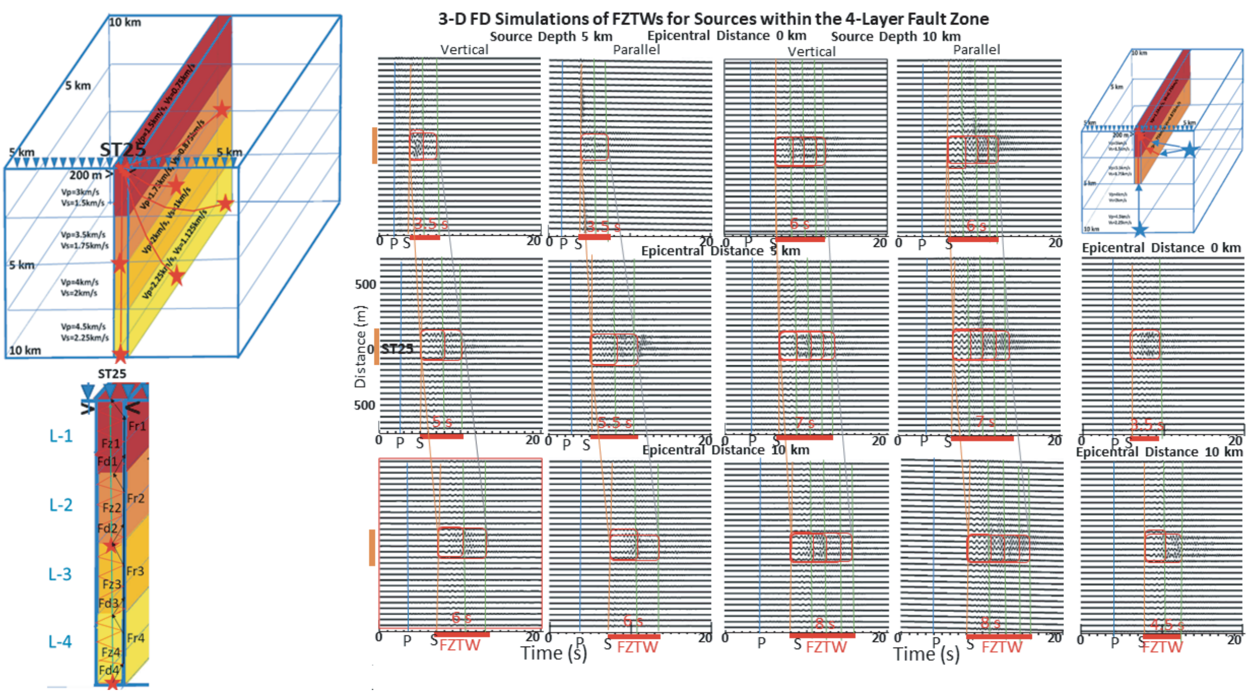


Fig. 10 left: The **s**chematic diagram of the four-layer model with a LVWG. P- and S-wave velocities in surrounding rocks for each layer are plotted. Red stars denote source locations. Blue triangles denote receiver stations. Lines with arrows denote ray-paths between the source and receiver within the LVWG. The diagram in lower left shows the ray-paths from a source located within the LVWG at 10-km depth beneath the center of receiver array. **middle:** 3-D finite-difference synthetic seismograms (<5 Hz filtered) in vertical and fault-parallel components, using this model with sources within the LVWG at 5 km and 10 km depths and epicentral distances of 0, 5 km and 10 km, respectively. Vertical red bars denote the width of LVWG. Vertical blue and brawn lines align the first P- and S-arrivals. Vertical green lines align arrivals of direct S-waves traveling within multiple layers of the LVWG. FZTWs (in red boxes) show multiple amplitude peaks corresponding to multiple layers of the LVWG. The entire durations of FZTWs (denoted by red horizontal bars), within which amplitudes above twice noise level, increases as the epicentral distance and depth increase. Grey lines align the entire durations of FZTWs, showing the travel time move-out versus distance. **Right:** The schematic diagram of the model containing a shallow LVWG having only two top layers to 5-km depth. Ray-paths in blue and green color denote waves traveling outside and within the LVWG. FZTWs show 3-4 s post-S durations for the source at 10 km depth and epicentral distances of 0 and 10 km.

Fig.10 (middle) demonstrates 3-D finite-difference synthetic seismograms using this multi-layered model for sources located within the LVWG at depths of 5 km and 10 km with epicentral distances of 0. 5 and 10 km from the array center, respectively. FZTWs with multiple amplitude peaks in the post-S coda are shown at stations located within the 200-m-wide LVWG. The FZTWs are produced by interference of coherent reflected waves with an amplitude peak at the Airy phase within each layer of the LVWG. The early portion of FZTWs with the maximum amplitude is produced within the top layer of LVWG when the refracted waves from the deep source enter the layer bottom as shown in the ray-path diagram (Fig.10, lower left). The late portion of FZTWs with minor amplitude is produced by interference of coherent reflected waves within deeper layers as refracted waves enter the bottom of these layers. PSSP ratios of the entire FZTWs are 1.8-2.0, showing that FZTWs are well generated as the source and receiver are both located within the LVWG. The duration of FZTWs increases as the travel distance within the LVWG increases. These features are consistent with observations of FZTWs at Landers and HM rupture zones. The smaller amplitude peaks in the later portion of FZTWs could help depict the depth extension of a realistic fault damage zone. We then tested the model with a shallow LVWG containing only two top layers extending to 5-km depth and sources located at 10-km depth beneath the LVWG. When seismic waves travelling from the deep source through surrounding rocks first and then enter the bottom of the shallow LVWG at 5 km, the coherent reflected waves form FZTWs within its two-layers. The post-S duration of FZTWs for the source at 10-km depth and 0-km epicentral distance is 3 s as the same as that shown in seismograms for the model with 10-km deep LVWG and the source at 5-km depth and 0 km epicentral distance, because waves travel the same distance within two top layers of the LVWG. However, the duration of FZTWs for the source at 10-km epicentral is 4 s, much less than that (6-8 s) for the model with 10-km deep LVWG and the source at 5- or 10-km depth and 10 km epicentral distance, because waves travel shorter distance within the shallow LVWG. These numerical tests show that the early portion with the maximum amplitude peak of the entire FZTWs is mainly produced in the top layer of the LVWG for a deep source.

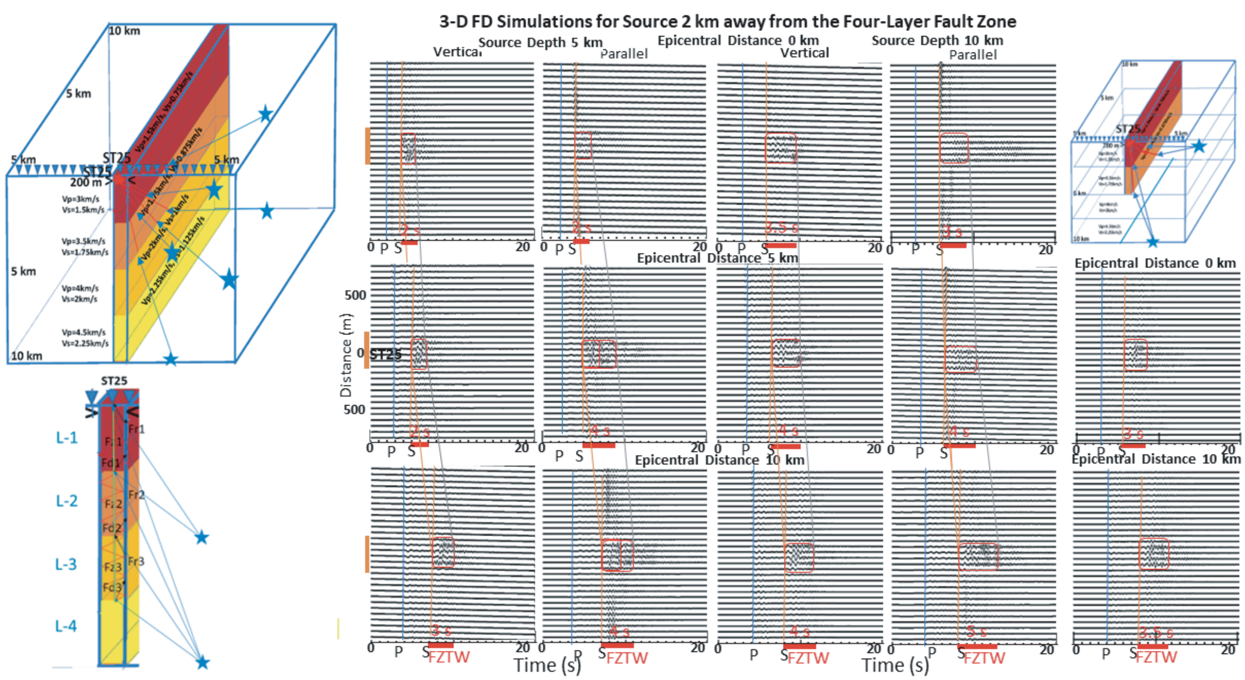


Fig. 11 Same as in Fig.11 but for sources (blue stars) located 2-km outside the LVWG. FZTWs show large amplitude but short post-S duration appear in seismograms recorded at stations located within the LVWG.

In contrast, Fig.11 shows synthetic seismograms in terms of this multi-layered model for sources located 2 km away from the LVWG at depths of 5 km and 10 km with epicentral distances of 0 km, 5 km and 10 km from the center of the array, respectively. Seismic waves from off-fault sources travel in the high-velocity surrounding rocks first, and then enter the bottom of upper layers of the LVGW within which FZTWs are produced. Although these FZTWs show large amplitude but have shorter post-S duration than those for sources at the same depths and epicentral distances within the LVWG because they travel shorter distance within the upper layers of the LVWG only. The PSSP ratios for these FZTWs are around 1.0, much smaller than those for on-fault sources. We also tested the model with a shallow LVWG containing two top layers extending to 5 km depth for sources at 10 km depth with 2-km offset from the LVWG (Fig.11, right). Seismic waves from the deep source enter the bottom of the shallow LVWG at 5 km, within which FZTWs are produced. The post-S durations of FZTWs for the source at 0 and 10 km epicentral distances is 3-3.5 s nearly same to those shown in Fig.10 (right), because FZTWs travel the similar distances within the two-layer LVWG in these examples. These numerical tests show that the FZTWs with large amplitude and short duration can be produced within the top few layers of the LVWG despite the source location within or out of the LVWG. However, the duration of these FZTWs is much shorter than that from sources at 10-km depth within the 10-km deep LVWG. Therefore, when we analyze the recorded FZTWs, using the early portion of FZTWs showing large amplitudes, which are dominant in seismograms, but neglecting the late portion of FZTWs with small amplitudes might lead to underestimate the depth extension of a realistic multi-layered fault damage zone.

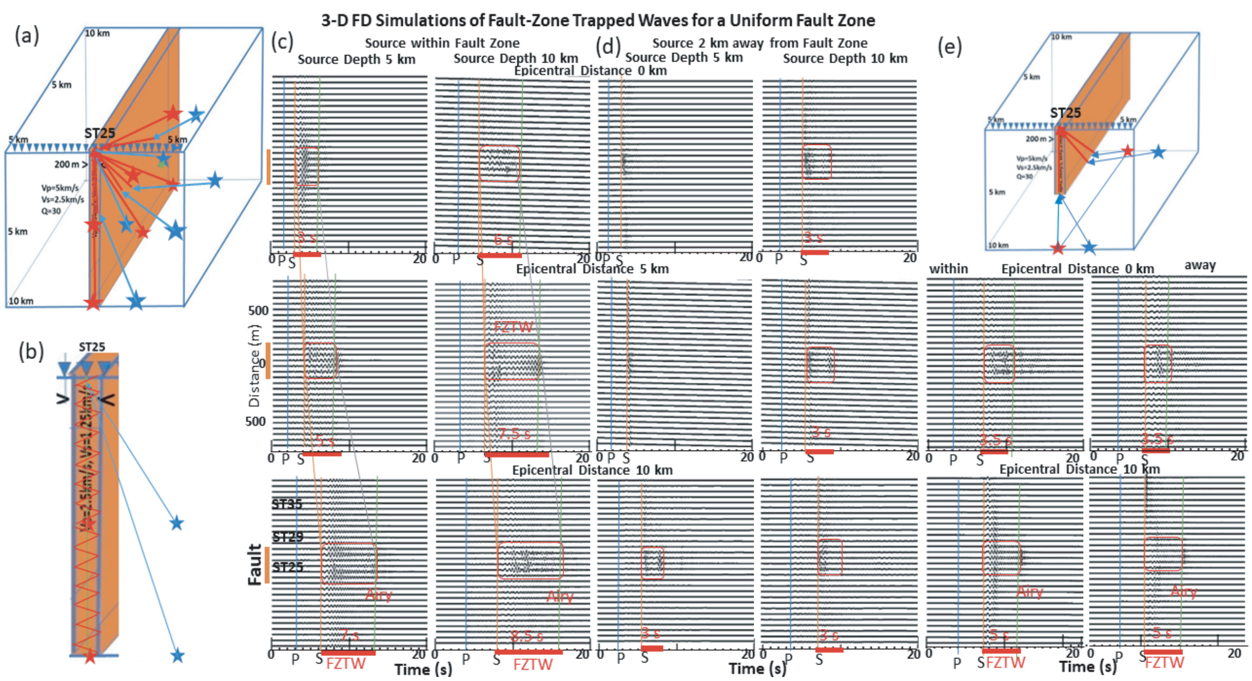


Fig. 12 Same as in Fig.10 and Fig.11 but for a uniform LVWG. FZTWs with a single large amplitude peak in long duration appear as the source and station both within the LVWG.

For a comparison, we tested the trapping effect for a uniform LVWG, within which seismic velocities are reduced by 50% from wall-rock velocities (Fig.12, left). As the source is located within the LVWG extending to 10-km depth, FZTWs with large amplitude and dispersion appear at stations located within the 200-m wide LVWG (Fig.12, middle). The appearance of these FZTWs is relatively simple and with a single amplitude peak, different from those with multiple amplitude peaks produced by a multi-layered LVWG. The duration of FZTWs increase as the travel distance within the LVWG between the source and receiver increases. As the source is located 2 km away from the deep LVWG and above its bottom at 10-km, no FZTWs are observed in all stations. However, for a shallow uniform LVWG extending to 5-km depth, FZTWs are produced within the LVWG as waves from 10-km deep sources enter the bottom of shallow LVWG at 5-km depth (Fig.12, right). The durations of these FZTW are 3.5-5 s, similar to those produced by sources located at 5-km depth within the 10-km deep LVWG (Fig.12, middle). Using a uniform LVWG to simulate recorded FZTWs with the maximum amplitude peak might lead to an underestimation of the depth extension of a depth-dependent multi-layer fault damage zone.

5. Conclusion and Discussion

We used FZTWs generated by explosions and aftershocks and recorded at dense seismic arrays deployed across and along the rupture zones of the 1992 M7.4 Landers and 1999 M7.1 Hector Mine earthquakes to investigate the internal structure and physical properties of the fault damage zones in 4-D. The damage zones are interpreted as being a process (breakdown) zone as suggested by fracture mechanics. This zone is partly the transient result of the dynamic rupture in the 1992 M7.4 and 1997 M7.1 mainshocks although it probably also represents the accumulated damage from previous earthquakes. The width of fault damage zone on the LLF at Hector Mine is about half of that on the JVF at Landers. The entire length of HM rupture zone is about half of the Landers rupture zone, consistent with the proportional scaling of process zone size to rupture length as predicted in the published dynamic rupture models. Velocity reductions within Landers and HM rupture zones are approximately the same, indicating that fault rocks were damaged to the similar magnitude in the two earthquakes. At Landers, the rupture is segmented by stepovers between pre-existing faults. At HM, the northern rupture zone bifurcates with the similar slip displacement release and damage magnitude, but one exposed at the surface and another buried.

We observed the healing of the JVF with time after it ruptured in the 1992 M7.4 Landers earthquake, and this healing was interrupted in 1999 by the nearby M7.1 HM earthquake rupture. Although the fault healing due to chemical process may have a significant effect over longer time period at seismogenic depth (Angevine and Turcotte, 1983), we prefer to the 'crack dilatancy' mechanisms associated with the post-seismic fault healing at shallow depth even if other processes are active too. The HM earthquake both strongly shook and permanently strained the JVF, adding damage discernible as a temporary reversal of the healing process. The fault has since resumed the trend of strength recovery that it showed after the Landers earthquake. These observations lead us to speculate that fault damage could bd caused by strong seismic waves, and may help to explain earthquake clustering and seismicity triggering by shaking, and friction reduction during faulting. Repeated surveys at the HM rupture zone showed that the seismic velocities increased with varying rate on different rupture segments after the 1999 mainshock, suggesting that a fault may regain strength at a different rate that depends upon the heterogeneity of rock type, slip and stress distributions, and rheology. The healing rate decreases with time, showing that the damaged fault zone can regain strength rapidly in the early stage of the interseismic period, but may take a long time to fully recover the strength for the next earthquake on it.

The depth extension of damage zone on an active fault zone is still a crucial and challenging issue in the view point of fracture mechanics and in the seismic profiling. We observed multiple amplitude peaks of FZTWs in long duration post-S coda recorded at Landers and HM rupture zones, which might correspond multiple layers of the fault damage zone at depth. Our recent numerical tests for the FZTWs produced by a depth-dependent layered LVWG show multiple amplitude peaks in the post-S long-duration coda, consistent with the observed FZTWs at Landers and HM rupture zones. Although the early portion of FZTWs with the maximum amplitude peak is most likely produced within the top layer of the fault damage zone, the later-portion of FZTWs with smaller amplitude peaks might come from its deeper layers. For a layered fault damage zone, FZTWs with large amplitudes could be produced within its upper layers as only as the source is located below the bottom of these upper layers. However, these FZTWs show shorter post-S durations than those for the source located within deeper layers of the fault damage zone. Because most aftershocks occur out of the narrow (several hundred meters wide fault core zone, FZTWs with large amplitudes but short post-S durations are assumed to be dominant in recorded seismograms. It might lead us to underestimate the depth extension of fault damage zone.

Acknowledgments

This project was supported by grants of the U.S. National Science Foundation and Geological Survey, and the Southern California Earthquake Center (SCEC). Acknowledge to the IRIS for the use of PASSCAL instruments in our field experiments. The author specially thanks many scientists and students from multiple institutions for their collaborations in this research project and participating in field experiments, including the late Professor K. Aki, J. Vidale, C. Marone, S. Day, D. Oglesby, E. Cochran, R. Graves, P. Chen, A. Hasemi, D. Bowman, D. Adams, K. Mayeda, Fei, Xu, J. Chen, J. McRaney, M. Robertson, R. Laney, WHK Lee, T. Burdette, K. Gross, R. White, M. Alvarez, R. Russell.

References

Aki, K., 1984, Asperities, barriers, characteristic earthquakes, and strong motion prediction, J. Geophys. Res., 89, 5867-5872.

Angevine, C. L and D. L. Turcotte, 1983. Porosity reduction by pressure solution: A theoretical model for quartz arenites, Geol. Soc. Am. Bull., 94, 1129-1134.

Beroza, G., A. Cole, and W. Ellsworth, 1995. Stability of coda wave attenuation during the Loma Prieta, California, earthquake sequence, J. Geophys. Res., 100, 3977-3987.

Blanpied, M. L., C. J. Marone, D. A. Lockner, J. D. Byerlee, and D. P. King, 1998. Quantitative measure of the variation in fault rheology due to fluid-rock interactions, J. Geophys. Res., 103, 9691-9712.

Byerlee, J., 1990. Friction, overpressure and fault-normal compression, Geophys. Res. Lett., 17, 2109-2112.

Chester, F. M., J. P. Evans, and R. L. Biegel, 1993. Internal structure and weakening mechanisms of the San Andreas fault, J. Geophys. Res., 98, 771-786.

Cochran, E. S., Y. G Li, P. M. Shearer, S. Barbot, Y. Fialko, and J. E. Vidale, 2009. Seismic and geodetic evidence for extensive, long-lived fault damage zones, Geology, v9.37; no. 4, 315–318, doi:10.1130/G25306A.1.

Das, S., and K. Aki, 1997. Fault plane with barriers: A versatile earthquake model, J. Geophys. Res., 82, 5658-5670.

Dieterich, J. H., 1978. Time-dependent friction and the mechanics of strike-slip. Pure Appl. Geophys. 116, 790-806.

Fialko, Y. et al. 2002. Deformation on nearby faults induced by the 1999 Hector Mine earthquake. Science 297, 1858-1862.

Garbin, H. D., and L. Knopoff, 1975. Elastic moduli of a medium with liquid-filled cracks, Quart. Appl. Math., 32, 301-303.

Graves, R., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, Bull. Seism. Soc. Am., 86, 1091-1106.

Hauksson, E., L. M. Jones, and K. Hutton, 2002. The 1999 Mw7.1 Hector Mine, California earthquake sequence: complex conjugate strike-slip faulting, Bull. Seis. Soc. Am., 92, 1154-1170, doi:10.1785/0120000920.

Hickman, S., R., Sibson, and R. Bruhn, 1995. Introduction to special section: mechanical involvement of fluids in faulting, J. Geophys. Res., 100, 12831-12840.

Ji, C., D. J. Wald, and D. V. Helmberger, 2002. Source description of the 1999 Hector Mine, California earthquake; Part II: Complexity of slip history, Bull. Seismol. Soc. Am. V92, No. 4, 1208–1226.

Johnson, A. R. Fleming, K. Cruikshank, S. Martosudarmo, N. Johnson, and K. Johnson, 1997. Analecta of structures formed during the 1992 Landers-Big Bear, California earthquake sequence, USGS Open File Rep., 97-94, 59 pp.

Li, Y. G., P. C. Leary, K. Aki, and P. E. Malin, 1990. Seismic trapped modes in the Oroville and San Andreas fault zones, Science, 249, 763-766.

Li, Y. G., and P. C. Leary, 1990. Fault-zone trapped seismic waves, Bull. Seismol. Soc. Am., 80, 1245-1271.

Li, Y. G., K. Aki, D. Adams, A. Hasemi, and W. H. K. Lee, 1994a. Seismic guided waves trapped in the fault zone of the Landers, California, earthquake of 1992, J. Geophys. Res., 99, 11,705-11,722.

Li, Y. G., J. E. Vidale, K. Aki, C. J. Marone, and W. H. K. Lee, 1994b. Fine structure of the Landers fault zone: Segmentation and the rupture process, Science, 256, 367-370.

Li, Y. G., W. L. Ellsworth, C. H. Thurber, P. E. Malin, and K. Aki, 1997a. Observations of fault-zone trapped waves excited by explosions at the San Andreas fault, central California, Bull. Seism. Soc. Am., 87, 210-221.

Li, Y. G., F. L. Vernon, and K. Aki, 1997b. San Jacinto fault-zone guided waves: A discrimination for recently active fault strands near Anza, California, J. Geophys. Res., 102, 11,689-11,701.
Li, Y. G., J. E. Vidale, K. Aki, F. Xu, and T. Burdette, 1998a. Evidence of shallow fault zone strengthening after the 1992 M7.5 Landers, California, earthquake, Science, 279, 217-219.
Li, Y. G., K. Aki, J. E. Vidale, and M. G. Alvarez, 1998b. A delineation of the Nojima fault ruptured in the M7.2 Kobe, Japan, earthquake of 1995 using fault-zone trapped waves, J. Geophys. Res., 103, 7247-7263.
Li, Y. G., K. Aki, J. E. Vidale, and F. Xu, 1999. Shallow structure of the Landers fault zone from explosion-generated trapped waves, J. Geophys. Res., 104, 20,257-20,275.
Li, Y. G., J. E. Vidale, and K. Aki, 2000. Depth-dependent structure of the Landers fault zone using fault zone trapped waves generated by aftershocks, J. Geophys. Res., 105, 6237-6254.
Li, Y. G., Mapping fault zones, 2000. IRIS Newsletter, V2000, No.1, 26.
Li, Y. G. and J. E. Vidale, 2001. Healing of the shallow fault zone from 1994-1998 after the 1992 M7.5 Landers, California, earthquake, Geophys. Res. Letter, 28, 2999-3002.
Li, Y. G., J. E. Vidale, S. M. Day and D. Oglesby, 2002. Study of the M7.1 Hector Mine, California, earthquake fault plan by fault-zone trapped waves, Hector Mine Earthquake Special Issue, Bull. Seism. Soc. Am., 92, 1318-1332.
Li, Y. G., J. E. Vidale, D. Oglesby, S. M. Day and E. Cochran, 2003a. Multiple-fault rupture of the M7.1 Hector Mine, California, earthquake from fault-zone trapped waves, J. Geophys. Res., 108, ESE 11, 1-25.
Li, Y. G., J. E. Vidale, S. M. Day, D. D. Oglesby, and E. Cochran, 2003b. Post-seismic fault healing on the 1999 M7.1 Hector Mine, California earthquake, Bull. Seism. Soc. Am., 93, 854-869.
Li, Y. G., J. E., Vidale, and S. E. Cochran, 2004. Low-velocity damaged structure of the San Andreas fault at Parkfield from fault-zone trapped waves, Geophy. Res. Lett., 31, L12S06.
Li, Y. G., P. Chen, E. S. Cochran, J. E., Vidale, and T. Burdette, 2006. Seismic evidence for rock damage and healing on the San Andreas fault associated with the 2004 M6 Parkfield earthquake, Special issue for Parkfield M6 earthquake, BSSA, 96, S1-15.
Li, Y. G., P. Chen, E. S. Cochran, and J. E. Vidale, 2007. Seismic velocity variations on the San Andreas Fault caused by the 2004 M6 Parkfield earthquake and their implications, Earth, Planets and Space, 59, 21-31.
Li, Y. G., G. De Pascale, M. Quigley, and D. Gravely, 2014. Fault damage zones of the M7.1 Darfield and M6.3 Christchurch earthquakes characterized by fault-zone trapped waves, Tectonophysics, 10.1016/j.tecto.2014.01.029.
Li, Y. G., R. D. Catchings, and M. R. Goldman, 2016. Subsurface fault damage zone of the 2014 Mw 6.0 South Napa, California, earthquake viewed from fault-zone trapped waves, Bull. Seism. Soc. Am. 106, doi: 10.1785/0120160039.
Li, Y. G., R. D. Catchings, and M. R. Goldman, 2019. Rupture branching Structure of the 2014 Mw 6.0 South Napa, California earthquake inferred from explosion-generated fault-zone trapped waves, Bull. Seismol. Soc. Am. 109, doi: 10.1785/0120180181.
Li, Y. G., 2021. Spatial-temporal characterization of the San Andreas Fault by fault-zone trapped waves at seismic experiment site, Parkfield, California, Earthquake Science, 34(3): 261–285, doi: 10.29382/eqs-2021-0014.
Li, Y. G., 2022a. Study of fault-zone damage and heal: 30-years retrospective of the 1992 M7.4 Landers Earthquake and numerical tests of FZTWs from a layered fault zone, 052, SCEC Annual Meeting, Palms Spring, September.
Li, Y. G., 2022b. Post-seismic heal of fault zones damaged during the 1992 M7.3 Landers, 1999 M7.1 Hector Mine and 2004 M6 Parkfield Earthquakes in California, T24A-03, American Geophysics Union Fall Meeting, Chicago.
Marone, C., 1998. The effect of loading rate on static friction and the rate of fault healing during the earthquake cycle, Nature, 391, 69-72.
Massonnet, D, W. Thatcher, and H. Vadon, 1996. Detection of postseismic fault-zone collapse following Landers earthquake, Nature, 382, 612-616.
Mooney, W. D. and A. Ginzburg, 1986. Seismic measurements of the internal properties of fault zones, Pure Appl. Geophys., 124, 141-157.
Noda H., E. M. Dunham, and J. R. Rice, 2009. Earthquake ruptures with thermal weakening and the operation of major faults at low overall stress levels, J. Geophys. Res., 114 (B7), B07302.
Nur, A., 1972. Dilatancy, pore fluid, and premonitory variations of ts/tp travel times, B. S. Soc. Am., 62, 1217-1222.
O'Connell, R. J. and B. Budiansky, 1974. Seismic velocities in dry and saturated cracked solids, J. Geophys. Res., 79, 5412-5426.
Oglesby, D. D., S. M. Day, Y. G. Li, and J. E. Vidale, 2003. The 1999 Hector Mine earthquake: The dynamics of a branched fault system, Bull. Seism. Soc. Am., 93, 2459-2476.
Peltzer, G., P. Rosen, F. Rogez, and K. Hudnut, 1996. Postseismic rebound in fault step-overs caused by pore fluid flow, Science, 273, 1202-1204.
Rice, J. R., 1980. The mechanics of earthquake rupture, in Physics of the Earth's Interior, edited by A. M. Dziewonski and E. Boschi, pp. 555-649, Amsterdam.
Scholz, C. H., 1990. The Mechanics of Earthquakes and Faulting, Cambridge Univ. Press, New York.
Sibson, R., 1977. Fault rocks and fault mechanisms, J. Geol. Soc. London, 133, 191-213.
Sieh, K., et al., 1993. Near-field investigations of the Landers earthquake sequence, April to July 1992, Science, 260, 171-176.
Vidale, J. E., 1989. Finite-Difference seismograms for SH waves, Bull. Seismo. Soc. Am., 78, 1765-1782.
Vidale, J. E., W. L. Ellsworth, A. Cole, and C. Marone, 1994. Rupture variation with recurrence interval in eighteen cycles of a small earthquake, Nature, 368, 624-626.
Vidale, J. E. and Y. G. Li, 2003. Damage to the shallow Landers fault from the nearby Hector Mine earthquake, Nature, 421, 524-526.

S100 蛋白家族在中枢神经系统疾病中的研究进展

张沐捷[1]，王希萌[2]， 梁嘉欣[1]，田晓燕[1]，陈宝珠[1]，彭亮[1]*
[1]广州医科大学附属第五医院，广东 广州，510515，中国； [2] Hefei Normal University, Hefei , Anhui, China, California Chemical Inc, Los Angeles, CA 91776, USA
*通讯作者：彭亮，邮箱：pl_206@126.com

摘要：S100 蛋白是脊椎动物组织中广泛表达的小分子 EF-手型（EF-handtype）结构蛋白，参与细胞的增殖、凋亡、分化、炎症、钙离子稳态和能量代谢等多种过程。许多神经系统疾病都归因于细胞内钙稳态的破坏，作为钙离子结合蛋白的 S100 在如阿尔兹海默症、帕金森等疾病的发生发展中具有重要作用，提示 S100 蛋白可能为中枢神经系统疾病的诊断标志及治疗靶点。本文对 S100 蛋白家族成员在不同神经系统疾病中的意义进行综述，并就临床治疗中针对 S100 蛋白的潜在干预策略进行讨论。

关键词：中枢神经系统；S100 蛋白家族；钙稳态

Advances in the study of S100 protein family in central nervous system diseases

Mujie Zhang, Ximeng Wang , Jiaxin Liang, Xiaoyan Tian, Baozhu Chen, Liang Peng*
The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510515, Guangdong, China， Hefei Normal University, Hefei , Anhui, China, California Chemical Inc, Los Angeles, CA 91776, USA
* Corresponding author: Liang Peng, e-mail: pl_206@126.com

Abstract: S100 protein is a small molecular EF-handtype structural protein widely expressed in vertebrate tissues, which is involved in many processes such as cell proliferation, apoptosis, differentiation, inflammation, calcium homeostasis and energy metabolism. Many nervous system diseases are attributed to the destruction of intracellular calcium homeostasis. S100, as a calcium binding protein, plays an important role in the occurrence and development of diseases such as Alzheimer's disease and Parkinson's disease, it is suggested that S100 protein may be a diagnostic marker and therapeutic target for central nervous system diseases. This article reviews the significance of S100 protein family members in different nervous system diseases,

and discusses the potential intervention strategies for S100 protein in clinical treatment.

Key words: central nervous system; S100 protein family; calcium homeostasis

一、S100 蛋白家族概述

1965 年，Blake W. Moore 从牛大脑中分离出一种蛋白质，由于该蛋白在中性 pH 条件下可 100%溶于饱和硫酸铵溶液，因此被命名为 S100[1]。Isobe 等[2]后来发现 S100 蛋白为二聚体结构，由两个同源但不同的分子组成，定义为 S100A 和 S100B。随后，更多 S100 蛋白被发现，目前已知的 S100 蛋白家族包括 25 个成员[3]。大量研究表明，这些成员参与细胞内多种不同的途径，在基本的细胞过程如增殖、凋亡、分化和炎症中发挥关键作用[4, 5]，并在癌症、自身免疫性疾病、慢性炎症等疾病中发挥重要作用[6]（图 1）。

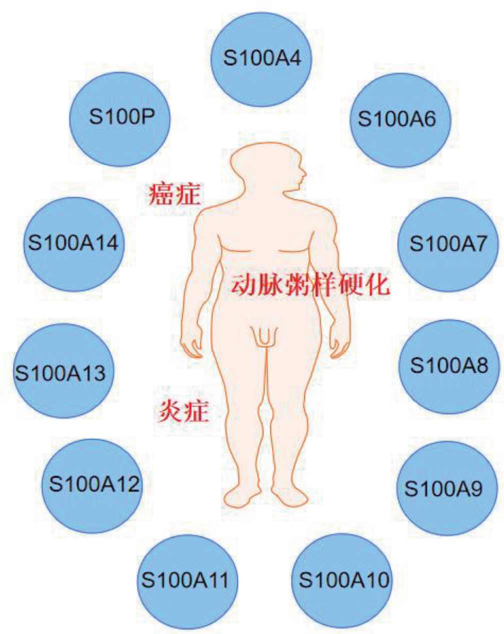


图 1 S100 蛋白在各类疾病发病机制中的作用

Fig.1 Role of S100 proteins in pathogenesis of various diseases

1. S100 蛋白家族的表达

目前，S100 蛋白只被发现在脊椎动物组织中广泛表达。在人类中，S100A 亚家族(S100A1-S100A16)的基因聚集在染色体 1q21 上，而其他 S100 成员的基因位于染色体 21q22(S100B)、Xp22(S100G)、4p16(S100P)和 5q13(S100Z)[7]。S100 蛋白家族成员在不同的细胞和组织呈现出特异性表达模式，并且表达水平因细胞

类型而异，例如 S100A7 主要在角化细胞表达，而 S100A10 则主要在巨噬细胞表达[8]。

同时，S100 基因的表达受到氧化应激、细胞因子、生长因子和 Toll 样受体（Toll-like receptor，TLR）的严格调节，以维持免疫稳态。例如，Calprotectin（S100A8/A9）在如单核细胞、中性粒细胞和树突状细胞等免疫细胞中构成性表达[9]，激活后也可在成纤维细胞[10]或成熟巨噬细胞[11]中表达。此外，表观遗传机制在调控 S100 基因表达中也起着至关重要的作用，DNA CpG 岛甲基化是一种常见的转录抑制方法，DNA 低甲基化在前列腺癌和胃癌中显著诱导 S100 蛋白家族几个成员的表达[12, 13]。与其他 EF-手型（EF-handtype）蛋白不同，S100 蛋白不仅存在于细胞内，还可分泌至细胞外，并与受体相互作用[4]。然而由于 S100 蛋白缺乏经典的内质网-高尔基体分泌途径的结构序列，其潜在的主动分泌机制尚不完全清楚。

2. S100 蛋白家族的结构

S100 蛋白为 EF-handtype 结构的钙结合蛋白，为 EF 家族中最大的亚群[14]。通常，S100 成员形成同源二聚体，每个单体由两个 EF-Hand 基序组成，其中每个 EF-Hand 基序由一个 Ca^{2+}结合环组成，其两侧为α-螺旋（螺旋-环-螺旋）[15]。N 端是 S100 特异性的，有 14 个氨基酸残基，具有较低的 Ca^{2+}结合亲和力，而具有 12 个氨基酸残的典型 C 端则具有较高的 Ca^{2+}结合亲和力[16, 17]。除 Ca^{2+}外，一些 S100 蛋白如 S100A7、S100A12 和 S100B 能够结合其他二价金属离子，如锌（Zn^{2+}）、铜（Cu^{2+}）和锰（Mn^{2+}）等。尽管 S100 家族成员具有高度的序列相似性并表现出相似的折叠行为，但它们的形状和电荷则不尽相同[18]。

3. S100 蛋白家族的生物学功能

S100 蛋白与大量不同靶点相互作用，从而参与各种细胞过程，包括 Ca^{2+}稳态、增殖、分化、凋亡、炎症和细胞迁移等[4]。又因 Ca^{2+}几乎参与每一个细胞过程，且升高的 Ca^{2+}浓度会导致细胞死亡。因此，严格调节 Ca^{2+}水平至关重要[19]。S100 蛋白不仅可通过将胞内游离 Ca^{2+}与质膜结合并运输，而且可通过与跨膜蛋白相互作用，来确保细胞内 Ca^{2+}稳态[20]。大多数细胞外 S100 蛋白与晚期糖基化终产物受体（advanced glycosylation end product-specific receptor，RAGE）和 Toll 样受体 4（Toll-like receptor4，TLR4）相互作用[21]，S100 与这些受体结合后刺激

活化核因子κB（nuclear factor kappa-B，NF-κB），导致细胞因子和促炎因子的表达上调，如白介素-1β（IL-1β）和肿瘤坏死因子-α（TNF-α）等[8, 22]。

此外，研究表明 S100A7、S100A8、S100A9 和 S100A15 具有趋化特征，吸引中性粒细胞和淋巴细胞[23-26]。S100 蛋白对细胞增殖的促进也通常以 RAGE 依赖性方式介导，诱导 NF-κB 和丝裂原活化蛋白激酶（MAPK）信号传导，从而介导生长因子和生存蛋白的转录[27-29]。一些 S100 蛋白如 S100A2、S100A4、S100A14 和 S100B 还可与 p53 相互作用[30-32]，调节 DNA 复制并诱导促凋亡蛋白的转录[33]，因此 S100 蛋白也与细胞凋亡的调节紧密相关（图 2）。

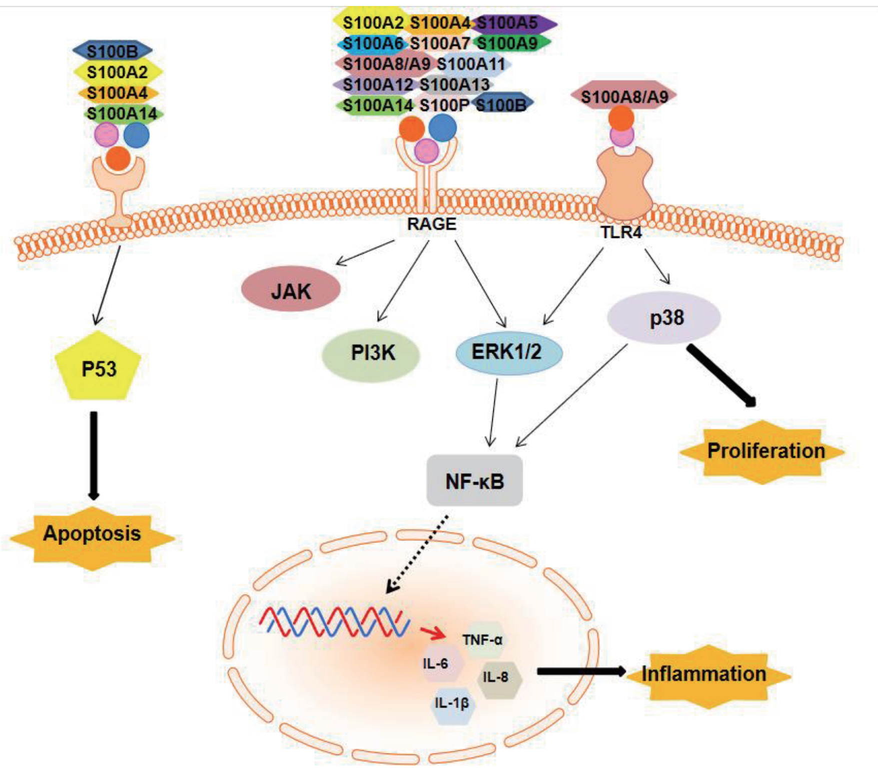


图 2 S100 蛋白参与信号通路调控示意图

Fig.2 Schematic diagram of S100 proteins participating in signal pathway regulation

二、S100 蛋白家族在中枢神经系统疾病中的意义

1. S100 蛋白与阿尔兹海默症

阿尔茨海默病（Alzheimer's disease，AD）是一种慢性进行性神经退行性疾病，影响大脑皮层和海马的广泛区域。大多数 AD 患者是特发性的，该疾病的特点是发病晚，无法从大脑中清除淀粉样β肽（Aβ）[34]。主要症状表现为进行性记忆缺陷、认知障碍和性格改变。神经炎症、氧化应激和钙水平失调是 AD 病理学中的重要特征[35]。迄今为止描述的 25 种 S100 蛋白中，有一些表达于大脑中，

其中七种与 AD 通路有关：S100B、S100A1、S100A6、S100A7、S100A8、S100A9 和 S100A12。这些蛋白作为 Ca^{2+}传感器，调节参与 Ca^{2+}释放、摄取、以及 Ca^{2+}通道的活性，也为 AD 中维持异常 Ca^{2+}信号提供反馈和前馈机制。

S100B 是 AD 范围内研究最多的 S100 蛋白。研究表明，AD 患者和 AD 小鼠模型[36, 37]血清中 S100B 水平明显高于正常对照组，并可在海马、颞叶和皮层的 I 层中观察到 S100B 水平的大量增加[38]。在一项临床研究中，患有中度和重度痴呆的 AD 患者血清 S100B 分别增加了 60 倍和 37 倍；在患有中度痴呆症的 AD 患者中，S100B 水平增加后，自身抗体增加 10 倍；有趣的是，在患有严重痴呆症的患者中，自身抗体的水平则降至与对照组保持一致，这表明在患有重度痴呆症的 AD 患者中，在此阶段没有针对 S100B 浓度升高的免疫保护，可能是由于形成了不溶性免疫复合体[39]。S100A1 主要在神经元中表达，与其相关的一些靶点在 AD 中发生了改变，例如赖氨酸受体（RyR），tau 以及 RAGE 等，且 S100A1 与 tau 磷酸化、神经元细胞对 Aβ的敏感性有关[40]。此外，在人类 AD 患者和 AD 小鼠模型的斑块样沉积物中也观察到细胞外 S100A1 的沉积[41]。S100A6 在 AD 相关基因名中被鉴定为与 AD 表型显著正相关的蛋白质之一[42]，与其他 S100 蛋白一样，S100A6 在 AD 患者和 AD 小鼠模型中上调，并在脑灰质中围绕 Aβ淀粉样蛋白沉积的星形胶质细胞阳性簇中被发现[43]。据报道，S100A7 在轻度认知障碍患者中增加[44]，且在 AD 痴呆患者的大脑和脑脊液中显著增加[45]。S100A8/A9 与“非淀粉样”α-分泌酶 ADAM-10 之间也存在联系，S100A8/A9 在过度表达 ADAM-10 的 AD 小鼠模型中表达较低[46]。AD 患者脑中可见 S100A9 沉积物形成，研究指出在体外，S100A9 能够形成类似淀粉样结构的聚合结构并结合淀粉样荧光团，通过两步成核发生聚合反应，导致 S100A9 的初始错误折叠和β片形成[47]。S100A12 是 AD 背景下研究最少的 S100 蛋白，仅有一项研究报告了 S100A12 六聚体可能与散发性 AD 患者大脑中的老年斑块、反应性胶质细胞和神经元相关[48]。综上，这些研究结果表明 S100 蛋白在一些 AD 相关的生理病理过程中起着关键作用。

2. S100 蛋白与帕金森病

与 AD 类似，帕金森病（Parkinson’s disease，PD）也为一种进行性神经退行性疾病，病理特征为α-突触核蛋白在皮质或脑干区聚集，患者常表现出运动迟缓、

僵硬和静息时产生震颤[49]。据研究，低表达的 S100B 蛋白通过减少小胶质细胞增生和 TNF-α的表达而起到神经保护作用；而 S100B 高表达则激活促炎细胞因子的释放，导致多巴胺能神经元的损伤，临床报道帕金森病患者死后黑质中 S100B 蛋白的水平明显高于正常组[50]。S100A6 最早被发现并在脑中大量表达的靶标之一是 CacyBP/SIP 蛋白，结合基于质谱学的定量蛋白质组学，已证明 AD 患者嗅球中 CacyBP/SIP 表达减少，提示 S100A6 在 AD 中扮演重要角色[51]。

3. S100 蛋白与创伤性脑损伤

创伤性脑损伤（traumatic brain injury, TBI）在各类创伤性损伤中最为常见，多与其他各种创伤合并存在，以高死亡率、高致残率而位居各类创伤之首。目前在 TBI 的研究中，CSF 和血液中的 S100B 水平被认为是由心血管疾病或创伤性损伤引起的急性脑损伤的可靠生物标志物，同时也与疾病的预后密切相关[52]。在 S100B 基因敲除小鼠及使用 S100B 拮抗剂小鼠的受控皮质撞击模型中，两种干预措施均显著降低了 TBI 诱发的病变体积，改善了记忆功能，并减弱了小胶质细胞活化；S100B 拮抗剂组还显著减少了感觉运动缺陷并改善了皮层中的神经元存活，提示 S100B 在 TBI 诱导的神经炎症、细胞损伤和神经功能障碍中发挥重要作用，可能为 TBI 的潜在治疗靶标[53]。另有学者研究发现，在创伤性脑损伤后，大鼠海马和皮层组织中 S100A6 蛋白明显下降至最低值，随着创伤后时间推移，S100A6 蛋白表达又逐渐回升，直至伤后一周时与正常对照组比较差异无统计学意义。这样的先降低而后回升的规律，与临床上 TBI 患者意识深而后恢复，认知功能降低而后回升的规律一致，提示脑组织中 S100A6 的变化与神经细胞的损伤与恢复有一定关系[54]。

4. S100 蛋白与脑肿瘤/脑转移瘤

越来越多研究表明 S100B 在脑肿瘤及脑转移瘤方面具有诊断、治疗及预后等临床应用价值。胶质瘤是中枢神经系统最常见的癌症，S100B、S100A11、S100A16 在胶质瘤组织中表大水平显著高于正常组织，表明 S100 家族基因可能对脑胶质瘤患者具有重要的预后价值[55]。黑色素瘤为最容易导致脑转移的恶性肿瘤之一，高达 75%的晚期黑色素瘤患者在死亡时存在脑转移。在黑色素瘤中，用 siRNA 减少 S100B 的表达可激活 p53 转录活性和 p53 依赖性凋亡途径，表明 S100B 可能通过下调抑癌蛋白 p53 而促进癌症发展[56]。在非小细胞肺癌

（non-small cell lung cancer，NSCLC）脑转移的患者中，血清 S100B 与 NSCLC 脑转移显著正相关，且血清高水平 S100B 与脑转移瘤数量和预后显著相关，表明血清 S100B 是早期检测 NSCLC 脑转移的敏感和特异性标志物，可作为预后评估的监测工具[57]。

5. S100 蛋白与癫痫

癫痫是一种神经和全身性疾病，以反复发作为特征，与异常或过度的脑神经活动有关，目前其分子生物学机制仍未阐明，尚无有效的靶向治疗[58]。神经炎症在癫痫发生及其持续状态中起重要作用，其中先天免疫的关键受体 TLR4 在癫痫发生中有重要价值。报道，TLR4 在癫痫中的作用与其通路中配体的关系密切据，在刺激杏仁核诱导的癫痫大鼠模型中，大鼠大脑皮质区 S100A6 表达持续上调[59]；其次，S100A8 /A9 作为 TLR4 的内源性配体在 TLR4 介导的炎性通路上发挥重要生物学特性[60]。S100A8 /A9 所致内侧颞叶癫痫是一种常见的难治性癫痫，在急性、潜伏、慢性内侧颞叶癫痫的整个发展阶段，海马中 S100A8 /A9 的表达水平均高于正常人群。其原因在于 S100A8 /A9 与 TLR4 结合并触发 NF-κB 信号通路，促使 IL-1β产生，从而维持炎症微环境的持续存在，以持续刺激内侧颞叶皮质，使癫痫发作时间延长[61]。因此，S100A8 /A9 可能作为上调癫痫易感性的治疗靶点，但目前尚无通过拮抗 S100A8 /A9 控制癫痫的实验研究证据，故仍有待进一步研究。

三、S100 蛋白家族作为神经系统疾病治疗靶点的可能性

越来越多 S100 蛋白已被用作治疗多种疾病的潜在治疗靶点，S100 中和抗体和小分子抑制剂的开发已成为当前一些癌症、自身免疫性疾病和慢性炎症性疾病的治疗方法[62]。如 S100B 作为一种细胞内钙离子受体蛋白和分泌的神经肽，其表达加剧了 AD 的病理过程。与对照组相比，S100B 基因敲除的 AD 小鼠皮质斑块负荷、胶质增生和神经元功能障碍都显著减轻，提示 S100B 抑制剂可延缓阿尔茨海默病的进展[63]。S100 蛋白，特别是钙粒蛋白如 S100A8、A9 和 A12 等，在调节一系列疾病的免疫反应中发挥着重要作用，包括炎性关节炎、动脉粥样硬化和细菌感染等[64]。炎症性肠病常在肠道以外的器官系统引起并发症，在中枢神经系统可致患者认知障碍，而 S100A9 抑制剂帕奎莫特的预防性治疗可以减轻结肠和神经病变[65]。

许多 S100 蛋白的组织和血清水平与组织或局部炎症期间的疾病严重程度相关，针对受体和配体的功能阻断抗体已被广泛用作治疗多种疾病的治疗剂，应用在癌症和免疫疾病治疗的领域[66]。细胞外 S100 蛋白主要通过细胞受体信号传导介导许多病理条件下的炎症反应，因此，使用 S100 功能阻断抗体可能提供治疗这些条件的有效治疗策略，有助于为多种中枢神经系统疾病的治疗及预后判断提供新的方向。继续验证 S100 蛋白作为生物标志物在早期疾病检测和预后中的作用，以及开发基于抗 S100 治疗的新策略，将为治疗 S100 家族蛋白驱动的癌症、自身免疫性疾病和慢性炎症疾病等提供新的方法。

四、小结

综上，许多疾病状态与 S100 蛋白的表达水平改变有关，S100 蛋白作为中枢神经系统疾病发生发展过程中的关键参与者，已逐渐被应用为诊断和监测患者疾病的生物标志物，但仍需更广泛的临床研究来确定 S100 蛋白作为生物标记物的灵敏度和准确度。目前虽已在 S100 蛋白的细胞内外功能以及它们在炎症、癌症、自身免疫等疾病发生发展中的作用方面取得重大进展，但对 S100 寡聚体形成和翻译后修饰在调节细胞内外中的作用，及介导细胞外 S100 信号的细胞表面受体等方面的了解尚有不足。继续深入探索 S100 蛋白的相关信号通路、分泌途径以及翻译后修饰等具体机制将为中枢神经系统疾病或其他系统疾病研究进展提供新的见解。进一步优化安全有效的 S100 疗法，验证 S100 蛋白作为生物标志物在早期疾病检测和预后中的作用，并开发基于抗 S100 治疗的新策略将是未来努力的方向。

参考文献

[1] Moore B W. A soluble protein characteristic of the nervous system[J]. Biochemical and Biophysical Research Communications, 1965,19(6):739-744.

[2] Isobe T, Okuyama T. The amino-acid sequence of the alpha subunit in bovine brain S-100a protein[J]. Eur J Biochem, 1981,116(1):79-86.

[3] Kozlyuk N, Monteith A J, Garcia V, et al. S100 Proteins in the Innate Immune Response to Pathogens[J]. Methods Mol Biol, 2019,1929:275-290.

[4] Donato R, Cannon B R, Sorci G, et al. Functions of S100 proteins[J]. Curr Mol Med, 2013,13(1):24-57.

[5] Singh P, Ali S A. Multifunctional Role of S100 Protein Family in the Immune System: An Update[J]. Cells, 2022,11(15).

[6] Heizmann C W. S100 proteins: Diagnostic and prognostic biomarkers in laboratory medicine[J].

Biochim Biophys Acta Mol Cell Res, 2019,1866(7):1197-1206.
[7] Allgower C, Kretz A L, von Karstedt S, et al. Friend or Foe: S100 Proteins in Cancer[J]. Cancers (Basel), 2020,12(8).
[8] Gonzalez L L, Garrie K, Turner M D. Role of S100 proteins in health and disease[J]. Biochim Biophys Acta Mol Cell Res, 2020,1867(6):118677.
[9] Averill M M, Barnhart S, Becker L, et al. S100A9 differentially modifies phenotypic states of neutrophils, macrophages, and dendritic cells: implications for atherosclerosis and adipose tissue inflammation[J]. Circulation, 2011,123(11):1216-1226.
[10] Rahimi F, Hsu K, Endoh Y, et al. FGF-2, IL-1beta and TGF-beta regulate fibroblast expression of S100A8[J]. FEBS J, 2005,272(11):2811-2827.
[11] Ingersoll M A, Spanbroek R, Lottaz C, et al. Comparison of gene expression profiles between human and mouse monocyte subsets[J]. Blood, 2010,115(3):e10-e19.
[12] Wang Q, Williamson M, Bott S, et al. Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer[J]. Oncogene, 2007,26(45):6560-6565.
[13] Wang X H, Zhang L H, Zhong X Y, et al. S100A6 overexpression is associated with poor prognosis and is epigenetically up-regulated in gastric cancer[J]. Am J Pathol, 2010,177(2):586-597.
[14] Gilston B A, Skaar E P, Chazin W J. Binding of transition metals to S100 proteins[J]. Sci China Life Sci, 2016,59(8):792-801.
[15] Donato R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles[J]. Int J Biochem Cell Biol, 2001,33(7):637-668.
[16] Spratt D E, Barber K R, Marlatt N M, et al. A subset of calcium-binding S100 proteins show preferential heterodimerization[J]. FEBS J, 2019,286(10):1859-1876.
[17] Melville Z, Aligholizadeh E, McKnight L E, et al. X-ray crystal structure of human calcium-bound S100A1[J]. Acta Crystallogr F Struct Biol Commun, 2017,73(Pt 4):215-221.
[18] Calderone V, Fragai M, Luchinat C. Reviewing the Crystal Structure of S100Z and Other Members of the S100 Family: Implications in Calcium-Regulated Quaternary Structure[J]. Methods Mol Biol, 2019,1929:487-499.
[19] Bagur R, Hajnoczky G. Intracellular Ca(2+) Sensing: Its Role in Calcium Homeostasis and Signaling[J]. Mol Cell, 2017,66(6):780-788.
[20] Madureira P A, O'Connell P A, Surette A P, et al. The biochemistry and regulation of S100A10: a multifunctional plasminogen receptor involved in oncogenesis[J]. J Biomed Biotechnol, 2012,2012:353687.
[21] Chen H, Xu C, Jin Q, et al. S100 protein family in human cancer[J]. Am J Cancer Res, 2014,4(2):89-115.
[22] Carvalho A, Lu J, Francis J D, et al. S100A12 in Digestive Diseases and Health: A Scoping Review[J]. Gastroenterol Res Pract, 2020,2020:2868373.
[23] Wolf R, Ruzicka T, Yuspa S H. Novel S100A7 (psoriasin)/S100A15 (koebnerisin) subfamily: highly homologous but distinct in regulation and function[J]. Amino Acids, 2011,41(4):789-796.
[24] Son E D, Kim H J, Kim K H, et al. S100A7 (psoriasin) inhibits human epidermal differentiation by enhanced IL-6 secretion through IkappaB/NF-kappaB signalling[J]. Exp Dermatol, 2016,25(8):636-641.
[25] Krop I, Marz A, Carlsson H, et al. A putative role for psoriasin in breast tumor progression[J]. Cancer Res, 2005,65(24):11326-11334.

[26] Goyette J, Geczy C L. Inflammation-associated S100 proteins: new mechanisms that regulate function[J]. Amino Acids, 2011,41(4):821-842.

[27] Dahlmann M, Kobelt D, Walther W, et al. S100A4 in Cancer Metastasis: Wnt Signaling-Driven Interventions for Metastasis Restriction[J]. Cancers (Basel), 2016,8(6).

[28] Riuzzi F, Sorci G, Beccafico S, et al. S100B engages RAGE or bFGF/FGFR1 in myoblasts depending on its own concentration and myoblast density. Implications for muscle regeneration[J]. PLoS One, 2012,7(1):e28700.

[29] Prica F, Radon T, Cheng Y, et al. The life and works of S100P - from conception to cancer[J]. Am J Cancer Res, 2016,6(2):562-576.

[30] Pan S C, Li C Y, Kuo C Y, et al. The p53-S100A2 Positive Feedback Loop Negatively Regulates Epithelialization in Cutaneous Wound Healing[J]. Sci Rep, 2018,8(1):5458.

[31] Basnet S, Sharma S, Costea D E, et al. Expression profile and functional role of S100A14 in human cancer[J]. Oncotarget, 2019,10(31):2996-3012.

[32] Fernandez-Fernandez M R, Veprintsev D B, Fersht A R. Proteins of the S100 family regulate the oligomerization of p53 tumor suppressor[J]. Proc Natl Acad Sci U S A, 2005,102(13):4735-4740.

[33] Aubrey B J, Kelly G L, Janic A, et al. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression?[J]. Cell Death Differ, 2018,25(1):104-113.

[34] Masters C L, Bateman R, Blennow K, et al. Alzheimer's disease[J]. Nat Rev Dis Primers, 2015,1:15056.

[35] Wang X, Wang W, Li L, et al. Oxidative stress and mitochondrial dysfunction in Alzheimer's disease[J]. Biochim Biophys Acta, 2014,1842(8):1240-1247.

[36] Marshak D R, Pesce S A, Stanley L C, et al. Increased S100 beta neurotrophic activity in Alzheimer's disease temporal lobe[J]. Neurobiol Aging, 1992,13(1):1-7.

[37] Yeh C W, Yeh S H, Shie F S, et al. Impaired cognition and cerebral glucose regulation are associated with astrocyte activation in the parenchyma of metabolically stressed APPswe/PS1dE9 mice[J]. Neurobiol Aging, 2015,36(11):2984-2994.

[38] Simpson J E, Ince P G, Lace G, et al. Astrocyte phenotype in relation to Alzheimer-type pathology in the ageing brain[J]. Neurobiol Aging, 2010,31(4):578-590.

[39] Gruden M A, Davidova T B, Malisauskas M, et al. Differential neuroimmune markers to the onset of Alzheimer's disease neurodegeneration and dementia: autoantibodies to Abeta((25-35)) oligomers, S100b and neurotransmitters[J]. J Neuroimmunol, 2007,186(1-2):181-192.

[40] Zimmer D B, Chaplin J, Baldwin A, et al. S100-mediated signal transduction in the nervous system and neurological diseases[J]. Cell Mol Biol (Noisy-le-grand), 2005,51(2):201-214.

[41] Afanador L, Roltsch E A, Holcomb L, et al. The Ca2+ sensor S100A1 modulates neuroinflammation, histopathology and Akt activity in the PSAPP Alzheimer's disease mouse model[J]. Cell Calcium, 2014,56(2):68-80.

[42] Wruck W, Schroter F, Adjaye J. Meta-Analysis of Transcriptome Data Related to Hippocampus Biopsies and iPSC-Derived Neuronal Cells from Alzheimer's Disease Patients Reveals an Association with FOXA1 and FOXA2 Gene Regulatory Networks[J]. J Alzheimers Dis, 2016,50(4):1065-1082.

[43] Boom A, Pochet R, Authelet M, et al. Astrocytic calcium/zinc binding protein S100A6 over expression in Alzheimer's disease and in PS1/APP transgenic mice models[J]. Biochim Biophys Acta, 2004,1742(1-3):161-168.

[44] Mueller C, Zhou W, Vanmeter A, et al. The heme degradation pathway is a promising serum

biomarker source for the early detection of Alzheimer's disease[J]. J Alzheimers Dis, 2010,19(3):1081-1091.
[45] Qin W, Ho L, Wang J, et al. S100A7, a novel Alzheimer's disease biomarker with non-amyloidogenic alpha-secretase activity acts via selective promotion of ADAM-10[J]. PLoS One, 2009,4(1):e4183.
[46] Prinzen C, Trumbach D, Wurst W, et al. Differential gene expression in ADAM10 and mutant ADAM10 transgenic mice[J]. BMC Genomics, 2009,10:66.
[47] Iashchishyn I A, Sulskis D, Nguyen N M, et al. Finke-Watzky Two-Step Nucleation-Autocatalysis Model of S100A9 Amyloid Formation: Protein Misfolding as "Nucleation" Event[J]. ACS Chem Neurosci, 2017,8(10):2152-2158.
[48] Shepherd C E, Goyette J, Utter V, et al. Inflammatory S100A9 and S100A12 proteins in Alzheimer's disease[J]. Neurobiol Aging, 2006,27(11):1554-1563.
[49] Gatev P, Darbin O, Wichmann T. Oscillations in the basal ganglia under normal conditions and in movement disorders[J]. Mov Disord, 2006,21(10):1566-1577.
[50] Sathe K, Maetzler W, Lang J D, et al. S100B is increased in Parkinson's disease and ablation protects against MPTP-induced toxicity through the RAGE and TNF-alpha pathway[J]. Brain, 2012,135(Pt 11):3336-3347.
[51] Lachen-Montes M, Gonzalez-Morales A, Iloro I, et al. Unveiling the olfactory proteostatic disarrangement in Parkinson's disease by proteome-wide profiling[J]. Neurobiol Aging, 2019,73:123-134.
[52] Ye H, Wang L, Yang X K, et al. Serum S100B levels may be associated with cerebral infarction: a meta-analysis[J]. J Neurol Sci, 2015,348(1-2):81-88.
[53] Kabadi S V, Stoica B A, Zimmer D B, et al. S100B inhibition reduces behavioral and pathologic changes in experimental traumatic brain injury[J]. J Cereb Blood Flow Metab, 2015,35(12):2010-2020.
[54] 房博. 创伤性脑损伤后大鼠脑组织钙结合蛋白 S100A6 的表达及意义[D]. 延安大学, 2013.
[55] Hu Y, Song J, Wang Z, et al. A Novel S100 Family-Based Signature Associated with Prognosis and Immune Microenvironment in Glioma[J]. J Oncol, 2021,2021:3586589.
[56] Lin J, Yang Q, Wilder P T, et al. The calcium-binding protein S100B down-regulates p53 and apoptosis in malignant melanoma[J]. J Biol Chem, 2010,285(35):27487-27498.
[57] Chen L, Hu X, Wu H, et al. Over-expression of S100B protein as a serum marker of brain metastasis in non-small cell lung cancer and its prognostic value[J]. Pathol Res Pract, 2019,215(3):427-432.
[58] Yuen A, Keezer M R, Sander J W. Epilepsy is a neurological and a systemic disorder[J]. Epilepsy Behav, 2018,78:57-61.
[59] Jurewicz E, Bednarczyk J, Bot A, et al. Status epilepticus induces long lasting increase in S100A6 expression in astrocytes[J]. Neurochem Res, 2013,38(9):1941-1948.
[60] Wang S, Song R, Wang Z, et al. S100A8/A9 in Inflammation[J]. Front Immunol, 2018,9:1298.
[61] Gan N, Yang L, Omran A, et al. Myoloid-related protein 8, an endogenous ligand of Toll-like receptor 4, is involved in epileptogenesis of mesial temporal lobe epilepsy via activation of the nuclear factor-kappaB pathway in astrocytes[J]. Mol Neurobiol, 2014,49(1):337-351.
[62] Bresnick A R. S100 proteins as therapeutic targets[J]. Biophys Rev, 2018,10(6):1617-1629.
[63] Roltsch E, Holcomb L, Young K A, et al. PSAPP mice exhibit regionally selective reductions in gliosis and plaque deposition in response to S100B ablation[J]. J Neuroinflammation,

2010,7:78.
[64] Foell D, Roth J. Proinflammatory S100 proteins in arthritis and autoimmune disease[J]. Arthritis Rheum, 2004,50(12):3762-3771.
[65] Talley S, Valiauga R, Anderson L, et al. DSS-induced inflammation in the colon drives a proinflammatory signature in the brain that is ameliorated by prophylactic treatment with the S100A9 inhibitor paquinimod[J]. J Neuroinflammation, 2021,18(1):263.
[66] Tardif M R, Chapeton-Montes J A, Posvandzic A, et al. Secretion of S100A8, S100A9, and S100A12 by Neutrophils Involves Reactive Oxygen Species and Potassium Efflux[J]. J Immunol Res, 2015,2015:296149.

Environmental Commentary: Paradox in human versus nature

Yue Rong, Ph.D. (SCCAEPA president 2006 – 2009 & CESASC president 2015)

First of all, I would like to congratulate the USA Chinese Scholars Association (USACSA) for its journal of Academic Prospects serving scientific community and Chinese American community. Thank all of Association colleagues at USACSA for making this Journal available.

In last several years, people around the world start building up some momentum to curb global environmental problems such as climate change and alternative green energy, and trying to work together for solutions. However, the environmental problem has entered a more complicated social, economic, and political chapter. To cope with the problem, we may have to balance the economic development and environmental protection in order not to cause social and economic conflict and turmoil.

Yes. We are the human being living on earth, and responsible for all of pollutions, global climate changes, and the recent pandemic, etc. All of the environmental problems seem to point to the same reason of increasing human demand in living. Today, we have more people than ever in the human history. Global population had grown from 3 billion in 1960 to 6.8 billion in 2010 and almost 7.8 billion in 2020. Energy consumption in the U.S. has more than doubled, and the municipal waste in the U.S. has tripled during that 60 years. These increases are no doubt translating into environmental burden and damages.

At the United Nations 2009 Copenhagen Climate Change Conference, a Chinese delegate argued that China's biggest contribution to the solution of global environmental program is its population control policy. According to him, China has been implementing family planning policy (to reduce birth rate) for the last 3 decades in counting to reduce approximately a total of 400 million people, which is translated to reduce about 2 billion tons of CO_2. From that perspective, the environmental pollutions and global problems may all be attributed to human growth.

Maybe the human growth is the primary reason for our global environmental problems? If we do not see it this way, our new efforts, technologies, energy, resources, and new ideas may quickly be outweighed and overwhelmed by the quicker growth of human population. However, if we see it this way, the population control will become a social and political conflict among religious believers, which may cause road blocks to slow down environmental protection and solution. Ironically, China has changed its population control policy to relax the birth control in 10 years after the above-mentioned Copenhagen Climate Change Conference. This is because China recognizes this kind of population control policy could not sustain in a society.

This is, I call it, an environmental paradox between human beings and the nature. Human beings may be the most intelligent living things on earth. They will and have right to reproduce. However, one day, we may run out resources and space to hold all of us on earth. Reckless growth and endless consumptions are definitely unsustainable on earth. Therefore, thinking about a long term solution now should be commensurate to our current environmental efforts. For us, we should not fight with the nature and try to take more from it, rather we should start behaving, protecting the environment, saving energy, stopping wasting resources, and reusing all we can. So, what do you do for the environment today?

The commentary I wrote and submitted to your journal was the one I wrote for SCCAEPA 30-year anniversary journal. SCCAEPA is Southern California Chinese American Environmental Protection Association (南加华人环保协会).

Therefore, please indicate this commentary is a repost from SCCAEPA Journal.

中国教育：成就、问题与展望

陈岳云 美国洛杉矶西来大学

自 1978 年改革开放以来，中国教育事业发生了翻天覆地的变化，取得了巨大的进步和成果。正本清源，教育回到了教书育人，传授知识，培育人格，探索自然，科学及其他未知的世界，长见识增技能的本职，而不是阶级斗争，与人奋斗其乐无穷的舞台。

尊重知识，尊重知识分子，尊重教师，已成为社会的文化。刻苦学习，积极努力向上，力争上游，已成为普遍的社会风气。教育特别是高考已成为青年人实现梦想成材的必经之路，更是无数穷人的孩子翻身解放，改变人生道路的主要途径。高考和大学教育，选拔培养造就了无数人才，亦为改革开放和现代化建设事业，奠定了人才基础。教育体制改革，教育政策的制订和实施，特别是<<九年义务教育制>>和<<高等教育法>>的实施，保证了教育事业稳步健康发展。

四十多年来，中国经济文化科技等领域的全面发展和快速变化，得益于思想解放，得益于伟大的改革开放，也直接来自于教育事业的丰硕成果。在高考改革后培养的各种人才已成为中国各领域的中流砥柱。教育事业的进步保证并促进了中国现代化建设事业的成功。

然后与整个社会的主要矛盾-人民日益增长的美好生活需要和不平衡不充分的发展之间的矛盾一样，教育领域也存在严重的不平衡不充分的教育发展不能充分满足人民大众对教育的日益增长的要求，包括对各类教育的多样性高质量的要求。教育资源分配分布的不均衡性，更使这种矛盾复杂化，并影响社会经济的稳定发展。老百姓所讲的三座大山，教育不仅是其中之一，还影响决定其他因素，比如住房问题，也与教育相关。因此深刻认识教育问题的本质，并采取积极可行的措施，不仅是解决教育本身的问题，也能化解其他社会问题。

教育战线取得的成就有目共睹，而问题也累积许久，日益明显，否则国家也不会采取如此严厉措施，来整顿解决民众所关注的教育问题。然后，这些新政和措施是否符合中国实情，是否能治标亦治本，也是许多人所思考和关心的。不破不立，矫枉必须过正，是革命的哲学和实践。现代化建设需要这种敢于直面问题，敢于争斗的勇气，但更需实事求是，科学分析，精准施策，高瞻远瞩的顶层规划。

关于中小学教育

中小学教育中的主要问题，是学业压力大，课外辅导补课重，竞争激烈，考试导向式的教学和训练，使学习不是培养兴趣爱好，不是使人的知识和素养的全面成长。功利主义盛行，排名至上，学校排名，学生排名，青少年失去童年的天真，美好和快乐。中小学教育资源分布分配不均，城市中对农民工子女排拆，以户藉定教学机会，准入门槛，而不是以住地。由于出生率下降，青少年人数大幅下降，许多中小学招不到足够的人，难以为继。现在准备减少普通初高中，早日分流，安排更多的人去读初、高中级的职业学校，既有一定的合理性，但势必造成普通高中，特别是优质高中入学更困难，竞争更激烈，内卷势必更严重！

虽然过度竞争过早竞争，内卷化对教育事业健康发展不利，但适当的竞争是必要的有益的。人的潜力能力提高往往是在内在动力和外在压力下实现的，特别是适度的竞争环境，相互激励和促进，是使人，特别是青少年不满足于已取得的成绩，设定更高更具挑战性的目标，从而更加投入和努力的有效途径。矫枉过正，取消各种考试，包括升学考试，取消所有的重点学校及重点班等，不符合教育特别是人才培养和成长的规律，更会影响创新型人才，优秀人才的成长和培育。

其实，在美国这样以居住地为择校原则的教育体制下，许多学区仍会有一所重点学校，一般是初中高中联读。而进入这样的重点学校，需小学毕业前，统一考试，有名额分配的择优录取。比如，此重点中学录取 100 优秀小学毕业生，学区共有 7 所小学。其中 70 名额分配给这 7 所学校，按照学生人数，每所学校有 7-12 最低名额，并按照考试成绩，并参考平时学习成绩等各方面表现，由每所学校决定或推荐。而所剩的 30 个名额，则由学区或重点学校，在剩余的学生中择优录取。最后结果是，最差的小学，只有 7 位学生升入重点中学，而好的小学，可能有 15 学生顺利进入此重点中学。另外美国许多地方设立联合学区，若干个小城市联合办一个学区，统管所在地区的所有小学或所有中学或所有中小学。这样规模经济高，并能更具条件设立重点学校。

考虑到中国的国情，为避免人情等因素的可能影响，此类重点学校的录取可严格按照加权总分，比如统一考试成绩占 60%，学生小学时的平均成绩占 40%。中国县或县级市，应具备条件设立一重点初中和高中。而中大型城市，可根据人口及学校数量，以一区或若干区办一重点初高中。大众对高质量教育的要求是非常强烈且刚性的。取消所有重点初高中，会促进更多父母把小孩送到相应的私立学校甚至国外。

关于考试，过度密集的各种考试，会误导学生，更使学生成为应试教育的受害者，浪费大量青春时光，并导致许多学生只死记硬背，缺乏独立思考,创造性思维的习惯和能力,害人误国。但不能从一个极端走向另一个极端。明文禁止必要的考试，不仅使教育缺少行之有效的衡量其成效的工具和方法，也使学生和老师，缺乏必要的教与学的动力和方向。况且科学研究表明，适当的考试可提高学生的学习成果。

教育的改革，必须充分考虑中国的国情，充分考虑社会大众对教育问题的忧虑和本质要求。不仅要在主体上以满足极大多数民众的需求和要求为工作的核心，出发点和重心，也要充分考虑到部分

民众的特殊情况和特殊需求。比如美国许多地区有专门学校，招收特殊学生，象自闭症的学生等。应该承认和意识到，有各种天才型及非常优秀的青少年，他们需要特殊的课程, 教学安排，早学，快学，多学。合理安排满足这类学生的需求，也是社会的责职和光荣使命，因为下一个钱学森或钱钟书可能就来自于他们。中国的乔布斯，马斯克或比尔盖茨更可能出自于他们。

关于中小学生课外辅导补习，目前所采取的一刀切的限制措施，是为了回应大众多年的忧虑和深度关切，也是为了扼制资本的无序大量流入，扰乱辅导补习市场，扰乱干涉正常的中小学教育，避免家长和学生更多更深更广地纠入这种不必要的内卷。这也是为了使青少年能正常健康地成长。但任何一个市场是由供需俩侧所决定的。目前所采取的这些措施，只针对供给侧的问题，市场需求仍存在，以至出现各种变相的辅导和补习。社会文化, 价值导向上不改变，社会用人报酬等方面实际不改变，很难从根本上解决问题。一方面提倡鼓励随缘而安，以住地择校。而所有顶尖大学仍主要面向若干重点高中密集招生。更显著的是，包括北京和上海这样的非常吸引人的超大城市，公开发布政策说，只对若干重点大学毕业生，自动开放户籍。这是什么样的社会导向？

韩国作为有与中国一样非常重视小孩教育的国家，也曾经历了教育高度内卷，补习班盛行，家庭用于教育的开支日益居高，然后政府采取严厉限补政策等。但根本问题似乎从未解决。真心希望中国有关新政能长期有效，术到病除。

九年义务教育制已实施几十年，成绩斐然。若干城市和地区，已在执行十二年义务教育，并初显成效。应有更多的省市试行十二年义务教育制。中高收入的家庭，应能承担得起小孩高中学习的费用，而且一般不会让他们不完成高中的学习; 所以向推动十二年义务教育制迈进，应重点支助低收入及贫困家庭与低收入及贫困地区, 首先实现十二年义务教育制。

关于职业教育

中国各类职业院校，几十年来经过各种调整，取得了长足的进步；办学特色更明确，方向更清楚。院校总数，在校学生及毕业生总数达到了前所未有的水平。目前中国本、专科学生，在校总人数，招生数及每年毕业生数, 已基本接近、持平。

职业院校主要是为社会培养生产岗位、管理岗位、服务岗位等面向一线工作的技术应用型专业人才。主要有两种类型, 为学生提供制造业工作的职业课程, 包括贸易和工业项目，如建筑、机械和修理、精密生产、运输和材料搬运等；与为学生在服务行业和信息产业工作做好准备的职业课程包括医疗卫生、技术和通信等。各类职业技术学院，满足社会需求，培养了众多技术型工程型的能工巧匠，为中国制造业的进步乃至整个经济的发展作出了特殊的贡献。制造业强则国强，制造业兴则国兴。而制造业尤其是中高端产业的进步，离不开大批优秀的技工。因此加强提高职业院校特别是职业技术院校，关系到整个工业及整个经济的平稳持续发展。中国似乎在试探是否向德国学习，提高在中职高职院校学生的比例，并尽早分流，从而培养更多有社会需求，实用型的有技能的毕业生，为现代化建设特别是制造业迈向中高端发展奠定人才基础。

就国际比较而言，中国职业教育的规模与整个经济的规模，应已大致相适宜。今后中国职业教育，特别是职业技术教育，应以提高质量，培养满足社会需求的各类毕业生为主要目标，而不应偏面追求量的快速增长。特别是私立或民营的各类职业院校，不能盲目兴建及扩建，否则误人子弟，过度内卷竞争，造成社会资源的极大浪费。另外有的省区，没有必需的足够大的相关产业，特别是制造业，若有太多的这类职业学院，学生无法实习，毕业时，更无合适的工作。自然无法生存发展。

美国高中毕业生中，直接上四年制本科的约 45%，二年制约 25%。德国多年前，每年加入职业学院学习的新生大大超过四年制的大学生，但从 2014 年开始,发生了变化,现在入学职业教育的人数在持续下降而进入四年制本科学习的在增加中。主要原因是工资报酬和失业率。中国目前，职业教育当然仍需要改进提高,并有进一步扩大发展的潜力，但过渡强调和扩展职业学院,会造成社会资源的极大浪费，更耽误了大批年青学子的前途和事业。许多职业学院，特别是职业技术学院在工科或工程类的专业方面有优势， 在制造和技术领域有实际操作经验和技能,其毕业生主要去生产制造企业。中国仍是世界制造中心，制造业对中国不仅重要仍有稳定发展的潜力。但经济和行业发展有其客观的规律，随着中国富裕程度和工资福利的提高，许多制造企业和行业的优势会逐渐下降甚至消失，从而转移到其他国家和地区，对各种人才的需求会显著下降。不认识到这些变化，而继续扩展职业学院，会造成各种严重后果。美国职业技术教育在 20 世纪中叶曾有较好的增长和发展,但随着制造业的外流及衰退，职业技术学院的市场需求越来越少，加上二年制公立社区学院的兴起，职业技术学院所剩廖廖无几。

总之，中国职业教育，尤其是职业技术院校，应注重强身增质而非扩量，切忌盲目发展。同时，随着新兴产业和技术的兴起和成长，职业技术学院也要增设新的学科和课程，如与人工智能，机器人，无人驾驶技术及现代医疗实验技术相关的学科，以更好地服务于社会，同时开拓扩展新的领域。

关于大学

中国高等教育过去几十年发展很快,规模巨大,每年大学毕业生约 800 万,每年研究生招生超 100 万，其中博士生 12 万。在校研究生超 300 万。每年有 70 多万研究生毕业，其中博士生近 7 万。中国 STEM 科技工程数学类的大学生已超美国规模。在美外国留学生

100 多万，其中中国留学生超 37 万。中国到海外留学生总数约 70 万。目前在中国有海外留学生约 50 万,亚洲占近 60%，但读学位比例较低,52.44%。

中国高等教育中，一是尖端世界超一流大学缺失，前 10 名中缺席,前 20 名中,有的排名榜中能列入。尽管科研进步神速,但离顶尖大学仍有很大的距离,特别是每年诺贝尔奖的缺席,与邻国日本亦有很大的差距。美国 25 岁或以上人中，有大学本科学历的约 35%。高中毕业生中，直接上四年制本科的约 45%，所以美国人口中有四年制本科文凭的比例仍在逐年上升。美国联邦政府教育开支占当年 GDP 6.2%，加上卅和地方政府，则应超过 10%。中国本科学历占总人口不足 4%，专科占约 14%， 本科占 25 岁及以上人口应在 6%左右,专科约 24%。所以中国仍应巩固发展本科教育。中国中央政府教育开支占 GDP 4%多点,近几年稳定并有所增加,但比例仍不高,且分配不均,教育资源，特别比较优秀的大学,高度集中,对大众不公,也不利地区间平衡发展。

目前私立大学不少，但总数及学生总规模仍不高，特别是有较高质量和水平的私立大学缺失。是否应彻底全面解放思想。允许鼓励青年人到国外留学，接受资本主义思想、制度的教育，为什么不能容许这些学生在中国接受优质私立大学的教育?

对大学在科研，经济社会发展中的作用，片面要求大学为经济社会建设的主战场直接服务，解决紧迫的科技经济问题，直接或间接地导致大学，特别是研究型大学的长远规划和工作重点的偏离。大学特别是顶尖的大学，应该立足于，着重于研究发现科学规律，从事于众多的基础科学研究，重大技术的发明和创造，也应从事一定程度的应用研究和开发，帮助解决实际问题，但工作研究重点和中心不能偏离。

大学文化, 领导及管理中，官本位主义盛行，教授成功的标志，不是本专业的成果，而往往是能否升官，当院长、校长。学校缺乏足够的自主权，学校管理缺乏民主制，教授和学生没有什么参与权，学术自由没有足够保证，甚至让学生举报老师是否有出格言行。有的教授，教书不育人, 甚至根本不教书, 忙于捞钱, 扬名, 缺师德甚至人格品德，败坏了校园风气，毒害甚至残害学生。同时大学过度的商业导向，唯利是图，一切向钱看，像名利场，缺乏价值观, 社会公理, 缺乏社会良知观和作为。

大学评鉴考核制度标准化, 公式化, 简单化, 细致化, 似乎公平科学合理但产生许多矛盾和问题，包括短期科研行为导向, 行政控制能力增加, 弄虚作假猛增, 大量时间精力浪费于申报评鉴及达标上, 对提升整体科研水平和能力, 特别是对创新型, 基础科学的研究, 没有正面作用而起反作用, 与创建世界一流大学的目标不符。

教授中近亲繁殖普遍且严重, 似乎有利于团队的形成及和谐发展, 但排拆不同意见, 小团体, 小帮派, 缺乏新鲜思想, 影响整体科研水平的多样化发展和提高。学阀学霸当道，阻碍年轻学者的正常成长。

在学生培养中，过度注意强调专业知识技能和水平，对人文，科学，社会科学没有足够的要求；特别是培养了许多精致, 精明有能力的利己主义者, 缺乏社会担当, 没有雄心壮志, 浪费了其优秀才华和能力，未成大器。

总之, 中国高等教育, 要继续解放思想, 深化改革, 改变学风教风, 保证学术自由，鼓励独立思考精神，提高学生创造性思维的能力，减少对教授不必要的考核和评鉴。需要指出，并引起足够重视的是，改革并非必须市场化。中国四十多年的改革开放，逐步确立了有中国特色的市场经济，推动和促进经济和社会的全面进步。但教育包括大学，带有公共产品/服务的属性, 而非私有产品, 所以教育产品

不能完全依赖市场,不能推行市场化的改革。这是简单的经济学原理和市场规律。

十年树木，百年树人。教育政策影响广大民众，直接影响当代青少年的成长和发展,并对社会产生长远的影响。教育事业的发展方向和目标必须与整个社会的导向相一致，来自于社会服务于社会。但同时必须认识到，教育事业像任何一个学科或事业一样，有其独特性。教育的改革,政策的制定必须了解掌握其规律和特点，科学客观地分析存在的各种主要问题，解剖其原因及起因等，充分认识解决问题的条件和要求，探索解决问题的可能的不同方案，及具体实施的方法和步骤。行有行规，业有专术，应充分征求各级专家学者的独立意见，同时广泛听取大众的意见和建议，集思广益，发扬民主集中制，使各项政策更精准，更合实际，真正能解决逼切的问题，并能获得大众的支持。

各地教育水平，经济发展水平不同，社会文化和习惯等也有异，在统一整个国家基本教育制度和政策的前提下，应允许及鼓励各地根据当地实情，采取有别的具体措施，试验解决教育热点问题的不同方案。中国的国情不同，但各国的教育也有共性-满足民众需求,培养造就各种人才，探索了解未知的世界。所以中国也应更好更多地学习了解各国在教育改革和发展中的经验和教训，少走弯路。

全球化是时代趋势，教育的国际化发展也是不可逆转的潮流。中国教育要继续走国际化全球化的道路。要继续对外开放，扩大国际交流与合作。不仅要继续鼓励支持更多的青年人，到国外留学深造，支持鼓励更多的外国人到中国留学攻读学位，也要吸引并留住更多的世界各国优秀人才，到中国来从事科研和教育工作。中国高校中，师资的国际化程度极低，有待加强提高。

教育是崇高的事业，“为天地立心，为生民立命，为往圣继绝学，为万世开太平”。崇高的事业需要崇高的思想和并有愿为之奉

献的崇高的人，包括师生和各种职员等。而这一切，在市场经济中，在功利化盛行的社会中，往往时有矛盾和冲突。这也是当代教育面临困境的主要原因。人无完人，金无足赤。教育制度，教育事业也必然在异化，进化中逐步改变，与时代同行，伴历史齐进。相信中国的教育事业定能在进一步改革开放，经济和社会全面发展的大潮中稳健发展, 不负大众, 不负历史, 不负时代!

美国华裔教授学者协会学术期刊

USA CHINESE SCHOLARS ASSOCIATION

Los Angeles, California, USA, September 12, 2023