核酸条形码应用
—MAP-seq 和 BRIC-seq 技术的原理与研究现状

2024 年生物信息学课程第 12 组结课论文

小组成员：成修屹，王韵淇，凌翔宇

摘要：核酸条形码起源于 Paul Hebert 教授于 2003 年提出的 DNA 条形码（DNA barcoding）概念，该概念旨在利用一段较短的 DNA 序列作为物种快速鉴定的标记，并希望通过国际生命条形码计划建立 DNA 序列和生物物种之间的一一对应关系。DNA 条形码技术的步骤通常包括采集样品、提取 DNA、PCR 扩增、测序、序列校正及建立系统进化树等。由于核酸具有的独特生命性质，能够从概念上拓展到细胞条形码。核酸条形码在使用过程中需要兼顾多样性、稳定性和可识别性，这样的权衡过程从体内条形码的生成过程中可以体现。本文聚焦于 MAP-seq 及其衍生的 BRIC-seq 技术。这些技术通过引入核酸条形码来标记神经元，实现了对神经元跨区域投射的高精度追踪，通过测序绘制神经连接，为神经解剖学的研究提供了新的视角。在上述的背景下，本文记录了 BRIC-seq 技术的数据分析过程，最终尝试利用投射强度矩阵构建三维连接网络、脑联接图谱或热图等可视化图像，实现能够进一步研究相关环路的目的。随着基因测序技术的进步，核酸条形码技术将在生物多样性保护、生态监测、食品安全等方面发挥越来越重要的作用。DNA 条形码在物种鉴定与区分、食物链等生态学有关研究以及蛋白质条形码技术等方面的应用，展示了核酸条形码技术的广泛性和多样性。然而，核酸条形码在遗传差异不足、基因数据库不完善和高成本方面仍面临挑战。未来，随着测序技术进步和国际合作加深，该技术将在生物多样性保护、生态监测和精准医疗等方面发挥更大作用。通过对这些内容的深入探讨，不仅揭示了核酸条形码技术的原理和应用现状，也为其未来的发展方向提供了有益的参考。

关键词：核酸条形码，DNA 条形码，MAP-seq 技术，BRIC-seq 技术。

Nucleic Acid Barcoding Applications—

Principles and research status of MAP-seq and BRIC-seq techniques

CHENG Xiuyi. WANG Yunqi, LING Xiangyu

(*University of Chinese Academy of Sciences, Beijing* 100049,*China*)

Abstract Nucleic acid barcoding originated from the concept of DNA barcoding proposed by Prof. Paul Hebert in 2003, which aims to utilize a short DNA sequence as a marker for the rapid identification of species, and hopes to establish a one-to-one correspondence between DNA sequences and biological species through the International Barcode of Life Program. The steps of DNA barcoding technology usually include sample collection, DNA extraction, PCR amplification, sequencing, sequence correction and establishment of a phylogenetic tree. Nucleic acids are capable of conceptually expanding into cellular barcoding due to their unique life-giving properties. Nucleic acid barcodes need to balance diversity, stability and recognizability in their use, and such a trade-off process can be reflected in the generation of in vivo barcodes. This paper focuses on MAP-seq and its derived BRIC-seq techniques. These techniques enable high-precision tracking of neuronal projections across regions by introducing nucleic acid barcodes to label neurons, and mapping neural connectivity by sequencing, which provides new perspectives for neuroanatomical studies. Against the above background, this paper records the data analysis process of BRIC-seq technology, and ultimately attempts to use the projection intensity matrix to construct visual images such as 3D connectivity networks, brain connectivity maps, or heat maps, to achieve the purpose of being able to further study the relevant loops. With the progress of gene sequencing technology, nucleic acid barcoding technology will play an increasingly important role in biodiversity conservation, ecological monitoring, food safety, etc. The application of DNA barcoding in species identification and differentiation, food chain and other ecology-related studies, as well as protein barcoding technology, etc., demonstrates the extensiveness and diversity of nucleic acid barcoding technology. However, nucleic acid barcoding still faces challenges in terms of insufficient genetic differences, imperfect genetic databases and high costs. In the future, with the advancement of sequencing technology and the deepening of international cooperation, the technology will play a greater role in biodiversity conservation, ecological monitoring and precision medicine. The in-depth discussion of these contents not only reveals the principle and application status of nucleic acid barcoding technology, but also provides a useful reference for its future development direction.

Keywords Nucleic acid barcoding, DNA barcoding, MAP-seq, BRIC-seq.

一、核酸条形码简介

1.1 从条形码到 DNA 条形码（DNA barcoding）

条形码是一个由宽度不等的多个黑条与空白按照一定编码规则排列得到、可以表达一组信息的图形标识。除了商品名称、价格等，条形码还可表示物品制造厂家、生产日期、图书分类号、邮件起止地等众多信息，被广泛运用在商品流通、图书、邮政、银行等领域，可以说条形码的使用大大便利了我们的生活。

加拿大 Paul Hebert 教授受到商品条形码的启发，于 2003 年提出了 DNA 条形码（DNA barcoding）的概念，即利用一段较短的 DNA 序列作为物种快速鉴定的标记，并发起了国际生命条形码计划（International Barcode of Life Project），并希望以此建立 DNA 序列和生物物种之间一一对应的关系。DNA 条形码技术一经推出便引起广泛关注，Hebert 教授因此被誉为“DNA 条形码之父”。这一想法也和 1960 年代初期，二战老兵吉恩·罗登贝里（Gene Roddenberry）带来的著名科幻剧《*星际迷航*》中用来扫描和识别外星生命形式的手持式“三叉戟”设备相通（www.startrek.com）。

DNA 条形码，即一小段具有普适性的能够区分生物物种的 DNA 序列，研究人员可以基于这段 DNA 序列揭示生物多样性。动物中比较普适性的是线粒体细胞色素 C 氧化酶 1 基因 5’端约 650bp 的一段序列，其优点是容易扩增并有效区分物种。

图 1 商品条形码和 DNA 条形码的对比[1]

[1] Letchuman, Sarvananda. (2018). Short Introduction of Dna Barcoding. International Journal of Research. 05.

1.2 利用 DNA 条形码的步骤

利用 DNA 条形码进行物种鉴定大致可分为以下步骤：采集样品，做好采集时间、地点等记录并提取样品 DNA；利用相关 DNA 条形码引物对目的片段进行 PCR 扩增，得到该个体的富集的 DNA 条形码片段；对扩增片段进行测序；利用软件校正序列，建立系统进化树，分析结果；将研究样品的 DNA 序列、图像、采集地点和日期、采集人和鉴定人等信息提交 DNA 条形码数据库。

当然，不能随便选一段 DNA 序列便拿过来做条形码。优质的 DNA 条形码需要具有以下优点：广泛分布于各个物种中，并且可用通用引物进行扩增；该序列检测到的种间遗传差异要明显大于种内遗传变异。目前广泛运用的条形码有线粒体细胞色素 c 氧化酶 I (COI)、核糖体 RNA 的 18S、28S 和 ITS 基因。

1.3 从 DNA 条形码到细胞条形码

2005 年 2 月，第一届关于生命条形码的国际科学会议在伦敦自然历史博物馆举行2

在这之后全球范围内的科学家们建立了多个 DNA 条形码数据库，不仅促进科学家们的协作，也成为全球公民的重要公共资源。其中最著名的是加拿大生物多样性基因组学中心开发的 BOLD 数据库（The Barcode of Life Data System，http://www.boldsystems.org/），目前已有 826.1 万个 DNA 条形码，覆盖了 22.2 万种动物、6.9 万种植物、2.3 万种真菌及其他生物。

但条形码给科学家带来的启示不止于此，除了 DNA 条形码，还有 RNA 条形码甚至蛋白条形码，在谱系追踪和筛选等应用领域还有细胞条形码。

细胞条形码的主要构成部分是几乎无限多的短核苷酸序列，在最简单的情况下，每个细胞都使用给定的长度的特定序列进行标记的话，条形码总共可能的数量是 4N，N 是序列的长度，这个数量是相当可观的。除了完全随机的序列，细胞条形码还可以是半随机的序列，也可以是序列片段的随机组合，或是对一段已知序列中的随机切除。对于每一种情况，通常是条形码多样性和条形码稳定性或可识别性间的权衡。

从原理上讲，想把条形码递送到细胞内最简单的方法就是手动将单个条形码逐个分配到细胞，但这种方法虽然保证了唯一性但确实劳动密集型的，逐个标记仅在非常有限的条件下

---

[2] Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B Biol Sci. 2005 Oct 29;360(1462):1805-11. doi: 10.1098/rstb.2005.1730. PMID: 16214739; PMCID: PMC1609222.

使用，。目前比较有效且可靠的方法是在体外大量生产条形码载体，如病毒或质粒。载体库在合适的条件下进行转染可以实现向每个被转染的细胞递送少量条形码，且仅转染所需数量的靶细胞，方法包括：逆转录病毒、Sindbis 病毒、质粒注射或电穿孔等。

1.4 体内条形码的生成

体内条形码最初的生成方法是基于 DNA 重组酶。第一个通过这种途径的方法是 Brainbow。在 Brainbow 中，Cre 重组酶可以切除或翻转侧翼有特异性识别序列的 DNA 序列，作用于一系列荧光蛋白开放阅读框。重复的 Cre 作用导致目标阵列的随机调换，在每个细胞中产生不同的荧光组合，这些荧光组合可以通过成像进行区分。然而，由于 Cre 本质上倾向于切除而不是翻转，因此目标阵列的大小会随着时间的推移而缩小。另一种方法是使用 RciDNA 重组酶，该重组酶可翻转但不切除识别位点之间的 DNA 片段。避免切除，显著增加潜在的条形码多样性。也可以利用 CRISPR-Cas9 介导的断裂修复或重复来创造序列多样性。

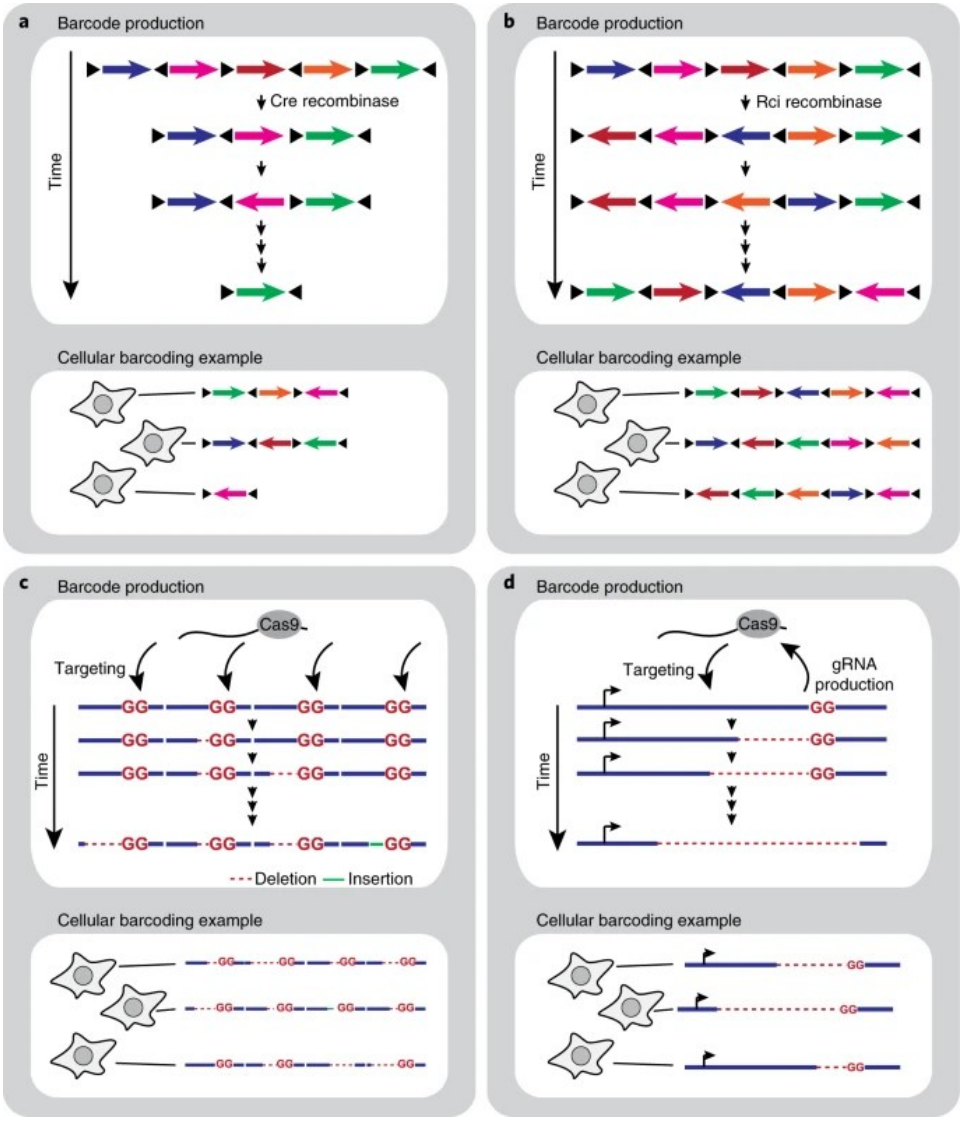

图 2 体内条形码生成原理[3]

如图 2 所示，a 中，当 Cre 重组酶作用于两侧是 loxP 位点（黑色三角形）的一系列靶位点（彩色箭头）时，它将切除或翻转这些靶标，从而产生序列多样性。过度切除会将数组折叠为单个（a 上方图），但 Cre 活动如果受到限制则会生成高度多样化的随机条形码（a 下方图）；b 中，Rci 重组酶仅翻转一系列目标位点中的片段，使得多样性随着时间的推移而增加的同时保持长度；c 中，由 Cas9 介导的双链断裂的非同源末端链接（NHEJ）修复产生的插入和缺失，CRISPR-Cas9 活性将随着时间的推移逐渐将序列多样性引入一系列靶位点；d 中，也可以将 CRISPR 向导 RNA（gRNA）设计为重复靶向自身，从而在该位点建立序列多样性。

1.5 体内条形码的读取和利用

至于条形码的读取方法，目前几乎所有关于细胞条形码的工作都依赖于核酸的提取，然后在体外进行定量检测，读取的方法包括定量 PCR、微阵列检测、Sanger 测序和高通量测序等。单细胞测序的方法目前已经用于在解离细胞的同时读取细胞条形码和转录组。

在本文中，我们关心的是如何运用条形码来映射神经连接（身份标签+谱系示踪）。神经解剖学最终的目标是确定大脑在单细胞和单突触分辨率下的完整神经网络图，但细胞条形码在这方面和传统方法的局限是类似的，都要在通量和分辨率之间受到一定的权衡，需要选择是通过大批量追踪来快速绘制多个神经元之间的连接，还是通过跟踪单神经元一次映射一个神经元的连接。为了克服这种不可避免的权衡，研究者开发了 MAPseq，并在这个基础上开发了 BRICseq 技术。

二、MAPseq 和 BRIC-seq 技术简介

2.1 将核酸条形码应用到神经元跨区域跟踪的背景

传统的长程投射示踪技术主要依赖于注射病毒或化学试示踪剂，并通过光学成像来直接观察远端的轴突末端，如果同一区域许多神经元都被大量标记了相同的荧光或酶，则单个神经元的不同投射模式通常难以识别。稀疏标记技术能够从单个神经元高精度地重建全脑轴突

[3] Letchuman, Sarvananda. (2018). Short Introduction of Dna Barcoding. International Journal of Research. 05. Kebschull JM, Zador AM. Cellular barcoding: lineage tracing, screening and beyond. Nat Methods. 2018 Nov;15(11):871-879. doi: 10.1038/s41592-018-0185-x. Epub 2018 Oct 30. PMID: 30377352.

投射4，但追踪通量很低，一个大脑只有一个或几个神经元可被涉及。使用不同颜色来标记单个神经元支持多通路的追踪和在同一大脑中投射模式的比较[5]，但它的处理量仍然受到成像光谱的分辨率。与可以读出大脑中数百到上千个细胞的高通量测序和神经活动监测技术相比，上述这些基于光学的单神经元示踪技术的通量总体上很有限，跨区域的远程追踪是很有挑战性的。

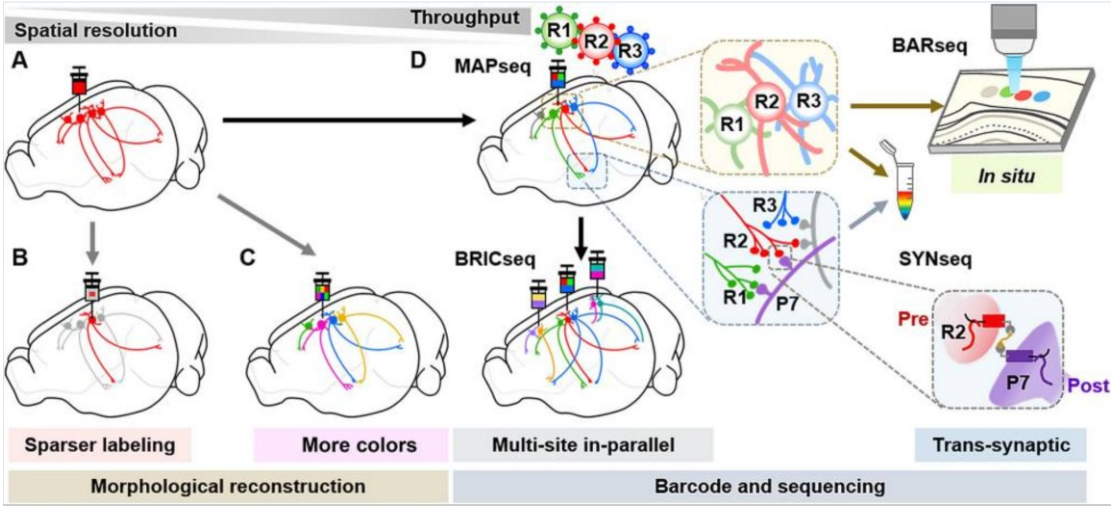

图 3 MAPseq 衍生技术对比[6]

2.2 MAPseq 技术

为了实现高通量的单神经元追踪，Zador 研究团队利用短的随机 RNA 条形码对单个神经元及其轴突进行了独特的标记，并通过在远端目标区域对这些条形码进行测序来读出投射模式。这些 RNA 条形码标记是高度多样化的 30ntRNA 片段，并利用 Sindbis 病毒来进行递送，同时也能表达出一种突触前蛋白 MAPP-nk，可以结合并运输 RNA 条形码到轴突末端，从而标记投射通路。这种方法被称作 MAPseq (Multiplexed Analysis of Projections by Sequencing) 。但这种方法也有明显的弊端，即在对这些神经元进行标记的过程中，必须先将这些脑细胞混合在一起，然后分离它们并进行标记，这会导致较差的空间分辨率。这个过程产生了模糊的图谱，这使得科学家们很难观察到神经元如何与特定的大脑功能（比如基因

[4] Jia F, Zhu X, Lv P, Hu L, Liu Q, Jin S, et al. Rapid and sparse labeling of neurons based on the mutant virus-like particle of Semliki forest virus. Neurosci Bull 2019, 35: 378–388.

[5] Abdeladim L, Matho KS, Clavreul S, Mahou P, Sintes JM, Solinas X, et al. Multicolor multiscale brain imaging with chromatic multiphoton serial microscopy. Nat Commun 2019, 10: 1662.

[6] Wu X, Zhang Q, Gong L, He M. Sequencing-Based High-Throughput Neuroanatomy: From Mapseq to Bricseq and Beyond. Neurosci Bull. 2021 May;37(6):746-750. doi: 10.1007/s12264-021-00646-3. Epub 2021 Mar 8. PMID: 33683648; PMCID: PMC8099946.

表达）相关联在一起。这种低分辨率还阻止了人们准确地确定神经元在大脑中的位置。[7]

2.3 BARseq 技术

为了克服这些缺点，Zador 小组进一步开发了 BARseq（Barcoded Anatomy Resolved by sequencing）。这种新技术利用了 MAPseq 与注射部位的原位测序相结合，可用于通过精确地指出神经元所在的位置来扩展大脑图谱。这使得 BARseq 不仅可以确定神经元的连接，还可以确定其基因表达模式和生理活性，因为原为测序可以与遗传驱动因子或荧光原位杂交结合。这是 MAPseq 不能解决的两个难题。[8]

2.4 SYNseq 技术

MAPseq 的另一个拓展是 SYNseq 技术，与其他三种已经成功用于绘制小鼠大脑连通性的技术不同，该技术的主要思路是通过测序来绘制突触连接。除了在 MAPseq 中使用的突触前条形码成分，它还增加了一个突触后条形码成分，并交联形成一个跨突触复合物，即条形码对。具体过程为首先在单独的突触前和突触后神经元群中表达随机 mRNA 条形码和修饰的突触蛋白。修饰的突触蛋白通过 RNA 结合结构域特异性结合 mRNA 条形码，从而将条形码分别运输到突触前或突触后区室。蛋白质在突触处相遇，并通过接头原位交联突触前和突触后蛋白的细胞外结构域。所得复合物由共价结合的突触前和突触后蛋白对组成，通过 RNA 结合结构域与其各自的条形码结合，然后通过免疫沉淀（IP）进行纯化。相关的条形码对（代表连接的神经元对）被连接、扩增和测序，最后可以对数据进行测序来重建连接矩阵。但目前该应用在体内还不够有效。

---

[7] Zador AM, Dubnau J, Oyibo HK, Zhan H, Cao G, Peikon ID. Sequencing the connectome. PLoS Biol 2012, 10: e1001411.

Kebschull JM. DNA sequencing in high-throughput neuroanatomy. J Chem Neuroanat 2019, 100: 101653.

[8] Xiaoyin Chen et al. High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. Cell, 2019, doi:10.1016/j.cell.2019.09.023.

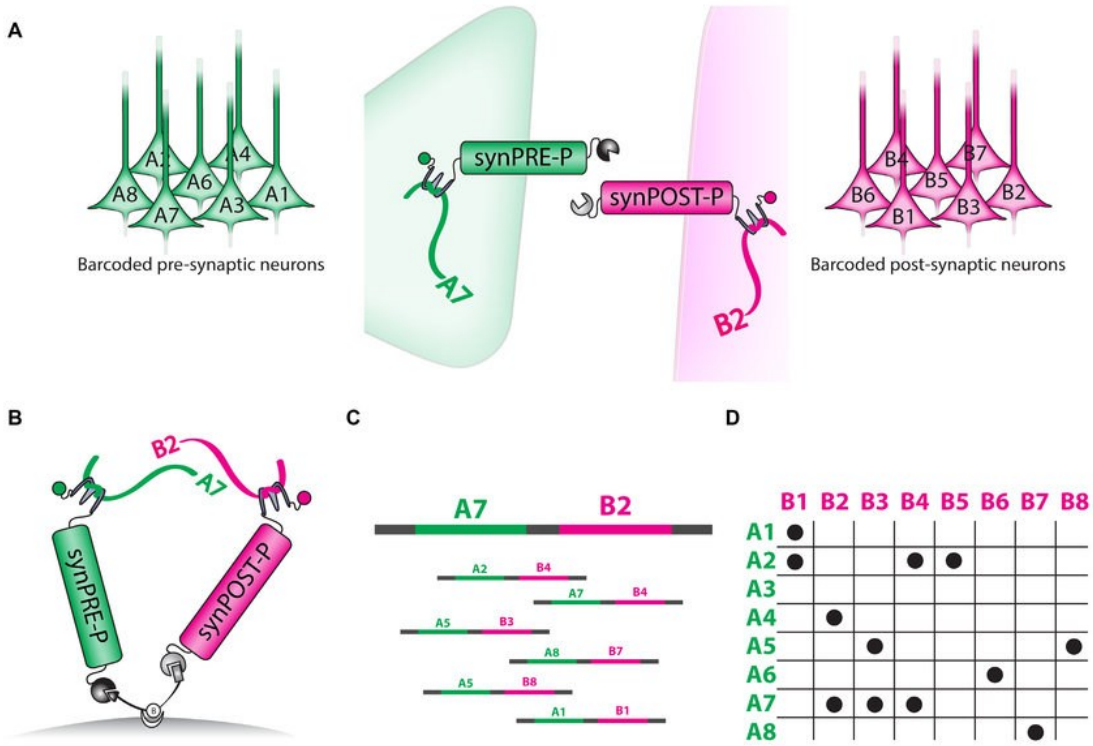

图 4 SYNseq 技术图示[9]

2.5 BRIC-seq 技术

最后就是我们本文的关键，BRIC-seq (Brain-wide Individual Animal Connectome Sequencing) 是基于细胞转录组测序技术建立的脑连接组绘图技术。BRIC-seq 具有高通量，耗材少的特点，并且在理论上可以拥有无限种标记细胞的“颜色”。同样是利用传播能力受限的 Sindbis 病毒将特异性 barcode 导入脑细胞使得细胞胞体带上 barcode 标记。随着含有 barcode 的 RNA 在细胞内不断扩增，barcode 会逐渐从胞体通过浆轴转运扩散至树突与轴突。如果一种 barcode 代表一种颜色，那么此时被 barcode 特异性的 Sindbis 病毒感染的细胞就被染上了一种颜色。并且染色深度——Barcode 浓度随着与胞体的距离增加而减小。通过对脑不同区域进行转录组测序就可以得到该区域的 barcode 丰度信息。将全脑多个区域的信息相互联用就可以得到每个被标记的胞体的具体位置和投射路径。尽管如此 BRIC 依然存在不足：其绘制脑联接组图谱的空间分辨率完全取决于 cublets 的大小，cublets 越大其分辨率越小，越容易丢失连接信息.

[9] Ian D. Peikon, Justus M. Kebschull, Vasily V. Vagin, Diana I. Ravens, Yu-Chi Sun, Eric Brouzes, Ivan R. Corrêa, Dario Bressan, Anthony M. Zador, Using high-throughput barcode sequencing to efficiently map connectomes, Nucleic Acids Research, Volume 45, Issue 12, 7 July 2017, Page e115

三、Bric-seq 流程实例

3.1 基本流程

BRIC-seq，大致流程主要分为六步，首先利用随机生成标签库和质粒克隆的方法构建带有特异性 barcode 的 Sindbis 病毒和病毒表达测序文库. 随后使用立体定位注射的方法将病毒注射到目标脑区位置. 进一步的待动物实验完成后对目标小鼠进行解剖,取出其大脑并进行低温冷冻切片并利用激光显微技术将脑片切割为大小合适的组织块(cublets). 最后将上述 cublets 单独进行单细胞转录组测序得到测序数据进行进一步的数据分析。

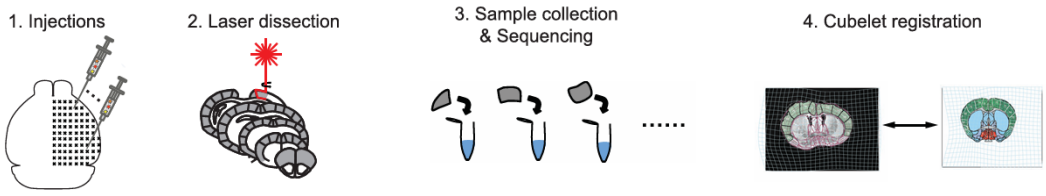

图 5 BRIC-seq 基本流程[10]

3.2 数据处理

3.2.1 数据矫正

数据处理主要分为误差校正,连接强度矩阵构建,结果分析,结论可视化四大部分. 在测序后得到两个。fastq 双端测序文件，其中一个文件为 32 个核苷酸的条形码(Barcode)序列结果，另一个为 12 个核苷酸的 UMI 和 8 个核苷酸的 CSI 序列。首先将量两段测序合并，随后使用 Fastqc 对数据进行质量控制,筛掉质量不佳的数据(含有显示为 N 的含模糊碱基的序列，读段次数小于相关阈值的序列等)。

为了消除 PCR 过程中的替换问题，通过代码可以将所有条形码聚类成大量的簇，使得对于给定簇中的任何条形码（BC1），在同一簇中存在另一个条形码（BC2），与其错配数小于 3。作为一种简单的算法，理论上它可能导致非常不同的条形码被聚类到同一个簇中；然而，由于使用的条形码之间的汉明距离较大（Kebschull 和 Zador，2015 年），在实际场景中很难发生这种情况。每个簇中具有最高 UMI 计数的条形码被用来表示该簇，并计算该簇中所有条形码的 UMI 计数总和作为该条形码的校正 UMI 计数。

[10] Huang et al., 2020

3.2.2 细胞投射强度构建

为了评估细胞的投射强度，我们以 UMI 丰度作为对投射强度的衡量:

$$UMI(i,*,k) = \frac{\sum_{j=1}^{N(i)} UMI(i,j,k)}{N(i)}$$

其中: i 代表源立方体，k 代表目标立方体，而 j 代表源立方体中的神经元; N(i) 表示立方体 i 中的投影神经元数量，UMI（i; j; k） 表示立方体 k 中来自立方体 i 中第 j 个神经元的 UMI 计数，

为了能够有效区分胞体与轴突需要对 UMI 设置相关丰度阈值范围: 我们需要提前预设相关阈值范围: 对于胞体,我们要求其 UMI 计数最高的丰度值大于 250，而对于轴突,我们要求其第二高的 UMI 丰度在 20 在 250 之间。从而我们可以进一步区分源区域与目标区域的细胞。

3.2.3 结果分析

接下来，将模板切换、重复使用条形码以及基线污染所导致的噪声考虑在内,用:

$$NOise(i,k) = UMI_{ts}(i,*,k) + UMI_{r\theta} + UMI_{ba}$$

其中第一项为模板切换，第二项为重复使用的条形码，第三项为基线污染所致的噪声。其中模板替换噪声是指由于在 RT-PCR 的过程中所有来自同一个 cublet 的 cDNA 被汇聚在一起 PCR，且共用一个引物。因此扩增结果可能存在扩增到一半突然更换模板的而产生的杂交链，所幸该现象较为罕见，并且可以通过为分子设定读取阈值来纠正。

与此同时的是重复使用条形码是 BRICseq 的另一个主要的假阳性误差来源，尤其是当条形码多样性不够高时。为了扩大 MAPseq 的规模，使用具有足够高多样性的条形码库至关重要。否则，相同的条形码可能会标记两种（或更多）不同的细胞，导致数据的误解。重复使用条形码的比率由条形码库的多样性和感染神经元的总数决定。然而，由于存在大量“非投射”神经元(弱投射,短程投射神经元或者胶质细胞等，它们所携带的条形码分子往往只存在于一个 cublet 中，并且可能具有丰度低的特征)，表达条形码的神经元总数远高于回收神经元的数量。尽管这些“非投射”神经元未被纳入数据分析，但它们可能包含与其他投射神经元共享的重复使用的条形码，从而导致错误的投射。

将条形码按照之前的阈值标准分为四类:

1. 最大值大于胞体阈值，第二大值大于 UMI 阈值(背景值)，且第二大值小于轴突阈值；
2. 最大值大于胞体阈值，第二大值小于 UMI 阈值；

3. 第二大值大于轴突阈值；

4. 第一大值小于轴突阈值但大于胞体阈值。

为了减少重复使用条形码的影响，我们只对 1 型条形码进行投影模式分析。据估计，约有 8%的 1 型条形码是重复使用的条形码，因此可以看出大部分条形码并未被浪费。在完成相关分析后计算其 P 值以估计计算结果的显著性(P 小于 0.05 为非显著连接,反之为显著连接)。

3.2.4 结论可视化

利用完成的投射强度矩阵我们可以构建三维连接网络，脑联接图谱，热图等可视化图像 从而进一步研究相关环路。

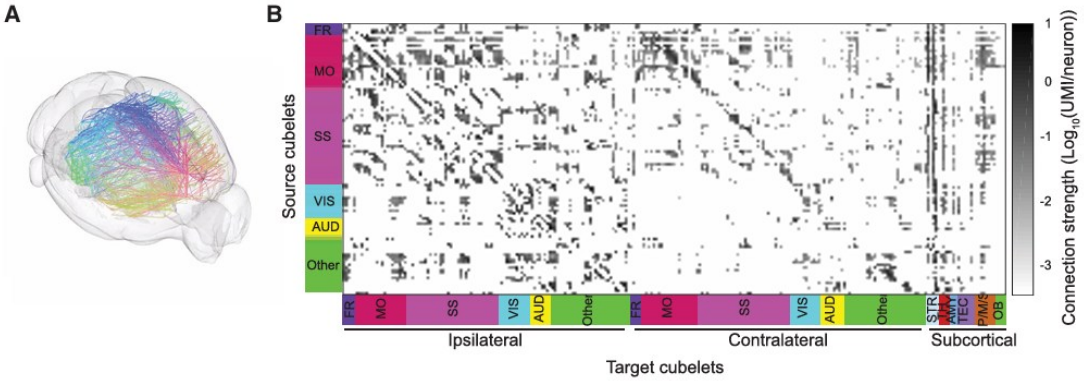

图 6 相关结论可视化效果[11]

四、核酸条形码应用示例

核酸条形码具有广阔的应用案例与前景。作为一种利用生物体核酸序列的特定区域作为识别标志来确定物种身份的方法，核酸条形码通过分析具有物种特异性的基因片段标记，可用于解决系统生物学、生态学、进化生物学、环境科学中的基本问题。目前核酸条形码的应用包括新物种的鉴定、食品的安全评估、隐蔽物种的鉴定和评估、外来物种的检测、濒危和受威胁物种的鉴定、阐明摄食生态位等。

2.1 物种鉴定与区分

核酸条形码可以通过测定生物体中特定基因片段的 DNA 序列来确定物种身份。在不同物种间，某些基因序列存在显著差异，而在同一物种内，这些序列则相对保守。因此，利用

[11] Huang et al., 2020

这些具有足够变异性的基因区域，可以对物种进行有效的识别和分类。不同突变率的基因也可以被用于鉴定分离共同祖先距今不同年代的物种。

根据不同的生物类群，常选用的条形码基因区域包括，动物界生物中的 COI 基因（细胞色素 c 氧化酶亚单位 I）、植物界生物中的 rbcL 基因（Rubisco 大型亚基基因）和 matK 基因（核糖体内转录区基因）、真菌界的 ITS 区（内转录间隔区）等。这些基因具有足够的序列多样性，能有效地区分不同物种，同时在同一物种内的变异较小，适合用于大规模物种鉴定。通过对不同物种的条形码基因序列进行存储和整理，建立一个大型核酸条形码数据库（如 BOLD 等），为物种的快速鉴定提供了重要支持。

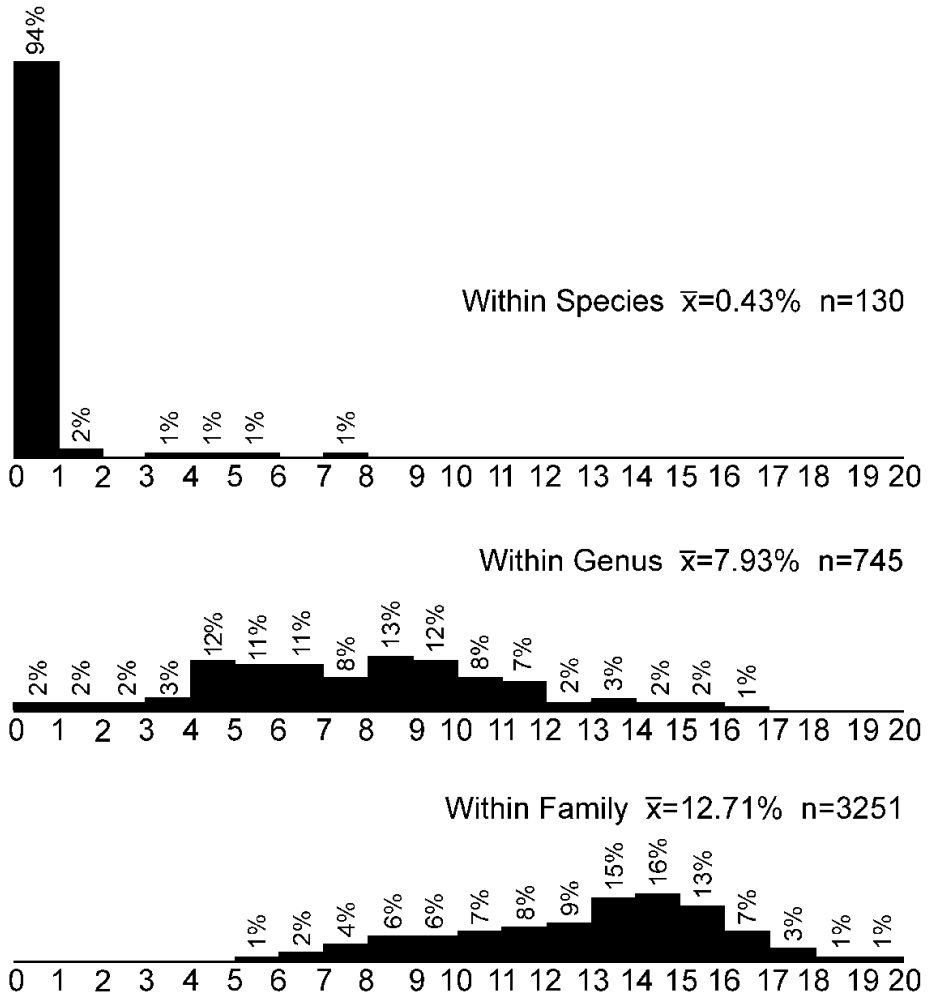

图 7 同种生物之间，不同种物种之间，不同属之间鸟类的序列比对结果[12]

利用 DNA 条形码进行物种鉴别一般包括以下几个步骤：

[12] Hebert, Paul D. N.; Stoeckle, Mark Y.; Zemlak, Tyler S.; Francis, Charles M. (October 2004). "Identification of Birds through DNA Barcodes". PLOS Biology. 2 (10): e312. doi:10.1371/journal.pbio.0020312. ISSN 1545-7885. PMC 518999. PMID 15455034.

首先，从待鉴定的生物体中提取 DNA。不同物种的 DNA 提取方法有所不同，通常采用商业化的 DNA 提取试剂盒或传统的 CTAB 法等。提取出的 DNA 需要进行质量检测，确保其没有降解或断裂，并适合后续的 PCR 扩增。其次，需根据物种类别选择适当的条形码基因区域，如前文所述，通常选择 COI、rbcL、matK、ITS 等区域。在合适的引物作用下，使用聚合酶链反应（PCR）技术扩增目标基因区域。PCR 扩增后的 DNA 片段需要进行纯化和测序。常用的测序方法包括 Sanger 测序和高通量测序技术等，测得的 DNA 序列将作为条形码序列。将测得的 DNA 序列与数据库中的已知物种条形码序列进行比对。通常使用 BLAST（基本局部比对搜索工具）等工具来进行序列比对，依据相似度来确定待鉴定样本的物种。如果匹配度高于某个阈值（一般为 97%-99%），则可以确定该样本的物种身份。此外，可结合形态学特征、生态信息等进行交叉验证，以确保鉴定的准确性。

例如，Hebert 等人于 2004 年曾利用 DNA 条形码对阿斯特拉普特斯·富尔格拉托尔蝶属进行了全新的物种划分。他利用了线粒体 COI 基因的一个短片段，作为划分新物种的方法，并重新分析了先前使用 DNA 条形码识别阿斯特拉普特斯·富尔格拉托尔蝶属中 10 个可能新物种的数据。分析显示，数据最多支持 7 个不同的线粒体 DNA 支系，即最多支持具有 7 个种级分类单元。

2.2 食物链等生态学有关研究

近年来，DNA 条形码已广泛应用于生态学研究，特别是在食物链研究中。食物链研究旨在揭示生态系统中物种之间的能量流动与物质循环，通过了解不同物种的相互关系，科学家能够深入探讨生态平衡、物种相互作用以及环境变化对生态系统的影响。传统的食物链研究多依赖形态学观察和捕捉样本的方法，但这些方法存在一定的局限性，例如鉴定困难、取样偏差以及样本的时间与空间限制等。而 DNA 条形码技术为研究食物链提供了更加精准和全面的手段。

在食物链研究中，DNA 条形码技术的核心原理是利用 DNA 序列的特征差异来鉴定某种生物的食物来源，并通过分析不同物种之间的相互关系，揭示生态系统的食物链结构。DNA 条形码技术具有高效、灵敏、跨领域等特点，使得食物链研究可以突破传统方法的局限，特别是在环境 DNA（eDNA）研究和微小物种的检测方面，提供了更加深入和广泛的数据支持。

利用 DNA 条形码进行食物链研究的基本过程包括样本采集、DNA 提取、基因扩增与测序、数据分析与食物链构建等几个步骤。研究食物链时，首先需要采集生态系统中的不同物种样本。这些样本可以来源于捕食者、被捕食者以及其共生体的体内，代谢废物，遗体或环境。

从样本中提取 DNA 时，对环境 DNA 提取通常需要使用专门的提取方法，以去除样本中的杂质，获得高质量的 DNA。提取的 DNA 经常是混合型、片段型的，可能包含多种物种的 DNA 信息。在食物链研究中，通常会选择多种基因区域进行扩增，以确保物种鉴定的准确性。扩增后的基因片段需要进行测序，测序后的 DNA 片段将作为条形码序列，用于后续的物种鉴定和食物链分析。测序结果需要通过数据库比对， 通过对不同样本的 DNA 条形码进行比对与鉴定，可以得到生态系统中各个物种食物组成及其占比。结合生态学数据（如捕食行为、物种共存关系等），可进一步构建食物链模型。

通过分析物种之间的相互作用关系（捕食、被捕食、竞争等），我们便可以揭示生态系统中的能量流动路径，分析物种在食物链中的作用与贡献，评估生态系统中能量的流动效率与物质的循环情况。

例如，弗朗索瓦·庞帕农在 2011 年的一篇文章中提到，基于 DNA 的方法可能为饮食研究提供更准确的方法。文章介绍了已经得到实验验证的一套基于通过表征肠道或粪便样本中存在的 DNA 来鉴定消耗物种的方法。最初，由于该方法涉及对 PCR 产物的克隆进行测序，由于所需的成本和精力，研究的规模受到限制。新一代测序 （NGS） 的最新发展使这种方法更加强大，允许直接表征数十个样品，每个 PCR 产物有数千个序列，并且有可能同时揭示许多消耗的物种（DNA 宏条形码）。NGS 技术的持续改进、成本的持续降低以及当前参考数据库的大规模扩展使这种方法前景广阔。梅丽莎·弗里德曼也于 2017 提出了一种用 DNA 条形码是评估食品质量的方法。其通过 DNA 条形码提供了从食物产品或残渣中识别导致雪卡毒鱼中毒的石斑鱼的可能性，并在食物中毒患者的案例中得到验证，为保障食品安全提供了一种新的检测手段。

2.3 蛋白质条形码技术

蛋白质条形码（Protein Barcode）是一种基于蛋白质或肽片段的分子识别技术，旨在通过独特的蛋白质序列、修饰模式或结构特征对生物样本进行标识和分类。

蛋白质条形码的设计通常涉及这些层面：使用特定的肽片段或全蛋白质作为标识符，并通过化学修饰、荧光标记或同位素标记使其易于检测。需要检测时，可采用质谱（MS）、液相色谱（LC）或荧光显微镜等技术对条形码进行解读。

蛋白质条形码技术可被广泛应用于疾病诊断、药物开发和生物分子追踪等领域。

通过分析患者血液或组织样本中的蛋白质条形码，可以发现疾病相关的生物标志物。例如，特定癌症类型通常伴随独特的蛋白质表达模式，蛋白质条形码技术可以用来检测这些特

征蛋白，从而实现早期诊断。此外，如阿尔茨海默病等疾病患者，其蛋白质条形码可能包含异常折叠或聚集的蛋白。

此外，蛋白质条形码也可用于药物筛选和疗效评估。通过蛋白质条形码鉴定关键的分子靶点，为待检测药物的实际效果提供依据；也可通过患者治疗前后样本的蛋白质条形码变化，评估药物的效果及副作用。

与核酸条形码类似，蛋白质条形码也可用于监测环境中的微生物群落变化。例如，在水质监测中，通过识别微生物的蛋白质条形码，可以评估水体污染程度和生态健康状况。

总之，蛋白质条形码作为一项前沿技术，为生命科学研究和医学应用提供了新的工具。通过不断优化条形码的设计与检测技术，其潜力将进一步释放，为疾病诊断、药物研发和基础研究带来深远影响。

2.4 核酸条形码技术面临的挑战

虽然 DNA 条形码技术在多方面具有显著优势，但其应用仍然面临一些挑战和局限性。

例如，不同物种之间的遗传差异在某些基因标记中可能不够显著，从而影响条形码的准确性和有效性。尤其是在一些基因变异较小的物种或较为接近的物种间，核酸条形码的辨识度可能受到限制，因此需要根据物种的特征和研究需求选择合适的基因片段。然而，在针对不同基因片段产生的数据中可能会出现整合难度大，出现差异等情况，使得构建出的进化树呈现出碎片化，局部化等趋势。如何有效科学地进行数据整合或寻找使用更广泛的条形码基因仍然是核酸条形码研究的核心问题。

核酸条形码技术的准确性高度依赖于基因数据库的建设与更新。随着物种多样性的不断发现，现有的基因数据库中尚存在许多物种未被收录，尤其是在一些欠发达地区和数量稀少的物种数据尚不完善。这使得核酸条形码技术的普遍应用受到了一定限制。

虽然核酸条形码技术已逐渐成熟，但在部分地区，尤其是发展中国家的实验设施和技术条件可能不够完善，导致其应用成本较高，数据量受到了严重限制。尽管随着技术的进步，核酸检测与测序成本有逐步下降的趋势，但要实现全球范围内广泛的应用，还需进一步降低技术门槛。

随着基因测序技术的不断进步和价格的不断下降，核酸条形码技术在未来有望在更多领域中发挥作用。尤其是高通量基因测序技术的进步，使得核酸条形码不仅能够识别单一物种，还能在复杂的生态系统中实现多物种的同时鉴定。此外，基因数据库的不断扩展与国际合作也将推动核酸条形码技术的发展与应用，使其在生物多样性保护、生态监测、食品安全等方

面发挥越来越重要的作用。

总之，DNA 条形码技术作为一项革命性的生物学工具，已经在全球范围内得到了广泛应用，并且随着科学技术的发展，其应用范围还将进一步扩展。

参考文献（应用部分 1-4，流程部分 5-8，简介部分 9-14）：

1. Friedman MA, Fernandez M, Backer LC, Dickey RW, Bernstein J, Schrank K, Kibler S, Stephan W, Gribble MO, Bienfang P, Bowen RE, Degrasse S, Flores Quintana HA, Loeffler CR, Weisman R, Blythe D, Berdalet E, Ayyar R, Clarkson-Townsend D, Swajian K, Benner R, Brewer T, Fleming LE. An Updated Review of Ciguatera Fish Poisoning: Clinical, Epidemiological, Environmental, and Public Health Management. Mar Drugs. 2017 Mar 14;15(3):72.
2. POMPANON, F., DEAGLE, B.E., SYMONDSON, W.O.C., BROWN, D.S., JARMAN, S.N. and TABERLET, P. (2012), Who is eating what: diet assessment using next generation sequencing. Molecular Ecology, 21: 1931-1950.
3. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci U S A. 2008 Feb 26;105(8):2923-8.
4. Brower, Andrew. (2006). Problems with DNA barcodes for species delimitation: 'Ten species' of Astraptes fulgerator reassessed (Lepidoptera: Hesperiidae). Systematics and Biodiversity. 4. 127-132. 10.1017/S147720000500191X.
5. _FastQC 测序质量—Xiaojikuaipao—博客园_. (n.d.). Retrieved December 19, 2024,
6. Huang, L., Kebschull, J. M., Fürth, D., Musall, S., Kaufman, M. T., Churchland, A. K., & Zador, A. M. (2020). BRICseq Bridges Brain-wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals. _Cell_, _182_(1), 177-188.e27.
7. Imamachi, N., Tani, H., Mizutani, R., Imamura, K., Irie, T., Suzuki, Y., & Akimitsu, N. (2014). BRIC-seq: A genome-wide approach for determining RNA stability in mammalian cells. _Methods_, _67_(1), 55–63.
8. _五分钟教你观看 FASTQ 文件质量评估结果 - 云生信_. (n.d.). Retrieved December 19, 2024, from http://www.biocloudservice.com/wordpress/?p=25283
9. Letchuman, Sarvananda. (2018). Short Introduction of Dna Barcoding. International Journal of Research. 05.
10. Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philos Trans R Soc Lond B Biol Sci. 2005 Oct 29;360(1462):1805-11. doi: 10.1098/rstb.2005.1730. PMID: 16214739; PMCID: PMC1609222.

11. Kebschull JM, Zador AM. Cellular barcoding: lineage tracing, screening and beyond. Nat Methods. 2018 Nov;15(11):871-879. doi: 10.1038/s41592-018-0185-x. Epub 2018 Oct 30. PMID: 30377352.
12. Jia F, Zhu X, Lv P, Hu L, Liu Q, Jin S, et al. Rapid and sparse labeling of neurons based on the mutant virus-like particle of Semliki forest virus. Neurosci Bull 2019, 35: 378–388.
13. Abdeladim L, Matho KS, Clavreul S, Mahou P, Sintes JM, Solinas X, et al. Multicolor multiscale brain imaging with chromatic multiphoton serial microscopy. Nat Commun 2019, 10: 1662.
14. Wu X, Zhang Q, Gong L, He M. Sequencing-Based High-Throughput Neuroanatomy: From Mapseq to Bricseq and Beyond. Neurosci Bull. 2021 May;37(6):746-750. doi: 10.1007/s12264-021-00646-3. Epub 2021 Mar 8. PMID: 33683648; PMCID: PMC8099946.