

DOI: 10.13376/j.cbls/2023016

文章编号: 1004-0374(2023)02-0107-07

睿 ● 观 ● 家

生命科学研究面临的实验可重复性之危机

吴家睿^{1,2}

(1 中国科学院分子细胞科学卓越创新中心, 上海 200031; 2 上海交通大学安泰经济与管理学院, 上海 200030)

“实验结果必须重复”是笔者做学生时就学到的研究工作之基本标准,也是笔者在指导学生做实验时提出的基本要求。但让人没有想到的是,近些年这个基本常识却成为科学共同体高度关注的一个问题。早在2011年,美国著名生物技术公司“Amgen”选择了肿瘤研究领域53篇标志性研究论文进行验证,发现其中仅有6篇论文中的结果可以被重复^[1]。*Nature*杂志曾对多个国家的数、理、化、生等领域1576位科研人员在线问卷调查,其中52%的受访者认为目前发表的研究论文的“可重复性”(reproducibility)出现了显著的危机,38%的受访者认同这种危机的存在,但认为其程度还不是很严重^[2]。2014年,时任美国国立卫生研究院(National Institutes of Health, NIH)院长的柯林斯(Collins F)和同事以“NIH计划加强可重复性”为题发表了他们的担忧:“临床前研究,尤其是动物实验,看上去就正好是当前对可重复性问题最敏感的领域”^[3]。

为了更好地理解和分析实验的可重复性,美国“开放科学中心”(Center for Open Science)专门发起对已经发表的研究论文进行可重复性检验的研究项目。该中心牵头组织了来自世界各国的250多名科学家,自2011年起对2008年三本顶级的心理学杂志上发表的文章系统地进行了可重复性检验工作,其检验结果不容乐观——被检验的100篇研究论文中,只有39篇的心理学研究工作可被重复^[4]。此外,该中心和美国肿瘤生物学家艾隆丝(Iorns E)在2013年联合发起“癌症生物学可重复性项目”(Reproducibility Project: Cancer Biology),选择了2010—2012年间多篇高影响力的肿瘤生物学研究论文,利用7种方法进行可重复性检验。目前已经对23篇论文涉及的50个实验进行了验证研究,其中只有46%的实验可以被成功地重复出来^[5]。为此,*Nature*发表了题为“科学结果的重复并不容易,但很关键”的社论,明确指出“高影响力肿瘤生物学论文的可

重复性研究给出了令人失望的结果,科学家必须加倍努力去找出原因”^[6]。

1 实验结果的可重复性

20世纪中叶诞生的现代生命科学是一门实验科学,研究人员必须在严格控制的环境下开展实验,并获得各种实验数据来解答科学问题或证实科学假设。因此,一篇研究论文所陈述的各种实验过程及其相应的实验结果是该论文的核心部分,也是可重复性检验的主要目标。例如,“癌症生物学可重复性项目”的研究者对2011年发表在*Nature*的一篇研究论文进行了验证工作,证明该论文中关于一种新型小分子抑制剂对白血病细胞生长抑制的实验可以被重复,但关于该小分子能够让白血病小鼠存活时间显著增加的实验则没有得到证实^[7]。

需要指出的是,做一个实验不容易,重复一个实验也并不容易。“癌症生物学可重复性项目”最近总结了关于29篇临床前肿瘤研究论文的验证进展情况:共完成并发表了18篇可重复实验报告。在其余11篇论文的可重复工作中,4篇论文的可重复实验没有得到结果而被终止;2篇论文的可重复实验遇到了意想不到的问题,所以实验结果并不完整;剩下的5篇论文都只完成了一部分可重复实验,其他的可重复实验由于技术限制或者方法学上的难度而没有完成^[8]。

可重复实验不容易的原因有很多。首先,论文所公开的实验方案通常描述得不够详细,可重复实验项目的实施者往往需要联系论文作者以了解实验方案的细节。其次,发表在论文上的实验数据是经过分析和整理的,往往还需要论文作者提供原始数据。再次,如果可重复实验需要用到论文涉及的质粒、细胞和小鼠等实验材料,项目的实施者还要与论文作者及其相关研究单位签订实验材料的转移协议并得到相应的实验材料。由此可见,论文作者的

配合对可重复实验的实施起着重要的作用。可众所周知, 论文作者在当今功利主导的科研文化和科研环境下做到这样的配合并非易事。此外, 进行可重复实验还可能遇到某些不可抗力, 如试剂耗材的停产。另外, 还存在一个“循环论证”的问题: 可重复实验本身还需要再重复吗?

科学实验的可重复性危机一方面与人有关, 另一方面与物有关。从人的角度来看, 首先要注意到当前研究人员在科研诚信方面呈现的诸多问题; 这些问题显然会影响到实验的可重复性。其次, 研究和有关流程规范也会影响到实验的可重复性。NIH 的柯林斯院长指出, 动物实验等临床前研究是对可重复性问题最敏感的领域; 他同时却认为, 针对人的临床研究在可重复性方面的问题不大, 因为这类研究已经被严格的实验设计和独立的监管等各种实验规范所支配^[3]。

从物的角度来看, 对生物学实验可重复性影响较大的是常用的生物学实验材料, 尤其是细胞系和抗体。细胞作为生物体的基本结构单元和功能单元, 是生命科学研究人员做活体实验时最常用的实验材料。因此, 大多数生命科学实验室都保存有各种细胞系, 可以通过培养这些“种子”细胞来不断扩增细胞数量并用于研究, 而且可以在不同的实验室之间很方便地进行交换和共享。但是, 多项调查发现, 实验室使用的细胞系中有 20%~36% 被其他不同种类的细胞污染或被错误标识。不久前发表的一份科学文献分析报告指出: 有 3 万多份论文的研究工作采用了错误标识的细胞, 这些论文又被近 50 万篇文章引用^[9]。美国遗传学家科赫 (Korch C) 通过对两种常用的细胞系 HEP-2 (喉鳞癌细胞) 和 INT 407 (肠胚性细胞) 的文献分析发现, 被一种人宫颈癌细胞“HeLa”污染了的 HEP-2 涉及了近 5 800 篇研究论文, 而 HeLa 细胞污染了的 INT 407 则涉及了 1 336 篇研究论文^[10]; 也就是说, 认为自己正在研究 HEP-2 或者 INT 407 细胞的科学家, 实际上可能正在研究 HeLa 细胞! 目前, 越来越多的科技期刊要求论文作者对所使用的细胞系进行“身份”鉴定并提交相应的证据。

细胞系影响实验可重复性问题还不仅仅是污染和错误标识, 同一细胞系在实验过程中也会自发地产生各种变化。不久前, 研究人员利用蛋白质组学和其他组学技术系统地分析了 HeLa 细胞系, 发现在同一实验室内连续培养一株 HeLa 细胞系 3 个月就可以造成不同培养时间的细胞之基因表达谱产生

大约 7% 的变化, 且不同实验室相同命名的 HeLa 细胞系存在亚系, 它们在染色体数目、基因表达、蛋白质表达和细胞扩增等方面均有显著的差异^[11]。此外, 一项大规模的基因组研究发现, 来自两个实验室的 106 个人类细胞系显示出了广泛的亚克隆变异, 而一株实验室常用的乳腺癌细胞系 MCF7 衍生的 7 个品系 (strains) 的细胞之间也表现出迅速的遗传多样化 (genetic diversification)^[12]。

另一个对生物学实验可重复性有重大影响的是在众多实验室中广泛使用的抗体。Nature 于 2015 年发表了题为“抗体的无政府状态: 呼唤次序”的评论文章, 明确指出研究中使用的抗体经常导致模糊的结果^[13]。抗体在生物学实验中最常见的问题是: 抗体的特异性不强, 常常出现假阳性或假阴性。如 2019 年的一项研究证实 2006 年在 *EMBO Journal* 发表的一篇论文使用的抗体有问题, 该抗体并未与论文所说的目标蛋白结合, 而是结合了另一个蛋白, 从而致使该论文被撤回^[14]。研究人员还发现, 有时同一个抗体在几次平行实验里得出的结果都不同, 或者用新抗体重复以前的实验时发现结果无法重复。

抗体的可重复性差主要是与抗体的生产过程和批次差异等质量控制有关。据估计, 实验室每年因购买质量差的抗体导致的直接损失达到 8 亿美金。学术界和产业界为保证抗体的质量做了大量的努力。从 2017 年起, 美国的基金申请人被允许在预算中专门列支一些经费用于抗体验证。抗体生产商和供应商也开发出各种验证抗体特异性的技术和方法。例如, 英国生命科学试剂供应商 Abcam 采用 CRISPR-Cas9 技术来验证抗体, 然后把每个抗体的验证结果发布出来。

研究人员通常很难评估他们购买的抗体之特异性和选择性, 也很难保证一个给定抗体产品的每个批次都同样有效。2015 年 9 月, 由瑞典科学家乌能 (Uhlén M) 领导的“国际抗体验证工作组” (International Working Group for Antibody Validation, IWGAV) 成立。该工作组于 2016 年提出了验证抗体特异性的 5 个基本策略^[15], 用以指导“免疫印迹” (Western blot) 等实验涉及的抗体验证, 并在 2018 年报告了应用这 5 个基本策略对 6 000 多个抗体的验证结果, 表明这些抗体验证策略既适用于抗体生产方, 也适用于抗体使用者^[16]。

2 研究结论的可重复性

研究论文通常由一组相互关联的实验组成, 其

目标是要解答科学问题或证实科学假设；研究人员通过系统地分析这些实验获得的数据/结果来形成研究结论并整理成为论文。因此，人们不仅要考虑实验结果的可重复性，更要关注研究结论的可重复性。要指出的是，这二者的验证是在不同的层面上进行，意义也不一样，有的实验是决定研究结论是否成立的关键性实验，而有的则是支持或完善研究结论的补充性实验。时不时我们会看到某篇论文的作者这样说：某实验结果虽然有点问题，但本论文的研究结论不受影响，依然成立。

科技出版界有一个特点，即编辑偏好把研究结论一致或高度相关的两篇论文放在一本杂志的同一期里发表。这被形象地称为“背靠背”。“背靠背”发表的两篇研究论文通常来自两个独立的实验室，有时他们做的实验内容都很相近，有时两篇论文虽然得到了一致的研究结论，但具体的实验方式和实验材料彼此却有很大的差异。显然，如果两个独立的实验室通过“背靠背”的论文表述了相同的研究结论，那么二者之间的实验工作无论相似还是有差别，该研究结论的可重复性就已经得到了彼此的验证。

从以上简单的讨论可以看到，研究结论的可重复性要比实验结果的可重复性更为重要，对前者的验证工作也比后者更为复杂。以神经科学为例，成人大脑发育成熟后是否还保持形成新的神经元之能力——神经发生(neurogenesis)——是最具争议的问题之一，其中一些研究者认为“是”，另外一些研究者则给出“否”。2018年，*Nature*发表了一篇支持“否”的论文——成人大脑海马区(hippocampal)没有新生神经元产生^[17]；而2019年，*Nature Medicine*却发表了一篇支持“是”的论文——成人大脑海马区有新生神经元产生^[18]。这两篇文章的作者采用的实验策略和实验材料很相近，他们围绕着彼此实验采用的材料和技术展开了争论，包括实验用的人脑组织样本在固定液中浸泡的时间长短，以及免疫组织化学标记检测结果存在假阳性和假阴性的可能性；争论结果是“公说公有理，婆说婆有理”，谁也没有说服谁。

要强调的是，生命体系的构造和其生理活动的调控非常复杂，需要回答的科学问题可以采用不同的研究思路来开展研究。这一点在有关成人大脑是否能够进行神经发生的科学争论中有着很好的体现。例如，2019年发表在*Cell*上的一篇研究论文就是通过成年小鼠的海马区域神经祖细胞的起源来

探讨神经发生的问题，发现成年小鼠海马区域的神经祖细胞来源于胚胎期海马齿状回的神经干细胞，发育结束后这类神经干细胞依然一直保留在海马特定区域，从而使成体海马的神经再生能够维持下去^[19]。也就是说，这项关于小鼠胚胎神经干细胞的研究工作从一个新的维度为成人脑的神经发生之间题提供了不同类型的实验证据。这应该也属于对支持“是”的论文之研究结论的可重复性验证。

研究人员除了关注成人大脑是否在自然状态下进行神经发生，还关注能否在人为干预下实现神经再生(neuro-regeneration)。这个科学问题同样存在着巨大的争议，一批研究人员给出肯定的答案，如人为增加 NeuroD1 基因表达或下调 Ptbp1 基因表达可以将胶质细胞转分化为神经元细胞；而另外一批研究人员却给出否定的答案：单独改变 NeuroD1 基因或 Ptbp1 基因表达均不能让胶质细胞实现转分化^[20-21]。美国约翰·霍普金斯大学医学院周峰泉教授不久前系统地评述了这两派研究人员在神经再生领域所开展的实验工作和造成研究结论不一致的可能因素，并在文中总结到：“由于面临的问题难度大而且实验相对复杂并涉及到大量活体动物实验，该领域经常有‘突破性发现’，但实际上许多发表的文章难以让人信服”^[22]。针对脊髓损伤与再生研究领域研究结论可重复性问题，国际上在该领域的70多位专家组成的“Minimum Information about a Spinal Cord Injury experiments”联合体(MIASCI Consortium)于2014年发表了一篇文章，明确提出了研究者在开展脊髓损伤与再生研究时需要提供的最基本实验信息之标准，涉及到实验条件、实验方法和实验动物等方面^[23]。

生命的复杂性往往会产生许多不在研究人员预期框架里的问题。例如，不久前的一项研究发现，在条件完全一样的实验中，进行实验的男性研究人员和女性研究人员分别会让小鼠的大脑对氯胺酮产生不同的响应，进而导致小鼠出现不同的行为方式^[24]。利用小鼠模型或癌细胞系得到的研究结论通常被直接推广到人体内相应的肿瘤。为了检验这种推广是否可靠，美国研究人员开发了一种称为“CancerCellNet”的算法，可以通过比较RNA转录数据来分析真实的肿瘤和实验室研究模型之间的异同程度；他们的基因表达分析结果表明，基因工程小鼠和肿瘤类器官(tumorooids)要比癌细胞系和人源肿瘤异种移植模型(patient-derived tumor xenograft, PDX)更接近真实的肿瘤^[25]。2022年发表的一篇文章

章进一步揭示了这种实验模型与真实生物体出现差异的一种原因：人体肿瘤细胞在机体内通常生长于缺氧的病理环境下，一旦把肿瘤样本从体内取出进行培养或者其他实验操作时，肿瘤细胞就被暴露在富氧的自然环境下，进而导致肿瘤细胞的基因表达和信号通路等发生明显的改变；药物实验表明，缺氧环境下的癌细胞明显比暴露于富氧的正常空气中的更加耐药；而把在正常空气中暴露过的癌细胞注射给小鼠，这些癌细胞的成瘤比例更高，生长速度也更快^[26]。显然，在体外实验中获得的研究结论不能简单地推广到活体实验。

3 数据密集型研究的可重复性

20世纪中叶诞生的生命科学主要是采用一种针对个别基因或蛋白质的“小数据”实验科学范式。随着世纪之交人类基因组计划的实施，生命科学进入了“后基因组时代”，表现出了一种“数据密集型研究”的新范式，如2022年*Science*发表的一篇研究论文就涉及了英国12 000多名癌症患者的全基因组序列，从中发现了海量的基因组变异，包括近3亿个单碱基置换(substitutions)、260多万个双碱基置换(double substitutions)、1亿5千多万个插入或缺失(indels)和近200万个重排(rearrangements)^[27]。据统计，世界范围内每年所产生的生物医学数据在2013年大约为153 EB (1 EB = 10^{18} Byte)，而在2020年则估计达到了2 314 EB。

显然，生命科学领域的数据密集型研究范式所面临的重复性与经典的小数据实验科学范式有着很大的区别。从实验结果可重复性的角度来看，数据密集型研究采用的样本之来源、标准和质量需要考 虑，实验过程中的具体路径以及获得的数据质量等也需要考 虑。这类研究还可能面临着一类特有的挑战——“规模”，即一项研究往往涉及到成百上千的研究人员，他们来自不同的研究单位甚至不同的国家。例如，在2001年2月*Nature*发表的“人类基因组草图”中，列入论文的主要作者就有近300人，涉及到6个国家的48个研究机构或组织。

为了保证数据密集型实验的可重复性，学术界有针对性地发布了各种相关实验工作的建议性标准或规范。早在2001年，欧洲多国和美国的一批研究人员就针对全基因组范围基因表达的芯片实验制定了一个“芯片实验的基准信息”(minimum information about a microarray experiment)，提出基于芯片技术获得的基因表达数据的记录和报告之基本标准^[28]。

这种“minimum information”的方式随后被普及到了各种数据密集型实验工作，蛋白质组学实验有“minimum information about a proteomics experiment”(2007)；代谢组学实验有“Metabolomics Standards Initiative”(2007)；基因组学实验有“minimum information about a genome sequence”(2008)；糖组学实验有“minimum information required for a glycomics experiment”(2017)。

虽然研究人员通过各种努力来保证数据密集型实验的可重复性，但这类实验依然存在难以重复的问题。以“新一代测序技术”(next-generation sequencing, NGS)为例，NGS目前在数据密集型实验中应用最为广泛，仅仅美国的生物医学数据库“MEDLINE”目前所收录的NGS相关的论文数量就已近10万篇。但是，一篇分析文章指出，研究人员对NGS数据的可重复性验证面临着广泛的障碍，“他们经常发现，许多需要用来分析NGS数据的信息学组件很难用于重复已经发表的文章”；其可能的原因包括文章中使用的分析技术缺乏细节，或者是文章作者所使用的计算方法不能被其他研究者再次使用^[29]。此外，一篇分析文章也指出，“批次影响”(batch effects)广泛存在于各种高通量数据的实验中；因此，“除了把标准化、验证分析和统计显著性计算用于高通量数据分析中，还有必要针对批次影响把合并和调整方法作为分析这类数据的标准步骤”^[30]。

从研究结论可重复性的角度来看，数据密集型研究范式表现出了一种特殊性质——“迭代”(iterate)，即论文给出的研究结论可以是一种未完成的或不完备的，允许研究人员在前期研究结论的基础上进行完善，然后提出更新了的研究结论。人类基因组计划就是这种数据密集型研究结论进行“迭代”的典型。2001年2月，*Nature*发表的人类基因组测序成果明确指出，这是一份仅仅覆盖了基因组90%核酸序列的“草图”；2004年10月，*Nature*再次发表了人类基因组测序论文，其标题也只是说完成了常染色质区域内核酸序列的测定；2020年9月，*Nature*发表了人类第一条染色体的完整核酸测序结果；2022年4月，以“人类基因组完整序列”为题的研究论文在*Science*上发表^[31]，时隔“草图”发表已经22年。在这篇最新的论文里，研究人员终于完整地测定了人类基因组的30亿对碱基序列，比2004年发表的版本增加了近2亿个碱基对和1 956个新基因^[31]。显然，这种对研究结论进行“迭代”的特性一定程度上消解了数据密集型研究的可

重复性问题！

从数据密集型实验中获得的研究结论离不开研究人员所采用的数据分析方法，因此有可能不同的数据分析方法导致研究结论的不一致。这里讲一个比较“极端”的案例。一般认为，DNA在转录过程中按照碱基配对原则指导RNA的合成，二者的碱基序列具有严格的对应关系。但是，一项基于高通量测序技术的比较研究发现，人类的DNA序列和相对应的RNA序列之间具有很大的差异，仅在10 000多个编码蛋白质的外显子位点上就发现了大量的碱基差异；其论文题目“人类转录组的RNA与DNA序列之间存在着广泛的差异”清晰表明了实验所得到的研究结论^[32]。然而，就在该论文发表3个月之后出现了一篇题为“人类转录组的RNA与DNA序列之差异非常少”的论文^[33]。需要指出的是，在这篇后发的论文中，作者并没有做任何新的实验，仅仅是采用不同的数据分析方法分析了前文中的原始数据，然后就得出了完全相反的结论^[33]。显然，由此引出数据密集型实验的研究结论可重复性特有的问题：主要是源于“干实验”数据分析而得到的研究结论究竟可信用度有多高？

在数据密集型实验中，从“大数据”到形成研究结论往往不能像经典生命科学实验那样对实验获得的“小数据”和研究结论进行非常严格和全面的验证。因此，从“大数据”推导出研究结论的过程中存在着很大的“解释空间”。有关环状RNA的生物学功能之争论就是一个典型案例。环状RNA是指在mRNA剪接过程中，外显子序列被反向剪接(back splicing)而形成的封闭环状RNA结构。环状RNA发现之初被认为属于剪接错误导致的副产物，不具备生物学功能。但2013年发表在*Nature*的研究论文提出，环状RNA是一类具有生理调控作用的功能分子^[33]。需要强调的是，研究人员在形成这个研究结论的过程中，只是从其检测到的数千个环状RNA中挑出一个来进行详细的功能验证实验^[34]。然而，*Cell Report*在2021年发表的一篇文章对这一研究结论进行了挑战。该文的作者通过对人类、恒河猴和小鼠11个组织的RNA测序数据分析，估算出大多数环状RNA皆因剪接错误而产生，其中有害的比例超过97%；因此，大多数环状RNA属于没有功能的垃圾RNA^[35]。还要指出的是，该文作者称他们也的确找到了若干个具有生物学功能的候选环状RNA^[35]。也就是说，前文发现一个有功能的环状RNA之实验结果与后文的结论并不冲突，

关键问题是，前文没有对所发现的数千个环状RNA都进行功能实验就推导出它们可能都具有功能的研究结论。这种“由点及面”的推导模式常常出现在数据密集型实验中，导致了一种新型的研究结论可重复性之问题。

4 小结：解决实验可重复性危机首先需要恢复科学精神并重建学术生态

科学研究的可重复性涉及到许许多多的客观因素，包括实验设计、实验技术、实验材料、科学理论和分析方法，等等。但最需要考虑的应该是主观因素，即研究者及其所处的学术环境。可以说，当今“引爆”科学研究可重复性危机的正是整个学术界高度功利化的学术环境以及被其影响乃至控制的研究人员。因此，要想从根本上解决当前的科学可重复性危机，关键在于恢复失去的科学精神，重建健康的学术环境。正如*Science Signaling*杂志主编亚法(Yaffe MB)在题为“科学的可重复性”的社论中所说：“如果不改变整个科学文化，我相信很难去解决这些实验技术引发的可重复性问题”^[36]。

科学研究的主要目标是探索未知，在这个探索过程中出现错误在所难免；科学独特的性质之一就在于其错误可以被系统地进行分析和批评，并在大多数时候都得到及时的改正。也就是说，科学研究的精神从本质上说是“自我纠偏”(self-correction)——研究人员在实验过程中发现和改正失误是其研究工作成功的一个重要组成部分；这正是我们常说的“失败乃是成功之母”。在“可重复性”概念之背后体现的也正是这种“自我纠偏”精神。可是，今天的学术界显然背离了这种“自我纠偏”精神，研究人员、学术期刊和科技管理层习惯于“报喜不报忧”，通常只关注那些研究取得的“阳性结果”(positive results)，而明显地忽略那些研究中产生的“阴性结果”(negative results)。为此，亚法主编在社论中明确指出：“只有当项目资助过程得到根本的改变，当科学影响力以不同的方式评估，当我们想出办法来发表重要的、具有很好对照实验的阴性结果——包括那些对‘高影响力’发现重复失败的数据，人们才有可能改进可重复性的状况”^[36]。

学术环境的改变重点应体现在对研究人员和研究工作的评价。美国科学院为解决这些问题组织了一个由前院长阿尔伯特(Albert B)领导的工作小组；在其工作报告中他们提出了一系列解决措施，其中也提到，“我们认为研究动机应该进行改变，要以

发表论文的质量而非数量来奖励研究者。大学的长聘评审应该像基金申请一样,对候选人的评价是根据其代表性工作的重要性,而不是用发表文章的数量或者杂志的影响因子作为质量的标识”^[37]。我国有关部门最近也明确提出了“反四唯”等改进科技评价的一系列措施,如2020年2月国家科技部发布了《关于破除科技评价中“唯论文”不良导向的若干措施(试行)》的通知。

当前的学术界还有一个重要的特征——封闭,即存在大大小小的“学术圈”(笔者对“学术圈”的详细讨论见参考文献[38])。“学术圈”的封闭性进而导致了科学研究的不透明:不仅技术路线、实验材料和数据分析等实验过程中的关键信息模糊,而且“阴性结果”基本上不予披露,甚至“阳性结果”也往往是经过选择的。可以说,当前学术界形成的封闭和不透明在科学可重复性危机中也扮演了重要的角色。正是基于这样的考虑,牵头对高影响力论文进行可重复性检验研究的美国“开放科学中心”,专门成立了“透明与开放促进委员会”(Transparency and Openness Promotion Committee);该委员会制定了促进科学研究透明与开放的8项标准:引用的标准、数据的透明、分析方法(代码)的透明、设计和分析的透明、研究的预登记、分析计划的预登记、重复实验^[39]。Cell杂志社在2016年也启动了一个旨在让研究者的实验材料和技术等信息更透明、更清晰、更容易获取的“STAR”(Structured, Transparent, Accessible Reporting)计划^[40],要求作者把其研究论文内有关实验方法的详细内容提交到线上发布(<http://www.cell.com/star-methods>);该杂志社为此把其发表文章中原来的方法部分之要求进行了修改,从“Experimental Procedures”改为“STAR + METHODS”。

显然,要想改善整个学术环境进而解决科学的可重复性问题并非易事,需要学术界、政府和社会各个方面的共同努力,仅靠某个单位或某个部门的力量是远远不够的。在“NIH计划加强可重复性”的评论文中,前院长柯林斯对此有着明确的认识:“仅凭NIH一家之力不足以让这个不健康的环境发生真正的改变”^[3]。

[参 考 文 献]

[1] Begley CG, Ellis LM. Raise standards for preclinical cancer research. *Nature*, 2012, 483: 531-3
 [2] Baker M. Is there reproducibility crisis? *Nature*, 2016, 533: 452-4
 [3] Collins FS, Tabak LA. NIH plans to enhance reproducibility.

Nature, 2014, 505: 612-3
 [4] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 2015, 349: aac4716
 [5] Errington TM, Mathur M, Soderberg CK, et al. Investigating the replicability of preclinical cancer biology. *eLife*, 2021, 10: e71601
 [6] Editorial. Replicating scientific results is tough--but crucial. *Nature*, 2021, 600: 359-60
 [7] Shan X, Fung JJ, Kosaka A, et al. Replication study: inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *eLife*, 2017, 6: e25306
 [8] Errington TM, Denis A, Allison AB, et al. Experiments from unfinished registered reports in the reproducibility project: cancer biology. *eLife*, 2021, 10: e73430
 [9] Horbach SPJM, Halffman W. The ghosts of HeLa: how cell line misidentification contaminates the scientific literature. *PLoS One*, 12: e0186281
 [10] Neimark J. Line of attack. *Science*, 2015, 347: 938-40
 [11] Liu Y, Mi Y, Mueller T, et al. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nat Biotechnol*, 2019, 37: 314-32
 [12] Ben-David U, Siranosian B, Ha G, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*, 2018, 560: 325-30
 [13] Baker M. Antibody anarchy: a call to order. *Nature*, 2015, 527: 545-51
 [14] Lyst MJ, Nan X, Stancheva I. Retraction: regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *EMBO J*, 2019, 38: e103220
 [15] Uhlén M, Bandrowski A, Carr S, et al. A proposal for validation of antibodies. *Nat Method*, 2016, 13: 823-7
 [16] Edfors F, Hober A, Linderbäck K, et al. Enhanced validation of antibodies for research applications. *Nat Commun*, 2018, 9: 4130
 [17] Sorrells SF, Paredes MF, Cebrian-Silla A, et al. Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature*, 2018, 555: 377-81
 [18] Moreno-Jiménez EP, Flor-García M, Terreros-Roncal J, et al. Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nat Med*, 2019, 25: 554-60
 [19] Berg DA, Su Y, Jimenez-Cyrus D, et al. A common embryonic origin of stem cells drives developmental and adult neurogenesis. *Cell*, 2019, 177: 1-15
 [20] Rao Y, Du S, Yang B, et al. Neuro1 induces microglial apoptosis and cannot induce microglia-to-neuron cross-lineage reprogramming. *Neuron*, 2021, 109: 4094-108
 [21] Guo T, Pan X, Jiang G, et al. Downregulating PTBP1 fails to convert astrocytes into hippocampal neurons and to alleviate symptoms in Alzheimer's mouse models. *J Neurosci*, 2022, 42: 7309-17
 [22] 周峰泉. 真假难辨的神经再生, 科学争议该如何解决? [EB/OL]. 知识分子, 2021-10-13. <https://mp.weixin.qq.com/s/xoHQnyZ55rjMgK6mlFncwQ>
 [23] Lemmon VP, Ferguson AR, Popovich PG, et al. Minimum

- information about a spinal cord injury experiment: a proposed reporting standard for spinal cord injury experiments. *J Neurotrauma*, 2014, 31: 1354-61
- [24] Georgiou P, Zanos P, Mou TC, et al. Experimenters' sex modulates mouse behaviors and neural responses to ketamine via corticotropin releasing factor. *Nat Neurosci*, 2022, 25: 1191-200
- [25] Peng D, Gleyzer R, Tai WH, et al. Evaluating the transcriptional fidelity of cancer models. *Genome Med*, 2021, 13: 73
- [26] Kumar B, Adebayo AK, Prasad M, et al. Tumor collection/processing under physioxia uncovers highly relevant signaling networks and drug sensitivity. *Sci Adv*, 2022, 8: eabh3375
- [27] Degasperi A, Zou X, Amarante TD, et al. Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science*, 2022, 376: eabl9283
- [28] Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet*, 2001, 29: 365-71
- [29] Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat Rev Genet*, 2012, 13: 669-72
- [30] Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*, 2010, 11: 733-9
- [31] Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. *Science*, 2022, 376: 44-53
- [32] Li M, Wang IX, Li Y, et al. Widespread RNA and DNA sequence differences in the human transcriptome. *Science*, 2011, 333: 53-8
- [33] Schrider DR, Gout JF, Hahn MW. Very few RNA and DNA sequence differences in the human transcriptome. *PLoS One*, 2011, 6: e25842
- [34] Memczak S, Jens M, Elefsinioti A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 2013, 495: 333-8
- [35] Xu C, Zhang J. Mammalian circular RNAs result largely from splicing errors. *Cell Rep*, 2021, 36: 109439
- [36] Yaffe MB. Reproducibility in science. *Sci Signal*, 2015, 8: eg5
- [37] Alberts B, Cicerone RJ, Fienberg SE, et al. Self-correction in science at work. *Science*, 2015, 348: 1420-2
- [38] 吴家睿. “精英中心化”科研范式的特征及其面临的挑战. *科学通报*, 2021, 66: 3509-14
- [39] Nosek BA, Alter G, Banks GC, et al. Promoting an open research culture. *Science*, 2014, 348: 1422-5
- [40] Editorial. A STAR is born. *Cell*, 2016, 166: 1059-60