

描绘扰动图谱，一窥天书奥秘

荆征宇^{1,3}，张效铭^{1,3}，康立清，池天^{1,2,4}

¹上海科技大学生命科学与技术学院

²耶鲁大学医学院免疫生物学系

³ 共同一作

⁴ 通信作者

chitian@shanghaitech.edu.cn

微信 13817022262

引言

虎啸狮吼，莺歌燕舞，花香草翠，松挺柳依……壮观绚丽的生命现象，归根到底由基因决定。基因的物质基础是 DNA。一个生物体的全部 DNA（基因组），构成了其生命现象的蓝图和剧本。人类基因指导我们健康地生长发育并成为万物之灵，其功能障碍则带来各种病痛。20 年前，人类基因组计划正式完成，这是彪炳史册的伟大科学成果。但是，誊抄生命密码，只是了解基因组“外貌”；破译这个密码，使我们对其“知人知面也知心”，其难度远超密码誊抄。人类 2 万个基因在 400 多种细胞中的功能，至今鲜为人知，人类基因组还是一部神秘天书。

要揭秘基因功能，需对其进行扰动。工欲善其事，必先利其器。我们八年磨一剑，打造了大规模扰动技术 iMAP（induced Mosaic Animal for Perturbation），描绘了世界首张“扰动图”，它展示 90 个基因分别在 39 种组织细胞的基本功能，并揭示了多个潜在药靶。iMAP 的解码速度比传统单基因扰动技术高出 100 倍（类似汽车与步行的区别），并有可能再升高 10 倍（类似变汽车为飞机）。毫无疑问，终有一天，生命奥秘将被彻底破解，疑难病症的药靶也将浮出水面。这一梦寐以求的时刻，随着 iMAP、高通量测序、计算生物学、人工智能等各种相辅相成的尖端技术的涌现、发展，已隐隐出现在地平线上，它也许很快会扑面而来，将我们带进一个激动人心的新时代。

■ 研究背景

1866年，孟德尔提出“遗传因子”是遗传变异的决定因素；1909年，约翰森将“遗传因子”命名为“基因”；1944年，艾弗里证明了基因本质是DNA；1953年，沃森和克里克阐明了DNA结构；1990年，在美国启动“人类基因组计划”，先后共有6国（包括中国）2000多名科学家加入。在花费13年、38亿美金后，6国政府首脑于2003年联名发表《关于完成人类基因组序列图的联合声明》，正式宣告人类基因组测序基本完毕。这是生命科学研究史上最重要的里程碑之一，开启了“后基因组时代”，其任务是破解基因组功能。人类基因组计划被誉为生命科学的阿波罗登月计划。如果测序基因组是登陆月球，那么揭秘其功能也许比移民火星还要难。

人类基因组有~30亿个碱基对，含~2万个编码蛋白质的基因（简称基因），但蛋白质编码序列（基因外显子）仅占基因组序列的1.5%。其余的序列，多种多样：有些是内含子，有些调控基因表达（如启动子、增强子、沉默子），有些转录成“非编码RNA”，还有假基因、重复序列（包括转座子）等。这些的序列中，基因是生命蓝图的核心成分，其研究也最为深入。人体细胞有37万亿个，起码可分为400种类型^[1]。但是，2万个基因在各种细胞中的功能，至今大多未知，更遑论基因组其他序列。基因在疾病中的作用，同样鲜为人知，大多疾病因此缺乏药靶，尤其是高效、特异的“优质靶点”。事实上，人类起码有1万种病^[2]，而临床使用的基因药靶仅667个，不到基因总数的3%^[3]。这些药靶无一为中国发现，凸显了我国与制药领域先进国家间亟待缩短、消除的差距。

那么，如何鉴定基因（或其他DNA序列）功能，从而挖掘药物靶点？可用以下故事解释。据说，马斯克（MUSK）有一秘书，要求加薪。马斯克让她休假2周，然后就炒了她，因为他发现，她不在时公司照样正常运行。同样，鉴定DNA功能的基本方法就是使其“休假”（将其从基因组剔除即“敲除”），再观察细胞表型。除了敲除，也可对基因进行其他形式的扰动（例如降低或提高基因表达水平、改变蛋白质序列），但敲除是最基本、最常用的扰动方法，也是鉴定基因功能的金标准。基因敲除实验当然不能在人体进行；违法伦理会受到学界谴责乃至刑法严惩，如“贺建奎事件”。因此，解码人类基因组，模式动物不可缺少。小鼠是生命科学研究中应用最广泛的动物。小鼠体积小、繁殖快，而且其蛋白序列与人类高度（85%）一致，其一些

非编码序列如启动子、增强子也有一定的保守性，因此小鼠实验有重要参考价值。

小鼠基因敲除技术问世于 20 世纪 80 年代末，其发明人于 2007 年获诺贝尔生理学医学奖^[4]。该技术利用同源重组的方法，在小鼠胚胎干细胞(ESC)中剔除一个靶基因，再通过复杂的方法，将干细胞发育成小鼠。这种小鼠全身缺失靶基因，称为全身性单基因敲除（“单敲”）品系，可用于鉴定该基因在个体水平的功能。单敲是目前最常用、最基本的小鼠基因解码策略。基因敲除技术问世后 10 多年，科学家们各自独立地敲除、研究基因。这种“各自为政”、“零敲碎打”的研究模式，很像人类基因组测序前，人类基因的研究局面。为了加快小鼠基因解码，2007 年，柯林斯（Collins）等人发起成立“国际基因敲除小鼠联盟”（International Knockout Mouse Consortium, IKMC），以便有组织、系统性地 ESC 中敲除 2 万基因^[5, 6]。有趣的是，柯林斯也正是当年人类基因组计划的领导人，这反映了基因研究从测序（20 世纪）到功能解码（21 世纪）这一由表及里、由浅及深的历史行程。2011 年，“国际小鼠表型鉴定联盟”（International Mouse Phenotyping Consortium, IMPC）成立，它规模宏大，由 21 个研究机构的数百科学家组成，旨在联合 IKMC，用 10 年时间制备 2 万个单敲品系并对其进行初步、简单的表型分析（<https://www.mousephenotype.org/>）。但是，IMPC 成立至今 10 余年，制备、分析的品系不到 1 万，而且其中 1/3 品系胎死腹中，表明小鼠基因中，有 1/3 为“胚胎必需基因”（embryonic essential gene, EEG）。EEG 也常表达于成体，但由于敲除导致胎死腹中，其成体功能无法研究。即使对非 EEG，也难以用单敲品系揭示基因功能。这是因为，要理解基因功能，必须鉴定其敲除对细胞的直接效应，但全身性敲除可影响多个组织器官，后者又可继而互相影响，从而混淆直接效应。为鉴定直接效应，人们研发了小鼠“组织特异性”敲除技术，它基于 Cre-LoxP 重组系统^[7]。Cre 和 LoxP 分别是噬菌体产生的重组酶和其识别序列。将一对 LoxP 分别插入基因片段两端，再利用 Cre 使 LoxP 之间发生重组，可剔除 LoxP 之间的序列。在插入了 LoxP 的小鼠中，把 Cre 选择性地表达在特定组织，即可将靶基因敲除局限在这个组织，这不但能揭示基因的直接功能，而且还常能避免胎死腹中，从而克服全身性敲除的缺点。但是，该技术流程复杂（需制备 LoxP 和 Cre 品系，再将其杂交），而且每次只能研究一个组织。全身性和组织特异性敲除还都有一个关键缺点：两者都只能逐个敲除、分析基因，通量极低。

为了弥补单敲策略的缺点，人们研发了基于“CRISPR-Cas9 基因编辑”^[8]的高通

量遗传筛选技术（简称“CRISPR 筛选”）^[9]。什么是 CRISPR-Cas9？20 世纪 80 年代末，发现细菌基因组携带一种奇怪的 DNA 序列，将其命名为 CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats，成簇的规律性间隔的短回文重复序列)，其谜底于 2007 年揭开：CRISPR 是细菌免疫防御体系的一个成分。原来，某些细菌在遭到病毒入侵后，能够捕获病毒基因的一小段并将其插入自己基因组中的 CRISPR 位点，在此存储病毒信息，相当于将病毒列入“黑名单”。细菌随后将病毒基因片段转录成 RNA（相当于读出黑名单），并将其转化成“向导 RNA”（gRNA），再将其与细菌免疫防御体系的另一关键成分——Cas9（CRISPR associated protein 9）——组装成复合物；Cas9 是能切割 DNA 的核酸酶，俗称“魔剪”。这个复合物像导弹，其中 gRNA 和 Cas9 分别负责识别（通过碱基互补）和摧毁（通过切割）病毒基因。因此，经历病毒感染后，细菌就制备、部署了导弹严阵以待，以便迅速而精准地歼灭再犯之敌。真核细胞 CRISPR-Cas9 基因编辑技术就是通过人工设计 gRNA，使导弹靶向细胞本身基因组中的特定基因，精准地将其剪断，诱发细胞固有的修复反应，使研究者得以趁机引入所需突变，从而改造、修饰（“编辑”）靶基因^[8]。最简单、常见的突变是移码，常导致基因敲除，它源于修复时自发产生的碱基随机插入或丢失（而如提供特定的修复模版，则可获得其他突变）。CRISPR-Cas9 基因编辑技术用途广泛，已风靡全球，并且催生了 2020 年获诺贝尔化学奖。值得指出，2013 年，Jaenisch 和我国

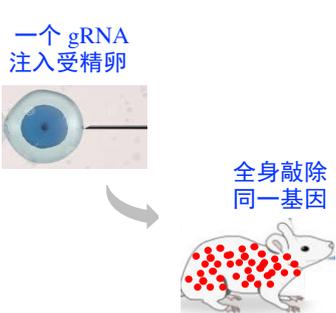
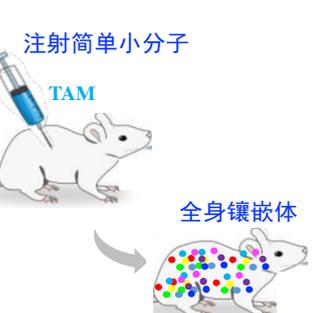
	单基因敲除	原位CRISPR 筛选	iMAP
	 <p>一个 gRNA 注入受精卵</p> <p>全身敲除同一基因</p>	 <p>注射昂贵、高滴度病毒gRNA文库</p> <p>局部镶嵌体</p>	 <p>注射简单小分子 TAM</p> <p>全身镶嵌体</p>
敲除通量	低（1基因/品系）	高（但通常<120基因/小鼠）	中（100基因/品系）
gRNA递送	高效	低效、不均	天生携带，TAM诱导，分布全身
靶细胞	全身细胞	实质器官（脑肝肺）	全身细胞
小鼠反复使用	是	否	是
快速解码基因	否	难	易
描绘扰动图谱	不合适	不合适	合适

图 1. 几种代表性的基于 CRISPR-Cas9 的小鼠体内基因解码策略。研发 iMAP 是本工作的研究目标之一。

黄行许等发现，将 Cas9 mRNA 和 gRNA 注射入小鼠受精卵，就能有效地敲除靶基因，产生基因敲除小鼠^[10, 11]。该技术已取代基于 ESC 的传统技术，成为大家制备单敲品系的首选方法（图 1 左）。

与单敲不同，CRISPR 筛选技术通常是利用病毒载体，将多种 gRNA 的混合物（“pooled gRNA 文库”）同时引入一群细胞，但每个细胞只随机引入文库中的一个 gRNA、敲除一个靶基因；对这群细胞同时、“平行”（而不是先后、逐个）地进行表型筛选，即可从大量基因中一举查出调控特定表型的基因^[9]。CRISPR 筛选通常在体外进行。例如，要寻找调控肿瘤细胞耐药性的基因，可在体外培养药敏肿瘤细胞，引入 gRNA 文库，再用药杀；如果发现有细胞存活，那么只要鉴定其 gRNA，就可推测出介导药物敏感性的基因。此外，如果引入文库后，对细胞群进行单细胞水平的高通量 RNA 测序，即可一举表征文库中各个靶基因对转录组的调控作用，从而较全面地了解基因功能。因此，CRISPR 体外筛选通量远高于小鼠单敲策略。遗憾的是，多数生理过程难以在体外再现、研究，而且每次筛选只能在一个细胞类型里进行。鉴此，人们已经试图在小鼠体内进行筛选^[12]。

体内 CRISPR 筛选分为 2 类。第一可称为“移植筛选”，即从小鼠（表达 Cas9）分离靶细胞，在体外培养，用病毒引入 gRNA 文库，再输回体内分析。该方法比体外筛选有一定优势，但只适合少数既可移植又可感染的细胞类型（主要是免疫细胞），而且每次只能分析一类细胞，操作也相当复杂。第二类体内筛选可称为“原位筛选”：将病毒文库直接注射到小鼠器官，使每个细胞随机感染上一个病毒颗粒、敲除一个基因，从而将局部组织转化为“镶嵌结构”（mosaic structure，即不同细胞含不同基因组），再分析各基因敲除分别导致的细胞表型（图 1 中）。原位筛选操作简单（无需分离、移植细胞），结果可靠（靶细胞处于天然环境中），而且每次可“平行”地感染、分析多种细胞。值得指出，文库中通常只有少数 gRNA 能严重破坏细胞，所以整个文库只能破坏组织中小部分细胞，不会影响整个组织器官而混淆基因敲除对靶细胞的直接效应（除非 gRNA 导致细胞大量增殖甚至肿瘤；详后，图 4A 左）。也就是说，原位筛选一般能揭示基因在细胞里的直接功能，正像其他形式的 CRISPR 筛选。遗憾的是，原位筛选亦有其短板：体内病毒感染效率低下，范围也窄（限于注射针头附近），目前只被用于脑、肺、肝这 3 个易触及的实质性器官。此外，所有 CRISPR 筛选都有严重的共同局限性：其珍贵试剂（病毒文库、感染了病毒的小鼠）不能重复

使用，必须反复制备，从而使实验昂贵、繁难并遭受“批次效应”的干扰。这与小鼠单敲品系相反，后者一旦制备，即可自我繁殖，成为永久资源，可反复使用。除了 CRISPR，转座子、致癌物等也能诱导小鼠基因组突变，已被用于大规模体内筛选，但各自也有其局限性。

由于各种筛选技术的缺点，目前解码小鼠基因主要还得依赖单敲策略；快速小鼠体内基因解码技术，至今还是求而不得的圣杯。

■ 研究目标与研究内容

1. 长远目标

- 小鼠有 2 万个基因、400 多种细胞，共 800 万种组合。我们将在单细胞水平，鉴定每个基因敲除后分别对每种细胞的效应，即描绘“全景扰动图”，从而解码每个基因在各种细胞的基本功能。全景图将成为探索生命奥秘必备的“世界地图”，其问世将成为生命科学研究史上的重要里程碑和分水岭，因为在此后，探索基因基本功能将如查字典般简单，这对破解人类生命天书有重大意义。
- 挖出重大疾病的优质药靶，助力我国在该领域零的突破，为治疗疑难病症奠定基础。

2. 短期目标/研究内容

千里之行，始于足下。本研究聚焦以下内容：

- 研发小鼠快速原位基因解码技术 iMAP (induced Mosaic Animal for Perturbation)，使人们用简单的小分子化合物泰莫昔芬 (TAM) 即可将整个小鼠转化为镶嵌体 (图 1 右)。iMAP 汇集了单敲和 CRISPR 筛选的优点而避免了其关键缺点，填补了该领域的空白，是崭新、突破性技术。
- 利用 iMAP 解码 90 个基因分别在 39 种器官/组织/细胞 (以下简称“组织”) 的基本功能，从而描绘世界首张“微型扰动图”。可以预测，微型图将催生一个崭新的国际前沿研究方向：描绘全景图。这是本工作前瞻性之所在。

■ 研究成果

6代学生前赴后继，历经8年，取得了以下成果，发表于《细胞》(Cell)杂志 [13]。

1. 研发 iMAP

iMAP 原理 iMAP 核心是一种新型 gRNA 文库，由多达 100 个 gRNA 编码序列 (“guide”) 串联而成。将它插入受精卵基因组，制备转基因鼠，使每个细胞都天然携

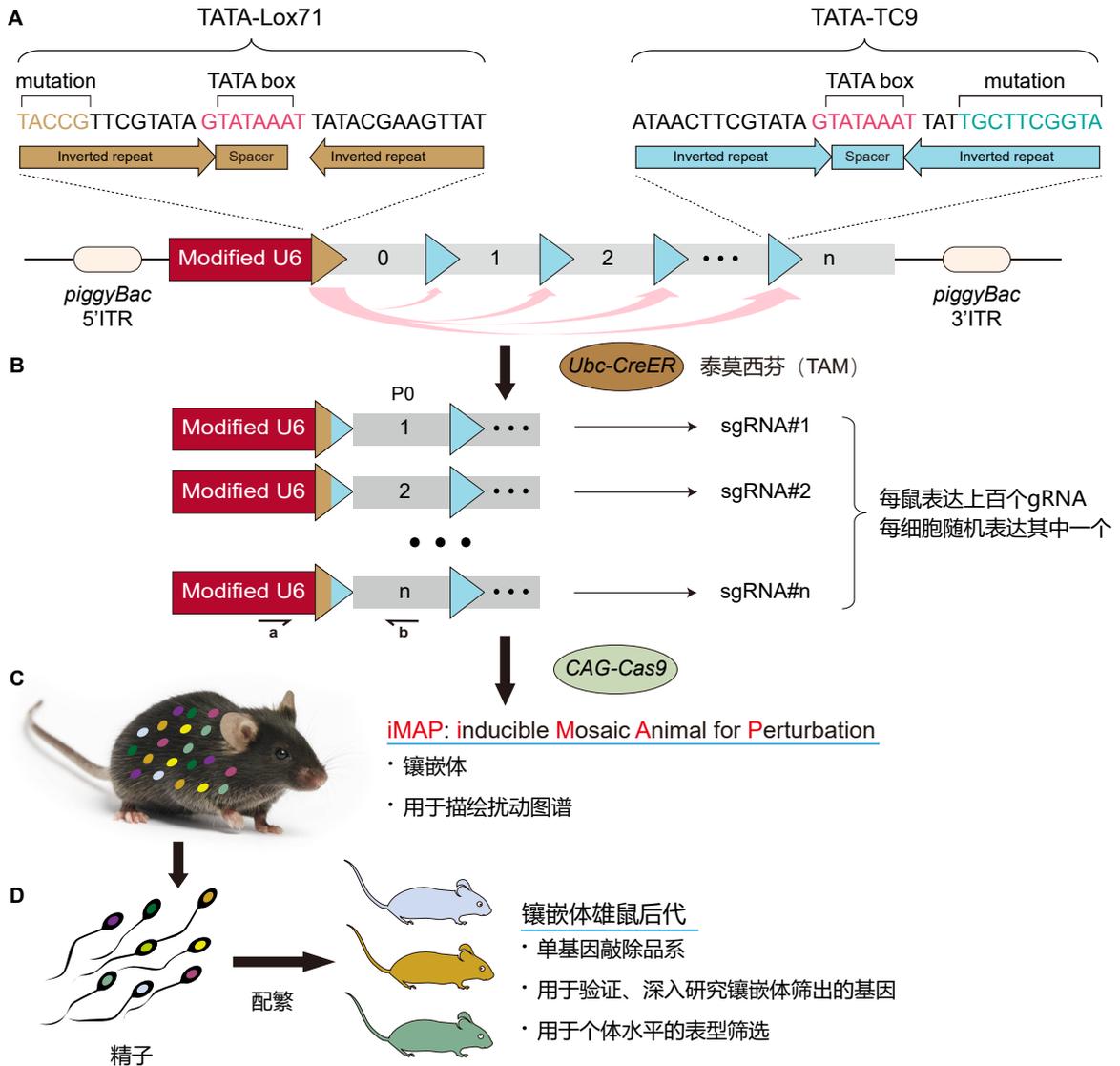


图 2. iMAP 原理。(A) 重组前的转基因。(B) 重组后的转基因。(C) 镶嵌体。(D) 镶嵌体衍生出单基因敲除品系。U6、Ubc 和 CAG 都是启动子，分别将 gRNA、CreER 和 Cas9 广泛表达于全身各部位。CreER 是 Cre 和 ER (改造过的雌激素受体) 的融合蛋白，其活性受泰莫西芬(TAM)诱导。a/b 是一对 PCR 引物，可放大出转基因重组后，所有位于 P0 的 guide。对 PCR 产物进行二代高通量测序，可鉴定出其中每种 guide 的相对丰度。比较实验组 (有转基因、CreER 和 Cas9) 和对照组 (缺 Cas9) 的 guide，可鉴定敲除靶基因对小鼠组织中细胞丰度的影响，从而揭示靶基因在细胞存活、增殖、分化过程中的生理功能。

带 gRNA 文库，从而无需用病毒递送。具体地说，iMAP 转基因融合了 CRISPR-Cas 和 Cre-Lox 技术，由四个元件组成：1) 一对 *piggyBac* 转座子的反向末端重复序列 (ITR)，位于转基因两侧，介导随机单拷贝转基因整合，这是每个细胞只表达一个 gRNA 的先决条件；2) 镶嵌了 LoxP 变体 TATA-lox71 的 U6 启动子，它具有双重功能，即启动 gRNA 转录和进行 Cre 介导的重组；3) 由多达 100 个 guide 串联而成的 array，每 guide 后有转录终止序列；4) 另一个 LoxP 变体 (TATA-TC9)，分隔各个 guide。静息状态下，只有紧邻 U6 启动子，即坐落在 Position 0 (P0) 的 guide (g0) 才能表达 (图 2A)。在 TAM 诱导下，CreER 被激活，使 TATA-Lox71 和 TATA-TC9 发生重组 (图 2A 三角形下的箭头)，剔除其间的序列，g1 到 gn 中的 guide 前移到 P0 而得以表达 (每细胞随机前移一个 guide；图 2B)，表达出来的 gRNA 再引导 Cas9 敲除靶基因，将小鼠转为镶嵌体。本工作中，CreER 和 Cas9 都被广泛表达，使小鼠全身细胞 (包括生殖细胞) 都转为镶嵌结构。因此，这种全身镶嵌体可用于普查上百基因在全身各种细胞的功能和挖掘疾病药靶 (即描绘扰动图；图 2C)，也能通过配繁，衍生出上百传统的单敲品系，因此大大提高其制备通量、降低其制备成本 (图 2D)。图 3 试以艺术形式再现 iMAP 的工作原理。

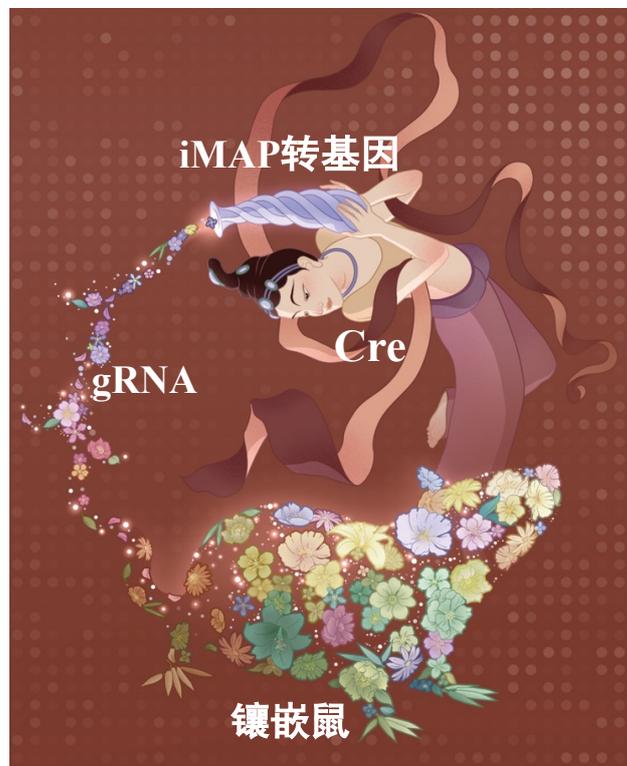


图 3. iMAP 原理的“敦煌版”，基于“飞天散花”典故。图示飞天 (Cre) 从花瓶 (双螺旋状转基因) 中释放出花朵 (gRNA)，后者再形成镶嵌鼠。设计：李自豪、池天

iMAP 的特色、优势总结如下：

- 操作简单。用 TAM 就可把整个小鼠转化为镶嵌体。
- 应用灵活。（1）可控制基因敲除的时空特征。例如，用广泛表达的 Cre，可使敲除发生在各种细胞，以普查基因功能，而用组织特异性表达的 Cre，则可选择性地研究特定组织。（2）用简单的二代测序技术鉴定 guide 丰度，可判断靶基因的基本生理功能（调控存活、增殖、分化，详图 2 的图注），而“高通量单细胞 RNA 测序”等高端技术则能提供更丰富信息。
- 结果可靠。细胞处于天然环境，无体外因素干扰。
- 易于普及。iMAP 品系一旦制成，即可像常规转基因鼠一样维持、扩增、分享，一劳永逸。

实验结果 为验证上述策略的可行性，选取 3 类功能已知的基因，即细胞必需基因、抑癌基因和细胞表面受体基因，其敲除应分别导致细胞丢失、细胞扩增和细胞标记物丢失。每类基因选取了两个基因，针对每个基因设计 10 个 gRNA（共 60 个），将其串联成“61-guide”转基因，发现其至少能稳定地传 13 代。在携带 61-guide 转基因并广泛表达 CreER 和 Cas9 的“iMAP-61”品系里，TAM 能使转基因重组，释放各种 guide。为鉴定 guide 的生理效应，检测了全身 10 种组织的表型，发现其与预期的基本一致（例如敲除抑癌基因能使某些组织明显扩增乃至发生肿瘤），表明了 iMAP 的可行性。另外，靶向同一基因的 10 个 guide，平均 9 个可测到活性，尽管这些 guide 未经事先验证，而较明显的脱靶效应仅见于一个数据点（靶向胚胎必需基因的一个 guide 在睾丸中被富集 2 倍）。这些数据不但表明 iMAP 切实可行，并且提示用一个（而不是多个）gRNA 靶向一个基因将是最佳设计，因为它使 iMAP 通量最大化但不明显增加假阳/阴性事件。最后，我们证明镶嵌体雄鼠可通过交配，衍生出单敲后代。这些结果提示，iMAP 研发成功。

2.用 iMAP 解码胚胎必需基因（EEG）的成体功能

如前所述，EEG 的成体功能难以用单敲品系研究。为回应这个挑战，我们挑选了 87 个在成体内广泛表达的 EEG，制备了新品系 iMAP-100，它共携带 100 个 guide，其中 87 个靶向 EEG，3 个靶向对照基因（包括 *Morf4l2*，其敲除不产生任何表型），8 个不靶向任何基因（NC）。用 TAM 诱导幼年鼠，3-4 个月后检查，发现部分小鼠出现多

囊肾，所有雄鼠睾丸萎缩（图 4A）。同时，检测上述 98 个 guide 在 39 种组织中的丰度，以揭示各基因分别在各组织的基本功能（3822 个数据点）。有 60% 的数据点，guide 丰度明显下降 (>2 倍)，远远超过丰度升高 (0.9%) 和不变 (39%) 的数据点。只有靶向 EEG（而不是非必需基因或不靶向任何基因)才可影响组织丰度。一个 EEG guide 至少影响一种组织，一般影响多个组织，有时乃至全部组织，表明所有 EEG 在成体依然都有重要功能。

我们用二维热图（“扰动图”；图 4B）展示这 3822 个数据点，它们簇聚成 6 群组织 (A-F)、5 群 guide (I-V)，其形式有明显生理意义：

- 功能相关的组织倾向于簇聚，如大脑与小脑（样本#1-2）、肠道组织（#19-23）、免疫细胞（#24-33）和男性生殖器官/细胞（#34-39）。
- 同样，功能相关的 guide 倾向于簇聚。例如，全部 NC (g1, g3-9) 和 g2- *Morf4l2*

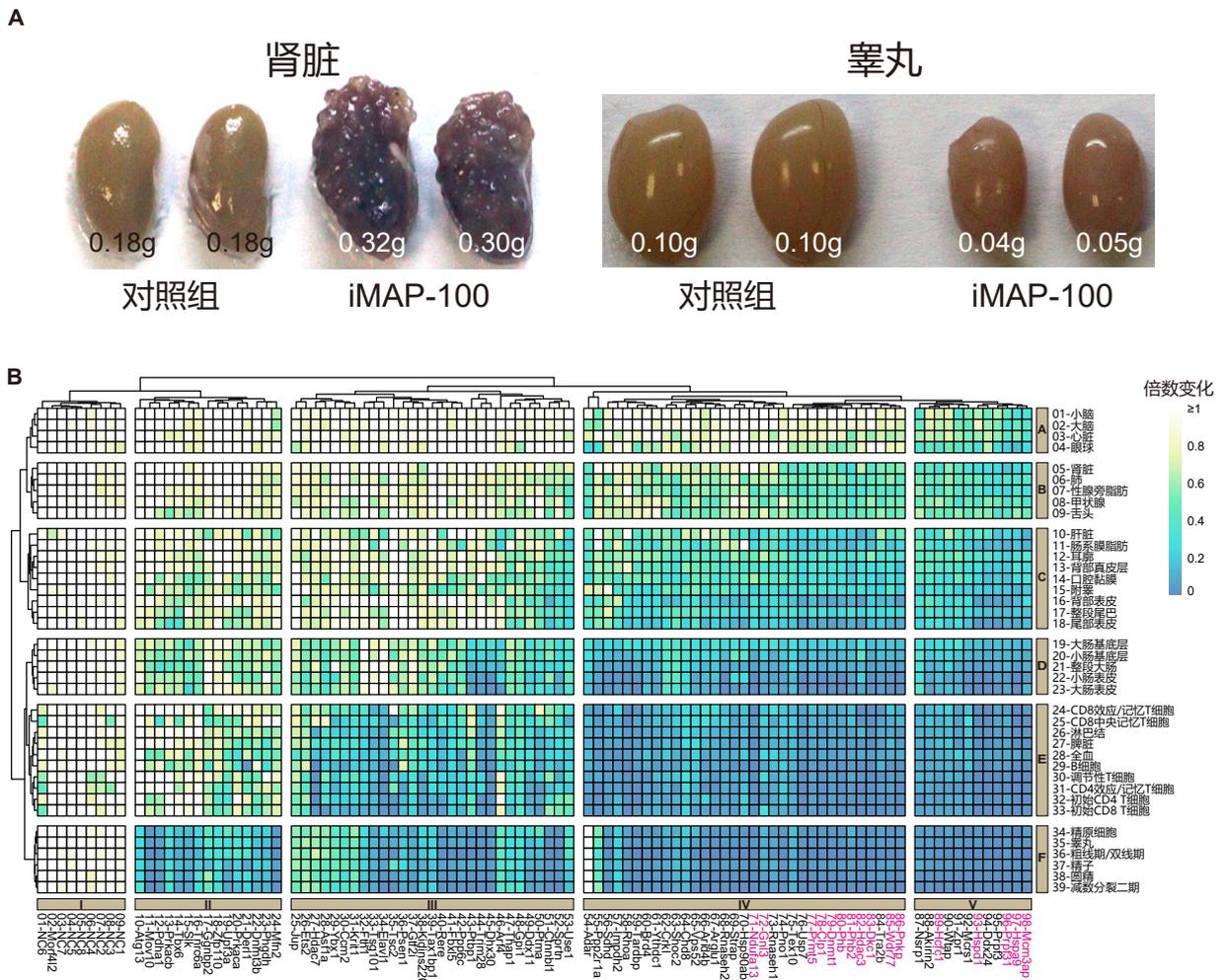


图 4. iMAP-100 表型。(A) 多囊肾和睾丸萎缩 (B) 扰动图。展示 98 个 guide 在 39 种组织中的丰度，其中丰度下降远比上升普遍。为清晰展示丰度下降，将丰度上升视为丰度不变（白色）。Guide 和组织样本都按其在热图上的排列次序编号。紫色显示人类 CEG 的同源基因。

（单敲无表型）簇聚成第 I 组。此外，有 3 对 *guide* 紧密簇聚，即 *g30-Ccm2* 和 *g31-Krit1/Ccm1*, *g80-Phb* 和 *g81-Phb2*, 以及 *g95-Prpf3* 和 *g96-Prpf31*。文献报道，这 3 对 *guide* 的靶基因分别编码各自蛋白复合物里的一对必需亚基（即敲除任一亚基足以摧毁复合物，因此两亚基分别敲除后表型同一），完美解释了簇聚形式^[14-16]。

- **Guide 簇聚反映单敲鼠表型。***g2-Morf4l2*（第 I 组）不影响细胞，正如 *Morf4l2* 单敲不产生任何表型，而 II-V 这 4 组 *guide* 导致逐渐加重的细胞丰度降低（可反映细胞死亡），正如其靶基因敲除导致逐渐提前的胚胎死亡（分别死于受精后第 12、10、7、6 天）。
- **Guide 簇聚反映人类基因功能。**相关基因是“核心必需基因”（core essential gene, CEG），它们调控细胞的基本生理过程，为细胞生存所普遍需要，在合成生物学和肿瘤研究领域有重要意义。人类有 684 个 CEG^[17]。我们发现，扰动图上的 90 个基因，有 16 个与人 CEG 同源（图 4B，紫色基因）。重要的是，这 16 个基因不是随机分布，而是富集于图谱右端的 27 个基因中 (*g71-98*)，表明它们也为小鼠多种乃至全部组织所必需，从而佐证了 *guide* 簇聚的生理意义。

我们尚未充分挖掘 3822 个数据点，但已发现很多重要信息，其中多为文献未曾报道，而已报道的一般符合我们的发现。以下仅描述部分结论。

- **EEG 促进细胞生成、存活，但有例外。**具体地说，靶向 EEG 的 *guide* 一般都降低细胞丰度，但有 5 个 *guide* 在部分组织中被富集，其代表是 *g35-Tsc2*（靶向 *Tsc2* 的 *guide*，热图横坐标第 35 号），它在肾中被富集 95 倍。同时，肾脏变成多囊肾（图 4A 左），提示 *Tsc2* 敲除能促进肾细胞大量增殖，以至于镶嵌体中极少数细胞最终产生器官水平的表型（出现多囊肾）。这与文献报道吻合：*Tsc2* 在小鼠敲除后，可诱发肿瘤和多囊肾^[18, 19]。
- **胚胎发育所必需的基因一般也为精子发生所必需，提示精子发生与胚胎发育有某种神奇、意外的共性，也解释了睾丸独特的表型（萎缩；图 4A 右）。**
- **EEG 中隐藏着潜在医疗靶点。**

***Mov10* 和 *Mfn2*：男性避孕药。**文献报道，两基因广泛表达，但其敲除不影响体细胞存活。与此一致，*g11-Mov10* 和 *g24-Mfn2* 不影响各种体细胞，但出乎意料的是，两者强烈抑制精子生成（使精原细胞和由其衍生的各种生殖细胞

丰度降低 33-37 倍)。因此, *Mov10* 和 *Mfn2* 是潜在的男性避孕药靶点, 其靶向药有可能高效而安全地阻断精子生成。为什么 *Mov10* 和 *Mfn2* 如此“青睐”生殖细胞而“忽视”体细胞? 这依然是谜, 因为 *Mov10* 和 *Mfn2* 功能分别是抑制转座子和调控线粒体, 两者都是基本生理过程, 对体细胞也很重要。也许体细胞里, 有“备胎”基因, 其冗余作用遮盖了 g11-*Mov10* 和 g24-*Mfn2* 的破坏效应。

***Der11*: 肿瘤靶向疗法。**人类 *DERL1* 在多种肿瘤中都是重要的致癌基因^[20]。扰动图显示, g21-*Der11* 丰度减少只见于 B 细胞和男性生殖细胞, 提示 *Der11* 在生理条件下的主要功能是促进 B 细胞和精子发育, 虽然其突变可导致多种肿瘤。因此, *Der11* 小分子抑制剂可能是相对安全的广谱抗肿瘤药物。

***Hdac7*: 肿瘤免疫疗法。**治疗肿瘤的一个重要策略是从人体分离 T 细胞, 在体外改造(插入人工抗原受体 CAR 等), 强化其功能, 再回输病人^[21]。另一策略是利用分子药物, 在病人体内刺激免疫反应^[22]。两种策略对多数病人无效, 但对少数病人有特效(导致 2018 相关诺奖), 显示了巨大的医疗潜力。如何优化免疫疗法, 是热点医学问题。CD8 T 细胞是关键的反瘤细胞。在肿瘤抗原的刺激下, “初始” CD8 T 细胞被激活, 分化成“效应/记忆”细胞, 以杀死肿瘤。g27-*Hdac7* 在初始 CD8 T 细胞严重减少, 但在效应/记忆 CD8 T 细胞中明显恢复(比较组织#33 vs #24-25), 提示 *Hdac7* 的敲除能促进 CD8 T 细胞激活、分化。我们推测, 在人类 CAR T 细胞敲除 *HDAC7*, 能提高其抗癌功能(细胞因子分泌和肿瘤杀伤等)。令人兴奋的是, 这被实验证实(图 5A-D)。同时, 我们发现, g27-*Hdac7* 对内脏细胞丰度无明显影响(图 5E), 提示 *Hdac7* 小分子拮抗剂可抑制肿瘤而无明显毒性。的确, 文献报道, 在乳腺癌小鼠模型中, *Hdac7* (以及其他几个同类酶) 的拮抗剂 **TMP196** 可安全有效地抑制肿瘤, 这过程伴随着 CD8 T 细胞的激活^[23]。值得指出, 这项工作还发现 **TMP7** 导致的 CD8 T 细胞激活需要巨噬细胞介导, 但这当然不能排除 **TMP7** 对 CD8 T 细胞还有直接作用, 因此与我们的数据不矛盾。

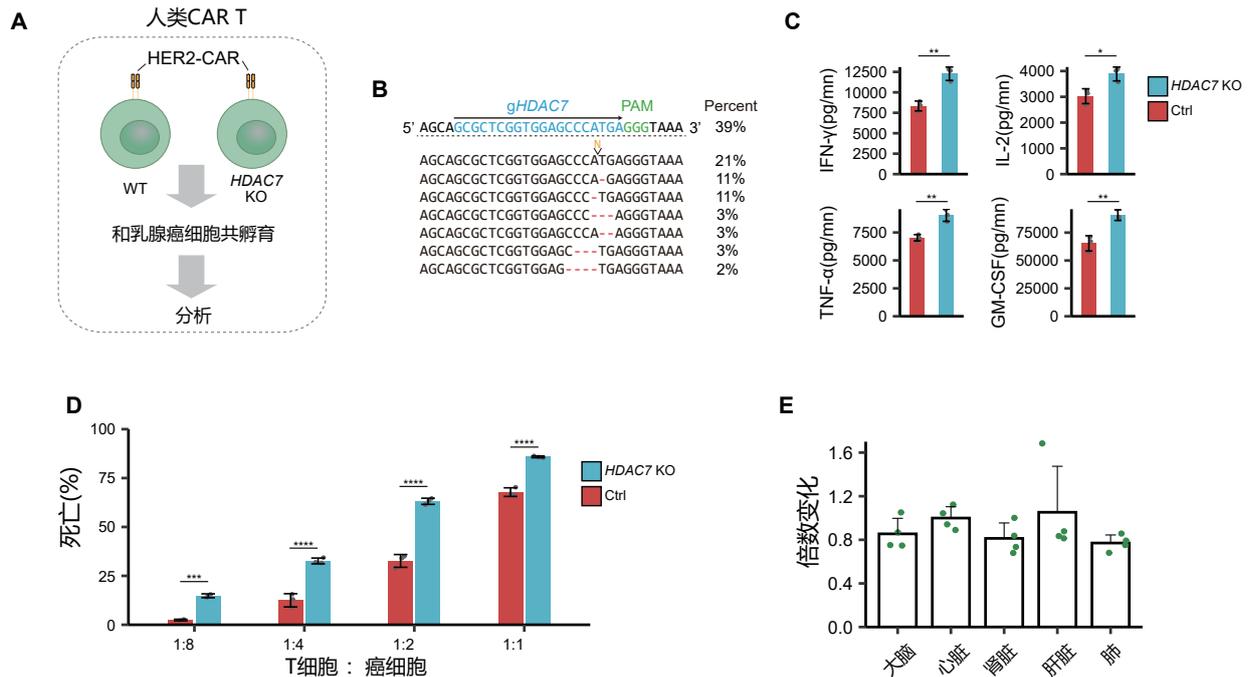


图 5. *HDAC7* 敲除强化人类 CAR T 细胞。(A) 实验流程。从健康人外周血分离 T 细胞，制备 HER2-CAR T，利用 Cas9 敲除 *HDAC7*，再将细胞与癌细胞共培养，分析 CAR T 细胞的细胞因子表达水平与肿瘤杀伤力。(B) *HDAC7* 位点被成功编辑。敲除 *HDAC7* 后，用 PCR 放大相关序列，进行二代测序，发现只有 39% 的序列未必编辑，其余序列都出现插入或缺失突变，其中大部份是移码突变。(C) *HDAC7* 敲除增强了 CAR T 细胞因子的表达。(D) *HDAC7* 敲除提高了 CAR T 杀癌功能（表现为癌细胞死亡率）。(E) *HDAC7* gRNA 不明显影响 iMAP-100 鼠内脏细胞活性。数据取自图 4B 的扰动图。

■ 总结

我们将 CRISPR-Cas 和 Cre-Lox 融合成突破性的基因解码技术 iMAP，并利用 iMAP 描绘了微型扰动图，快速揭示了 90 个基因分别对 39 种组织存活、增殖、分化的调控功能，发现了多个潜在药靶。简单、通俗地说，iMAP 的解码速度比传统单敲技术高出大约 100 倍，相当于汽车（时速 100 公里）与慢行（时速 1 公里）的区别。论文发表后，引起了广泛关注。原位 CRISPR 筛选的先驱、斯坦福大学教授 Monte Winslow 等在《细胞报告-方法》（*Cell Reports Methods*）上发表题为“用以分析基因功能的一种镶嵌体小鼠模型”的文章，全面评述 iMAP，强调了其创新性以及在揭示发育、稳态和疾病机制方面的价值和前景^[24]。《自然-生物技术》（*Nature Biotechnology*）和《自然-结构与分子生物学》（*Nature Structural and Molecular Biology*）都在“研究亮点”

(Research highlights) 中介绍了 iMAP。在中文世界，学术杂志《遗传学》和大量科技、公众媒体也对 iMAP 进行了报道，包括《麻省理工科技评论》、《中国科学报》，iNature, BioArt、奇点网、《新民晚报》、《文汇报》、中新社、澎湃新闻、浦东发布等。此外，已有著名药企对 iMAP 表示兴趣（专利授权洽谈中）。

■ 发展前景和展望

iMAP 应用广泛，至少有以下未来研究方向。首先是描绘单细胞水平的全景扰动图。在战略上，随着 iMAP 和微型扰动图的问世，随着高通量单细胞测序、计算生物学、人工智能等尖端技术的飞速发展，全景图已呼之欲出。但是，在战术上，全景图的描绘是规模宏大的工程，需各国科学家的通力合作，而中国应能起关键作用。回顾人类基因组计划，我国仅贡献了 1% 的序列，但在过去 20 年取得了惊人的进步，必定能在后基因组时代做出更大贡献。其次，iMAP 可用于揭秘非编码序列、改善家畜和农作物性状等。最后，可将 iMAP 偶联到不同的 CRISPR 技术平台，以拓展其应用场景；这些平台包括 CRISPRa（激活基因表达）、CRISPRi（沉默基因表达）、碱基编辑器（引入或修正点突变）和 Cas13（降解 RNA），后者很适合揭秘非编码 RNA^[25-28]。

原创技术刚问世时，一般都不成熟，iMAP 也不例外。例如，iMAP 通量目前是 100 基因/品系，需多达 200 个品系才能覆盖全基因组的 2 万基因，这妨碍了全景图的描绘。我们试图将通量提高 10 倍，这将极大地赋能 iMAP，类似变汽车为飞机（时速 100 变为 1000 公里）。

破解生命天书，是 21 世纪生命科学光荣而艰巨的历史使命。在大家（尤其是年轻一代）共同努力下，也许在不久的将来，人类就能完成这一使命，最终使人类健康发生质的飞跃，实现其千古梦想。

参考文献

- [1] Jones R C, Karkani J, Krasnow M A, et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans[J]. *Science*, 2022,376(6594):14896.
- [2] Smith C, Bergman P, Hagey D W. Estimating the number of diseases - the concept of rare, ultra-rare, and hyper-rare[J]. *iScience*, 2022,25(8):104698.
- [3] Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets[J]. *Nature Reviews Drug Discovery*, 2017,16(1):19-34.
- [4] Hall B, Limaye A, Kulkarni A B. Overview: generation of gene knockout mice[J]. *Current Protocols in Cell Biology*, 2009,Chapter 19:12-19.
- [5] Collins F S, Finnell R H, Rossant J, et al. A new partner for the international knockout mouse consortium[J]. *Cell*, 2007,129(2):235.
- [6] Collins F S, Rossant J, Wurst W. A mouse for all reasons[J]. *Cell*, 2007,128(1):9-13.
- [7] Kim H, Kim M, Im S K, et al. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes[J]. *Laboratory Animal Research*, 2018,34(4):147-159.
- [8] Wang J Y, Doudna J A. CRISPR technology: A decade of genome editing is only the beginning[J]. *Science*, 2023,379(6629):d8643.
- [9] Doench J G. Am I ready for CRISPR? A user's guide to genetic screens[J]. *Nature Reviews Genetics*, 2018,19(2):67-80.
- [10] Wang H, Yang H, Shivalila C S, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering[J]. *Cell*, 2013,153(4):910-918.
- [11] Shen B, Zhang J, Wu H, et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting[J]. *Cell Research*, 2013,23(5):720-723.
- [12] Kuhn M, Santinha A J, Platt R J. Moving from in vitro to in vivo CRISPR screens[J]. *Gene and Genome Editing*, 2021,2:100008.
- [13] Liu B, Jing Z, Zhang X, et al. Large-scale multiplexed mosaic CRISPR perturbation in the whole organism[J]. *Cell*, 2022,185(16):3008-3024.
- [14] Baranoski J F, Kalani M Y, Przybylowski C J, et al. Cerebral Cavernous Malformations: Review of the Genetic and Protein-Protein Interactions Resulting in Disease Pathogenesis[J]. *Frontiers in Surgery*, 2016,3:60.
- [15] Graziotto J J, Farkas M H, Bujakowska K, et al. Three gene-targeted mouse models of RNA splicing factor

- RP show late-onset RPE and retinal degeneration[J]. *Investigative Ophthalmology & Visual Science*, 2011,52(1):190-198.
- [16] Ross J A, Nagy Z S, Kirken R A. The PHB1/2 phosphocomplex is required for mitochondrial homeostasis and survival of human T cells[J]. *The Journal of Biological Chemistry*, 2008,283(8):4699-4713.
- [17] Hart T, Tong A, Chan K, et al. Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens[J]. *G3 (Bethesda, Md.)*, 2017,7(8):2719-2727.
- [18] Onda H, Lueck A, Marks P W, et al. Tsc2(+/-) mice develop tumors in multiple sites that express gelsolin and are influenced by genetic background[J]. *The Journal of Clinical Investigation*, 1999,104(6):687-695.
- [19] Ren S, Luo Y, Chen H, et al. Inactivation of Tsc2 in Mesoderm-Derived Cells Causes Polycystic Kidney Lesions and Impairs Lung Alveolarization[J]. *The American Journal of Pathology*, 2016,186(12):3261-3272.
- [20] Fan J, Tian L, Huang S, et al. Derlin-1 Promotes the Progression of Human Hepatocellular Carcinoma via the Activation of AKT Pathway[J]. *OncoTargets and Therapy*, 2020,13:5407-5417.
- [21] Singh A K, McGuirk J P. CAR T cells: continuation in a revolution of immunotherapy[J]. *The Lancet Oncology*, 2020,21(3):e168-e178.
- [22] Kerr W G, Chisholm J D. The Next Generation of Immunotherapy for Cancer: Small Molecules Could Make Big Waves[J]. *Journal of immunology (Baltimore, Md. : 1950)*, 2019,202(1):11-19.
- [23] Guerriero J L, Sotayo A, Ponichtera H E, et al. Class IIa HDAC inhibition reduces breast tumours and metastases through anti-tumour macrophages[J]. *Nature*, 2017,543(7645):428-432.
- [24] Cai H, Winslow M M. A new system for multiplexed mosaic analysis of gene function in the mouse[J]. *Cell Reports Methods*, 2022,2(9):100295.
- [25] Kampmann M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine[J]. *ACS Chemical Biology*, 2018,13(2):406-416.
- [26] Chen P J, Liu D R. Prime editing for precise and highly versatile genome manipulation[J]. *Nature Reviews Genetics*, 2023,24(3):161-177.
- [27] Liu Y, Mao S, Huang S, et al. REPAIRx, a specific yet highly efficient programmable A > I RNA base editor[J]. *The EMBO Journal*, 2020,39(22):e104748.
- [28] Huang X, Lv J, Li Y, et al. Programmable C-to-U RNA editing using the human APOBEC3A deaminase[J]. *The EMBO Journal*, 2020,39(22):e104741.