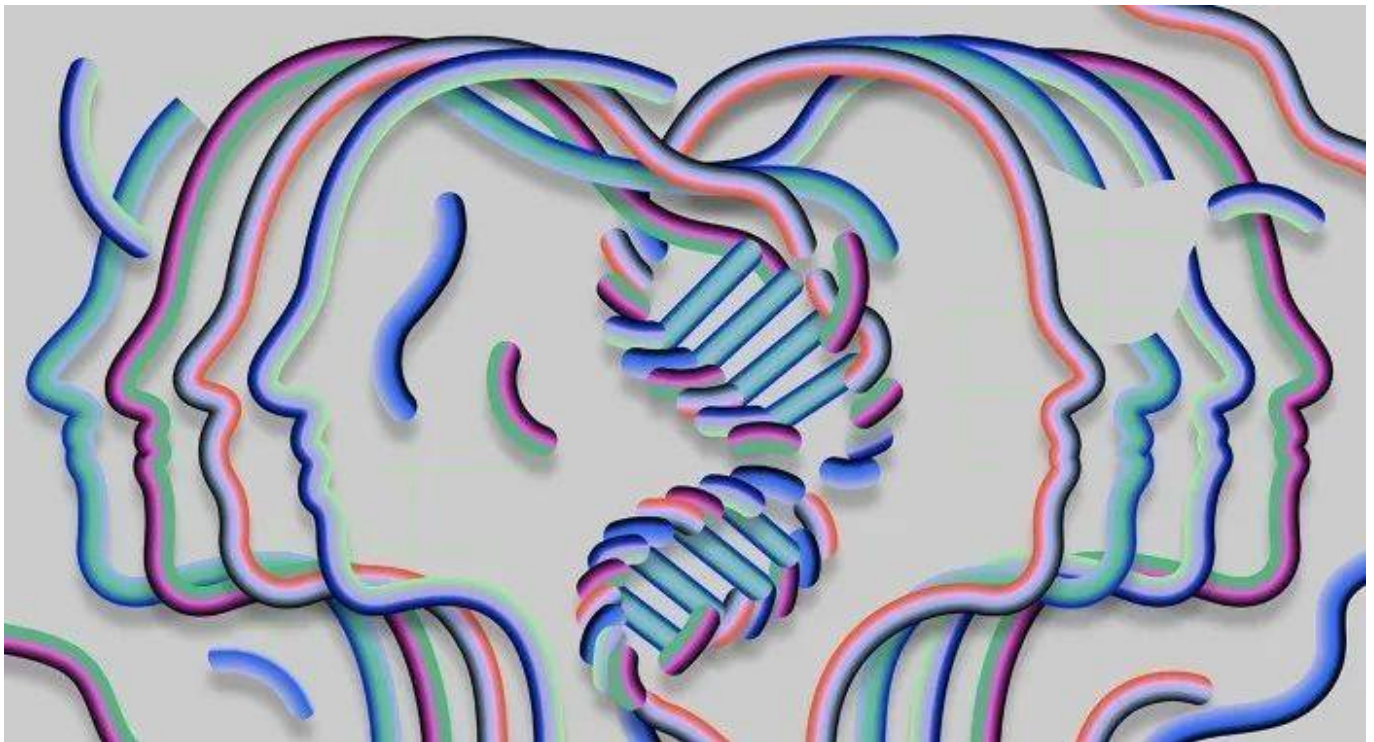


“人类泛基因组计划”来了，能体现人类的全部多样性吗？

几年前，美国华盛顿大学的遗传学家Evan Eichler彻底搜索了人类基因组中的未知变异，偶然获得了一些不寻常的发现。他的研究团队发现了一段长约40万个碱基的DNA，其中包含了额外的基因副本——很可能是从一个名为丹尼索瓦人的古人类群体遗传下来的[1]。这段DNA在约80%的巴布亚新几内亚居民体内存在，但别处几乎没有。

“这个长度太惊人了，”Eichler说，“我们一直知道，基因组中会有古老的片段。”但是这一片段的长度以及它在别处的缺失，“刷新了我们的想法”。

包括这一发现在内，一个个意想不到的发现，使得Eichler和其他遗传学家越来越无法满足于现有人类基因组图谱的广度和深度。第一版基因组图谱源自耗资27亿美元的人类基因组计划，在2001年发布，当时旨在成为未来基因研究的参考系。但**图谱中93%的序列仅来自11个人**，这些人大多是通过纽约州水牛城的报纸广告招募的；高达70%的DNA都来自同一名男子。



插图：Ana Kova

****到了2003年，新版的参考基因组图谱GRCh38在技术上完成了，但它仍然包含数百段空缺和含有大量错误的部分。这些缺陷并非无关痛痒。Eichler与本校医学中心的临床遗传学家合作发现，该参考基因组缺少一个区域，其中包含与Baratela-Scott综合征相关的变异，该综合征可导致儿童认知迟滞和骨骼畸形。因为缺失这一块，医生无法快速检测出这里的DNA错误。**

尽管基因组图谱有所改进，但还不能充分体现人类的巨大多样性。例如，2018年，一组研究人员对910名非洲裔个体进行测序，发现了一段由3亿个DNA字母——或称为碱基——组成的陌生序列[2]，这大概是整个基因组的10%。

为了获得更完整、更有代表性的参考，Eichler与许多著名科学家（大多在美国）携手合作。**他们的目标是尽量查明全部的人类基因变异——人类基因中数不胜数的“混音”，包括添加、缺失和其他类型的突变。**

基因组不是来自单个个体的线性读数，它更像是包含着无穷分支与融合的多重路径，就像错综复杂的伦敦地铁线路图。这体现在不同人群中发现的序列多变性，例如巴布亚新几内亚许多人身上发现的那一大段DNA。

2019年，在美国国家人类基因组研究所（NHGRI）资助下，Eichler的团队启动了耗资3000万美元的**人类泛基因组计划**（Human Pangenome Project）。最初的目标是对大约350名不同背景的人进行参考级质量的详细基因组测序，并尽量自由共享这些数据。

这项工作提出了重大技术挑战，不过幕后的科学家们（包括加州大学的Karen Miga和华盛顿大学医学院的Ting Wang）感到大有可为。他们认为这是使基因组医学实现公平的关键[3]。“阐明多样性是为了以人为本，”Wang说，“宗旨是为了兼顾公正与平等，是为了建立服务人类的、更具包容性的基因组资源。”

泛基因组计划的研究人员反思了人类基因多样性研究的历史，有些研究项目采集边缘人群的数据却不尊重他们的需求与愿望，从而被视为“吸血鬼”项目。对此，**泛基因组计划全程聘请生物伦理学家，而不是像过去那样隔段时间才邀请他们一回。**“伦理学家们不会再闭目塞听、单独工作，而是参与到项目的每一步，包括所有的技术决策。”Wang说。

然而，一些关注原住民社区需求的遗传学家对这项倡议持谨慎态度。他们没有呼吁终止人类泛基因组计划本身，但他们认为边缘人群理应掌控自己的基因数据和测序者。“我们自己掌控这些技术，就是在赋予社群权益，”加州大学遗传学家、夏威夷原住民Keolu Fox解释说，“谁也没我们这些自己出身社群的人靠谱。”

初见成效

泛基因组的理念可以追溯到对“无乳链球菌”（*Streptococcus agalactiae*，又称B族链球菌）的研究，这种细菌可引发致命的新生儿感染。2005年，科学家分析了该菌的6个菌株并发表了论文，试图阐明同种细菌的所有基因细微差别[4]。他们发现了一个由全部6个菌株共享的核心基因组，一个部分共享的“非必需”基因组，以及一些菌株独有的基因。

这是一项棘手的任务，因为凭借名为水平基因转移的机制，细菌能互换和分享DNA片段，甚至能跨物种进行。“细菌会出的状况可多了。”明尼苏达大学的植物遗传学家Candice Hirsch说。因此，生物学家要持续更新细菌的参考基因组。相比之下，人类要增加新变异就不那么容易。这让分析人类泛基因组的可行性更高了，Hirsch说。



遗传学家Krystal Tsosie。来源：Tomás Karmelo Amaya/New York Times/Redux/eyevine

不过人类基因组变化力不足，长度和重复性却有余。例如，人类24条染色体中最长的1号染色体长度超过2.5亿个碱基对，这是无乳链球菌的100多倍。该染色体上散布着简单重复的序列和其他更复杂的重复片段。直到最近十年里，科学家们进行测序的主要手段还是把DNA打碎，再一小段一小段地解读。这使他们能相对容易地检测到DNA中的单个碱基变化。但在一段DNA包含某个基因的副本超过一个时，这种短读方式就会难以识别。作为鉴定基因结构变异（如基因重复和缺失）的专家，**Eichler选择了一种名为“长读测序”的新方法，一次能分析更长的DNA片段。*这使他能发现巴布亚新几内亚人身上先前被忽视的变异。

2018年，Eichler等科学家在NHGRI共聚一堂，讨论人类泛基因组计划。在那里，Eichler重逢了一位同样热衷于长读测序技术的同行——纽约洛克菲勒大学的神经科学和分子生物学家Erich Jarvis。

“我们那时老是举手对人家说，‘没有高质量的参考基因组，您这事儿就成不了’。”Jarvis回忆道。但是长读测序更花钱，也不是每个人都喜欢用它。Jarvis还记得在某些辩论中备受打击。“我上餐厅吃饭时因为咬得太用力，门牙都被叉子崩掉一块。”他说。最后，他们这些推行长读测序的人赢了。

Miga那时已经在用长读测序技术了，后来她因完成DNA难读片段测序而使人类泛基因组计划名声大噪。她和Jarvis、Eichler等人共同发表了有史以来首次完全测序的人类基因组，检测了全部30亿个碱基，包括覆盖在染色体末端、混乱又高度重复的部分——端粒[5]。**这是第一个端粒到端粒的基因组序列**，纠正了先前参考中的许多错误，并发现了大约100个被忽视的可能编码蛋白的基因（参见：[《胜利在望：即将完整的人类基因组》](#) | [《自然》技术特写](#)）。

然而，这成果来之不易。通常，人类细胞包含两组23条染色体——一组来自卵子，另一组来自精子。但当机器试图同时读取两组序列时，重复序列和其他DNA结构变异就会混淆。为了解决这个问题，科学家们使用来自所谓“葡萄胎”的细胞系进行DNA分析。葡萄胎是无细胞核的卵子受精后形成的，其DNA只含有一组染色体。

相比之下，人类泛基因组计划的350个基因组将来自二倍体细胞系，**即包含父母双亲DNA的细胞**，这样科学家就必须使用复杂的计算工具来分离基因组，以确保能够准确检测出结构变异。

泛基因组计划已经完成了大约70个详细的基因组。他们的目标是在2024年中项目结束前，完成全部350个端粒到端粒的基因组序列。

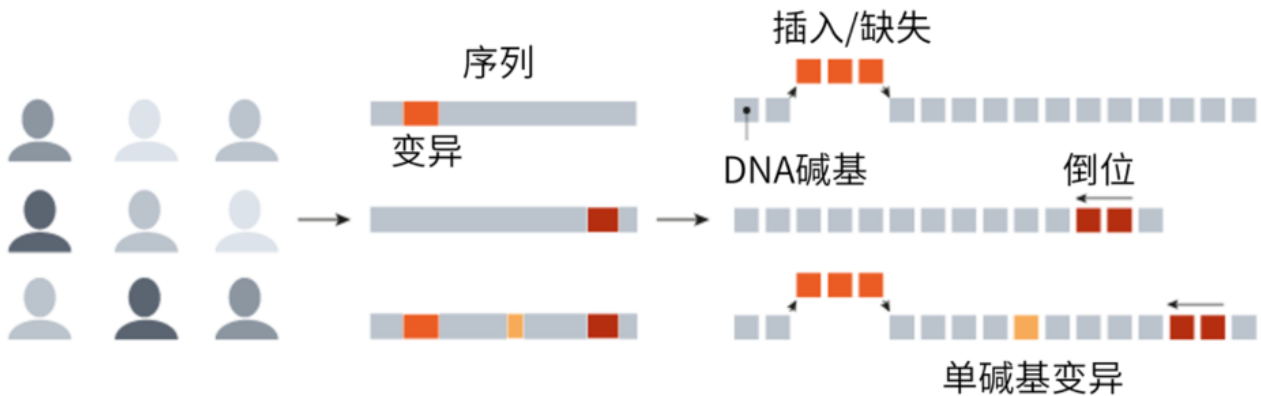
科学家们已经开始尝试将DNA的多样性可视化，呈现这些变异。迄今为止，包括GRCh38参考基因组在内，常规方式都是一个简单的线性呈现，在相伴的数据库中罗列变异在序列里的不同位置，如单碱基变异。“科学界使用这种简化版的线性参考序列已有20年了。”加州大学的计算生物学家Benedict Paten说。Paten的办公室就在Miga的隔壁，他的团队正在协作改善泛基因组可视化的精细程度。在这种新的可视化中，彩色线条代表不同的变异。变异频率越高则线条越粗。“使人类的变异更直观易懂，也是我们整合泛基因组的任务之一。”Paten说（见“泛基因组可视化”）。

泛基因组可视化

人类泛基因组计划旨在获得全世界人类基因组中的所有变异度。这项工作对这些变异进行分析，并以全新方式呈现，从而打破了人类基因组只能以固有形式呈现的成见。

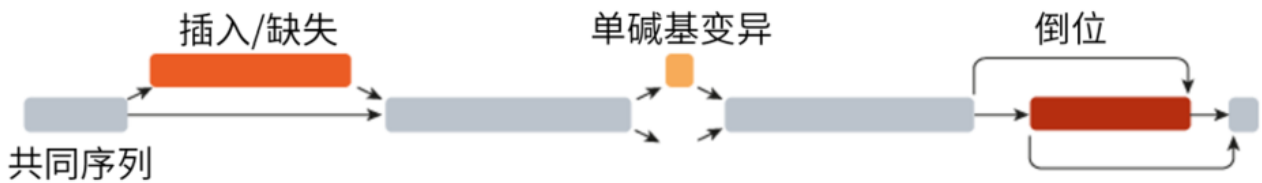
收集样本

研究者必须对数百人进行高质量测序并统计变异类型，包括单碱基变异、插入、缺失和倒位。



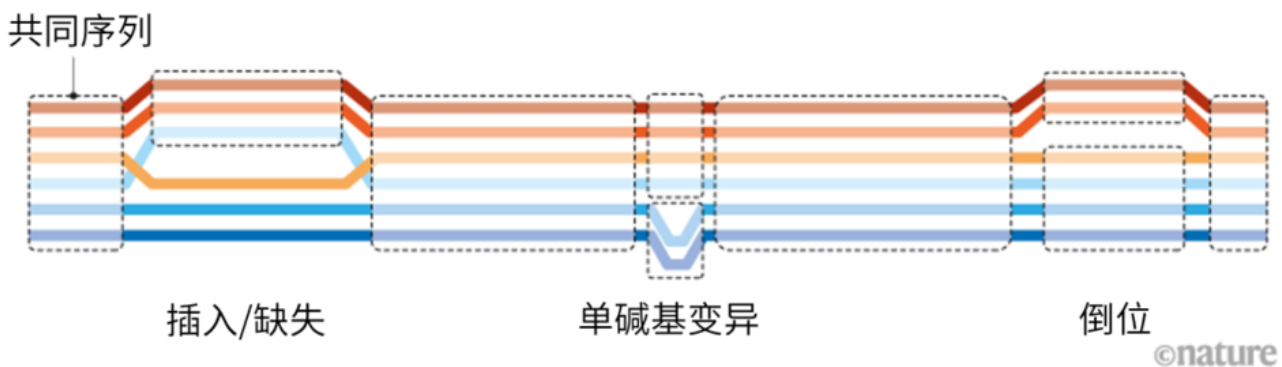
变异可视化

以图像模型呈现变异数据，无需预设标准或默认的参考基因组。



探索泛基因组

地铁线路图般的表现形式能帮助研究者在测序水平比较人群中的变异。



错误与进步

在即将参加人类泛基因组计划的350个受试者中，有许多人的基因组已在2008年启动的“千人基因组计划”中得到分析——该项目旨在对26个不同人群中的常见和罕见变异进行分类统计。该项目中采集冻存的DNA样本将被取出，重新用于更详细的泛基因组计划的长读测序。这些人几年前签署的知情同意书中覆盖了新项目继续使用其DNA数据。但是人类泛基因组计划采取进一步措施确保对基因数据的收集和使用符合伦理。Eichler说，例如，

其他大型基因测序项目中都是科学家做决策，随后只由机构评审委员会审查，但人类泛基因组计划的社会伦理学家们深度参与决策过程，并持续审查项目。

正如Wang所言：“主要是，如何指导那些不太考虑社会问题的科学专才们，以最合情理的方式做好他们的科学研究。”

泛基因组计划的领导者试图多管齐下，解决以往项目遗留的棘手伦理问题。例如，1991年启动的、旨在收集全球人类DNA信息的“**人类基因组多样性计划**”，就曾受到一些社区强烈反对。其中，原住民群体尤其感到该项目对待他们的方式好像在看濒临灭绝的活化石[6]。

“科学家收集原住民基因组数据主要是为了其他非原住民的利益，这项工作如果无视原住民的数据主权，就成了一种无限攫取数据的手段。”Krystal Tsosie说，她是美国范德堡大学的遗传学和生物伦理学家，纳瓦霍族的成员。

那之后十年里，基因研究对弱势群体的伦理侵犯甚至引发了更大的风波，特别是在2004年，哈瓦苏派部落起诉了亚利桑那董事会和亚利桑那州立大学的研究人员。该部落成员曾捐献DNA用于2型糖尿病的遗传研究，却发现现在未征得他们同意的情况下，该DNA被用于精神分裂症和人口迁移研究[7]。

Tsosie说，这些研究人员还使用了“近亲繁殖”等污蔑性词汇来解释实际上由人口瓶颈导致的遗传现象——**而这种人口瓶颈实际上和种族灭绝有关**。她还说，过去开展测序工作的遗传学家经常使用种族歧视性语言，而没有正确地认识到殖民主义对科学造成的漫长后遗症，以及它对原住民造成的威胁。

几年来，Fox等人一直在呼吁大范围纠正这种做法。他们认为，原住民群体在接受基因数据采集时，应该有更大的话语权。Fox是Eichler实验室的一名研究生，他说，他不认为泛基因组计划之类的项目能以真正促进权利的方式惠及提供样本的各个人群。“我爱Evan，伙计。当我有问题时会打电话给他寻求建议，”他说，“不过，你看，我们也不是每件事情想得都一样。”



包括Joseph Yracheta（最右）在内的原住民生物数据联盟（Native BioData Consortium）成员正在使用基因组测序仪。来源：Joseph Yracheta

Fox主张**把测序权力交到人民手中**。他和Tsosie参与了原住民生物数据联盟，这是一个由美国原住民科学家和部落成员领导的非盈利性研究机构，一直致力于帮助原住民群体获得DNA测序仪，并在他们自己的地区测序。该联盟的共同创始人Joseph Yracheta是约翰霍普金斯大学彭博公共卫生学院的公共卫生遗传学家，他表示，首台测序仪已于2020年12月被送到夏延河苏族保留地。今年2月，Yracheta加入了人类泛基因组计划的一个工作组，研究该计划的伦理、法律和社会影响。

Fox目前致力于研究太平洋岛屿上发现的遗传复杂性。他和同事们使用一种整体的方法，同时对农业物种和周围环境中的其他生物进行串联测序。并正在着手建立一个基因组学研究所，用于服务这里的社区。Fox指出，最新的技术，比如“分布式记账”计算机系统可以让个人安全连接他们的基因数据，给予人们更高的自主权来决定允许谁访问和使用他们的信息。“现在数据科学的诸多进步，确实使受试者的权利达到了新高度。”Fox说。

Eichler对Fox选择的道路表示支持。“我赞赏他邀请原住民科学家参与基因组学研究的努力——我们需要更多这种努力，”Eichler说，“不过我觉得，这不是非此即彼的情况。”他补充说，人类泛基因组计划鼓励原住民科学家创建他们自己的参考基因组。在这种情况下，“我们会提供必要的专业知识和设备，以共同实现这一目标。”

一切突变，皆得呈现

Tsosie说，原住民们未来可能会与大型多样性项目合作，但其形式必须能保障这些社区可以进行自己的测序。另外，尽管这类大型基因组项目通常是数据开放的，Tsosie却认为**最好对原住民DNA序列的存储数据进行额外保护**，例如只能通过访问请求获得数据，以避免剥削利用。她说：“要做就要做到最好，能够真正代表原住民。”

顾虑代表性和数据所有权问题的，不只是美国原住民社区的活动家。Jarvis（他是该计划的取样委员会成员）说，也有人批评泛基因组计划没有充分吸纳美国以外的研究人员。他承认，有些人认为这项计划主要是美国的工作，但Jarvis说他们团队正在努力扩大影响，让世界各地的科学家和受试者都参与进来。例如，他们联系了非洲人类遗传与健康（H3Africa）项目的领导人，希望吸纳非洲科学家，让他们能在非洲国家进行测序。（然而，看来没有哪个测序项目能够避免伦理难题——比如说，H3Africa项目也不得不勉力适应各国管理受试者数据使用的法律法规。）

Jarvis说，他希望人类泛基因组计划能更好地体现人类基因多样性。“我是有色人种，作为非裔美国人，在科学界属于少数派弱势群体，”他说，“我这边的多样性没有得到体现。因此，我于公于私都希望确保这个泛基因组能真正代表人群。”

在推进研究的同时，科学家们也承认**350个基因组并不能体现人类的全部多样性**。想要做到这点，最终实际需要的基因组数量还难以确定，而遗传学经验告诉我们，罕见的差异也可能至关重要。“我不认为有什么一劳永逸的神奇数字。”Adam Phillippy说。他是NHGRI基因组信息学部门的负责人，也是泛基因组项目的研究员。

在忙于繁重科研任务的同时还要避开伦理隐患，让泛基因组的研究人员倍感压力。“我知道，我们将做的事情里总会有一些在未来的5到10年内受到批评——这点我几乎100%肯定，”Eichler说，“但如果我们能一直问心无愧，能说我们尽己所能地做好每件事，我觉得那就挺好。”