
基于机器学习建立精确的人类遗传重组图谱

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/17266.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

基于机器学习建立精确的人类遗传重组图谱。

1月20日，中国科学院上海营养与健康研究所李海鹏研究组等在Human Genetics上，在线发表了题为Fine human genetic map based on UK10K data set的研究论文。遗传重组是生命进化的基础，在有性生物形成配子的过程中，来自父方和母方的染色体相互交换遗传物质，从而增加了相邻基因间不同等位基因的组合，丰富了遗传多样性。精确度量基因组不同区域的遗传重组率是生物研究的热点。精确的遗传重组图谱对探究遗传重组的发生机制、杂交育种、准确定位致病突变和某一性状的决定基因，均非常重要。

遗传重组率估值的精确度与数据囊括的遗传重组次数成正比，若数据囊括了越多的遗传重组事件，则遗传重组率估值越精确，反之亦然。无论是基于家系或单精子测序的研究，还是基于群体遗传数据的研究，这一原则均成立。基于群体遗传数据的分析，局限于已有的分析方法，较难运用以分析大样本。本研究中，科研人员扩展了前期开发的机器学习方法，运用新开发的FastEPRR 2.0分析了公开的UK10K共3,781个非相关个体（n=7,562个基因组）测序数据，基于Out-of-Africa群体历史模型，准确估计了遗传重组率，构建出精确的遗传重组图谱。总体上看，少数已知的遗传重组热点在UK10K遗传图谱中依然存在，但在UK10K遗传图谱中，遗传重组率估值波动较为平缓，遗传重组异质性较低（如图）。为了剖析样本大小对估值的影响，研究从UK10K数据中随机选取2,000、400和200个基因组测序数据，分析结果显示随着样本量的降低，遗传重组率的估值波动加大。上述结论不依赖分析时所用的群体历史模型，研究人员在使用群体数量恒定模型中也观察到同样的现象。该研究提供了精确的人类遗传重组图谱，并发现遗传重组在基因组上的分布或比预期的更均匀。

理论群体遗传学领域拥有完善的数学基础，与机器学习中的黑盒子概念相反。2008年初，研究人员预见有监督的机器学习对群体遗传学的促进作用，因而在2011年与合作者首次将有监督的机器学习引入了群体遗传学（Genetics），并在2013年（Genetics）、2016年（G3）持续发展这一新范式。在某些方面，虽然有监督的机器学习比极大似然法、贝叶斯等方法更好，但是这一新范式能否为进化生物学领域带来新发现，依然未知。结果表明，新范式带来了新发现，也回应了领域中某些质疑意见。

研究工作得到国家自然科学基金、中科院战略性先导科技专项、国家重点研发计划和营养与健康所的支持。（来源：中国科学院上海营养与健康研究所）

相关论文信息：<https://doi.org/10.1007/s00439-021-02415-8>

作者：李海鹏等 来源：《人类遗传学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发