
动物所建立灵长类特异新基因数据库并系统预测新基因功能

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/4321.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！

动物所建立灵长类特异新基因数据库并系统预测新基因功能。灵长类特异乃至人特异新基因起源是推动人类表型演化的重要推动力，但只有少量新基因功能已知。近年来基因编辑技术和类器官技术的发展使得系统揭示新基因在人之所以为人这一演化过程中的贡献成为可能。然而，由于基因年龄推断方法的差别以及较低的新基因注释质量(很难区分新蛋白编码基因和假基因)，文献中已经发表的灵长类特异新基因(PSG)数据集相互之间差别极大。一个相对可靠的PSG数据集的缺乏阻碍了深入的功能研究。

中国科学院动物研究所张勇课题组通过整合进化基因组和功能基因组数据开发了人类新基因数据库(GenTree, <http://gentree.ioz.ac.cn>)。GenTree可辅助用户分析基因何时起源、如何起源以及可能的功能(图1)。

基于GenTree数据的整合分析发现：1)以基因组共线性为基础的基因年龄推断方法与其它方法相比对近期起源的基因有较高的准确率;2)蛋白水平的选择压力推断和蛋白基因组方法(proteogenomics)只能注释部分灵长类特异基因的蛋白编码能力。该研究考虑了这些方法各自的特点之后鉴定了846个PSG(图2)。其中，254个基因有不同程度的蛋白支持证，该集合也包含了41个错误注释的假基因(如MYH16)。

基于GenTree中所整合的转录组数据，该研究通过共表达分析推测了846个PSG的功能。与此前的认识相符，PSG经常呈现睾丸特异或睾丸偏好的表达;但PSG也呈现骨髓或胎盘偏好的表达谱。有意思的是，虽然PSG在成年脑中的表达量较低，但在孕中期胎脑表达上调(图3)。富集于各共转录网络并占据网络较核心位置的PSG可能推动了精子发生、免疫反应、母胎互作以及胎脑发育等快速演化的生命过程。

总体来看，该工作提供了一个专门的人类新基因数据库，产生了相对高质量的PSG列表并推测了这些基因的功能。结果中所展示的年龄推断方法、基因注释方法的特点以及新基因器官特异的功能偏好性对研究其它物种种系特异基因的工作具普遍参考意义。

该研究于3月12日在线发表于Genome Research(doi:10.1101/gr.238733.118)。博士研究生邵毅、陈春燕为共同第一作者，张勇为通讯作者。龙漫远、何斌、沈浩、朱政霖、余大奇、降帅、赵石磊、高志强、付岩、陈希、高歌、陈华等合作者在文章写作、网页开发、数据分析方面提供了大力支持。该研究得到中科院先导专项、国家自然科学基金委员会和科技部的资助;计算需求得到中科院北京基因组研究所基因组测序与数据分析中心和动物所高性能平台的支持。

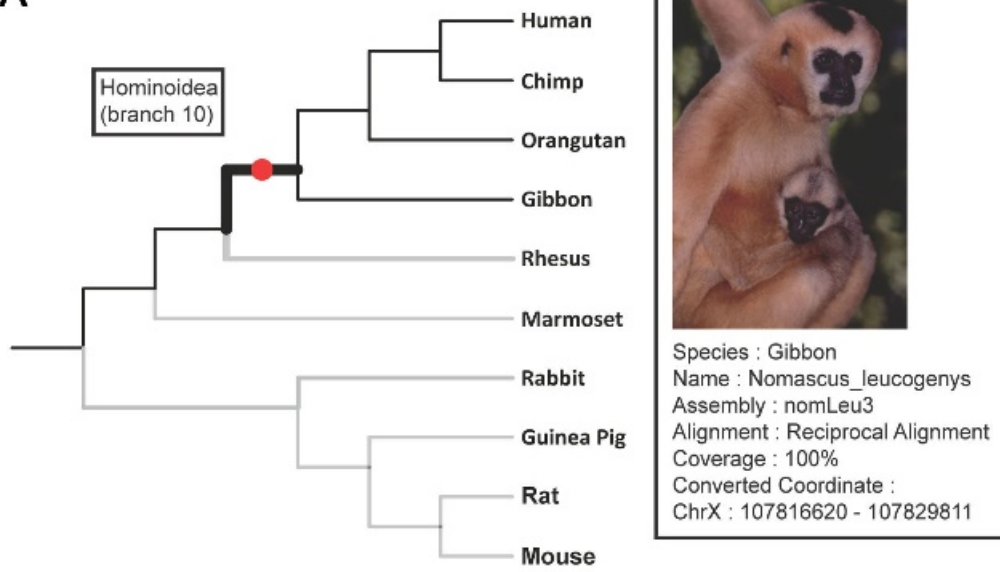
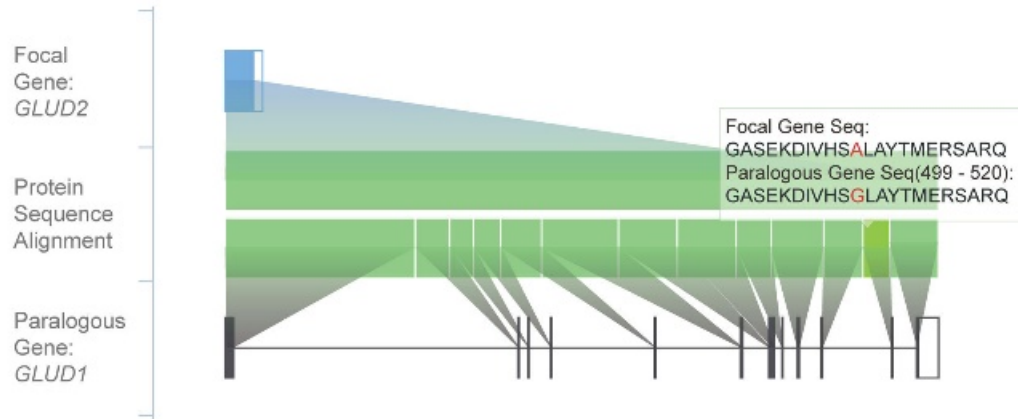
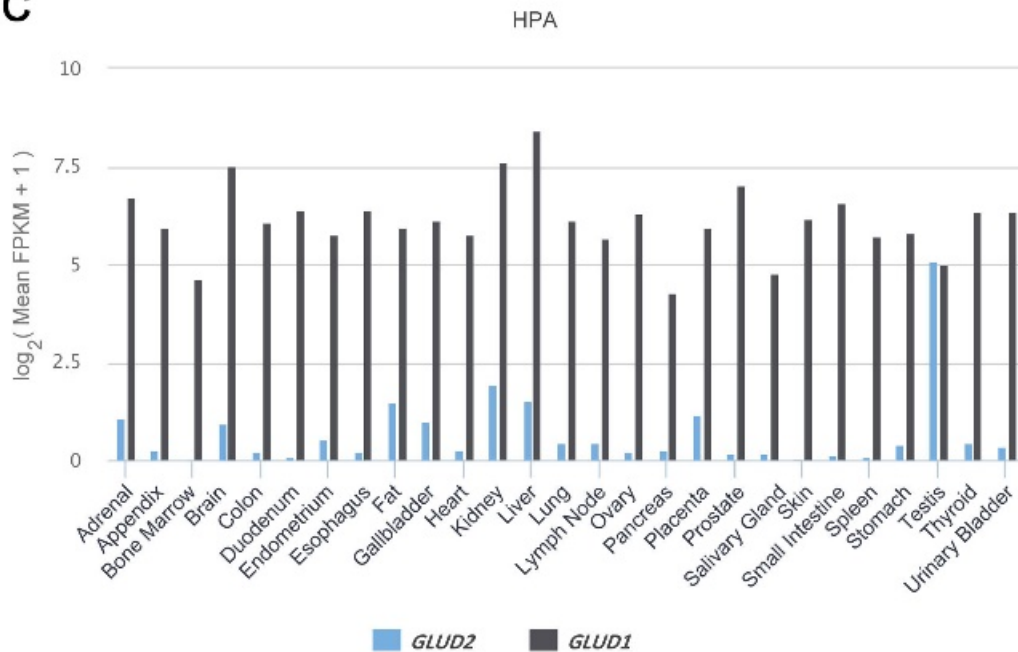
A**B****C**

图1：GenTree网页界面示例。A)谷氨酸脱氢酶2 (GLUD2)在人和长臂猿的共同祖先起源;B)GLUD2只编码一个外显子，而其最相似的同源基因GLUD1编码多外显子，暗示着GLUD2起源于GLUD1的mRNA水平的重复事件(逆转录);C)GLUD1广谱高表达，GLUD2除了睾丸之外，其它器官转录水平较低，暗示着后者功能可能比较局限。

图2：灵长类特异基因列表。灰色箭头说明蛋白编码能力可信度的下降，例如85个Ensembl注释的蛋白编码基因不仅有蛋白表达的证据(UniProt)也有功能研究(GO)，而90个PSG只有蛋白表达、质谱和选择压力中的一种证据。值得一提的是846个灵长类特异基因包含192个人类特异的新基因。

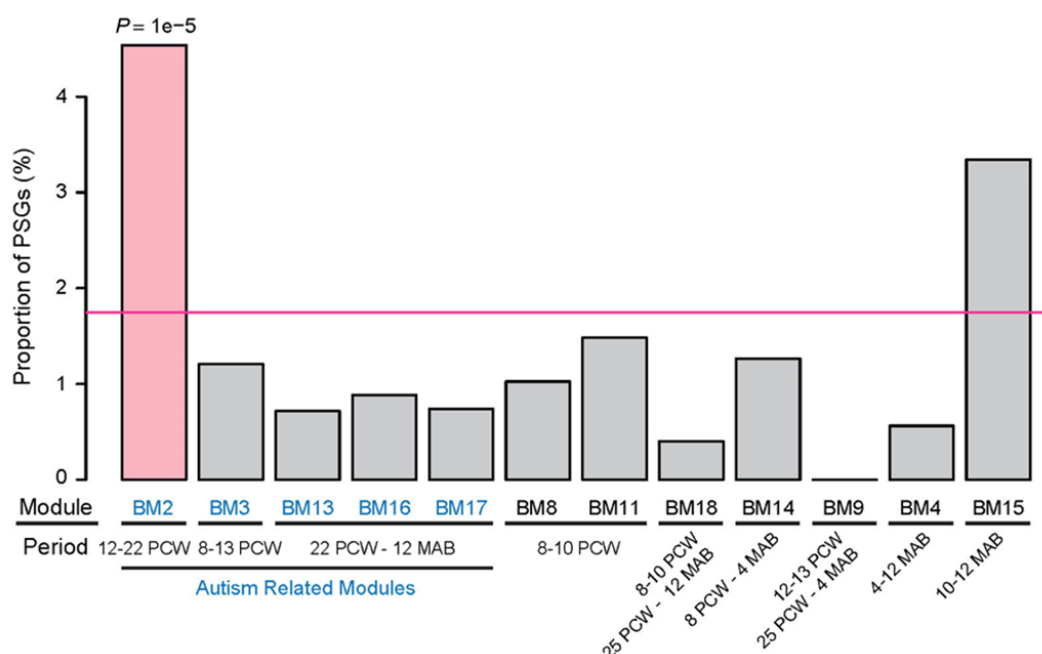


图3：不同脑发育阶段共转录模块(横轴)编码灵长类特异基因的比例(纵轴)。PCW和MAB分别代表孕周和出生后月份。共转录模块2(BM2)与基因组背景(粉红色线标记)相比统计上显著富集灵长类特异基因。该模块与其它四模块(蓝色标记)都富集自闭症易感基因，暗示它们的重要性。

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](http://www.iikx.com)转发