

探索发现 · 交大智慧

上海交大医工交叉研究取得可喜成果 ——成功构建中国汉族人泛基因组分析流程

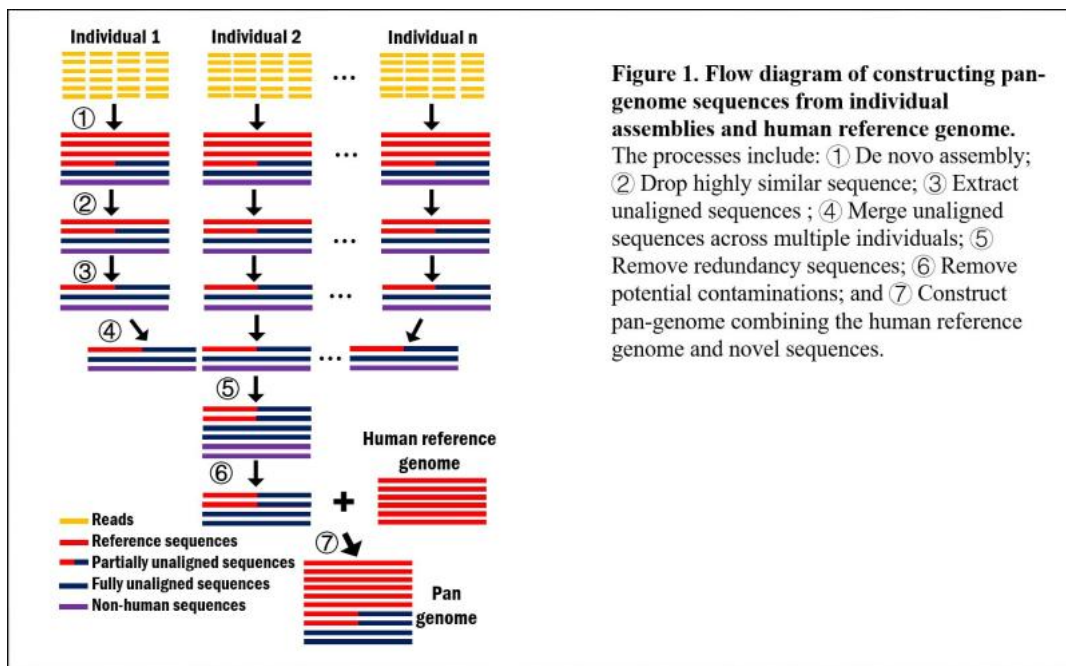
2019年08月01日 责任编辑: 刘环 王冰煜



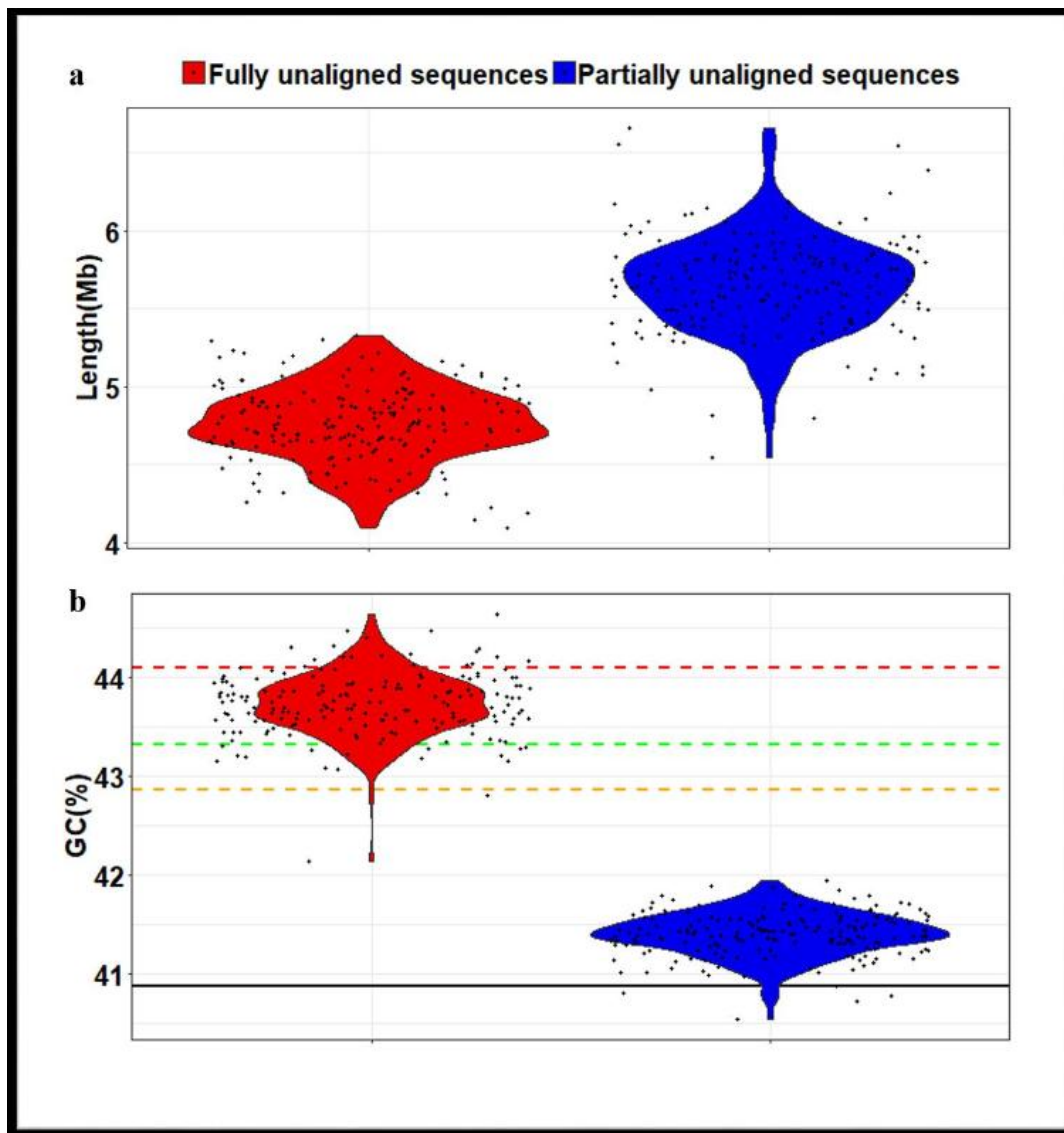
近日,来自上海交通大学的两个跨学科、跨专业科研团队,通过数年的协作科技攻关,开发了一套针对大规模人类基因组深度测序数据的全新分析系统——人类泛基因组分析系统(Human Pan-genome Analysis system,简称HUPAN)。该方法学的建立为解析人类基因组中尚未被发现的“暗物质”提供了重要研究工具。该项研究成果于2019年7月31号在线发表于国际基因组学研究权威刊物《Genome Biology》(影响因子14.028)。该研究工作是由上海交大生命科学技术学院韦朝春教授课题组和医学院附属瑞金医院于颖彦教授课题组联合完成。韦朝春教授与于颖彦教授为该文章共同通讯作者,博士生段忠取为该文章第一作者。

众所周知,人类基因组计划宣告完成后产生了一个基于有限个体的人类参考基因组序列,这也是之后众多分子生物学实验研究的参照基础。但随着对人类基因组测序研究的广泛开展、测序个体数量的不断增加,科学家们发现,现有的人类基因组参考序列尚不够完整,特别是在一些特定的人群或个体基因组中被测序到现有人类基因组参考序列中缺失的片段,也就是说,人类基因组的序列其实比已知的基因组参考序列要复杂,尚有很多未知序列(或者说暗物质)有待于科学家们通过不断的深化研究加以发现。

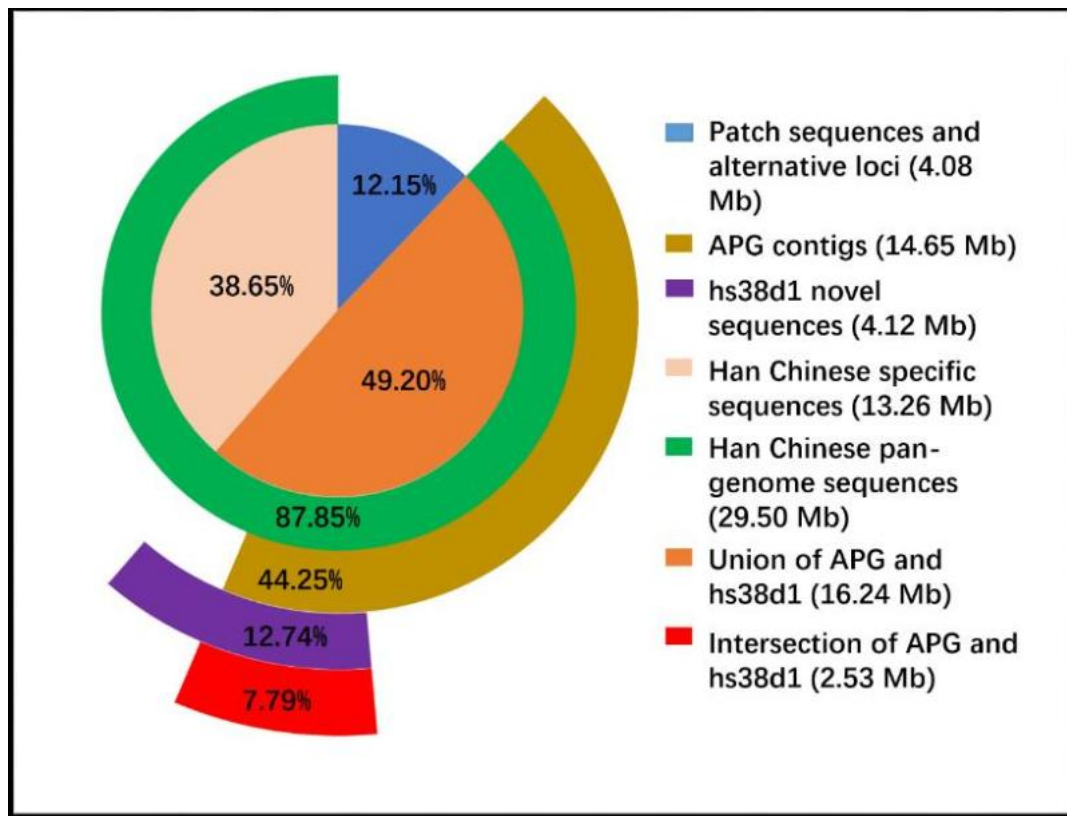
泛基因组(Pan-genome)是指某个群体中所有个体基因组的总和。随着测序技术的进展,针对人类某个群体的多个个体基因组的测序数据不断积累增加,给泛基因组研究提供了前所未有的契机。然而,人类全基因组测序数据量庞大,现有针对如此大型基因组数据进行拼接研究的方法学有限,如果研究方法不加以创新,在分析过程中容易引入较多的拼接错误,且分析速度缓慢。为此,该联合研究团队进行了分析方法的创新,他们首先对原有真核生物泛基因组分析流程进行了改进,包括引入节约内存的拼接方法,可直接对每个个体的所有测序数据进行拼接以降低拼接错误,优化了泛基因组分析步骤,明显提高了海量测序数据的分析速度和准确率。



新型分析方法对185个中国汉族人的全基因组深度测序分析，同时对开放数据库内90个中国汉族人全基因组深度测序数据测试显示，在中国汉族人全基因组测序数据中至少存在29.5Mb不同于人类参考基因组的新序列，暨人类基因组参考序列中漏掉的序列。通过新基因预测分析，发现188个新基因，且新序列中约40%仅见于中国汉族人群。



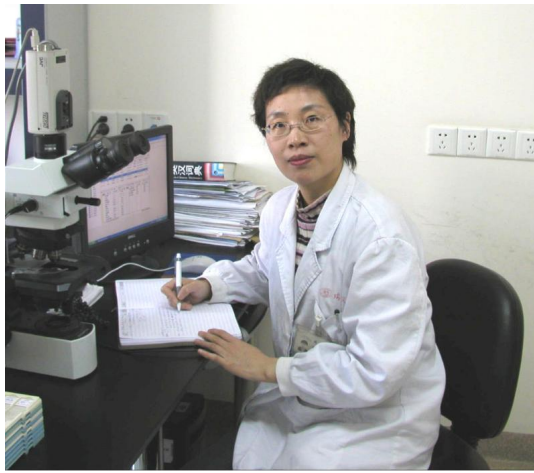
该新型方法的创立不仅仅为深入研究人类进化、人类迁徙规律、种族基因组差异以及新基因与人类疾病相关性提供了重要工具，还为其它具有较大基因组的高等动物泛基因组研究提供了重要分析工具。



该项研究是在上海交通大学医工交叉重点项目、国家科技部及卫健委重点研发计划、国家自然科学基金委以及上海市科委重点项目支持下完成。项目实施过程中还得到上海市转化医学协同创新中心和上海交通大学高性能计算中心的大力支持。



韦朝春，上海交通大学生命科学技术学院生物信息学与生物统计学系教授 / 博士生导师。先后于北京大学和美国华盛顿大学（圣路易斯）获得数学学士、信息处理硕士和计算机科学博士学位。主要研究方向为基因组学和进化基因组学。具体研究内容包括基因组中的功能因子的识别及其进化分析、真核生物泛基因组学、肿瘤基因组学和宏基因组学等。



于颖彦，上海交通大学医学院附属瑞金医院教授/博士生导师，上海消化外科研究所副所长，中国抗癌协会胃癌专业委员会委员，中国医药生物技术协会生物样本库分会及慢病管理分会常委。上海市浦江人才与上海市优秀学术带头人。从事消化病理、肿瘤分子分型、生物标志物和转化医学研究。承担国家重点研发计划精准医学专项和慢病专项，国家自然科学基金、上海市科委重点项目及上海交大医工交叉重点项目等。

论 文 链 接 : <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1751-y>
(<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1751-y>)

作者： 韦朝春
供稿单位： 生命科学技术学院

沪ICP备05052060 (<http://www.beian.miit.gov.cn/>) 沪举报中心 版权所有© 上海交通大学 新闻网编辑部维护

地址：上海市东川路800号 邮编：200240 查号：86-21-54740000