

What genetic changes made us uniquely human?

人类独特表型遗传基础的研究进展

罗鑫^{1,2}, 张栋秦^{1,2}, 宿兵^{1*}

1. 中国科学院昆明动物研究所, 遗传资源与进化国家重点实验室, 昆明 650223;

2. 中国科学院大学昆明生命科学学院, 昆明 650223

* 联系人, E-mail: sub@mail.kiz.ac.cn

2016-12-13 收稿, 2017-03-01 修回, 2017-03-02 接受, 2017-03-28 网络版发表

摘要 大约500万年前人类(*Homo sapiens*)和黑猩猩(*Pan troglodytes*)分歧之后, 人类在进化过程中逐渐形成了很多独特的表型. 什么样的遗传变异造就了如此独特的人类呢? 近10年来, 通过比较人类和其他非人灵长类的基因组序列, 研究者已经发现了许多基因组区域在人类中受到自然选择而发生了人类特异的遗传改变, 包括基因编码区的改变、基因表达调控的改变、基因的丢失和基因的重复等. 这些不同水平的遗传改变在进化过程中通过自然选择的作用在人群中固定下来, 对人类独特表型的产生有着重要贡献. 本文将从这4个遗传层面介绍人类特异表型与遗传变异的关系. 另外, 文中也介绍了利用基因编辑技术建立人源化的诱导多能干细胞及转基因动物模型研究人类特异突变和表型的关系, 以及新的测序方法在筛选群体中天然突变个体中的应用.

关键词 人类独特表型, 编码基因, 基因表达调控, 基因丢失, 基因重复

“裸猿”是对人类(*Homo sapiens*)的一种独特称呼. 这一称呼的由来是因为人类与其近亲——非人灵长类相比具有裸露的皮肤这一独特的表型. 除此之外, 人类还具有直立行走、巨大的脑容量、复杂的语言、超长的童年期以及文化和社会性等特点. 那么是什么导致了人类特异表型的产生呢? 长期以来研究人员试图从古人类学、解剖学、心理学、行为学和社会学等各个方面来研究和解释这一人类自身的根本问题. 从遗传学的角度来讲, 是什么样的遗传变异使人类进化出如此独特的表型呢? 人类特异的表型性状的产生通常涉及多个环节, 是一个非常复杂的过程. 例如, 人类大脑的进化涉及神经前体细胞的分化和迁移、轴突的导向、突触的数量, 以及新的细胞类型的产生等, 每个环节的改变都伴随着很多遗传变异的积累. 通过比较人类和其他非人灵长类基因组数据, 研究人员已经发现了众多人类特异的遗传变异^[1~8]. 归结起来, 大致可以分为4类: 基因编码水平

的改变、基因表达调控水平的改变、基因丢失和基因重复(gene duplication). 这4个遗传控制水平的改变都有可能引起人类特异表型的产生. 另外, 随着基因组编辑技术的进步, 研究人员可以利用如CRISPR/Cas9这样简单高效的技术将人类特异的遗传变异引入到实验动物(如小鼠(*Mus musculus*)和猕猴(*Macaca mulatta*))细胞系的基因组中(如诱导多功能干细胞(induced pluripotent stem cells, iPSCs)), 或者建立人源化的转基因小鼠和猴子, 进而对这些变异进行系统全面的研究. 另外, 研究人员也可以利用新近发展的方法(如分子倒位探针(molecular inversion probe, MIP))结合二代测序在人类群体中进行大样本量的突变筛选, 并对筛选到的自然发生突变个体进行后续的表型研究.

1 基因编码水平的改变

在基因组研究的早期, 人们认为只有编码蛋白

引用格式: 罗鑫, 张栋秦, 宿兵. 人类独特表型遗传基础的研究进展. 科学通报, 2017, 62: 1324–1332

Luo X, Zhang D Q, Su B. Genetic bases of human unique traits (in Chinese). Chin Sci Bull, 2017, 62: 1324–1332, doi: 10.1360/N972016-01063

的基因序列才会起到功能作用,而这些相应蛋白功能的变化才会引起表型的变化,即蛋白质进化假说.这个假说认为基因编码区序列的改变会导致编码蛋白质的改变,产生不同的生物学效应,进而受到自然选择的作用被固定下来.通过比对人类和非人灵长类基因组序列,研究发现了许多基因确实在人类中受到达尔文正选择的作用^[2].这些基因主要参与人类的免疫防御、细胞信号传导、氨基酸代谢、嗅觉、大脑发育和语言形成等各个方面^[9~11].

在人类进化的过程中,脑容量的增大是最显著的特征之一.通过遗传分析,研究人员发现了一系列脑容量相关基因.其中最受关注的是*MCPH1*和*ASPM*基因,这两个基因在人类中的功能丧失突变都会引发原发性小头症(primary microcephaly).序列比对结果显示,*MCPH1*和*ASPM*基因都受到正选择作用^[12,13].例如,在人类和黑猩猩(*Pan troglodytes*)分歧之后,*MCPH1*基因在人类中产生了9个人类特异的氨基酸突变,这些突变可能在人类大脑进化过程中发挥重要作用^[14].

另一个人类特异的蛋白质序列变异的例子是被称为“语言基因”的*FOXP2*基因.*FOXP2*基因在人类中的突变会导致遗传性的语言缺陷.通过比对不同物种间的*FOXP2*序列,研究者发现该基因的序列总体上非常保守,但在人类和黑猩猩分歧之后,人类中却发生了两个特异的突变^[15].有趣的是,在两个古人类物种尼安德特人和丹尼索瓦人也存在这两个突变,表明这两个位点可能在人类进化的较早时期就被固定下来了^[16].研究者将小鼠对应的氨基酸位点人源化,发现携带这两个突变位点的小鼠神经突触的数量和对照组相比显著增多^[17].

通过比对现代人和其他古人类以及非人灵长类的基因组,研究人员已经报道了87个基因在现代人中存在特异的氨基酸序列变异^[18].这些基因大部分都在胚胎阶段大脑的神经发生区域-侧脑室中表达.其中的3个基因(*CASC5*, *KIF18A*和*SPAG5*)不仅在侧脑室区域表达,而且与大脑皮层发育中神经前体细胞有丝分裂时期纺锤体的分裂方向相关.因为在有丝分裂时期,纺锤体的分裂方向决定神经前体细胞分裂后继续产生具有增殖能力的神经前体细胞还是产生分化的神经元,对大脑容量起到决定性的作用.因此,这些基因可能在人类大脑进化过程中具有重要作用^[19].

2 基因表达调控水平的改变

1969年, Britten和Davidson^[20,21]首先提出物种间表型的差异可能主要是由于调控元件的改变所形成的.1975年, King和Wilson^[22]将这一假设延伸至人类进化领域,并提出人类和他的近亲——非人灵长类之间的表型差异可能是主要由基因的表达调控差异所决定的.研究人员通过搜索在非人灵长类中高度保守,但是在人类中却由序列改变的非编码区域来寻找人类特异的调控序列. Pollard等人^[23,24]在2006年通过比对不同灵长类的基因组序列,定义了一批在人类中加速进化的区域(human accelerated regions, HARs),这些HARs大多数都处在基因的调控区域. Prabhakar等人^[25,26]随后又针对非编码序列进行了类似的研究并找到了一批在其他物种中高度保守,但在人类中快速进化的非编码序列(human accelerated conserved noncoding sequences, HACNSs).随后,一些研究又发现一些新的在人类中快速进化的非编码序列^[27,28].另一方面,研究发现HAR区域的一些人类特异突变可能是由于AT到GC的基因偏好性转换(biased gene conversion, BGC)造成的,这种AT到GC的频率升高与基因受到正选择的作用类似,因此很难区分开来^[23].通过开发新的方法, Kostka等人^[29]已经确认大部分的HARs区域突变是自然选择的结果,而有些是自然选择和基因偏好性突变共同作用的结果.

人类快速进化的非编码序列如何影响人类特异的表型进化呢?多个研究组通过功能实验对以上发现的HARs和HACNSs进行了研究. Prabhakar等人^[26]通过小鼠实验证明*HACNS1*(即Pollard定义的*HAR2*)作为一个增强子在下肢的发育过程中具有重要作用.结果显示,与黑猩猩以及猕猴的同源序列相比,人类的增强子活性更强.这也暗示*HACNS1*可能在人类直立行走的起源中扮演重要角色.同样的,2015年Boyd等人^[30]通过小鼠的功能试验发现*HAR5*能够增强*Fzd8*的表达.与黑猩猩的同源序列相比,人类特异的*HAR5*能够加快神经前体细胞的对称分裂,最终产生更多的神经元和更大的大脑容量.这项研究表明*HAR5*可能在人类进化过程脑容量的增大中发挥作用.

近年来,随着二代测序成本的降低以及各种实验技术与二代测序的结合,发展出了许多新的测序技术,如ChIP-Seq和ATAC-Seq等.通过分析组蛋白修饰和DNase I超敏感位点测序有助于找到基因转

录的活性区域,寻找可能的增强子区域.例如,Gitelman等人^[31]在近期研究中用一种DNase I超敏位点测序的方法研究人类中快速进化的调控区域在1800万个超敏位点中检测到了524个人类特异性快速进化的区域.这些序列可能是转录中的活性区域,可能和人类特异性的调控功能有关.

近年来,研究者通过比较人类和非人灵长类的多种组学数据发现表观遗传修饰、小RNA(microRNA)和长非编码RNA(lncRNA)等在人类进化过程中可能也具有重要作用.甲基化是一种重要的DNA修饰方式,也是一种重要的表观遗传调控机制,影响基因的转录调控.DNA甲基化可以改变DNA的构象,从而影响DNA与转录因子等蛋白质的结合,从而引起基因失活.人和非人灵长类在许多基因的甲基化水平上存在着显著差异,导致基因表达显著差异.人类特异性的甲基化也对人类独特的表型产生了很大影响.最近,研究者对人类、黑猩猩和猕猴3个物种的大脑前额叶进行了大脑全基因组甲基化测序,发现了85个人类特异的甲基化变化区域(differentially methylated region, DMR).这些区域可能通过影响一些神经特异转录因子与之结合参与神经发育调控过程.该研究提示在进化中表观遗传变化对人类大脑起源的重要作用^[32].

MicroRNA的海绵调控机制是近年来发现的一种全新的基因表达调控方式,它是指lncRNA或者假基因可以竞争性地结合相关蛋白编码基因对应的microRNA从而削弱microRNA对靶基因蛋白合成的抑制,提高蛋白的表达量.近期研究表明,一种灵长类特异的lncRNA (lncND)与miR-143-3p存在海绵调控机制,可能对灵长类大脑皮层的扩张具有重要作用^[33].虽然目前还没有发现具有海绵调控效应的人类特异的lncRNA,但是作为一种新的基因表达精细调控机制,lncRNA的海绵效应可能对人类特异表型也有重要贡献.

通常研究基因的表达调控关注的都是一维水平.事实上,一个基因的两个表达调控元件也许在物理距离上相隔甚远,但在三维空间中却靠得很近.如何去研究基因的三维远程调控呢?随着技术的进步,研究人员现在可以利用以染色体构象捕获技术(chromosome conformation capture, 3C)为基础的多种技术手段检测染色体的三维结构,构建立体的基因调控网络^[34,35].目前,使用最广泛的是高通量染色体

构象捕获(high-throughput chromosome conformation capture, Hi-C)技术.研究人员利用甲醛固定活细胞,然后用限制性内切酶(如Hind III)酶切使其片段化,在DNA末端加上生物素接头,而后建库并进行高通量测序^[36].由此,研究人员可以对细胞进行染色体三维结构水平的测定.不久前,加州大学洛杉矶分校的Geschwind实验室^[37]利用Hi-C技术对人类胚胎神经发育高峰期的神经元和神经前体细胞进行了染色体三维结构水平的基因调控网络分析,发现了人类特有的增强子在人类大脑发育中的重要作用.未来,可以对其他非人灵长类的对应发育阶段的脑组织进行分层的Hi-C测序,检测物种间远程调控的差异,进而揭示这种染色体三维水平的基因调控在人类进化中的作用.

3 基因丢失

在进化过程中一些基因发生基因片段的丢失,或者通过添加或替换碱基导致基因不翻译或提前终止翻译,使得基因失去功能或假基因化(gene loss).1999年Olson博士^[38]提出less-is-more假说,他认为基因丢失在人类进化过程中发挥重要作用.该假说提出人类通过一些关键基因功能的丢失,使人类失去很多与猿类相似的特征(如体表的毛发和宽大的下颌骨等)^[39].

研究发现,在人类进化过程中一些基因的丢失导致了人类特异表型的产生.例如,与其他猿类相比人类的咀嚼肌退化严重.研究发现在大约240万年前早期人类的MYH16基因就已经发生假基因化,导致基因丧失功能^[40].进一步的研究发现,MYH16基因丢失的时间在人类进化的更早期(约500万年前),即人类和黑猩猩分歧之后就产生了^[41].MYH16基因的丢失使得人类进化过程中咀嚼肌退化,下颌骨缩小,进一步导致人类颅面的改变,为脑容量的增大提供了空间.另一方面,也有学者提出MYH16的丢失在其他物种也有发生,并且咀嚼肌的退化并不是人类特异的表型,在一些有蹄类,如兔子(Leporidae)和袋鼠(Macropus)中也有类似的表型.因此,人类中MYH16基因的丢失或许并不是脑容量增大的直接原因^[42].研究发现,人类的角蛋白基因KRTHAP1基因大约在24万年前失活,使得人类体表的毛发大大减少,有利于在快速奔跑过程中迅速降温^[43].

CMAH基因编码一种细胞表面分子——唾液酸

(Neu5Gc), 它参与细胞与细胞以及细胞与病原体之间的相互作用。研究发现, 大约在280万年之前由于Alu序列的插入导致了人类*CMAH*基因的假基因化, 造成人类细胞表面大量分子相互作用的重新配置, 形成了人类特异的免疫系统^[44]。后续的研究发现在尼安德特人中这一基因也像现代人一样失去了功能^[45]。

另外一个重要的例子是*APOC1*基因在人类中的丢失。*APOC1*是一种载脂蛋白, 主要参与脂蛋白的代谢。在大猿中有两个*APOC1*拷贝, 分别编码正负两种形式的载脂蛋白。但是, 编码负价载脂蛋白的*APOC1*基因在人类进化过程中发生突变, 产生了一个提前终止密码子, 使得在人类中只有正价的载脂蛋白发挥作用^[46]。*APOC1*基因的丢失在人群中存在多态性, 研究表明这种多态性与阿尔茨海默症和动脉粥样硬化等疾病相关^[47,48]。

4 基因重复

基因重复在人类进化过程中也具有重要作用。通过对比人类和非人灵长类的基因组, 研究人员发现了大量在人类中特异的基因拷贝数变异^[49]。归结起来, 主要分为3类: (i) 最常发生的一类, 即基因重复之后, 其中一个拷贝退化, 失去功能。因而对表型不会带来影响。(ii) 通过拷贝数的增加带来基因剂量效应(gene dosage effect)。其中, 最为显著的一个例子是*DUF1220*。*DUF1220*是*NBPF*基因家族具有的一个结构域(domain)。研究表明, 在灵长类进化过程中, *DUF1220*结构域的拷贝数和脑容量大小呈现正相关^[50]。古人类颅骨化石研究表明, 尼安德特人的脑容量要比现代人大一些^[51]。通过比对尼安德特人基因组, 研究人员发现这一古老人种中具有约350个*DUF1220*拷贝, 比现代人具有的272个*DUF1220*拷贝数量还多, 进一步证实了*DUF1220*拷贝数和脑容量大小正相关的假说^[52]。(iii) 通过复制祖先型基因的部分或者全部, 并产生具有全新功能的基因。例如, *SRGAP2*基因, 该基因在神经元的迁移、分化以及突触的发育等方面具有重要作用。*SRGAP2*在人类进化过程发生过3次重复事件。在340万年前*SRGAP2*发生第一次重复形成了*SRGAP2B*。随后在240万年前*SRGAP2B*又发生重复形成了*SRGAP2C*。最近一次重复发生在100万年之前, *SRGAP2B*通过部分片段的重复形成截短型的*SRGAP2D*^[53]。研究人员将人类特异的*SRGAP2C*拷贝转入小鼠中, 发现小鼠呈现类似人

类幼态持续(neoteny)的表型。转入*SRGAP2C*拷贝延缓了小鼠的神经元的成熟, 并且使神经棘的数量显著增加^[54]。最近的研究表明, 即使*SRGAP2C*只有部分*SRGAP2A*的结构序列, 它却能够全面抑制*SRGAP2A*的功能, 共同调节锥体神经元兴奋和抑制性突触的发育^[55]。另一个类似的例子是*ARHGAP11B*基因。*ARHGAP11B*基因是在大约500万年前人类和黑猩猩分歧之后通过部分重复祖先拷贝*ARHGAP11A*基因形成的。*ARHGAP11B*缺少*ARHGAP11A*蛋白末端的756个氨基酸, 取而代之的是C端有47个氨基酸组成的新功能域。因为*ARHGAP11B*只有267个氨基酸组成, 不具备RhoGAP的功能, 因此科研人员推测*ARHGAP11B*可能进化出了新的功能。转录组的分析表明, *ARHGAP11B*在人类大脑SVZ区域的神经前体细胞中高度表达。小鼠实验证明, *ARHGAP11B*能够增强神经前体细胞的对称分裂, 产生更多的神经前体细胞, 使SVZ的厚度增加30%; 转入*ARHGAP11B*基因的小鼠皮层厚度增加, 并且大脑出现一定程度的沟回。这些发现说明*ARHGAP11B*基因在人类脑容量和结构的进化过程中发挥非常重要的作用^[56]。

目前为止, 研究者发现的大多数基因重复都是相对古老的, 即大多发生在人类和黑猩猩分离之后的较早时期。随着测序技术的进步, 人们现在可以对现代人基因组复杂区域的拷贝数变异进行研究。2016年, Eichler实验室^[57]通过对人类染色体16p11.2区域进行遗传分析, 找到了现代人特异的*BOLA2*基因的重复, 这一重复大约发生在28.2万年之前。通过一系列的细胞实验证明, *BOLA2*基因在人类的胚胎干细胞和神经干细胞都呈现高表达量, 而在神经元细胞中却低表达, 这说明*BOLA2*基因在人类大脑神经发生阶段可能发挥作用。

5 人类特异遗传变异的功能验证

人类特异遗传变异的生物学功能需要一系列的功能实验去阐明。随着基因编辑技术的快速发展, 对基因组的精确编辑变得越来越简单高效。研究人员现在可以通过iPSCs和建立人源化转基因动物(小鼠和猕猴等)的途径验证这些变异带来的功能后果。

自2006年日本科学家山中伸弥在世界上首次利用成纤维细胞诱导出多能干细胞以来, 近10年的发展使科研人员可以利用已分化的细胞诱导产生各种类型的细胞^[58], 如神经前体细胞和神经元等。研究

人员可以利用各种非人灵长类(尤其是大猿)的成纤维细胞诱导产生多能干细胞,然后再根据实验需求,使其定向分化为特定类型的细胞.2013年, Marchetto 等人^[59]将2只黑猩猩和2只倭黑猩猩(*Pan paniscus*)的皮肤成纤维细胞成功诱导为多能干细胞.2015年, Gallego实验室^[60]分别建立了7个黑猩猩和7个人类诱导多能干细胞系.接下来, Prescott等人^[61]利用人类和黑猩猩的iPSCs培养出神经脊细胞.剑桥大学的 Otani等人^[62]利用人类、黑猩猩和猕猴的iPSCs在2D和3D的体系中培养出神经前体细胞,并且揭示了这3个物种间神经前体细胞增殖分化的差异及其对脑容量大小的影响.

利用iPSCs可以克服灵长类动物采样的难题,尤其是脑组织.然而,在细胞水平无法很好地呈现人类特异突变所引起的表型效应.采用CRISPR/Cas9等基因编辑技术可以将人类特异的突变引入到小鼠或者猕猴的基因组中,建立人源化的动物模型,从而在活体水平分析人类特异遗传变异的表型和功能效应^[63,64].

除了通过基因编辑的手段验证人类特异突变带来的功能后果,研究人员也可以利用群体遗传学的手段在人类群体中筛选携带突变的个体,并通过影像学、行为学和细胞实验等各个水平验证这些基因突变带来的表型效应.目前,研究人员已发展了多种较为成熟的技术手段用来大规模筛选人类乃至其他动物群体中天热发生的突变.例如,最近发展的MIP的方法利用预先设计好的探针序列捕获感兴趣的基因组区域,然后进行扩增和高通量测序.这个方法可以极大降低大规模群体筛选的成本,比较高效地在人类群体中筛选偶发的基因功能缺失突变^[65].该方法除了能获得单核苷酸突变信息外,还能得到小片段的缺失或重复的信息,并同样适用于在非人灵长类群体中进行大规模突变筛选.

6 展望

人类与非人灵长类在遗传上的差异导致了人类特异表型的产生,造就了人类这个独特的物种.从目前的研究进展来看,人类特异的遗传变异可以在不同的水平发生,且不是相互排斥的,很多遗传变异也不能简单地归为一类.例如, *FOXP2*基因发生的两个人类特异的氨基酸变异也会影响下游基因表达量的变化,因为*FOXP2*本身就是一个转录因子.研究发现, *FOXP2*非编码区中的一个非常保守的位点在现代

人中发生了突变.由于这个位点处于*POU3F2*结合处,因而会导致下游基因表达量的变化^[66].

与组装比较完整的人类基因组测序数据相比,非人灵长类(尤其是大猿)的基因组序列的组装和功能注释还有很大的差距.目前,大多数基因组测序数据都是基于二代测序技术,因为二代测序有很多优点,如测序成本低和技术成熟等.但目前二代测序技术最大读取片段比较短.对于全基因组组装来说,基因组中重复片段对组装有很大影响.目前的二代测序技术并不能解决这个问题.三代测序作为目前最新的测序技术在读长的问题上有了很大的改善.随着技术的成熟,三代测序的错误率也在不断下降.目前二代测序和三代测序结合的方法使基因组序列的可靠性进一步提高,这是寻找更多人类特异性的突变,以及调控序列的基础.随着三代测序技术的成熟以及基因组序列组装的改善,对灵长类基因组序列数据的挖掘会更加深入.

随着染色体构象捕获技术的发展,研究人员可以对各个物种的各种组织和细胞进行Hi-C测序,进而找出物种间在染色体三维结构水平基因调控的差异.然而由于Hi-C测序需要活细胞,这一技术在人类和非人灵长类(尤其是大猿)中的应用仍然受到限制.如果想获得高精度的数据需要千万级的细胞数量,这样的取材对于黑猩猩等大猿物种来说是比较困难的.但是Hi-C测序在猴子(如猕猴和食蟹猴(*Macaca fascicularis*))上是是可以实现的.可以获取准确胎龄的猕猴胚胎脑组织进行Hi-C测序,进而和已有的人类胎脑Hi-C数据进行比较,找出人类特异的三维水平的调控模式及其分子机制.

迄今研究人员已经发现了许多人类特异的遗传变异,然而对这些变异导致的功能后果仍然所知甚少.未来利用基因编辑技术将人类特异的突变引入到iPSCs和活体动物(如小鼠和猕猴等)中将有助于阐明这些变异的表型效应及其发挥作用的分子机制.利用非人灵长类构建基因编辑动物模型研究人类独特表型(如大脑)的遗传基础具有小鼠等传统的模式动物所不具备的优势,有望取得突破性的进展.另外,利用MIP等方法结合二代测序,可以在人群中进行大样本的突变筛选.同时这一技术也可以应用到猴子群体中,可以对圈养猴群进行大样本筛选,找出那些携带基因功能缺失突变的个体并进行后续表型和功能方面的研究.

参考文献

- 1 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860–921
- 2 Sequencing T C, Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 2005, 437: 69–87
- 3 Khaitovich P, Hellmann I, Enard W, et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, 2005, 309: 1850–1854
- 4 Gibbs R A, Rogers J, Katze M G, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, 2007, 316: 222–234
- 5 Locke D P, Hillier L D W, Warren W C, et al. Comparative and demographic analysis of orangutan genomes. *Nature*, 2011, 469: 529–533
- 6 Scally A, Dutheil J Y, Hillier L D W, et al. Insights into hominid evolution from the gorilla genome sequence. *Nature*, 2012, 483: 169–175
- 7 Prüfer K, Munch K, Hellmann I, et al. The bonobo genome compared with the chimpanzee and human genomes. *Nature*, 2012, 486: 527–531
- 8 Lindblad-Toh K, Garber M, Zuk O, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 2011, 478: 476–482
- 9 Cargill M, Clark A, Glanowski S, et al. Genome-wide comparative sequence analysis between human, mouse, and chimpanzee identifies human genes under positive selection. *Am J Hum Genet*, 2003, 73: 430
- 10 Nielsen R, Bustamante C, Clark A G, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*, 2005, 3: e170
- 11 Lai C S L, Fisher S E, Hurst J A, et al. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 2001, 413: 519–523
- 12 Evans P D, Anderson J R, Vallender E J, et al. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum Mol Genet*, 2004, 13: 1139–1145
- 13 Evans P D, Anderson J R, Vallender E J, et al. Adaptive evolution of *ASPM*, a major determinant of cerebral cortical size in humans. *Hum Mol Genet*, 2004, 13: 489–494
- 14 Shi L, Li M, Lin Q, et al. Functional divergence of the brain-size regulating gene *MCPH1* during primate evolution and the origin of humans. *BMC Biol*, 2013, 11: 1
- 15 Enard W, Przeworski M, Fisher S E, et al. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature*, 2002, 418: 869–872
- 16 Krause J, Lalueza-Fox C, Orlando L, et al. The derived *FOXP2* variant of modern humans was shared with Neanderthals. *Curr Biol*, 2007, 17: 1908–1912
- 17 Enard W, Gehre S, Hammerschmidt K, et al. A humanized version of *Foxp2* affects cortico-basal ganglia circuits in mice. *Cell*, 2009, 137: 961–971
- 18 Prüfer K, Racimo F, Patterson N, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 2014, 505: 43–49
- 19 Fietz S A, Huttner W B. Cortical progenitor expansion, self-renewal and neurogenesis—a polarized perspective. *Curr Opin Neurobiol*, 2011, 21: 23–35
- 20 Britten R J, Davidson E H. Gene regulation for higher cells: A theory. *Science*, 1969, 165: 349–357
- 21 Britten R J, Davidson E H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol*, 1971, 46: 111–138
- 22 King M C, Wilson A C. Evolution at two levels in humans and chimpanzees. *Science*, 1975, 188: 107–116
- 23 Pollard K S, Salama S R, King B, et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet*, 2006, 2: e168
- 24 Pollard K S, Salama S R, Lambert N, et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 2006, 443: 167–172
- 25 Prabhakar S, Noonan J P, Pääbo S, et al. Accelerated evolution of conserved noncoding sequences in humans. *Science*, 2006, 314: 786–786
- 26 Prabhakar S, Visel A, Akiyama J A, et al. Human-specific gain of function in a developmental enhancer. *Science*, 2008, 321: 1346–1350
- 27 Bird C P, Stranger B E, Liu M, et al. Fast-evolving noncoding sequences in the human genome. *Genome Biol*, 2007, 8: 1

- 28 Kim S Y, Pritchard J K. Adaptive evolution of conserved noncoding elements in mammals. *PLoS Genet*, 2007, 3: e147
- 29 Kostka D, Hubisz M J, Siepel A, et al. The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. *Mol Biol Evol*, 2012, 29: 1047–1057
- 30 Boyd J L, Skove S L, Rouanet J P, et al. Human-chimpanzee differences in a *FZD8* enhancer alter cell-cycle dynamics in the developing neocortex. *Curr Biol*, 2015, 25: 772–779
- 31 Gittelman R M, Hun E, Ay F, et al. Comprehensive identification and analysis of human accelerated regulatory DNA. *Genome Res*, 2015, 25: 1245–1255
- 32 Mendizabal I, Shi L, Keller T E, et al. Comparative methylome analyses identify epigenetic regulatory loci of human brain evolution. *Mol Biol Evol*, 2016, 33: 2947–2959
- 33 Rani N, Nowakowski T J, Zhou H, et al. A primate lncRNA mediates Notch signaling during neuronal development by sequestering miRNA. *Neuron*, 2016, 90: 1174–1188
- 34 Dekker J, Rippe K, Dekker M, et al. Capturing chromosome conformation. *Science*, 2002, 295: 1306–1311
- 35 Van Steensel B, Dekker J. Genomics tools for unraveling chromosome architecture. *Nat Biotech*, 2010, 28: 1089–1095
- 36 Lieberman-Aiden E, Van Berkum N L, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 2009, 326: 289–293
- 37 Won H, de la Torre-Ubieta L, Stein J L, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*, 2016, 538: 523–527
- 38 Olson M V. When less is more: Gene loss as an engine of evolutionary change. *Am J Hum Genet*, 1999, 64: 18–23
- 39 Olson M V, Varki A. Sequencing the chimpanzee genome: Insights into human evolution and disease. *Nat Rev Genet*, 2003, 4: 20–28
- 40 Stedman H H, Kozyak B W, Nelson A, et al. Myosin gene mutation correlates with anatomical changes in the human lineage. *Nature*, 2004, 428: 415–418
- 41 Perry G H, Verrelli B C, Stone A C. Comparative analyses reveal a complex history of molecular evolution for human *MYH16*. *Mol Biol Evol*, 2005, 22: 379–382
- 42 McCollum M A, Sherwood C C, Vinyard C J, et al. Of muscle-bound crania and human brain evolution: The story behind the MYH16 headlines. *J Hum Evol*, 2006, 50: 232–236
- 43 Winter H, Langbein L, Krawczak M, et al. Human type I keratin pseudogene phi hHaA has functional orthologs in the chimpanzee and gorilla: Evidence for recent inactivation of the human gene after the Pan-Homo divergence. *Hum Genet*, 2001, 108: 37–42
- 44 Chou H H, Hayakawa T, Diaz S, et al. Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc Natl Acad Sci USA*, 2002, 99: 11736–11741
- 45 Hayakawa T, Satta Y, Gagneux P, et al. Alu-mediated inactivation of the human CMP-N-acetylneuraminic acid hydroxylase gene. *Proc Natl Acad Sci USA*, 2001, 98: 11399–11404
- 46 Puppione D L, Ryan C M, Bassilian S, et al. Detection of two distinct forms of *apoC-I* in great apes. *Comp Biochem Physiol*, 2010, 5: 73–79
- 47 Lucatelli J F, Barros A C, Silva V K, et al. Genetic influences on Alzheimer's disease: Evidence of interactions between the genes *APOE*, *APOC1* and *ACE* in a sample population from the South of Brazil. *Neurochem Res*, 2011, 36: 1533–1539
- 48 Hansen J B, Fernández J A, Deguchi H, et al. The apolipoprotein CI content of very low density lipoproteins is associated with fasting triglycerides, postprandial lipemia, and carotid atherosclerosis. *J Lipids*, 2011, doi: 10.1155/2011/271062
- 49 Sudmant P H, Kitzman J O, Antonacci F, et al. Diversity of human copy number variation and multicopy genes. *Science*, 2010, 330: 641–646
- 50 Popesco M C, MacLaren E J, Hopkins J, et al. Human lineage-specific amplification, selection, and neuronal expression of *DUF1220* domains. *Science*, 2006, 313: 1304–1307
- 51 Holloway R L. The poor brain of *Homo sapiens neanderthalensis*: See what you please. In: *Ancestors: The Hard Evidence*, 1985. 319–324
- 52 Dumas L J, O'Bleness M S, Davis J M, et al. *DUF1220*-domain copy number implicated in human brain-size pathology and evolution. *Am J Hum Genet*, 2012, 91: 444–454
- 53 Dennis M Y, Nuttle X, Sudmant P H, et al. Evolution of human-specific neural *SRGAP2* genes by incomplete segmental duplication. *Cell*, 2012, 149: 912–922
- 54 Charrier C, Joshi K, Coutinho-Budd J, et al. Inhibition of *SRGAP2* function by its human-specific paralogs induces neoteny during spine maturation. *Cell*, 2012, 149: 923–935
- 55 Fossati M, Pizzarelli R, Schmidt E R, et al. *SRGAP2* and its human-specific paralog co-regulate the development of excitatory and inhibitory synapses. *Neuron*, 2016, 91: 356–369
- 56 Florio M, Albert M, Taverna E, et al. Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. *Science*, 2015, 347: 1465–1470

-
- 57 Nuttle X, Giannuzzi G, Duyzend M H, et al. Emergence of a *Homo sapiens*-specific gene family and chromosome 16p11. 2 CNV susceptibility. *Nature*, 2016, 536: 205–209
- 58 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 2006, 126: 663–676
- 59 Marchetto M C N, Narvaiza I, Denli A M, et al. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature*, 2013, 503: 525–529
- 60 Romero I G, Pavlovic B J, Hernando-Herraez I, et al. A panel of induced pluripotent stem cells from chimpanzees: A resource for comparative functional genomics. *Elife*, 2015, 4: e07103
- 61 Prescott S L, Srinivasan R, Marchetto M C, et al. Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. *Cell*, 2015, 163: 68–83
- 62 Otani T, Marchetto M C, Gage F H, et al. 2D and 3D stem cell models of primate cortical development identify species-specific differences in progenitor behavior contributing to brain size. *Cell Stem Cell*, 2016, 18: 467–480
- 63 Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc*, 2014, 9: 1956–1968
- 64 Yang H, Wang H, Shivalila C S, et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*, 2013, 154: 1370–1379
- 65 Hiatt J B, Pritchard C C, Salipante S J, et al. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res*, 2013, 23: 843–854
- 66 Maricic T, Günther V, Georgiev O, et al. A recent evolutionary change affects a regulatory element in the human *FOXP2* gene. *Mol Biol Evol*, 2013, 30: 844–852



宿兵

中国科学院昆明动物研究所研究员，理学博士，博士生导师；国家杰出青年科学基金获得者。毕业于武汉大学生物系并于中国科学院昆明动物研究所获博士学位，曾在美国德克萨斯大学休斯顿人类遗传学中心从事博士后研究，在美国辛辛那提大学医学院任助理教授。从事灵长类比较基因组学及现代人类起源和迁徙的研究。已在 *Science*, *Nature*, *Nature Reviews Genetics* 等国际核心刊物上发表论文 150 余篇。

Summary for “人类独特表型遗传基础的研究进展”

Genetic bases of human unique traits

LUO Xin^{1,2}, ZHANG DongQin^{1,2} & SU Bing^{1*}¹ State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China;² Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650223, China

* Corresponding author, E-mail: sub@mail.kiz.ac.cn

Being human, we possess unique biological features, such as the big brain, upright walking and sophisticated language, which set us apart from our close relatives, the non-human primates since our divergence from the Pan lineage approximately 5 million years ago. Scientists have attempted to solve the mystery of our unique phenotypic traits with approaches from paleoanthropology, anatomy, psychology, behavior and social cognition etc. However, the answers are embedded in our genetic makeup. What are the genetic changes in our genome that make us human? By comparing the genomes of humans and nonhuman primates, scientists have identified a lot of genetic changes that might have contributed to the human-specific traits. These genetic changes work at different genetic levels, including sequence changes in protein-coding genes, gene expression changes, gene losses and gene duplications. Natural selection is the driving force pushing fixation of these genetic changes during human evolution, and they may have important contribution to human's unique phenotypes. In this review, we mainly discuss the relationship between human-specific phenotypes and genetic changes. We summarize our current findings and understandings of genes with human-specific mutations and their possible functional effects. For example, a set of brain-size regulating genes have been identified and many of them showed strong signatures of Darwinian positive selection during primate evolution and human origin. These human-specific mutations likely work their ways through human brain development and eventually form a large brain. Another fascinating example is the language gene *FOXP2* that has two human specific mutations, which may lead to a unique developmental pattern change of the human brain region responsible for human language ability. Furthermore, besides protein coding genes, there are other genetic changes, and many identified fast-evolving regions in the human genome are not located in the coding sequences. Instead they are distributed in regulatory sequences likely involved in gene expression regulation which affect the temporal and spatial pattern of human development. Finally, we introduce the recent advance of technologies that can be used to study the functional outcomes of the human specific mutations. These technologies provide useful tools that may reproduce the scenes of human evolution in the laboratory. The *in vitro* analysis of humanized iPSCs can to some extent mimic the *in vivo* development of the human brain. Notably, the newly developed gene editing tools including CRISPR-Cas9 and TALEN are promising in studying the human-specific traits using transgenic animal models, especially the transgenic monkey models which may provide much more mechanistic information than the traditional mouse model. The human-specific genetic changes accumulated during human evolution occurred at different levels, but they must work together to produce the human-specific traits. Future studies may establish the links by figuring out how changes at different genetic levels coordinate with each other to make us human.

human unique phenotype, coding gene, gene expression regulation, gene loss, gene duplication

doi: 10.1360/N972016-01063