



人群变异的分子基础: 从单核苷酸多态性到单氨基酸多态性

苏智端^①, 林旭^②, 曾嵘^{①*}, 吴家睿^{①*}^① 中国科学院上海生命科学研究院, 生物化学与细胞生物学研究所, 系统生物学重点实验室, 上海 200031;^② 中国科学院上海生命科学研究院, 营养科学研究所, 营养与代谢重点实验室, 上海 200031

* 联系人, E-mail: zr@sibs.ac.cn; wujr@sibs.ac.cn

收稿日期: 2012-08-07; 接受日期: 2012-11-20

国家重大科学研究计划(批准号: 2011CB910200, 2011CB910601)和国家自然科学基金(批准号: 31130034)资助项目

doi: 10.1360/052012-278

摘要 单核苷酸多态性可以划分为位于基因编码区的 SNP 和非编码区的 SNP 两大类; 而在基因编码区的 SNP 还可以进一步划分为两个亚类: 不改变氨基酸序列的同义 SNP 和改变氨基酸序列的非同义 SNP. 显然, 非同义 SNP 将导致氨基酸序列的改变, 即形成单氨基酸多态性. 基于蛋白质组学方法, 对亚洲人群血浆样本中的 SAP 进行了系统研究, 发现某一特定 SAP 在纯合人群和杂合人群中可能与生理或病理性状有着不同的关联. 更为重要的是, 近期有研究发现, 在生物体中广泛存在着 RNA 序列与 DNA 序列不一致的现象. 导致这种差异的主要原因是在转录水平上存在着规模化的 RNA 编辑(被称为 RNA 编辑组, RNA editome). 该发现表明, 个体拥有的 SAP 中可能有一部分与基因组 SNP 无关, 而是源于 RNA 编辑组. 进一步推论, 可能在翻译水平上存在着不依赖 DNA 和 RNA 序列的全新的 SAP.

关键词单核苷酸多态性(SNPs)
单氨基酸多态性(SAPs)
RNA 编辑组
蛋白质编辑
中心法则

在 2003 年人类基因组计划完成后, 人们开始致力于人群水平的遗传变异研究, 以求更好地了解及抗击癌症、糖尿病等复杂性疾病. 2003 年开始的国际人类基因组单体型图计划, 研究重点就在于揭示非洲、亚洲及欧洲人群的基因组变异谱图. 由于单个碱基变异引起的单核苷酸多态性(single-nucleotide polymorphism, SNP)是人类基因组变异的最普遍方式, 与众多生理病理性状相关. 据统计, 人类基因组中约有上千万个 SNP 位点. 2005 年, 国际人类基因组单体型图计划组织公布了第一版的人类基因组

单体型图谱, 该图谱包含了超过 100 万个可供分析的 SNP 位点^[1]. 此后, 该组织又在 2007 年公布了第 2 版人类基因组单体型图谱, 共包含了超过 300 万个 SNP 位点^[2].

当前, 大量的 SNP 位点可用于生理病理性状关联分析. 一般来说, SNP 可分为以下 2 类: 改变氨基酸序列或转录因子结合元件的功能型 SNP 以及不能改变蛋白质表达的非功能型 SNP. 在不同的人群中, 非功能型 SNP 可能与功能型 SNP 发生不同程度的连锁不平衡, 所以为了避免在分析时其可能引发的潜

引用格式: 苏智端, 林旭, 曾嵘, 等. 人群变异的分子基础: 从单核苷酸多态性到单氨基酸多态性. 中国科学: 生命科学, 2013, 43: 16–22, doi: 10.1360/052012-278
Su Z D, Lin X, Zeng R, et al. SNPs and SAPs are molecular bases of population variation. SCIENTIA SINICA Vitae, 2013, 43: 16–22, doi: 10.1360/052012-278

在问题, Humphries 等人^[3]强烈建议只针对功能型 SNP 进行关联分析. 因此, 在蛋白质组水平对功能型 SNP 进行研究十分必要.

1 单氨基酸多态性

根据是否位于基因编码区将 SNP 分为编码 SNP 和非编码 SNP. 由于遗传密码子存在简并性的特征, 一部分编码 SNP 并不能改变其编码的蛋白质氨基酸序列, 另一部分则可以改变. 因此, 编码 SNP 可以进一步划分为同义 SNP 和非同义 SNP. 显然, 非同义 SNP 属于功能型 SNP 类别. 最近, Schaefer 等人^[4]建立了一个非同义 SNP 数据库(SNPdb), 其中整合了 3 个 SNP 数据库信息, 包括 130 多万个可引起单氨基酸替换的非同义 SNP 信息.

大量研究工作揭示, 导致氨基酸序列变异的非同义 SNP 通常会对基因的功能造成显著影响^[5-7]. Yoshiura 等人^[5]发现, 只有 *ABCC11*(ATP-binding cassette transporter sub-family C member 11)基因上的非同义 SNP(rs17822931, 180G/R)可以决定人类的耳垢类型, 而位于内含子或者 Alu 重复序列的 SNP 并不会影响该性状, 尽管在关联分析中这些 SNP 与 rs17822931 具有相似的显著性 *P* 值. 在另一项研究中, Kubo 等人^[6]利用日本人群样本, 通过对 5 万多个基因的 SNP 位点进行分析, 发现在编码蛋白质激酶 C 家族 *PRKCH*(protein kinase C eta)基因的第 9 外显子上, 存在着一个可增加脑梗塞患病风险的非同义 SNP(rs2230500, 1425G/A). 最近的一项中国人群病例-对照配对及其相应的生化实验分析发现, 位于 *CTLA-4*(cytotoxic T-lymphocyte antigen 4)基因的非同义 SNP(rs231775, 49G/A)参与了 T 细胞激活的调节过程, 可增加多种癌症的易感性, 如肺癌、乳腺癌、食道癌、胃癌等^[7]. 当然, 也存在着某些不影响基因功能的非同义 SNP. 例如, 有研究发现, 位于 *DNase1L1* 基因的非同义 SNP(rs34952165, 122V/I)对 DNase 的酶活性无明显影响^[8].

为了研究人群中非同义 SNP 基因型分布, Takeshita 等人^[9]在亚洲人群、非洲人群及高加索人群等 13 个不同的健康人群中系统地分析了位于 3 个细胞凋亡相关的 DNA 内切酶基因 *ENDOG*(endonuclease G), *DFFB*(DNA fragmentation factor beta)和 *FEN-1*(Flap endonuclease-1)上的 13 个非同义 SNP 位

点. 他们发现, 只有 *ENDOG_S12L*, *DFFB_R196K* 及 *DFFB_K277R* 3 个非同义 SNP 具有多态性. 基于该实验室的前期工作, Takeshita 等人^[9]推断, 核酸内切酶具有非同义 SNP 的低水平遗传异质性的特点. 另一方面, 中国科学家和丹麦科学家通过对 200 个丹麦人外显子的深度测序, 发现了位于基因编码区 18654 个注释基因上的 53081 个 SNP 位点, 其中有一半的 SNP 为非同义 SNP^[10]. 此外, 他们还发现, 在低分布频率范围(次要等位基因频率 2%~5%), 有害的非同义 SNP 数量超过同义 SNP 数量的 2 倍以上. 这一结果说明, 低分布频率非同义 SNP 可能对进化适应性产生重要影响^[10].

既然位于基因编码区的非同义 SNP 一定会在蛋白质翻译过程中造成氨基酸序列的变异, 那么在蛋白质水平直接对它们进行分析就应该更有利于揭示其特有属性. 尽管在最近 10 年内已经广泛开展了 SNP 研究, 但目前在人群中针对蛋白质氨基酸序列多态性的研究尚未引起重视. 通过对国际生物医学文献库“PubMed Database”的搜索, 找到了 4 篇有关“单氨基酸多态性”研究的文章, 其中有 2 篇文章是建立“单氨基酸多态性”的数据库, 另外 2 篇文章则研究了心肌组织和一个蛋白质的“单氨基酸多态性”. 最近, 本实验室提出, 这种“单氨基酸多态性”应该用英文字母“SAP”(single amino-acid polymorphism)表示, 以便与“SNP”相对应, 并首次利用定量蛋白质组学方法, 在群体水平对汉族人群的 SAP 进行了系统性的研究^[11].

为了在人群血浆样本中对 SAP 进行系统的检测, 本实验室利用 Swiss-Prot 高可信数据库中的有关数据, 首次构建了一个包含 439 个 SAP 位点, 对应于 29 个血浆蛋白质的血浆相关的 SAP 数据库. 基于这个 SAP 数据库, 对 33 例亚洲人血浆样本进行了“鸟枪法”蛋白质组学分析^[11], 共鉴定到 2154 条 SAP 肽段, 对应 46 个特异性 SAP 位点. 为了进一步揭示 SAP 在人群水平的分布特性, 挑选出分别对应补体系统 C5 (complement component 5)、补体系统 C7(complement component 7)以及补体因子 H(complement factor H, CFH)等补体系统中 3 个不同蛋白的 3 对 SAP 肽段, 并利用靶标蛋白质组学策略(选择反应监控技术, 简称“SRM(selected reaction monitoring)技术”)^[12]在 290 例亚洲人群血浆样本中对这些 SAP 肽段的绝对丰度进行了定量检测.

2 SAP 的定性及定量分析

对于人类及其他二倍体物种来说, SNP 对于某一等位基因可引起两种基因型结构: 纯合子和杂合子^[13]. 有报道称, 编码 SNP 的不同基因型结构与多种生理病理性状密切相关^[5-7]. 例如, Yoshiura 等人^[5]在日本人群中对 *ABCC11* 基因上的非同义 SNP (rs17822931, 180G/R) 研究后发现, AA 型纯合个体与干耳垢相关, 而 GG 型纯合个体以及 GA 型杂合个体与湿耳垢相关. 此外, 中国科研人员发现, 在中国肺癌患者人群中, *CTLA-4* 基因 AA 型(非同义 SNP, rs231775)纯合基因型频率显著高于正常人群, 但该基因 GA 型杂合基因型却并不与肺癌相关^[7]. 另一方面, Chess 博士的研究小组发现, 在人类常染色体上普遍存在着单等位基因表达现象^[14]. 另一个研究组也发现, *FEN-1* 基因的 6 个非同义 SNP 在 13 个检测人群中均表现为单等位基因^[9]. 这些研究工作表明, 某些 SNP 并不会出现基因型的异质性.

通过 SRM 技术对这 3 对 SAP 肽段的定量分析发现, 在中国糖尿病人群和中国超重/肥胖人群中, C5_D966Y 的杂合个体分布频率显著高于正常人群, 而在同一人群 C7_T587P 及 CFH_V62I 的研究中并未发现类似的杂合子分布差异(文献[11]图 2A). 进一步研究发现, 在正常人群中 C7_587P 型纯合个体的分布频率显著高于其他 3 组代谢紊乱人群(文献[11]图 2B). 总体来说, 在人群中蛋白质组水平上的 SAP 结构与基因组水平的 SNP 基因型结构有着相似的分布特性.

从定量的角度来看, 对于拥有杂合型 SNP 的个体, SNP 等位基因的差异性表达在转录组水平已得到了广泛的检测^[15,16]. 据报道, 在 4 种人源细胞系中, 约有 11%~22% 的 SNP 等位基因都有差异性表达, 而另有 4.3%~8.5% 的 SNP 等位基因则表现为组织特异性^[17]. 这种 SNP 等位基因的差异性表达往往与个体间的生理病理性状差异相关. 近期的一项研究发现, 具有最为常见的 H2 单倍型以及 *TGFBR1* (transforming growth factor beta receptor 1) 等位基因差异性表达的个体对于结肠癌有着更为显著的易感性^[18]. 虽然杂合 SNP 的定量信息可以在转录组水平进行检测, 但对于多态性相关的蛋白质组水平表达则需要利用蛋白质组学方法进行定量分析. 有一项研究工作值得关注, 哈佛大学谢晓亮博士的研究小组发展了一种

荧光蛋白质检测系统, 能够在单分子水平定量分析蛋白质及 mRNA 表达的相关性, 发现蛋白质的表达量与 mRNA 的拷贝数并无关联^[19]. 这一工作表明, 在蛋白质组水平定量分析 SAP 是很有必要的, 因为在蛋白质组水平的 SAP 差异性表达与在转录组水平研究 SNP 的 mRNA 差异性表达可能不是线性相关的, 因此, SAP 差异性表达方面的研究也不能被 SNP 相关的 mRNA 差异性表达的研究所简单替代.

利用 SRM 定量分析方法, 在 290 例亚洲人群血浆样本中对 3 对 SAP 的绝对丰度进行了检测^[11]. 研究表明, 在纯合子中 C7_578T, CFH_62V 以及 CFH_62I 与代谢紊乱表现出更显著的相关性(文献[11]图 3). 此外, 对 3 对 SAP 在 4 组人群的杂合个体中的分布进行了定量分析, 分别揭示了它们各自与不同生理和病理性状的相关性(文献[11]图 4).

本实验室工作首次在人群水平揭示了蛋白质多态性的定量特征, 并从中总结出以下 3 点: (i) Western blotting 和 ELISA 等分析蛋白质表达量的传统方法只能检测蛋白质的总体表达量, 并不能在杂合子中分别检测 SAP 蛋白质亚型各自的丰度, 而 SRM 方法则可以很好地满足这一研究需求; (ii) 在杂合子中, 不同 SAP 蛋白质亚型可能与不同生理病理性状表现出不同的相关性, 如 C7_587T 在杂合子中与糖尿病及肥胖显著相关, 而 C7_587P 则并未表现出类似的相关性^[7](图 4); 最近利用 SRM 方法在 1000 多例亚洲人群血浆样本中对载脂蛋白 A4 的某个 SAP 进行了定量检测, 发现其中一个蛋白质 SAP 亚型与甘油三酯的丰度呈显著正相关, 而与之对应的另一个蛋白 SAP 亚型与甘油三酯的丰度则表现为显著负相关(未发表数据); (iii) 基于 SRM 技术的血浆 SAP 定量分析将从蛋白质组水平上揭示蛋白质变异与生理病理性状的关联, 因此, SAP 将成为群体的蛋白质多态性特性分析中的一类新的生物标记物.

3 3 种可能的 SAP 来源

最初提出 SAP 这一概念是基于 1958 年 Crick 提出的分子生物学中心法则, 即 SAP 是源于基因组的非同义 SNP(图 1). 中心法则的关键在于描述了遗传信息如何从基因组水平传递到蛋白质组水平, 即位于 DNA 序列中的遗传信息可通过转录过程传递给 mRNA 的核苷酸序列, 后者又可以通过遗传密码子

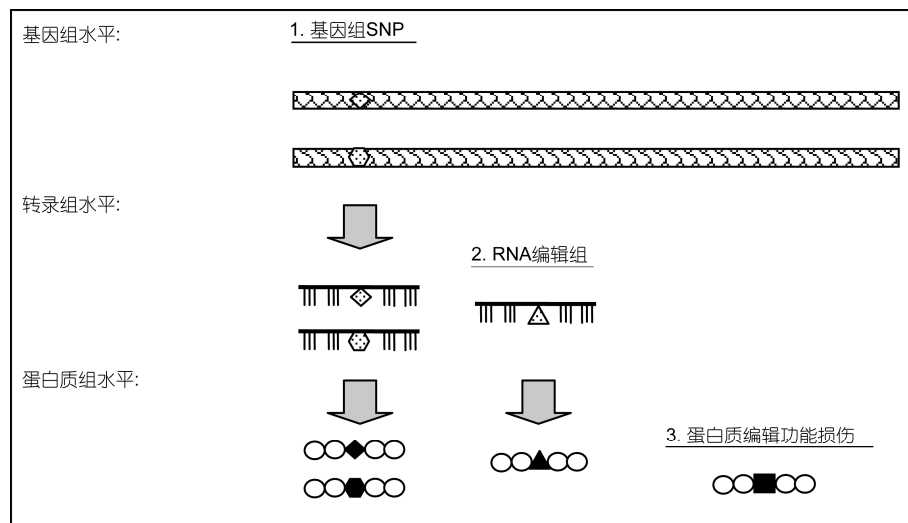


图 1 SAP 的 3 种可能的来源

第 1 种是基因组的非同义 SNP(有阴影的几何图形), 这是 SAP 的主要来源; 第 2 种 SAP 来源于 RNA 编辑组(有阴影的几何图形); 第 3 种 SAP 可能是源自蛋白质编辑功能损伤, 这类 SAP 不依赖 DNA 或 mRNA 序列

经翻译过程传递给蛋白质的氨基酸序列。由于中心法则的存在, 长久以来人们相信, mRNA 序列必定严格由 DNA 序列决定, 而蛋白质的氨基酸序列也必定经遗传密码子严格由 mRNA 序列决定。显然, 基因组的非同义 SNP 应该是个体的 SAP 的主要来源。

尽管早在 20 世纪 80 年代就已经发现, 在转录过程中存在引起 mRNA 的核苷酸变异的“RNA 编辑”现象^[20], 但这并没有改变人们对于中心法则的坚定信念, “RNA 编辑”现象被认为是在转录过程中遗传信息的罕见变异。但是, Cheung 的研究小组最近发现, 在人类转录组上广泛存在着 RNA 序列与 DNA 序列不一致的现象^[21]。他们共发现超过 1 万个 RNA 外显子序列差异, 涉及所有 12 种可能的 RNA-DNA 差异位点^[21]。有些科学家质疑 Cheung 的这一结果可能包含了他们在分析测序数据时所产生的大量假阳性结果^[22]。Schridder 等人^[23]甚至通过重新分析 Cheung 的数据集, 得到了完全相反的结论。但是, 现在越来越多的研究工作提供了强有力的证据, 支持 Cheung 的结论。ADAR (adenosine deaminases acting on RNA) 是一种已知的可催化腺嘌呤转变为次黄嘌呤的 RNA 编辑酶^[24], Bahn 等人^[25]利用 siRNA 技术敲除 ADAR 蛋白, 发现在基因组水平 RNA-DNA 序列差异现象被明显抑制。Bahn 等人的这一发现证明, 这些广泛存在的转录组水平上的核苷酸变异是由 RNA 编辑这一机制引起的, 他们把这种现象称为“RNA 编辑组”^[25]。Bahn 等人进一

步发现, 绝大多数 RNA 编辑属于腺嘌呤-次黄嘌呤型, 而且可能发生腺嘌呤-次黄嘌呤 RNA 编辑的基因显著富集在癌症相关的基因类别, 提示可能存在癌症特异的 RNA 编辑现象^[25]。其他一些研究工作同样支持 Bahn 等人的结论, 甚至在 microRNA 领域也发现了 RNA 编辑组^[26]。有趣的是, Snyder 领导的研究小组^[27]通过对一个正常个体的外周血液单核细胞在 14 个月时间周期内的动态变化研究, 不仅发现了大量 RNA 编辑现象, 并且首次揭示 RNA 编辑水平在该时间周期内是动态变化的。这一发现表明, RNA 编辑组的组成可能会随着外部环境的变化而改变^[27]。显然, 这些关于 RNA 编辑组的研究工作对中心法则提出了巨大的挑战: 在遗传信息从基因组到转录组的传递表达过程中, 可能存在着大规模的核苷酸变异。正如 Hayden 所指出: “若经证实, 这些发现势必改写分子生物学的中心法则”^[22]。

既然 RNA 编辑组导致了 mRNA 的单核苷酸变异, 因此就能够推断出, RNA 编辑组是产生 SAP 的另一个来源。显然, SAP 的 RNA 编辑组来源独立于 SNP 这种来源类型(图 1)。Cheung 领导的研究小组^[21]在利用蛋白质组学方法对人类 B 细胞蛋白质进行总体分析后, 发现部分肽段的氨基酸序列并不是由 DNA 编码序列精确编码的。另外, Snyder 领导的研究小组^[27]利用质谱方法, 在一个包含 4586 条单核苷酸变异序列(等同于非同义 SNP)及全部 30385 条 RNA 编辑序

列的数据库里搜索后, 共鉴定到 48 条单核苷酸变异序列及 51 条 RNA 编辑序列. 该结果说明, 包含 RNA 编辑的转录本的确可以作为模板指导蛋白质的合成.

“蛋白质编辑”的概念与“RNA 编辑”的概念正好相反, 主要是指在蛋白质合成过程中对氨基酸的组装进行校正从而减少在多肽链上的氨基酸变异的作用机制^[28,29]. 早在 20 世纪 70 年代, 研究人员就发现氨酰 tRNA 合成酶可利用其编辑结构域对错误组装的氨酰 tRNA 进行水解^[30]. 这些编辑结构域的结构特征和作用方式也得到了系统地研究^[31,32]. 此外, 其他类型的蛋白质编辑机制也陆续被发现, 包括核糖体能够选择识别正确的氨酰 tRNA 并丢弃错误的氨酰 tRNA, 即所谓的诱导适应机制^[33]. 从理论上讲, 蛋白质编辑的功能紊乱将会导致翻译过程中蛋白质序列保真度的降低. Lee 等人^[34]研究发现, 氨酰 tRNA 合成酶编辑功能的缺失会导致错配的 tRNA 增加, 并会在神经元细胞中累计这些错误翻译的蛋白质. 此外, Nangle 等人^[35]在哺乳动物细胞中采用转基因技术表达一个编辑功能缺失的氨酰 tRNA 合成酶, 发现 EGFP (enhanced green fluorescent protein) 受体的某些特定氨基酸的错误插入有明显的增加. 重要的是, 通过抑制蛋白质编辑功能而增加翻译过程出错几率的机制在进化过程中可能具备选择优势. 研究人员最近发现, 支原体类寄生虫的氨酰 tRNA 合成酶编辑结构域在自然条件下存在点突变或缺失, 并能够导致蛋白质的错误翻译^[36]. 他们推测, 这些在自然条件下存在的寄生虫蛋白质编辑突变可能是在进化过程中演变形成的, 成为一种用来产生抗原多样性的机制, 从而有利于逃避宿主的防御系统^[36].

据估计, 蛋白质翻译错误的发生几率大约在万分之一至千分之一^[32,37]. 基于这个出错概率, 认为在

平均长度为 400 个氨基酸的蛋白质群体中, 大约有 18% 的蛋白质包含至少一个氨基酸的错义替换^[38]. 但是, 由于人们尚未在蛋白质组水平对氨基酸变异进行广泛的分析, 因此, 蛋白质翻译过程中的氨基酸变异几率很可能被低估, 在人群水平上很可能存在着更为广泛的氨基酸变异. 如果在蛋白质翻译过程中存在广泛的氨基酸变异, 一个合理的推论是, 很可能存在着 SAP 的第 3 种来源, 即蛋白质编辑功能的紊乱或其他在蛋白质合成过程中及合成过程后增加氨基酸变异的未知机制(图 1). 这第 3 种 SAP 来源与基因组水平的 SNP 来源或转录水平的 RNA 编辑组来源均无关, 即细胞中可能存在着一类与 DNA 序列和 mRNA 序列都没有直接关系的全新的 SAP(*de novo* SAP). 显然, 一旦人们通过研究发现这类不依赖于核酸编码序列的 SAP, 将在蛋白质组水平又一次对中心法则提出挑战.

4 展望

蛋白质单氨基酸多态性研究是继基因多态性研究之后的一个重要研究新领域, 但目前尚未引起足够的重视, 尤其在定量分析方面属于空白. 要利用目前已经成熟的深度测序技术和靶向蛋白质定量分析技术等研究方法, 研究源于 SNP 或源于 RNA 编辑组的 SAP 之间的比例和特征. 同时, 还需要发展基于从头测序的蛋白质多态性位点发现新方法, 以检测不依赖于核酸序列的全新 SAP 的存在和分布特点. 在这些技术的帮助下, 将能够深入研究 SAP 在不同人群中的分布特点, 以及与不同生理病理性的关联性, 从而为人们抗击重大慢性病提供新型的分子靶标和分子标记物.

参考文献

- 1 The International HapMap Consortium. A haplotype map of the human genome. *Nature*, 2005, 437: 1299–1320
- 2 The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 2007, 449: 851–861
- 3 Humphries S E, Ridker P M, Talmud P J. Genetic testing for cardiovascular disease susceptibility: a useful clinical management tool or possible misinformation? *Arterioscler Thromb Vasc Biol*, 2004, 24: 628–636
- 4 Schaefer C, Meier A, Rost B, et al. SNPdb: constructing an nsSNP functional impacts database. *Bioinformatics*, 2012, 28: 601–602
- 5 Yoshiura K, Kinoshita A, Ishida T, et al. A SNP in the *ABCC11* gene is the determinant of human earwax type. *Nat Genet*, 2006, 38: 324–330
- 6 Kubo M, Hata J, Ninomiya T, et al. A nonsynonymous SNP in *PRKCH* (protein kinase C eta) increases the risk of cerebral infarction. *Nat Genet*, 2007, 39: 212–217
- 7 Sun T, Zhou Y, Yang M, et al. Functional genetic variations in cytotoxic T-lymphocyte antigen 4 and susceptibility to multiple types of

- cancer. *Cancer Res*, 2008, 68: 7025–7034
- 8 Ueki M, Fujihara J, Takeshita H, et al. Genetic and expression analysis of all non-synonymous single nucleotide polymorphisms in the human deoxyribonuclease I-like 1 and 2 genes. *Electrophoresis*, 2010, 31: 2063–2069
 - 9 Takeshita H, Fujihara J, Ueki M, et al. Nonsynonymous single-nucleotide polymorphisms of the human apoptosis-related endonuclease—DNA fragmentation factor beta polypeptide, endonuclease G, and Flap endonuclease-1—genes show a low degree of genetic heterogeneity. *DNA Cell Biol*, 2012, 31: 36–42
 - 10 Li Y, Vinckenbosch N, Tian G, et al. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet*, 2010, 42: 969–972
 - 11 Su Z D, Sun L, Yu D X, et al. Quantitative detection of single amino acid polymorphisms by targeted proteomics. *J Mol Cell Biol*, 2011, 3: 309–315
 - 12 Addona T A, Abbatiello S E, Schilling B, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol*, 2009, 27: 633–641
 - 13 Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat Rev Genet*, 2010, 11: 533–538
 - 14 Gimelbrant A, Hutchinson J N, Thompson B R, et al. Widespread monoallelic expression on human autosomes. *Science*, 2007, 318: 1136–1140
 - 15 Yan H, Yuan W, Velculescu V E, et al. Allelic variation in human gene expression. *Science*, 2002, 297: 1143
 - 16 Pickrell J K, Marioni J C, Pai A A, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*, 2010, 464: 768–772
 - 17 Zhang K, Li J B, Gao Y, et al. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat Methods*, 2009, 6: 613–618
 - 18 Martinez-Canto A, Castillejo A, Mata-Balaguer T, et al. TGFBR1 intralocus epistatic interaction as a risk factor for colorectal cancer. *PLoS One*, 2012, 7: e30812
 - 19 Taniguchi Y, Choi P J, Li G W, et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 2010, 329: 533–538
 - 20 Lewin R. Editing mRNA precursors. *Science*, 1983, 222: 493
 - 21 Li M, Wang I X, Li Y, et al. Widespread RNA and DNA sequence differences in the human transcriptome. *Science*, 2011, 333: 53–58
 - 22 Hayden E C. Evidence of altered RNA stirs debate. *Nature*, 2011, 473: 432
 - 23 Schrider D R, Gout J F, Hahn M W. Very few RNA and DNA sequence differences in the human transcriptome. *PLoS One*, 2011, 6: e25842
 - 24 Barraud P, Allain F H. ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr Top Microbiol Immunol*, 2012, 353: 35–60
 - 25 Bahn J H, Lee J H, Li G, et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res*, 2012, 22: 142–150
 - 26 Peng Z, Cheng Y, Tan B C, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol*, 2012, 30: 253–260
 - 27 Chen R, Mias G I, Li-Pook-Than J, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*, 2012, 148: 1293–1307
 - 28 Rodnina M V, Wintermeyer W. Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem Sci*, 2001, 26: 124–130
 - 29 Cochella L, Green R. Fidelity in protein synthesis. *Curr Biol*, 2005, 15: R536–R540
 - 30 Fersht A R. Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry*, 1977, 16: 1025–1030
 - 31 Guo M, Schimmel P. Structural analyses clarify the complex control of mistranslation by tRNA synthetases. *Curr Opin Struct Biol*, 2012, 22: 119–126
 - 32 Ogle J M, Ramakrishnan V. Structural insights into translational fidelity. *Annu Rev Biochem*, 2005, 74: 129–177
 - 33 Pape T, Wintermeyer W, Rodnina M. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J*, 1999, 18: 3800–3807
 - 34 Lee J W, Beebe K, Nangle L A, et al. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature*, 2006, 443: 50–55
 - 35 Nangle L A, Motta C M, Schimmel P. Global effects of mistranslation from an editing defect in mammalian cells. *Chem Biol*, 2006, 13: 1091–1100
 - 36 Li L, Boniecki M T, Jaffe J D, et al. Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in *Mycoplasma* parasites. *Proc Natl Acad Sci USA*, 2011, 108: 9378–9383
 - 37 Kramer E B, Farabaugh P J. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA*, 2007, 13: 87–96
 - 38 Drummond D A, Wilke C O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*, 2008, 134: 341–352

SNPs and SAPs are Molecular Bases of Population Variation

SU ZhiDuan¹, LIN Xu², ZENG Rong¹ & WU JiaRui¹

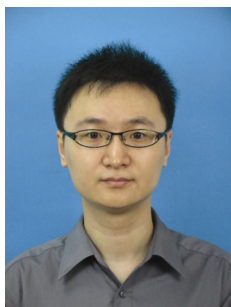
1 Key Laboratory of Systems Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China;

2 Key Laboratory of Nutrition and Metabolism, Institute of Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

There are two types of single nucleotide polymorphisms (SNPs): within coding regions and within non-coding regions. SNPs within coding regions should be further divided into two categories: synonymous and non-synonymous. The non-synonymous SNPs can result in variation of amino-acid sequences, which are single amino-acid polymorphisms (SAPs). Recently, using quantitative proteomic approaches, SAPs in the plasma proteomes were identified at population level for the first time, which showed that heterozygous and homozygous proteins with various SAPs have different associations with particular traits in the population. RNA editome that has been uncovered recently might be a new source of SAPs. In addition, there is another possible source of SAPs that is *de novo* one independent on both DNA and RNA sequences.

single nucleotide polymorphisms, single amino-acid polymorphism, RNA editing, protein editing, the central dogma

doi: 10.1360/052012-278



苏智端, 2007年毕业于山东大学, 获生物技术专业理学学士学位; 2012年毕业于中国科学院上海生命科学研究院生物化学与细胞生物学研究所, 获生物化学与分子生物学理学博士学位; 现于中国科学院上海生命科学研究院系统生物学重点实验室从事博士后研究工作. 研究领域: 利用基于质谱技术的蛋白质组学方法, 进行单氨基酸多态性(SAPs)鉴定及定量的方法学研究, 并对SAPs在多种生物临床样本中的生物学意义进行深入分析.



吴家睿, 1982年毕业于广州中山大学生物系; 1985年在中国科学院遗传与发育生物学研究所获硕士学位; 1994年在瑞士苏黎世联邦理工学院获博士学位; 1994年10月在纽约州立大学卫生科学中心进行博士后研究; 1997年10月起在中国科学院上海生命科学研究院生物化学与细胞生物学研究所从事细胞分子生物学研究工作. 现任中国科学院系统生物学重点实验室主任; 中国生物化学与分子生物学会副理事长; 中国科学院“百人计划”和国家杰出青年科学基金项目获得者; *Journal Molecular Cell Biology* 主编; *BMC Systems Biology* 副主编; *Frontiers in Systems Physiology* 副主编; *Journal of Biological Chemistry* 编委; 《科学通报》副主编.