

遗传基因组学 (Genetical genomics) 的研究进展*

刘秀艳^{1,2)**} 谢正苗^{1,2)} 陈惠哲³⁾

¹⁾杭州电子科技大学, 杭州 310018;

²⁾浙江大学环境与资源学院, 环境修复与生态健康教育部重点实验室, 杭州 310029;

³⁾中国水稻研究所, 杭州 310006)

摘要: 遗传基因组学 (genetical genomics) 是将微阵列技术和数量性状座位 (QTL) 分析结合起来, 全基因组水平上定位基因表达的 QTL (eQTL). 它为研究复杂性状的分子机理和调控网络提供全新的手段. 遗传基因组这个概念和研究策略在 2001 年由 Janson 和 Nap 首先提出, 到目前为止, 遗传基因组学已应用于酵母、老鼠、人以及玉米等植物. 研究结果表明: 基因表达水平的差异是可遗传的复杂性状; eQTL 可以分为顺式作用 eQTL 和反式作用 eQTL, 顺式作用 eQTL 就是某个基因的 eQTL 定位到该基因所在的基因组区域, 表明可能是该基因本身的差别引起 mRNA 水平的差别, 反式作用就是 eQTL 定位到其他基因组区域, 表明其他基因的差别控制该基因 mRNA 水平的差异. 将 eQTL 结果、基因功能注解以及多种统计分析方法相结合, 不仅能更准确地鉴别控制复杂性状及其相关基因表达的候选基因, 而且能构建相应的基因调控网络.

关键词 微阵列技术, 基因表达的数量性状座位 (QTL), 基因调控网络

学科分类号 Q55

生物的许多重要性状, 如作物的产量、动物的寿命, 以及人类肥胖、心脏病、精神分裂症的机理等, 是非常复杂的. 这些复杂性状的生物学过程由多个基因控制, 并受到环境因子的影响. 如何鉴别控制复杂性状的基因及基因网络并揭示其分子机理, 是一直困扰着人们的大问题. 现代分子生物学、基因组学、功能基因组学技术的飞速发展, 使得研究复杂性状的分子机理和调控网络成为可能. 20 世纪 90 年代发展的分子标记技术以及多种遗传作图和统计的算法和软件, 可以将控制复杂性状的基因 (数量性状座位, QTL) 定位到特定基因组区域并分析其效应, 定位的 QTL 可以进一步通过图位克隆等手段得到克隆分离. 另外, 在基因组水平上, 采用高通量的 DNA 微阵列技术对不同材料、不同处理或不同时间点进行全基因组基因表达谱的分析, 也成为研究复杂性状基因调控机理的另一个重要手段.

一直以来, QTL 分析和基因表达谱研究都是相互独立的, 其实验设计、生物材料和统计分析手段都是完全不同的. 直到 2001 年, Janson 和 Nap^[1] 提出了遗传基因组学 (genetical genomics) 这个概念

以及相应的研究策略, 将 DNA 微阵列这个基因组分析手段和 QTL 分析这个遗传学方法结合起来, 用来更深入地揭示复杂性状的分子机理和调控网络. 在随后的时间里, 遗传基因组学已成功地应用到酵母、鼠类、人、植物的研究中, 获得了大量的研究结果.

1 什么是遗传基因组学

所谓遗传基因组学, 就是在分析分离群体每个个体的基因型、构建分子标记连锁图谱的基础上, 采用 DNA 微阵列技术分析该群体每个个体的全基因组表达谱, 将每个基因的表达水平作为数量性状, 采用 QTL 分析定位控制该基因的表达 QTL (eQTL). 遗传基因组学不仅可以直接应用于自花授粉植物如拟南芥、水稻、玉米等的分离群体, 而且也能应用于实验室动物、家畜和人等的分离家系.

*浙江大学环境修复与生态健康教育部重点实验室开放基金资助项目(050101), 浙江省教育厅科研基金项目资助项目.

** 通讯联系人. Tel: 0571-88492362, Fax: 0571-86919055

E-mail: liuxiuyan1970@yahoo.com.cn

收稿日期: 2006-05-08, 接受日期: 2006-06-02

eQTL可以分为顺式作用和反式作用. 顺式作用就是某个基因的 eQTL 定位到该基因所在的基因组区域, 表明可能是该基因本身的差别引起 mRNA 水平的差别; 反式作用就是某个基因的 eQTL 定位到其他基因组区域, 表明其他基因的差别控制该基因 mRNA 水平的差异^[1].

1.1 酵母的遗传基因组学

2002 年, Brem 等^[2]首次以酵母为材料证明遗传基因组学的这个研究策略是可行的. 他们采用寡核苷酸微阵列, 鉴定由实验室酵母菌株和野生酵母菌株杂交产生的 40 个单倍体分离后代的基因型, 构建了一个含 3 312 个分子标记的图谱, 覆盖 99% 的基因组. 同时采用包含 6 215 个可读框的 cDNA 微阵列, 对亲本和分离后代进行表达谱分析, 通过 Wilcoxon-Mann-Whitney 检验建立基因表达和分子标记之间的关联 (即 eQTL), 570 个基因的表达水平至少检测到一个 eQTL, 在亲本之间有差异表达的 1 528 个基因中, 有 308 个基因能检测到 eQTL, 在双亲之间没有差异表达的基因中, 262 个基因检测到 eQTL, 在 570 个 eQTL 中, 185 个 (32%) 为顺式作用 eQTL, 385 (68%) 个为反式作用 eQTL, 他们还发现 8 个 eQTL 聚集的区域 (eQTL 热点), 分别控制功能相关的基因.

Yvert 等^[3]采用扩大到 86 个单倍体分离后代群体, 不仅证实了 Brem 等^[2]的结果, 而且还检测到更多的 eQTL, 共有 2 294 个基因检测到 eQTL, 其中 578 (25%) 个基因检测到顺式 eQTL, 共检测到 13 个 eQTL 热点, 这些热点区域控制 1 265 个基因的表达, 许多功能相似的基因定位在同一 eQTL 上. 值得注意的是, 大部分反式 eQTL 并没有定位在转录因子上, 反式 eQTL 位点也不富集于某类分子功能或生物学过程的基因中.

1.2 鼠类的遗传基因组学

Schadt 等^[4]用小鼠标准自交系 C57BL/6J 和 DBA/2J 的 111 个 F2 个体的肝组织, 采用微卫星分子标记构建分子标记连锁图谱. 用小鼠基因的寡核苷酸微阵列检测 23 574 个基因的表达, 采用区间作图法进行 QTL 定位. 共检测到 4 339 个 eQTL, 大约 34% 的 eQTL 为顺式作用, 在第 2、6、7、9、10、16 和 17 个染色体的某些区域检测到 eQTL 热点.

Bystrykh 等^[5]从 DBA/2(D2)和 C57BL/6(B6)的 30 个重组自交系, 构建了一个含有 779 个分子标记连锁图谱, 用流式细胞分离仪分离出每个重组自

交系的造血干细胞, 用 Affymetrix U74Av2 寡核苷酸微阵列分析其表达谱, 共检测到 162 个顺式作用 eQTL 和 136 个反式作用的 eQTL.

Chesler 等^[6]采用与 Bystrykh 等相同重组自交系群体的前脑组织, 检测到 101 个 eQTL ($P < 0.05$), 在 88 个最显著的 eQTL 中, 83 个是顺式 eQTL, 同时他们还检测到 7 个 eQTL 热点.

Hubner 等^[7]采用 30 个大鼠重组自交系, 在脂肪组织检测到 2 118 个 eQTL, 在肾脏组织中检测到 2 490 个 eQTL, 并在两种组织中分别检测到 1 个 eQTL 热点.

另外, Vazquez-Chona 等^[8]分析了视网膜受伤后的小鼠重组自交系的视网膜、大脑和脊髓组织的基因表达谱, 对视网膜受伤后急性反应基因进行遗传基因组学研究. Lum 等^[9]采用 1 200 个 SNP 标记分析了小鼠的 300 个 F2 个体的基因型, 同时测量了脑组织的基因表达谱, 对脑组织的转录调控网络进行遗传基因组学分析.

1.3 人的遗传基因组学

Schadt 等^[4]分析了 Centre d'Etude du Polymorphisme Humain (CEPH)的 4 个家族共 56 个人的淋巴细胞系基因表达谱, 发现 29% 差异表达基因是可遗传的, 但由于样本太少, 不能进行 eQTL 分析.

Morley 等^[10]分析了来自 CEPH 的 14 个大家族免疫 B 细胞的表达谱, 其中 3 554 个基因有差异表达, 这些基因中大约 1 000 个基因能检测到 eQTL, 在检测到最显著 eQTL 的 142 个基因中, 27 (19%) 个基因检测到一个顺式 eQTL, 110 (77.5%) 个基因检测到一个反式 eQTL, 5 (3.5%) 个基因检测到多个 eQTL; 在检测到的 eQTL 中, 反式 eQTL 占多数, 许多基因有多个 eQTL, 有 2 个 eQTL 热点, 受 eQTL 热点区域调控的基因在功能上具有相关性.

Monks 等^[11]分析了 15 个 CEPH 家族淋巴细胞的表达谱, 只有 2 340 个基因在淋巴细胞中表达, 其中 31% 表达基因的转录水平是可遗传的. 通过 QTL 分析, 有 33 个基因能检测到 eQTL, 其中 13 个基因检测到顺式 eQTL.

1.4 植物的遗传基因组学

Schadt 等^[4]采用硬秆型玉米和兰客斯特型玉米杂交的 76 个 F3 株系的叶组织, 共检测到 7 322 个 eQTL, 控制 6 481 个基因的表达.

Kirst 等^[12]并采用 2 608 个基因的 cDNA 微阵

列, 分析了 91 个桉树种间回交群体木质部的表达谱, 检测到一些木质素相关基因的 eQTL, 这些 eQTL 与直径生长速度的 QTL 在同一区域, 其中 S-腺苷甲硫氨酸合酶基因检测到顺式 eQTL.

Vuytsteke 等^[13]采用 cDNA-AFLP 分析了拟南芥 50 个重组自交系的 912 个基因的表达丰度. 一共检测到 198 个 eQTL, 包括顺式 eQTL 和反式 eQTL, 同时也检测到 eQTL 热点.

2 基因表达的遗传力

基因表达水平在不同基因型之间存在显著的差异^[14]. 目前遗传基因组学研究结果表明, 基因的转录水平的确是可遗传的. 在酵母中, 9.2% 的基因能检测到 eQTL, 而 20.2% 的双亲差异表达基因能检测到 eQTL, 具有最显著 QTL 的 1 083 个基因的中位遗传力为 27%^[2, 15]. 在小鼠中, 基因表达丰度的遗传力中值为 11%, 基因表达的最高遗传力达 78%, 有 608 个基因表达的遗传力超过 33%^[6]. 在人的淋巴细胞系中, 31% 的差别表达基因可遗传, 遗传力中值为 34%^[11]. 在拟南芥中, 82% 的 792 个差异表达基因是可遗传的, 其遗传力从 11% 到 93% 中值为 30%^[13].

然而, 从 eQTL 的检测结果来看, 基因表达的遗传是非常复杂的. 在酵母中, 40% 的高度遗传转录子检测不到 eQTL, 只有 3% 的高度遗传转录子符合单基因遗传, 17%~18% 的转录子由 1 个或 2 个位点控制, 大约 50% 的转录子由 5 个以上的位点控制, 一半以上的基因表达强度呈超亲分离, 17% 的高度遗传转录子受互作的位点调控^[15], 67% 的互作位点的加性效应不能被检测到^[16]. 在小鼠中, 40% 能检测 eQTL 的基因受多个 eQTL 调控, 其中 4% 的基因有 3 个以上的 eQTL^[4]. 在玉米中, 还检测到一种非常特殊的互作, 2 个表达丰度不相关的基因分别受不同的 eQTL 调控, 在特定 eQTL 等位基因组合中, 这 2 个基因的表达具有相关性^[4].

3 顺式作用 eQTL 和反式作用 eQTL

基因的表达丰度不仅受该基因本身的启动子和其他调控元件的控制(顺式作用), 而且还受参与转录、mRNA 剪接和降解以及复杂的分子级联放大、反馈环和大规模的网络等基因的调控(反式作用)^[6]. 在遗传基因组学研究中检测到的 eQTL, 相应地分为顺式作用 eQTL 和反式作用 eQTL.

顺式 eQTL 比反式 eQTL 具有更高的显著

性^[2, 4, 6, 16]. 虽然在几乎所有检测到顺式 eQTL 的基因都含有调控元件的多态性^[5], 并且有报道说基因中的 SNP 不显著影响转录丰度的测量^[17], 但这并不能完全排除由于 SNP 等多态性产生错配所导致的假顺式 eQTL. 另外, eQTL 和其调控的基因有可能位于同一染色体区域, 以及基因通过反式反馈环调控自身的表达水平等, 都会产生假顺式 eQTL^[18]. 因此, 有必要采用等位基因特异的定量 RT-PCR 对顺式 eQTL 进行鉴定^[10]. Doss 等^[19]通过测量等位基因的相对表达水平, 发现有 64% 的小鼠顺式作用 eQTL 是真实的. 另外他们发现大于 96% 的顺式 eQTL 发生在双亲间具有较多 SNP 的区域. Ronald 等^[18]发现, 52%~78% 的极显著顺式 eQTL 能得到确认, 全基因组序列分析表明, 这些基因的启动子区域的多态性明显增多, 然而并不是所有的多态性都位于转录因子结合位点, 这些基因的 3' 非翻译区的多态性也增多.

在大多数研究中, 都能检测到更多的反式作用 eQTL^[4, 6]. 反式 eQTL 不仅没有局限于转录因子, 也不集中在参与特定的分子功能或生物学过程的基因中^[2]. 几乎所有的遗传基因组学研究结果都检测到聚集了大量的 eQTL 基因组区域, 推测这些 eQTL 热点区域拥有控制大量基因表达的主“调节者”. Chesler 等^[6]的结果表明, 许多定位在同一 eQTL 热点的“靶”基因都是转录因子, 虽然这些“靶”基因的上游“调节者”尚不明确.

永久性的重组自交系群体不仅可以提供高程度的技术重复, 而且将多种组织的数据整合在一起, 进行组织水平表型的比较. 在小鼠的脑组织和血液之间, 297 个基因的 QTL 是重叠的, 基本上都是顺式-QTL^[5, 6]. Hubner 等^[7]发现, 大约 15% eQTL 在大鼠的肾和脂肪组织是共同的, 主要是顺式作用位点. 这些结果表明反式作用 eQTL 好象具有组织特异性.

4 候选基因和调控网络

传统的微阵列数据、蛋白质相互作用数据以及基因注解信息等已经用来构建基因调控网络, 遗传基因组学研究获得的 eQTL 信息将为调控网络的构建提供有力帮助. 顺式 eQTL 能直接提供候选基因的信息, 反式 eQTL 的信息不仅能与其他方法获得的关联网络相结合, 而且还提供了一个减少调节网络的候选节点的强有力工具. 将基因表达性状和表型性状 QTL 分析结果进行综合分析, 并结合相关

分析等统计工具, 能初步构建基因的调控网络。

酵母中, Brem 等^[2]分析了 6 个 eQTL 的热点候选基因, 并初步确定了转录因子 Hap1 是其中一个 eQTL 的热点主调节基因。Yvert 等^[3]根据热点局部的顺式 eQTL 确定了 GPA1 和 AMN1 这两个主调节基因, 并初步验证了这两个基因分别影响信息素反应和子细胞的分离相关基因的转录水平。Bing 等^[20]根据反式 eQTL 的候选基因和该 eQTL 所调控基因的相关性, 将 eQTL 候选基因和该 eQTL 控制的基因相关联; 他们分析了 Brem 等^[2]的数据, 获得了 768 个调节关联, 并构建了一些初步的网络模型。

在小鼠中, Zhu 等^[21], Schadt 等^[22]根据 Schadt 等^[4]的 eQTL 结果, 构建了多个基因调控网络, 同时将 eQTL 和临床 QTL 进行对比分析, 推断出 4 个肥胖相关候选基因。Ghazalpour 等^[23]根据 Schadt 等^[4]的数据, 鉴定了 13 个与肥胖相关的基因调控途径, 发现位于染色体 3、6、16、19 的 4 个基因组区域参与这些途径的调控, 并发现了可能与这些途径相关的新基因。Bystrykh 等^[5]根据 eQTL 结果, 在控制干细胞繁殖 QTL 的区域中检测到 8 个顺式作用的 eQTL, 并分析了它们的“靶”基因, 为确定控制干细胞繁殖的候选基因提供信息。Vazquez-Chona 等^[8]发现, DNA 结合抑制剂 II 是调控视网膜伤害急性期基因的候选基因。Lum 等^[9]发现顺式调控的脑垂体肿瘤转移 I 基因调节是主“调节者”, 他们还建立了代谢疾病相关的大脑中的转录调节网络。Li 等^[24]根据 Chesler 等^[6]的数据, 在 209 个反式 QTL 的基础上, 构建了 66 个 QTL 构成的候选调控网络。

5 展 望

近几年遗传基因组学的飞速发展, 使人们对复杂生物学性状的分子机理和调控网络有了更深入的了解。在酵母中已建立多个基因调控网络模型, 部分网络模型已通过实验验证, 鼠类和人类的研究结果对揭示肥胖、糖尿病、癌症等疾病的分子机理提供有力的帮助。

目前遗传基因组学研究尚处于初步发展阶段, 还具有很大完善空间。可以相信, 随着实验设计、QTL 分析、候选基因分析和调控网络构建等新方法的不断发展, 蛋白质组、代谢物组等的检测手段的进一步完善, 遗传基因组学的内涵将从表达 QTL 扩展到蛋白质 QTL 以及代谢物 QTL。结合传

统微阵列数据, 基因功能信息, 遗传基因组学将在复杂生物学性状的研究以及构建代谢和发育途径的基因调控网络中发挥更大的作用。

参 考 文 献

- 1 Jansen R C, Nap J P. Genetical genomics: the added value from segregation. *Trends Genet*, 2001, **17** (7): 388~391
- 2 Brem R B, Yvert G, Clinton R, *et al.* Genetic dissection of transcriptional regulation in budding yeast. *Science*, 2002, **296** (5568): 752~755
- 3 Yvert G, Brem R B, Whittle J, *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet*, 2003, **35** (1): 57~64
- 4 Schadt E E, Monks S A, Drake T A, *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature*, 2003, **422** (6929): 297~302
- 5 Bystrykh L, Weersing E, Dontje B, *et al.* Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat Genet*, 2005, **37** (3): 225~232
- 6 Chesler E J, Lu L, Shou S, *et al.* Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet*, 2005, **37**(3): 233~242
- 7 Hubner N, Wallace C A, Zimdahl H, *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*, 2005, **37** (3): 243~253
- 8 Vazquez-Chona F R, Khan A N, Chan C K, *et al.* Genetic networks controlling retinal injury. *Mol Vis*, 2005, **11**: 958~970
- 9 Lum P Y, Chen Y, Zhu J, *et al.* Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J Neurochem*, 2006, **97** (Suppl 1): 50~62
- 10 Morley M, Molony C M, Weber T M, *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature*, 2004, **430** (7001): 743~747
- 11 Monks S A, Leonardson A, Zhu H, *et al.* Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet*, 2004, **75**(6): 1094~1105
- 12 Kirst M, Myburg A A, De Leon J P, *et al.* Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol*, 2004, **135** (4): 2368~2378
- 13 Vuylsteke M, Daele H, Vercauteren A, *et al.* Genetic dissection of transcriptional regulation by cDNA-AFLP. *Plant J*, 2006, **45**(3): 439~446
- 14 Cheung V G, Conlin L K, Weber T M, *et al.* Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet*, 2003, **33** (3): 422~425
- 15 Brem R B, Kruglyak L. The landscape of genetic complexity across 5 700 gene expression traits in yeast. *Proc Natl Acad Sci USA*, 2005, **102** (5): 1572~1577
- 16 Brem R B, Storey J D, Whittle J, *et al.* Genetic interactions between

- polymorphisms that affect gene expression in yeast. *Nature*, 2005, **436** (7051): 701~703
- 17 Hughes T R, Mao M, Jones A R, *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol*, 2001, **19** (4): 342~347
- 18 Ronald J, Brem R B, Whittle J, *et al.* Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS. Genet*, 2005, **1** (2): 213~222
- 19 Doss S, Schadt E E, Drake T A, *et al.* *Cis*-acting expression quantitative trait loci in mice. *Genome Res*, 2005, **15** (5): 681~691
- 20 Bing N, Hoeschele I. Genetical genomics analysis of a yeast segregant population for transcription network inference. *Genetics*, 2005, **170** (2): 533~542
- 21 Zhu J, Lum P Y, Lamb J, *et al.* An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet Genome Res*, 2004, **105** (2~4): 363~374
- 22 Schadt E E, Lamb J, Yang X, *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet*, 2005, **37** (7): 710~717
- 23 Ghazalpour A, Doss S, Sheth S S, *et al.* Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol*, 2005, **6** (7): R59
- 24 Li H, Lu L, Manly K F, *et al.* Inferring gene transcriptional modulatory relations: a genetical genomics approach. *Hum Mol Genet*, 2005, **14** (9): 1119~1125

Advance in Genetical Genomics*

LIU Xiu-Yan^{1,2)**}, XIE Zheng-Miao^{1,2)}, CHEN Hui-Zhe³⁾

⁽¹⁾Hangzhou Dianzi University, Hangzhou 310018, China;

⁽²⁾Ministry of Education Key Laboratory of Environmental Remediation and Ecological Health, College of Natural Resources and Environmental Science, Zhejiang University, Hangzhou 310029, China;

⁽³⁾China National Rice Research Institute, Hangzhou 310006, China)

Abstract Genetical genomics is combining of microarray technology and quantitative trait loci (QTL) analysis, mapping expression QTL (eQTL) in global level. It provides a novel procedure for revealing the molecular mechanism and regulatory network of complex trait. The conception and strategy of genetical genomics were brought forward by Janson and Nap in 2001. By now, genetical genomics has been applied on yeast, mouse, human and plant such as maize. These results indicated that the variation of expression level of genes is heritable complex trait. Expression QTLs are classified into *cis*-acting and *trans*-acting. *Cis*-acting eQTL is located on the same genomic region of the regulated gene, meaning the variation of mRNA level could be due to the polymorphism in the gene itself. *Trans*-acting eQTL is located on other genomic region, meaning other gene controls the variation of mRNA level. The integration of eQTL results, gene function annotation and statistic analysis will not only more precisely identify candidate genes controlling complex traits and the expression of related genes, but also construct regulatory networks for these complex traits.

Key words microarray, expression QTL (eQTL), regulatory network

*This work was supported by grants from Ministry of Education Key Laboratory of Environmental Remediation and Ecological Health, Zhejiang University (05010), and Science Research Foundation of Department of Education of Zhejiang Province.

**Corresponding author. Tel: 86-571-88492362, Fax: 86-571-86919055, E-mail: liuxiuyan1970@yahoo.com.cn

Received: May 8, 2006 Accepted: June 2, 2006