

实时定量 PCR 的数据归一化方法

吴云鼎 综述 解保生, 丁仲鹃 审校
(昆明医科大学附属口腔医院, 云南 昆明 650031)

[摘要] 在生命科学研究中, 随着实时定量 PCR (Real-time PCR) 成为高通量快速、定量检测 mRNA、micRNA 等的常用技术, 选择合适的数据归一化处理基准, 对提高结果分析的准确性尤为重要. 新近研究显示应根据不同研究目的或研究对象选用合适的 Real-time PCR 数据归一化方法, 如单个内参基因归一化、多个内参基因归一化、micRNA 归一化以及体外人工合成基因归一化等. 就近年实时定量 PCR 的数据归一化方法进展做一综述.

[关键词] 实时定量 PCR; 归一化; 管家基因; 内参基因

[中图分类号] Q789 **[文献标识码]** A **[文章编号]** 1003 - 4706 (2013) 03 - 0160 - 05

Normalization Method in Real-time PCR Data

WU Yun - ding, XIE Bao - sheng, DING Zhong - juan

(The Affiliated Hospital of Stomatology of Kunming Medical University, Kunming Yunnan 650031, China)

[Abstract] In biological research, gene analysis with real-time PCR is a routine tool becoming increasingly important for high throughput and accurate profiling of mRNA, micRNA, etc. In real-time PCR, a right normalization count is particularly important for improving the accuracy of data analysis and results. In recent studies, appropriate normalization reference in real-time PCR should be selected according to different objectives, such as single internal control gene normalization, multiple internal control genes normalization, microRNA normalization and artificial molecular normalization and so on. In this paper, an overview of the progress of normalization in real-time PCR was presented.

[Key words] Real-time PCR; Normalization; Housekeeping genes; Internal control gene

Real-time PCR^[1]是基于普通 PCR 的衍生技术, 通过连续监测 PCR 指数扩增期间的荧光信号强弱的变化来测定特异性产物的量. 1992 年 Higuchi 和其同事提出这一理念, 并在 1996 年由美国应用生物系统公司 (Applied Biosystems) 推出这一技术, 为生命科学研究提供了一个强有力的工具. Real-time PCR 具有快速、高效、灵敏、特异性强的特点, 同时封闭的反应环境有效地解决了 PCR 污染问题. 在 real-time PCR 中, 准确定量重要前提是合适的数据归一化处理基准^[2]. 多数基因表达研究是在比对 2 个或者多个基因的基础上进行的, 为了使各个基因的表达结果具有可比性, 这就要求目标基因样本起始量一致. 而在样本制备的过

程中, 许多因素都会造成样本之间的差异, 如样本提取的质量和数量, 药物使用情况, 实验条件影响以及逆转录的效率等. 数据归一化处理即是消除这些差异的过程^[3]. 鉴于合理的归一化是正确测量基因 mRNA 表达极为重要的前提, 制定更有效的归一化策略也是 real-time PCR 技术应用中的难点, 现就 real-time PCR 中数据归一化方法做一综述.

1 归一化的方法

1.1 单个内参基因归一化 (single internal control gene for normalization)

管家基因 (housekeeping gene), 又称看家基

[基金项目] 国家自然科学基金资助项目 (30960424)

[作者简介] 吴云鼎 (1987~), 男, 白族, 云南大理市人, 在读硕士研究生, 主要从事口腔修复学临床及研究工作.

[通讯作者] 解保生. E-mail: bobxie2003@yahoo.com

因, 是一类维持细胞正常生长、增殖所必须的基因, 对维持细胞的正常生理功能有重要作用^[4]. 如微管蛋白基因、糖酵解酶系基因与核糖体蛋白基因等. 以往的研究认为, 管家基因的表达水平是恒定的, 不受各种条件的影响, 故在 real-time PCR 实验中, 通常选择单个的管家基因进行数据的归一化处理, 消除样本与样本间的起始差异, 以使样本之间具有可比性. 然而, 管家基因的表达水平并非恒定不变^[5]. 例如常用的 2 个人类管家基因 β 肌动蛋白 (β -actin) 和甘油醛-3-磷酸脱氢酶 (GAPDH 或 G3PDH), 在不同组织和不同条件下 mRNA 表达水平存在差异. Glare E M 等在一次横向研究中, 对 92 例哮喘患者和 26 例正常对照患者支气管肺泡灌洗液和支气管内组织切片标本行竞争性 RT-PCR 定量, 测定标本中编码 β -actin、GAPDH 和白细胞介素 II 的 mRNA 量. 结果显示: 不使用吸入性皮质类固醇类患者, 其支气管肺泡灌洗液和支气管组织中 GAPDH 的表达量低于正常对照组 10 倍, 而肺泡灌洗液中 β -actin 同样低于正常对照组 10 倍^[6]. 与此同时, 研究还发现, 不仅是在肺组织中, 在人体其他组织中, β -actin 和 GAPDH 的表达水平也存在着较大差异. Ryutaro Mori 等对 15 种人体癌变组织及其周边正常组织的研究结果表明: GAPDH 和 β -actin 在癌变组织中的表达都高于正常组织^[7]; Sikand 等的研究显示 GAPDH 和 β -actin 的表达水平同时也受到 miR-644a 的影响和调控, 与阴性对照比较, miR-644a 使 GAPDH 和 β -actin 表达水平下降 50%到 90%^[8].

文献综合分析表明: GAPDH 无论在体外还是体内, 其 mRNA 的水平会随着环境改变而进行调节. 且最近对内参基因的研究指出假基因 (Pseudogenes) 的存在也是 GAPDH 和 β -actin 作为内参的一个缺陷, 人的 GAPDH 和 β -actin 分别存在 64 个和 67 个高度同源的假基因, 其 mRNA 的大小也较原基因相近, 难以辨别^[9].

尽管在上述研究中发现一些管家基因并非一直恒定地表达, 且有假基因存在, 但这并不表示它们就失去了成为归一化所需基因的资格. 在一定条件下, 其中不乏表达稳定一致、能够达到实验需求的基因, 如 GAPDH. 在经过免疫球蛋白 IgE, 环孢菌素、地塞米松处理的人体肾脏组织微切片样本当中, 即使在不同实验条件下, GAPDH 也有着稳定一致的表达^[10]. 有研究建议: 在实验中涉及对低氧, 癌细胞株、个体发生、急性胰腺炎、致癌作用的研究时, 应该避免使用 GAPDH 作为归一化基因. 而 β -actin 及与它类似的 γ -actin (表达度相

对 β -actin 低) 在给定的实验条件下展示了良好的稳定性. 同样地, 在涉及处于不同发育时期的组织样本, 或者组织样本形态跨度较大时, 不建议采用 β -actin 来作为归一化基因^[11].

故此, 尽管使用单个的管家基因来进行归一化处理并不完全可靠, 但是在一定条件下, 其可行性仍然存在.

1.2 多个内参基因归一化 (multiple internal control genes for normalization)

Vandesompele 等^[12]推荐使用的一种精确的归一化方法是同时采用多个 (管家) 基因, 在选定其作为内参之前, 以 geNorm 先判断其在实验条件下表达的稳定性 (测定每个基因的 M 值大小, 越小则该基因表达越稳定), 选择表达最稳定的几个基因作为内参, 同时应用在实验中进行归一化, 取其几何平均值作为最终归一化的基准. 当然, 选择的内参基因个数取决于多个条件, 实际选取的个数是在实际条件和实验精确性之间进行权衡的. 如果目标基因数较少, 或者只有很少量的 RNA 存在时, 那么就没有必要选用多个 (如 7~8 个) 内参基因. 此外, 如果目标基因的表达相对稳定, 或者采用更多的内参基因对于归一化精确性的提高没有明显的帮助, 此时使用除必要以外的内参基因就属多余, 对有限的实验室资源而言是一种浪费.

对于 RT-PCR 而言, 文中推荐使用的内参基因数最少为 3 个, 使用的内参基因数可以逐步增加到新增加的内参基因对于增加归一化的精确性没有明显的意义为止, 但最多不超过 9 个^[12]. 确定是否有必要增加一个基因作为内参的方法是使用 geNorm 比较内参基因之间的成对差异 (Pairwise variation), 以数值表示. 这个差异值反映了增加一个基因作为内参对归一化的影响, 二者的差异值越大, 则越有必要选取更多的基因作为内参, 来增加归一化的准确性. 文献中以 0.15 作为一个阈值, 当某一对内参基因间的成对差异值低于 0.15 时, 则可认为增加该基因以后的基因作为内参对于归一化精确性的提高已经不存在明显的意义和作用. 同时使用多个基因进行归一化取其几何平均值的方法是现在广为接受的归一化金标准^[13]. McCulloch R S 等通过 geNorm 及其他两种类似软件 (BestKeeper 和 NormFinder) 测定在小型猪关节软骨中 10 个看家基因的表达稳定性, 结果是 3 种软件都得出了相似的结果: 肽基脯氨酰异构酶 A, 黄素蛋白琥珀酸脱氢酶, β 肌动蛋白, 甘油醛-3-磷酸脱氢酶四个基因于该组织中表达最稳定, 并建议在相关的研究中采用以上 4 个内参基因的几何平

均值作为归一化的依据^[14]。S.J Park 等对犬的 13 个大脑样本和 10 种其他组织中的 16 个候选管家基因进行了表达稳定性的测定, 测定过程中同样使用了以上 3 种软件, 为涉及犬脑组织和其他组织内参基因的选定提供了指导^[15]。

1.3 微小 RNA 归一化 (microRNA for normalization)

MicroRNA 是一种非编码 RNA (non-coding RNA), 在细胞分化、细胞周期调控等方面起着重要作用。microRNA 对 RT-PCR 进行归一化, 是使用所有在样本表达的 microRNA 的平均值作为 microRNA RT-PCR 的归一化依据。结果表明该法能更多减少因技术差异导致的误差, 更加准确地反应出基因表达方面的变化^[13]。在该项研究中, 研究者设立 5 个包含不同组织的分组, 其中正常组包含了 8 种正常组织 (脑, 心脏、肝脏、肾脏、肺、结肠、肾上腺、乳腺), 5 组共计 147 个样本。实验测量了每个组中, 每个样本间 miRNA 的表达水平之间的差异, 得出一个差异系数 (coefficient of variation), 并使用这个差异系数作为归一化的依据。测量基因表达数据的误差是生物性误差和实验性误差的集合。差异系数越低, 表明由于实验过程造成的误差就越小。

由于 microRNA 在细胞发育分化等方面的重要调控作用, 相应的 microRNA 表达变化与疾病进程和癌症病变密切相关。Wotschovsky Z 等对肾细胞癌、肾细胞癌周边正常组织和远处骨转移组织中的 4 种 microRNA (miR-28, miR-103, miR-106a, miR-151) 和 2 种小片段 RNA (RNU6B, RNU44) 进行了表达稳定性的测定, 发现 miR-28, miR-103, miR-106a, RNU48 的表达最为稳定^[16], 并推荐单个内参基因归一化时使用 miR-28 作为内参, 若进行多个内参基因归一化, 则使用 miR-28, miR-103 或 miR-28, miR-103, miR-106a 组合, 取其平均值作为归一化的基准。

1.4 体外人工合成基因归一化 (artificial molecular for normalization)

另外一种归一化的方法是在体外合成 RNA, 这种人工合成的 RNA 分子可以克隆后从其他物种转录而得到, 或者人工合成。尔后在提取样本的过程中以一定的浓度加入其中, 这样做的好处有二: 其一是在所有会影响目标 RNA 的实验失误中它会一直存在; 其二是不会因为表达水平浮动的原因对内参基因造成影响 (加入时浓度是已知确定的)。这种方法提供了一种可靠且独立于样本之外的归一化方法^[17]。然而这种方法也并非完美无缺。由于是

人工合成后被导入到所提取样本中的, 与细胞中提取得到的 RNA 和内参基因不同, 在某些情况下可能会产生问题, 如在研究纤维性病变组织样本时。所以在成为归一化基因前, 仍需验证其可行性。而且对于一些只要进行少量定量试验的小型实验室而言, 人工合成基因进行归一化并不一定可行。虽然使用人工合成的 RNA 作为内参的标准已经确定^[18], 但是在人工合成基因更加的商业化和更多深入研究之前, 这仍只是在理论上可行。

2 归一化方法的应用和选择

以上所述的 real-time PCR 归一化的策略与方法, 可以对样本量大小、RNA、DNA (包括基因组 DNA) 进行归一化, 消除样本间或者组与组之间的一些差异, 使得实验数据具有可比性, 符合实验要求。然而正如前所述, 很多研究都表明一些常用的内参基因以及 RNA 的数量质量、转录效率都会发生变化, 造成组间差异, 从而影响实验结果。在 real-time PCR 成为现实可用的技术之前, 早在 1975 年即有研究发现在巨细胞病毒感染中, 18SrRNA 的表达水平有所上升^[19]。1984 年 Piechaczyk et al 发现 GAPDH 在大鼠不同组织中以相近的速率转录时, 其中包含的 mRNA 的量也高低不同^[20]。

在进行归一化时, 我们希望能够得到尽可能精确的结果, 能够将差异消除是最理想的状态。然而, 归一化的精确程度还是要取决于实验要求, 并非一定要完全彻底消除样本或者组间差异^[21]。也就是说, 即使一个用于归一化的内参基因其表达水平不够稳定, 但如果样本组间的差异和不稳定性超过该基因, 那么该基因作为内参的可行性仍然存在^[21]。虽然使用多个内参基因是归一化的金标准, 但由于样本和成本的限制, 同时选定多个基因作为内参并不总是可行。对于一些特殊的测定研究还是取决于所选基因的表达变化程度。下面以尿道上皮癌和小型猪心衰模型为例, 简单叙述归一化内参基因的选定过程。

在尿道上皮癌内参基因的选定研究中, 首先使用基因芯片在 24 个尿道上皮癌样本和 24 个正常样本中检出其中存在的 miRNA 种类, 并选定 8 个在不同组织间绝对表达变化量在 1.2 倍以下的 miRNA 基因作为研究对象。为了避免基因芯片技术人为归一化带来的误差和影响, 以相同标准增加了 8 个基因进行研究。同时将研究的范围扩展到了小片段 RNA 和临床样本 (9 个非恶性组织样本, 25 个恶

性尿道上皮癌样本)。而后分别以 geNorm、NormFinder 和 BestKeeper3 种软件分析所选基因的表达稳定性。最终 3 种软件得出了不同的归一化策略: geNorm 推荐使用多个内参取几何平均值行归一化的同时 (miR-101, miR-125a-5p, miR-148b, miR-151-5p 组合或 miR-125a-5p, miR-148b, miR-151-5p 组合), 也给出了最适合单一基因归一化的内参 (miR-148b); NormFinder 建议使用多个内参行归一化 (miR-125a-5p and Z30 组合或 miR-148b, miR-181b, and miR-874 组合); BestKeeper 则给出了在尿道上皮癌研究中最适合单一基因归一化的 miRNA 基因 (RNU48)^[2]。在此可见 3 种软件得出的结果不尽相同, 这是由 3 种软件不同的运行原理所决定的。然而考虑到不同样本间 RNA 在数量和质量上的差异, 有人认为是 geNorm 在理论上更为可靠, 更适合于涉及共

同调控基因研究的归一化内参选定^[23]。

而在小型猪的心衰研究中, 研究者将 9 只小型猪分为 2 组 (正常组 4 只, 心衰组 5 只), 模型建成后在无菌条件下收集 4 种心脏组织 (左心房、右心房、左心室、右心室)。分别提取 mRNA, 对每种组织中 8 个基因 (ACTB, GAPDH, B2M, YWHAZ, TOP2B, PPIA, TBP, HPRT-1) 进行实时定量 PCR, 得到实时定量 PCR 数据后以 geNorm 分析所选基因的表达稳定性。最终得到的结果是: 在小型猪左心房和右心房组织中, HPRT-1、TBP、PPIA 三个基因的表达是最稳定的; 左心室组织中 GAPDH、HPRT-1、TBP 的表达最稳定, 而右心室组织中则以 PPIA、GAPDH、ACTB 基因的表达最为稳定^[24]。

最后, 对在相应组织内已经确认表达稳定性的内参基因作一归纳 (表 1)。

表 1 相对应组织中表达稳定的内参基因

Fig. 1 The stable expression reference genes in tissues

组织类型	可用内参基因	筛选软件
人体肾组织 (处理后)	GAPDH	
小型猪关节软骨组织	β -actin、GAPDH、PPIA、SDHA	geNorm NormFinder BestKeeper
小型猪心脏组织	左、右心房: HPRT-1、TBP、PPIA; 左心室: GAPDH、HPRT-1、TBP; 右心室 PPIA、GAPDH、ACTB	geNorm
犬脑组织	所有脑组织: RPL13A、RPS19; 前脑和大脑: RPS5、RPS19; 脑干: RPL32、RPS19; 后脑: GAPDH、RPS19; 小脑: MEPS7、RPL13A	geNorm NormFinder BestKeeper
肾细胞癌组织	miR-28、miR-103、miR-106a、RNU48	geNorm NormFinder
尿道上皮癌组织	miR-125a-5p、miR-151-5p、miR-101、miR-148b	geNorm NormFinder BestKeeper

*注: 不同软件分析结果不完全相同, 以上列出的可用内参基因均为 geNorm 分析结果。

3 总结

综上所述, real-time PCR 进行归一化时, 在确定基因作为内参之前, 需要对该基因在实验体系和样本组织中表达水平的稳定性进行评估确认。准确的归一化对实验正确反映变化和数据的准确分析起着至关重要的作用。采用经过验证和未经验证的内参进行归一化后, 二者得到的结果可存在明显的差异, 合理的归一化策略对于 real-time PCR 实验结果有着明确而重要的意义。

[参考文献]

[1] HEID C A. Real time quantitative PCR[J]. Genome Research, 1996, 6(10): 986 - 994.

[2] ANDERSEN C L. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets[J]. Cancer Research, 2004, 64(15): 5 245 - 5 250.

[3] YANG Y H. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation [J]. Nucleic Acids Res, 2002, 30(4): 15.

[4] WARRINGTON J A. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes [J]. Physiol Genomics, 2000, 2 (3): 143 - 147.

[5] FERREIRA, EAND M J, CRONJE. Selection of Suitable Reference Genes for Quantitative Real-Time PCR in Apoptosis-Induced MCF-7 Breast Cancer Cells [J].

- Molecular biotechnology, 2012, 50(2): 121 – 128.
- [6] GLARE E M. Beta-Actin and GAPDH housekeeping gene expression in asthmatic airways is variable and not suitable for normalising mRNA levels [J]. *Thorax*, 2002, 57(9): 765 – 770.
- [7] MORI R. Both beta-actin and GAPDH are useful reference genes for normalization of quantitative RT-PCR in human FFPE tissue samples of prostate cancer [J]. *Prostate*, 2008, 68(14): 1 555 – 1 560.
- [8] SIKAND K. Housekeeping Gene Selection Advisory: Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) and beta-Actin Are Targets of miR-644a [J]. *PLoS One*, 2012, 7(10): 47 510.
- [9] SUN Y. Pseudogenes as weaknesses of ACTB (Actb) and GAPDH (Gapdh) used as reference genes in reverse transcription and polymerase chain reactions [J]. *PLoS One*, 2012, 7(8): 41 659.
- [10] WILLIAMS, CAND J. Coleman, Induced expression of mRNA for IL-5, IL-6, TNF-alpha, MIP-2 and IFN-gamma in immunologically activated rat peritoneal mast cells: inhibition by dexamethasone and cyclosporin A [J]. *Immunology*, 1995, 86(2): 244.
- [11] STURZENBAUM, SRAND P, KILLE. Control genes in quantitative molecular biological techniques: the variability of invariance [J]. *Comp Biochem Physiol B Biochem Mol Biol*, 2001, 130(3): 281 – 289.
- [12] VANDESOMPELE J. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes [J]. *Genome Biol*, 2002, 3(7): 34.
- [13] MESTDAGH P. A novel and universal method for microRNA RT-qPCR data normalization [J]. *Genome Biology*, 2009, 10(6): 64.
- [14] MCCULLOCH R S. Identification of stable normalization genes for quantitative real-time PCR in porcine articular cartilage [J]. *J Anim Sci Biotechnol*, 2012, 3(1): 36.
- [15] PARK S J. Selection of Internal reference genes for normalization of quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis in the canine brain and other organs [J]. *Mol Biotechnol*, 2012, 4(25): 363..
- [16] WOTSCHOFSKY Z. Reference genes for the relative quantification of microRNAs in renal cell carcinomas and their metastases [J]. *Anal Biochem*, 2011, 417(2): 233 – 241.
- [17] SMITH R D. Exogenous reference RNA for normalization of real-time quantitative PCR [J]. *Biotechniques*, 2003, 34(1): 88.
- [18] CRONIN M. Universal RNA reference materials for gene expression [J]. *Clinical chemistry*, 2004, 50(8): 1 464 – 1 471.
- [19] TANAKA S, TFURUKAWA, AND S.A. PLOTKIN, Human cytomegalovirus stimulates host cell RNA synthesis [J]. *Journal of virology*, 1975, 15(2): 297 – 304.
- [20] PIECHACZYK M. Post-transcriptional regulation of glyceraldehyde-3-phosphate-dehydrogenase gene expression in rat tissues [J]. *Nucleic Acids Research*, 1984, 12(18): 6 951 – 6 963.
- [21] HUGGETT J. Real-time RT-PCR normalisation; strategies and considerations [J]. *Genes and Immunity*, 2005, 6(4): 279 – 284.
- [22] RATERT N. Reference miRNAs for miRNAome analysis of urothelial carcinomas [J]. *PLoS One*, 2012, 7(6): 39 309.
- [23] ANDERSEN C L, JENLEN J L, RNTTOFT T F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets [J]. *Cancer Research*, 2004, 64(15): 5 245 – 5 250.
- [24] MARTINO A. Selection of reference genes for normalization of real-time PCR data in minipig heart failure model and evaluation of TNF- α mRNA expression [J]. *Journal of Biotechnology*, 2011, 153(3): 92 – 99.

(2013-02-14 收稿)