

Gene-family profiling: a normalization-free real-time RT-PCR approach with increased physiological resolution

Stian Ellefsen¹ and Kåre-Olav Stensløyken^{2,3}

¹Lillehammer University College, Lillehammer; ²Institute of Clinical Medicine, and ³Department of Molecular Biosciences, University of Oslo, Oslo, Norway

Submitted 16 November 2009; accepted in final form 8 March 2010

Ellefsen S, Stensløyken K. Gene-family profiling: a normalization-free real-time RT-PCR approach with increased physiological resolution. *Physiol Genomics* 42: 1–4, 2010. First published March 9, 2010; doi:10.1152/physiolgenomics.00196.2009.—Here we present gene-family profiling, an approach for improved real-time RT-PCR analyses. It is based on recently published data, and we argue that it brings solutions to two major problems. First, it is normalization-free and therefore unbiased by variation in normalization agents such as reference gene expression. This strengthens data validity and also increases data resolution, reducing coefficients of variation by ~48% in our data sets. Second, it includes all members of a particular gene family, treating individual genes as constituting fractions of collective gene-family expression rather than as unrelated entities. Because different family members typically fulfill similar, but complementary roles, this increases the physiological relevance. Gene-family profiling is particularly useful for evaluation of cellular adaptations to physiological challenges and for comparison of properties between different experimental systems such as species, tissues or tissue regions. In addition, it seems suitable for analyses of inherent patterns of gene expression in singular biological samples. In our opinion, the approach is valuable for both research and diagnostic purposes, and should be readily applicable for many studies of gene expression. Its value is likely to increase as science continues to unravel gene function.

physiology; drug library screening; external control gene; reference gene; housekeeping gene

BACKGROUND

Real-time RT-PCR analyses are dependent on accurate procedures for data normalization, and it is well established that the potential inaccuracy of prevailing strategies could lead to misinterpretations (8, 13, 35). Although significant efforts such as applying external RNA control genes (13) and geometric averaging of internal RNA control genes (36) have been made to minimize this problem, there are still unresolved prospects (13). In addition, real-time RT-PCR analyses are complicated by the difficult interpretation of data, primarily caused by a limited selection of evaluated genes. However, this can easily be resolved by increasing the number of analyzed genes, a notion that is gradually being realized. This will increase the physiological value of data. Unfortunately, multiple-gene analyses have so far been limited to investigations of genes that share gene ontology or pathway association (26), or for verification of microarray data (27). Typically, such studies do not assess genes that exert complementary physiological roles but, rather, assess genes with different functions. As a consequence

such analyses still address singular genes, only on a larger scale. Instead, the expression of singular genes should be put into context of complementary genes. This will enable discussion of how particular physiological tasks are performed and will allow more adequate analyses of cell physiology. The physiological insight gained by such analyses will increase as insight of gene function increases.

From a physiological point of view, gene families are ideal targets for multigene analyses of gene expression. They consist of a set of genes that share baseline properties and fulfill similar physiological roles. However, there are important differences that give them slightly diverging and often complementary roles (17). These differences provide a genetic toolkit for developmental and spatial differentiation of particular physiological tasks (17). This concept applies to gene families that contain genes that function as singular units, such as voltage-gated ion channels and neurotransmitter transporters (3, 4, 9, 29), as well as to families that contain genes that function as subunits of multipolypeptide complexes, such as neurotransmitter receptors like *N*-methyl-D-aspartic acid receptors (NMDARs) (24) and γ -aminobutyric acid_A receptors (GABA_ARs) (14), motor proteins like myosin (32), and protein kinases like AMP-activated protein kinase (AMPK) (19).

Gene family-based analysis will permit comparison of the role of single genes to the roles of other family members. This enables analysis of inherent physiological properties within and between experimental systems. We recently utilized the approach in studies of NMDAR and GABA_AR function and regulation in a teleost fish (11, 12) [commented on by Prentice (30)]. In this study, we further explore the method and give it the name gene-family profiling.

GENE-FAMILY PROFILING: THE APPROACH

Gene-family profiling allows evaluation of the expression of singular genes by considering them as constituting fractions of overall gene-family mRNA abundance. Gene expression of singular genes is calculated using Eq. 1:

$$\frac{100(\text{Gene}_x^{\text{E}C_p})^{-1}}{(\text{Gene}_1^{\text{E}C_p})^{-1} + (\text{Gene}_2^{\text{E}C_p})^{-1} + \dots + (\text{Gene}_n^{\text{E}C_p})^{-1}} = \text{Gene}_x \text{ in percentage of total gene-family expression (1)}$$

E = reaction efficiency and Cp = crossing point. In our recent studies (11, 12), E values were calculated separately for each real-time RT-PCR reaction using LinRegPCR software (31). In the final calculations, average reaction efficiencies were used for each transcript. Cp values were obtained using the second derivative maximum method (Lightcycler 2.0 software). In the figures, all gene-family members are presented together, appearing as a collective gene-family expression of 100% (see

Address for reprint requests and other correspondence: S. Ellefsen, Lillehammer Univ. College, N-2618 Lillehammer, Norway (e-mail: stian.ellefsen@hil.no).

Figs. 1–3). Gene expression data for single genes are evaluated as percentage of total gene-family expression. This provides data that will not fluctuate together with changes in reverse transcription efficiency, contrasting other normalization-free strategies such as absolute quantification. To obtain accurate gene-family profiles, particular care must be taken to avoid artifacts caused by differences in real-time RT-PCR performance (28). For example, every gene and every experimental group must be included in every real-time RT-PCR run.

GENE-FAMILY PROFILING: AN EVALUATION

Compared with reference gene-based normalization of single gene real-time RT-PCR data, gene-family profiling offers several advantages. It allows easily applicable normalization-free and accurate analyses of gene expression. It permits a more thorough discussion of physiology, giving way for evaluation of inherent patterns of expression in biological samples such as tissue regions, organs, and species, and evaluation of adaptations to physiological or pathophysiological challenges.

In general, the physiological accuracy of gene-family profiling depends on a close correlation between polypeptide and mRNA abundance. Importantly, this premise has been suggested to be met by genes acting as subunits of multipolypeptide complexes (18, 22). The alternative to targeting mRNA is to target polypeptides directly. However, at the present time gene-family profiles cannot be obtained through protein quantification because of a lack of methods with sufficient specificity and sensitivity.

Review of Current Data Sets

Gene-family profiling provides increased sensitivity and accuracy to real-time RT-PCR data compared with reference gene normalized data. In Ellefsen et al. (11, 12), this was evident from a 39, 40, and 64% reduction in coefficients of variation of the expression of NMDAR subunits, α -amino-3-hydroxyl-5-methyl-4-isoxazole-propionate receptor (AMPA) subunits, and GABA_AR subunits in crucian carp brain. Gene-family profiles provided coefficients of variation as low as 16, 11, and 10%, respectively. Correspondingly, we recently found a 39 and 49% reduction in coefficients of variation of the expression of NMDAR and AMPAR subunits in turtle brains (Couturier C, Ellefsen S, Stensl okken K-O, Nilsson GE, unpublished observations). Interestingly, by defining total gene-family mRNA expression to 100%, gene-family profiles eliminate covariation of subunit expression as a source of technical variation. It basically corrects for the fluctuating nature of gene expression (16). This may explain the reduced coefficients of variation. Given the likely scenario that the function of multipolypeptide complexes is regulated by the stoichiometric availability of subunits (18), this will provide more accurate data analyses, resulting in increased physiological resolution. Notably, in the described studies, data normalization was achieved using an external RNA control gene, which means that the observed variation in target gene expression cannot be ascribed to variation in the expression of internal reference genes (13).

The suitability of the gene-family profiling approach is supported by analyses of physiological aspects of available data sets (11, 12). In these studies, the approach was utilized to study adaptations and responses to anoxia (oxygen depletion) in an anoxia-tolerant teleost, the crucian carp (*Carassius car-*

assius). The data support current knowledge of anoxic survival. First, the NR2 subfamily of NMDARs was dominated by NR2B and NR2D subunit (11) (Fig. 1), a subunit composition that has been ascribed an essential role in the hypoxia tolerance of neonatal rats (2). Second, GABA_AR family expression was dominated by subunits that provide tonic inhibition of neuronal activity (12) (Fig. 2), supporting a vital role for such inhibition in the anoxic fish (20). In addition, when the approach was utilized to analyze AMPK α -subunit data (α_1 - α_2) from Stensl okken et al. (33), the brain showed higher expression of α_1 than α_2 , whereas the heart was predominated by α_2 (Fig. 3), being similar to the tissue-specific α -subunit composition reported for mammals (1, 6).

Muscle Tissue as an Area of Utility

Gene-family profiling should be particularly well suited for analyses of physiologically adaptable tissues such as skeletal and cardiac muscle. In these tissues, the approach has a range of applications, such as 1) comparison of innate patterns of mRNA abundance shown by different muscles or between muscles of different species, 2) disclosure of changes that occur in developing myocytes, and 3) disclosure of changes that occur in response to physiological interventions.

The myosin heavy chain (MyHC) gene family stands out as a suited case for gene-family profiling. The abundance of MyHC polypeptide and mRNA shows close correlation in both skeletal and cardiac muscle (21, 25, 37), and changes in MyHC composition are seen at both polypeptide and mRNA levels (21, 23, 34), an observation that is underlined by the transcriptional control of MyHC during development (5). Overall, gene-family profiling can arguably be more appropriate for measurement of MyHC content than the traditionally applied immunohistochemistry, as real-time RT-PCR provides volumetric (3D) analysis that takes into account hybrid fibers and longitudinal variation. Indeed, Eizema et al. (10) showed that this is the case in tissues containing high degrees of hybrid fibers.

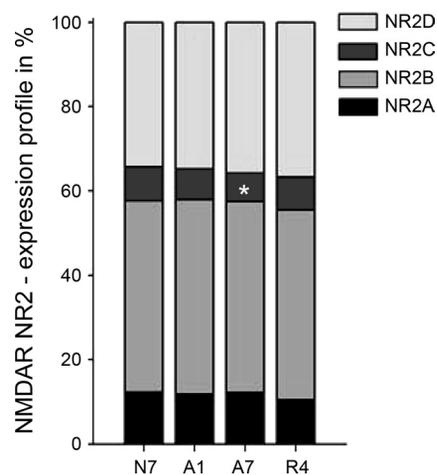


Fig. 1. Expression profile for *N*-methyl-D-aspartic acid receptors (NMDAR) subunits NR2A-D in brain of crucian carp exposed to normoxia, anoxia, and reoxygenation. N7, normoxia 7 days ($n = 11$); A1, anoxia 1 day ($n = 11$); A7, anoxia 7 days ($n = 11$); R4, anoxia 7 days followed by reoxygenation 4 days ($n = 12$). * $P < 0.05$ (1-way ANOVA followed by Tukey Kramer's posttest). Values are means. Modified from Ellefsen et al. 2008 (11).

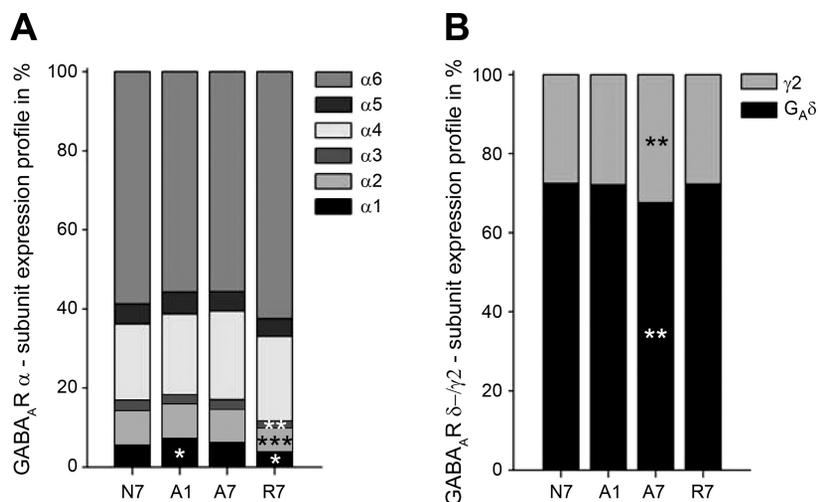


Fig. 2. Expression profile for γ -aminobutyric acid_A receptor (GABA_AR) α - (A) and δ -/ γ 2-subunits (B) in brain of crucian carp exposed to normoxia, anoxia, and reoxygenation. α 4-, α 6-, and δ -Subunits are known to exert tonic, extrasynaptic GABAergic activity. N7, normoxia 7 days ($n = 6$); A1, anoxia 1 day ($n = 6$); A7, anoxia 7 days ($n = 6$); R7, anoxia 7 days followed by reoxygenation 7 days ($n = 6$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (1-way ANOVA followed by Tukey Kramer's posttest). Values are means. Modified from Ellefsen et al. 2009 (12).

Criteria and Limitations

For gene-family profiles to allow proper physiological analyses, they must include all relevant gene-family members. Failure to do so will lead to overestimation of the expression of included genes. However, sound judgment is important. For example, in adult mammalian skeletal muscle, only four out of eight MyHC genes are expressed, ruling out the need for assessing all eight variants. Furthermore, the required inclusion of all gene family members is a particular challenge in studies of nonmodel organisms. As an example, fish species share at least one genome duplication event, which for most genes has resulted in the coexistence of multiple paralogs. This put forth a demand for gene screening. Because gene family profiling relies on concurrent analyses of multiple genes, its area of application is naturally restricted to gene families that consist of more than one gene. The potential problem of splice variants is overcome by designing primer sets that detect all variants.

For data sets that show substantial changes in gene-family profiles, it may be necessary to keep track of quantitative changes in gene expression (11, 12). This can be achieved by

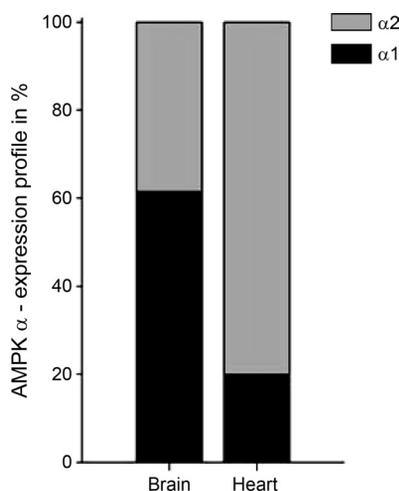


Fig. 3. Expression profile for AMP-activated protein kinase (AMPK) α -subunits in brain and heart of normoxic crucian carp. Values are means ($n = 6$). Data from Stensl kken et al. 2008 (33).

simultaneous presentation of data normalized to an external RNA control gene (13). This is particularly important when the expression of gene family members are spatially dispersed, functioning as distinct physiological units. Changes in the activity of such genes/proteins cannot be compensated for by other family members, and alterations in their expression arguably represent loss-of-physiological function or gain-of-physiological function events. As an example, in Ellefsen et al. (12), we found decreased expression of extrasynaptic GABA transporters (GATs) but not of synaptic GATs in the anoxic crucian carp brain. From a gene-family perspective, extrasynaptic GATs would seem to be replaced by synaptic GATs, a scenario that is physiologically unlikely. Nevertheless, also for GABA transporters, gene-family profiling remained a valuable tool for evaluation of function (12). Similarly, in cases of shut down of physiological processes, such as during developmental arrest in a wide range of arthropods (7), quantitative tracking of gene expression is required alongside gene-family profiling.

Although it seems clear that gene-family profiling will provide increased physiological resolution to most studies of gene expression, it shares the same limitations as any mRNA-based analyses. It does not give insight into processes such as regulation of translation efficiency (e.g., by microRNAs) or posttranslational modification (e.g., chemical modification by phosphorylation or glycosylation, or alterations in protein turnover rates). Nor does it account for physiological complexities such as target genes that serve multiple physiological roles, physiological tasks that are solved by the concerted action of multiple genes, or gene functions that can be replaced by other genes. Again, these limitations also apply to other mRNA-based analyses and importantly also to polypeptide-based analyses. These challenges can, however, to some extent be met or corrected for by the approach itself. As an example, gene-family profiling of microRNA abundance could give valuable indications to how they will impact the translation process. At present, microRNA remains an emerging and highly uncharacterized group of RNAs (15).

Future Prospects

Gene-family profiling is likely to provide increased accuracy to analyses of mRNA expression and should be valuable for

analyses of innate patterns of gene family expression, as well as for evaluation of changes occurring between experimental groups. It seems particularly well suited for medical purposes such as screening of drug libraries, where the aim is to identify specific cellular phenotypes induced by chemical compounds.

ACKNOWLEDGMENTS

We thank Christine Couturier for technical assistance.

GRANTS

This work was supported by Lillehammer University College, University of Oslo, and the Norwegian Research Council.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

REFERENCES

- Arad M, Seidman CE, Seidman JG. AMP-activated protein kinase in the heart: role during health and disease. *Circ Res* 100: 474–488, 2007.
- Bickler PE. Clinical perspectives: neuroprotection lessons from hypoxia-tolerant organisms. *J Exp Biol* 207: 3243–3249, 2004.
- Catterall WA. Structure and regulation of voltage-gated Ca²⁺ channels. *Annu Rev Cell Dev Biol* 16: 521–555, 2000.
- Conti F, Minelli A, Melone M. GABA transporters in the mammalian cerebral cortex: localization, development and pathological implications. *Brain Res Brain Res Rev* 45: 196–212, 2004.
- Cox RD, Buckingham ME. Actin and myosin genes are transcriptionally regulated during mouse skeletal muscle development. *Dev Biol* 149: 228–234, 1992.
- Culmsee C, Monnig J, Kemp BE, Mattson MP. AMP-activated protein kinase is highly expressed in neurons in the developing rat brain and promotes neuronal survival following glucose deprivation. *J Mol Neurosci* 17: 45–58, 2001.
- Denlinger DL. Regulation of diapause. *Annu Rev Entomol* 47: 93–122, 2002.
- Dhedra K, Huggett JF, Chang JS, Kim LU, Bustin SA, Johnson MA, Rook GAW, Zumla A. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. *Anal Biochem* 344: 141–143, 2005.
- Diss JKJ, Fraser SP, Djamgoz MBA. Voltage-gated Na⁺ channels: multiplicity of expression, plasticity, functional implications and pathophysiological aspects. *Eur Biophys J* 33: 180–193, 2004.
- Eizema K, van den Burg M, Kiri A, Dingboom EG, van Oudheusden H, Goldspink G, Weijs WA. Differential expression of equine myosin heavy-chain mRNA and protein isoforms in a limb muscle. *J Histochem Cytochem* 51: 1207–1216, 2003.
- Ellefsen S, Sandvik GK, Larsen HK, Stensløykken KO, Hov DAS, Kristensen TA, Nilsson GE. Expression of genes involved in excitatory neurotransmission in anoxic crucian carp (*Carassius carassius*) brain. *Physiol Genomics* 35: 5–17, 2008.
- Ellefsen S, Stensløykken KO, Fagernes CE, Kristensen TA, Nilsson GE. Expression of genes involved in GABAergic neurotransmission in anoxic crucian carp brain (*Carassius carassius*). *Physiol Genomics* 36: 61–68, 2009.
- Ellefsen S, Stensløykken KO, Sandvik GK, Kristensen TA, Nilsson GE. Improved normalization of real time RT-PCR data using an external RNA control. *Anal Biochem* 376: 83–93, 2008.
- Farrant M, Nusser Z. Variations on an inhibitory theme: phasic and tonic activation of GABA receptors. *Nat Rev Neurosci* 6: 215–229, 2005.
- Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9: 102–114, 2008.
- Flück M, Däpp C, Schmutz S, Wit E, Hoppeler H. Transcriptional profiling of tissue plasticity: role of shifts in gene expression and technical limitations. *J Appl Physiol* 99: 397–413, 2005.
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531–1545, 1999.
- Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 4: 117, 2003.
- Hardie DG. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nat Rev Mol Cell Biol* 8: 774–785, 2007.
- Hylland P, Nilsson GE. Extracellular levels of amino acid neurotransmitters during anoxia and forced energy deficiency in crucian carp brain. *Brain Res* 823: 49–58, 1999.
- Izumo S, Lompre AM, Matsuoka R, Koren G, Schwartz K, Nadalgirard B, Mahdavi V. Myosin heavy-chain mRNA and protein isoform transition during cardiac hypertrophy - interaction between hemodynamic and thyroid hormone-induced signals. *J Clin Invest* 79: 970–977, 1987.
- Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. *Genome Res* 12: 37–46, 2002.
- Jaschinski F, Schuler M, Peuker H, Pette D. Changes in myosin heavy chain mRNA and protein isoforms of rat muscle during forced contractile activity. *Am J Physiol Cell Physiol* 274: C365–C370, 1998.
- Kohr G. NMDA receptor function: subunit composition versus spatial distribution. *Cell Tissue Res* 326: 439–446, 2006.
- Lefaucheur L, Milan D, Ecolan P, Le Calennec C. Myosin heavy chain composition of different skeletal muscles in Large White and Meishan pigs. *J Anim Sci* 82: 1931–1941, 2004.
- Leong HS, Kipling D. Text-based over-representation analysis of microarray gene lists with annotation bias. *Nucl Acids Res* 37: e79, 2009.
- Morey J, Ryan J, Van Dolah F. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biol Proc Online* 8: 175–193, 2006.
- Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 1: 1559–1582, 2006.
- Perez-Reyes E. Molecular physiology of low-voltage-activated T-type calcium channels. *Physiol Rev* 83: 117–161, 2003.
- Prentice HM. The major contribution of brain GABAergic function to anoxic survival. *Physiol Genomics* 36: 59–60, 2009.
- Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 339: 62–66, 2003.
- Sellers JR. Myosins: a diverse superfamily. *BBA Mol Cell Res* 1496: 3–22, 2000.
- Stensløykken KO, Ellefsen S, Stecyk JAW, Dahl MB, Nilsson GE, Vaage J. Differential regulation of AMP-activated kinase and AKT kinase in response to oxygen availability in crucian carp (*Carassius carassius*). *Am J Physiol Regul Integr Comp Physiol* 295: R1803–R1814, 2008.
- Stevens L, Sultan KR, Peuker H, Gohlsch B, Mounier Y, Pette D. Time-dependent changes in myosin heavy chain mRNA and protein isoforms in unloaded soleus muscle of rat. *Am J Physiol Cell Physiol* 277: C1044–C1049, 1999.
- Tricarico C, Pinzani P, Bianchi S, Paglierani M, Distanto V, Pazzagli M, Bustin SA, Orlando C. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Anal Biochem* 309: 293–300, 2002.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paeppe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034, 2002.
- Wright C, Haddad F, Qin AX, Baldwin KM. Analysis of myosin heavy chain mRNA expression by RT-PCR. *J Appl Physiol* 83: 1389–1396, 1997.