Eszter Szantai[1]
Zsuzsanna Elek[1]
András Guttman[2]
Maria Sasvari-Szekely[1]

[1]Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary
[2]Horváth Laboratory of Bioseparation Sciences, Institute of Analytical Chemistry and Radiochemistry, University of Innsbruck, Austria

Received November 18, 2008
Revised December 5, 2008
Accepted December 12, 2008

Short Communication

Candidate gene copy number analysis by PCR and multicapillary electrophoresis

Genetic polymorphisms are often considered as risk factors of complex diseases serving as valuable and easily detectable biomarkers, also stable during the whole lifespan. A novel type of genetic polymorphism has been identified just recently, referred to as gene copy number variation (CNV) or copy number polymorphism. CNV of glycogen synthase kinase 3 beta and its adjacent gene, Nr1i2 (pregnane X receptor isoform), has been reported to associate with bipolar depression. In our study we introduced multicapillary electrophoresis for gene copy number analysis as an affordable alternative to real-time PCR quantification with TaqMan gene probes. Our results show the reliability of the developed method based on conventional PCR followed by separation of products by multicapillary electrophoresis with quantitative evaluation. This method can be readily implemented for the analysis of candidate gene CNVs in high throughput clinical laboratories and also in personalized medicine care of depression-related risk factors.

Keywords:
Copy number variation / Multicapillary gel electrophoresis / Nr1i2
DOI 10.1002/elps.200800755

Utilization of biomarkers in the estimation of disease risk has increased in the last couple of decades. Genetic polymorphisms are considered as important biomarkers also having the advantage of being stable during the whole life of a patient. As these predisposing genetic variations can be analyzed years or decades before the onset of the disease, early mapping of genetic biomarkers could be crucial in disease prevention. Genetic variants spread throughout the whole genome and involve SNPs as well as insertion/deletion polymorphisms of genomic segments varying in size from a single nucleotide to large, microscopically visible chromosomal alterations [1]. These variations can be utilized as markers in linkage [2] and genome-wide analyses [3, 4]. Moreover, they were investigated as functional variants both in coding [5] and non-coding [6, 7] gene regions.

Current focus of genetic polymorphism research has shifted towards a novel type of insertion/deletion polymorphism: the copy number variations (CNVs). These submicroscopic variations of DNA segments range from kilobases to megabases in size. Previously such changes have been detected as chromosomal abnormalities with severe outcome. Recently, multiple studies demonstrated that CNVs are quite frequent in humans [8–10] and other mammals [11]. The first comprehensive map of gene dosage variations was given in 2006 [1], demonstrating that CNVs can be identified in all human chromosomes. According to recent theories the impact of CNVs may be far more extensive than that of single SNPs, since important genes could be eliminated, or extra copies of the gene may cause overproduction of the corresponding protein, affecting the finely balanced biochemistry of the cell. Therefore, it is not surprising that copy number gains (duplications) or losses (deletions) play a role as risk factors in complex diseases, such as in autism [12] and schizophrenia [13].

There are various techniques for CNV determination. Array-based comparative genomic hybridization is suitable to show alterations in 5–10 kilobases of DNA sequences providing genome-wide data; however, the number of false positives and false negatives is also significant, as the accurate identification of the gene triplicates (1.5-fold change) is near the limit of sensitivity.

Quantitative PCR (qPCR) is a frequently used method for gene dosage determination applicable for candidate genes but not for genome wide studies. It can be carried out in the presence of SYBR Green measuring the fluorescence signal produced by the binding of SYBR Green to the studied amplicons and compared with a reference sample [14]. Another highly reliable but more expensive qPCR method applies allele-specific TaqMan® probes labeled by different reporter fluorophores (VIC and FAM) in a single reaction. This method was previously applied in our laboratory to determine the CNV of

Correspondence: Dr. Eszter Szantai, Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, SOTE – Chemistry, Tuzolto u. 37–47, Budapest, H-1094, Hungary
E-mail: szantai@puskin.sote.hu
Fax: +36-1-2662615

Abbreviations: CNV, copy number variation; **C_t,** cycle threshold; **Gsk3β,** glycogen synthase kinase 3 beta; **qPCR,** quantitative PCR

human complement C4A and C4B genes [15]. Application of three sequence-specific probes for a single gene gave much higher specificity to the TaqMan assay compared with the SYBR Green method.

Multiplex ligation-dependent probe amplification is an alternative for detection of copy number changes [16]. MLPA® probes consist of two oligonucleotides, hybridizing adjacent to each other and a sequence complementary to the target, known as the hybridization sequence. When the probes correctly hybridize to the target sequence, they are ligated by a thermo-stable ligase enzyme and amplified. One of the primers is labeled with a fluorescent dye to visualize the amplified products.

CGE is an excellent tool for identification of double-stranded DNA fragments such as the products of PCRs. If the PCR is stopped in the exponential phase, quantitative analysis of gene copy number is also possible since the peak areas are then directly proportional to the amount of DNA in the original sample [17].

Recently, CNV of glycogen synthase kinase 3 beta (Gsk3β) and its adjacent gene, Nr1i2 (pregnane X receptor isoform) has been shown by SYBR Green qPCR method to be associated with bipolar depression [14]. Figure 1 shows the position of these genes on chromosome 3. Since they are strongly in linkage disequilibrium and the CNV affects Nr1i2 gene in full but only the 3′-flanking part of Gsk3β, it is assumed that the real effect on bipolar depression is caused rather by changes of the full Nr1i2 gene and not so much by variations in the 3′-flanking part of Gsk3β gene.

Here we compare the two methods of TaqMan real-time PCR assay and a cost-effective analysis based on conventional PCR followed by separation of products by multi-CGE for the determination of gene copy number of the Nr1i2 gene and a control gene (RNase P). The ratio of these two genes defines the copy number for Nr1i2, considering that the control gene (RNase P) is always present in two copies.

DNA samples were obtained by a non-invasive method of collecting buccal swabs. To isolate the genomic DNA the Gentra DNA isolation kit (Minneapolis, MN, USA) was used. The study was approved by the Hungarian Research Ethics Committee with signed informed consent from all participants.

Relative gene dosage determination was carried out by an ABI 7500 Real Time PCR System. The 15 μL reaction mixture contained FAM-labeled 1 × TaqMan® Nr1i2 kit (ABI, Hs02515976_s1), VIC-labeled 1 × TaqMan® Human RNase P detection mix (ABI, Cat. No. 4316844),

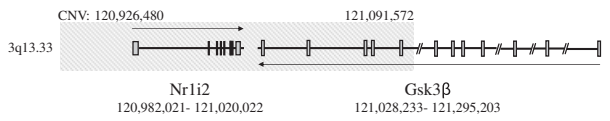

Figure 1. Localization of the Nr1i2 and Gsk3β genes on chromosome 3. The shaded area depicts the proposed CNV (http://projects.tcag.ca/variation); the arrows show the direction of transcription. Positions of the genes are given in accordance with the NCBI database (http://www.ncbi.nlm.nih.gov/).

1 × TaqMan Universal PCR Master Mix (AmpliTaq Gold® DNA Polymerase, dNTPs with dUTP, Passive Reference, No AmpErase UNG®) and 3 ng genomic DNA template. Thermocycling was initiated at 95°C for 10 min, which was followed by 40 cycles of 15 s denaturation at 95°C, and 1 min annealing-extension at 60°C. PCR primers were designed for exon 9 of Nr1i2 gene and for exon 1 of RNase P gene since the ABI probes used for real-time PCR also anneal to these exons. The Qiagen® HotStarTaq™ DNA polymerase kit (Qiagen, Valencia, California, USA) was used for the PCR of the Nr1i2 and Rnase P genes. The reaction mixtures contained 200 μM of each deoxyribonucleotide-triphosphate (Promega, Madison, WI, USA); 1 μM of Rnase P forward (5′GCA AGY AAG TTT CTC CGA ATC C3′) and Rnase P reverse (5′GCG CAG AGC CTT CAG GT3′) primers, 0.5 μM of Nr1i2 forward (5′GAT GGC AGG GCA GGA AGA T3′) and Nr1i2 reverse (5′GCA TGG GCT CCA GTA GAA GTT G3′) primers; approximately 1 ng DNA template, 0.25 U DNA polymerase, 1 × reaction buffer, and 1 × Q solution in a total volume of 10 μL. The primers were designed by the Oligo 5.0 software (Plymouth, MN, USA). Thermocycling was initiated at 95°C for 15 min followed by 28 cycles of 30 s denaturation at 94°C, 30 s annealing at 60°C and 1 min extension at 72°C. A final 10 min extension step at 72°C was followed by cooling the samples to 8°C.

The PCR products were analyzed on QIAxcel multi-capillary electrophoresis system (Qiagen; formerly eGene's HDA-GT12 system) with a 12-capillary gel-cartridge DNA Screening kit. Separations were performed by using method AL320 with 6000 V separation voltage and 20 s injection time at ambient temperature.

The aim of the presented work was to compare real-time PCR quantification and the multi-CGE analysis in conjunction with conventional PCR to analyze gene CNVs. During the screening process for the Nr1i2 gene, amplification of copy 3 instead of 2 was found in several individuals, while in most of the cases the copy number was equal to the copy number of the control gene (two copies). For comparative purposes, we present the results of sample 1 (amplified) and sample 2 (normal) by both methods. Figure 2 depicts the logarithmic representation of real-time PCR curves of the samples possessing copy numbers 2 and 3 of Nr1i2. Cycle threshold (C_t) is the intersection between amplification curve and threshold line; that is, the cycle at which a predetermined amount of product is generated. Therefore, the C_t is a relative measure of the concentration of target in the PCR. ΔC_t of a sample is the difference of C_t values of Nr1i2 and RNase P genes, which arises from the efficacy difference of the two kits in the case when both genes have identical copy numbers. Thus, $\Delta\Delta C_t$ means the difference of ΔC_t values of two samples, whose value is proportional to the CNV between two individuals. The theoretical $\Delta\Delta C_t$ value in the case of 1.5-fold copy number change is $\log_2(3/2) = 0.58$ or $\log_2(2/3) = -0.58$ as the amount of the amplicon gets doubled in every cycle. As one can see in Fig. 2, ΔC_t of sample 1 is 0.64 and sample 2 is 0.04, applying automatic threshold restriction. Therefore,

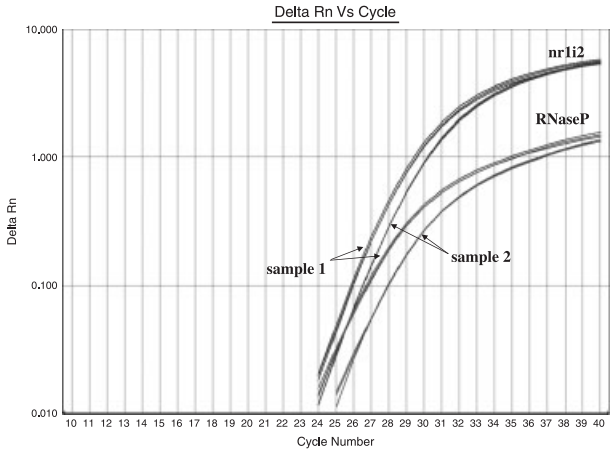

Figure 2. Logarithmic representation of real-time PCR curves of samples from two individuals with three parallel measures. The lines depict the Nr1i2 and the control Rnase P genes. Normalized reporter (Rn) is the ratio of the fluorescence emission intensity of the reporter dye to the fluorescence emission intensity of the passive reference dye.

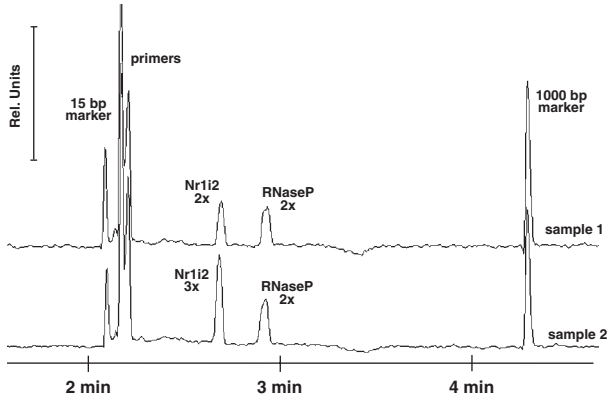

Figure 3. Multi-CGE of patient samples 1 and 2 with Nr1i2 copy numbers 2 and 3, respectively. Separations are aligned by means of bracketing markers (15 and 1000 bp) for highly accurate size determination. DNA fragments were visualized by ethidium-bromide; excitation wavelength was 524 nm. Size of Nr1i2 and RNase P fragments are 143 and 196 bp long, respectively.

$\Delta\Delta C_t$ is 0.60, which corresponds to Nr1i2 copy numbers 2 (sample 1) and 3 (sample 2).

Figure 3 depicts the multicapillary electrophoresis analysis of the same samples. Concentrations of the PCR primers were adjusted so that in the case of normal copy numbers the peak areas of the two genes were approximately the same. In this way a 1.5 Nr1i2/RNase P normalized area ratio refers to an Nr1i2 copy number 3, while 0.5 refers to Nr1i2 copy number 1. The normalized areas of the Nr1i2 peaks were 4.13E-03 and 8.07E-03 in samples 1 and 2, respectively, while the normalized area of the RNase P peaks were 4.49E-03 and 5.26E-03 in samples 1 and 2, respectively. The Nr1i2/RNase P ratio of normalized areas in sample 1 was 0.92, which refers to Nr1i2 copy number 2. Nr1i2/RNase P ratio of normalized areas in sample 2 was 1.53, which refers to Nr1i2 copy number 3, assuming that RNase P copy number was 2. Our assumption about the RNase P copy number was previously verified [15].

Our results show that conventional PCR combined with multicapillary electrophorsis can be considered as a good alternative for gene CNV determination. The main advantages of multicapillary electrophoresis are that its sample consumption is very low and still offers high detection sensitivity. Speed and automation are also considered to be important, especially in the case of high throughput applications utilizing 96-well plates. Applying a multicapillary electrophoreses system, automated loading of the PCR products from 96-well plates and computer-directed analysis of samples allow the fast and reproducible quantitative analysis of gene CNV. It is also important to note that besides this technique is high-throughput, it is also more cost effective than other CNV analyzing tools applied (for comparison: costs of the widely used TaqMan probe system are about 8-to-10-fold higher). Therefore, the applied multicapillary eletrophoresis system seems to be a good candidate for wide application in basic and clinical research to identify genetic markers for complex diseases.

This work was supported by the Hungarian National fund OTKA T048576.

The authors have declared no conflict of interest.

References

[1] Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H. *et al.*, *Nature* 2006, *444*, 444–454.

[2] Amos, C. I., Chen, W. V., Lee, A., Li, W., Kern, M., Lundsten, R., Batliwalla, F. *et al.*, *Genes Immun.* 2006, *7*, 277–286.

[3] John, S., Shephard, N., Liu, G., Zeggini, E., Cao, M., Chen, W., Vasavda, N. *et al.*, *Am. J. Hum. Genet.* 2004, *75*, 54–64.

[4] Middleton, F. A., Pato, M. T., Gentile, K. L., Morley, C. P., Zhao, X., Eisener, A. F., Brown, A. *et al.*, *Am. J. Hum. Genet.* 2004, *74*, 886–897.

[5] Ollerenshaw, M., Page, T., Hammonds, J., Demaine, A., *Cancer Genet. Cytogenet.* 2004, *153*, 122–126.

[6] Kereszturi, E., Kiraly, O., Barta, C., Molnar, N., Sasvari-Szekely, M., Csapo, Z., *BMC Mol. Biol.* 2006, *7*, 18–26.

[7] Okuyama, Y., Ishiguro, H., Toru, M., Arinami, T., *Biochem. Biophys. Res. Commun.* 1999, *258*, 292–295.

[8] Feuk, L., Carson, A. R., Scherer, S. W., *Nat. Rev.* 2006, *7*, 85–97.

[9] Iafrate, A. J., Feuk, L., Rivera, M. N., Listewnik, M. L., Donahoe, P. K., Qi, Y., Scherer, S. W. *et al.*, *Nat. Genet.* 2004, *36*, 949–951.

[10] Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S. *et al.*, *Science* 2004, *305*, 525–528.

[11] Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., Aburatani, H. *et al.*, *Genome Res.* 2006, *16*, 949–961.

[12] Miller, D. T., Shen, Y., Weiss, L. A., Korn, J., Anselm, I., Bridgemohan, C., Cox, G. F. *et al.*, *J. Med. Genet.* 2008.

[13] International Schizophrenic Consortium, *Nature* 2008, 455, 237–241.

[14] Lachman, H. M., Pedrosa, E., Petruolo, O. A., Cockerham, M., Papolos, A., Novak, T., Papolos, D. F. *et al.*, *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 2007, *144B*, 259–265.

[15] Szilagyi, A., Blasko, B., Szilassy, D., Fust, G., Sasvari-Szekely, M., Ronai, Z., *BMC Genet.* 2006, *7*, 1–9.

[16] den Dunnen, J. T., White, S. J., in: Haines J. L. *et al.* (Eds.), *Current Protocols in Human Genetics/Editorial Board*, Wiley: Hoboken, NJ 2006, Chapter 7, Unit 7.14

[17] Szilagyi, A., Blasko, B., Ronai, Z., Fust, G., Sasvari-Szekely, M., Guttman, A. *et al.*, *Electrophoresis* 2006, *27*, 1437–1443.

