Allele Frequency Estimation of Single Nucleotide Polymorphisms in the Promoter and 5'untranslated Region of Vascular Endothelial Growth Factor in Thai Breast Cancer Patients

Doonyapat Sa-nguanraksa MD, PhD*,
Prida Malasit MD**,***, Pornchai O-charoenrat MD, PhD*

** Division of Head Neck and Breast Surgery, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand*
*** Division of Dengue Hemorrhagic Fever Unit, Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*
**** Medical Biotechnology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Bangkok, Thailand*

Background: *Polymorphisms in the promoter and 5'untranslated region of vascular endothelial growth factor (VEGF) had been reported to be associated with increased risk and aggressiveness of breast cancer via alteration of promoter activity.*
Objective: *To screen for novel polymorphisms in the promoter and 5'untranslated region of VEGF and estimate allele frequency.*
Material and Method: *Screening of novel single nucleotide polymorphisms was performed by direct sequencing of pooled genomic DNA from 100 Thai breast cancer patients. Estimated allele frequency was calculated according to adjusted peak height compared to that of heterozygous individual chromatogram.*
Results: *Estimated allele frequency of the known polymorphisms was comparable with the observed allele frequency obtained by genotyping of individual genomic DNA. However, no novel polymorphisms could be identified.*
Conclusion: *This method might be employed to screen for novel informative single nucleotide polymorphisms or estimate allele frequency as the preliminary data before large scale genotyping.*

Keywords: *Breast cancer, Vascular endothelial growth factor, Polymorphisms, DNA sequencing*

J Med Assoc Thai 2017; 100 (Suppl. 2): S48-S53
Full text. e-Journal: *http://www.jmatonline.com*

Angiogenesis, the formation of the new blood vessels had been implicated in carcinogenesis[1,2]. Vascular endothelial growth factor (VEGF), the most important moderator in this complex process, had significant role in aggressiveness of breast cancer[3].

Promoter and 5' untranslated region (UTR) of VEGF span3 kilobases. Several single nucleotide polymorphisms (SNPs) in the promoter and 5' UTR had been identified[4,5]. Several studies demonstrated associations between VEGF polymorphisms and levels of VEGF expression, as well as susceptibility and aggressiveness of breast cancer. However, the results were still inconclusive[6].

VEGF 634G/C and 634C/C were associated with an increased risk for breast cancer in Thai population. The VEGF 634G/C was also associated with larger tumor size, presence of perineural invasion and more advanced stage. In addition, the patients with 634C/C had lower disease free survival[7]. In vitro model demonstrated that VEGF promoter containing -1498T/-634C haplotype had increased in promoter activity[8]. This finding indicated that the highly polymorphic VEGF promoter and 5' UTR had haplotypic effect in alteration of promoter activity.

To identify novel SNPs located on promoter and 5' UTR of VEGF, screening of polymorphisms was performed by direct sequencing of pooled genomic DNA. Allele frequencies were estimated based on peak height of chromatogram from heterozygous individual.

Correspondence to:
O-charoenrat P, Division of Head-Neck and Breast Surgery, Department of Surgery, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.
Phone: +66-2-4198016, Fax: +66-2-4198929
E-mail: pornchai.och@mahidol.ac.th

Material and Method

Patients

Breast cancer patients were recruited from the Division of Head-Neck and Breast Surgery, Department of Surgery, Faculty of Medicine, Siriraj Hospital, Mahidol University, Thailand. Patients with newly diagnosed breast cancer, aged 18 years or older with ability to give informed consent were included. Patients with histories of other cancer were excluded. At recruitment, informed consent was obtained from each subject. This study was approved by the Siriraj Ethical Committee on Research.

DNA amplification

Genomic DNA was isolated from peripheral blood lymphocytes of the subjects using standard phenol-chloroform method. One hundred DNA samples were randomly picked. All DNA samples were quantified twice by spectrophotometry and then diluted to 50 ng/µl. The diluted DNA samples were re-quantified to confirm the concentrations. To check DNA integrity, all DNA samples were used as templates for polymerase chain reaction (PCR) utilizing primers specific to β-globin gene (Table 1). DNA pool was consisted of 200 ng aliquots from each sample. The DNA pool was mixed by rocking at 4°C for 16 hours. Overlapping fragments, 281 to 375 base pairs (bp) in length, were PCR amplified to cover the chosen 2,935-bp promoter region and 5' UTR. Primers were used to amplify 800 ng of individual genomic DNA or pooled genomic DNA in 100 µl reaction volumes (Table 1).

Purification of PCR products and DNA sequencing

PCR products were separated on 1% agarose in 1X TAE buffer. The gel bands containing the desired DNA fragments were excised and purified using Qiaprep (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. The purified PCR products were estimated for concentration by agarose gel electrophoresis. The intensity of PCR products was compared to the standard DNA marker. The purified PCR products were used as sequencing template

Table 1. PCR primer pairs and reaction conditions for SNPs screening

	Primers	Tm (°C)	Product size (bp)	PCR conditions (T_{ann}, $MgCl_2$)[a]
Frag 1	F: GGCTCTTTTAGGGGCTGAAG	62	345	60, 2.5
	B: ACCCCAACCCTCCCTTTC	58		
Frag 2	F: CAGGACTAGTGCACGAATGA	60	281	60, 2.5
	B: AGATCGTGCCAGGGTCTG	58		
Frag 3	F: GAGGCTATGCCAGCTGTAGG	64	329	60, 2.5
	B: AGAAATGGGGGTTTCGTCTC	60		
Frag 4	F: AAGCTGTGAGCCTGGAGAAG	62	339	60, 2.5
	B: CACTGTGGAGTCTGGCAAAA	60		
Frag 5	F: TCAGTTCCCTGGCAACATCT	60	320	60,1.5
	B: TGACAATGTGCCATCTGGAG	60		
Frag 6	F: CTGCCGCTCACTTTGATGT	58	314	60, 2.5
	B: CACACACGTCCTCACTCTCG	64		
Frag 7	F: GGAAGCTGGGTGAATGGAG	60	338	60, 1.5
	B: CAGCCTGAAAATTACCCATCC	62		
Frag 8	F: GCGTGTCTCTGGACAGAGTTT	64	303	60, 1.5,
	B: ACGACCTCCGAGCTACCC	60		Q-solution[b]
Frag 9	F: GCTGGTAGCGGGGAGGAT	60	312	60, 2.5
	B: CCACAGTGATTTGGGGAAGT	60		
Frag 10	F: CATTGATCCGGGTTTTATCCC	62	375	60, 2.5
	B: CTGTCTGTCTGTCCGTCAG	60		
Frag 11	F: ATCCCGCAGCTGACCAGT	58	339	60,1.5,
	B: CCCCTCTCCTCTTCCTTCTC	64		Q-solution
Control	F: ACACAACTGTGTTCACTAGC	58	110	63, 2.5
	B: CAACTTCATCCACGTTCACC	60		

[a] Annealing temperature (°C) and $MgCl_2$ concentrations (mM); [b] with Q-solution additive (Qiagen)

without further characterization or manipulations. Direct sequencing of the PCR products were performed at Macrogen Inc. (Gasan-dong, Geumcheon, Seoul, Korea).

Identification of sequence variations

The DNA sequencing traces from 3 individual control DNA samples and the pooled DNA were analyzed. Peak heights of the sequencing traces were measured using BioEdit software and normalized by computation from the average peak height for the same base chosen from two 25-base windows upstream and downstream (-28 to -4 and 4 to 28), excluding peaks that were off-scale, less than 10% of full-scale, or within 3 bases of the base in question. Any base with a decrease in normalized peak height with a concomitant appearance of a new base-peak underneath when compared to its homozygous counterpart was considered polymorphisms[9].

Estimation of allele frequency in a population

The sequencing traces from the pooled DNA that revealed polymorphisms were compared with the sequencing traces from the individual. If there was heterozygous individual, the relative peak heights of the corresponding bases between these two sequencing traces can provide an estimate of the allele frequencies. The estimated allele frequencies (p and q) were determined by using the formula $p = 0.5x (A_{pool}/A_{het})$ and $q = 1 - p$, where A_{pool} and A_{het} are normalized peak heights of the major allele seen in the sequencing traces of the pooled DNA and the heterozygote, respectively[9].

Results

One hundred DNA samples were randomly picked from breast cancer patients. The DNA samples were diluted to 50 ng/µl and performed PCR reaction using primers specific to β-globin gene. The presence of PCR products indicated the integrity of the DNA (Fig. 1).

A total of ten overlapping fragments were sequenced. The third fragment which was known to contain -2549 18 bp insertion/deletion and -2447 G insertion/deletion was skipped. To validate this assay, the sequencing results of fragment 7, 8, and 10 which contained common polymorphisms, VEGF -1498C/T, -1154G/A, and -634G/C, were analyzed whether they exhibited decrement of normalized peak with a concomitant appearance of a new base peak (Fig. 2). When compared to the individual DNA samples, the pooled DNA sample exhibited the decrement of normalized peak of major allele with appearance of minor allele underneath (Fig. 3). VEGF -1498C/T,

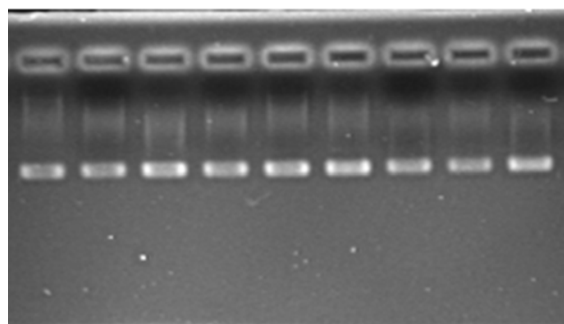

Fig. 1 Representative agarose gel electrophoresis of control PCR reaction for confirmation of DNA integrity. The genomic DNA was amplified using a primer pair that specific to β-globin gene. All samples displayed 110 bp PCR products with comparable intensity.

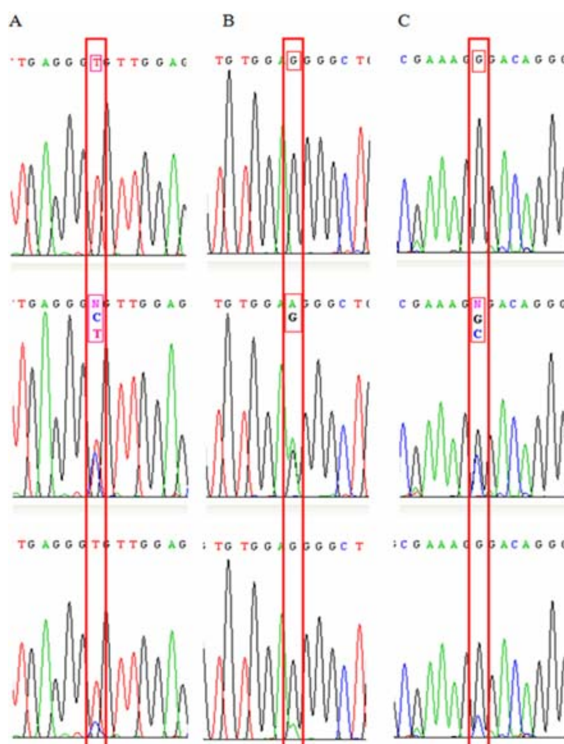

Fig. 2 Validation of pooled DNA sequencing. The sequencing traces from the fragments that corresponded to the polymorphisms at -1,498, -1,154, -634 were analyzed. Column A) -1,498C/T; column B) -1,154G/A; column C) -634G/C. Upper row: homozygote; middle row: heterozygote; lower row: pooled DNA sample.

-1154G/A and -634G/C exhibited decrement of 34.03, 19.32, and 30.14%, respectively (Table 2). The estimated allele frequencies of the major allele were 76.0, 83.2, and 71.9 for -1498T, -1154G, and -634G allele (Table 3).

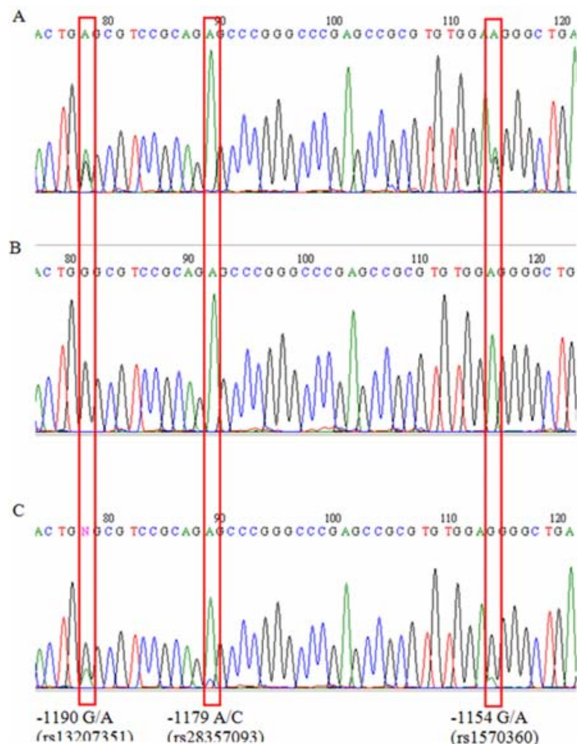

Fig. 3 Representative comparison between sequencing traces of pooled DNA and individual DNA. A) Individual DNA sample 1 revealed -1,190GA, -1,179AA, and -1,154GA genotype. B) Individual DNA sample 2 revealed -1,190GG, -1,179AA, and-1,154AA genotype. C) Pooled DNA sample exhibited decrement of the major allele base peak with a concomitant appearance of the minor allele base peak.

Analysis of the sequencing results revealed other 2 loci with decreased in normalized peak height with appearance of new base peak at -1190 and -1179 (Fig. 3). The -1190 displayed 21.96% decrement of normalized peak height of G peak with A peak underneath (Table 2). The sequencing results showed that one of individual DNA sample is heterozygote at this locus. The estimated frequency of -1190G allele was 76.7%. The -1179 exhibited 26.89% decrement of normalized peak height of A peak with C peak underneath. There was no heterozygote at this locus, thus, the estimated frequency cannot be calculated. These two SNPs were compared with SNPs database of NCBI. The -1190 G/A and -1179 A/C were equivalent to rs13207351 and rs28357093, respectively. No novel SNPs was identified in the studied population.

Discussion

Direct sequencing of pooled DNA in Thai breast cancer patients could not identify novel SNPs in VEGF promoter. Nevertheless, estimated allele frequencies of the known SNPs were concordant with the frequencies derived from individual genotyping.

SNPs screening by direct sequencing of pooled DNA not only identify the SNPs but also allele frequency. A study by Kwok et al showed that this method was reliable and reproducible(9). Different methods for determination of allele frequency in pooled DNA using high throughput strategy was tested and compared. Albeit the accuracy of each method was different, all methods had sufficient accuracy to determine allele frequency in large-scale association study(10). When compared with real-time PCR, direct sequencing of pooled DNA exhibited the advantages in terms of simple and flexible primer design and possibility to assess proximate polymorphisms(11).

Table 2. Comparison of peak intensity between pooled DNA, homozygote and heterozygote

Locus	Major allele	Normalized peak heights			
		Pooled DNA	Homozygote	Heterozygote	Change (%)[a]
-1498 C/T	T	0.698337	1.058554	0.459401	34.03
-1190 G/A	G	0.781887	1.001937	0.509736	21.96
-1179 A/C	A	1.306755	1.787349	N/A[b]	26.89
-1154 G/A	G	0.813216	1.007946	0.488586	19.32
-634 G/C	G	1.164195	1.666365	0.809085	30.14

[a] Calculated from peak height of major allele of homozygote and the pooled DNA; [b] Among individual DNA samples, no heterozygote was found.

Table 3. Estimated allele frequency calculated from peak heights of the pooled DNA sequencing traces

	Allele	Estimated
-1498 C/T	T	76.0
	C	24.0
-1190 G/A	G	76.7
	A	23.3
-1179 A/C	A	N/A[a]
	C	N/A
-1154 G/A	G	83.2
	A	16.8
-634 G/C	G	71.9
	C	28.1

[a] Cannot be estimated due to lack of heterozygous individual sample sequencing

Utilization of next generation sequencing technique combined with specialized software had been proofed to facilitate large scale association study with satisfied performance[(12,13)].

Direct sequencing of pooled DNA by conventional technique is cost effective, not require special instrument, can identify polymorphisms and determine allele frequency simultaneously. However, because of depending on the peak height underneath the common allele, this method may not suitable for rare allele with frequency less than 10%. Another limitation is when genomic DNA contained heterozygous of deletion/insertion, the sequencing traces exhibited overlapping DNA sequences resulting in complicated interpretation. This method is useful in identifying informative SNPs and deriving preliminary data before large scale genotyping.

What is already known on this topic?

VEGF polymorphisms altered promotor activity and were associated with breast cancer susceptibility and aggressiveness.

What this study adds?

Pooled DNA sequencing for polymorphisms screening is cost-effective and can be used in estimation of allele frequency.

Acknowledgements

This study received grant support from AstraZeneca (Thailand) for data collection process. The authors would like to express their appreciation to Onchira Acharayothin for data collection.

Potential conflicts of interest

None.

References

1. Folkman J. Tumor angiogenesis: therapeutic implications. N Engl J Med 1971; 285: 1182-6.
2. Folkman J. What is the evidence that tumors are angiogenesis dependent? J Natl Cancer Inst 1990; 82: 4-6.
3. Gasparini G. Prognostic value of vascular endothelial growth factor in breast cancer. Oncologist 2000; 5 (Suppl 1): 37-44.
4. Brogan IJ, Khan N, Isaac K, Hutchinson JA, Pravica V, Hutchinson IV. Novel polymorphisms in the promoter and 5' UTR regions of the human vascular endothelial growth factor gene. Hum Immunol 1999; 60: 1245-9.
5. Watson CJ, Webb NJ, Bottomley MJ, Brenchley PE. Identification of polymorphisms within the vascular endothelial growth factor (VEGF) gene: correlation with variation in VEGF protein production. Cytokine 2000; 12: 1232-5.
6. Sa-Nguanraksa D, Charoenrat P. The role of vascular endothelial growth factor a polymorphisms in breast cancer. Int J Mol Sci 2012; 13: 14845-64.
7. Sa-Nguanraksa D, Chuangsuwanich T, Pongpruttipan T, Kummalue T, Rojananin S, Ratanawichhitrasin A, et al. Vascular endothelial growth factor 634G/C polymorphism is associated with increased breast cancer risk and aggressiveness. Mol Med Rep 2013; 8: 1242-50.
8. Sa-Nguanraksa D, Kooptiwut S, Chuangsuwanich T, Pongpruttipan T, Malasit P, Charoenrat P. Vascular endothelial growth factor polymorphisms affect gene expression and tumor aggressiveness in patients with breast cancer. Mol Med Rep 2014; 9: 1044-8.
9. Kwok PY, Carlson C, Yager TD, Ankener W, Nickerson DA. Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. Genomics 1994; 23: 138-44.
10. Shifman S, Pisante-Shalom A, Yakir B, Darvasi A. Quantitative technologies for allele frequency estimation of SNPs in DNA pools. Mol Cell Probes 2002; 16: 429-34.
11. Wilkening S, Hemminki K, Thirumaran RK, Bermejo JL, Bonn S, Forsti A, et al. Determination of allele frequency in pooled DNA: comparison of three PCR-based methods. Biotechniques 2005;

39: 853-8.
12. Jin SC, Benitez BA, Deming Y, Cruchaga C. Pooled-DNA sequencing for elucidating new genomic risk factors, rare variants underlying Alzheimer's disease. Methods Mol Biol 2016; 1303: 299-314.
13. Vallania F, Ramos E, Cresci S, Mitra RD, Druley TE. Detection of rare genomic variants from pooled sequencing using SPLINTER. J Vis Exp 2012; pii: 3943.

การประมาณความถี่ของอัลลีลบนโพรโมเตอร์และส่วนนำก่อนแปรรหัสของยีน vascular endothelial growth factor ในผู้ป่วยมะเร็งเต้านมชาวไทย

ดุลยพัฒน์ สงวนรักษา, ปรีดา มาลาสิทธิ์, พรชัย โอเจริญรัตน์

ภูมิหลัง: *ความแปรผันของโพรโมเตอร์และส่วนนำก่อนแปรรหัสของจีน vascular endothelial growth factor (VEGF) มีความสัมพันธ์กับการเพิ่มความเสี่ยงและความรุนแรงของมะเร็งเต้านมโดยเปลี่ยนแปลงระดับการทำงานของโพรโมเตอร์*
วัตถุประสงค์: *เพื่อคัดกรองหาความแปรผันของโพรโมเตอร์และส่วนนำก่อนแปรรหัสของจีน VEGF*
วัสดุและวิธีการ: *การตรวจหาความแปรผันของจีน VEGF ทำโดยการตรวจลำดับนิวคลีโอไทด์ในสายพันธุกรรมที่ได้รับการผสมรวมกันของผู้ป่วยมะเร็งเต้านมชาวไทย 100 ราย ความถี่ของอัลลีลได้จากการคำนวนความสูงของกราฟ*
ผลการศึกษา: *ความถี่ของอัลลีลที่ได้ใกล้เคียงกับวิธีการตรวจโดยใช้สารพันธุกรรมของผู้ป่วยแต่ละรายแต่ไม่สามารถพบการแปรผันในตำแหน่งใหม่ ๆ ได้*
สรุป: *การตรวจลำดับนิวคลีโอไทด์ในสายพันธุกรรมที่ได้รับการผสมรวมกันอาจใช้ในการตรวจหาความแปรผันของจีนในตำแหน่งใหม่ และสามารถใช้ประมาณความถี่ของอัลลีลก่อนทำการศึกษาในกลุ่มประชากรขนาดใหญ่*