

Candidate gene polymorphisms for diabetes mellitus, cardiovascular disease and cancer are associated with longevity in Koreans

Ji Wan Park^{1*}, Yong Ick Ji^{2*}, Yoon-Ho Choi³,
Mi-Yeon Kang², Eunhyun Jung², Se Young Cho²,
Hee-Youn Cho², Byung Kyu Kang²,
Yoo Sook Joung⁴, Duk-Hwan Kim^{2,5},
Sang Chul Park^{6,7} and Joobae Park^{2,5,7}

¹Department of Medical Genetics
Hallym University College of Medicine
Chuncheon 200-702, Korea

²Center for Genome Research
Samsung Biomedical Research Institute

³Department of Internal Medicine

⁴Department of Psychiatry

⁵Department of Molecular Cell Biology
Sungkyunkwan University School of Medicine
Seoul 135-710, Korea

⁶Aging and Apoptosis Research Center
Seoul National University Institute on Aging
Seoul 110-799, Korea

⁷Corresponding authors: Tel, 82-2-3410-3600;
Fax, 82-2-3410-3620; E-mail, jbpark@med.skku.ac.kr (J.B. Park),
Tel, 82-2-740-8244; Fax, 82-2-744-4534;
E-mail, scpark@snu.ac.kr (S.C. Park)

*These authors contributed equally to this work.

DOI 10.3858/emm.2009.41.11.083

Accepted 22 June 2009

Abbreviations: *ABCC8*, ATP-binding cassette, sub-family C (CFTR/MRP), member 8; *ACE*, angiotensin I converting enzyme (peptidyl-dipeptidase A) 1; *ADCY2*, adenylate cyclase 2 (brain); *ALDH2*, aldehyde dehydrogenase 2 family (mitochondrial); *APOE*, apolipoprotein E; *CDK10*, cyclin-dependent kinase 10; *CETP*, cholesteryl ester transfer protein, plasma; CI, confidence interval; CR, call rate; CVD, cardiovascular disease; DM, diabetes mellitus; HWE, Hardy-Weinberg equilibrium; *IL12RB2*, interleukin 12 receptor, beta 2; LD, linkage disequilibrium; LRM, logistic regression model; *LYN*, V-yes-1 Yamaguchi sarcoma viral related oncogene homolog; MAF, minor allele frequency; *MGP*, matrix Gla protein; *MTTP*, microsomal triglyceride transfer protein; OR, odds ratio; *PAX4*, paired box 4; *PCSK1*, convertase subtilisin/kexin type 1; *PON1*, paraoxonase 1; *PPP1R14C*, protein phosphatase 1, regulatory (inhibitor) subunit 14C; *PPP1R1A*, protein phosphatase 1, regulatory (inhibitor) subunit 1A; SNP, single nucleotide polymorphism; *TNFSF11*, tumor necrosis factor (ligand) superfamily, member 11

Abstract

Long-lived people may have a unique genetic makeup that makes them more resistant than the general population to prevalent age-related diseases; however, not much is known about genes involved in the longevity. To identify susceptibility variants controlling longevity, we performed a high-throughput candidate gene study using 137 Koreans over 90 yr old and 213 young healthy Koreans. We evaluated 463 informative markers located in 176 candidate genes mostly for diabetes mellitus, cardiovascular disease and cancer under five genetic models. We estimated the odds ratios for each allele, genotype, haplotype, and gene-gene interaction using logistic regression analysis. Associations between 13 genes and longevity were detected at a *P*-value less than 0.01. Particularly, the rs671 (A) allele of the aldehyde dehydrogenase 2 family (mitochondrial) (*ALDH2*) gene was associated with longevity only in men (OR 2.11, *P* = 0.008). Four genes, proprotein convertase subtilisin/kexin type 1 (*PCSK1*, *P* = 0.008), epidermal growth factor receptor (*EGFR*, *P* = 0.003), paired box 4 (*PAX4*, *P* = 0.008), and V-yes-1 Yamaguchi sarcoma viral related oncogene homolog (*LYN*, *P* = 0.002) consistently yielded statistical evidence for association with longevity. The findings of the current study may provide a starting point for future studies to unravel genetic factors controlling longevity in Koreans.

Keywords: case-control studies; genome-wide association study; longevity; polymorphism, single nucleotide

Introduction

In the past 50 yr, the mean life expectancy in most developed countries has increased from around 50 yr to 80 yr. This change suggests that environmental changes may influence on extending the human lifespan. At the same time, variation in maximum lifespan among species and genetic variations in lifespan across human populations clearly points toward a genetic basis for lifespan (Christensen *et al.*, 2006; Bergman *et al.*, 2007).

According to a birth cohort study from north-east America, having a centenarian sibling increased the chance of survival beyond ninety years old by four times, indicating a strong familial aggregation to longevity (Perls *et al.*, 1998). The heritability of lifespan was around 33% and did not vary by gender in Danish twins (McGue *et al.*, 1993). Linkage study is often considered inappropriate for complex traits; however, a region of chromosome 4 was shown to be linked with exceptional longevity (Puca *et al.*, 2001). Gene-environment interactions (*e.g.* ACE, angiotensin I converting enzyme (peptidyl-dipeptidase A) 1 and exercise; IL6, interleukin 6 (interferon, beta 2) and zinc) also contribute to the development of aging-related phenotypes (*e.g.* physical decline and inflammation, respectively) (Capri *et al.*, 2006).

The frequency of centenarians is approximately 1/100,000 people worldwide. According to the Korea National Statistical Office (KNSO) Report in 2005 (<http://www.kosis.kr>), the frequencies of nonagenarians and centenarians were 131 and 2/100,000 Koreans, respectively (KNSO, 2005). Centenarians may not be free of disease, but manage to survive with or without treatment owing to their resistance to disease (Franceschi *et al.*, 2000). The incidences of major common diseases slow down or even decline around age 85-90 yr (Bergman *et al.*, 2007). Gender also accounts for important differences in longevity as well as in the occurrence of a variety of age-related disease. The male/female ratios were about 1:4 in nonagenarians and 1:8.2 in centenarians in Korea (KNSO, 2005). Such data raise the possibility of gender-specific genetic determinants of human longevity.

Approximately 30 candidate genes for longevity have been reported in previous studies since Herman (1956) proposed the free radical theory of aging, where oxidative stress causes aging (Salvioli *et al.*, 2006). Genes involved in inflammation and the immune response (*e.g.* IL6; LMP2, low molecular mass protein 2; TNFA, tumor necrosis factor (TNF superfamily, member 2); TGFB1, transforming growth factor, beta 1; and PPAR γ , peroxisome proliferator-activated receptor gamma), genome maintenance and repair (*e.g.* P53), lipid metabolism (*e.g.* APOE, apolipoprotein E; CETP, cholesteryl ester transfer protein, plasma; MTTP, microsomal triglyceride transfer protein, and PON1, paraoxonase 1), glucose metabolism (*e.g.* IGF1, insulin-like growth factor 1 (somatomedin C); and HFE, hemochromatosis), oxidative stress (*e.g.* SOD1, superoxide dismutase 1, soluble; PARP, poly (ADP-ribose) polymerase 1; and GSTT1, glutathione S-transferase theta 1), mitochondrial mutation, premature aging syndromes such as Werner

syndrome (*i.e.* WRN), and telomere length have been considered with respect to both age-related diseases and longevity (Atzmon *et al.*, 2006; Capri *et al.*, 2006; Christensen *et al.*, 2006; Salvioli *et al.*, 2006).

Genes involved in increased lifespan may play protective roles in the development of age-related diseases. The presence of deleterious genotypes of age-related diseases may be a surrogate indicator of aging. Genes associated with aging-related diseases often have pleiotropic effects, while epistatic interactions affect human longevity (Christensen *et al.*, 2006). For instance, significant decreases of paraoxonase (*i.e.* PON1) activity on the surface of HDL have been observed in both diabetes mellitus (DM) and cardiovascular disease (CVD) (Mackness *et al.*, 2004). Although association between DM and cancer remains controversial, an increased risk of lung cancer has been observed among diabetic Korean women (hazard ratio 1.39, 95% confidence interval 1.10-1.76) (Jee *et al.*, 2005) and such experimental evidence has suggested that both insulin and insulin-like growth factors (IGFs) can stimulate tumor cell proliferation (Rousseau *et al.*, 2006).

Because the difficulty in finding centenarians or nonagenarians limits the replication of genome-wide association (GWA) studies using an extremely large number of randomly spaced markers, we adopted a high-throughput candidate gene approach to identify susceptibility variants controlling lifespan in Koreans.

Results

Among 565 polymorphic SNPs analyzed, 102 SNPs (18%) with MAF < 2%, HWE $P < 10^{-4}$, or CR < 95% in the young age group did not pass the threshold of our quality test for genotypes. Thus, a total of 463 informative markers located in 176 genes were included in the subsequent analyses. For the comparison of the individuals above age 90 versus the young healthy controls, our study had a sample size adequate to reach a power of 80% to significantly detect a genotypic OR over 1.7/2.4 (heterozygote/ homozygote) under the assumptions shown above (see the Methods). The result obtained from centenarian women was shown in Supplemental Data Tables.

Allelic- and genotypic association

A total of 43 SNPs located in 33 genes (18.8% of 176 candidate genes) had a nominally significant association with lifespan ($P < 0.05$) in the allelic χ^2

tests (data not shown). Among these, 34 SNPs of 28 genes showed a significant allelic OR and 7 genes had a P value less than 0.01 in at least one age/gender group (*i.e.* *IL12RB2*, interleukin 12 receptor, beta 2; *PPP1R14C*, protein phosphatase 1, regulatory (inhibitor) subunit 14C; *EGFR*, epidermal growth factor receptor; *PAX4*, paired box 4; *MMP1*, matrix metalloproteinase 1 (interstitial collagenase); *PPP1R1A*, protein phosphatase 1, regulatory (inhibitor) subunit 1A; and *ALDH2*, aldehyde dehydrogenase 2 family (mitochondrial)). A total of 128 SNPs located in 72 genes (41% of 176 genes) were statistically associated with one or more age/gender groups ($P < 0.05$) using regression models for five modes of inheritance (data not shown). Among these, the genotypic ORs of 95 SNPs located in 68 genes were nominally significant in at least one of age/gender groups and 13 genes had a P value less than 0.01 (Table 1). We present the P values corrected by genomic control where the λ is greater than 1 (*i.e.* female centenarians, log-additive model, $\lambda = 1.05$). As shown in Table 1, the genotypic effect on longevity was stronger than the corresponding allelic effect in either direction (*i.e.* either increased or decreased lifespan).

A total of 27 SNPs located in 25 genes showed evidence for association with lifespan in both gender groups (data not shown). Although the allelic effects of the matrix Gla protein (*MGP*) and the tumor necrosis factor (ligand) superfamily, member 11 (*TNFSF11*) were insignificant, these genes showed notable genotypic effects in the overdominant model. As shown in Table 1, TG heterozygotes of the rs1054016 were observed at lower frequencies among the old age group compared to the young age group (OR 0.45 and 0.32 and 0.30; $P = 4 \times 10^{-4}$ and 0.001 for each group of nonagenarians adjusted for gender and males ≥ 90 , respectively). All 27 genes consistently showed evidence for association with lifespan in both allele- and genotype-based analyses. In particular, six genes showing significant association with a P value less than 0.01 (*ADCY2*, adenylyl cyclase 2 (brain), *PPP1R1A*; *LYN*, V-yes-1 Yamaguchi sarcoma viral related oncogene homolog; *PAX4*, *ALDH2*, *IL12RB2*) attracted considerable attention as candidate genes affecting lifespan in Koreans. The results revealed significant gender differences in genetic associations with human longevity.

Nonagenarian men

A total of 43 SNPs located in 31 genes including four missense mutations located in each of euka-

ryotic translation initiation factor 4 gamma, 1 (*EIF4G1*), ATP-binding cassette, sub-family C (CFTR/MRP), member 8 (*ABCC8*), *ALDH2*, and telomerase-associated protein 1 (*TEP1*) showed evidence for genotypic association and 9 genes yielded P values less than 0.01 (Table 1B). By way of example, a male with an A allele of the rs671 (Lys504Glu) located in the *ALDH2* increased the chance of living longer than 90 yr by 2.11 times compared to men having a G allele ($P = 0.009$) and this SNP fitted best under a dominant model (OR 2.63, $P = 0.004$). No allelic effect of the rs757110 (Ser1369Ala) in the *ABCC8* gene was identified; however, the effect of the TG heterozygous genotype was greater than that of the rare GG homozygous type (OR 3.12, $P = 6 \times 10^{-4}$).

Nonagenarian women

A total of 37 SNPs located in 31 genes showed evidence for genotypic associations in either age group of women (*i.e.* women aged over 90 yr and centenarian women). Genes such as *PPP1R14C* ($P = 0.005$) and the *EGFR* ($P = 0.007$) had lower allelic P values in the centenarian group with a small sample size (data not shown), while *IL12RB2* was significant only in the younger group with more subjects ($P = 0.008$). However, the rare T allele of rs2030071 was not sufficiently common to determine the genetic model (Table 1C).

Haplotypic association

Among the 69 genes that showed significant genotypic association, we found significant association with longevity in 12 genes by using tagSNPs for 32 genes with at least 3 genotyped SNPs. Among nine genes that consistently yielded significant evidence of allelic, genotypic, and haplotypic associations with longevity (*i.e.* *HK2*, hexokinase 2; *ERBB4*, v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian); *PCSK1*, proprotein convertase subtilisin/kexin type 1; *ITK*, IL2-inducible T-cell kinase; *EGFR*; *PAX4*; *LPL*, lipoprotein lipase; *LYN*; and *CDK10*, cyclin-dependent kinase 10), only the *PCSK1* ($P = 0.008$), *EGFR* ($P = 0.003$), *PAX4* ($P = 0.008$), and *LYN* ($P = 0.002$) remained significant after the Bonferroni correction for multiple testing of haplotypes in each gene (Table 2). Figure 1 summarizes the $-\log_{10}(P)$ value obtained from three different analyses of association for these four genes.

The presence of the G allele of rs155979 in the CGA haplotype of *PCSK1* decreased lifespan among the group of nonagenarian adjusted by sex (OR 0.54). The G allele of rs712700 increased

Table 1. Results from the allele- and genotype-based analyses for individual SNPs showing evidence for association with longevity ($P < 0.01$).

Gene Location	N ^a	SNP	Function ^b	Allele ^c Model ^d	Number		OR (95%CI)	P value
					Case	Control		
<i>A. All ≥ 90 adjusted by sex</i>								
<i>ADCY2</i>	2	rs326146	Ile881Val	G/A	171/101	305/123	1.46 (1.05-2.03)	0.024
5p15.3				Additive	58/55/23	108/86/18	1.43 (1.03-1.97)	0.030
		rs326175	3' UTR	C/G	161/113	293/137	1.53 (1.11-2.11)	0.009
				Additive	50/61/26	96/98/19	1.53 (1.11-2.12)	0.009
<i>PCSK1</i>	3	rs155979	2 kb ~ 5'	C/G	225/49	315/115	0.61 (0.42-0.90)	0.012
5q15-q21				Dominant	95/35/7	113/85/15	0.53 (0.33-0.83)	0.005
<i>PIK3C2G</i>	2	rs4237973	2 kb ~ 5'	T/C	190/82	328/100	1.51 (1.06-2.14)	0.022
12p12				Dominant	60/70/6	122/82/8	1.81 (1.16-2.82)	0.009
<i>MGP</i>	1	rs1800802	2 kb ~ 5'	T/C	197/77	281/149	0.77 (0.55-1.08)	0.132
12p13.1-p12.3				Overdominant	75/47/15	45/54/7	0.44 (0.25-0.77)	0.003
<i>PPP1R1A</i>	2	rs1691595	2 kb ~ 5'	T/C	244/30	405/21	2.44 (1.35-4.42)	0.003
12q13.2				Additive	110/24/3	191/21/0	2.35 (1.31-4.20)	0.004
<i>TNFSF11</i>	1	rs1054016	3' UTR	G/T	110/162	183/239	1.09 (0.80-1.50)	0.579
13q14				Overdominant	57/48/31	62/114/33	0.45 (0.28-0.70)	4×10^{-4}
<i>B. Male ≥ 90</i>								
<i>PCSK1</i>	3	rs155979	2 kb ~ 5'	C/G	92/20	186/76	0.53 (0.31-0.92)	0.025
5q15-q21				Dominant	39/14/3	63/56/10	0.42 (0.21-0.81)	0.008
<i>PAX4</i>	3	rs327519	3' UTR	A/G	101/9	256/4	5.70 (1.72-18.94)	0.004
7q32				Additive	47/7/1	124/4/0	5.00 (1.50-16.68)	0.005
		rs712700	3' UTR	A/G	62/44	184/70	1.87 (1.16-3.00)	0.010
				Additive	19/24/10	66/49/10	1.82 (1.13-2.93)	0.013
		rs327517	Synonymous	A/G	103/9	254/4	5.55 (1.67-18.42)	0.005
				Additive	48/7/1	123/4/0	4.86 (1.46-16.20)	0.005
<i>LYN</i>	10	rs1980042	Intron	T/C	80/32	220/42	2.10 (1.24-3.55)	0.006
8q13				Dominant	29/22/5	94/30/5	2.50 (1.30-4.80)	0.006
		rs1979128	Intron	A/G	80/32	220/42	2.10 (1.24-3.55)	0.006
				Dominant	29/22/5	94/30/5	2.50 (1.30-4.80)	0.006
<i>JAK2</i>	1	rs2274471	Intron	T/C	83/29	213/47	1.58 (0.93-2.68)	0.088
9p24				Recessive	34/15/7	84/41/3	5.95 (1.48-23.95)	0.008
<i>ABCC8</i>	2	rs757110	Ser1369Ala	T/G	63/45	150/112	0.96 (0.61-1.51)	0.848
11p15.1				Overdominant	13/37/4	47/53/29	3.12 (1.59-6.12)	6×10^{-4}
<i>PPP1R1A</i>	2	rs1691595	2 kb ~ 5'	T/C	97/15	248/12	3.20 (1.44-7.07)	0.004
12q13.2				Additive	44/9/3	117/12/0	2.77 (1.28-5.96)	0.008
<i>ALDH2</i>	1	rs671	Lys504Glu	G/A	83/27	227/35	2.11 (1.20-3.70)	0.009
12q24.2				Dominant	30/23/2	98/29/2	2.63 (1.35-5.13)	0.004
<i>TNFSF11</i>	1	rs1054016	3' UTR	G/T	48/62	111/143	1.00 (0.64-1.57)	0.991
13q14				Overdominant	23/16/16	36/70/19	0.32 (0.16-0.64)	0.001
<i>CDK10</i>	3	rs2075880	Intron	G/A	38/70	120/142	1.56 (0.98-2.48)	0.061
16q24				Dominant	25/20/9	34/70/25	0.42 (0.21-0.81)	0.009
<i>C. Female ≥ 90</i>								
<i>IL12RB2</i>	1	rs2030071	Intron	C/T	159/3	156/16	0.18 (0.05-0.64)	0.008
1p31.3-p31.2				Codominant	78/3/0	68/16/0	0.16 (0.05-0.59)	0.001

^aNumber of informative markers for the gene; ^bUTR, un-translated region; ^cAncestral alleles were shown in bold; ^dBest fitting mode of inheritance: Only the SNPs showing a significant genotypic OR from the most significant regression model ($P < 0.01$) for the gene were shown.

Table 2. Results from the haplotype-based analyses for candidate genes showing evidence for association with longevity ($P < 0.05$ /Number of haplotypes).

Gene	SNPs	Group P value ^b	Haplotypes	Freq	OR	95% CI
<i>PCSK1</i> 5q15-q21	rs3811952-rs155979 -rs3762985 (3/3) ^a	Age \geq 90	CCA	0.302	1.00	Reference
		0.628	CCC	0.247	0.89	0.56-1.43
		0.008	CGA	0.228	0.54	0.34-0.85
		0.752	GCA	0.156	0.92	0.53-1.58
		0.996	GCC	0.060	1.00	0.48-2.09
		0.246	Rare	0.007	4.59	0.35-60.17
<i>EGFR</i> 7p12	rs2072454-rs2293347- rs3807362-rs884225 (4/4)	Female \geq 100	CACA	0.310	1.00	Reference
		0.420	TGCG	0.247	0.67	0.25-1.77
		0.003	CGCA	0.187	4.11	1.64-10.31
		0.339	CGCG	0.150	1.60	0.61-4.18
		0.000	TACA	0.052	0.00	—
		0.329	TGCA	0.033	0.31	0.03-3.27
		0.000	CATA	0.022	0.00	N.a.
<i>PAX4</i> 7q32	rs327519-rs712700 (2/3)	Male \geq 90	AA	0.683	1.00	Reference
		0.130	AG	0.282	1.49	0.89-2.51
		0.008	GG	0.035	5.14	1.54-17.14
<i>LYN</i> 8q13	rs1450740-rs1450741- rs1562818- rs1980042- rs1027989-rs1546518- rs3968371 (7/10)	Age \geq 90	CTCTGCA	0.504	1.00	Reference
		0.141	ATCCGCA	0.119	1.46	0.88-2.41
		0.741	CTTTGCA	0.114	1.10	0.64-1.89
		0.325	CCCTGCA	0.077	1.37	0.73-2.58
		0.155	ATCTGCA	0.052	1.77	0.81-3.88
		0.002	CTCCGCA	0.048	3.72	1.64-8.45
		0.900	CTCCGGG	0.037	0.95	0.40-2.24
		0.438	ATCTCCA	0.016	0.57	0.14-2.36
		0.262	ATCTCGA	0.014	0.31	0.04-2.41
		0.227	ATTTGCA	0.012	2.92	0.51-16.59
		0.576	Rare	0.010	1.61	0.31-8.44
		Male \geq 90	CTCTGCA	0.519	1.00	Reference
		0.787	ATTTGCA	0.014	1.36	0.15-12.76
		0.411	ATCTCGA	0.024	0.42	0.05-3.29
0.000	ATCTCCA	0.012	0.00	—		
0.109	ATCTGCA	0.053	2.52	0.81-7.82		
0.181	ATCCGCA	0.123	1.63	0.80-3.35		
0.390	CTTTGCA	0.105	1.46	0.62-3.43		
0.150	CTCCGGG	0.028	2.48	0.72-8.56		
0.004	CTCCGCA	0.040	6.20	1.80-21.30		
0.518	CCCTGCA	0.076	1.38	0.52-3.67		
0.516	Rare	0.007	2.45	0.16-36.74		

^aNumber of tagSNPs among all SNPs genotyped for a candidate gene, ^bThe P value for each haplotype using a general linear model (GLM).

lifespan in the GG haplotype of *PAX4* (OR 5.14) among nonagenarian men. Although only the rs1980042 (C) allele in haplotypes of *LYN* seemed

to confer longevity, the effect size of 2.5 for the CC genotype increased to 6.2 for the CTCCGCA haplotype among nonagenarian men. Among female

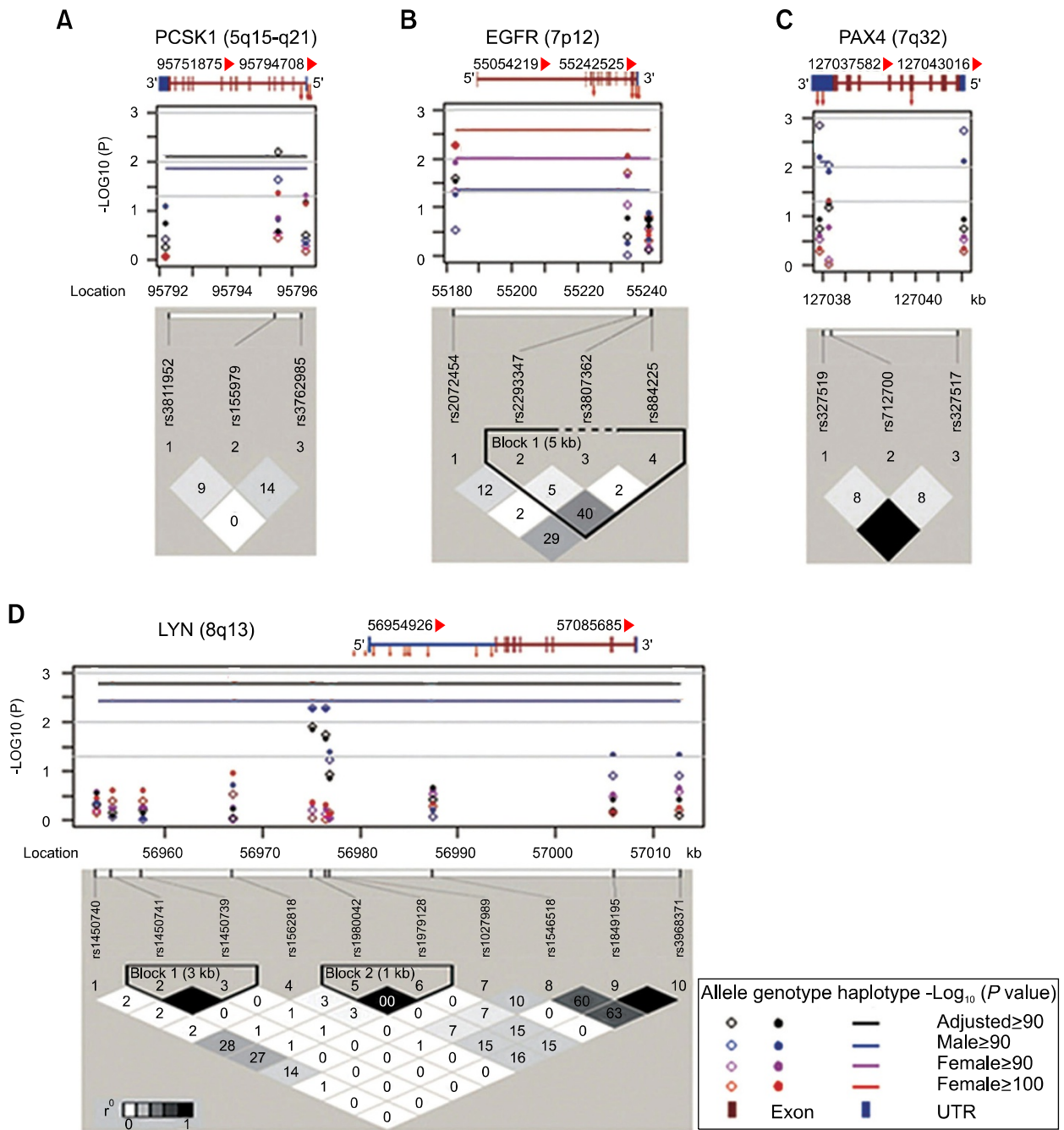


Figure 1. Summary results with gene and SNP information for four genes that consistently showed evidence for association with lifespan in allelic, genotypic, and haplotypic analyses.

centenarians, the OR of 3.12 for the rs2293347 (GG) genotype increased to 4.11 for the CGCA haplotype of the *EGFR* gene when compared with the CACA reference haplotype.

Gene-gene interactions

Supplement Data Table S2 presents significant *P*

values for the epistatic interaction LRT under a best fitting model among five genetic models tested for each age/gender group. Supplement Data Table S3 shows the ORs and their 95% CIs for the genotype combinations composed of each candidate SNP pair for two genes along with the numbers of cases and controls. For each age/gender group, 6, 1, 3, and 2 pairs of loci for each group

yielded statistical significance in both the regression model and ORs for interaction. In particular, the interaction of platelet derived growth factor C (*PDGFC*) and glycogen synthase 2 (liver) (*GYS2*) for males older than 90 yr as well as the interaction of *TNFSF11* and *CDK10* in the group adjusted by gender showed a dose-response relationship with increasing copies of the target allele for each SNP.

Discussion

It is well known that genetic variation exists across ethnic groups. Indeed, significant association of genes for human leukocyte antigen (*HLA*), *ACE*, *APOE*, *MTTP*, and *CETP* with human longevity in Caucasian populations have not been replicated in studies using Korean centenarians (Choi *et al.*, 2003). The Korean longevity study carried out since 1998 reported significant differences in the occurrence of centenarians by gender, location, lifestyle, and diet (Lee *et al.*, 2005). In 2005, the ratios of the Jeolla province to the Kyongsang province were 1:0.97 for nonagenarians and 1:1.27 for centenarians (KNSO, 2005). Interestingly, a reverse relationship between the incidence of cancer and degree of longevity has been observed across provinces, indicating that prevention of cancers may be an important strategy to enhance lifespan (Kwon and Park, 2005). This notion could be applied to other age-related diseases such as DM and CVD.

According to the theory of evolution, differential survival reflects the effect of deleterious genotypes of age-related diseases. However, some longevity genes (*e.g.* *klotho* that is associated with low HDL) buffer the deleterious effect of age-related diseases genes (*e.g.* lipoprotein, Lp(a), a susceptibility gene for vascular disease), which may explain an observed paradoxical increase of deleterious genotypes in centenarians (Bergman *et al.*, 2007). Lunetta *et al.* (2007) conducted a genome-wide association study of age-related phenotypes using Affymetrix 100 K SNP Chips in the context of a cross-sectional study design. However, none of the associations with 5 aging traits achieved genome-wide significance using 1,345 Framingham Study participants.

In the current study, a total of 60 genes out of 179 candidates were nominally associated ($P < 0.05$) with longevity in at least one of age/gender groups. Among eleven genes that consistently yielded statistical evidence of allelic, genotypic and haplotypic association with lifespan, only four genes remained significant after the Bonferroni correction for haplotypes. The protein encoded by

the *PCSK1* gene, located in 5q15-q21, is a proinsulin-processing enzyme that plays a key role in regulating insulin biosynthesis (Ohagi *et al.*, 1996). Mutations in this gene are associated with obesity, tumorigenesis, and metastasis (Tzimas *et al.*, 2005). The *PAX4* gene, located in 7q32, is a member of the paired box family of transcription factor that mediates differentiation of insulin-producing beta-cells in the pancreas and plays critical roles during cancer growth (Li *et al.*, 2006). *EGFR* is a transmembrane glycoprotein that functions as a tyrosine protein kinase (Kondo and Shimizu, 1983). Mutations in the *EGFR* gene, located in the 7p12.3-p12.1 region, have oncogenic potential in the development of non-small cell lung cancer (Tai *et al.*, 2006). The *LYN*, located on 8q13, encodes a tyrosine kinase that is essential in immunoglobulin-mediated signaling, particularly in establishing B cell tolerance (Parravicini *et al.*, 2002). Even aging mice with a Lyn up/up phenotype do not display hematologic malignancies, while Lyn^{-/-} mice are more susceptible to tumorigenesis (Harder *et al.*, 2001). To the best of our knowledge, this is the first reported evidence associating these genes with human longevity.

Several limitations of this study deserve mention. First, the influence of environmental factors could not be controlled in this study due to lack of information in the control group. Second, we could not evaluate the function of any SNP that was not a missense mutation but did have statistical evidence for association. Third, given the small number of centenarians, the statistical power may not be sufficient to significantly detect genes with weak effect, particularly, for the gene-gene interaction. There are further issues concerning the interpretation of the findings using the over-dominant model and the control group consisted of 12-14 yr old middle school students. To eliminate the impact of different substructure between the long-lived group and the young control group, the *P* values were corrected by genomic control for each age-gender group. In the nonagenarian group, the Bonferroni threshold for $P < 0.05$ with 463 SNPs was 1.08×10^{-4} and the significance level was relaxed by 1000 permutations to 1.69×10^{-4} , 1.56×10^{-4} , 2.16×10^{-4} , 1.50×10^{-4} , and 2.42×10^{-4} under the co-dominant, dominant, recessive, over-dominant, and log-additive models, respectively. None of the SNPs achieved genome-wide significance. However, the underlying LD structure among the markers reduced the number of effectively independent tests, and so the genome-wide threshold of significance may be too stringent particularly in the context of candidate gene approach (Rao and Gu, 2001).

For genetic variants with strong effects, relatively small sample size is sufficient to achieve statistical power to detect association. The results in this study warrant further replication studies. However, the expectation for replication of results may need to be relaxed in studies of extremely old subjects, who are rarely observed. Under such circumstances, additional information gathered from laboratory techniques, bioinformatics, and a priori insights of biological pathways could be used to provide plausibility for interpreting genetic association findings (Chanock *et al.*, 2007). To increase statistical power of a test for gene by gene or gene by environment interaction, international collaboration is essential due to the rarity of centenarians. Continuing efforts may be necessary to find a functional variant in the associated region.

Nevertheless, the results from various statistical analyses consistently supported that the presence of the target allele of a potent SNP conferred a greater likelihood of longevity in Koreans. The effects of gender mediated genetic association with the probability of achieving longevity, as shown in previous studies (Franceschi *et al.*, 2000) and the observed gender difference in our findings supports, in part, the gender-specific probability of achieving longevity. Interestingly, a few candidate gene variants with no or insignificant allelic effect revealed strong evidence for the allele with a large heterozygote effect on longevity (*e.g.* *MGP*, *TNFSF11*, and *ABCC8*).

As far as we know, this is the first association study of longevity susceptibility genes adopting a high-throughput candidate gene approach in nonagenarians and centenarians who rarely appear in Korea. An understanding of the mechanisms underlying aging and longevity by discovering novel susceptibility genes may lay the foundations for prediction, prevention, and treatment of age-related diseases. Thus, the findings of the current study may provide a starting point to unravel genetic factors controlling longevity for future studies.

Methods

Subjects

We identified 137 long-lived individuals over 90 yr old including 35 centenarians from the Korean Centenarian Study, which began in 1999 (Choi *et al.*, 2003). We also selected 213 healthy middle school students from individuals who visited the Samsung Seoul Hospital for psychiatric tests in 2003 to be used for comparison. Eight control samples from a three-generation family were included to test genotype quality. After informed consent was obtained from all subjects, a clinician interviewed study subjects and

drew a 10 ml fasting venous blood sample in an EDTA tube. The institutional review board (IRB) of the Samsung medical center reviewed the research protocols and approved this study. The gender and age composition of study subjects were shown in Supplemental Data Table S1.

Candidate genes and SNP selection

Candidate genes, mostly for DM, CVD, and cancers, were selected based on searches for their functions in public databases (OMIM, Gene, *etc.*) and previous literature reviews. SNP markers within or near candidate genes were identified from the literature, the dbSNP database (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), the HapMap database (<http://www.hapmap.org/>), and Vector NTI[®] software (Invitrogen, *Inc.*, Carlsbad, CA). We included 565 informative single nucleotide polymorphism (SNP) markers with high design scores provided by Illumina, *Inc.* (San Diego, CA) and heterozygosity above 0.05 using 300 independent healthy Korean samples. Information for these 565 SNPs located in 194 genes is presented in Supplemental Data Table S2.

DNA preparation and SNP genotyping

Genomic DNA samples were prepared from peripheral blood using the QIAamp DNA Blood Maxi kit following the manufacturer's instructions (Qiagen, *Inc.*, Valencia, CA). For each sample, an average concentration of DNA from three repeated measures was determined using a SpectraMax Plus 384 spectrophotometer (Molecular Devices, *Inc.*, Sunnyvale, CA). A 250 ng aliquot of DNA was genotyped for SNP markers using Golden-Gate chemistry on Sentrix[®] Array Matrices (Illumina, *Inc.*) at the Samsung Biomedical Research Institute. Genotypes showing significant statistical evidence for association with longevity were confirmed using an ABI PRISM 3100 sequencer (Applied Biosystems, *Inc.*, Foster City, CA).

Statistical analysis

At each SNP, the genotyping call rate (CR > 95%), minor allele frequency (MAF > 2%), and deviation from Hardy-Weinberg equilibrium (HWE $P < 0.0001$) were separately computed in the long-lived group and the young healthy group. Ancestral alleles for each SNP were determined using the dbSNP database. The SNPs-based whole genome association studies (SNPassoc) program implemented in the free downloadable statistical software environment R v. 2.6.2 (<http://www.r-project.org>) was used to perform the quality tests for genotypes (González *et al.*, 2007). We performed preliminary analyses of linkage disequilibrium (LD) patterns using both D' and r^2 and then selected a set of tagging SNPs within a gene for the haplotype-based analysis using Haploview v. 4.0 (<http://www.broad.mit.edu/mpg/haploview/index.php/>).

We performed subsequent statistical analyses for each of four groups stratified by age and gender (nonagenarians, male nonagenarians, female nonagenarians, and female centenarians). We assessed allelic association with the χ^2 -tests for individual SNPs and calculated the allelic

odds ratio (OR) and 95% confidence interval (CI) in both adjusted and stratified analyses using publicly available subroutines 'genassoc' and 'gamenu' in the STATA/SE v. 9 software package (<http://www.gene.cimr.cam.ac.uk/clayton/software/stata/>).

We performed genotypic association tests and logistic regression analyses implemented in the SNPAssoc package in five different genetic models: codominant, dominant, recessive, over-dominant, and log-additive. To account for the impact of substructure or genetic relatedness among the subjects, we estimated the 'inflation factor (λ)' for each age-gender group and corrected *P* values by genomic control (Devlin and Roeder, 1999).

We used general linear models for the regression of longevity on ambiguous haplotypes. Rare haplotypes with an allele frequency of less than 1% were combined in a group to be analyzed. We applied logistic regression models (LRMs) to perform two-way interaction analyses between two SNPs for each gene (Briollais *et al.*, 2007). For each of four age/gender groups, we screened the most promising pairs of SNPs that yielded $P < 0.01$ for the interaction likelihood ratio test (*i.e.* $LRT_{ij} = -2(\log L_{full(i,j)} - \log L_{additive(i,j)})$) under the five genetic models. Subsequently, we obtained the ORs and 95% CIs under the best fitting model for both marginal effects and interactions of the pair of SNPs.

We computed the statistical power of our sample sizes for each genotype using the web browser 'Genetic Power Calculator' (<http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html>) under an additive model with the assumption of disease allele frequencies of 0.3, *D'* of 1, and a longevity prevalence of 0.13% and 0.002% for the Korean nonagenarians and centenarians, respectively (Purcell *et al.*, 2003).

Supplemental data

Supplemental Data include three tables and can be found with this article online at http://e-emm.or.kr/article/article_files/SP-41-11-01.pdf.

Acknowledgements

We appreciate the technical support of our staffs: En Jin Han, Hanees Kwon and Yonghun Lim at the Center for Genome Research of the Samsung Biomedical Research Institute and also gratefully acknowledge the hard work of those involved in collecting samples. This work was supported by the Research Program of dual regulation mechanisms of aging and cancer from the Korea Science and Engineering Foundation (M1075604000107N5604-00110); the Korea Research Foundation for Health Science; and Hallym University Research Fund, 2009 (HRF-2009-000).

References

Atzmon G, Rincon M, Schechter CB, Shuldiner AR, Lipton RB, Bergman A, Barzilai N. Lipoprotein genotype and conserved pathway for exceptional longevity in humans. *PLoS Biol* 2006;4:e113

Bergman A, Atzmon G, Ye K, MacCarthy T, Barzilai N. Buffering mechanisms in aging: a systems approach toward uncovering the genetic component of aging. *PLoS Comput Biol* 2007;3:e170

Briollais L, Wang Y, Rajendram I, Onay V, Shi E, Knight J, Ozcelik H. Methodological issues in detecting gene-gene interactions in breast cancer susceptibility: a population-based study in Ontario. *BMC Med* 2007;5:22

Capri M, Salvioli S, Sevini F, Valensin S, Celani L, Monti D, Pawelec G, De Benedictis G, Gonos ES, Franceschi C. The genetics of human longevity. *Ann NY Acad Sci* 2006; 1089:516-37

Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G., Hirschhorn JN, Abecasis G., Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF Jr, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LJ, Phimister EG., Rice JP, Roberts J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS. Replicating genotype-phenotype associations. *Nature* 2007;447: 655-60

Choi YH, Kim JH, Kim DK, Kim JW, Kim DK, Lee MS, Kim CH, Park SC. Distributions of ACE and APOE polymorphisms and their relations with dementia status in Korean centenarians. *J Gerontol A Biol Sci Med Sci* 2003;58: 227-31

Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006;7:436-48

Devlin B, Roeder K. Genomic Control for Association Studies. *Biometrics* 1999;55:997-1004

Franceschi C, Motta L, Valensin S, Rapisarda R, Franzone A, Berardelli M, Motta M, Monti D, Bonafé M, Ferrucci L, Deiana L, Pes GM, Carru C, Desole MS, Barbi C, Sartoni G., Gemelli C, Lescai F, Olivieri F, Marchegiani F, Cardelli M, Cavallone L, Gueresi P, Cossarizza A, Troiano L, Pini G, Sansoni P, Passeri G, Lisa R, Spazzafumo L, Amadio L, Giunta S, Stecconi R, Morresi R, Viticchi C, Mattace R, De Benedictis G, Baggio G. Do men and women follow different trajectories to reach extreme longevity? Italian Multicenter Study on Centenarians (IMUSCE). *Aging (Milano)* 2000;12: 77-84

González JR, Armengol L, Solé X, Guinó E, Mercader JM, Estivill X, Moreno V. SNPAssoc: an R package to perform whole genome association studies. *Bioinformatics* 2007;23: 644-5

Harder KW, Parsons LM, Armes J, Evans N, Kountouri N, Clark R, Quillici C, Grail D, Hodgson GS, Dunn AR, Hibbs ML. Gain- and loss-of-function Lyn mutant mice define a critical inhibitory role for Lyn in the myeloid lineage. *Immunity* 2001;15:603-15

Jee SH, Ohrr H, Sull JW, Yun JE, Ji M, Samet JM. Fasting serum glucose level and cancer risk in Korean men and women. *JAMA* 2005;293:194-202

Kondo I, Shimizu N. Mapping of the human gene for epidermal growth factor receptor (EGFR) on the p13 leads to q22 region of chromosome 7. *Cytogenet Cell Genet*

1983;35:9-14

Korea National Statistical Office: Statistical Annual Report 2005. Daejeon, Republic of Korea

Kwon IS, Park SC. Gender-Specific and Age-Dependent Changes in Health Status and Medical Characteristics of Korean Centenarians. *Kor J Gerontol* 2005;15:10-25

Lee MS, Yeo EJ, Kwak CS, Kim K, Choi YH, Kwon IS, Kim CH, Park SC. Gender differences in health and nutritional status of Korean centenarians. *Kor J Gerontol* 2005;15:65-75

Li Y, Nagai H, Ohno T, Ohashi H, Murohara T, Saito H, Kinoshita T. Aberrant DNA demethylation in promoter region and aberrant expression of mRNA of PAX4 gene in hematologic malignancies. *Leuk Res* 2006;30:1547-53

Lunetta KL, D'Agostino RB Sr, Karasik D, Benjamin EJ, Guo CY, Govindaraju R, Kiel DP, Kelly-Hayes M, Massaro JM, Pencina MJ, Seshadri S, Murabito JM. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. *BMC Med Genet* 2007;8(Suppl 1):S13

Mackness M, Mackness B. Paraoxonase 1 and atherosclerosis: is the gene or the protein more important? *Free Radic Biol Med* 2004;37:1317-23

McGue M, Vaupel JW, Holm N, Harvald B. Longevity is moderately heritable in a sample of Danish twins born 1870-1880. *J Gerontol* 1993;48:B237-44

Ohagi S, Sakaguchi H, Sanke T, Tatsuta H, Hanabusa T, Nanjo K. Human prohormone convertase 3 gene: exon-intron organization and molecular scanning for mutations in Japanese subjects with NIDDM. *Diabetes* 1996;45:897-901

Parravicini V, Gadina M, Kovarova M, Odom S, Gonzalez-Espinosa C, Furumoto Y, Saitoh S, Samelson LE, O'Shea JJ, Rivera J. Fyn kinase initiates complementary signals required for IgE-dependent mast cell degranulation. *Nature*

Immun 2002;3:741-8

Perls TT, Bubrick E, Wager CG, Vijg J, Kruglyak L. Siblings of centenarians live longer. *Lancet* 1998;351:1560

Puca AA, Daly MJ, Brewster SJ, Matisse TC, Barrett J, Shea-Drinkwater M, Kang S, Joyce E, Nicoli J, Benson E, Kunkel LM, Perls T. A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc Natl Acad Sci USA* 2001;98:10505-8

Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149-50

Rao DC, Gu C. False positives and false negatives in genome scans. *Adv Genet* 2001;42:487-98

Rousseau MC, Parent ME, Pollak MN, Siemiatycki J. Diabetes mellitus and cancer risk in a population-based case-control study among men from Montreal, Canada. *Int J Cancer* 2006; 118:2105-9

Salvioli S, Olivieri F, Marchegiani F, Cardelli M, Santoro A, Bellavista E, Mishto M, Invidia L, Capri M, Valensin S, Sevini F, Cevenini E, Celani L, Lescai F, Gonos E, Caruso C, Paolisso G, De Benedictis G, Monti D, Franceschi C. Genes, ageing and longevity in humans: problems, advantages and perspectives. *Free Radic Res* 2006;40:1303-23

Tai AL, Sham JS, Xie D, Fang Y, Wu YL, Hu L, Deng W, Tsao GS, Qiao GB, Cheung AL, Guan XY. Co-overexpression of fibroblast growth factor 3 and epidermal growth factor receptor is correlated with the development of nonsmall cell lung carcinoma. *Cancer* 2006;106:146-55

Tzimas GN, Chevet E, Jenna S, Nguyen DT, Khatib AM, Marcus V, Zhang Y, Chrétien M, Seidah N, Metrakos P. Abnormal expression and processing of the proprotein convertases PC1 and PC2 in human colorectal liver metastases. *BMC Cancer* 2005;5:149