EGFR L858R Mutation and Polymorphisms of Genes related to Estrogen Biosynthesis and Metabolism in Never-smoking Female Lung Adenocarcinoma Patients

Shi-Yi Yang[1,2], Tsung-Ying Yang[3,4], Kun-Chieh Chen[3,5], Yao-Jen Li[1,2] , Kuo-Hsuan Hsu[3], Chi-Rne Tsai[6,7], Chih-Yi Chen[8,9], Chung-Ping Hsu[10,11], Jiun-Yi Hsia[10], Cheng-Yen Chuang[3,10], Ying-Huang Tsai[12], Kuan-Yu Chen[13], Ming-Shyan Huang[14], Wu-Chou Su[15], Yuh-Min Chen[11, 16], Chao A. Hsiung[17], Chen-Yang Shen[18,19,20,21], Gee-Chen Chang[3,5,9,11], Pan-Chyr Yang[13], Chien-Jen Chen[1,2]

Affiliation of authors:

[1]Genomics Research Center, Academia Sinica, Taipei, Taiwan; [2]Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan; [3]Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan; [4]Institute of Medical and Molecular Toxicology, Chung Shan Medical University, Taichung, Taiwan; [5]Institute of Biomedical Sciences, National Chung-Hsing University, Taichung, Taiwan; [6]Department of Pediatrics, Taichung Veterans General Hospital, Taichung, Taiwan; [7]Institute of Molecular Biology, National Chung-Hsing University, Taichung, Taiwan; [8]Cancer Center, China Medical University Hospital, Taichung, Taiwan; [9]School of Medicine, China Medical University, Taichung, Taiwan;[10]Division of Thoracic Surgery, Department of Surgery, Taichung Veterans General Hospital, Taichung, Taiwan; [11]School of Medicine, National Yang-Ming University, Taipei, Taiwan; [12]Department of Pulmonary and Critical Care, Chang Gung Memorial Hospital, Lincou, Taiwan; [13]Department of Internal Medicine, National Taiwan University Hospital and

National Taiwan University College of Medicine, Taipei, Taiwan; [14]Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan; [15]Department of Internal Medicine, National Cheng Kung University Hospital and College of Medicine, Tainan, Taiwan; [16]Chest Department, Taipei Veterans General Hospital Taipei, Taiwan; [17]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan; [18]Institute of Biomedical Sciences,[19]Taiwan Biobank and [20]Life Science Library, Academia Sinica, Taipei, Taiwan; [21]Graduate Institute of Environmental Science, China Medical University, Taichung, Taiwan;

Footnote:

Shi-Yi Yang, and Tsung-Ying Yang contributed equally to this work and should be considered joint first authors.

Running title:

EGFR L858R Mutation, Estrogen-related Genetic Polymorphisms

Keywords:

EGFR mutation, polymorphisms, estrogen biosynthesis and metabolism, never-smoker, lung adenocarcinoma

Correspondence to:

Gee-Chen Chang, Division of Chest Medicine, Department of Internal Medicine, Taichung Veterans General Hospital, No. 160, Section 3, Chung-Kang Rd, Taichung 40705, Taiwan. Phone: +886-4-23592525 ext. 3256, Fax: +886-4-23552590, e-mail: august@vghtc.gov.tw

and Chien-Jen Chen, Genomics Research Center, Academia Sinica, No. 128, Section 2, Academia Rd, Nankang, Taipei 11529, Taiwan, e-mail: cjchen@ntu.edu.tw

The sources of grant support:

Funding Supported by grants (DOH96-TD-G-111-012) from the Department of Health, Executive Yuan, and Academia Sinica, Taipei, Taiwan

Word count: 3755 words

Translational Relevance.

The efficacy of targeted therapy using gefitinib or erlotinib to non-small cell lung cancer patients depends on somatic mutation status of epidermal growth factor receptor (*EGFR*). This study explored whether genetic polymorphisms of genes related to estrogen biosynthesis and metabolism are associated with *EGFR* hotspot mutations. The number of risk alleles of *CYP17*, *CYP19A1*, *ERα* and *COMT* was found to be associated with an increasing odd ratio of *EGFR* L858R mutation in never-smokers. The genetic polymorphism of *ERα* was associated with various *EGFR* mutations in female never-smokers. The findings provide a clue for the genesis of *EGFR* mutations.

Abstract

Purpose:

To assess whether polymorphisms of genes related to estrogen biosynthesis and metabolism are associated with *EGFR* mutations.

Experimental Design:

We studied 617 patients with lung adenocarcinoma, including 302 never-smoking women. On the basis of multiple candidate genes approach, the effects of polymorphisms of *CYP17*, *CYP19A1, ERα* and *COMT* in association with the occurrence of *EGFR* mutations were evaluated using logistic regression analysis.

Results:

In female never-smokers, significant associations with *EGFR* L858R mutation were found for the $(TTTA)_n$ repeats in *CYP19A1* (odds ratio [OR], 2.6; 95% confidence interval [CI], 1.2-5.7 for one or two alleles with $(TTTA)_n$ repeats >7 compared to both alleles with $(TTTA)_n$ repeats ≦7), and the rs2234693 in *ERα* (OR,2.1; 95% CI, 1.1-4.0 for C/T and C/C genotypes compared to T/T genotype). The C/C genotype (vs. T/T genotype) of *ERα* was significantly associated with EGFR L858R mutation (OR, 3.0; 95% CI, 1.1-8.1), in-frame deletion (OR, 2.9; 95% CI, 1.1-7.6) and other mutations (OR, 4.3; 95% CI, 1.3-14.0). The genotype of *COMT* rs4680 was significantly associated with *EGFR* L858R mutation in female and male never-smokers showing OR's (95% CI) of 1.8 (1.0-3.2) and 3.6 (1.1-11.3), respectively, for genotypes G/A and G/G compared to genotype A/A. The number of risk alleles of *CYP17*, *CYP19A1*, *ERα* and *COMT* was associated with an increasing OR of *EGFR* L858R mutation in female never-smokers (P=.0002 for trend).

Conclusions:

The L858R mutation of *EGFR* is associated with polymorphisms of genes related to estrogen biosynthesis and metabolism in never-smoking female lung adenocarcinoma patients.

Word count of the Abstract: 248 words.

Introduction

Somatic mutations in the tyrosine kinase domain of epidermal growth factor receptor gene (*EGFR*), located from exon 18 to exon 23, were found in lung adenocarcinoma (1-3). The most common mutations are the in-frame deletion in exon 19 and a substitution of lysine for arginine mutation at amino acid position 858 (L858R) in exon 21. The efficacy of targeted therapies such as gefitinib or erlotinib to non-small cell lung cancer (NSCLC) patients depends on the presence of *EGFR* somatic hotspot mutations including L858R in exon 21 or in-frame deletion in exon 19. *EGFR* mutations have been found to be more frequent in adenocarcinoma than other NSCLC, in females than males, and in never-smokers than smokers (4). Furthermore, there was a sex difference in *EGFR* mutation subtypes (5, 6). The exon 19 in-frame deletion was more frequently associated with male gender, while exon 21 mutations were more frequent in females. The sex difference in *EGFR* mutation subtypes may result from the influence of sex hormone.

Estrogen and its metabolites may play some roles in cell proliferation and genotoxicity to induce breast cancer (7), and estrogen could also potentially promote lung cancer (8). Estrogen receptors were found to express in normal human lung tissue and NSCLC tissues (9-11). *EGFR* promoter containing imperfect estrogen-response elements could be bound by ER (12), and the administration of 17β-estradiol was found to increase the EGFR level and tyrosine kinase activity *in vivo* (12, 13). These findings suggest that sex hormones play a role in lung carcinogenesis.

One cohort study has found a higher proportion of *EGFR* hotspot mutation in female patients having experienced a longer fertile life, measured by an early age of menarche, the delayed menopause or the late first parity (14). This finding implies the period of exposure to endogenous hormone may be correlated with the occurrence of *EGFR* hotspot mutations.

Many studies have used candidate genes approach to assess whether the biosynthesis and metabolism of estrogen is associated with the susceptibility to hormone-related cancer, such as breast

cancer. Among these genes, the genetic polymorphisms of *CYP17* rs743572, *CYP19A1* $(TTTA)_n$, *COMT* rs4680, ER*α* rs2234693 are widely studied.

This study aimed to evaluate the associations between *EGFR* hotspot mutations in lung adenocarcinoma and genetic polymorphisms of *CYP17*, *CYP19A1*, *ERα,* and *COMT*.

Materials and Methods

This study included patients affected with lung adenocarcinoma, who were diagnosed between September 2002 and December 2007. This study was approved by the Joint Institutional Review Board (JIRB) at six teaching hospitals and tertiary referral centers in Taiwan which participated in the Genetic Epidemiological Study of Lung Adenocarcinoma (GELAC) study and IRB of Taichung Veterans General Hospital, Taiwan. Lung cancer histology was classified according to World Health Organization criteria (15). Demographic characteristics and lifestyle variables, including age, gender, and cigarette smoking status (never-smokers were defined as patients who had never smoked cigarette, while ever-smokers defined as those who were current or former smokers), were collected by questionnaire interview. Clinical data were abstracted using a special abstract form according to a standard protocol.

The Selection of Genes

On the basis of multiple candidate genes approach in the estrogen biosynthesis and metabolism pathways, and the selection of candidate genes in this study is based on their physiological importance and absence of isoforms. To this end, four genes, including *CYP17*, *CYP19A1*, *ERα* and *COMT*, were chosen. *CYP17* and *CYP19A1* are known to be key genes involved in the biosynthesis of estrogen (16, 17). *COMT* is the major gene to deactivate catecholestrogen (7). No isoforms of these three genes have been identified, and. thus, the contribution of these three genes in the biosynthetic and metabolic pathways could be considered unique and independent. The *ERα* serves as a receptor for estrone/estrodiol and facilitates the activation of downstream genes (18).

The selection of SNPs:

CYP17 rs743572, *CYP19A1* $(TTTA)_n$, *ERα* rs2234693, and *COMT* rs4680 were chosen as representative polymorphisms of these genes. Details of SNPs selection were described in the Supplementary methods.

DNA Extraction, Sequencing and Genotype Analysis:

Genomic DNA from tumor frozen specimen and paraffin-embedded tissues were extracted by QIAamp DNA Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's protocols. DNA sequencing of exons 18 to 21 of *EGFR* gene was amplified by polymerase chain reaction, subsequently DNA sequencing reaction and subjected to electrophoresis in the ABI PRISM 3130XL as described preciously (4).

The TTTA tetra-nucleotide repeats of the *CYP19A1* were determined by polymerase chain reaction using primers as described previously(19), and TaqMan assays were used for genotyping of the *CYP17*, *ERα*, and *COMT*. Details of genotyping procedures were described in the Supplementary methods and Supplementary Table 1.

Statistical Analysis

The associations between *EGFR* mutation status and demographic and clinical characteristics, including gender, age, cigarette smoking status (never vs. ever smoker), disease stage and study genotypes, were examined by chi-square test. The association between *EGFR* mutation status and sex after adjustment for smoking status was examined by Cochran-Mantel-Haenszel Statistics. The odds ratio (OR) with the corresponding 95% confidence intervals (CI) for each variable was estimated by the unconditional logistic regression model. Multivariate-adjusted OR (aOR) of *EGFR* mutation status associated with individual genotypes was assessed after adjustment for age.

Patients were classified into four groups by their *EGFR* mutation status: those without mutation in exons 18-21 were defined as "wild type", those with in-frame deletion amino acid positions 746-753 in exon19 as "in-frame deletion", those with a substitution of lysine for arginine at amino acid position 858 as "L858R mutation", and those with various mutations other than "in-frame deletion" and "L858R

mutation" as "other mutations".

The effect of genotype was initially evaluated under a codominant model in which each genotype was considered separately. However, the estimates of ORs were relatively imprecise due to the small numbers of cases. The four genetic polymorphisms were further classified into two groups by pooling the heterozygous group with either homozygous variant or wild type groups based on the estimated OR of heterozygous genotype: The TTTA repeat in intron 4 of *CYP19A1* was dichotomized as "S/L (one short allele ≤ 7 repeats and one long allele > 7 repeats) and L/L (two long alleles > 7 repeats)" vs. "S/S (two short alleles ≤ 7 repeats)". The genotype of *ERα*, rs2234693, as "C/C and T/C" vs. "T/T"; genotype of *COMT*, rs4680, as "A/A and G/A" vs. "G/G", and genotype of *CYP17*, rs743572, was grouped as "C/C" vs. "T/T and T/C". This grouping scheme has been commonly used in previous studies (20-22).

The above-mentioned four genotypes were further combined to examine whether patients harboring more risk alleles had higher occurrence of *EGFR* hotspot mutations in a dose-response relationship. The OR for each combinatory group was estimated using women had one or less high-risk genotype as the referent group. The statistical significance of the biological gradient was assessed by the test for trend. Two-sided P values less than 0.05 were considered statistically significant. All analyses were performed using SAS statistical software for Windows version 9.2 (SAS institute, Cary, NC, USA)

Results

Patient Characteristics and Distribution of EGFR Mutations

The clinical and demographic characteristics of 617 patients are shown in Table 1. The mean age was 62 years, and the female-to-male ratio was about one. There were only 32% (195/617) cigarette smokers, and female patients were less likely to be smokers than males (2.9% vs. 63.3%). With regard to *EGFR* mutation status, there were 146 (23.7%) in-frame deletions, 139 (22.5%) L858R mutations, and 71 (11.5%) other mutations.

The associations with *EGFR* mutations for sex, cigarette smoking status and disease are shown in Table 2. There were significant associations with *EGFR* mutation status for sex and cigarette smoking in univariate analysis, The frequency of L858R and in-frame deletion was higher in females than males (28.1 % vs. 16.3% and 24.0% vs. 22.7 %, respectively), and higher in never-smokers than ever-smokers (26.6% vs.13.3 % and 26.8% vs. 16.9 %, respectively). Sex was still significantly associated with *EGFR* mutation status after adjustment for smoking status (p=0.04). There was no significant difference in *EGFR* mutation status among patients at different tumor stages.

In order to exclude the possible effect of active smoking status, further analyses were limited to never-smokers. The *EGFR* mutation status among never smokers stratified by sex is shown in Table 3. The percentage of wild type was similar in males (35.2%) and females (33.1%). The percentage of in-frame deletion was higher in males than females (36.1% vs. 23.5%), and the percentage of L858R mutation was higher in females than males (28.8% vs. 20.4%). Sex was significantly associated with *EGFR* mutation status in never-smokers (p=0.03). The sex difference suggests the hotspot mutations of *EGFR* may be related to genes involved in estrogen biosynthesis and metabolism.

Associations between CYP17, CYP19A1, ERα and COMT Genotypes and EGFR Mutations

Table 4 shows the associations with *EGFR* mutations for genetic polymorphisms of *CYP17*, *CYP19A1*, *ERα* and *COMT* in never-smokers stratified by sex.

Genotype of *CYP17* rs743572 was not significantly associated with any type of *EGFR* mutation in female and male never-smokers. The aOR (95% CI) of L858R mutation in female never-smokers was 1.5 (0.8-2.7) for C/C genotype vs. T/T and T/C genotypes.

The (TTTA) repeat polymorphism in the intron4 of *CYP19A1* was significantly associated with the L858R mutation in female never-smokers based on either codominant model (aOR, 2.9 with 95% CI, 1.3-6.5 for S/L genotypes vs. S/S genotype) or dominant model (aOR, 2.6 with 95% CI, 1.2-5.7 for S/L and L/L genotypes vs. S/S genotype), but not in male never-smokers. The (TTTA) repeat polymorphism was not significantly associated with in-frame deletion or other mutations in both female and male

never-smokers.

With regard to the *ERα* rs2234693 polymorphism, the C/C genotype was significantly associated with an increased proportion of L858R mutation (aOR, 3.0 with 95% CI, 1.1-8.1), in-frame deletion (aOR, 2.9 with 95% CI, 1.1-7.6), and other mutations (aOR, 4.3 with 95% CI, 1.3-14.0) compared to the T/T genotype in female never-smokers. Based on the dominant model, the *ERα* rs2234693 polymorphism was still significantly associated with L858R mutation (aOR, 2.1 with 95% CI, 1.1-4.0 for T/C and C/C genotypes vs. T/T genotype) and other mutations (aOR, 2.6 with 95% CI, 1.1-6.0 for T/C and C/C genotypes vs. T/T genotype) in female never-smokers. This *ERα* rs2234693 polymorphism was not significantly associated with in-frame deletion in female never-smokers or any type of *EGFR* mutation in male never-smokers.

The *COMT* rs4680 genetic polymorphism was associated with the L858R mutation (aOR, 6.2 with 95% CI, 1.6-23.5 for A/A genotypes vs. G/G genotype) and other mutations (aOR, 4.7 with 95% CI, 1.1-19.6 for A/A genotypes vs. G/G genotype) in female never-smokers under the codominant model. Based on the dominant model, the *COMT* rs4680 genetic polymorphism was associated with the L858R mutation in both female (aOR, 1.8 with 95% CI 1.0-3.2 for G/A and A/A genotypes vs. G/G genotype) and male (aOR, 3.6 with 95% CI 1.1-11.3 for G/A and A/A genotypes vs. G/G genotype) never-smokers. This SNP was not significantly associated with in-frame deletion and other mutations in both female and male never-smokers.

Effect of Gene-gene Interaction on EGFR mutations

Based on known interactions of individual proteins involved in the estrogen biosynthesis and metabolism pathway shown in Figure 1, we further assessed the associations with L858R mutation and in-frame deletion of *EGFR* gene for the combination of genotypes which were significantly associated with *EGFR* mutation singly. As shown in Table 5, there was a significant dose-response relationship of the number of risk alleles with L858R mutation (P=0.0002), but not with in-frame deletion in female never-smokers or any type of *EGFR* mutation in male never-smokers. The aOR (95% CI) of L858R

mutation increased from 2.0 (0.8-4.9), 3.2 (1.3-7.9), to 10.1 (2.6-38.6) for patients carrying two, three, and four risk alleles compared to those with one or less risk allele in female never-smokers.

Discussion

In the subgroup analysis for never-smokers only, a significant difference was found between *EGFR* mutation subtypes and sex in never-smokers (p=0.03). The percentage of L858R mutation was higher in females than males (28.8% vs. 20.4%) and the percentage of in-frame deletion was higher in males than females (36.1% vs. 23.5%), which were in agreement with previous findings (5). This sex difference implies different mechanisms for the mutagenesis *EGFR* in males and females.

On the basis of multiple candidate genes approach, the L858R mutation of *EGFR* was found to be significantly associated with genetic polymorphisms of *CYP19A1, ERα* and *COMT* in female never-smokers based on the dominant model. In contrast, the in-frame deletion of *EGFR* was significantly associated with *ERα* polymorphism only.

Based on the gene-gene interaction shown in Table 5, the number of risk alleles of *CYP17*, *CYP19A1*, *ERα* and *COMT* was significantly associated with an increasing OR of *EGFR* L858R mutation but not in-frame deletion. The aOR (95% CI) of L858R mutation in male never-smokers also increased from 2.8 (0.6-12.8) to 3.6 (0.8-15.6) for patients carrying two and three risk alleles compared to those carrying one or less risk allele. But the increasing trend was not statistically significant due to the small number of cases. Thus, the possibility that estrogen biosynthesis and metabolism pathway may also contribute to the occurrence of L858R in male never-smokers could not be excluded. In contrast, the in-frame deletion in exon19 was not through such a mechanism.

The difference between L858R mutation and in-frame deletion in carcinogenicity was also found. In-frame deletion has a higher oncogenic ability than that of L858R mutation (23). The mice bearing in-frame deletion has longer tumor latency of tumor formation than those with L858R (24).Tanaka et al.

proposed that chromosomal recombination that involves DNA repair mechanisms may involve in the occurrence of the in-frame deletion in exon19 (6). The instability of the *EGFR* gene might link to the deletion of various regions in exon19. The cytosine and adenine repeats in the *EGFR* intron 1 were reported to influence the expression of *EGFR* (25). It is thus hypothesized that CA repeats in intron 1 of *EGFR* may be associated with the in-frame deletion (unpublished data).

The occurrence of a specific mutation may involve many genes. In this study, four genes including *CYP17*, *CYP19A1*, *ERα* and *COMT* were chosen. The *CYP17* gene codes for cytochrome steroid 17 hydroxylase and C17,20 lyase. They catalyze the conversion of pregnenolone and progesterone to dehydroepiandrosterone (DHEAS) and androstenedione (26). The rs743572 polymorphism is located -34 in the promoter region of *CYP17* with a T to C substitution, which was hypothesized to increase the *CYP17* m-RNA transcriptional activity by creating a SP1 binding site (27). Patients with C/C homozygous genotype were found to have an increased level of testosterone or estrogen (28, 29). In this study, the C/C genotype was associated with a statistically non-significant increase in L858R mutation, and with a statistically non-significant decrease in in-frame deletion in both male and female never-smokers.

The *CYP19A1* gene codes for an aromatase enzyme, which involves in the conversion of androstenedione and testosterone to estrone (E1) and estradiol (E2), respectively. Aromatase expression is elevated in human lung tumor cell lines as well as lung adenocarcinoma tissues (30, 31). Furthermore, an aromatase inhibitor was reported to suppress cell growth and aromatase activity in lung A549 cells *in vitro* (30). A (TTTA)n repeat polymorphism in intron 4 of *CYP19A1* was found to be associated with the risk of breast cancer (32) and the prognosis of patients affected by premenopausal breast cancer (20). Women with one 7 repeat allele (S allele) had lower plasma levels of estrone to androstenedione ratio (32). Our study found a significant association between L858R mutation and S/L and L/L genotype of *CYP19A1*. Mah et al. found higher levels of aromatase was associated with a worse disease prognosis in female never-smokers but not male (31), which is consistent with our finding for the sex disparity.

Interestingly, in-frame deletion was associated with a statistically non-significant association with L/L genotype of *CYP19A1* in male never-smokers, which agrees with previous observations that exon19 in-frame deletions were more frequently found in males (5, 6).

The *COMT* gene codes for catechol-O-methyl-transferase which involves in the methylation of catechol estrogens into inactive metabolites such as 2-methoxyestradiol (7), which was found to inhibit the proliferation of cancer cells (33). A G to A substitution in exon 4 of *COMT,* resulting in the substitution of valine to methionine, may decrease the enzyme activity *in vitro* (34, 35). Thus the subsequent accumulation of hydroxyl-catecholestrogens may generate reactive quinone and semiquinone, which may lead to site-specific oxidative DNA damage such as the binding to the N-7 of guanine to become a depurinating agent (4-OH-E1(E2)-1-(αβ)N7 Gua). The cleavage of the glycosidic bond may result in apurinic sites (36, 37) and genetic mutation. In this study, the G/A+A/A genotypes of *COMT* rs4680 were associated with *EGFR* L858R mutation in both male and female never-smokers.

The *ERα* gene on chromosome 6q25.1 may act as a nuclear receptor protein with an estrogen-binding domain and a DNA-binding domain. Estrone or estradiol bound to ERα play a role as a transcriptional factor, which binds to estrogen response element (ERE) upstream of target genes and affects the gene activity. The rs2234693 T/C (often called *Pvu*II or IVS1-401) polymorphism is located in the intron1 of the *ERα* gene. There was an interactive effect of this SNP and ER status on the survival of breast cancer patients (38). It was also found to increase the risk of myocardial infarction (39). C allele was reported to produce a myeloblastosis (myb) binding site and to increase the expression of downstream reporter gene by co-transfection of B-myb (40). Raso et al. found a significantly elevated expression of ERα in females, never-smokers and patients with all types of *EGFR* mutations (11), which are consistent with our findings that the C/C genotype of *ERα* rs2234693 was significantly associated with all types of *EGFR* mutations in female never-smokers. In contrary, there were no significant associations for males. However, the number of male never-smokers with *EGFR* mutations was too small (n=70) to conclude no association in males. One study with a limited number of cases reported that ERα

expression occurred more frequently in female than male lung tissues (41), another study reported that estrodiol increased the phospho-serine-118-ERα in H1793 but not A549 lung adenocarcinoma cells derived from a female and male patient, respectively (42). Nevertheless, two studies by Stabile et al. (8) and Hershberger et al. (43) reported that lung adenocarcinoma cell lines from males could be responsive to estrogen and antiestrogen. In addition, another study (30) reported aromatase inhibitors could suppress cell growth of male lung A549 cells. Moreover, Hershberger et al. suggested that ERβ is sufficient to mediate both genomic and non-genomic signaling responses to estrogen, and ERβ can cooperate with EGF to promote cell proliferation (43). In addition, a significant correlation between ERβ expression and *EGFR* mutations was also observed (11). Thus, the mechanisms by which the *ERα* and *ERβ* genes might be related to *EGFR* mutations deserve further exploration.

Cytochrome p450 genes are involved in estrogen hydroxylation, such as *CYP1A1*. In an association study between estrogen-related genetic polymorphisms and *EGFR* mutation in lung cancer, the frequency distribution of *CYP1A1* genotypes was different among those with different *EGFR* mutation status (44). But the difference was not statistically significant due to small sample size. Moreover, the CYP1A1, CYP1B1, CYP2C9, CYP3A4 are involved in estradiol hydroxylation (45), which could dilute the associations between *CYP1A1* genotypes and *EGFR* mutation.

In this study, polymorphisms of *CYP17* and *CYP19A1, ERα* and *COMT* were found to be associated with an increased occurrence of *EGFR* mutations in female never-smokers. There are several possible explanations for this finding. Firstly, the robust production of estradiol and estrone due to the high expression level of *CYP17* and *CYP19A1* in combination with the dominant *ER* genotype may increase the binding of estradiol and estrone to the ER, stimulate the proliferation of cell, and increase the risk of DNA replication errors that may result in the burden of DNA repair. Meanwhile, the excessive amount of estrogen hormones might induce the production of hydroxyl-E1 or E2 through hydroxylation catalyzed by the phase I metabolism enzymes, such as the cytochrome p450. If the unstable hydroxyl-E1 or E2 were not methylated by *COMT*, they may lead to the formation of semi-quinone or quinone and related DNA

adducts resulting in mutation, such as the formation of the L858R mutation. This hypothesis may be supported by our finding that number of risk alleles of *CYP17*, *CYP19A1*, *ERα* and *COMT* was significantly associated with an increasing OR of *EGFR* L858R mutation.

Raso et al. found that ERα expression was associated with *EGFR* mutations and a worse survival (11). Coombes et al. (46) reported a lower incidence of second primary lung cancer among breast cancer patients treated with exemestane after tamoxifen than those treated with tamoxifen alone. In addition to examine the associations between genetic polymorphisms and *EGFR* mutations in this study, it is also important to explore their associations with the survival of lung adenocarcinoma patients. Nevertheless, in our study, patients underwent surgical procedures and subsequently treated with chemotherapy rather than gefitinib. We were thus unable to assess the association between genetic polymorphisms and the responsiveness to treatment with EGFR inhibitors. Furthermore, previous studies have reported that *EGFR* mutations were more frequent in East Asians than in non-Asians. In addition to genetic background, environmental factors common among East Asians may contribute to the development of *EGFR* mutations. Both lifestyle and dietary habit, including eating animal viscera with hormone residues, using plastic bags for hot soup to increase the exposure to low level of bisphenol A, and cooking with high exposure to oil fumes, may increase the mutagenic substrates . As the information of these environmental exposures was not available in this study, the gene-environment interactions could not be assessed. In addition to the genes investigated in this study, *ERβ*, *EGFR* and DNA repair genes mentioned above deserve further comprehensive investigation. This study was limited by the wide confidence interval of estimation, resulting from a relatively small sample size. Further multi-disciplinary prospective studies based on a large number of never-smoking patients are needed to elucidate the gene-environment interactions involved in the formation of the hotspots mutation of *EGFR*.

In summary, our results indicate that L858R mutation of *EGFR* is associated with polymorphisms of genes related to estrogen biosynthesis and metabolism in never-smoking female lung adenocarcinoma patients. The number of risk alleles of *CYP17*, *CYP19A1*, *ERα* and *COMT* is associated with an increasing

OR of *EGFR* L858R mutation in never-smokers, especially in females. The genetic polymorphism in *ERα* may be important for various *EGFR* mutations in female never-smokers. The findings provide a clue for the genesis of *EGFR* mutations.

References

1. Lynch TJ, Bell DW, Sordella R*, et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 2004;350: 2129-39.

2. Paez JG, Janne PA, Lee JC*, et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 2004;304: 1497-500.

3. Pao W, Miller V, Zakowski M*, et al.* EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A 2004;101: 13306-11.

4. Shigematsu H, Lin L, Takahashi T*, et al.* Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 2005;97: 339-46.

5. Tokumo M, Toyooka S, Kiura K*, et al.* The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. Clin Cancer Res 2005;11: 1167-73.

6. Tanaka T, Matsuoka M, Sutani A*, et al.* Frequency of and variables associated with the EGFR mutation and its subtypes. Int J Cancer 2010;126: 651-5.

7. Yager JD, Liehr JG. Molecular mechanisms of estrogen carcinogenesis. Annu Rev Pharmacol Toxicol 1996;36: 203-32.

8. Stabile LP, Davis AL, Gubish CT*, et al.* Human non-small cell lung tumors and cells derived from normal lung express both estrogen receptor alpha and beta and show biological responses to estrogen. Cancer Res 2002;62: 2141-50.

9. Kawai H, Ishii A, Washiya K*, et al.* Estrogen receptor alpha and beta are prognostic factors in non-small cell lung cancer. Clin Cancer Res 2005;11: 5084-9.

10. Mollerup S, Jorgensen K, Berge G, Haugen A. Expression of estrogen receptors alpha and beta in human lung tissue and cell lines. Lung Cancer 2002;37: 153-9.

11. Raso MG, Behrens C, Herynk MH*, et al.* Immunohistochemical expression of estrogen and

progesterone receptors identifies a subset of NSCLCs and correlates with EGFR mutation. Clin Cancer Res 2009;15: 5359-68.

12. Yarden RI, Lauber AH, El-Ashry D, Chrysogelos SA. Bimodal regulation of epidermal growth factor receptor by estrogen in breast cancer cells. Endocrinology 1996;137: 2739-47.

13. Mukku VR, Stancel GM. Regulation of epidermal growth factor receptor by estrogen. J Biol Chem 1985;260: 9820-4.

14. Matsuo K, Ito H, Yatabe Y*, et al.* Risk factors differ for non-small-cell lung cancers with and without EGFR mutation: assessment of smoking and sex by a case-control study in Japanese. Cancer Sci 2007;98: 96-101.

15. Travis W, Brambilla, E, Muller-Hermelink, HK, Harris, CC et al (eds). . World Health Organization Classification of Tumours. Pathology and Genetics of tumours of the Lung, Pleura, Thymus, and Heart. IARC Press, Lyon, France 2004.

16. Zuber M, Simpson E, Waterman M. Expression of bovine 17 alpha-hydroxylase cytochrome P-450 cDNA in nonsteroidogenic (COS 1) cells. Science 1986;234: 1258-61.

17. SIMPSON ER, MAHENDROO MS, MEANS GD*, et al.* Aromatase Cytochrome P450, The Enzyme Responsible for Estrogen Biosynthesis. Endocr Rev 1994;15: 342-55.

18. Paech K, Webb P, Kuiper GG*, et al.* Differential Ligand Activation of Estrogen Receptors ER{alpha} and ER at AP1 Sites. Science 1997;277: 1508-10.

19. Polymeropoulos MH, Xiao H, Rath DS, Merril CR. Tetranucleotide repeat polymorphism at the human aromatase cytochrome P-450 gene (CYP19). Nucleic Acids Res 1991;19: 195.

20. Huang CS, Kuo SH, Lien HC*, et al.* The CYP19 TTTA repeat polymorphism is related to the prognosis of premenopausal stage I-II and operable stage III breast cancers. Oncologist 2008;13: 751-60.

21. Yim DS, Parkb SK, Yoo KY*, et al.* Relationship between the Val158Met polymorphism of catechol O-methyl transferase and breast cancer. Pharmacogenetics 2001;11: 279-86.

22. Spurdle AB, Hopper JL, Dite GS*, et al.* CYP17 promoter polymorphism and breast cancer in

Australian women under age forty years. J Natl Cancer Inst 2000;92: 1674-81.

23. Greulich H, Chen TH, Feng W*, et al.* Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. PLoS Med 2005;2: e313.

24. Politi K, Zakowski MF, Fan PD, Schonfeld EA, Pao W, Varmus HE. Lung adenocarcinomas induced in mice by mutant EGF receptors found in human lung cancers respond to a tyrosine kinase inhibitor or to down-regulation of the receptors. Genes Dev 2006;20: 1496-510.

25. Gebhardt F, Zanker KS, Brandt B. Modulation of Epidermal Growth Factor Receptor Gene Transcription by a Polymorphic Dinucleotide Repeat in Intron 1. J Biol Chem 1999;274: 13176-80.

26. Sharp L, Cardy AH, Cotton SC, Little J. CYP17 gene polymorphisms: prevalence and associations with hormone levels and related factors. a HuGE review. Am J Epidemiol 2004;160: 729-40.

27. Carey AH, Waterworth D, Patel K*, et al.* Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet 1994;3: 1873-6.

28. Feigelson HS, Shames LS, Pike MC, Coetzee GA, Stanczyk FZ, Henderson BE. Cytochrome P450c17alpha gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. Cancer Res 1998;58: 585-7.

29. Zmuda JM, Cauley JA, Kuller LH, Ferrell RE. A common promotor variant in the cytochrome P450c17alpha (CYP17) gene is associated with bioavailability testosterone levels and bone size in men. J Bone Miner Res 2001;16: 911-7.

30. Weinberg OK, Marquez-Garban DC, Fishbein MC*, et al.* Aromatase inhibitors in human lung cancer therapy. Cancer Res 2005;65: 11287-91.

31. Mah V, Seligson DB, Li A*, et al.* Aromatase expression predicts survival in women with early-stage non small cell lung cancer. Cancer Res 2007;67: 10484-90.

32. Haiman CA, Hankinson SE, Spiegelman D*, et al.* A tetranucleotide repeat polymorphism in CYP19 and breast cancer risk. Int J Cancer 2000;87: 204-10.

33. Zhu BT, Conney AH. Is 2-methoxyestradiol an endogenous estrogen metabolite that inhibits

mammary carcinogenesis? Cancer Res 1998;58: 2269-77.

34. Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. Pharmacogenetics 1996;6: 243-50.

35. Dawling S, Roodi N, Mernaugh RL, Wang X, Parl FF. Catechol-O-methyltransferase (COMT)-mediated metabolism of catechol estrogens: comparison of wild-type and variant COMT isoforms. Cancer Res 2001;61: 6716-22.

36. Liehr JG. Hormone-associated cancer: mechanistic similarities between human breast cancer and estrogen-induced kidney carcinogenesis in hamsters. Environ Health Perspect 1997;105 Suppl 3: 565-9.

37. Cavalieri EL, Stack DE, Devanesan PD*, et al.* Molecular origin of cancer: catechol estrogen-3,4-quinones as endogenous tumor initiators. Proc Natl Acad Sci U S A 1997;94: 10937-42.

38. Boyapati SM, Shu XO, Ruan ZX*, et al.* Polymorphisms in ER-alpha gene interact with estrogen receptor status in breast cancer survival. Clin Cancer Res 2005;11: 1093-8.

39. Shearman AM, Cupples LA, Demissie S*, et al.* Association between estrogen receptor alpha gene variation and cardiovascular disease. JAMA 2003;290: 2263-70.

40. Herrington DM, Howard TD, Brosnihan KB*, et al.* Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein. Circulation 2002;105: 1879-82.

41. Fasco MJ, Hurteau GJ, Spivack SD. Gender-dependent expression of alpha and beta estrogen receptors in human nontumor and tumor lung tissue. Mol Cell Endocrinol 2002;188: 125-40.

42. Ivanova MM, Mazhawidza W, Dougherty SM, Klinge CM. Sex differences in estrogen receptor subcellular location and activity in lung adenocarcinoma cells. Am J Respir Cell Mol Biol 2010;42: 320-30.

43. Hershberger PA, Stabile LP, Kanterewicz B*, et al.* Estrogen receptor beta (ERbeta) subtype-specific ligands increase transcription, p44/p42 mitogen activated protein kinase (MAPK) activation and growth

in human non-small cell lung cancer cells. J Steroid Biochem Mol Biol 2009;116: 102-9.

44. Bell DW, Brannigan BW, Matsuo K*, et al.* Increased prevalence of EGFR-mutant lung cancer in women and in East Asian populations: analysis of estrogen-related polymorphisms. Clin Cancer Res 2008;14: 4079-84.

45. Modugno F, Knoll C, Kanbour-Shakir A, Romkes M. A potential role for the estrogen-metabolizing cytochrome P450 enzymes in human breast carcinogenesis. Breast Cancer Res Treat 2003;82: 191-7.

46. Coombes RC, Hall E, Gibson LJ*, et al.* A Randomized Trial of Exemestane after Two to Three Years of Tamoxifen Therapy in Postmenopausal Women with Primary Breast Cancer. New England Journal of Medicine 2004;350: 1081-92.

Acknowledgements

Thanks for the support of clinical data by Comprehensive Cancer Center of Taichung Veterans General Hospital, Taichung, Taiwan

Table 1. Demographics and Clinical Characteristics of 617 patients affected with lung adenocarcinoma

Demographic or Clinical Characteristic	No. of Patients	%
Age		
<30	2	0.3
30-34	2	0.3
35-39	15	2.4
40-44	31	5.0
45-49	59	9.6
50-55	76	12.3
55-59	59	9.6
60-64	83	13.4
65-69	90	14.6
70+	200	32.4
Mean± SD (range)	62±12 (28-90)	
Gender		
Female	313	50.7
Male	304	49.3
Cigarette smoking status‡		
Never smoker	410	67.8
Ever Smoker	195	32.2
Cigarette smoking status‡ by sex		
Female		
Never smoker	302	97.1
Ever Smoker	9	2.9
Male		
Never smoker	108	36.7
Ever Smoker	186	63.3
EGFR mutation status*		
Wild type	261	42.3
L858R †	139	22.5
In-frame Deletion	146	23.7
Others	71	11.5
Disease stage ‡		
IA	121	19.8
IB	173	28.4
IIA	5	0.8
IIB	50	8.2
IIIA	113	18.5
IIIB	37	6.1

IV	111	18.2
Total	617	

*EGFR, epidermal growth factor receptor gene.

†L858R, a substitution of Lysine for Arginine mutation at amino acid position 858 point mutation in exon 21.

‡12 patients with missing data on cigarette smoking status, and 7 with missing data on disease stage

Table 2. Demographics and clinical characteristics of lung adenocarcinoma patients by EGFR mutation status

	Wild type*		L858R*		In-frame Deletion*		Others*			
Variable	No	%	No	%	No.	%	No.	%	p-value†	p-value‡
Gender										
Female	104	33.2	88	28.1	75	24.0	46	14.7		
Male	157	50.1	51	16.3	71	22.7	25	8.0	<0.0001	0.04
Cigarette smoking status										
Never smoker	138	33.6	109	26.6	110	26.8	53	12.9		
Ever smoker	118	60.5	26	13.3	33	16.9	18	9.2	<0.0001	
Disease stage										
IA/IB	113	38.4	79	26.9	70	23.8	32	10.9		
IIA/IIB	34	61.8	9	16.4	6	10.9	6	10.9		
IIIIA/IIIB	63	42	32	21.3	38	25.3	17	11.3		
IV	46	41.4	19	17.1	30	27.0	16	14.4	0.06	

*12 patients (5 with wild type, 4 with L858R, 3 with In-frame deletion) with missing data on cigarette smoking status, and 7 patients (5 with wild type, 2 with In-frame deletion) with missing disease stage

†p-value based on $\chi 2$ test for gender, cigarette smoking status, and disease stage

‡ p-value indicated that the association between mutation status and sex after adjusting for smoking status based on Cochran-Mantel-Haenszel Statistics.

Table 3. Frequency of EGFR mutation status among never-smokers by sex.

	Male		Female	
EGFR mutation	No.	%	No.	%
Wild	38	35.2	100	33.1
L858R	22	20.4	87	28.8
In-frame Deletion	39	36.1	71	23.5
Others	9	8.3	44	14.6
Total	108		302	p=0.03

†p-value based on $\chi 2$ test for sex

Table4. Odds ratios of developing lung adenocarcinoma for EGFR mutation status among never smokers by sex

Genetic polymorphism	Wild	L858R			In-frame Deletion			Others		
	No (%)	No (%)	aOR (95% CI)*		No (%)	aOR (95% CI)†		No (%)	aOR (95% CI)‡	
Female										
CYP17_rs743572**										
T/T	16 (16.8)	14 (16.3)	1.0 (referent)	1.0 (referent)	13 (19.1)	1.0 (referent)	1.0 (referent)	12 (28.6)	1.0 (referent)	1.0 (referent)
T/C	45 (47.4)	34 (39.5)	1.0 (0.4-2.3)		35 (51.5)	1.0 (0.4-2.3)		11 (26.2)	0.3 (0.1-1.1)	
C/C	34 (35.8)	38 (44.2)	1.4 (0.6-3.4)	1.5 (0.8-2.7)	20 (29.4)	0.7 (0.3-1.8)	0.7 (0.4-1.5)	19 (45.2)	0.8 (0.3-2.1)	1.5 (0.7-3.1)
CYP19A1(TTTA)$_n$ §,‖										
S/S	26 (27.7)	11 (12.8)	1.0 (referent)	1.0 (referent)	17 (25.0)	1.0 (referent)	1.0 (referent)	14 (32.6)	1.0 (referent)	1.0 (referent)
S/L	46 (48.9)	58 (67.4)	2.9 (1.3-6.5)††	2.6 (1.2-5.7)††	34 (50.0)	1.1 (0.5-2.4)	1.1 (0.6-2.3)	18 (41.8)	0.7 (0.3-1.7)	0.8 (0.4-1.8)
L/L	22 (23.4)	17 (19.8)	1.9 (0.7-5.0)		17 (25.0)	1.2 (0.5-2.9)		11 (25.6)	0.9 (0.4-2.5)	
ERα _rs2234693**										
T/T	39 (41.0)	22 (25.6)	1.0 (referent)	1.0 (referent)	24 (35.3)	1.0 (referent)	1.0 (referent)	9 (21.4)	1.0 (referent)	1.0 (referent)
T/C	47 (49.5)	49 (57.0)	1.9 (0.9-3.7)	2.1 (1.1-4.0)††	28 (41.2)	1.0 (0.5-1.9)	1.3 (0.7-2.4)	24 (57.2)	2.2 (0.9-5.4)	2.6 (1.1-6.0) ††
C/C	9 (9.5)	15 (17.4)	3.0 (1.1-8.1)††		16 (23.5)	2.9 (1.1-7.6)††		9 (21.4)	4.3 (1.3-14.0)††	
COMT _rs4680**										
G/G	56 (58.9)	39 (45.3)	1.0 (referent)	1.0 (referent)	36 (52.9)	1.0 (referent)	1.0 (referent)	28 (66.7)	1.0 (referent)	1.0 (referent)
G/A	36 (37.9)	35 (40.7)	1.4 (0.8-2.6)	1.8 (1.0-3.2)††	29 (42.6)	1.3 (0.7-2.4)	1.3 (0.7-2.4)	7 (16.7)	0.4 (0.2-1.0)	0.7 (0.3-1.5)
A/A	3 (3.2)	12 (14.0)	6.2 (1.6-23.5)††		3 (4.4)	1.5 (0.3-8.0)		7 (16.7)	4.7 (1.1-19.6)††	
Male										
CYP17_rs743572**										
T/T	6 (16.2)	4 (19.0)	1.0 (referent)	1.0 (referent)	9 (23.7)	1.0 (referent)	1.0 (referent)	2 (22.2)	1.0 (referent)	1.0 (referent)
T/C	20 (54.1)	9 (42.9)	0.6 (0.1-2.8)		20 (52.6)	0.7 (0.2-2.2)		4 (44.4)	0.6 (0.1-4.3)	
C/C	11 (29.7)	8 (38.1)	1.0 (0.2-4.9)	1.4 (0.5-4.5)	9 (23.7)	0.5 (0.1-2.1)	0.7 (0.3-2.1)	3 (33.3)	1.0 (0.1-1.9)	1.5 (0.3-8.0)
CYP19A1(TTTA)$_n$ §,‖										
S/S	11 (29.7)	7 (35.0)	1.0 (referent)	1.0 (referent)	9 (23.7)	1.0 (referent)	1.0 (referent)	2 (22.2)	1.0 (referent)	1.0 (referent)
S/L	19 (51.4)	8 (40.0)	0.7 (0.2-2.7)	0.9 (0.3-2.8)	17 (44.7)	1.1 (0.4-3.3)	1.3 (0.5-3.8)	6 (66.7)	2.5 (0.4-17.1)	1.9 (0.3-12.1)
L/L	7 (18.9)	5 (25.0)	1.1 (0.3-5.3)		12 (31.6)	2.0 (0.6-7.4)		1 (11.1)	0.9 (0.1-12.5)	

ERα _rs2234693**												
T/T	13	(35.1)	8 (38.1)	1.0 (referent)	1.0 (referent)	12 (31.6)	1.0 (referent)	1.0 (referent)	3 (33.3)	1.0 (referent)	1.0 (referent)
T/C	18	(48.7)	9 (42.9)	0.8 (0.2-2.6)	0.8 (0.2-2.4)	16 (42.1)	0.9 (0.3-2.7)	1.1 (0.4-3.0)	6 (66.7)	1.5 (0.3-7.8)	1.1 (0.2-5.1)
C/C	6	(16.2)	4 (19.0)	0.7 (0.1-3.7)		10 (26.3)	1.7 (0.5-2.3)		0 (0.0)	0	
COMT _rs4680**												
G/G	27	(73.0)	9 (42.9)	1.0 (referent)	1.0 (referent)	22 (57.9)	1.0 (referent)	1.0 (referent)	5 (55.6)	1.0 (referent)	1.0 (referent)
G/A	9	(24.3)	10 (47.6)	3.1 (0.9-10.5)	3.6 (1.1-11.3) ††	14 (36.8)	1.9 (0.7-5.2)	1.9 (0.7-5.2)	4 (44.0)	2.0 (0.4-9.4)	1.9 (0.4-8.7)
A/A	1	(2.7)	2 (9.5)	8.6 (0.6-124.0)		2 (5.3)	2.7 (0.2-32.7)		0 (0.0)	0	

95% CI, 95% confidence interval.

*aOR, age-adjusted ORs of L858R vs. wild type

†aOR, age-adjusted ORs of in-frame deletion vs. wild type

‡aOR, age-adjusted ORs of other mutations vs. wild type

§S/S, two alleles ≤ 7 repeats; L/L, two alleles > 7 repeats

‖3 patients (1 with wild type among female never-smokers, 1 with wild type and 1 with L858R among male never-smokers) with missing data on *CYP19A1* $(TTTA)_n$

**2 patients (1 with other mutations among female never-smokers, 1 with wild type among male never-smokers) with missing data on *CYP17*, rs743572, *COMT*, rs4680 and *ERα*, rs2234693.

†† $P < .05$

Table 5. Multiple logistical regression analysis of associations between the number of risk alleles involved in estrogen biosynthesis and metabolism and the EGFR mutations in never-smoking affected with lung adenocarcinoma

No. of risk alleles*	Wild type†		L858R†			In-frame Deletion		
	No.	%	No	%	aOR (95% CI)‡	No	%	aOR (95% CI)‡
Female								
0-1	26	27.7	10	11.6	1.0 (referent)§	19	27.9	1.0 (referent)§
2	36	38.3	28	32.6	2.0 (0.8-4.9)	24	35.3	0.9 (0.4-2.0)
3	28	29.8	34	39.5	3.2 (1.3-7.9)‖	20	29.4	1.0 (0.4-2.2)
4	4	4.2	14	16.3	10.1 (2.6-38.6)**	5	7.4	1.7 (0.4-7.3)
			Test for trend		P=0.0002			P=0.69
Male								
0-1	15	41.7	4	20.0	1.0 (referent)§	9	23.7	1.0 (referent)§
2	10	27.8	7	35.0	2.8 (0.6-12.8)	17	44.7	2.8 (0.9-8.8)
3	9	25.0	9	45.0	3.6 (0.8-15.6)	10	26.3	1.8 (0.5-6.2)
4	2	5.5	0	0.0	-	2	5.3	1.7 (0.2-14.0)
			Test for trend		0.2			0.38

*Number of risk alleles of *CYP17* rs743572 (C/C genotype), *CYP19A1* TTTA repeats (S/L and L/L alleles), *COMT* rs4680 (G/A and A/A genotype) and *ERα* rs2234693 (T/C+ C/C genotypes).

†4 patients (1 with wild type among female never-smokers, 2 with wild type and 1 with L858R among male never-smokers) with missing data

‡aOR (95% CI), age-adjusted odds ratio (95% confidence interval)

[§] Zero or one risk allele as the referent group

[‖] $P<.01$

[**] $P<.001$.

Figures

Fig.1. Biosynthetic, metabolic and signaling pathways of estrogen and estrogen receptors

Supplementary Table 1. Sequences of primers and Taqman probes for genotyping:

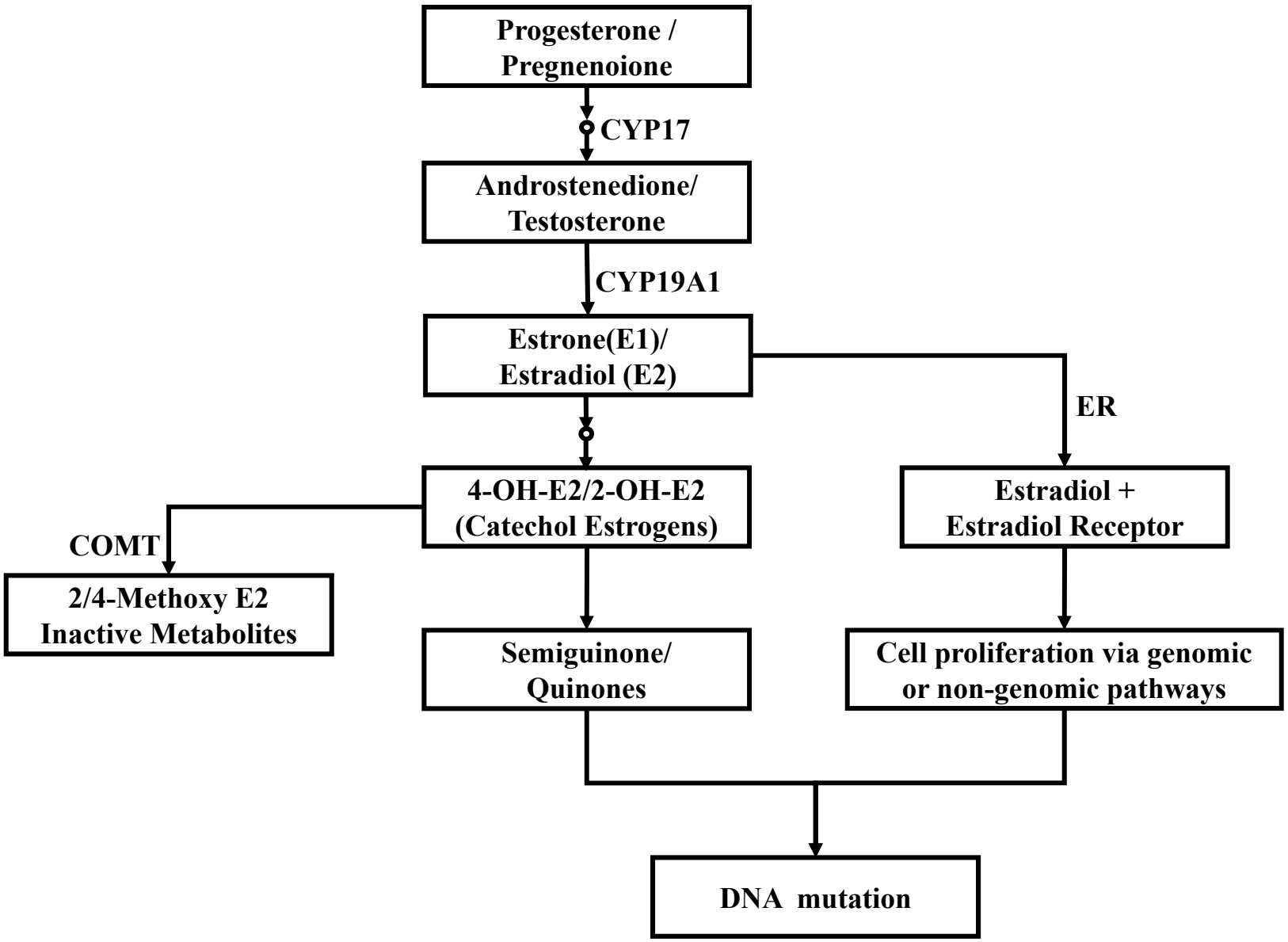

Fig.1. Biosynthetic, metabolic and signaling pathways of estrogen and estrogen receptors