Genomic Profiles Specific to Patient Ethnicity in Lung Adenocarcinoma

Philippe Broët1,7,8, Cyril Dalmasso1, Eng Huat Tan2, Marco Alfano8, Shenli Zhang3, Jeanie Wu4, Ming Hui Lee6, Jean-François Régnard8,6, Darren Lim2, Heng Nung Koong6, Thirugnanam Agastian6, Lance D. Miller1,10, Elaine Lim5, Sophie Camilleri-Broët1, and Patrick Tan1,4,6

Abstract

Purpose: East-Asian (EA) patients with non–small-cell lung cancer (NSCLC) are associated with a high proportion of nonsmoking women, epidermal growth factor receptor (EGFR)-activating somatic mutations, and clinical responses to tyrosine kinase inhibitors. We sought to identify novel molecular differences between NSCLCs from EA and Western European (WE) patients.

Experimental Design: A total of 226 lung adenocarcinoma samples from EA (n = 90) and WE (n = 136) patients were analyzed for copy number aberrations (CNA) by using a common high-resolution SNP (single nucleotide polymorphism) microarray platform. Univariate and multivariate analyses were carried out to identify CNAs specifically related to smoking history, EGFR mutation status, and ethnicity.

Results: The overall genomic profiles of adenocarcinomas from EA and WE patients were highly similar. Univariate analyses revealed several CNAs significantly associated with ethnicity, EGFR mutation, and smoking, but not to gender, and KRAS or p53 mutations. A multivariate model identified four ethnicity-specific recurrent CNAs—significantly higher rates of copy number gain were observed on 16p13.13 and 16p13.11 in EA tumors, whereas higher rates of genomic loss on 19p13.3 and 19p13.11 were observed in tumors from WE patients. We identified several potential driver genes in these regions, showing a positive correlation between cis-localized copy number changes and transcriptomic changes.

Conclusion: 16p copy number gains (EA) and 19p losses (WE) are ethnic-specific chromosomal aberrations in lung adenocarcinoma. Patient ethnicity should be considered when evaluating future NSCLC therapies targeting genes located on these areas. Clin Cancer Res; 17(11); 1–9. ©2011 AACR.

Introduction

Lung cancer is a common malignancy worldwide and the leading cause of global cancer mortality in both males and females (1). Clinical and epidemiologic studies on non–small-cell lung carcinomas (NSCLC) of Asian and Caucasian origins have revealed substantial clinical and molecular differences between the two populations, with the former exhibiting a higher proportion of non-smoking females, adenocarcinomas, and epidermal growth factor receptor (EGFR) mutations (2–5).

Some studies have shown that EGFR mutations, which are currently used as predictive markers of clinical response to EGFR tyrosine kinase inhibitors (TKI), are linked with EGFR gene amplifications (6, 7). Because Asian patients are known to exhibit higher response rates to EGFR TKIs, these findings raise the possibility that in addition to DNA mutations, recurrent copy number aberrations (CNA; amplifications/deletions) specific to Asian and Caucasian NSCLC patients may also exist. To date, most reported large CNA analyses of NSCLC have largely comprised tumors of Caucasian origin, identifying important oncogenic loci such as Myc, TTF-1 (NKX2-1), and E2F (8–12). In contrast, similar CNA data on NSCLCs from Asian patients have been sparse, unavailable, or only available as small patient populations or low-resolution technologies. Identifying discrete CNA differences between NSCLCs from Asian and Caucasian patients would suggest that these 2 conditions represent distinct biological and clinical entities. The discovery of ethnic-specific CNAs would also argue for patient ethnicity being an important consideration in the design and evaluation of future targeted therapies in lung cancer if the targeted genes involve ethnic-specific CNAs.

Here, we report the first large-scale comparative analysis of genomic alteration profiles between lung adenocarcinomas from East-Asian (EA) and Western Europe (WE). We found, using a common high-resolution SNP (single nucleotide polymorphism) microarray platform, that the genomic profiles of lung adenocarcinomas from both...
Translational Relevance

East-Asian (EA) and Western European (WE) patients with non–small-cell lung cancer (NSCLC) exhibit several epidemiologic and molecular differences, including a high proportion of nonsmoking women, EGFR (epidermal growth factor receptor)-activating somatic mutations, and clinical responses to tyrosine kinase inhibitor therapy in cancers in EA patients. However, whether differences exist in the pattern of copy number aberrations (CNA) between EA and WE populations is currently unknown.

We report the first large-scale comparative analysis of genomic alteration profiles between lung adenocarcinomas from EA and WE populations. We identify 2 ethnic-specific CNAs: 16p13 amplification (Asian specific) and 19p13 deletion (Caucasian specific). We discuss candidate copy number–driven genes located on these areas. Our findings emphasize the need to take into account ethnicity in the design of future NSCLC clinical trials evaluating targeted therapies involving copy number–driven candidate genes located on those ethnic-specific genomic regions.

Material and Methods

Patients and procedures

Frozen samples of 318 lung carcinomas were processed. Tumors from Western European patients (n = 157, WE series) were collected at the Hotel-Dieu Hospital (Paris, France) between 2000 and 2005. Tumors from EA patients (n = 162, predominantly Asian Chinese, EA series) were collected at 4 centers (National University Hospital, Tan Tock Seng Hospital, National Cancer Centre Singapore, and Singapore General Hospital) between 2002 and 2006. All samples were surgical specimens and underwent pathologic review. Tumor samples were immediately frozen at −80°C until retrieved. Only tissue samples with tumor cells more than 50% were selected.

The 2 series were initially collected from different perspectives (WE series: prognostic impact for stage I adenocarcinomas; EA series: chromosomal aberrations in lung tumors). Thus, the inclusion for the WE series was restricted to stage I adenocarcinomas whereas the EA series was heterogeneous with respect to histology and disease stage (only half of the tumors were stage I).

A preliminary analysis on both variables (stage, histology) on the EA series was done. Although genome-wide significant differences in CNAs were observed for histology (e.g., higher copy number gains in 3q for squamous carcinomas), there was no genome-wide significant relationship between CNAs and disease stage. These results suggest that CNAs observed in lung cancers from different disease stages are likely to be highly similar. Thus, for this work, we restricted our analysis only to lung adenocarcinomas but included all stages to preserve statistical power.

The present study included 226 adenocarcinoma samples: 136 from WE patients (all stage I) 90 from EA patients (48 stage I). All study protocols were approved by both institutional and ethics review committees.

DNA was extracted from frozen samples by using a Nucleon DNA extraction kit (BACC2; Amersham Biosciences). All samples were processed and hybridized to Affymetrix Genome-Wide Human SNP 6.0 arrays according to the manufacturer’s specifications in the same center (13). All 318 patients were genotyped for EGFR, KRAS, and p53 mutations. Primers (EGFR exons 18–21, KRAS exon 2, and p53 exons 5–9) were used to amplify the relevant regions, and DNA sequencing was conducted on an ABI3730xl Sanger sequencer. All mutations were confirmed by bidirectional sequencing.

For MYH11 and LKB1 sequencing, we sequenced 43 samples (22 from WE and 21 from EA patients) amplified using a whole-genome amplification (WGA) protocol using the Qiagen REPLI-g Midi Kit (catalogue no. 150045; Qiagen; ref. 14). Amplified DNA was diluted 1:100, and 2 to 3 μL of diluted DNA was used for each PCR. DNA sequencing was conducted on an ABI3730xl Sanger sequencer. All mutations were confirmed by bidirectional sequencing.

For gene expression analyses, total RNA was extracted from frozen (−80°C) tumor samples using TRIzol. Samples were processed and hybridized to Human U133Plus 2.0 oligonucleotide arrays (Affymetrix) according to the manufacturer’s protocol. In this study, RNA from 128 (WE series) and 71 (EA series) samples were of sufficient quality to enable reliable gene expression analysis.

Statistical analysis

To detect DNA copy number changes, we considered the log ratio between the probe signal and a reference signal estimated from a pool of hybridized HapMap samples that were matched for ethnicity to rule out the hypothesis that the alterations identified could be normal copy number variations specific to the different ethnic populations. Systematic sources of variation (GC content, fragment length) were removed by subtracting for each log ratio its estimated predicted values inferred from the least-squares estimates of the parameters. Data from the sex chromosomes were not considered. To reduce measurement variability while keeping a sufficient level of resolution, we averaged the intensities over 50...
Consecutive probes leading to a total of 17,769 genomic segments (average size of 150 kb) providing tiling coverage of the human genome. Inferences about the copy number status of each genomic segment were obtained using the modified CGHmix classification procedure (14).

We employed a novel methodology for chromosomal pattern analysis to distinguish between "passenger" and the "driver" chromosomal aberrations (15). From this classification, driver genomic areas were identified as genomic segments, having the highest frequency for amplification while showing the lowest frequency for deletion (exclusively amplified CNAs), or vice versa (exclusively deleted CNAs).

A set of univariate nonparametric analyses was used to compare the marginal trinomial distribution of genomic CNAs to clinical (ethnicity, smoking history, gender) and molecular (EGFR, KRAS) factors (Fig. 1). For each genomic segment, we computed the $\chi^2$ statistics (with Yates' correction for continuity) testing the null hypothesis of no difference between the considered groups for the chromosomal aberration distributions. Selected genomic segments were reported while controlling the false discovery rate (FDR) for a classical threshold of 5% (16).

To disentangle relationships between ethnicity and CNAs from other confounding factors, we conducted multinomial logistic regression analyses. We compared the log-likelihood ratios of a null model (without ethnicity but including all clinicomolecular variables shown to be significantly associated to CNAs in univariate analysis) with those of a full model including ethnicity by using an iterative least-squares estimation procedure (VGAM R-package, http://cran.r-project.org/web/packages/VGAM; ref. 17). We calculated the likelihood ratio statistics corresponding to twice the difference in the log-likelihood ratios that test the hypothesis of the lack of influence of ethnicity on the propensity for deletion and amplification. We then selected genomic segments exhibiting an FDR threshold of 5%. To focus on driver areas, these segments were further filtered only to regions that were exclusively amplified or deleted in only one population.

To overcome any confounding effect related to disease stage, we also conducted additional analyses restricted on stage I lung adenocarcinomas (136 from WE and 48 from EA patients). Because of the reduction in the sample size of the EA series, we decided to relax the FDR threshold for
feature selection and increase it from 5% to 25% that keeps a good balance between not inflating the number of false-positive discovery rates and not sacrificing the detection of true discovery rates.

For gene expression, the microarray data were standardized and normalized using the robust multiarray average (RMA) procedure (18) implemented in the Bioconductor open source software (http://www.bioconductor.org). We tested the null hypothesis that no correlation exists between cis-localized copy number changes and transcriptomic changes for the subset of genes located in the selected genomic areas. To address the multiple testing problems, we selected genes for a 5% FDR level. These genes were considered as copy number-driven candidate genes.

Results

Clinical and molecular features from the 2 series are detailed in Table 1. As expected (19, 20), the EA series was associated with a higher rate of females, nonsmokers, and EGFR mutations. In both series, smoking history showed a significant relationship with EGFR mutation, with the highest mutation rate among nonsmokers (\(P < 10^{-10}\), 52.8% and 30.7% in EA and WE series, respectively). There was a significant relationship between gender and smoking, with a higher rate of females, nonsmokers, and EGFR mutations (exons 20 and 21) were synonymous nucleotide mutations (\(P < 0.02\)). We identified 2 cases with a co-mutation of EGFR and KRAS, but, in these cases, the EGFR mutations (exons 20 and 21) were synonymous nucleotide alterations and thus not expected to influence EGFR protein function.

For each series, we evaluated overall patterns of chromosomal aberration complexity by calculating the fraction of genomic regions altered in each sample. The median chromosomal aberration rate was of 38.7% and 38.1% for EA and WE series, respectively. When comparing the 2 series, the fraction of genome altered (divided on quartiles) was not significantly different between the 2 series (\(P = 0.14\)).

Table 1. Clinical characteristics and mutational status in WE and EA series

<table>
<thead>
<tr>
<th></th>
<th>Paris series ((n = 136))</th>
<th>Singapore series ((n = 90))</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>100</td>
<td>49</td>
<td>0.002</td>
</tr>
<tr>
<td>Female</td>
<td>36 (26.5%)</td>
<td>41 (46.7%)</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>63 (42–84)</td>
<td>66 (30–85)</td>
<td>NS</td>
</tr>
<tr>
<td><strong>Smoking history</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smokers</td>
<td>123</td>
<td>37</td>
<td>(9 \times 10^{-10})</td>
</tr>
<tr>
<td>Nonsmokers</td>
<td>13 (9.5%)</td>
<td>35</td>
<td>8 \times 10^{-5}</td>
</tr>
<tr>
<td>NA</td>
<td></td>
<td>48.6%</td>
<td></td>
</tr>
<tr>
<td><strong>Mutational status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>15 (11%)</td>
<td>29 (32.2%)</td>
<td></td>
</tr>
<tr>
<td>KRAS</td>
<td>28 (20.5%)</td>
<td>9 (10.0%)</td>
<td>0.03</td>
</tr>
<tr>
<td>TP53</td>
<td>41 (30.1%)</td>
<td>27 (30.0%)</td>
<td>NS</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not available.

As previously described (21, 22), there was a significant relationship between smoking history and KRAS mutation status, with a higher rate of mutation among smokers (\(P = 0.002\)). KRAS mutations were not associated with gender. As expected from previous data (23, 24), there was a significant negative correlation between EGFR and KRAS mutations (\(P = 0.02\)). We identified 2 cases with a co-mutation of EGFR and KRAS, but, in these cases, the EGFR mutations (exons 20 and 21) were synonymous nucleotide alterations and thus not expected to influence EGFR protein function.

For each series, we evaluated overall patterns of chromosomal aberration complexity by calculating the fraction of genomic regions altered in each sample. The median chromosomal aberration rate was of 38.7% and 38.1% for EA and WE series, respectively. When comparing the 2 series, the fraction of genome altered (divided on quartiles) was not significantly different between the 2 series (\(P = 0.14\)).

Figure 2 displays the frequencies of copy gains and losses across the whole genome for the 2 series. Except for a limited number of genomic areas that are discussed later, visual inspection of the 2 series revealed a similar pattern of CNAs. From chromosomal pattern analysis, 15.5% (WE) and 14.5% (EA) of genomic segments were considered as exclusively amplified with a mixture of broad (1q, 5p, 6p, 7p, 8q) and focal (11q, 14q, 20q) contiguous areas. Exclusively deleted areas represent 25.3% (WE) and 28.7% (EA) of genomic sequences with large contiguous areas (3p, 6q, 8p, 9q, 13p).

As previously reported (8–11), 5p15 and 1q21 were the most common exclusively amplified genomic regions, which are found in more than 60% of total samples. With a rate higher than 50%, 3p14 and 9q21 are the most common exclusively deleted genomic regions, which are found in more than 60% of total samples. The most recurrent deletions occurred at similar frequencies between the 2 series. KRAS mutation rates were significantly higher in the WE series than in the EA series. As previously described (21, 22), there was a significant relationship between smoking history and KRAS mutation status, with a higher rate of mutation among smokers (\(P = 0.002\)). KRAS mutations were not associated with gender. As expected from previous data (23, 24), there was a significant negative correlation between EGFR and KRAS mutations (\(P = 0.02\)). We identified 2 cases with a co-mutation of EGFR and KRAS, but, in these cases, the EGFR mutations (exons 20 and 21) were synonymous nucleotide alterations and thus not expected to influence EGFR protein function.

For each series, we evaluated overall patterns of chromosomal aberration complexity by calculating the fraction of genomic regions altered in each sample. The median chromosomal aberration rate was of 38.7% and 38.1% for EA and WE series, respectively. When comparing the 2 series, the fraction of genome altered (divided on quartiles) was not significantly different between the 2 series (\(P = 0.14\)).

Table 1 displays the frequencies of copy gains and losses across the whole genome for the 2 series. Except for a limited number of genomic areas that are discussed later, visual inspection of the 2 series revealed a similar pattern of CNAs. From chromosomal pattern analysis, 15.5% (WE) and 14.5% (EA) of genomic segments were considered as exclusively amplified with a mixture of broad (1q, 5p, 6p, 7p, 8q) and focal (11q, 14q, 20q) contiguous areas. Exclusively deleted areas represent 25.3% (WE) and 28.7% (EA) of genomic sequences with large contiguous areas (3p, 6q, 8p, 9q, 13p).

As previously reported (8–11), 5p15 and 1q21 were the most common exclusively amplified genomic regions, which are found in more than 60% of total samples. With a rate higher than 50%, 3p14 and 9q21 are the most common exclusively deleted genomic regions, which are found in more than 60% of total samples. The most recurrent deletions occurred at similar frequencies between the 2 series. KRAS mutation rates were significantly higher in the WE series than in the EA series. As previously described (21, 22), there was a significant relationship between smoking history and KRAS mutation status, with a higher rate of mutation among smokers (\(P = 0.002\)). KRAS mutations were not associated with gender. As expected from previous data (23, 24), there was a significant negative correlation between EGFR and KRAS mutations (\(P = 0.02\)). We identified 2 cases with a co-mutation of EGFR and KRAS, but, in these cases, the EGFR mutations (exons 20 and 21) were synonymous nucleotide alterations and thus not expected to influence EGFR protein function.

For each series, we evaluated overall patterns of chromosomal aberration complexity by calculating the fraction of genomic regions altered in each sample. The median chromosomal aberration rate was of 38.7% and 38.1% for EA and WE series, respectively. When comparing the 2 series, the fraction of genome altered (divided on quartiles) was not significantly different between the 2 series (\(P = 0.14\)).

Figure 2 displays the frequencies of copy gains and losses across the whole genome for the 2 series. Except for a limited number of genomic areas that are discussed later, visual inspection of the 2 series revealed a similar pattern of CNAs. From chromosomal pattern analysis, 15.5% (WE) and 14.5% (EA) of genomic segments were considered as exclusively amplified with a mixture of broad (1q, 5p, 6p, 7p, 8q) and focal (11q, 14q, 20q) contiguous areas. Exclusively deleted areas represent 25.3% (WE) and 28.7% (EA) of genomic sequences with large contiguous areas (3p, 6q, 8p, 9q, 13p).

As previously reported (8–11), 5p15 and 1q21 were the most common exclusively amplified genomic regions, which are found in more than 60% of total samples. With a rate higher than 50%, 3p14 and 9q21 are the most common exclusively deleted genomic regions, which are found in more than 60% of total samples. The most recurrent deletions occurred at similar frequencies between the 2 series. KRAS mutation rates were significantly higher in the WE series than in the EA series. As previously described (21, 22), there was a significant relationship between smoking history and KRAS mutation status, with a higher rate of mutation among smokers (\(P = 0.002\)). KRAS mutations were not associated with gender. As expected from previous data (23, 24), there was a significant negative correlation between EGFR and KRAS mutations (\(P = 0.02\)). We identified 2 cases with a co-mutation of EGFR and KRAS, but, in these cases, the EGFR mutations (exons 20 and 21) were synonymous nucleotide alterations and thus not expected to influence EGFR protein function.
regions. We also confirmed well-known exclusive amplifications at 7p11 (EGFR), 8q24 (Myc), 11q13 (CCND1), 14q13 (NKX2-1/JTF-1), and 20q11 (E2F). Exclusive deletions were identified at 3p14 (FHIT), 8p21 (DUSP4), 9p21 (CDKN2A), 10q, 15q, 16q23 (WWOX), and 17p13 (p53). All these genomic aberrations were common to both series.

Using univariate analyses, we then identified genomic segments associated with the following factors (Table 2, Fig. 3): ethnicity, smoking history, gender, EGFR, and KRAS.

**Table 2. Cytobands selected from univariate analysis**

| Ethnicity | 1p36.23; 1p36.22; 1p36.21; 1p36.13; 1p36.12; 1p36.11; 1p35.3; 1p31.1; 2p22.3; 4p15.1; 5q35.3; 6p21.32; 6q14.1; 16p13.3; 16p13.2; 16p13.13; 16p13.11; 16p13.3; 16p13.2; 16p13.13; 16p13.12; 16p13.11 |
| Smoking status | 1p34.3; 1p34.2; 5q31.1; 5q31.2; 5q31.3; 5q35.2; 5q35.3; 6p12.3; 6p12.2; 6p12.1; 6p11.2 |
| Gender | None |
| EGFR | 1p36.32; 1p36.31; 1p36.23; 1p36.22; 1p36.21; 1p36.13; 1p36.12; 1p36.11; 1p35.3; 1p35.2; 1p35.1; 1p34.3; 1p34.2; 1p34.1; 1p31.3; 2q12.2; 4p16.1; 4p15.33; 4p15.32; 4p15.31; 4p15.2; 4p15.1; 4p13; 7p22.2; 7p22.1; 7p21.3; 7p21.2; 7p21.1; 7p15.3; 7p15.2; 7p15.1; 7p14.3; 7p14.2; 7p14.1; 7p13; 7p12.3; 7p12.2; 7p12.1; 7p11.2; 12q21.2; 14q24.3; 14q31.1; 14q31.2; 14q31.3; 14q32.11; 14q32.12; 14q32.13; 15q26.12; 15q26.2; 15q26.3; 16p13.3; 16p13.2; 16p13.13; 16p13.12; 16p13.11; 16p12.3; 16p12.2; 21q21.3; 21q22.11 |
| KRAS | None |

**NOTE:** In bold, ethnic-specific cytobands, selected by multivariate analysis.

*Ethnic-specific cytobands that were exclusively amplified or deleted in one series but not in the other.
Ethnicity
Twenty-eight chromosomal cytobands exhibited CNAs significantly related to ethnicity. Significantly higher rates of copy number gain on chromosomes 1 (1p36) and 16 (16p13, 16p12, 16p11) were observed in tumors from EA patients. In contrast, higher rates of copy loss on chromosome 19 (19p13) were identified in tumors from WE patients.

EGFR status
Fifty-nine chromosomal cytobands exhibited CNAs distributions significantly related to EGFR mutation. In EGFR mutant tumors, we identified significantly higher rates of copy number gain on chromosomes 1 (1p36, 1p35), 7 (7p22-21, 7p15-12), 16 (16p13, 16p12), and 14 (14q31-32). Conversely, higher rates of copy loss on 21q (21q21-22) were observed in tumors with EGFR wild-type status. As previously reported, we observed a strong positive correlation between EGFR amplification and mutation (P < 10^{-5}).

Smoking history
Sixteen chromosomal cytobands with distributions of CNAs significantly related to smoking history were selected. Significant higher rates of copy gains were observed on 1p (1p34) and 16p (16p13, 16p12, 16p11) among nonsmokers.

Figure 3. Univariate analysis. The test statistics [−log_{10}(P)] testing the null hypothesis of identical chromosomal distributions for the 2 tested groups are shown according to genome position order from 1pter to 22qter. Genomic segments with a FDR below 5% are highlighted in orange. Test statistics are shown for ethnicity (A), EGFR (B), smoking status (C), and gender (D).
Gender and KRAS status

No significantly associated CNAs were identified for gender and KRAS mutation status.

Some CNAs associated with ethnicity were also related to smoking history and EGFR mutation status, such as chromosomes 1 (1p36, 1p35) and 16 (16p13, 16p12, 16p11). This finding suggests a complex interplay between these distinct factors that ultimately combine to shape the NSCLC genome.

To unravel the genomic contributions of ethnicity, smoking, and EGFR mutation status, we applied multivariate modeling to identify CNAs significantly related to ethnicity while accounting for the potentially confounding effects of EGFR mutation status and smoking history. We identified 7 chromosomal cytobands significantly associated with ethnicity in the multivariate analysis (Table 2, in bold). It is worth noting that for the multivariate model, genomic regions such as 1p36 and 1p35, initially identified as related to ethnicity in the univariate analysis, were no longer retained, suggesting that these regions are mainly related to other factors. Specifically, the high rate of amplification on 1p, previously observed among EA patients, is likely related to EGFR mutation. Similarly, several CNAs on 16p were not retained in the multivariate model because they are mainly related to smoking history (16p11.2) or EGFR mutation (16p13.12, 16p12.2, 16p12.1).

We then focused on ethnic-specific genomic segments that were either exclusively amplified or deleted in only one series but not the other (contrasting chromosomal aberration patterns). Using this filter, we identified 4 ethnic-specific genomic areas: 16p13.13, 16p13.11, and 19p13.3 on 19p deletion. The finding of only one LKB1 mutation (W308L) occurring in a WE tumor concurrently exhibiting 19p deletion. The finding of only one LKB1 mutation (W308L) confirms previous findings showing that LKB1 inactivation in lung cancer is more often linked with copy loss than somatic mutation (25). We also identified 4 cases of MYH11 genetic alterations occurring in EA tumor samples—of these, 2 were germ line or constitutional alterations (N312S in both samples) whereas 1 was a somatic mutation (R280C).

Discussion

In this study, we analyzed 226 histologically homogeneous lung tumors (adenocarcinomas) from 2 distinct series. To the best of our knowledge, this is the first study to specifically compare CNAs from lung tumors of EA and WE origins. Our use of a common technology platform,
with standardized DNA extraction protocols, quality parameters, and array profiles generated at a single center avoids the inherent difficulty of comparing results from different publications that are often based on different technology platforms, data analysis methods, and study designs.

The 2 series considered in this study recapitulated several previously known clinical differences, with EA patients being associated with higher proportions of females, non-smokers, and tumors with EGFR mutations. Given these clinical differences and the complexity of chromosomal aberrations observed in individual lung tumors with more than a quarter of genomic regions showing nonrandom propensities for either copy loss or gain, it is perhaps remarkable that the overall genomic profiles of the 2 series were strikingly similar, with classical recurrent copy gains (1q, 5p) and copy losses (3p, 9q) observed in both series. The high similarities in genomic profiles suggest that the basic underlying oncogenic pathways regulating lung carcinogenesis are similar in both populations. Consistent with this notion, aberrations in well-known lung cancer oncogenes such as MYC, CCND1, and TFF-1 were observed in both series.

We analyzed the relationship between CNAs frequency and clinicomolecular characteristics. Univariate analyses revealed that ethnicity, EGFR mutation, and smoking history were significantly associated with several chromosomal aberrations whereas no significant relationships were identified for gender and KRAS or TP53 mutation. Taking into account EGFR mutation and smoking status, 4 ethnic-specific CNAs were revealed on 16p13.13, 16p13.11, 19p13.3, and 19p13.11. These areas are still selected (but considering a higher rate of false-positive discovery rates) when focusing only on stage I lung adenocarcinomas. However, it is worth noting that the top association signals are mainly located on 3 areas: 16p13.13, 16p13.11, and 19p13.3, whereas 19p13.11 shows weaker signals. It is worth noting that, in our study, 1p36-35 was selected in univariate analysis but not in the multivariate model. It suggests that the high rate of 1p36-35 copy gain in EA patients is more likely explained by its relationship with EGFR mutation than with ethnicity. This idea is supported by recent work from a Japanese series from Shibata and colleagues (26), which showed that 1p36 amplifications are correlated with EGFR mutation status. Among the candidate genes located on 1p36 is mTOR/FRAP kinase, which controls cell growth and is deregulated in many human cancers. mTOR inhibitors are currently being tested in cancer clinical trials (27), and it will be of interest to test the relationship between mTOR amplifications and resistance to EGFR TKIs.

19p13.3 is exclusively deleted in the WE series. This strongly suggests the activity of a potential tumor suppressor gene in this population, with a highly plausible candidate gene being the LKB1/STK11 gene. Constitutional mutations in LKB1/STK11 have been associated with Peutz-Jeghers syndrome, an autosomal-dominant disorder characterized by the growth of polyps in the gastrointestinal tract, and a high frequency of somatic mutations in LKB1 have been described in NSCLC. Consistent with LKB1 acting as an important lung cancer tumor suppressor gene in Caucasian patients, LKB1 mutation rates are significantly higher in NSCLC tumors from the United States than from Korea (17% vs. 5%; ref. 28).

16p13.13, 16p13.11, and 16p12.3 exhibited exclusively amplified patterns in the EA series, strongly suggesting the activity of an oncogene. Although 16p13 amplifications have been suggested to be linked with nonsmoking histories (11, 29), we show here that even taking tobacco history into account, a high level of amplification is still observed in EA patients, suggesting that 16p13 amplifications are also related to ethnicity.

Among the potential oncogene candidates located on 16p13, the MYH11 gene is notable, as activating MYH11 mutations have been shown to occur in human colorectal cancer. Moreover, germ line mutations in MYH11, similar to LKB1, have been shown to cause features similar to Peutz–Jeghers syndrome (30). In a preliminary analysis, we have observed MYH11 somatic mutations in a subset of cases of our EA series. Specifically, one somatic mutation [R280C mutation (c–>t)] was found in an EA case that was also amplified for 16p13. Additional results are needed to confirm MYH11 as the driver gene. Among other candidate genes, we also note the RRN3 gene that is required for efficient transcription initiation by RNA polymerase I. However, to the best of our knowledge, no amplification has been described in carcinomas.

Our discovery of prominent, ethnic specific CNAs in NSCLC indicates that patient ethnicity status should be an important consideration in the design of clinical trials evaluating NSCLC-targeted therapies acting on biological pathways that involve genes located on these areas. Moreover, our results strongly argue that the recruitment and analysis of biological samples from diverse ethnic populations should represent an explicit goal of large-scale cancer genome projects such as the International Cancer Genome Consortium (31).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

This work was supported by grants from the Agency for Science, Technology, and Research, the Duke-NUS Graduate Medical School and the French Cancer Research Association (ARC).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 16, 2010; revised April 3, 2011; accepted April 6, 2011; published OnlineFirst April 26, 2011.
References

Genomic Profiles Specific to Patient Ethnicity in Lung Adenocarcinoma

Philippe Broët, Cyril Dalmasso, Eng Huat Tan, et al.

Clin Cancer Res  Published OnlineFirst April 26, 2011.

Updated version  Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-10-2185

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/early/2011/05/13/1078-0432.CCR-10-2185. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.