Introduction

Lung cancer is a disease with significant burden, with nearly 2.5 million new diagnoses in 2011 contributing to almost 1.5 million deaths worldwide (1). However, no longer is lung cancer managed by distinguishing non-small cell lung cancer (NSCLC) and the associated subtypes from small cell lung cancer (SCLC), but as variety of distinct, although related, diseases each with requiring their own treatment options.

NSCLC make up approximately 85% (2) of lung cancers, which is then further broken down into three distinct histological subtypes (3): adenocarcinoma, squamous cell carcinoma and large cell carcinoma (LCC). Adenocarcinoma comprises the majority of all new lung cancer diagnosed with an associated fall in the proportion of squamous cell cancers (4,5).

Epidermal growth factor receptor (EGFR), is one of several somatic mutations, in NSCLC (6), which is seen more frequently in certain population groups. This population group is classically described as Asian, non-smoking females with adenocarcinoma (7-9). The interest in these mutations is due to the small molecule targeted therapies (such as erlotinib and gefitinib) available and in development, which can have significant prognostic benefits (10,11).

The role of EGFR in NSCLC

The EGFR is a 170 kdalton member of the ErbB family of cell surface tyrosine kinases (12) and is encoded on chromosome 7. The receptor belongs to the HER/erbB family of tyrosine kinases, which include HER1 (EGFR/erbB1), HER2 (neu, erbB2), HER3 (erbB3), and HER4 (erbB4) (13). The function of the receptor is to regulate both cell proliferation and apoptosis via signal transduction pathways (14).

The EGFR is a transmembrane receptor consisting of three portions; an extracellular ligand-binding domain, a transmembrane domain and an intracellular tyrosine kinase domain (15). Activation of EGFR is achieved by the binding of a ligand [such as epidermal growth factor, transforming growth factor and neuregulins (16)] to the extracellular portion. The binding of a ligand results in receptor dimerization or heterodimerisation with related receptors [especially HER2/neu (17)] (18). Without a ligand bound to the receptor and the subsequent dimerisation there is no activity at the enzymatic site of the intracellular portion (16).

Once dimerisation occurs there is disruption of the autoinhibitory activity of the intracellular domain resulting in rapid autophosphorylation at tyrosine residues located on the intracellular portion (15,19). The phosphorylated...
receptor then functions to allow assembly and activation of intracellular messenger proteins (18), especially through the mammalian target of rapamycin (mTOR) (20).

Dysregulation of the EGFR leads to increased intracellular pathways activity, via tyrosine kinase autophosphorylation, resulting in directly or indirectly, cell proliferation, angiogenesis, invasion and metastasis (12).

Overexpression of the EGFR gene has been identified in a variety other cancers including: head and neck, ovary, cervix, bladder, oesophagus, stomach, brain, breast, endometrium, colon and lung (21). EGFR overexpression has been identified in between 40% to 89% of NSCLC (6,22), with highest rates seen in squamous tumours (89%) and lowest in adenocarcinomas (41%) (22).

Tyrosine kinase domain mutations

As EGFR was noted to be overexpressed in NSCLC, it was felt that regulating the receptor with an tyrosine kinase inhibitor (TKI), gefitinib, would be an effective treatment for NSCLC, however this was not shown to be case (23). However, during the initial trial of gefitinib, a subgroup of patients were identified that had significant improvement in their lung and metastatic lesions (6). The identification of a particular subgroup of patients with dramatic response to TKI treatment led to molecular investigation of the EGFR pathway. This subgroup of patients was analysed separately by both Lynch et al. and Paez et al. who each showed that patients who possessed mutations in the tyrosine kinase domain of the EGFR (6,24). These mutations were shown to occur in exons 18, 19 and 21.

Analysis of the tyrosine kinase domain of the EGFR of 617 unselected lung cancer specimens by Shigematsu et al. identified that all mutations occurred within exons 18-21, with a prevalence of 21% (7). These mutations (listed in Table 1) provide sensitivity to targeted therapies, known as TKIs, such as erlotinib and gefitinib (21).

The majority of mutations in exon 21 are point mutations whereas exon 19 consists of almost entirely in-frame deletions (20). The L858R point mutation and ∆E746-A750 comprise up to 86% of all EGFR mutations in some studies (25). Both the aforementioned mutations result in changes near the ATP cleft, which results in enhanced catalytic activity and autocatalysis of the tyrosine kinase when the receptor is not stimulated by EGF (or other ligands), with up to a three-fold increase in activity compared to the wild-type EGFR (6).

Whilst most tyrosine kinase domain mutations lead to sensitivity to TKIs (Table 1), mutations in exon 20 are associated with intrinsic resistance (26-31) which may account for up to 9% of all EGFR mutations (31).

While squamous cell carcinomas (SCC) rarely possess mutations in the tyrosine kinase domain of the EGFR receptor, about one-third of SCCs demonstrate amplification of the EGFR protein (2). Approximately 5% of SCC possess deletion mutations in exons 2-7 (EGFRvIII) which code for the extracellular domain of the protein (32). In the same series no adenocarcinomas possessed EGFRvIII mutation, however the extracellular domain mutations are frequently seen in SCCs of head and neck cancers (33).

Histology

Amongst the various forms of NSCLC, adenocarcinoma, is

| Table 1 List of EGFR mutations in NSCLC resulting in sensitivity or resistance to first generation TKIs (21) |
|---|---|---|---|---|
| TKI sensitivity | Exon 18 | Exon 19 | Exon 20 | Exon 21 |
| Resistance | D761Y, D770_N771 (ins NPG), D770_N771 (ins SVQ), D770_N771 (ins G), N771T V769L, S768I, T790M |

EGFR, epidermal growth factor receptor; NSCLC, non-small cell lung cancer; TKIs, tyrosine kinase inhibitors.
most commonly identified in all comers tested for EGFR mutation status (34-40). Bronchioalveolar cell carcinoma (BAC), a subtype of adenocarcinoma, was associated in some of the early gefitinib studies (6,41) with response to treatment. As most NSCLCs do not respond to gefitinib, unless they have the activating mutation, then this would suggest that BAC is more commonly associated with EGFR mutation than other forms of adenocarcinoma. A retrospective audit of 139 NSCLC patients treated with gefitinib, by Miller et al., revealed that significantly more patients, who experienced response to TKI therapy, possessed BAC features than those that did not receive a benefit from drug therapy (38% vs. 14%, P<0.001) (41).

BAC was then further divided into mucinous, non-mucinous carcinomas and mixed non-mucinous and mucinous or indeterminate in the World Health Organisation histological classification of tumour guide (3). However, since 2004, further clarification of the term BAC has occurred and subsequently recommended the discontinuation of the term BAC in preference for the following categories; adenocarcinoma in situ, minimally invasive adenocarcinoma (mucinous and rarely mucinous), lepidic predominant (non-mucinous), adenocarcinoma predominantly invasive with some non-mucinous lepidic components and, finally, invasive mucinous adenocarcinoma (42). The latter two are the forms of BAC formerly referred to as nonmucinous BAC and mucinous BAC respectively (42). As the studies referenced below present their data using the original nomenclature, the data will be presented using the papers author’s original terms to ensure that no inappropriate interpretation is undertaken.

In analysing 141 primary NSCLC biopsies, of which 118 were adenocarcinomas from a Japanese population, Sakuma et al. demonstrated that 54% (P<0.0001) of the adenocarcinomas with EGFR mutations possessed histopathological features consistent with nonmucinous BAC (43). Similarly, Marchetti et al. found that the 56% (P=0.00002) of adenocarcinomas with EGFR mutation were BAC with all patients possessing a nonmucinous subtype (44). However, while Tam et al. also demonstrated that nonmucinous BAC was significantly associated with EGFR mutation (79%), only 13% of adenocarcinomas with an identified EGFR mutation were of BAC subtype (45).

Using the updated histological nomenclature from international association for the study of lung cancer (IASLC), the American Thoracic Society (ATS), and the European Respiratory Society (ERS) (42), Yoshizawa et al. analysed 440 resected lung adenocarcinomas. They demonstrated that 167 cases were positive for EGFR mutation with a high rate of features consistent with adenocarcinoma in situ (85.7%), minimally invasive adenocarcinoma (83.3%), lepidic (71.4%) and papillary predominant (68.5%), while there were no mutations identified in mucinous subtype tumours (46). Using the same criteria, Gahr et al. demonstrated that of the 101 patients with EGFR positive NSCLC (from a population of 1,122), 90% were nonmucinous adenocarcinoma, with only 22% poorly differentiated. Further divided down 65.3% of EGFR positive tumours had features consistent with non-lepidic-nonmucinous adenocarcinoma and 21.8% lepidic-nonmucinous histology (35). In a population of Korean smokers (n=249), of the 51% with EGFR positive NSCLC, when classifying the tumour on the major histological subtype, the most common finding was acinar (68.5%) followed by papillary (11.8%), solid (9%), lepidic (7.5%), micropapillary (1.4%) and only 1.8% falling into the invasive mucinous category (47).

While the vast majority of EGFR mutations in NSCLC are found in adenocarcinoma, the mutation is also seen in SCC and LCC. Comparing 15 studies (Table 2), the majority of which were in selected patient populations, the prevalence of EGFR mutation positive SCC lung cancer ranged between 0-14.6%, with an average of 4.9% when the 4,870 patients were combined into a single group.

### Epidemiology

The conventional phenotype of patients who develop a lung cancer that is positive for an EGFR mutation is the young, Asian, non-smoking, female with adenocarcinoma (7-9). While this does certainly appear to be the case, there are very few studies that have prospectively analysed non-selected populations of patients. The vast majority of papers, that examine the predictors and prevalence of EGFR mutations, recruit patients with advanced stage disease or who have failed alternate therapies (surgical or first-line chemotherapy). Even those studies that do not select for patient populations commonly have intrinsic selection bias, by the very fact that they recruit patients from a single country with homogenous ethnic populations.

Of the eight papers identified (35,36,39,58-62), which measured the frequency of EGFR mutations in NSCLC in unselected patients, only four clearly indicated that the data was gathered in a prospective manner (35,36,60,62). There was a range of mutation testing, with the majority of papers examining for mutations in exons 18-21, but some limiting...
their investigation to only the common mutations (exon 19 deletions and L858R substitution). None of the studies examined for the effect of race on the presence or absence of the EGFR mutation.

The findings of the eight studies are listed in Table 3 (individual studies were excluded from analysis in the case of missing data).

### EGFR prevalence

The prevalence of the various EGFR mutations tested for was 13.9% of the 7,595 patients with the highest prevalence (36.4%) of mutations seen in the single study conducted in Japan.

### Smoking

In those patients with the EGFR mutation, nearly 60% of patients were identified as never-smokers [or less than 20 years in one study (39)]. The prevalence of never smokers with the mutation was 42%, whereas the mutation was still identified in 10.7% of current or former smokers. The variation of the mutation presence was identified as significant in 5 of the 6 studies where statistical analysis was performed.

### Sex

In the EGFR mutation group, 64.9% of the patients were female, while the prevalence of the mutation overall was 25.8% for females but only 12.2% for males. This was statistically significant for all studies that tested for significance.

### Age

No correlation with age and the presence or absence of the mutation was identified.
Tumour type

In those studies where enrolment criteria were not limited to adenocarcinomas, only 3-9% of tumours identified did not possess histology consistent with adenocarcinoma. The analysis of exact tumour type could be limited in these studies as histological analysis can be difficult in cytology only specimens (such as obtained with fine needle aspiration). Only one of the studies indicated the source of tumour specimens used in mutation analysis.

Discussion

When Lindeman et al. analysed the EGFR mutation prevalence rate, divided by race from multiple previous studies, they found that amongst the Asian/Indian population the prevalence was 52% when compared to only 24% amongst Caucasians (63). Shigematsu et al., in multi-nationality study (primarily South East Asia and Caucasians) of patients with resectable disease found an overall mutation prevalence of 23% (7). However, when divided by race the mutation rate amongst Australians and North Americans was 7% and 14% respectively, whereas the mutation rate in Asian countries was as high as 34% in the Taiwanese population.

It is difficult to resolve the wildly varying prevalence of the EGFR mutation in the above studies. The analysis performed in this paper on an unselected population of 17,712 mainly European and American patients is close to Shigematsu et al. overall prevalence calculation. When Lindeman et al., Shigematsu et al. and this papers analysis are compared (Table 4) the overall prevalence in the unselected cohort (this study) is similar to that of the non-Asian population. The final row in Table 4 was obtained by calculating the prevalence of each stated factor in the

### Table 3 Studies which analyse all comers for EGFR mutation positive NSCLC

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Total patients</th>
<th>EGFR positive</th>
<th>Female</th>
<th>Males</th>
<th>Never smokers</th>
<th>Smokers or former smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gahr et al. (35)</td>
<td>2013</td>
<td>Germany</td>
<td>1,201</td>
<td>118</td>
<td>81 385</td>
<td>37 698</td>
<td>38 118</td>
<td>22 506</td>
</tr>
<tr>
<td>Locatelli-Sanchez et al. (36)</td>
<td>2013</td>
<td>French</td>
<td>753</td>
<td>121</td>
<td>76 210</td>
<td>45 422</td>
<td>73 NA</td>
<td>48 NA</td>
</tr>
<tr>
<td>Paik et al. (59)</td>
<td>2012</td>
<td>United States</td>
<td>675</td>
<td>164</td>
<td>77 329</td>
<td>48 182</td>
<td>110 183</td>
<td>54 328</td>
</tr>
<tr>
<td>D’Angelo et al. (62)</td>
<td>2011</td>
<td>United States</td>
<td>2,142</td>
<td>503</td>
<td>346 969</td>
<td>157 670</td>
<td>302 278</td>
<td>201 1,361</td>
</tr>
<tr>
<td>Tanaka et al. (39)</td>
<td>2010</td>
<td>Japan</td>
<td>308</td>
<td>112</td>
<td>60 41</td>
<td>52 155</td>
<td>59 37</td>
<td>43 150</td>
</tr>
<tr>
<td>Rosell et al. (60)</td>
<td>2009</td>
<td>Spain</td>
<td>2,105</td>
<td>350</td>
<td>244 570</td>
<td>106 1,181</td>
<td>231 381</td>
<td>116 1,266</td>
</tr>
<tr>
<td>Fontanini et al. (61)</td>
<td>2009</td>
<td>Italy</td>
<td>411</td>
<td>52</td>
<td>37 139</td>
<td>19 220</td>
<td>14 38</td>
<td>16 149</td>
</tr>
<tr>
<td>Beau-Faller et al. (58)</td>
<td>2014</td>
<td>France</td>
<td>10,117</td>
<td>1,047</td>
<td>NA NA</td>
<td>NA NA</td>
<td>NA NA</td>
<td>NA NA</td>
</tr>
</tbody>
</table>

EGFR, epidermal growth factor receptor; NSCLC, non-small cell lung cancer; WT, wild type; NA, not available.

### Table 4 Prevalence of EGFR mutation

<table>
<thead>
<tr>
<th>Study</th>
<th>Prevalence of EGFR mutation</th>
<th>Males (%)</th>
<th>Females (%)</th>
<th>Never-smokers (%)</th>
<th>Smokers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shigematasu et al. (7)</td>
<td></td>
<td>14</td>
<td>42</td>
<td>51</td>
<td>10</td>
</tr>
<tr>
<td>Lindeman et al. (non Asians) population (63)</td>
<td></td>
<td>18</td>
<td>28</td>
<td>45</td>
<td>15</td>
</tr>
<tr>
<td>Lindeman et al. (Asian population) (63)</td>
<td></td>
<td>32</td>
<td>58</td>
<td>58</td>
<td>26</td>
</tr>
<tr>
<td>Mitsudomi et al. (64)</td>
<td></td>
<td>10</td>
<td>38</td>
<td>47</td>
<td>7</td>
</tr>
<tr>
<td>Summary of Table 3</td>
<td></td>
<td>12.2</td>
<td>25.8</td>
<td>42.2</td>
<td>10.7</td>
</tr>
</tbody>
</table>

The final row contains a summary of the data obtained from the studies listed in Table 3. EGFR, epidermal growth factor receptor.
studies listed in Table 3 (studies with incomplete data were excluded from individual analyses).

The prevalence of the individual sub-populations (males, females, never-smokers and smokers) are similar between the three papers when Caucasian population is considered. This suggests that the EGFR mutation is far more prevalent in Asian populations than Caucasians. Logistic regressions have only demonstrated that a low smoking history and adenocarcinoma histology are significant independent predictors of EGFR mutation status, but not sex nor age (39).

**Conclusions**

EGFR mutations are significant drivers in NSCLC, especially amongst Asian females who are never-smokers with adenocarcinoma histology. However 10% of patients with EGFR mutant NSCLC have some degree of smoking history and 12% are male. Simple choosing to only mutation test patients who fit a single phenotype will miss a significant proportion of sufferers who may benefit from small molecule therapy.

Current studies on the prevalence of the mutation tend to focus on a single race and many do not test for the presence of the mutation in all lung cancer stages. Despite smoking remaining the highest risk for lung cancer (5,65), there is a rising incidence of adenocarcinoma in non-smokers (66). Having an accurate model of who may develop EGFR mutation NSCLC may allow prognostic benefits with targeted therapies.

**Acknowledgements**

**Disclosure:** The author declares no conflict of interest.

**References**


63. Lindeman NI, Cagle PT, Beasley MB, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer,


