# Somatic Mutations of the Tyrosine Kinase Domain of Epidermal Growth Factor Receptor and Tyrosine Kinase Inhibitor Response to TKIs in Non-small Cell Lung Cancer: An Analytical Database

*Samuel Murray, PhD,* Issa J. Dahabreh,† Helena Linardou, MD, PhD,‡ Menelaos Manoloukos, MSc,* Dimitrios Bafaloukos, MD,‡ and Paris Kosmidis, MD§*

**Background:** After the discovery of somatic mutations in the tyrosine kinase domain (exons 18–24) of the epidermal growth factor receptor (EGFR) correlating with responses of non-small cell lung cancer (NSCLC) to the tyrosine kinase inhibitors (TKIs) gefitinib and erlotinib, there has been increasing interest in utilizing this molecular marker for treatment selection. We aimed to analytically catalogue the mutational spectrum of somatic mutations in EGFR and format a database allowing correlation of specific mutations with clinico-pathologic factors and response to TKIs.

**Methods:** A computerized search of MEDLINE (January 1, 2004 to June 30, 2007) was performed to identify articles reporting on NSCLC patients harboring somatic mutations in EGFR. Demographic, clinico-pathologic, mutational, and response data were extracted and tabulated.

**Results:** A total of 202 eligible articles were identified. We report data on 12,244 patients with 3381 somatic EGFR mutations. The majority of mutations have been reported on only one occasion (158 of 254, 62.2%), and only nine mutations occur at a rate of ≥1%. L858R and delE746-A750 account for 32.84% and 24.28% of all mutations, respectively; with 50% of mutations being exon 19 deletions or "deletional-insertions." There is a clear association between the presence of mutations and response to TKI.

**Conclusions:** We have generated a free access, nonprofit online analytical database of somatic EGFR mutations in NSCLC. Cumulative information will be made available through a routine update of both database tables and associated graphical representations. Direct updates and submissions through the online site (www.somaticmutations-EGFR.org) are encouraged, as are comments and suggestions.

*Departments of Molecular Biology and Genetics Metropolitan Hospital, †Department of Pathophysiology, Medical School, National University Medical School of Athens, ‡1st Department of Oncology Metropolitan Hospital, §2nd Department of Oncology Hygeia Hospital, Athens, Greece.

Disclosure: The authors declare no conflicts of interest.

Address for correspondence: Samuel Murray, PhD, Department of Molecular Biology and Genetics, Metropolitan Hospital, N. Faliro, 18547, Athens, Greece. E-mail: smbhsam@hotmail.com or smurray@metropolitan-hospital.gr



ISSN: 1556-0864/08/0308-0832

**Key Words:** Gefitinib, Erlotinib, Non-small cell lung cancer, Smoking, Adenocarcinoma.

(*J Thorac Oncol.* 2008;3: 832–839)

The clinical course of non-small cell lung cancer (NSCLC) has seen potentially important changes over the last few years after the introduction of tyrosine kinase inhibitors (TKIs) as a treatment modality.[1–6] With this introduction, clinical perceptions on both treatment and outcome have also seen some significant changes. Oral bioavailability, and daily administrative schedules are hallmarks of these new molecules that target the epidermal growth factor receptor (EGFR).[7,8] Newfound enthusiasm in the treatment of metastatic disease has however diminished after the rapid completion of a number of rather disappointing phase III studies,[9–12] resulting in the clinical approval for only one of two current 4-anilinoquinazolines, single agent erlotinib for the second-line treatment of NSCLC in the United States and Europe.[4,6]

In early 2004, somatic mutations in the tyrosine kinase (TK) domain of EGFR were identified, correlating with responses to these agents.[1,13,14] Since then numerous groups have compiled data on the mutational spectrum of EGFR with respect to incidence, clinico-pathologic correlates, prognostic and predictive significance, response to TKIs and survival.[15–22] Several questions have been raised regarding the significance of their origin, Ras mutations, EGFR gene copy number, specific clinico-pathologic correlates (smoking status, gender, etc.), and more recently the role of mutations as predictive markers.

Since the first reports of EGFR mutations occurring in NSCLC, there has been an exponentially growing number of publications presenting data on the incidence of EGFR mutations in patients treated with TKIs correlating their presence with responses, and documenting the existence of such mutations in TKI-naive patients. In this systematic review, we aim to retrieve all published information on EGFR mutational status, categorize, and tabulate it according to the relevant clinico-pathologic parameters and finally, initiate the development of an independent patient level EGFR mutation

database supporting the EGFR mutational spectrum while tabulating treatment responses. The database is designed to be flexible for the inclusion of any additional fields that are deemed appropriate in relation to somatic mutations of EGFR. The database provides a graphical overview of the complexity of the somatic mutational spectrum, while allowing readers to gain a picture of the efficacy of TKIs in patients with specific somatic mutations.

## MATERIALS AND METHODS

### Identification of Published Material

We performed a systematic computerized search of the MEDLINE (PubMed) database (last search: July 1, 2007) to identify all published articles from January 1, 2004 to June 30, 2007 dealing with the identification of somatic mutations in EGFR pertaining to NSCLC, using the algorithm: (epidermal growth factor receptor OR EGFR OR gefitinib or non-small cell lung cancer OR NSCLC OR erlotinib OR iressa OR tarceva) AND (mutation OR gefitinib OR iressa). We hand searched journals known to publish data relevant to our search, the reference lists of all articles we recovered and those of relevant review articles were also cross-referenced. We contacted experts in the field of NSCLC to broaden our yield of potentially eligible articles. Because of glitches in the search engine and unintentional handling errors, there is the distinct possibility that one or more articles escaped our search. Through a process of planned author contact and updates it is expected that the database will become more complete.

### Eligibility Criteria

We considered only peer-reviewed published articles with data pertaining to the somatic mutation analysis of the TK domain of EGFR (with a minimum requirement of at least investigation of one of exons 18 through 21), or any specific mutation (i.e., L858R) in NSCLC samples. Abstracts and meeting proceedings were excluded and no language restriction was imposed. Single patient data (including case reports) were included where at all possible. There was no exclusion based on the number of patients screened or completeness of field term identifiers per study. No restriction was placed on the method of mutational analysis (e.g., specific mutations such as L858R by restriction fragment length polymorphism through to whole gene sequencing, bidirectional sequencing, TaqMan probes identifying specific mutations, were eligible); nor the source of biological material (e.g., paraffin embedded biopsy, fresh frozen biopsy, cytologic specimens); nor the nature of the molecular analysis (e.g., RNA or DNA).

When multiple reports on the same population were available, we retained only the report with the largest number of events or largest patient population (where appropriate) to avoid duplication of information. Because of the cumulative nature of the data reporting and size restrictions it was inappropriate to perform any statistical analysis at this time point; therefore, all tabulated data is only descriptive in nature. Furthermore, because of the possibility of repeated data reporting or overlapping data, all studies with fewer than five screened patients have been removed from the final descriptive tabulated data, unless otherwise stated. The mutational spectrum has been generated from cumulating all possible data sets, as has the response per mutation data (inclusive of case reports, wherever not obviously of an overlapping nature).

### Data Extraction

Information recorded about each recovered reference is listed in Table 1. All publications were reviewed independently by two of the authors (SM and IJD), discrepancies were resolved by these two authors. Manually extracted information included EGFR mutational rate, mutation type (spectrum), correlations with clinico-pathologic and demographic data (smoking status, gender, histologic type), and also for data linking specific somatic mutations to treatment outcome (complete response [CR], partial response [PR], stable disease, progressive disease [PD], non-evaluable) with the TKIs gefitinib and erlotinib when administered as single agent, i.e., monotherapy TKI. No stratification has been made according to TKI with respect to response data. Data sets including patient information from populations treated with a TKI in combination with another agent (either chemotherapy or another biological modifier) were not included in the response outcome calculations.

Data sets, particularly those including information overlapping partially with previous publications that allowed extraction of relevant information (see field terms in Table 1) were reported on more than one occasion to achieve the highest completeness of data for inclusion in the database. Subanalyses have taken into account repetitive data presentation after cross-analysis and comparisons based on the list of authors, host-institution, relevant clinical studies, and sources(s) of biological material(s). Updates and author confirmation will be conducted routinely to improve data accuracy. Because of the complexity of some published data sets, more than one data entry may have been included in the database to reflect completeness of data extraction.

### Terminology and Fields

Field terms corresponding to data extracted from each article regarding mutational spectrum and clinico-pathologic correlates are explained in Table 1. To simplify the current complexity with respect to somatic mutations of EGFR, we have reported mutations in a simplified manner (Table 2).

### Notation

Mutations were not labeled directly according to the Human Genome Variation Society (www.hgvs.org/mut-' nomen) guidelines. Mutations were labeled according to the amino acid (AA) sequence, (reference mRNA sequence gi: 29725608; GenBank Accession NM_005228.2) with the translation initiation site methionine (ATG) corresponding to nucleotide 167. Various authors have used the common Human Genome Variation Society nomenclature; however, we believe that the majority of clinicians, and therefore the majority of diagnostic laboratories, will ultimately report the mutational spectrum with respect to alterations in the AA sequence and not necessarily in the nucleotide sequence.

 

**TABLE 1.** Field Terms Extracted

| | |
|---|---|
| Reference | Data sets represented by a given reference, references may appear on more than one occasion depending on the extent of data extraction. i.e., treated and nontreated populations from the same article are presented separately if possible |
| Mutation type | Reported as per Table 2. All refer to the amino acid sequence number assigned according to reference sequence gi: 29725608. Amino acid alterations are reported as single letter abbreviations |
| Confirmed somatic | Confirmed to be somatic by analysis of normal cells; yes or no (in most instances not all samples have been proven somatic) |
| ≥2 mutations | Patients with ≥2 reported somatic mutation are noted, these include patients with T790M, BUT do not include patients with silent polymorphisms or germ line single nucleotide polymorphisms (SNPs). Patients with T790M mutations have typically developed the mutation following TKI-based therapy; therefore, the majority of these patients have already been included in the tables (both mutation and response). Readers are advised to investigate these cases with caution. All other cases of dual (or multiple) mutations occurring in one individual have been recorded in both the mutation and response tables as independent events according to each individual mutation. Therefore, there are additional patient numbers associated throughout the data set and the associated tables. Considering that they represent approximately <1.7% of the total population readers are advised to interpret the presented tables and data sets bearing this in mind |
| Gender | ♂ = male; ♀ = female |
| Ethnicity | Based on ethnicity of the biological specimens used in the study. More than one ethnicity may occur per reference |
| Pathology (histological type) | Adenocarcinoma = adenocarcinomas combined with any histology with BAC component (bronchioalviolar component, e.g., BACs, AWFB, BWFI), adenosquamous are included with other; Other, all other histologies |
| Smoking status | Smoker = current, previous with no limit on pack year; nonsmoker = assumed never smokers |
| TKI | Tyrosine kinase inhibitor (TKI); E = erlotinib (Tarceva); G = gefitinib (Iressa). As yet no data has been assigned to combination versus single agent therapy, nor treatment period, or disease stage; C = chemotherapy |
| Response criteria | Reported as RECIST (R), WHO, SWOG |
| Author confirmation and date | Date at which corresponding author confirmed the data set |
| Sample source | PET = paraffin embedded tissue; FF = fresh frozen tissue; Other = as indicated |
| Analyte | DNA = mutational analysis conducted on extracted DNA; RNA = mutational analysis conducted on tissue extracted RNA |
| Exons examined | Exons screened per study (individual mutation through to entire gene) |
| Technique | Seq = bidirectional sequencing; RFLP = restriction fragment length polymorphism; TaqMan = TaqMan probes (typically 13 different mutations, unless otherwise stated); surveyor = combined Taqman-like format with enhancement of mutant allele; SSCP = single strand conformational polymorphism; MEP = mutant-enriched PCR |
| ^ | Not reported, or not directly extractable from the data source |
| Treatment response | CR = complete response; PR = partial response; SD = stabilization of disease; PD = progressive disease; NE = nonevaluable |
| Treatment status | Status at the time of TKI initiation: *P* = previously treated with chemotherapy, determined to be ineligible to receive standard chemotherapy; N = chemotherapy naïve |
| No. of patients—WT | Number of samples with wild type EGFR sequence (WT) |
| No. of patients—mutant | Number of samples with a somatic mutation in EGFR (Mut) |

**TABLE 2.** Mutation Terminology

| | |
|---|---|
| Deletions | Consecutive nucleotide loss |
| Insertions | Consecutive nucleotide insertion and/or duplication |
| Duplications | Duplications, as in insertions |
| Deletional-insertions | There are a variety of possible scenarios for this event including (1) colocalization of an insertion and a deletion; (2) two nonadjacent (≥1nt) mutations being either dual insertions or deletions; (3) more complex rearrangements. All of these have been simplified to align directly with the coding region and are reported simply as deletions (across the region) plus the associated amino acid insertion |
| Mutation | Only those that change the reference amino acid sequence (gi: 29725608), i.e., missense and frameshift mutations that alter the amino acid sequence have been included, however, silent somatic mutations have not been included in the analysis and are separately reported in the database |

## Database Format

Clinico-pathologic and response information and the mutational spectrum were tabulated. These make up the backbone of the database (www.somaticmutations-EGFR.org) and are maintained in an Excel format as simple spreadsheets. All figures and tables have been compiled by direct data extraction utilizing Microsoft Power-Point and Excel. All data sets and subanalyses will be available as PDF files through the Web site (updated q6m), and as supplemental data through the journal.

## Other Mutations

All other mutations will be incorporated in a separate list of mutations, namely mutations occurring outside of exons 18–21, and also all silent mutations.

## Limitations

There are numerous limitations associated with such a database. Many of these are discussed in Supplementary Notes I.

## Updates

The database will be updated q6 months after a continuing literature search conducted as mentioned. Data will be cross-checked, authors contacted through the database for confirmation, and updates made of all figures and tables posted on the Web site.

## Data Submissions

An Excel-based worksheet containing all of the fields included in this database is available from the Web site (www.somaticmutations.org) for authors willing to directly submit their information. All information will be manually reviewed and all submitting authors will be contacted for completeness before inclusion within the database.

# RESULTS

## Data Extraction

In this review, we have compiled data according to the trial flow (Figure 1). During the search period, a total of 12,244 patients (extracted from 202 eligible articles from an original pool of 2385 articles of potential interest) have been screened for somatic mutations in EGFR. Over this time, the number of both publications and reported somatic mutations in EGFR has increased dramatically. A complete set of extracted data is presented within the primary database of the site. Data is available from a total of 2548 patients treated with a TKI for which somatic EGFR mutational data is available (131 eligible articles with treatment:response data). Considering that ethnicity is correlated with the incidence of EGFR mutations all tabulated data has been stratified for ethnicity, Asians versus Whites.

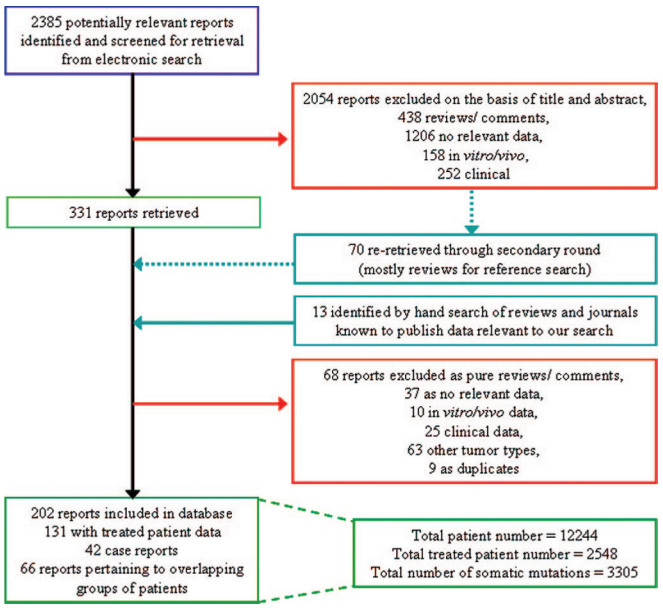


**FIGURE 1.** Trial flow of database generation and data extraction.

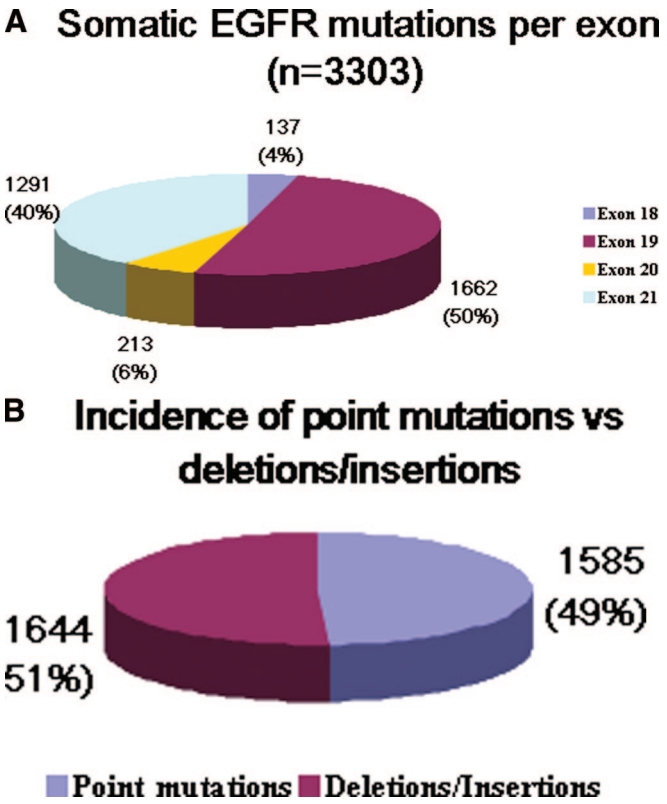


**FIGURE 2.** Distribution of somatic EGFR mutations. *A*, per exon and *B*, according to type (entire population).

## Mutational Spectrum and Incidence

We have catalogued the somatic mutational spectrum (principally of exons 18–21) of EGFR identifying a total of 3381 independently reported somatic mutations from 12,244 screened individuals. In some cases, because neither the specific mutation identification techniques nor for the lack of data reporting a number of these mutations have been localized within any specific exon, therefore, the total number of mutations is 3303 (Figure 2*A*). It must also be noted that because of the possibility of more than one somatic mutation occurring in any given sample, the actual number of individuals with somatic mutations of EGFR is 3188. The overall mutational spectrum and the reported incidence for each mutation are also graphically displayed per exon within the database. The "deletional" and "insertional" mutations that occur in exons 19 and 20 have been independently represented, as have single AA substitutions per exon.

The relative frequency of the different mutational types per exon is indicated in Table 3. From a total of 186 AAs that constitute exons 18–21, 109 (58.6%) are implicated in at least one form of somatic mutation. Figure 3 indicates the relative frequency of mutation reporting in relation to the number of occasions a specific mutation has been reported. The majority of the mutational spectrum occurs at a very low rate, with 158 of 254 (62.2%) different mutational types being reported on one occasion. Furthermore, by adding mutations that have only been reported in one article, this increases to 181 (71.3%). Only a

 

**TABLE 3.** Frequency and Types of Independent Somatic Mutations Identified

| | Exon | | | | Total |
|---|---|---|---|---|---|
| | 18 (41 AA's) | 19 (32 AA's) | 20 (61 AA's) | 21 (52 AA's) | (186 AA's) |
| Amino acids affected | 20 (48.8%) | 25 (78.1%) | 31 (50.8%) | 33 (63.5%) | 109 (58.6%) |
| Insertions and/or duplications | 0 | 4 | 27 | 0 | 31 |
| Deletions and/or deletional-insertions | 0 | 70 | 8 | 0 | 78 |
| Mutations | 38 | 24 | 39 | 44 | 145 |
| Total | 38 | 98 | 74 | 44 | 254 |

AA, amino acid. Two independent exon 22 somatic mutations have been identified, E884K and V897I (one occasion each).

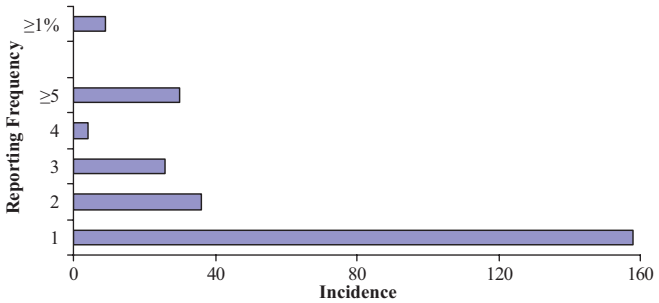


**FIGURE 3.** Frequency of reported incidence of somatic mutations in EGFR. Incidence of somatic mutations that have been reported on 1, 2, 3, 4, and ≥5 separate occasions. Also the number (incidence) of mutations that have a cumulative frequency of ≥1% of the total number of mutations reported ($n$ = 3305).

small proportion of all mutations have actually been reported at a frequency of ≥5 occasions.

The two most common mutations, L858R located in exon 21 representing 32.84% of all mutations and delE746-A750 representing 24.28%, account for the majority of all reported mutations. Deletions and "deletional-insertions" in exon 19 alone account for 47.35% of all reported mutations, with deletions and deletional-insertions combined from all exons constituting more than half of all mutations (Figure 2*B*). From the large number of different mutations occurring within these four exons, there are only a few that occur at a frequency of close to or ≥1% (Table 4).

## Efficacy

A cumulative representation of the overall response per exon and response per mutational type are shown (Figure 4). Mutations in exon 19 have an 86.21% overall response rate compared with only 33.33% for exon 20. Further subanalysis of these responses indicates that, in the absence of a T790M somatic mutation, the chance of an individual responding to TKI monotherapy in the presence of a mutation in exon 20 is approximately 68%, comparable to that of exon 18. Except for the T790M, few other exon 20 mutations actually display de novo resistance to TKI treatment. In fact, out of 115 different mutations for which response data were available, there are only 28 (24.35%) that are associated with lack of response, and 13 different mutations that have only demon-

**TABLE 4.** Incidence of the Most Common Specific Somatic Mutations of EGFR in NSCLC. Mutations and Incidence (Approximate %) are Reported as Determined Utilizing Independent Specific Mutation Data ($n$ = 2776)

| Mutation | Incidence (%) |
|---|---|
| E709 | 0.86 |
| G719 | 2.99 |
| delE746-A750 | 28.89 |
| delE746-T751insA | 0.86 |
| delE746-S752insV | 1.30 |
| delL747-A750insP | 1.73 |
| delL747-T751 | 1.62 |
| delL747-S752 | 0.79 |
| delL747-P753insS | 2.49 |
| S768 | 0.68 |
| T790M | 2.23 |
| L858 | 39.37 |
| L861 | 1.33 |

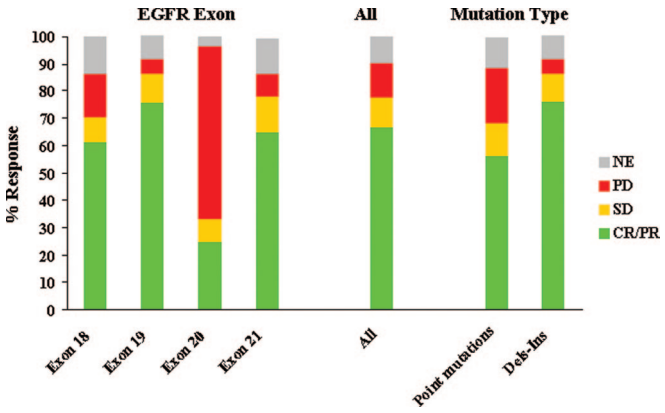


**FIGURE 4.** Response rates for somatic mutations per exon, total population, and per mutational type (cumulative data of treated patients for which response data exists). Responses are characterized as CR + PR, stable disease (disease stabilization), PD, nonevaluable. Total represents the percentage of each response group per stratification out of 100%. Point mutations represent missense mutations and frameshift mutations, Dels-Ins are a combination of deletions, insertions, and deletional-insertions (according to Table 2). All = all mutations.

strated PD as a response (8 of which are located in exon 20). In total, from 776 cases where response can be correlated to a specific somatic mutation, there are 102 (13.14%) cases of PD. Data presented in Figure 4 needs to be interpreted with caution because the two main mutations account for approximately 90% of all mutations, and furthermore, there are cases of individuals with dual mutations. In these cases, the response to one of the two mutations may be masked by that of the other, leading to some inconsistencies in the response data presented herein.

## DISCUSSION

Many aspects of the origin, incidence, response to TKI in the presence of such mutations, and associated molecular mechanisms of clinical response have been addressed by numerous reviews.[23–25] In this analytical database, we address issues pertaining to the incidence and clinical application of somatic EGFR mutations in NSCLC by providing a global picture, for use by both diagnostic laboratories and clinicians, in predicting the response to a TKI in the presence of distinct somatic EGFR mutations. The database also serves as a source of clinico-pathologic correlations and for the identification of differences according to ethnicity. It is believed that this database will allow for the generation of a more detailed appraisal of the somatic mutational spectrum of EGFR.

Gu et al.[26] have presented an updated version of a similar somatic EGFR mutation database; however, they have limited its scope to mutation carriers and have not collected information on wild-type cases. The mutational spectrum itself is also more readily presented herein, combined with information linking the mutational spectrum of TKI-treated individuals to response. Although cumulative data regarding response per specific mutation is somewhat misleading, it is hoped that this data will assist in treatment-based decisions that are more restricted when one relies on single reports.

Considering that virtually 90% of all mutations actually occur as either delEx19 and/or L858R, these two mutational "hot spots" have been extensively studied. Therefore, one can expect a degree of reporting bias concerning these two mutational types compared with all others in the mutational spectrum especially when we consider that many groups have chosen to selectively focus on the incidence of these two mutations.[27–29] A subset analysis according to the completeness of the mutational screen used per study will assist in more clearly identifying the actual mutational frequency of each given mutation. Nevertheless, because the majority of mutations are localized to two distinct regions of the gene, it is rather more likely that subsequent analyses will concentrate on these two regions, at the expense of all other mutations. Therefore, there is a clear need for an analytical database to include all data, especially when it is representative of rare mutations.

Several clinico-pathologic parameters have been shown to correlate with the presence of somatic mutations of the TK domain of EGFR as indicated in early studies, theses including gender, histology, ethnicity, and smoking history. These have been reproducibly correlated with the presence of somatic missense mutations in EGFR and also with better response(s) to TKIs.[2,3,13,14] As we do not have independent patient data, we are unable to effectively correlate any of the common clinico-pathologic factors with response. Rather, we have the opportunity to correlate these factors with mutational status.

In reviewing the existing data on the correlation of clinico-pathologic parameters with mutational spectrum in patients treated or naive to TKIs, it has become apparent that although there is sufficient information to perform retrospective (pooled) analyses, one of the major limitations includes the lack of adequate specimens, especially from some of the largest studies yet conducted in NSCLC. Recently, each of the five largest studies of TKIs have reported molecular analyses; TRIBUTE (carboplatin + paclitaxel with erlotinib or placebo, 1079 patients) reported on only 21% of enrolled patients[11,21]; INTACT-1 (cisplatin + gemcitabine with gefitinib or placebo, 1093 patients),[9,30] INTACT-2 (carboplatin + paclitaxel with gefitinib or placebo, 1037 patients)[10] and BR.21 (erlotinib or placebo, 731 patients)[4,31] each reported on only 28% of enrolled patients; and ISEL (gefitinib versus placebo, 1692 patients[32]) reported on less than 23% of enrolled patients for any biomarker tested.[33] Correlative studies from the TALENT trial have yet to be published.[12] Obviously, the availability of biological material from all of these studies would have greatly assisted in further unlocking the significance of EGFR and other potential molecular markers. Here, we tabulate the currently existing data concerning the four main clinico-pathologic factors (gender, ethnicity, smoking history, and histology). Based on our database, we believe that the tabulated data weighs in favor of supporting a correlation between mutation and clinico-pathologic parameters (a more comprehensive analysis is forthcoming).

In the reported series, for which all relevant data was extractable, the response rate for mutation carriers is significantly superior to that of wild type individuals, regardless of ethnicity (data not shown). Because of the relative low number of patients per study and also the overall limited experience with TKIs in populations for which mutational data is available, the most appropriate way to quantitatively determine the benefit obtained for TKI-treated mutation carriers versus wild type patients will be through a meta-analysis (preferably with independent patient data). There seem to be no large studies that could allow the formulation of a solid opinion on the clinical benefit offered by TKIs in mutation carriers, principally because of the low level of sample recruitment in large phase III studies. Nevertheless, the database leads us to speculate that the generally held view that TKIs do offer significant improvements in response to patients with EGFR mutant tumors compared to those with wild-type tumors.

Only two studies have addressed the association between EGFR mutation and survival, one in an independent patient data pooled analysis of 506 patients from mainland China (only 57 patients were eligible for survival), and the other in a pooled analysis of five prospective studies from Japan that reported on patients treated with gefitinib after having been selected on the basis of carrying EGFR muta-

 

tions (101 patients).[34,35] Although the response rate of exon 19 versus exon 21 (delExon19 versus L858R) may be different as ascertained from our database, they reported a response rate (CR/PR) for exon 19 deletion and L858R of 80.3% and 81.8%, respectively. It is unlikely that the data obtained from our database will change significantly over time, therefore, the pooled data from these 101 patients needs to be interpreted in light of the depth of the current data set.

Through initiation of this database we have (1) generated a user-friendly interface providing combined/cumulative data on mutational spectrum and incidence per AA of somatic mutations in EGFR; (2) tabulated data per study with respect to mutational rate for all informative clinico-pathologic fields (gender, histology, smoking status), and other relevant information including sample source, method(s) of analysis, specific exons analyzed, confirmation of somatic nature, etc.; and (3) combined cumulative data on mutational spectrum and response characteristics for TKI-treated patients per mutation and presented them in a user-friendly format facilitating access to mutation-specific response rate(s), for all such possible data. The EGFR somatic mutation database and associated subanalyses are available through a Web site (www.somaticmutations-EGFR.org) that will be updated q6 months. The generation of this database provides a first step towards a free access independent patient database from which significant improvements in NSCLC outcomes could arise. Through the development of additional collaborative efforts (e.g., www.TKI-CAN.org), we hope to see a broadening in our understanding of the origin of somatic mutations in EGFR.

We believe that the simplified representation of the data sets will be extremely useful not only for researchers in the field but more specifically for clinical teams who are utilizing the somatic mutational status of EGFR in treatment-based decisions for patients with NSCLC. We would like to invite all authors of published studies to confirm their datasets and welcome updates or new submissions to be made directly through our online site plus your comments and suggestions.

## REFERENCES


1. Pao W, Miller VA. Epidermal growth factor receptor mutations, small-molecule kinase inhibitors and non-small cell lung cancer: current knowledge and future directions. *J Clin Oncol* 2005;23:2556–2568.
2. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer (the IDEAL 1 trial) [corrected]. *J Clin Oncol* 2003;21:2237–2246.
3. Kris MG, Natale RB, Herbst RS, et al. Efficacy of gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: a randomized trial. *JAMA* 2003;290:2149–2158.
4. Shepherd FA, Rodrigues Peieira J, Ciuleanu T, et al. Erlotinib in previously treated non-small cell lung cancer. *N Engl J Med* 2005;353:123–132.
5. Cohen MH, Williams GA, Sridhara R, et al. United States Food and Drug Administration Drug Approval summary: Gefitinib (ZD1839; Iressa) tablets. *Clin Cancer Res* 2004;10:1212–1218.
6. Johnson JR, Cohen M, Sridhara R, et al. Approval summary for erlotinib for treatment of patients with locally advanced or metastatic non-small-cell lung cancer after failure of a least one prior chemotherapy regimen. *Clin Cancer Res* 2005;11:6414–6421.
7. Harari PM. Epidermal growth factor receptor inhibition strategies in oncology. *Endocr Relat Cancer* 2004;11:689–708.
8. Barker AJ, Gibson KH, Grundy W, et al. Studies leading to the identification of ZD1839 (IRESSA): an orally active, selective epidermal growth factor receptor tyrosine kinase inhibitor targeted to the treatment of cancer. *Bioorg Med Chem Lett* 2001;11:1911–1914.
9. Giaccone G, Herbst RS, Manegold C, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: a phase III trial—INTACT 1. *J Clin Oncol* 2004;22:777–784.
10. Herbst RS, Giaccone G, Schiller JH, et al. Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: a phase III trial—INTACT 2. *J Clin Oncol* 2004;22:785–794.
11. Herbst RS, Prager D, Hermann R, et al. TRIBUTE: a phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. *J Clin Oncol* 2005;23:5892–5899.
12. Gatzemeier U, Pluzanska A, Szczesna A, et al. Phase III study of erlotinib in combination with cisplatin and gemcitabine in advanced non-small cell lung cancer: the Tarceva Lung Cancer Investigation Trial. *J Clin Oncol* 2007;25:1545–1552.
13. 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;350:2129–2139.
14. Paez JG, Janne PA, Lee JC, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004;304:1497–1500.
15. 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 USA* 2004;101:13306–13311.
16. Marchetti A, Martella C, Felicioni L, et al. EGFR mutations in non-small cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *J Clin Oncol* 2005;23:857–865.
17. Huang SF, Liu HP, Li LH, et al. High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. *Clin Cancer Res* 2004;10:8195–8203.
18. Chou TY, Chiu CH, Li LH, et al. Mutation in the tyrosine kinase domain of epidermal growth factor receptor is a predictive and prognostic factor for gefitinib treatment in patients with non-small cell lung cancer. *Clin Cancer Res* 2005;11:3750–3757.
19. Shih JY, Gow CH, Yu CJ, et al. Epidermal growth factor receptor mutations in needle biopsy/aspiration samples predict response to gefitinib therapy and survival of patients with advanced non-small cell lung cancer. *Int J Cancer* 2006;118:963–969.
20. Han SW, Kim TY, Hwang PG, et al. Predictive and prognostic impact of epidermal growth factor receptor mutations in non-small-cell lung cancer patients treated with gefitinib. *J Clin Oncol* 2005;23:2493–2501.
21. Eberhard DA, Johnson BE, Amler LC, et al. Mutations in the epidermal growth factor receptor and KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. *J Clin Oncol* 2005;23:5900–5909.
22. Takano T, Ohe Y, Sakamoto H, et al. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. *J Clin Oncol* 2005;23:6829–6837.
23. Riely GJ, Politi KA, Miller VA, Pao W. Update on epidermal growth factor receptor mutations in non-small cell lung cancer. *Clin Cancer Res* 2006;12:7232–7241.
24. Shigematsu H, Gazdar AF. Somatic mutations of epidermal growth factor receptor signaling pathway in lung cancers. *Int J Cancer* 2006;118:257–262.
25. Chan SK, Gullick WT, Hill ME. Mutations of the epidermal growth factor receptor in non-small cell lung cancer—search and destroy. *Eur J Cancer* 2006;42:17–23.
26. Gu D, Scaringe WA, Li K, et al. Database of somatic mutations in EGFR with analysis revealing indel hotspots but no smoking-associated signature. *Hum Mutat* 2007;28:760–770.
27. Asano H, Toyooka S, Tokumo M, et al. Detection of EGFR gene mutation in lung cancer by mutant-enriched polymerase chain reaction. *Clin Cancer Res* 2006;12:43–48.

28. Nagai Y, Miyazawa H, Huqun, et al. Genetic heterogeneity of the epidermal growth factor receptor in non-small cell lung cancer cell lines revealed by a rapid and sensitive detection system, the peptide nucleic acid-locked nucleic acid PCR clamp. *Cancer Res* 2005;65:7276–7282.
29. Pan Q, Pao W, Ladanyi M. Rapid polymerase chain reaction-based detection of epidermal growth factor receptor gene mutations in lung adenocarcinoma. *J Mol Diagn* 2005;7:396–403.
30. Bell DW, Lynch TJ, Haserlat SM, et al. Epidermal growth factor receptor mutations and gene amplification in non-small-cell lung cancer: molecular analysis of the IDEAL/INTACT gefitinib trials. *J Clin Oncol* 2005;23:8081–8092.
31. Marchetti A, Felicioni L, Buttitta F. Assessing EGFR mutations. *N Engl J Med* 2006;354:526–528.
32. Thatcher N, Chang A, Parikh P, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomized, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). *Lancet* 2005;366: 1527–1537.
33. Hirsch FR, Varella-Garcia M, Bunn PA, et al. Molecular predictors of outcome with gefitinib in a phase III placebo-controlled study in advanced non-small-cell lung cancer. *J Clin Oncol* 2006;24:5034–5042.
34. Wu YL, Zhong WZ, Li LY, et al. Epidermal growth factor receptor mutations and their correlation with gefitinib therapy in patients with non-small cell lung cancer: a meta-analysis based on updated individual patient data from six medical centers in mainland China. *J Thorac Oncol* 2007;2:430–439.
35. Costa DB, Kobayashi S, Tenen DG, Huberman MS. Pooled analysis of the prospective trials of gefitinib monotherapy for EGFR-mutant non-small cell lung cancers. *Lung Cancer* 2007;58:95–103.

 