EGFR molecular testing in African-American non-small cell lung cancer patients - a review of discrepant data

Bradley T. Clifford[1], Pingfu Fu[1], Nathan A. Pennell[2], Balazs Halmos[3], Rom S. Leidner[1]

[1]Departments of Medicine and Biostatistics, University Hospitals - Case Medical Center, Case Western Reserve University, Cleveland, OH, USA; [2]Cleveland Clinic, Cleveland, OH, USA; [3]Department of Medicine, New York Presbyterian Hospital - Columbia University Medical Center, Columbia College of Physicians and Surgeons, USA

Corresponding to: Bradley T. Clifford, MD. Division of Hematology/Oncology, Case Western Reserve University, University Hospitals of Cleveland, 11100 Euclid Ave, WRB 3-126, Cleveland, OH 44106, USA. Email: bradley.clifford@uhhospitals.org.

Abstract: Substantial discrepancy in the literature has recently emerged regarding epidermal growth factor receptor (EGFR) mutational frequency in African American (AA) NSCLC. The first wave of tissue profiling studies, including by our group in 2009, consistently observed a significantly lower frequency of EGFR mutation in AA *vs.* White NSCLC, whereas three recent reports appear to directly contradict these findings. Reasons for this discrepancy are unclear, but one plausible explanation arises from Simpson's paradox, the consequence of aggregating heterogeneous study cohorts (in this case, the proportion of never-smokers in the study cohort). Our review of all prior studies (combined total 386 AA NSCLC cases) underscores the wide variation in the proportion of AA never-smokers among various studies (13-57%), calling inter-study comparisons into question. In parallel, we assessed objective response by RECIST to EGFR targeted therapy for AA NSCLC in the community setting, prior to the advent of routine EGFR testing. We observed a trend toward reduced response for community-based treatment of unselected AA NSCLC (5%; 3/57) as compared to overall response rates of 10% reported by large North American trials of primarily White NSCLC patients, but this was not significant (P=0.223).

Key Words: African-American; epidermal growth factor receptor (EGFR); non-small cell lung cancer; Simpson's paradox

Substantial discrepancy has recently emerged regarding the true frequency of activating epidermal growth factor receptor (EGFR) mutations in African American (AA) NSCLC subpopulations. Three recent reports appear to contradict earlier, and generally consistent, findings of a low mutation frequency in AA patients. The first round of sequencing studies by our group and others (2005 to 2009, *Table 1*) indicated that the frequency of activating EGFR mutations was significantly lower in African Americans (AA), ranging from 2-3% (4 of 160 cases combined), compared to a White cohort (1-5). However, in 2011 two large studies (*Table 1*: Reinersman, Cote) observed a much higher frequency of activating EGFR mutations in AA NSCLC cohorts, ranging

from 12-19% (31 of 188 cases combined), without significant differences compared to White NSCLC cohorts. A third 2011 study (*Table 1*: Harada) reported an even higher EGFR mutation frequency of 31% for AA NSCLC, though this is based on a limited cohort of only sixteen cases (6-8).

What could account for these discordant results? One explanation may lie in the heterogenous study cohorts. While not all prior studies have provided data on smoking status - and this is somewhat surprising given the disease at hand - comparison of the studies which do report this information, demonstrates that the proportion of never-smokers in the AA NSCLC cohorts varies widely from 13% to 57% (see *Table 1*). We would argue that the lower end of

Table 1 Summary of literature on EGFR in AA NSCLC

Study	Yang [2005] (1)	Riely [2006] (2)	Tsao [2006] (3)	Krishnaswamy [2009] (4)	Leidner [2009] (5)	Harada [2011] (6)	Reinersman [2011] (7)	Cote [2011] (8)
Sequencing Method	Sanger	Sanger (del19 and L858R only)	Sanger	Sanger	Sanger	Sanger	Sanger (del19 and L858R only)	Sequenom Mass Spec.
P-value (AA *vs.* White)	0.03	-	-	-	0.02	-	0.11	0.53
African-American (*n*)	41	14	8	66	53	16	121	67
Mutation frequency	3% (1/41)	43% (6/14)	0% (0/8)	3% (2/66)	2% (1/53)	31% (5/16)	19% (23/121)	12% (8/67)
95% CI	0-13%	n/a	0-40%	n/a	0-11%	n/a	13-27%	n/a
% Never smokers	n/a	57% (8/14)	n/a	n/a	13% (7/53)	50% (8/16)	n/a	16% (11/67)
Mutations in never smokers	0%	n/a	0%	0%	0%	100% (5/5)	n/a	88% (7/8)
EGFR mutation [*n*]	del 19 [1]	del 19 (NR) L858R (NR)	-	del 19 [1] S768I [1]	S768N [1]	del 19 [2] L858R [1] N771GY [1] A767-V769 dup [1]	del 19 [18] L858R [5]	del 19 [8]
White (*n*)	177	259	139	76	89	-	476	77
Mutation frequency	14% (25/177)	21% (55/259)	7% (10/139)	3% (2/76)	17% (15/89)	-	13% (61/476)	16% (12/77)
95% CI	9-20%	n/a	4-13%	n/a	11-26%	-	10-16%	n/a
% Never Smokers	n/a	n/a	n/a	n/a	15% (13/89)	-	n/a	27% (21/77)
Mutations in never smokers	48% (12/25)	n/a	n/a	0%	40% (6/15)	-	n/a	66% (8/12)
EGFR mutation [*n*]	del 19 [10] L858R [11] G778F [2] D770 del [2] G719R [1] H772L/ V773M [1]	del 19 (NR) L858R (NR)	n/a	P858L [1] P733T [1]	del 19 [7] L858R [8]	-	n/a	del 19 [7] L858R [3] E709A [1] G719S [1]

this range (13% Leidner) is more informative to community practice, and is more congruent with previously reported smoking rates in a large AA NSCLC cohort (7% in a series of 1,288 patients) (9); than the high proportion of never-smokers in recent studies indicating a higher frequency of EGFR mutation for AA NSCLC (57% Cote, 50% Harada). This skew may be due to archival specimen sourcing from referral-based tertiary/quaternary centers, and also points to the likely explanation for discrepancy - Simpson's paradox - a statistical phenomenon in which a correlation observed between heterogeneous groups in the aggregate, is reversed when groups are disaggregated (10-12).

As an illustration of Simpson's paradox, we can disaggregate smoking status and EGFR mutation findings based on results from the only two studies which published sufficient data to allow for this analysis (see *Table 2*, Leidner, Cote). Notably, these two studies had similar sized cohorts of AA NSCLC cases, but arrived at divergent conclusions regarding EGFR mutation frequency. *Table 2* presents *disaggregated* results according to ever/never smoking status for each study. In both studies, the ever-smoker group represents the large majority of the AA NSCLC cohort, (87% Leidner 46/53 and 84% Cote 56/67), as would be expected. Restricting analysis to AA ever-smokers, demonstrates good agreement between studies at 2% EGFR mutation frequency (1/46 Leidner and 1/56 Cote).

Table 2 Disaggregation of ever vs. never smokers to illustrate Simpson's paradox

Study	Ever smokers		Never smokers		Combined	
	Leidner [2009]	Cote [2011]	Leidner [2009]	Cote [2011]	Leidner [2009]	Cote [2011]
Mutation% AA NSCLC	2% (1/46)	2% (1/56)	0% (0/7)	64% (7/11)	2% (1/53)	12% (8/67)

In contrast, when the small AA never-smoker groups are compared, the real driver of divergence comes sharply into view, with a 64% discrepancy in EGFR mutation frequency between studies (0/7 Leidner 0% and 7/11 Cote 64%). Strikingly, 7 of the 8 total EGFR mutations identified by Cote *et al.*, were detected in the very limited AA never-smoker subgroup (n=11).

The discrepancy between the two studies above, which disappears for ever-smokers when data are disaggregated, illustrates of Simpson's paradox - a statistical phenomenon with implications both in the theoretical context and in practical application. In the famous Berkeley court case, alleged bias favoring males over females was claimed, based on analysis of admissions data *in the aggregate*. The alleged bias was subsequently shown to be driven by consistent and disproportionately higher rates of female applicants to more competitive departments, which came into clear view only after disaggregated analysis at the departmental level (13). This is also the likely explanation for apparent discrepancy between EGFR tissue profiling studies in AA NSCLC. In this case, disaggregation of the data according to smoking status (for the two studies where this is possible) demonstrates that the apparent divergence is driven entirely by results from very small never-smoker subgroups.

Additional discrepancies between the various studies which should be mentioned, include differences in the specific EGFR mutations being reported and the varying sensitivity of detection for the sequencing methods employed. Two activating EGFR mutations are routinely tested in clinical practice: a short oligonucleotide deletion in exon 19 (del 19) and a non-synonymous point mutation in exon 21 leading amino acid substitution at residue 858 (L858R). Together, these two mutations (del 19 and L858R) account for up to 90% of identified EGFR mutations. Several, much rarer mutations have been reported, including prior reports in the AA NSCLC tissue profiling literature (*Table 1*). Because these rare mutations are not routinely tested in clinical practice, the actual sensitivity they may, or may not, confer to EGFR targeted therapy is not known. In fact, there is evidence to suggest that some EGFR mutations may actually confer resistance. In the recent study by Harada *et al.*, five EGFR mutations were

observed in a cohort of 16 AA NSCLC cases, and notably, all were observed in AA never-smokers (6). Two of these five mutations represented rare insertions in exon 20 (N771GY and 767A-769V dup). Subsequent laboratory modeling using YFP-tagged MCF-7 cells expressing these mutations actually showed increased resistance to erlotinib *in vitro*, which may open an intriguing line of future investigation surrounding mechanisms of proclivity to rare variant EGFR mutations in specific population groups.

A further source of potential discrepancy between studies bears mention - the use of different sequencing technologies, which could influence the scope and threshold of mutations being detected. While most prior studies relied on standard Sanger sequencing for mutational profiling, the study by Cote *et al.* used a higher sensitivity platform (Sequenom mass spectrometry) which can detect mutation in as few as 5-10% of tumor cells (14). Whether response to EGFR TKI in a tumor consisting of >90% wild-type EGFR cells is clinically meaningful, remains to be determined, but may ultimately reveal that higher sensitivity is not a sine qua non of clinical benefit.

A final consideration must be given to patient self-reporting for determination of race which may lead to selection bias. Objective measure of genetic admixture is now theoretically possible, for example using ancestry SNP genotyping panels (15). The complex interaction of race and genetics in somatic tumorigenesis is not mechanistically well characterized or clinically interpretable at the present time. However, as hinted at by clustering of rare variant EGFR mutations in NSCLC, this may be an area ripe for future investigations as genomic advances proceed apace.

In the end, tissue-based EGFR mutational analysis is only a surrogate for what is actually of primary clinical interest: a reasonable prediction of treatment efficacy. In order to further assess the treatment effects of EGFR TKI's, we reviewed treatment response in an unselected AA NSCLC patient population treated in the community setting (Cleveland, OH). If indeed activating EGFR mutations are significantly rarer among AA NSCLC patients, a significantly lower rate of objective response would be expected in comparison to a White, North American counterpart where previous objective response

Table 3 Patient characteristics

Age, years	
Mean	64
Range	45-83
Gender, no of patients	
Female	37 (65%)
Male	20 (35%)
Histology	
Adenocarcinoma	65% (37/57)
Squamous	12% (7/57)
Large Cell	2% (1/57)
Undifferentiated	12% (7/57)
Unknown	9% (5/57)
Stage at time of EGFR TKI therapy	
IIIB	3% (2/57)
IV	97% (55/57) *
Smoking status	
Never smoker	11% (6/57)
Ever smoker (current or previous)	89% (51/57)
EGFR TKI therapy	
Erlotinib	93% (53/57)
Gefitinib	7% (4/57)
First line EGFR TKI therapy	37% (21/57)
*21% with brain metastases (12/57)	

rates of roughly 10% have been observed (12.3% Perez-Soler *et al.* 2004 and 8.9% Shepherd *et al.* 2005) (16,17). Self-identified AA patients with advanced NSCLC treated with empiric EGFR TKI (either erlotinib or gefitinib) prior to widespread mutation testing, were evaluated for radiographic response by RECIST criteria (18). Patient characteristics are summarized in *Table 3*.

We observed a 5% rate of response to EGFR TKI among unselected AA NSCLC patients treated in the community setting prior to the advent of routine EGFR mutation screening, using objective RECIST criteria and chart review of 57 cases. While this result did not reach significance (P=0.223), when compared against a 10% response rate as reported in large North American trials of unselected primarily White NSCLC patients (16,17), it represents a trend toward reduced rate of response. This trend is more in line with the results of early tissue profiling studies, including our own, which pointed to a lower frequency of activating EGFR mutations in AA *vs.* White NSCLC, and is consistent with current clinical practice of limiting EGFR

TKI to patients with activating EGFR mutations by tissue analysis, regardless of race. As EGFR mutations are more frequent in never-smokers, a true evaluation of mutational frequency by race should be studied in a never-smoker cohort to help clarify the relationship between race and EGFR mutational status.

Acknowledgements

References

1. Yang SH, Mechanic LE, Yang P, et al. Mutations in the tyrosine kinase domain of the epidermal growth factor receptor in non-small cell lung cancer. Clin Cancer Res 2005;11:2106-10.
2. Riely GJ, Pao W, Pham D, et al. Clinical course of patients with non-small cell lung cancer and epidermal growth factor receptor exon 19 and exon 21 mutations treated with gefitinib or erlotinib. Clin Cancer Res 2006;12:839-44.
3. Tsao AS, Tang XM, Sabloff B, et al. Clinicopathologic characteristics of the EGFR gene mutation in non-small cell lung cancer. J Thorac Oncol 2006;1:231-9.
4. Krishnaswamy S, Kanteti R, Duke-Cohan JS, et al. Ethnic differences and functional analysis of MET mutations in lung cancer. Clin Cancer Res 2009;15:5714-23.
5. Leidner RS, Fu P, Clifford B, et al. Genetic abnormalities of the EGFR pathway in African American Patients with non-small-cell lung cancer. J Clin Oncol 2009;27:5620-6.
6. Harada T, Lopez-Chavez A, Xi L, et al. Characterization of epidermal growth factor receptor mutations in non-small-cell lung cancer patients of African-American ancestry. Oncogene 2011;30:1744-52.
7. Reinersman JM, Johnson ML, Riely GJ, et al. Frequency of EGFR and KRAS mutations in lung adenocarcinomas in African Americans. J Thorac Oncol 2011;6:28-31.
8. Cote ML, Haddad R, Edwards DJ, et al. Frequency and type of epidermal growth factor receptor mutations in African Americans with non-small cell lung cancer. J Thorac Oncol 2011;6:627-30.
9. 9. Schwartz AG, Swanson GM. Lung carcinoma in African Americans and whites. A population-based study in metropolitan Detroit, Michigan. Cancer 1997;79:45-52.
10. Simpson EH. The Interpretation of Interaction in

Contingency Tables. J Roy Stat Soc B 1951;13:238-41.
11. Mittal Y. Homogeneity of subpopulations and Simpson's paradox. J Am Stat Assoc 1991;86:167-72.
12. Fu P, Panneerselvam A, Clifford B, et al. Simpson's paradox - aggregating and partitioning populations in health disparities of lung cancer patients. Stat Methods Med Res 2012. [Epub ahead of print].
13. Bickel PJ, Hammel EA, O'connell JW. Sex bias in graduate admissions: data from berkeley. Science 1975;187:398-404.
14. MacConaill LE, Campbell CD, Kehoe SM, et al. Profiling critical cancer gene mutations in clinical tumor samples. PLoS One 2009;4:e7887.
15. Via M, Ziv E, Burchard EG. Recent advances of genetic ancestry testing in biomedical research and direct to consumer testing. Clin Genet 2009;76:225-35.
16. Pérez-Soler R, Chachoua A, Hammond LA, et al. Determinants of tumor response and survival with erlotinib in patients with non--small-cell lung cancer. J Clin Oncol 2004;22:3238-47.
17. Shepherd FA, Rodrigues Pereira J, Ciuleanu T, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med 2005;353:123-32.
18. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-47.