Regional Disparities in EGFR Mutation Prevalence in Non-Small Cell Lung Cancer: A Global Comparative Analysis

Andreea Gabor[1], Sebastian Gabor[2], Constantina Bianca Vulpe[3], Emilian Onișan[4]

[1]Faculty of Animal Science and Biotechnologies, BUASVM, Timișoara, Romania · [2]Faculty of Mechanical Engineering, Politehnica University of Timișoara, Romania · [3]Faculty of Chemistry, Biology, Geography, West University of Timișoara, Romania · [4]Faculty of Agriculture, University of Life Sciences "King Mihai I" from Timișoara, Romania

1. Background

- Lung cancer ranks first in cancer mortality worldwide, causing almost a quarter of all cancer deaths. Non-small cell lung cancer (NSCLC) makes up roughly 60–85% of cases.
- Activating mutations in the EGFR kinase domain, mostly exon 19 deletions and L858R in exon 21, keep the receptor switched on without a ligand. This drives the Ras/Raf/MAPK and PI3K/Akt/mTOR pathways.
- Reported EGFR prevalence differs between regions. This matters for molecular testing and for access to targeted therapy.

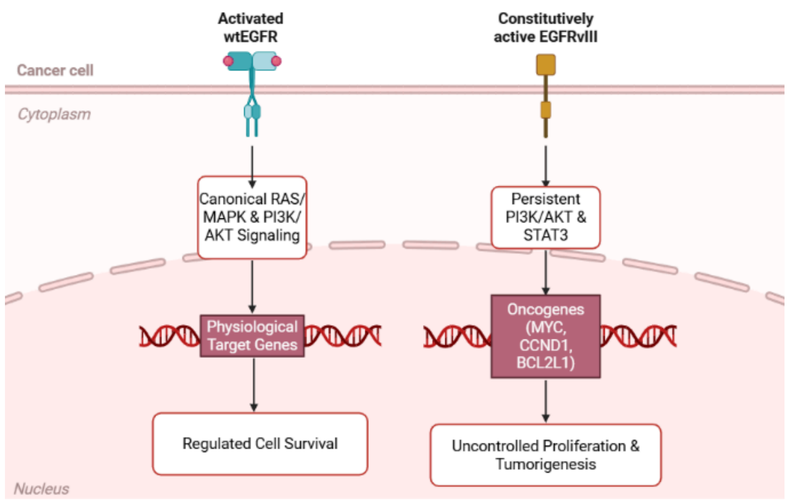


Figure 1. Wild-type EGFR signaling versus ligand-independent activation by mutant EGFR.

2. Aim

To describe how reported EGFR mutation prevalence in NSCLC differs across world regions in 2025, and how far those differences can be trusted.

3. Methods

- Descriptive analysis of 35 published values (% of NSCLC tumors with an EGFR mutation) from six regions: Europe, Asia, Africa, Oceania, Latin America and North America.
- Three datasets: data by continent (16 values), data by country (19 values), and a combined dataset with one value per source per continent, so that no source counts twice.
- 2025 papers were preferred. Nine entries from 2022, 2024 or 2026 filled gaps and were removed in a sensitivity analysis.
- Python in Google Colab. 95% confidence intervals: t-distribution for continent and country data (3 or more values), percentile bootstrap (10,000 resamples) for the combined data.
- Non-parametric tests only: Kruskal–Wallis, Mann–Whitney U (Cliff's δ) and Spearman correlation. Exploratory, α = 0.05, no adjustment for multiple comparisons.
- Global figure: unweighted mean of the six continental means, with a stratified bootstrap (5,000 resamples).

4. Results

Asia highest, Europe lowest

In the data by continent, Asia averages 50.1% (95% CI 45.9–54.3) and Europe 13.3% (10.7–15.9), a gap of about 3.8 times. The other regions lie between them.

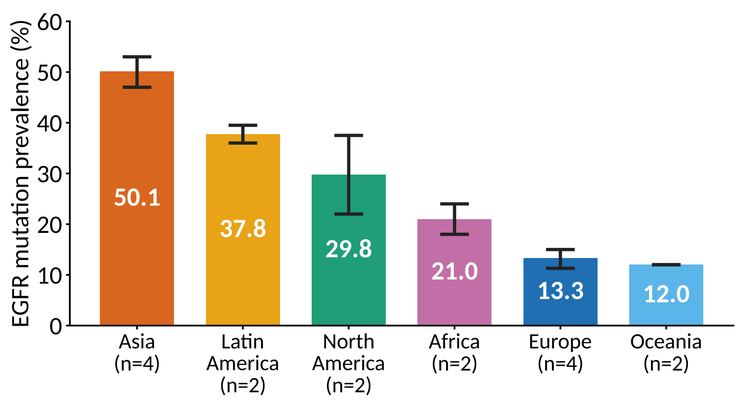


Figure 2. Mean EGFR prevalence per continent, data by continent. Bars show the mean, whiskers the minimum and maximum.

The six continents differ: Kruskal–Wallis H = 13.79, p = 0.017, ε^2 = 0.92. Asia against the rest of the world: p = 0.004 (δ = 1.00) in the continent data, p = 0.0007 (δ = 0.85) in the combined data, but p = 0.095 in the country data. Asia against Europe: p = 0.029.

A more cautious estimate

With one value per source, Asia is 41.0% (95% bootstrap CI 33.5–47.2, seven sources) and Europe 14.6% (11.8–18.4, nine sources). Latin America (28.6%), North America (25.3%), Africa (23.9%) and Oceania (16.8%) have heavily overlapping intervals and cannot be ranked reliably.

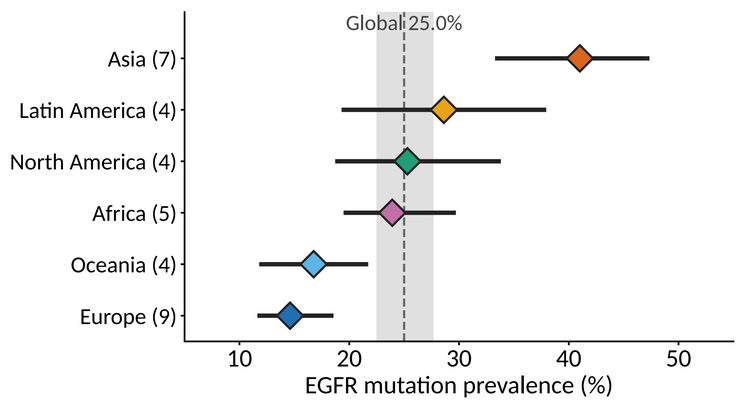


Figure 3. Combined estimate per continent (number of sources in brackets) with 95% bootstrap CI. Grey band: global estimate, 25.0% (95% CI 22.5–27.7).

Two data sources, two pictures

Continent and country estimates agree on the extremes and on little else (Spearman ρ = 0.31, p = 0.54). Asia differs by +20.8 percentage points, Latin America by +18.2 and Oceania by −9.5.

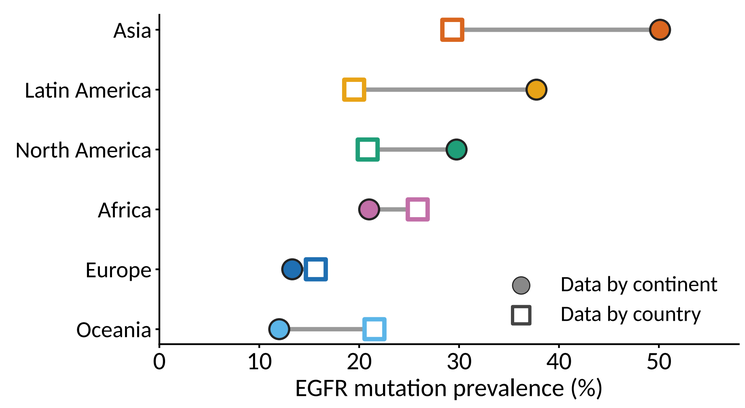


Figure 4. Mean per continent from continent values (circles) and from country values (squares).

- The countries available do not represent their continent: Latin America is Brazil and Argentina only, North America is the United States only.
- Within Asia, China (38.4%) and South Korea (41.2%) are far above Lebanon (11.3%).
- Country means are unweighted. Even two US studies differ by 7.7 points.

Robustness and coverage

Removing the nine non-2025 entries changes the continent means little (Asia 50.1 to 50.5%, Europe 13.3 to 13.2%). In the country data the shifts are larger (Asia 29.3 to 33.6%). The main conclusion survives; the fine detail does not.

Asian and non-Asian values overlap: the highest values elsewhere (39.5% in Latin America, 37.5% in North America) match the lower Asian ones. "Asia is higher" describes a shift in the typical value, not a clean split of every study.

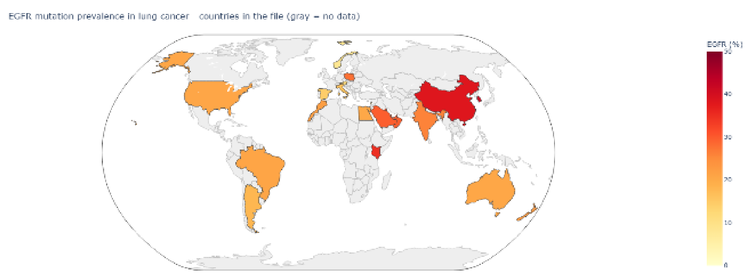


Figure 5. Country values on a world map. Grey: no value. Large parts of Africa and Latin America have no country-level data.

5. Conclusions

- Reported EGFR mutation prevalence follows a geographic gradient, not a set of exact national rates.
- Asia has the highest prevalence (roughly 41–50% depending on how the data are read) and Europe the lowest (roughly 13–16%).
- Latin America, North America, Africa and Oceania lie between them, but the evidence is too thin to rank them with confidence.
- The unweighted global figure of about 25% is a number for orientation, not a measurement of the world's prevalence.
- A single global rate hides differences that matter for molecular screening protocols and equitable access to targeted therapy.

Limitations

- Small samples: two to nine sources per region.
- Testing methods and populations differ between studies; the data cannot separate biology from method.
- Many countries have no data at all.

Keywords: EGFR mutation · Geographic prevalence · Global epidemiology · Non-small cell lung cancer

References

Du Z. et al., 2023. Biomolecules 13, 210. doi:10.3390/biom13020210
Galdadas I. et al., 2021. eLife 10, e65824. doi:10.7554/eLife.65824
Murphy C. et al., 2025. JAMA 334, 1836–1845. doi:10.1001/jama.2025.17695
Ngo L.V. et al., 2025. Oncology in Clinical Practice 21, 98–103. doi:10.5603/ocp.98161
Zhou F., Zhou C., 2018. Transl Lung Cancer Res 7, 450–463. doi:10.21037/tlcr.2018.05.14