Article

Molecular landscape of *ERBB2* alterations in 3000 advanced NSCLC patients

Check for updates

Lingzhi Hong [1,2], Sonia Patel[1], Leylah M. Drusbosky[3], Yuanyuan Xiong[4], Rongrong Chen[4], Ruixuan Geng[5], Simon Heeke [1], Monique Nilsson[1], Jia Wu [1,2], John V. Heymach [1], Yingyi Wang[6] ✉, Jianjun Zhang [1,7] ✉ & Xiuning Le [1] ✉

ERBB2 (*HER2*) represents a newly recognized actionable oncogenic driver in non-small cell lung cancer (NSCLC), with approved targeted therapy available. Understanding the landscape of *ERBB2* alterations and co-occurring mutations is essential for guiding treatment decisions. We conducted an analysis involving 3000 NSCLC patients with all types of *ERBB2* alterations, drawn from two extensive retrospective cohorts: 1281 from Geneplus (Chinese) and 1719 from Guardant360 (the United States, US). The incidence of all types of *ERBB2* alterations was found to be 5.6% in the Chinese group and 5.2% in the US group. In both cohorts, among oncogenic alterations of *ERBB2*, exon 20 insertion Y772_A775dupYVMA was the most frequent alteration (58% vs 41.6% in the Chinese vs the US), followed by G776delinsVC/LC/VV/IC (10.7% vs 9.7%), and S310X (10.5% vs 15.4%). *EGFR* ex20 insertions were identified in the A767-V774 region, whereas *ERBB2* ex20 insertions were observed in the Y772-P780 region. Notably, *EGFR* ex20 insertions exhibited greater insertion diversity. Clinical characteristics of *EGFR* and *ERBB2* ex20 NSCLC were similar, characterized by low tumor mutation burden (TMB), a predominant never-smoker population, and a majority of lung adenocarcinoma cases.

Genetic alterations in erythroblastic leukemia oncogene B (erbB) family, also known as human epidermal growth factor receptor (HERs), are important mediators of cellular proliferation, tumorigenesis, and apoptosis[1]. The HER family has four members: epidermal growth factor receptor (*EGFR*; also known as *HER1* or erbB1), *ERBB2* (erbB2 or neu), *HER3* (erbB3), *HER4* (erbB4). *EGFR* and *ERBB2* are major therapeutic targets in patients with non-small cell lung cancer (NSCLC)[2]. *ERBB2* has no ligand binding domain of its own, however, it binds to other ligand-bound EGF receptor family members to form a heterodimer, stabilizing ligand binding and enhancing kinase-mediated activation of downstream signaling pathways, such as mitogen-activated protein kinase (*MAPK*) and phosphatidylinositol-3 kinase (*PI3K*) pathways. Around 1–6% of NSCLC tumors harbor *ERBB2* mutations[3–5] leading to constitutive action of downstream proliferative pathways, and resulting in oncogenesis[6]. Similar to *EGFR* mutations, *ERBB2* mutations in NSCLC are most common in female non-smokers with a higher incidence of brain metastases[3,7,8]. Other reports showed that 6–35% of NSCLC tumors display *ERBB2* protein overexpression while between 10 and 20% of NSCLC tumors contain *ERBB2* amplification[9–14].

It has been characterized that *ERBB2* exon 20 insertions (3–12 bp) in the kinase domain between codons 775 and 881 are the most common oncogenic *ERBB2* alteration[3]. The most common *ERBB2* insertion is a YVMA duplication at the codon 775 (83% of cases)[3,8]. Some point mutations in *ERBB2* exon 20 are oncogenic as well, such as L755S and G776C[3]. Recurrent mutations in *ERBB2* transmembrane domain (TMD) and juxtamembrane domain (JMD) have been identified, including G660D, R678Q, E693K, and Q709L. These oncogenic mutations enhance *ERBB2* activity by improving the active dimer interface or stabilizing an activating conformation[15]. Exon 8 *ERBB2* mutation (S310X) is at the extracellular domain, also known as furin-like cysteine rich domain. S310F and S310Y mutations may result in hydrophobic interactions between the aromatic rings of the newly introduced 310F or 310Y with Y274 and F279 of the neighboring molecule, promoting noncovalent dimerization and kinase activation[16]. However, all previous

[1]Department of Thoracic/Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [2]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [3]Guardant Health, Redwood City, CA, USA. [4]Geneplus-Beijing, Beijing, China. [5]Department of International Medical Services, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [6]Department of Medical Oncology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. [7]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ✉e-mail: waltwyy@163.com; jzhang20@mdanderson.org; xle1@mdanderson.org

THE HORMEL INSTITUTE
UNIVERSITY OF MINNESOTA

studies had limited sample size, so the incidence of each of the *ERBB2* mutations in lung cancer are still largely unknown. Furthermore, *ERBB2* lung cancer's co-occurring mutation profile also require further investigation.

Currently, for metastatic lung cancer harboring *ERBB2* mutations, trastuzumab deruxtecan (EnHertu, T-Dxd) is the only approved targeted therapy. In the DESTINY-Lung01 and DESTINY-Lung02 trials, trastuzumab deruxtecan, a humanized anti-*ERBB2* monoclonal antibody-drug conjugate (ADC), showed an overall response rate (ORR) of 33.9%-72.5% with median PFS of 8.2 (95% CI, 6.0–11.9) months[17–19]. Small molecule tyrosine kinase inhibitors (TKIs) to target oncogenic *ERBB2* mutations have also been investigated. Studies using non-selective *ERBB2* TKIs like dacomitinib (NCT00818441), afatinib (NCT02597946), tarloxotinib (NCT03805841), and neratinib (NCT01827267) yielded ORR ranging 0–19%. More recently, mobocertinib (NCT02716116), poziotinib (NCT03318939) and pyrotinib (NCT04063462, NCT02834936)[20–23] demonstrated moderate efficacy. Currently, multiple anti-*ERBB2* agents are under active development, including TKIs such as BI1810631 (NCT04886804) and BAYER 2927088 (NCT05099172). Additionally, other *ERBB2* ADCs and monoclonal antibodies like trastuzumab with pertuzumab (NCT02925234, NCT02091141, NCT03845270) and T-DM1 (NCT02675829) have been conducted displaying clinical effect as well.

As *ERBB2* is established as a new actionable target oncogene in lung cancer, it is critical to understand the heterogeneity of its mutational landscape. Using next generation sequencing (NGS) approach from both patient blood and tissue sample, this study has identified a diversity of *ERBB2* alterations and the associated clinicopathological features from two independent retrospective cohorts from Geneplus (Chinese) and Guardant360 (the US). Furthermore, this study also examines the co-occurring mutations with *ERBB2* mutations and compared *EGFR* and *ERBB2* exon 20 insertions' distribution. This study represents a real-world dataset that incorporated patients with *ERBB2*-altered NSCLC from both China and the United States.

Results
Clinicopathologic characteristics of *ERBB2* altered NSCLC
In this study, two large retrospective cohorts of NSCLC patients were analyzed. In the Geneplus study, a total of 1281 Chinese patients diagnosed with NSCLC harboring *ERBB2* somatic mutations (including amplification) were enrolled from November 2016 to April 2022. The incidence of all types of *ERBB2* alterations in the tested NSCLC patients during the enrollment period was 5.6% (1281 out of 22,905). In the Guardant360 cohort, 1719 patients in the United States diagnosed with NSCLC harboring *ERBB2* somatic mutations were analyzed from June 2019 to October 2021, with an incidence of 5.2% (1719 out of 33,080). Guardant360 cohort tested patient samples from the United States from individuals with various ethnic background, however, this cohort did not collect racial data as a requirement. As Guardant360 is widely used in the United Stated, it is very likely that this cohort represents the US patient population. In total, 3000 *ERBB2*-altered NSCLC patients were identified and analyzed.

Among the 1281 patients with *ERBB2* alterations in the Geneplus cohort, 55% were female with a median age of 58 (18-89) years. The majority of patients with *ERBB2* alterations were never smokers (70.9%) and adenocarcinoma (94.9%) histology. Clinicopathologic and demographic features of patients with *ERBB2* somatic mutations including *ERBB2* amplifications are included in Table 1. Among the 1,719 patients with *ERBB2* alterations in the Guardant360 cohort, 51.8% were female with a median age of 70 (22–100) years. The majority of patients were adenocarcinoma (94.7%).

Mutation landscape of *ERBB2* alterations in NSCLC
We characterized the mutational landscape of *ERBB2* in NSCLC patients from these two cohorts. In Geneplus, *ERBB2* mutations or insertions were identified in 930 patients, *ERBB2* amplification(amp)-only in 351 patients, and concurrent mutation/amplification at 5.5% (70 out of 1281). Next, we annotated the mutation/insertions for their oncogenic potential, using OncoKB and COSMIC database ("Methods") and found that 920 mutation/insertions were oncogenic (Fig. 1a). Therefore, the incidence of *ERBB2*

Table 1 | Clinical features of patients with *ERBB2* somatic mutations including *ERBB2* amplifications in Geneplus cohort

Characteristics	*n* = 1281
Age, years	
Mean (SD)	56.8 (13.2)
Median (min-max)	58 (18-89)
Gender, *n* (%)	
Male	577 (45.0)
Female	704 (55.0)
Smoking, *n* (%)	
Never	479 (70.9)
Ever	197 (29.1)
Unknown	605 (100)
Tumor stage, *n* (%)	
I	68 (19.0)
II	25 (7.0)
III	42 (11.7)
IV	223 (62.3)
Unknown	923 (100)
Histology, *n* (%)	
Adenocarcinoma	1068 (94.9)
Squamous cell carcinoma	46 (4.1)
Adenosquamous carcinoma	8 (0.7)
Sarcomatoid carcinoma	2 (0.2)
Neuroendocrine neoplasms	1 (0.1)
Unknown	156 (100)
Sample type, *n* (%)	
Tissue	1155 (90.1)
Blood	97 (7.6)
Pleura effusion	24 (1.9)
Others	5 (0.4)
Panel, *n* (%)	
1021	663 (51.7)
73	371 (29.0)
36	132 (10.3)
Others	115 (9.0)

alterations was 5.6% (1,281/22,905) and *ERBB2* oncogenic mutations or insertions 4.0% (920/22,905), respectively.

Among the oncogenic mutations, the most common alterations were in the tyrosine kinase domain exon 20 (80.1% [737/920]), with Y772_A775dupYVMA being the most frequent (58% [534/920]), followed by G776delinsVC/LC/VV/IC (10.7% [98/920]), and S310F/Y (10.5% [97/920]) (Fig. 1a, c). In the 70 patients with co-occurring *ERBB2* amplification, oncogenic *ERBB2* alterations were identified in 61 patients (Fig. 1e). Among them, Y772_A775dupYVMA was the most common mutation (50.8% [31/61]) (Fig. 1f). 2.5% (32/1,281) of patients had *ERBB2* mutations which were reported as neutral by OncoKB or COSMIC or not reported (Supplemental Fig. 1a).

In the Guardant360 cohort, 634 mutation/insertions were analzyed as oncogenic by OncoKB and/or COSMIC (Fig. 1b). Therefore, the incidence of *ERBB2* alterations was 5.2% (1,719 / 33,080) and *ERBB2* oncogenic mutations or insertions 1.9% (634 / 33,080), respectively.

Among those, a similar pattern was observed with Y772_A775dupYVMA being the most frequent (41.6% [264/634]) (Fig. 1b, d), followed by G776delinsVC/LC/VV (9.7% [61/634]), and S310F/Y (15.4% [98/634]).
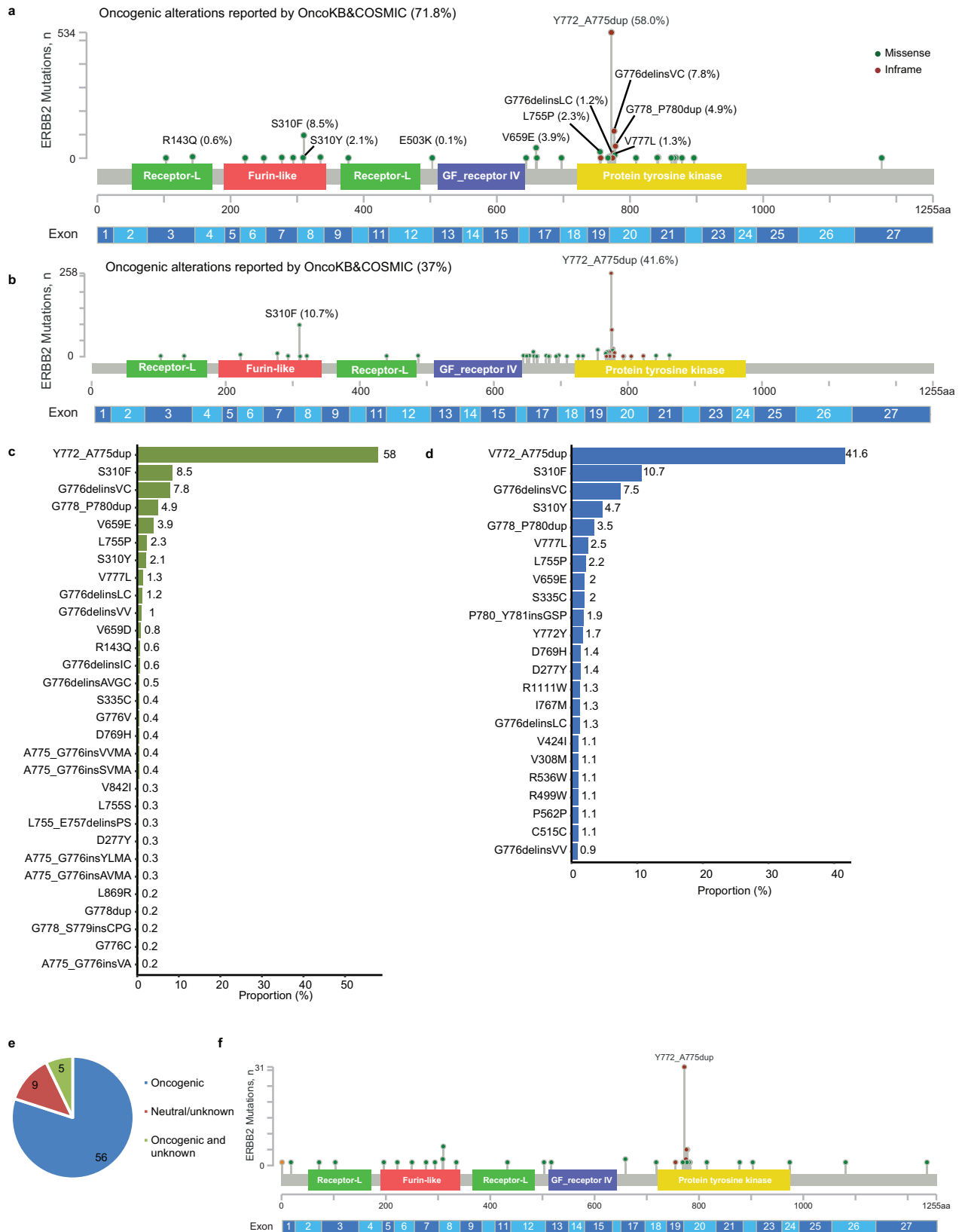
Fig. 1 | *ERBB2* mutational profile in newly diagnosed NSCLC patients. *ERBB2* oncogenic mutations and proportion reported by OncoKB and COSMIC in Geneplus (**a**, **c**) and Guardant360 (**b**, **d**). **e** Proportion of *ERBB2* oncogenic mutations for 70 patients with co-occurring *ERBB2* amplification in Geneplus cohort. **f** *ERBB2* mutational profile of the 70 patients with co-occurring *ERBB2* amplification in Geneplus cohort.
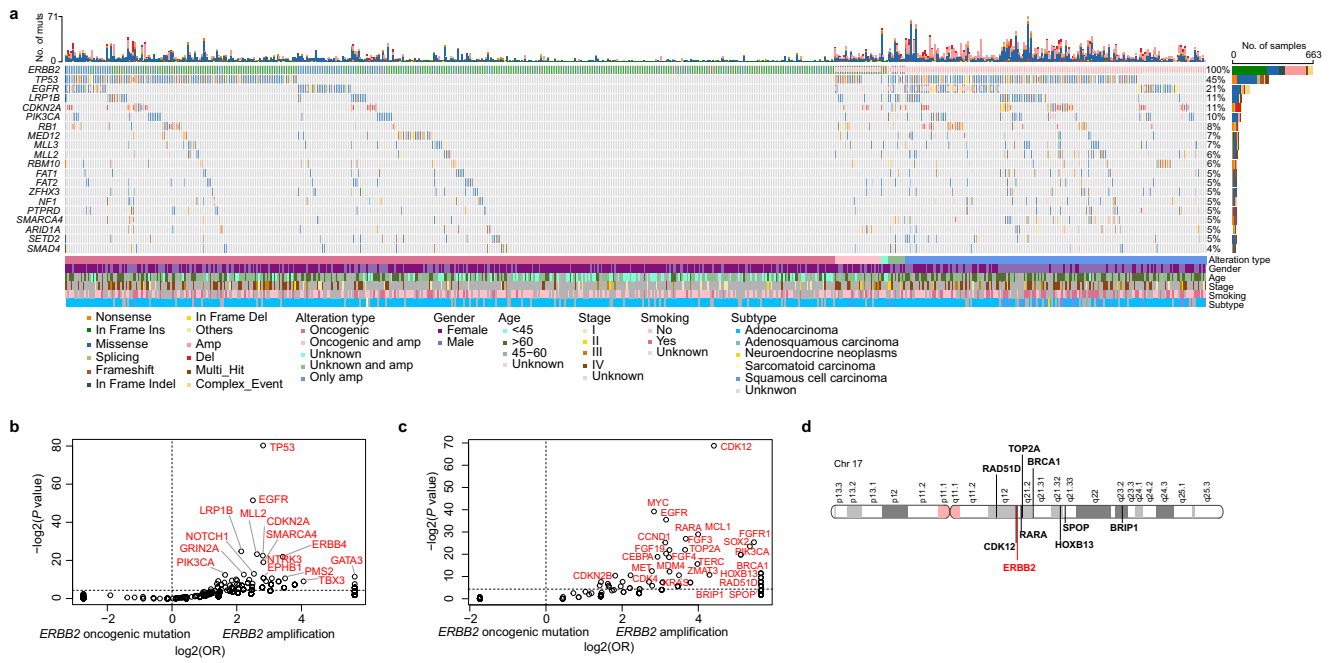
Fig. 2 | Differential co-mutated genes between patients with ERBB2 oncogenic mutations and with ERBB2 amplifications in Geneplus cohort. a Co-mutation plot showing co-occurring alterations between *ERBB2* oncogenic mutations and amplifications. Enrichment analysis of co-altered genes between patients with *ERBB2* oncogenic mutations and amplifications: (**b**) gene SNVs/Indels; (**c**) gene CNVs. **d** Map of chromosome 17 (Chr17) showing the location of *ERBB2* and co-amplified genes on the same chromosome.

48% (825 out of 1,719) of alterations had no known annotations in those two databases (Supplemental Fig. 1b) including 806 unique variants.

Among the 806 unique variants of unknown significance (VUS), 154 (19.1%) were detected in more than one patient and 50 (6.2%) in more than 3 patients. Using clonality analysis ("Methods"), the clonality of 309 (38.3%) variants were above 0.5. Overall, there were 29 VUSs having repeated occurrence (in more than one patient) and functioning as the predicted dominant clone. S335C, G222C, and D277Y, which were all located at the Furin-like region, were the top three frequent protein changes with dominant clonality (Supplemental Fig. 1c). Given this stringent filtering criteria, these 29 alterations may be potential oncogenic driver mutations and warrant further investigation (Supplemental Fig. 1c).

Co-occurring alterations with *ERBB2* in NSCLC
Next, we evaluated other co-occurring genetic alterations with *ERBB2* alterations in NSCLC. In Geneplus cohort, 663 patients' samples were analyzed using the 1021 gene panel (Table 1) from their tissue (91.4%), blood (6.8%), or pleural effusion (1.5%) samples. 451 (68.0%) samples had *ERBB2* mutations, 175 (26.4%) with *ERBB2* amplification only and 37 (5.6%) with *ERBB2* mutations and amplifications Fig. 2a. Clinicopathologic and demographic features of these patients are described in Supplementary Table 1. The median age at diagnosis was 57 years (range, 21-89), with 56.3% females and 70.3% non-smoker. Among patients with known histology types, most had adenocarcinoma histology (94.3%).

When the co-mutations were compared between *ERBB2* mutations to *ERBB2* amplification, *TP53*, *EGFR*, and *CDKN2A* were significantly more frequently co-occurring with amplifications in both the Geneplus (Fig. 2b) and Guardant360 cohorts (Supplemental Fig. 2a, b). Copy number gain in the genes of *CDK12*, *FGFR1*, and *EGFR* was significantly more frequently observed in *ERBB2* amplification compared to those who have *ERBB2* mutation (Fig. 2c). Given the proximal location of *CDK12* and *ERBB2* located on Chr17q12, it is not surprising that *ERBB2* amplified samples also contained an amplification of *CDK12* (Fig. 2d). This observation has also been reported for *ERBB2* amplified in other solid cancer types, including breast cancer, gastric cancer, biliary tract cancer, and colorectal cancer[24].

As *ERBB2* mutation versus amplification displayed distinct co-mutation landscape, we also compare the clinical features between the two groups (Table 2). Patients with *ERBB2* mutations are more significantly associated with being younger (mean age: 53.6 vs 63 years; $p < 0.001$), more likely female patients (66% vs 31%; $p < 0.001$) and being a never smoker (79% vs. 50%; $p < 0.001$). The majority of the *ERBB2* mutation tumors were adenocarcinomas (98.6%), compared to a higher proportion of *ERBB2* amplified tumors have a squamous cell carcinoma histology (15.4% vs 0.7%; $p < 0.001$). *ERBB2* mutation tumors has a low TMB (mean value: 3.5 vs 8.9; $p < 0.001$).

Clonality relationship between *ERBB2* and *EGFR* when co-occurring in NSCLC
Oncogene variant clonality can be deduced from VAF to infer dominant versus non-dominant clonal relationship. We annotated the relative clonality inferred by VAF for oncogenic alterations of *ERBB2* and *EGFR* to dissect the relationship between these two driver genes.

Among samples with *ERBB2* oncogenic alterations ($n = 474$), co-occurring *EGFR* oncogenic mutations were observed in 48 cases (10.1%), and the dominant clonality was identified as the following: *ERBB2* ($n = 7$), *EGFR* ($n = 16$), both ($n = 20$), and other gene alterations ($n = 5$) (Supplementary Fig. 3a). In the 27 cases with *ERBB2* dominant with or without *EGFR* dominant (22 S310F; 3 Y772_A775dup; 1 V659D; 1 S310Y), 18 had *EGFR* L858R and 7 had exon 19 deletion. In the 16 cases with *EGFR* only dominant, 11 cases had *ERBB2* S310X mutations, and 1 case had Y772_A775dup (Supplementary Fig. 3b–d). These results suggest that *EGFR* with specific documented dominant clones such as L858R and exon 19 deletion held the most dominant clonality in patients with concomitant *EGFR* and *ERBB2* oncogenic mutations.

Difference Between *EGFR* and *ERBB2* exon 20 insertions in NSCLC
As *EGFR* and *ERBB2* are homologous in the *ERBB* family and exon 20 insertion is one common mechanism of activation for both genes, we then made direct comparisons of the exon 20 insertion mutations for those two groups. In the Geneplus cohort, a comparison of clinical characteristics and molecular features between samples with *ERBB2* exon 20 insertions

Table 2 | Clinical features of patients with *ERBB2* somatic mutations or *ERBB2* amplification only in Geneplus cohort

Characteristics	*ERBB2* oncogenic mutation[a] (*n* = 474)	*ERBB2* amplification only (n = 175)	*P* value
Age, years			<0.01
Mean (SD)	53.6 (12.8)	63.0 (9.5)	
Median (min-max)	54 (21-83)	63 (37-89)	
Gender, *n* (%)			<0.001
Male	163 (34.4)	120 (68.6)	
Female	311 (65.6)	55 (31.4)	
Smoking, *n* (%)			<0.001
Never	222 (78.7)	58 (49.6)	
Ever	60 (21.3)	59 (50.4)	
Unknown	192 (100)	58 (100)	
Tumor stage, *n* (%)			0.04[b]
I	34 (23.8)	5 (6.8)	
II	6 (4.2)	5 (6.8)	
III	19 (13.3)	9 (12.4)	
IV	84 (58.7)	54 (74.0)	
Unknown	331 (100)	102	
Histology, *n* (%)			<0.001[c]
Adenocarcinoma	412 (98.6)	123 (82.6)	
Squamous cell carcinoma	3 (0.7)	23 (15.4)	
Adenosquamous carcinoma	2 (0.5)	2 (1.3)	
Sarcomatoid carcinoma	1 (0.2)	0	
Neuroendocrine neoplasms	0	1 (0.7)	
Unknown	56 (100)	26	
Germline mutation, *n* (%)			0.8
Yes	14 (3.0)	6 (3.4)	
No	460 (97.0)	169 (96.6)	
PD-L1 test, *n* (%)			0.8
TPS < 1%	54 (63.5)	34 (58.6)	
TPS,1%-50%	26 (30.6)	19 (32.8)	
TPS ≥ 50%	5 (5.9)	5 (8.6)	
Unknown	389 (100)	117 (100)	
TMB			<0.001
Mean (SD)	3.5 (4.6)	8.9 (9.2)	
Median (min-max)	1.9 (0-28.8)	5.8 (0-68.0)	

[a]*ERBB2* oncogenic mutation includes *ERBB2* oncogenic mutation accompanied with/without *ERBB2* amplification.
[b]Comparison between I-III and IV.
[c]Comparison between adenocarcinoma and non-adenocarcinoma.

Table 3 | Clinical features of patients with *ERBB2* exon 20 (ex20) insertions (ins) or *EGFR* ex20 ins using 1021 gene panel in Geneplus cohort

Characteristics	*ERBB2* ex20 ins (*n* = 370)	*EGFR* ex20 ins (*n* = 323)	*P* value
Age, years			<0.001
Mean (SD)	51.6 (12.4)	57.1 (11.8)	
Median (min-max)	52 (24-81)	58 (18-87)	
Gender, *n* (%)			0.03
Male	116 (31.3)	128 (39.6)	
Female	254 (68.7)	195 (60.4)	
Smoking, *n* (%)			0.24
Never	180 (80)	148 (75.1)	
Ever	45 (20)	49 (24.9)	
Unknown	145 (100)	126 (100)	
Tumor stage, *n* (%)			0.68[a]
0	0	11 (9.0)	
I	23 (21.7)	31 (25.4)	
II	6 (5.7)	0	
III	14 (13.2)	4 (3.3)	
IV	63 (59.4)	76 (62.3)	
Unknown	264 (100)	201 (100)	
Histology, *n* (%)			<0.001[b]
Adenocarcinoma	323 (99.4)	251 (94.0)	
Squamous cell carcinoma	1 (0.3)	5 (1.9)	
Adenosquamous carcinoma	1 (0.3)	0	
In situ adenocarcinoma	0	11 (4.1)	
Unknown	45 (100)	56 (100)	
Germline mutation, *n* (%)			0.48
Yes	8 (2.2)	10 (3.1)	
No	362 (97.8)	313 (96.9)	
PD-L1 test, *n* (%)			0.96
TPS < 1%	40 (64.5)	35 (64.8)	
TPS,1-50%	17 (27.4)	14 (25.9)	
TPS ≥ 50%	5 (8.1)	5 (9.3)	
Unknown	308 (100)	269 (100)	
TMB			0.007
Mean (SD)	2.8 (4.0)	3.6 (3.9)	
Median (min-max)	1.9 (0-28.8)	2.9 (0-28.8)	

[a]Comparison between 0-III and IV.
[b]Comparison between adenocarcinoma and non-adenocarcinoma.

(*n* = 370) and *EGFR* exon 20 insertions (*n* = 323) (Table 3). Patients with *ERBB2* exon 20 insertion are significantly associated with younger age (mean age: 51.6 vs 57.1 years; $p < 0.001$), and female gender (69% vs 60%; $p = 0.03$). A lower level of TMB (mean value: 2.8 vs 3.6; $p = 0.007$) was observed in patients with *ERBB2* exon 20 insertion than those with *EGFR* exon 20 insertion.

We then compared the exon 20 insertion patterns between *EGFR* and *ERBB2*. In the Geneplus dataset, 276 out of 370 (74.7%) *ERBB2* exon20 insertions were Y772_A755dup YVMA alteration in the alpha C-helix domain, and 25% of the cases were located in the loop following the alpha C-helix domain (Fig. 3a). Similar percentage of samples in each location was identified in the Guardant360 cohort (Fig. 3b). For *EGFR* exon 20 insertions, in the Geneplus cohort, 29.3% were A767_V769dup and 20.2% with S768_D770dup located in the near-loop. In the far loop, H773dup (4.6%) is the most frequently amino acid change, followed by H773_V774dup (3%) and H773_V774insAH (1.6%) (Fig. 3c). Similarly, In the Guardant360 cohort, in patients with *EGFR* exon 20 insertion, A767_V769dup and S768_D770dup are the most frequently insertions in the near loop, whereas H773dup and V774_C775insHV are more frequently occurred in the far loop (Fig. 3d). Overall, *EGFR* exon 20 insertions had a more diverse
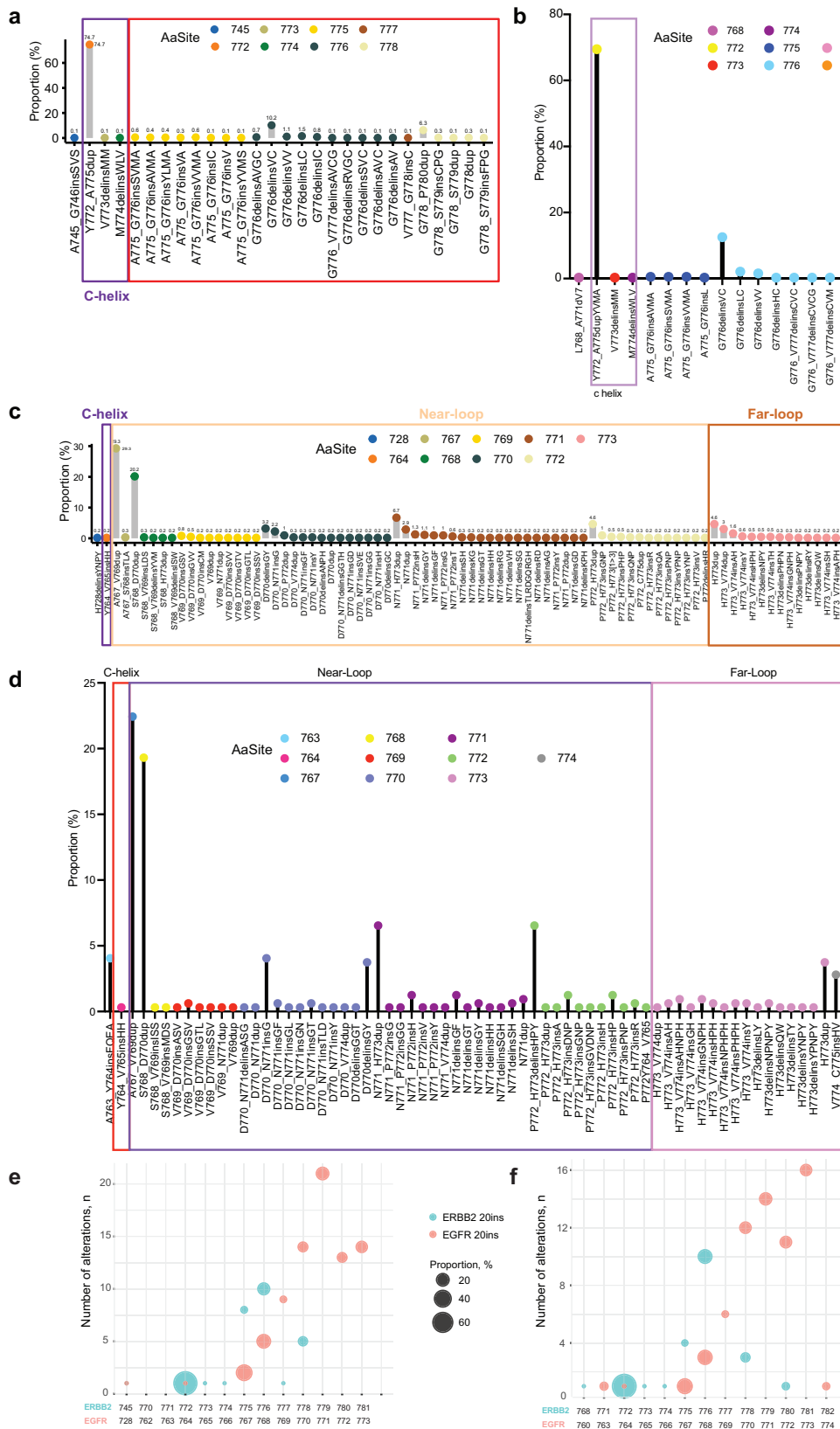
Fig. 3 | Difference between *EGFR* and *ERBB2* exon 20 insertions (ex20 ins) in NSCLC. *ERBB2* ex20 ins in Geneplus (**a**) and Guardant360 (**b**). *EGFR* ex20 ins in Geneplus (**c**) and Guardant360 (**d**). Bubble plots of ex20 ins comparison between *ERBB2* and *EGFR* in NSCLC patients in Geneplus (**e**) and Guardant360 (**f**).

insertion distribution than *ERBB2* exon 20 insertions in both cohorts. Most of the variants were diverse in-frame insertions or duplications of one to four amino acids spanning A767-V774 in *EGFR* and Y772-P780 in *ERBB2* (Fig. 3e, f).

We also compared the co-mutation landscape between *EGFR* and *ERBB2* exon20 insertions. In tumors with *ERBB2* exon 20 insertion, *TP53* is the most frequently co-mutated gene (29%), followed by *MED12* (7%), *LRP1B* (5%) and *RB1* (5%) (Fig. 4a). In *EGFR* exon 20 insertion tumors (Fig. 4b), top co-mutation genes were *TP53* (46%), *LRP1B* (9%), *RB1* (9%), and *MED12* (8%), sharing similarities. When directly compared, co-mutation in *TERT* were more frequently occurred in *ERBB2* than *EGFR* exon 20 insertion, whereas *TP53, RBM10, CTNNB1, TBX3, STAG2, RPTOR,* and *MLH3* were more frequently occurred in patients with *EGFR* exon 20 insertion (Fig. 4c and Supplementary Fig. 4a). *EGFR* and *ERBB2* each also enrich to have copy number gain with itself when copy number gain were compared (Fig. 4d). *ERBB2* co-mutations occurred in 13 samples with *ERBB2* exon 20 insertion and 3 samples with *EGFR* exon 20 insertion (Supplementary Fig. 4b), whereas *EGFR* co-mutations occurred in 8 cases with *ERBB2* exon 20 insertion and 42 cases with *EGFR* exon 20 insertion (Supplementary Fig. 4c). Regarding copy number variants (CNV), *MYC, MDM2, KRAS, IKZF1, CARD11,* and *PMS2* were associated with *EGFR* exon 20 insertion.

Discussion

In this study, we reported largest mutational profile dataset of NSCLC harboring *ERBB2* alterations, including both Geneplus (Chinese) patients and Guardant360 (US) patients and provided important insights into *ERBB2* alterations in NSCLC.

The incidence of *ERBB2* mutations in NSCLC has been controversial, often cited as 1%-6%. The incidence is dependent on the definition of the mutation, in the Guardant360 cohort, without filtering, 5.2% NSCLC samples had at least one *ERBB2* alterations; while after OncoKB and COS-MIC inferences, the incidence on annotated *ERBB2* oncogenic alterations decreased, and the oncogenic incidence was 1.9%. Unlike *EGFR* mutations with a strong predisposition in Asian patients, in previous reports, the prevalence of *ERBB2* oncogenic alterations appears to be comparable between patients from the US or France (~2%)[3,8] and Asian (4–8%)[14,25] population. Based on the *ERBB2* alterations' subtypes, previous studies have reported gene mutations (3–5%)[5,8,26], gene amplification (2–5%), and protein overexpression (2–30%)[27] in lung cancers. Here, we report oncogenic *ERBB2* mutations at 1.9–5.1% and amplification at 0.5–1.8%.

We characterized the demographics of NSCLC patients with *ERBB2* mutation or amplification. Patients with *ERBB2* mutation NSCLC were relatively young, with a female and never-smoker dominance. The tumors are mostly adenocarcinoma. Those features resemble *EGFR*-mutant NSCLC closely although with some difference. This demographics similarity was further confirmed in our comparison between *EGFR* and *ERBB2* exon 20 patient populations. However, patients with *ERBB2* amplification NSCLC had distinct features, including male and smoker dominance, as well as high TMB, similar to previous reports[24,28] and suggesting that *ERBB2* amplification might not be a strong oncogenic driver.

Previous studies had demonstrated variant-specific differences in patient outcomes on targeted treatment. Dacomitinib treatment resulted in an ORR of 11.5% for *ERBB2* mutant NSCLC but no response in patients with *ERBB2* exon 20 insertion Y772dupYVMA[29]. In a pan-*ERBB2* mutant NSCLC study testing efficacy of T-DM1, patients harboring *ERBB2* exon 20 insertion had an ORR of 54.5%, but patients with *ERBB2* exon 19 mutation (p.L755P) did not have responses[30]. Therefore, the understanding of mutational landscape is important for offering precision oncology treatment to the matching patients, and crucial for future drug development. We analyzed the frequency of mutations within the various regions reported by OncoKB and COSMIC. In cohorts from Asian and Western countries, *ERBB2* exon 20 insertions frequently occur (>50%), which had poor response to TKI (ORR: 0–18.8%)[29,31–33], while good response to ADCs (14.3–75%) reported by previous clinical studies[30,34–36].

Y772_V775dupYVMA is the most recurrent mutation in the kinase domain in Geneplus and Guardant360 datasets, 58% and 41.6%, respectively. Worse survival was associated with A775_G776YVMA in NSCLC when compared to other less common *ERBB2* alterations[37]. Conventional platinum-based chemotherapy has previously been found to have worse PFS in those with A775_G776YVMA compared with other *ERBB2* variants. However, studies reported that ADCs had better outcome in this group of patients. The most recurrent mutation variant in the non-TK domain is S310F, which located in the furin-like cysteine-rich domain, 8.5% in Geneplus and 10.7% in Guardant360, respectively. Furin-like cysteine-rich domain contain numerous cysteine residuals that participate in disulfide bond formation, and in homodimer and heterodimer formation with other ErbB family members. S310F mutation promoted *ERBB2* homodimerization and consequent auto-phosphorylation to activate the downstream *PI3K/AKT* and *MAPK* pathway, which was independent on *ERBB1, ERBB3,* and *ERBB4*, contributing to the growth and migration of cancer cells[38]. Responses to ADCs in transmembrane and extracellular domain *ERBB2* mutations (V659E and S310F) have been reported[30]. In summary, due to the high incidence of *ERBB2* exon 20 A775_G776YVMA, the population itself can represents an area of focused drug development. Drugs targeting trans-membrane and extracellular domain *ERBB2* mutations are also quite needed for the field.

Co-occurring genetic alterations with an oncogenic mutation can also associate with clinical response or resistance. Our findings indicated that concurrent driver mutations were mostly mutually exclusive with *ERBB2* kinase domain mutation, including *KRAS, ALK,* or *BRAF* actionable alterations. However, a relatively higher frequency of co-mutation with *EGFR* was observed in non-TK domain mutation, especially S310F, which was dominant or co-dominant with *EGFR* mutations. Furthermore, a significantly higher frequency of co-mutation with *EGFR* was showed in patients with *ERBB2* amplification group, indicating that S310F and *ERBB2* amplification could be a potential mechanism of resistance to *EGFR* TKI[39,40]. Dual targetable drugs such as afatinib or TKI plus ADCs could be a potential choice for this special group of patients with both *EGFR* and *ERBB2* actionable alterations.

Compared to *EGFR* classical mutations, *EGFR* exon 20 insertions amount to a small fraction of *EGFR* mutations and now has a distinct set of targeted therapy approvals in lung cancer patients. As described above, *ERBB2* exon 20 insertions are the dominant alterations in NSCLC patients harboring *ERBB2* alterations, different than *EGFR* classical mutations being the most common, exon 20 insertions less common. We assessed the characteristics between these two exon 20 insertions. We observed high heterogeneity in *EGFR* exon 20 insertions compared to *ERBB2* exon 20 A775_G776YVMA being highly dominant. In the past, some TKIs were developed to target both *EGFR* and *ERBB2* exon 20 insertions, including poziotinib, mobocertinib and pyrotinib. Poziotinib is an irreversible pan-HER TKI. ZENITH 20 trial demonstrated response rate of 14.8% in previously treated *EGFR* exon 20 insertion[41] and 27.8% in *ERBB2* exon 20 insertions[42]. Data from another trial treated with mobocertinib showed an ORR of 28% for patients with *EGFR* exon 20 insertion[22]. In patients with *ERBB2* exon 20 insertions treated with pyrotinib, a pan-HER TKI against HER1/2/4, a phase II study reported that the ORR was 31.7% and PFS was 6.8 months in *ERBB2* exon 20 insertion patients[43], which is similar to those patients with *EGFR* exon 20 insertion treated with pan-HER TKI. For this groups of patients with *ERBB2* exon 20 insertion, ADCs showed excellent results in DESTINY-Lung01 clinical trial and can probably change the clinical practice. Now, small molecules designed to specifically targeting *ERBB2* mutations are under investigation, which potentially spare targeting *EGFR* related toxicities.

There are strengths and weaknesses to our study. To our knowledge, this is the biggest cohort in comparison of NSCLC patients with *ERBB2* alterations between Chinese and the US datasets, however, the conclusion is limited by the heterogeneity of lab assays and sample source. Furthermore, due to the real-world study with incomplete information, clinical outcome data to certain drugs are not assessed in this study. Finally, for the VUS,
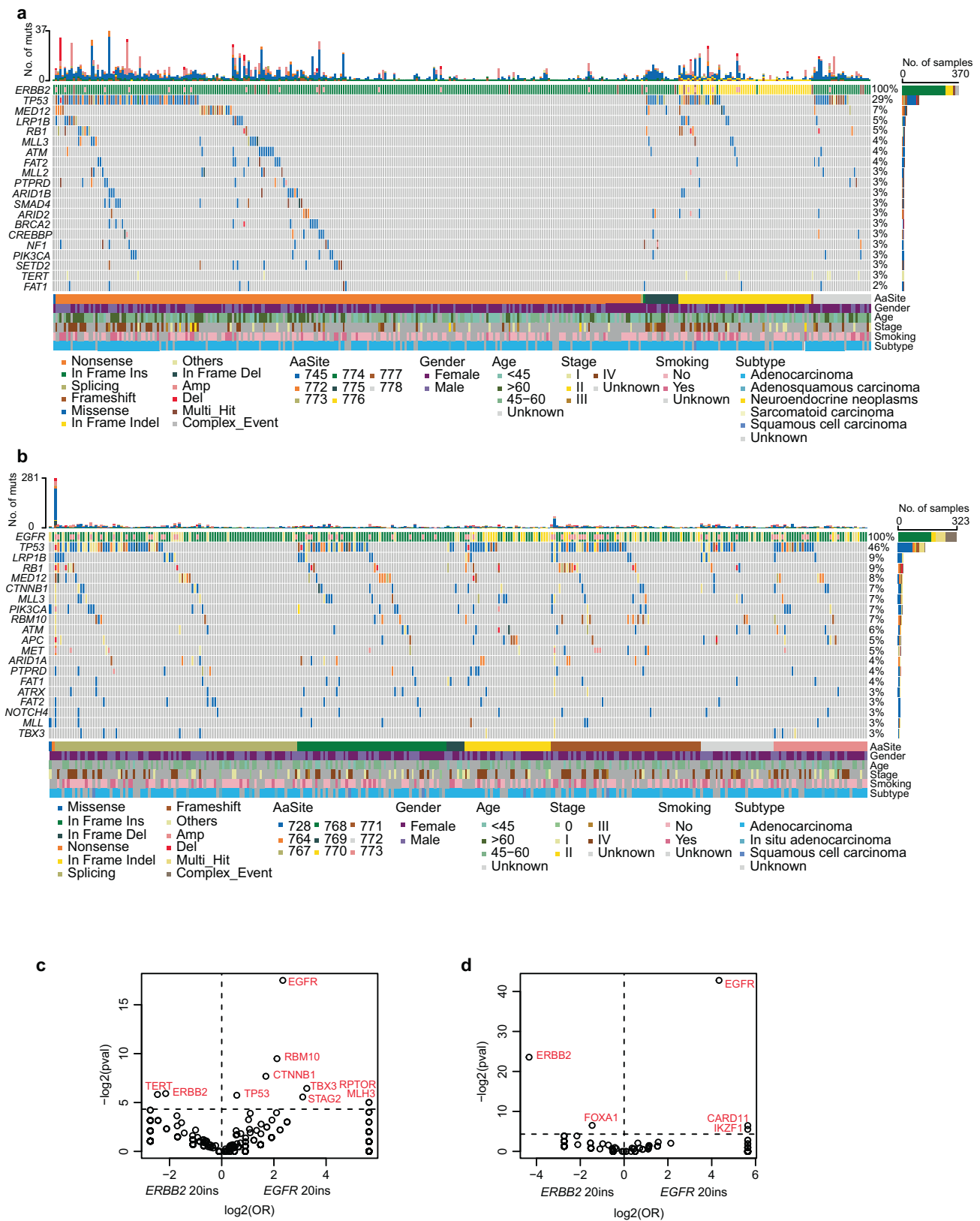
Fig. 4 | Differential co-mutated genes between patients with *EGFR* and *ERBB2* exon 20 insertions (ex20 ins) in Geneplus cohort. Co-mutation plots of *ERBB2* (**a**) and *EGFR* (**b**). Enrichment analysis of co-altered genes between patients with *ERBB2* ex20 ins and *EGFR* ex20 ins: (**c**) gene SNVs/Indels; (**d**) gene CNVs.

further technologies and bioinformatic approaches are needed to identify the function of rare variants. As such, future studies, both experimental and clinical, are warranted to validate these provocative genomic findings and their clinical implications.

In conclusion, in two large independent cohorts, Geneplus with patients from China and Guardant360 with patients from the United States, *ERBB2* mutation and co-mutation patterns were similar. *ERBB2* exon 20 insertions/mutations were dominant at over 80% with Y772_A775dupYVMA being the

most common driver mutation; *TP53* and *EGFR* were the most frequently co-occurred genes. *ERBB2* mutation lung cancers had low TMB and PDL1, as expected in female and never-smoker dominant lung adenocarcinomas, similar to *EGFR* exon 20 NSCLC.

Methods

Study population and platforms

Two retrospective cohorts of NSCLC patients were analyzed for *ERBB2* alterations: Geneplus (both ctDNA and tissue, November 2016 to April 2022); Guardant360 (ctDNA, June 2019 to Oct 2021). The panels used in tissue or blood samples from Geneplus were summarized in Supplementary Tables 2–6. Supplementary Tables 7–8 showed Guardant360 ctDNA 74 and 83 gene-panels. All patients in the study provided written informed consent. The study protocol was approved by Institutional Review Board of Peking Union Medical College Hospital (K3415) and was conducted in accordance with the Declaration of Helsinki and principles of Good Clinical Practice. This study is compliant with the Guidance of the Ministry of Science and Technology (MOST) for the Review and Approval of Human Genetic Resources, with the formal approval of publishing data in a scientific journal (2024BAT00891). The raw data from China will not leave the country and will not be disclosed publicly. The generation of de-identified data sets by Guardant Health for research purposes was approved by the Advarra Institutional Review Board (Pro00034566); patient identity protection was maintained throughout the study in a de-identified database.

DNA extraction and targeted next-generation sequencing (Geneplus)

DNA was isolated from tissue samples, peripheral blood, pleura effusion and hydrothorax using commercial kits (Qiagen, Hilden, Germany). Peripheral blood leukocytes were separated to extract germline genomic DNA using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). The KAPA Library Preparation Kit (Kapa Biosystems, Wilmington, MA, USA) was used to prepare indexed Illumina NGS libraries. Custom-designed 1021, 73, 36, 59 or 86 cancer-related gene panels were used to hybridize the DNA libraries, and their selected regions and genes are listed in Supplementary Tables 2–6. The hybridized libraries were sequenced using a 100-bp paired-end configuration on a DNBSEQ-T7RS sequencer (MGI Tech, Shenzhen, China). The minimal mean effective depth of coverage for tissue and germline DNA was 300× and for ctDNA was 1000×.

After the removal of terminal adaptor sequences and low-quality reads with FASTP[44], the remaining reads were mapped to the reference human genome (hg19) and aligned using the Burrows-Wheel Aligner (version 0.7.12-r1039) with default parameters. GATK (3.4–46-gbc02625) and MuTect2 (1.1.4) were used to call somatic single nucleotide variants and small insertions and deletions. Contra (2.0.8) was used to identify copy number variations[45]. NCSV (in-house algorithm 0.2.3) was employed to detect structural variants[46]. All candidate variants were manually confirmed by using the integrative genomics viewer browser. Variants were filtered to exclude clonal hematopoietic mutations with an inhouse database of clonal hematopoiesis variants of >10,000 pan-cancer patients and healthy individuals[47], germline mutations in dbSNP, as well as variants that occur at a population frequency of >1% in the Exome Sequencing Project. Oncogenic/neutral *ERBB2* mutations are those that are documented as pathogenic/neutral in the OncoKB (https://www.oncokb.org/) and COSMIC database (https://cancer.sanger.ac.uk/cosmic) databases. Unknown *ERBB2* mutations are those that are not reported by either OncoKB or COSMIC databases. Copy number alterations of *ERBB2* with copy number ≥ 2.6 were considered as amplification, and the alterations were manually confirmed with CNV plot.

Tumor mutation burden (TMB) and PD-L1 expression evaluation (Geneplus)

The TMB was determined as the number of somatic non-synonymous single nucleotide variants and small insertions/deletions per mega-base in the coding region (with VAF ≥ 0.03 for tumor tissues and ≥0.005 for ctDNA, respectively). Immunohistochemistry with the PD-L1 IHC 22C3 pharmDx assay (Agilent Technologies, Santa Clara, CA, USA) was performed to evaluate PD-L1 expression of tumor tissues.

Clonality analysis (Geneplus)

Somatic substitution/small insertions and deletions were applied to PyClone by default to analyze the clonal structure using a Bayesian clustering method[48]. Cancer cell fraction was calculated with the mean of predicted cellular frequencies. The cluster with the highest mean VAF was identified as the clonal cluster, and mutations in this cluster were clonal mutations. Meanwhile, other clusters and mutations were considered subclonal.

ctDNA sequencing and analysis (Guardant360)

ctDNA was evaluated using the commercially available Guardant360™ assay (Guardant Health, Inc., Redwood City, CA) to evaluate up to 83 cancer-related genes as previously described[49]. The Guardant360 assay is a comprehensive genomic profiling assay that identifies single-nucleotide variants (SNVs), insertions and deletions, fusions, and amplifications[50]. The assay covers complete exon sequencing of multiple genes, including *EGFR*, *ERBB2*, and *KRAS*. During the collection period, the assay included 74 to 83 genes. The NGS testing was performed as part of standard clinical care in a CLIA-certified and College of American Pathologists accredited laboratory[51]. Blood was collected in two 10 mL Streck tubes and processed plasma was evaluated for single-nucleotide variants (SNVs), insertions-deletions (indels), gene fusions/rearrangements, and copy number variants (CNVs)[51]. Mutations were annotated using OncoKB to define pathogenic variants[52]. Synonymous mutations and variants of unknown significance were not considered to be clinically relevant but were included as indicators of tumor shed in the plasma. *ERBB2* amplifications ≥2.2 copies were included in the Guardant360 cohort.

Clonality analysis (Guardant360)

Variant clonality was determined by normalizing VAF to the maximum somatic VAF in a sample. Variants were classified as clonal if the normalized value was ≥0.5.

Statistical analysis

Statistical analyses were performed using R statistical software (version 4.1.3 for Windows). A comparison of categorical variables was conducted with Pearson's $\chi2$ test or Fisher's exact tests. The Mann-Whiney U test and Student's t-test were used for nonnormally and normally distributed continuous variables, respectively. The Kruskal-Wallis test was used to compare non-normally distributed continuous variables among three or more independently sampled groups. All statistical tests were performed with two-sided methods, and $P < 0.05$ was considered to indicate statistical significance.

Data availability

The de-identified original data or variation data are available from corresponding author (Y. Wang for Geneplus; J. Zhang and X. Le for Guardant360) upon reasonable request, and qualified researchers can apply for access to the datasets with a data usage agreement. All other data may be found within the main manuscript or Supplementary Information.

References

1. Wee, P. & Wang, Z. Epidermal growth factor receptor cell proliferation signaling pathways. *Cancers* **9**, 52 (2017).
2. Friedlaender, A. et al. EGFR and HER2 exon 20 insertions in solid tumours: from biology to treatment. *Nat. Rev. Clin. Oncol.* **19**, 51–69 (2022).
3. Arcila, M. E. et al. Prevalence, clinicopathologic associations, and molecular spectrum of ERBB2 (HER2) tyrosine kinase mutations in lung adenocarcinomas. *Clin. Cancer Res.* **18**, 4910–8, (2012).

4. Shigematsu, H. et al. Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res*. **65**, 1642–6, (2005).

5. Tomizawa, K. et al. Prognostic and predictive implications of HER2/ERBB2/neu gene mutations in lung cancers. *Lung Cancer* **74**, 139–44, (2011).

6. Perera, S. A. et al. HER2YVMA drives rapid development of adenosquamous lung tumors in mice that are sensitive to BIBW2992 and rapamycin combination therapy. *Proc. Natl. Acad. Sci. USA* **106**, 474–9, (2009).

7. Li, C. et al. Lung adenocarcinomas with HER2-activating mutations are associated with distinct clinical features and HER2/EGFR copy number gains. *J. Thorac. Oncol.* **7**, 85–9, (2012).

8. Mazieres, J. et al. Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. *J. Clin. Oncol.* **31**, 1997–2003 (2013).

9. Pellegrini, C. et al. HER-2/Neu alterations in non-small cell lung cancer: a comprehensive evaluation by real time reverse transcription-PCR, fluorescence in situ hybridization, and immunohistochemistry. *Clin. Cancer Res.* **9**, 3645–52, (2003).

10. Buttitta, F. et al. Mutational analysis of the HER2 gene in lung tumors from Caucasian patients: mutations are mainly present in adenocarcinomas with bronchioloalveolar features. *Int. J. Cancer* **119**, 2586–91, (2006).

11. Stephens, P. et al. Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature* **431**, 525–6, (2004).

12. Heinmoller, P. et al. HER2 status in non-small cell lung cancer: results from patient screening for enrollment to a phase II study of herceptin. *Clin. Cancer Res*. **9**, 5238–43, (2003).

13. Rouquette, I. et al. Characteristics of lung cancer in women: importance of hormonal and growth factors. *Lung Cancer* **76**, 280–5, (2012).

14. Garrido-Castro, A. C. & Felip, E. HER2 driven non-small cell lung cancer (NSCLC): potential therapeutic approaches. *Transl. Lung Cancer Res*. **2**, 122–7, (2013).

15. Pahuja, K. B. et al. Actionable activating oncogenic ERBB2/HER2 transmembrane and juxtamembrane domain mutations. *Cancer Cell* **34**, 792–806 e5 (2018).

16. Greulich, H. et al. Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2. *Proc. Natl. Acad. Sci. USA* **109**, 14476–81, (2012).

17. Goto K., S-WK. et al. LBA55 Trastuzumab deruxtecan (T-DXd) in patients (Pts) with HER2-mutant metastatic non-small cell lung cancer (NSCLC): Interim results from the phase 2 DESTINY-Lung02 trial. *Ann. Oncol.* 2022

18. Li, B. T. et al. Trastuzumab deruxtecan in HER2-mutant non-small-cell lung cancer. *N. Engl. J. Med*. **386**, 241–251 (2022).

19. Goto, K. et al. Trastuzumab deruxtecan in patients with HER2-mutant metastatic non-small-cell lung cancer: Primary results from the randomized, phase II DESTINY-Lung02 trial. *J. Clin. Oncol.* **41**, 4852–4863 (2023).

20. Elamin, Y. Y. et al. Poziotinib for EGFR exon 20-mutant NSCLC: Clinical efficacy, resistance mechanisms, and impact of insertion location on drug sensitivity. *Cancer Cell* **40**, 754–767 e6 (2022).

21. Robichaux, J. P. et al. Mechanisms and clinical activity of an EGFR and HER2 exon 20-selective kinase inhibitor in non-small cell lung cancer. *Nat. Med*. **24**, 638–646 (2018).

22. Zhou, C. et al. Treatment outcomes and safety of mobocertinib in platinum-pretreated patients with EGFR exon 20 insertion-positive metastatic non-small cell lung cancer: A phase 1/2 open-label nonrandomized clinical trial. *JAMA Oncol.* **7**, e214761 (2021).

23. Riely, G. J. et al. Activity and safety of mobocertinib (TAK-788) in previously treated non-small cell lung cancer with EGFR exon 20 insertion mutations from a phase I/II trial. *Cancer Discov.* **11**, 1688–1699 (2021).

24. Wang, H. et al. Molecular landscape of erbb2 alterations in 14,956 solid tumors. *Pathol. Oncol. Res.* **28**, 1610360 (2022).

25. Oh, I. J. et al. Clinical activity of Pan-HER Inhibitors against HER2-mutant lung adenocarcinoma. *Clin. Lung Cancer* **19**, e775–e781 (2018).

26. Pillai, R. N. et al. HER2 mutations in lung adenocarcinomas: A report from the lung cancer mutation consortium. *Cancer* **123**, 4099–4105 (2017).

27. Mishra, R., Hanker, A. B. & Garrett, J. T. Genomic alterations of ERBB receptors in cancer: clinical implications. *Oncotarget* **8**, 114371–114392 (2017).

28. Wang, Q. et al. Characteristics of the immunogenicity and tumor immune microenvironment in HER2-amplified lung adenocarcinoma. *Front Immunol.* **13**, 1042072 (2022).

29. Kris, M. G. et al. Targeting HER2 aberrations as actionable drivers in lung cancers: phase II trial of the pan-HER tyrosine kinase inhibitor dacomitinib in patients with HER2-mutant or amplified tumors. *Ann. Oncol.* **26**, 1421–1427 (2015).

30. Li, B. T. et al. Ado-trastuzumab emtansine for patients with HER2-mutant lung cancers: Results from a phase II basket trial. *J. Clin. Oncol.* **36**, 2532–2537 (2018).

31. De Greve, J. et al. Phase II study of afatinib, an irreversible ErbB family blocker, in demographically and genotypically defined lung adenocarcinoma. *Lung Cancer* **88**, 63–9, (2015).

32. Dziadziuszko, R. et al. Afatinib in NSCLC with HER2 mutations: Results of the prospective, open-label phase II NICHE trial of European thoracic oncology platform (ETOP). *J. Thorac. Oncol.* **14**, 1086–1094 (2019).

33. Hyman, D. M. et al. HER kinase inhibition in patients with HER2- and HER3-mutant cancers. *Nature* **554**, 189–194 (2018).

34. Hotta, K. et al. A phase II study of trastuzumab emtansine in HER2-positive non-small cell lung cancer. *J. Thorac. Oncol.* **13**, 273–279 (2018).

35. Li, B. T. et al. HER2-mediated internalization of cytotoxic agents in ERBB2 amplified or mutant lung cancers. *Cancer Discov.* **10**, 674–687 (2020).

36. Peters, S. et al. Trastuzumab emtansine (T-DM1) in patients with previously treated HER2-overexpressing metastatic non-small cell lung cancer: Efficacy, safety, and biomarkers. *Clin. Cancer Res.* **25**, 64–72 (2019).

37. Offin, M. et al. Frequency and outcomes of brain metastases in patients with HER2-mutant lung cancers. *Cancer* **125**, 4380–4387 (2019).

38. Fan, K. et al. ERBB2 S310F mutation independently activates PI3K/AKT and MAPK pathways through homodimers to contribute gallbladder carcinoma growth. *Med Oncol.* **39**, 64 (2022).

39. Takezawa, K. et al. HER2 amplification: a potential mechanism of acquired resistance to EGFR inhibition in EGFR-mutant lung cancers that lack the second-site EGFR T790M mutation. *Cancer Discov.* **2**, 922–933 (2012).

40. Ohashi, K. et al. Impact of HER2 expression on EGFR-TKI treatment outcomes in lung tumors harboring EGFR mutations: A HER2-CS study subset analysis. *Lung Cancer* **150**, 83–89 (2020).

41. Le, X. et al. Abstract CT081: Poziotinib activity and durability of responses in previously treated EGFR exon 20 NSCLC patients - a Phase 2 study. *Cancer Res.* **80**, CT081–CT081 (2020).

42. Le, X. et al. Poziotinib in non-small-cell lung cancer harboring HER2 exon 20 insertion mutations after prior therapies: ZENITH20-2 trial. *J. Clin. Oncol.* **40**, 710–718 (2022).

43. Zhou, C. et al. Pyrotinib in HER2-mutant advanced lung adenocarcinoma after platinum-based chemotherapy: A multicenter, open-label, single-arm, phase II study. *J. Clin. Oncol.* **38**, 2753–2761 (2020).

44. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

45. Li, J. et al. CONTRA: Copy number analysis for targeted resequencing. *Bioinformatics* **28**, 1307–13, (2012).

46. Zhang, J. T. et al. Longitudinal undetectable molecular residual disease defines potentially cured population in localized non-small cell lung cancer. *Cancer Discov.* **12**, 1690–1701 (2022).

47. Zhang, Y. et al. Pan-cancer circulating tumor DNA detection in over 10,000 Chinese patients. *Nat. Commun.* **12**, 11 (2021).

48. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–8, (2014).
49. Nazha, B. et al. Blood-based next-generation sequencing in adrenocortical carcinoma. *Oncologist* **27**, 462–468 (2022).
50. Lanman, R. B. et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS One* **10**, e0140712 (2015).
51. Odegaard, J. I. et al. Validation of a plasma-based comprehensive cancer genotyping assay utilizing orthogonal tissue- and plasma-based methodologies. *Clin. Cancer Res.* **24**, 3539–3549 (2018).
52. Chakravarty, D. et al. OncoKB: A precision oncology knowledge base. *JCO Precis. Oncol.* **1**, 1–16 (2017).

Author contributions
Concept and design of the study: L.H., Y.W., J.Z., and X.L. Drafting of the manuscript: L.H., S.P., L.M.D., Y.X., Y.W., J.Z., and X.L. Acquisition, analysis, or interpretation of data: All authors. Critical revision of the manuscript: All authors. Final approval of manuscript: All authors. Y.W., J.Z. and X.L. led study supervision.

Competing interests
L.M.D. is employee and stockholder of Guardant Health. Y.X. and R.C. are employee of Geneplus-Beijing. S.H. received consulting fees from Guardant health and AstraZeneca. M.N. receives royalties and licensing fees from Spectrum Pharmaceuticals. J.V.H. reports receiving advisory/consulting fees from AstraZeneca, Boehringer-Ingeheim, Catalyst, Genentech, GlaxoSmithKline, Guardant Health, Foundation Medicine, Hengrui Therapeutics, Eli Lilly, Novartis, Spectrum, Sanofi, Takeda Pharmaceuticals, Mirati Therapeutics, Bristiol-Myers Squibb, BrightPath Biotherapeutics, Janssen Global Services, Nexus Health Systems, EMD Serono, Pneuma Respiratory, Kairos Venture Investments, Leads Biolabs, RefleXion, and research funding from GlaxoSmithKline, AstraZeneca, Spectrum, all outside of the submitted work. J.Z. reports grants from Merck, grants and personal fees from Johnson and Johnson and Novartis, and personal fees from Bristol Myers Squibb, AstraZeneca, GenePlus, Innovent, and Hengrui outside the submitted work. X.L. receives consulting/advisory fees from EMD Serono (Merck KGaA), AstraZeneca, Spectrum Pharmaceutics, Novartis, Eli Lilly, Boehringer Ingelheim, Hengrui Therapeutics, Janssen, Blueprint Medicines, Regeneron, Sensei Biotherapeutics, Abion, Pinetree Therapeutics and Abbvie, and Research Funding from Eli Lilly, EMD Serono, ArriVent, Teligene, Regeneron, Blackdiamond, and Boehringer Ingelheim. The other authors declare no competing interests in the submitted work.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-024-00720-9.

Correspondence and requests for materials should be addressed to Yingyi Wang, Jianjun Zhang or Xiuning Le.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.