EGFR protein editing by serine hydroxymethyltransferase SHMT2

Y. Eugene Chin (chinyue@suda.edu.cn) 
  Soochow University

Juanjuan Du  
  Zhejiang Provincial People's Hospital

Yuan-yuan Zeng  
  Zhejiang Proincial People's Hospital and Soochow University

Xue Li  
  Zhejiang Proincial People's Hospital

Fang Huang  
  Zhejiang Proincial People's Hospital

Chuanzan Zhou  
  Zhejiang Proincial People's Hospital

Rui-dong Zi  
  Zhejiang Proincial People's Hospital and Soochow University

Lu-ying Liu  
  Binzhou Medical University

Jianyuan Luo  
  Peking University Health Science Center  https://orcid.org/0000-0001-6057-2914

Ting Zhao  
  Brown University

Biological Sciences - Article

Keywords:

Posted Date: July 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3105455/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is NO Competing Interest.
EGFR protein editing by serine hydroxymethyl transferase SHMT2

Juanjuan K. Du1,2,3,4*, Yuan-yuan Zeng1,2,4*, Xue S. Li1,2, Fang F. Huang1, Chuanzan C. Zhou1, Rui-Dong Zi1,4, Lu-Ying Liu2, Jianyuan Luo5, Ting C. Zhao6, and Y. Eugene Chin1,2,3,4**

1Clinical Medicine Research Institute, Zhejiang Provincial People’s Hospital, Hangzhou Medical College, Hangzhou, Zhejiang 310014

2Yantai Peninsular Cancer Center, Binzhou Medical University, Yantai, Shandong 264003

3The Second Affiliated Hospital of Wenzhou Medical University, Wenzhou Zhejiang 325027

4The First Affiliated Hospital and Institutes of Biological and Medical Sciences, Soochow University, Suzhou, Jiangsu 2151006

5Department of Medical Genetics, Peking University Health Science Center, Beijing 100191

6Departments of Surgery and Plastic Surgery, Rhode Island Hospital, Alpert Medical School of Brown University, Providence, Rhode Island 02903

* These authors contributed equally.
**Corresponding author YEC email: chinyue@suda.edu.cn
ABSTRACT
The epidermal growth factor receptor (EGFR), a transmembrane tyrosine kinase, relies on phosphorylation for signal transduction\textsuperscript{1,2}. It is known that the \textit{EGFR} gene often harbors activating mutations for protein sequence change, leading to constitutive phosphorylation and activation of the EGFR protein in various cancers\textsuperscript{3}. This study reveals that EGFR proteins, isolated from lung cancer tissues or cell lines, possess multiple glycine-to-serine (G>S) protein sequence modifications, which are not associated with \textit{EGFR} gene mutations. These EGFR proteins enlist Serine Hydroxymethyl Transferase 2 (SHMT2) to induce glycine hydroxymethylation, resulting in G>S/GS>SG protein sequence editing. SHMT2 catalyzes the protein editing predominantly within the GS-islands of the EGFR protein, utilizing either free serine amino acids present within cells or the serine residue adjacent to glycine residue as the hydroxymethyl group donor. The G>S/GS>SG editing within GS islands G719S720, G810S811, and G1103S1104 causes significantly amplify of EGFR signaling activity, stimulates cancer cell growth, and induces drug resistance. In lung cancer cells, depletion of SHMT2 or inhibition with SHIN1 suppresses glycine hydroxymethylation, eliminating EGFR protein G>S/GS>SG editing and reducing resistance to EGFR-targeting drugs. Our findings not only identify a novel form of protein mutation through directly editing, but also underscore the importance of protein editing driving EGFR activation, thereby presenting potential avenues for EGFR-targeted therapy.
Post-translational modifications of proteins serve a paramount role in signal transduction. EGFR proteins, located on the membranes of epithelial cells, contain an inherent tyrosine kinase (TK), capable of being activated through tyrosine autophosphorylation and serine phosphorylation. This leads to the activation of EGFR-associated proteins, initiating signal transduction. The cytoplasmic region of the EGFR is comprised of the TK domain and the regulatory domain, which can either positively or negatively regulate TK activity. Frequently, the EGFR gene mutation-based EGFR protein sequence change or overexpression of the EGFR gene result in constitutive tyrosine autophosphorylation and serine phosphorylation, leading to unregulated activation in cancer cells. In non-small cell lung cancer (NSCLC), EGFR gene mutations occur at an incidence rate of 2.5 per 100,000 individuals. Despite the presence of multiple mutations throughout the EGFR gene, most mutations are in exons 18-21 of the cytoplasmic TK domain in NSCLC. Exon 19 deletion (Del19) and L858R of exon 21 are the most common EGFR gene mutations, while EGFR gene exon 18 mutations such as G719X (X = S, C, A) are less common but increasingly detected. These atypical EGFR gene mutations often lead to poorer patient outcomes. As increasingly sensitive DNA sequencing technologies have emerged, more canonical and noncanonical EGFR gene mutations have been documented in both extracellular and cytoplasmic regions.

EGFR TK inhibitor (TKI) targeted therapy has been developed for NSCLC patients with EGFR protein sequence change resulted from EGFR gene mutations, i.e., activating mutations. However, some NSCLC patients without detected EGFR gene mutations respond to EGFR TKI targeted therapy. This suggests the possibility of gene mutations and protein sequence change being uncoupled under certain conditions. To investigate this discrepancy, we have analyzed EGFR proteins in lung cancers. We found that EGFR protein carries multiple G>S and GS>SG sequence changes resulted from reversible hydroxymethylation editing. Serine hydroxymethyl-transferase (SHMT) family member SHMT2 is responsible for G>S or GS>SG editing along EGFR protein.

**EGFR protein carries gene mutation independent glycine-to-serine (G>S) protein sequence change.**

The EGFR protein, obtained through immunoprecipitation from lung cancer tissues of NSCLC patients, was subjected to mass spectrometry for analysis. Multiple G>S protein sequence changes within the EGFR protein's cytoplasmic domain in these lung adenocarcinoma tissues were detected (Fig.1a,b). Similarly, EGFR proteins procured from lung cancer cell lines exhibited multiple
G>S changes (Fig.1a,b). The human EGFR protein comprises 85 glycine residues, of which 28 are situated in the cytoplasmic domain (Fig.1a,b). This domain also contains 42 serine residues, with a significant proportion being phosphorylated like tyrosine residues\(^{15,16}\). Seven glycine residues in the cytoplasmic domain of the EGFR protein are succeeded by serine residues, referred as the "GS islands" that are conserved between human and mouse species (Fig.1b, Extended Fig.1a,b). Additionally, two "SG islands" are found within the TK domain of the EGFR protein: S695G696 and S720G721, with S720 being common to G719 GS island and G721 SG island. Among all the GS islands, G719S720, G810S811, and G1103S1104 were the most frequently detected GS islands with G>S/SG>SG protein sequence change but without DNA mutations (Fig.1b, Extended Fig.1c). These GS islands are highly conserved in EGFR proteins from human and mouse origins (Fig. 1b). Despite the noncanonical EGFR gene mutation resulting in G>X (X = C, T, A, Q, K)\(^{12}\), peptides with G>S/SG>SG protein sequence changes were predominantly observed in the lung adenocarcinoma samples and the analyzed lung cancer-cell lines (Fig.1c). Venn diagram analysis revealed that among all 52 lung cancer tissues and cell lines that carried G>S, GS>SG, or S>G protein sequence changes, G>S was far more frequent than S>G protein sequence change (Fig.1c). Peptides carrying G>S/SG>SG protein sequence change were identified and confirmed in the trypsin digestion coupled with mass-mass spectrometry analysis (Fig.1d). EGFR protein G719S resulted from EGFR gene mutation is known to activate the kinase by disrupting autoinhibitory interactions and accelerating catalysis\(^{17}\). The GS>SG switch in protein sequence of the G719 and G810 GS islands adopts conformations like the wildtype TK in its activated state (Extended Fig.1b).

Gene sequence (DNA and mRNA) analysis revealed no mutations of the EGFR gene corresponding to these GS-islands with G>S or S>G protein sequence change or we referred here as EGFR protein editing (Fig.1e, Extended Fig.1c,d). Therefore, the EGFR protein sequence can be changed without EGFR gene mutation.

**EGFR protein recruits SHMT2 protein.**

To investigate the discrepancy between EGFR protein and EGFR gene sequences of these GS islands, we began by examining EGFR protein binding partners. EGFR proteins, immunoprecipitated from lung cancer cell lines, were subjected to mass spectrometry for identification of binding proteins. Among the EGFR protein binding proteins identified, the SHMT family member SHMT2 protein, was associated with EGFR proteins in commonly-studied lung cancer
cell lines H1299 and A549, as well as other lung cancer cell lines under investigation (Fig.2a,b, Extended Data Fig.2a).

Although SHMT was originally identified during clostridial purine fermentation\textsuperscript{18}, SHMT1 and SHMT2 were subsequently discovered as mammalian members of the SHMT family. Although SHMT2 was visualized in mitochondria as immunostaining result indicated (Extended Data Fig.2b,c), a shorter form of SHMT2 was exclusively detected in the cytoplasm in both A549 and H1299 cells (Extended Data Fig.2b,c). It was reported that SHMT2 can be expressed either as the full length or as the shorter form of SHMT2, which lacks the N-terminal mitochondrial signal (SHMT2\textsubscript{α})\textsuperscript{19,20}. We constructed SHMT2 full-length and SHMT2\textsubscript{α} cDNAs for expression in HEK293T cells and noted that SHMT2\textsubscript{α} was the major form of SHMT2 in both A549 and H1299 cells (Extended Data Fig.2c)\textsuperscript{19,21}. SHMT forms a tetrameric enzyme complex within three isoforms: two cytosolic forms, SHMT1/SHMT2\textsubscript{α}, and one mitochondrial form, SHMT2\textsuperscript{22}. These enzyme complexes are all pyridoxal-5-phosphate (PLP) dependent while catalyzing the reversible conversion of L-serine and tetrahydrofolic acid (THF) into glycine and 5N,10N-methylene-THF (MTHF) during cellular amino acid synthesis and metabolism\textsuperscript{23}.

EGFR’s association with cytosolic SHMT2 was confirmed in H1299 and A549 cells with coimmunoprecipitation analysis (Fig.2c). SHMT2 has been reported to rely on lysine acetylation, succinylation or acylation for its catalytic activity\textsuperscript{21,24-26}. To analyze the effect of SHMT2 on EGFR, lysine mutated SHMT2 variants were constructed. Purified EGFR and SHMT2 proteins from HEK293T transfectants were applied for the protein-protein interaction constant (K\textsubscript{d}) analysis using a surface plasmon resonance (SPR) approach. Wild-type SHMT2 was found to bind to EGFR protein with a K\textsubscript{d} value of 6.16 µM (Fig.2d). Since K280 of SHMT2 is essential for SHMT2 PTP-mediated dimerization and tetramer formation\textsuperscript{19}, the failure of SHMT2-K280E to interact with EGFR (Fig.2d) indicates that K280-mediated tetramer formation is critical in EGFR association. SHMT2 carries multiple acetylable lysine residues within its C-terminal tail region\textsuperscript{20,21,26}. C-terminal lysine cluster mutated SHMT2 (5K>R) reduced its EGFR protein association ability nearing 3 folds (K\textsubscript{d} = 15.8 µM) comparing with wild type SHMT2 (Fig.2d,Extended Data Fig.2d)\textsuperscript{19,25,27}. Therefore, N-terminal 5K cluster of SHMT2 protein forms acetylation-dependent interaction with negatively charged residues of the EGFR protein (Fig.2e, Extended Data Fig.2d)\textsuperscript{25}. These results suggest that multiple lysine residues of SHMT2 such as K280 and the N-terminal lysine cluster play crucial roles not only in dimerization and tetramerization but also in substrate binding\textsuperscript{24-26}. 
SHMT2 catalyzes EGFR protein hydroxymethylation for G>S editing.

Our subsequent investigations aimed to unravel SHMT2 in catalyzing EGFR protein hydroxymethylation. In HEK293T cells, the transient overexpression of SHMT2 alongside EGFR resulted in the hydroxy-methylation of the G719 GS-island, leading to G719>S or G719S720>S719G720 protein sequence editing without causing the EGFR gene mutation responsible for G719 GS-island (Fig.2f,g). G>S or GS>SG protein sequence editing, stemming from glycine hydroxymethylation, were also identified within the G810 and G1103 GS-islands in HEK293T cells transfected with SHMT2 (Fig.2h). G719 carries a GSG protein sequence. S720G721>GS editing was also detected in both lung cancer tissues and cell lines (Fig.1c), the efficiency was lower as compared with G719S720>SG editing, suggesting that the efficacies of GS>SG and SG>GS editing may differ. To examine G>S editing of these GS islands in cancer samples, we prepared polyclonal antibodies, specifically recognizing the EGFR protein with hydroxymethylation G>S at G719, G810, and G1103 sites (Extended Data Fig.2g). Using these antibodies, we were able to visualize the induction of EGFR hydroxymethylation by EGF at these three glycine sites in H1299 cells. EGF clearly induced G>S hydroxymethylation on G719, G810, and G1103 sites of EGFR protein in a time-dependent manner without inducing any EGFR gene mutation corresponding to these GS-islands (Fig.2i, Extended Data Fig.2h). The pattern of EGF-induced EGFR G>S hydroxymethylation was a reminiscent of EGFR phosphorylation induction, which is a bell-shape pattern (Fig.2i, Extended Data Fig.2h). Stably introduced K280E and 5KR mutants of SHMT2 failed to induce EGFR protein hydroxymethylation on G719, G810, and G1103 in H1299 cells with SHMT2-knockout (SHMT2-KO) background (Fig.2j). Conversely, EGFR glycine hydroxymethylation was not detectable in the SHMT2-knockout background (Fig.2k). The co-localization of EGFR and SHMT2 proteins along the plasma membrane and cytoplasm, as observed in H1299 cells, further indicates a close relationship between them (Fig.2k).

Serine serves as the hydroxymethyl group donor.

SHMT was initially identified as facilitating serine and glycine metabolism in cells via PLP/THF-dependency\textsuperscript{20,28}. Therefore, we delineated how intracellular serine donates a hydroxymethyl group to the EGFR protein for glycine hydroxymethylation in cells. Parental H1299 cells, SHMT2-KO H299 cells, and increasing amount of SHMT2-Flag reintroduced SHMT2-KO H1299 cells were cultured in DMEM medium containing free isotope \textsuperscript{13}C-labeled serine. After 36 hours, EGFR proteins were isolated from these cells for
trypsinization-mass spectrometry analysis. We collected LC/mass and mass-mass spectra of the peptides covering G719 GS-island, which bears GSG (G719S720G721) sequence (Fig.3a,b). In addition to the wild type GSG peptide which was $^{13}$C-unlabeled, 4 sequence editing peptides i.e., GGG ($^{13}$C-unlabeled), $^{13}$C-SGG, $^{13}$C-SSG, and $^{13}$C-SS$^{13}$C-S in a SHMT2 expression dependent manner were detected (Fig.3a,b). The signals from $^{13}$C-SGG, $^{13}$C-SSG, and $^{13}$C-SS$^{13}$C-S peptides disappeared in SHMT2-KO H1299 cells but were restored in these SHMT2-KO H1299 cells with ectopic SHMT2 reintroduction, indicating that SHMT2 is responsible for G>S and/or S>G protein sequence editing induction. The recovery of $^{13}$C-SS$^{13}$C-S peptide suggests that G721 also underwent G>S protein sequence editing. The weak but persistent signal of GGG peptide was observed in all the cells, suggesting that a weak S720>G editing could occur spontaneously in cells independent of SHMT2 presence (Fig.3a, left).

In the LC-MS spectrum, the adduct ions of GSG and $^{13}$C-SGG peptide signals in the case of G719 peptide or the adduct ions of GS and $^{13}$C-SG peptide signals in the case of G810 peptide nearly overlapped (Fig.3a). However, peptide mass quantitative analysis revealed a 1-Dalton difference between the substrate GSG (1199 Daltons) and the product $^{13}$C-SGG (1200 Daltons) or between the substrate GS (1965 Daltons) and the product $^{13}$C-SG (1966 Daltons) (Fig.3a, right spectra associated with G719 and G810 peptides respectively). Mass-mass spectra confirmed all four product peptides of G719 GS island (GSG) occurred with G>S, S>G or GS>SG editing, i.e., GGG, $^{13}$C-SGG, $^{13}$C-SSG, and $^{13}$C-SS$^{13}$C-S as recovered by the LC-MS spectra (Fig.3c). All these 4 newborn peptides were differentiated as distinct peptides from parental GSG peptide in mass-mass spectra (Fig.3c). For the G>S hydroxymethylation reaction catalyzed by SHMT2 in cells, $^{13}$C-serine transfers its $^{13}$C-methylene group onto THF forming 5,10-methylene tetrahydrofolate, i.e., 5N,10N-$^{13}$C-CH2-THF (MTHF), which can then transfer its $^{13}$C-methylene in $^{13}$C-hydroxymethyl form onto the glycine residue of the EGFR protein followed with hydroxylation by H2O, as illustrated in Fig.3d. Even though S720 is common between G719 and G721, G719>S hydroxymethylation was somehow more frequent than G721>S hydroxymethylation to occur in H1299 cells with no GSS peptide was detected (Fig.3a, left), suggesting that G719 and G721 residues of G719 GS island differ in accepting methylene from MTHF.

We also analyzed G810 GS-island, which carries GS motif for G>S/GS>SG editing in detail. The LC-MS spectra of the EGFR G810-peptide (DNIG810SQYLL-NWCVQIAK) revealed the emergence of 3 new peaks, GG
([^{13}C]-unlabeled), [^{13}C]-SG, and [^{13}C]-SS, in addition to the ionic peaks of the parental GS of the G810 peptide in H1299 cells (Fig. 3a right, b). Like the emergence of ionic peak of the GGG signal in case of G719 peptide, the emergence of the GG signal was also independent of the presence of SHMT2. However, the GS>[^{13}C]-SG switch and the GS>[^{13}C]-SS editing signals were both SHMT2 dependent (Fig. 3a right, b). The mass-mass spectra confirmed the fingerprints of these three peptides. Detection of the [^{13}C]-S810S811-peptide or the [^{13}C]-S1103S1104-peptides (G>S single editing) and the [^{13}C]-S810G811-peptide or the [^{13}C]-S1103G1104-peptides (GS>SG switch editing) for the G810 and G1103 GS-islands respectively indicated that both the G810 and G1103 residues of the EGFR protein became [^{13}C]-S810 and [^{13}C]-S1103 upon receiving the [^{12}C]-hydroxymethyl group from free [^{13}C]-serine in cells. The G>S mono-editing and the GS>SG switch (i.e., double) editing occurred on the EGFR protein within almost all GS islands, as revealed in Fig. 1. The GS-islands accepted the [^{13}C]-hydroxymethyl group from free [^{13}C]-serine affording the G>[^{13}C]-S editing, meanwhile, the hydroxymethyl group of the serine residue of the GS-island could be removed to create the S>G editing in cells (Fig. 3a-c). In conclusion, a GS island can be switched into SG island in orientation by SHMT2 in cells.

To confirm the intrinsic catalytic activity of SHMT2 on GS island editing, we performed an in vitro assay. Bacterially purified GST-SHMT2 protein maintained its enzymatic activity and catalyzed a peptide with a 30.1-Daltons increase in mass, as confirmed by LC-MS and LC-MS/MS spectral analysis (Extended Data Fig. 3a-d). Product peptide fully overlapped with the synthetic S1103 peptide, indicating that the product peptide indeed had only a single G>S editing site (Extended Data Fig. 3c(ii),(iii) and d(ii),(iii)). Detection of G719>S, G810>S, and G1103>S in dot blot analysis was possible by incubating these synthetic peptides with purified rSHMT2 protein utilizing above antibodies (Fig. 3e). Both the dot blot and mass spectral results confirmed that the activity of the wild type rSHMT2 protein in G>S editing was markedly reduced by rSHMT2 with K280E or 5KR mutation introduction (Fig. 3e). We also performed the in vitro assay by including [^{13}C]-serine in the reaction buffer. Mass-mass spectral analysis revealed the recovery of [^{13}C]-SSG-peptide, [^{13}C]-SGG-peptide and [^{13}C]-SS[^{13}C]-S peptide with synthesized G719 peptide used as the substrate (Extended Data Fig. 3e,f), similar to the results obtained in the in vivo reaction of Fig. 3c. Again, for G719 peptide with G719SG721A sequence, G719 was more susceptible than G721 in G>S editing induction, even though [^{13}C]-SS[^{13}C]-S-peptide was detected (Extended Data Fig. 3e,f). The concentration of rSHMT2 protein was varied between 25-100 ng/ml.
In the SHMT2 catalyzed glycine-serine conversion reaction, rSHMT2 catalyzed 5N,10N-[\(^{13}\text{C}\)]-MTHF formation was detected in the in vitro reaction by incubating THF and [\(^{13}\text{C}\)]-serine with rSHMT2 (Fig.3f). ATP is required for many of those enzymes involved in glycine serine metabolism\(^{30}\). In our G>S/GS>SG editing in vitro assays, ATP was routinely added to reach an optimal reaction and ADP production from ATP consumption was accompanied with the wild type rSHMT2 protein but was reduced with rSHMT2 mutants (Fig.3g). We also calculated the catalytic values for G>S and S>G editing under the in vitro conditions. When G810 peptide was applied, the Kcat/Km for G810>S editing was 20 x 10\(^4\) and the Kcat/Km that for S811>G editing was 27 x 10\(^4\) (Fig.3h,i). Interestingly, S>G was slightly easier than G>S and GS>SG switch editing induction under the in vitro condition, suggesting removing hydroxymethyl group from serine may need lesser energy than adding the hydroxymethyl group onto glycine.

In addition to G>S, G>other residue (G>X) protein sequence changes in the lung cancer samples or cell lines were detected though with a much lower efficiency Fig.1a-c. To confirm G>X other than G>S changes, we committed both in vivo and in vitro assays to compare multiple isotope-labeled amino acids for glycine editing induction. While G>[\(^{13}\text{C}\)]-S was the most prominent one in SHMT2-dependnet manner under both in vivo and in vitro conditions, G>[\(^{13}\text{C}\)]-C, G>[\(^{13}\text{C}\)]-T as well as G>[\(^{13}\text{C}\)]-A were induced into decent levels but in an SHMT2-independent manner in cells (Fig.3j,k). The fact that SHMT2-depletion did not affect G>C, G>T, and G>A editing even indirectly (Fig.3j) simply indicates that enzyme(s) other than SHMT2 are responsible for their editing induction.

**Sidechain exchange between neighboring amino acid residues (SENA) editing between glycine and serine residues**

While the concentration of SHMT2 protein varied between 25-100 ng/ml, G>S single editing was predominantly induced using G719 peptide as the substrate in the in vitro reaction buffer containing [\(^{13}\text{C}\)]-serine (Extended Data Fig.3e). When SHMT2 protein was raised to 500 ng/ml and in the absence of free amino acid serine and THF in the reaction buffer, a switch editing (GS>SG) occurred on the substrate G810 or G1103 peptide (Extended Data Fig.3f), suggesting that the GS>SG switch editing could occur through a reaction of sidechain exchange between neighboring amino acid residues, referred as SENA editing here.

To further understand the potential implications of SENA editing, we constructed H1299 cells expressing EGFR-S720A or EGFR-A722S under an
EGFR-KO background. H1299 cells expressing EGFR-S720A were compared with H1299 cells expressing wild type EGFR in G719>S editing induction. The expression of EGFR-S720A led to a significant reduction in the EGFR protein G719>S editing (Fig.4a, Extended Data Fig.3g). In contrast, H1299 cells expressing EGFR-A722S greatly increased the EGFR protein G721>S editing rate (Fig.4b, Extended Data Fig.3g). These findings imply that the presence of a serine residue next to a glycine residue in protein could play a critical role in the induction of the GS>SG editing. To estimate the SENA editing reaction with EGFR protein, we synthesized S720-labeled G719 peptide (IKV[13C,15N]-SGAFGTVYK) and S811-labeled G810 peptide (YVREHKDN[13C,15N]-SQYLLNW)\(^3\). Since the 4C and 1N of the serine residue were all labeled, G[13C,15N]-S>[13C]-S[13C,15N]-G will be expected for detection providing SENA reaction occurs (Fig.4c). The results of 3 different doses of rSHMT2 were given in Fig.4d. When rSHMT2 concentrations from 25-100 ng/ml were applied, a newborn ionic peak of the GGG peptide emerged in addition to the ionic peak of the substrate G719 peptide (G[13C,15N]-SG), and a newborn peak of the GG peptide emerged in addition to the ionic peak of the substrate G810 peptide (G[13C,15N]-S) (Fig.4d). These GGG and GG peptides presumably originated from the G[13C,15N]-SG (G719) and G[13C,15N]-S (G810) substrate peptides via removing the [13C]-CH2OH away by rSHMT2 respectively. Thus, GGG here should be G[13C,15N]-GG and GG should be G[13C,15N]-G. As expected, there are 3 Daltons more for GSG peptide (1203 Daltons) or GGG peptide (1172 Daltons) appeared from SENA reaction in vitro of the spectra in Fig.4d than corresponding GSG peptide (1200 Daltons) or GGG peptide (1169 Daltons) detected in vivo of the mass spectra in Fig 3a.

Notably, the SENA editing reaction was readily induced by raising rSHMT2 concentration up to 500 ng/ml but, LC-MS or peptide mass quantitative analysis failed to distinguish between G[13C]-SG (substrate) and [13C]-SGG (product) or between G[13C]-S (substrate) and [13C]-SG (product) due to their identical molecular weights (Fig.4d). Whereas mass-mass spectra clearly confirmed that the peak (red color) under the condition of 500 ng/ml rSHMT2 in Fig.4d was a mixture of wild type G719-peptide (G[13C]-SG) and G719-peptide with SENA editing ([13C]-SGG) or a mixture of wild type G810-peptide (G[13C]-S) and G810-peptide (GS) with SENA editing ([13C]-SG) (Fig.4d,e). Given that SENA editing induction by SHMT2 was independent of THF and free amino acid serine, we calculated catalytic values for SHMT2 under the THF free condition\(^3\). Isotope labeled G719-peptide (IKV[13C]-SGAFGTYYK) was incubated with SHMT2 protein (500 ng/ml) under the in vitro condition lacking THF and free amino acid serine. The G719S720G>SGG editing (SENA-L) induction with
$K_{cat}/K_m$ 5.0 $\times 10^2$ was faster than the GS720G721$>$GGS editing (SENA-R) induction with $K_{cat}/K_m$ 2.6 $\times 10^2$ (Fig.4f,g), suggesting that the GS$>$SG (SENA-L) was more favored than the SG$>$GS (SENA-R) in the editing reaction. With G810-peptide (YVREHKDNIG$_{810}$$^{[13]C}$-SQYLLNW), we compared with and without THF for GS$>$SG switch editing induction by SHMT2. GS$>$SG editing in SENA was de facto restricted by THF for $K_{cat}/K_m$ reduced from $1.8 \times 10^4$ without THF to $0.45 \times 10^4$ with THF (Fig. 4f,g).

For EGFR G1054 GS island of human origin, G1054 was two residues away from S1057, as illustrated in Fig.1a. To examine the steric effect on SENA editing, we prepared the G1054 peptide with S1057 to be labeled with isotope (CIDRNG$_{1054}$LQ$^{[13]C}$-SCPIK) for the in vitro analysis. G1054 was significantly weaker in SENA editing mutation induction compared to G719 and G810, indicating that the distance between G and S might introduce a crucial steric impediment to an efficient GS mutation (Fig.4h). Both G695 and G721 reside within the TK domain and are the only two SG islands within the EGFR protein cytoplasmic region. Like the case of S720G721, S694G695-peptide with S694 to be isotope labeled (ELVEPLTP$^{[13]C}$-SG$_{695}$EAPNQALLR) responded poorly to rSHMT2 protein in SG$>$GS editing induction (Fig.4h). To confirm whether SENA editing in vitro was cis or trans, we mixed $^{[13]C}$-serine peptide with unlabeled serine peptide equally and incubated them with SHMT2 in the reaction buffer. Intriguingly, in the absence of THF, SHMT2 only catalyzed a cis reaction. Conversely, SHMT2 catalyzed both cis and trans reactions when THF was present in the reaction buffer (Fig.4i). Together, these results strongly support the conclusion that while SHMT2 switches serine residue hydroxymethyl group onto the adjacent upstream glycine residue in a cis SENA editing reaction induction, the steric orientation of G and S residues in the GS-islands are crucial for SENA editing to take place.

**SHMT2 promotes EGFR signaling activity via GS-island editing.**

Activating mutations in the EGFR gene enhance EGFR's inherent tyrosine kinase activity and the associated signaling activity. Notably, EGFR-Y1068 autophosphorylation and ERK1/2 activation following EGF treatment were amplified in cells co-transfected with EGFR and SHMT2 (Extended Data Fig.4a). Stable SHMT2 knockout colonies (KO1 and KO2) were created in both H1299 and A549 cell lines using the CRISPR-CAS9 method (Extended Data Fig.4b). EGF-stimulated EGFR autophosphorylation and ERK1/2 activation were significantly suppressed in these SHMT2-KO lung cancer cells (Extended Data Fig.4c), aligning with the inhibition of GS island mutation induction (Fig.2k). We reintroduced wild type SHMT2 and its mutants (K280E and 5KR) into SHMT2-
KO H1299 cells to establish stable colonies. In contrast to SHMT2-KO, overexpression of SHMT2 augmented EGFR autophosphorylation and ERK1/2 activation by EGF (Fig.5a,b). However, K280E or 5KR mutations hampered SHMT2 activity in this regard (Fig.5a,b).

SHMT inhibitor1 (SHIN1) is a specific inhibitor for the catalytic activity of SHMT1/2 in metabolic reactions via an enantiomer enzyme interaction\(^33\). SHIN1 (5 \(\mu\)M) pretreatment predictably inhibited EGF-induced EGFR autophosphorylation and ERK1/2 activation in these lung cancer cells (Extended Data Fig.4d). It is worth to note that SHIN1 treatment reduced SHMT2 protein level in cells (Extended Data Fig.4d). SHIN1 treatment significantly inhibited EGFR autophosphorylation and ERK1/2 activation in H1299 and A549 cells, as well as in SHMT2-overexpressing HEK293T cells via G>S editing inhibition, as evidenced by Western blot and fluorescent immunostaining analysis (Fig.5c,d, Extended Data Fig.4e-g). A direct inhibition of SHIN1 on SHMT2 was further validated in an in vitro assay, analyzed with LC-MS spectrometry and in dot blot (Fig.5e,f). SHIN1 significantly inhibited rSHMT2 protein in catalyzing G1103 peptide glycine hydroxymethylation in vitro in a dose-dependent manner (0.01 - 10 \(\mu\)M) (Fig.5e). As a result, the EGFR protein recruits cytosolic SHMT2 protein to enhance EGFR signaling, presumably through the induction of EGFR protein glycine hydroxymethylation by SHMT2. Although EGFR is a tyrosine kinase relying on tyrosine phosphorylation for signaling, EGFR serine/threonine phosphorylation has also been widely noted to play a critical role in EGFR signaling\(^7\). The G1103>S site, hydroxymethylated by SHMT2, could be phosphorylated in HEK293T cells with co-expression of EGFR and SHMT2 (Fig.5g). EGF treatment induced EGFR S1103 phosphorylation in H1299 cells (Fig.5h), but not in H1299 cells with SHMT2 knockout or in H1299 cells pre-treated with SHIN1 (Fig.5h, Extended Data Fig.4e,g). Together, these results suggest that SHMT2 plays a positive effect on EGFR signaling presumably via GS-island hydroxymethylation induction.

Since SHMT2's known role in regulating glycine and serine metabolism, as well as its potential involvement in mitochondrial translation initiation and tRNA formylation\(^34,35\), we sought to analyze the impact of SHMT2 variants on cellular metabolism. Previous studies have shown that reactive oxygen species (ROS) and \(\text{NAD}^+\)/\(\text{NADH}\) are involved in the SHMT-catalyzed anabolic metabolism pathway\(^18,23,36\). It was found that depletion of SHMT2 had a moderate effect on cytosolic ROS production in lung cancer A549 and H1299 cells (Extended Data Fig.4i,h). However, when comparing ROS levels in the cytoplasm and mitochondria of H1299 cells, mitochondrial ROS levels were more severely
affected (Extended Data Fig.4i). Similarly, depletion of SHMT2 had a moderate effect on cytosolic NAD$^+$ and NADH levels in both lung cancer cells tested (Extended Data Fig.4j,k). Although in EGF treated H1299 cells expressing SHMT2 wild type with a moderate effect on NAD$^+/\text{NADH}$ level alteration, mitochondrial membrane potential as indicated by the cationic carbocyanine dye JC-1 intensity was elevated in the cells expressing SHMT2 (Extended Data Fig.4k,l)$^{27}$. SHMT2 with K280E or 5KR mutation significantly reduced its activity in JC-1 intensity induction (Extended Data Fig.4l).

K280E and 5KR mutants of SHMT2 were compared with wild type SHMT2 with respect to their ability to catalyze the SHMT reaction in G$>\text{S}$ and S$>\text{G}$ conversion in vitro. The time course revealed a liner effect of rSHMT2 on G$>\text{S}$ and S$>\text{G}$ conversion, which was reduced with both SHMT2 K280E and SHMT2 5KR mutants (Extended Data Fig.4m,n). However, SHMT2 catalyzed G$>\text{S}$ conversion was much stronger than S$>\text{G}$ conversion (Extended Data Fig.4m,n), suggesting SHMT2 favors on glycine hydroxymethylation than serine dehydroxymethylation reaction. When cellular amino acid levels were compared in H1299 cells, it was found that glycine and serine levels were not significantly affected in the cells with or without SHMT2 expression (Extended Data Fig.4o). A moderate elevation of serine level in SHMT2-KO H1299 cells here could be due to reduced serine serving as the donor of the hydroxymethyl group to glycine residues of proteins like EGFR.

While there are multiple enzymes in cells involved in de novo serine synthesis and serine metabolism$^{37,38}$, our results strongly support the idea that the role of SHMT2 in cytosolic metabolism is relatively moderate and may be more strongly linked to EGFR signaling activation, especially in those cells highly active in proliferation$^{34,39}$. Indeed, overexpression of EGFR gene mutants (G$>\text{S}$ or GS$>\text{SG}$) alone in H1299 SHMT2-KO cells (Fig.5i,j) behaved quite similarly to EGFR and SHMT2 co-overexpression in terms of EGFR tyrosine auto-phosphorylation and/or ERK1/2 activation elevation (Fig.5i,j). Moreover, in cells expressing EGFR cDNA with GS$>\text{SG}$ double mutation of G810S811 and G1103S1104 were more active than cells expressing G$>\text{S}$ point mutation of G810 and G1103 in ERK1/2 activation (Fig.5i,j). Thus, expression of cDNA of EGFR gene mutants for these specific GS islands presumably mimics EGFR protein with constitutive G$>\text{S}$/GS$>\text{SG}$ editing for EGFR signaling activation.

EGFR protein GS island editing promotes cancer cell proliferation and TKI drug-resistance.

We then delve into the tumorigenic potential of cytosolic SHMT2 in a mouse
xenograft model. In both A549 and H1299 lung cancer cells, SHMT1 inhibition caused apoptosis. Remarkably, SHMT2 gene knockout A549 cells significantly reduced subcutaneous tumor formation in size and weight in mice (Fig.6a,b). Correspondingly, a decline in cell growth rate was observed in SHMT2 knockout A549 cells, as indicated by Ki67 staining, as compared to parental A549 cells expressing wild-type SHMT2 (Extended Data Fig.5a). Colony formation ability was notably diminished in SHMT2 knockout backgrounds for both H1299 and A549 cell lines (Fig.6b, Extended Data Fig.5b). In H1299 cells with SHMT2-KO background, stably reintroduced SHMT2 wild type increased cell growth rate and tumor formation ability (Fig.6c,d), whereas K280E and 5KR mutants of SHMT2 reduced such activity (Fig.6c,d).

We subsequently evaluated the EGFR gene mutations corresponding to these GS islands in tumor formation activity. The weight of subcutaneous tumors in mice implanted with A549 cells expressing EGFR-G719S, G810S, and G1103S were compared with A549 cells expressing EGFR-G719SS720G, EGFR-G810SS811G, and EGFR-G1103SS1104G double mutations. Tumor sizes were significantly greater with the cells expressing GS>SG double mutations than those cells expressing G>S point mutations (Fig.6e). In A549 cells, overexpression of EGFR-G1103S cDNA enhanced cellular viability and resistance to apoptosis compared to cells expressing wild-type EGFR, presumably due to the stimulatory effect of the EGFR-G1103S gene product on cellular proliferation (Extended Data Fig.5c). Both SHMT2α and SHMT2 FL promoted cell cycle progression in H1299 cells (Extended Data Fig.5d, left), presumably due to their catalytic activity in EGFR protein G>S editing induction (Extended Data Fig.5d, right). These results were consistent with the findings that EGFR gene mutation of GS island stimulate cell proliferation (Extended Data Fig.5e). In addition to the widespread activation of the MAPK pathway, the EGF-induced activation of the STAT pathway is also critical for cancer cell growth regulation. Luciferase reporter activity assay in HEK293T cells revealed that only wild-type SHMT2, not the SHMT2 K280E or 5KR variants, could significantly augment STAT3 transcriptional activity in response to EGF treatment (Extended Data Fig.5f). Transient overexpression of EGFR gene variants with G>S or GS>SG switch mutations at G810 and G1103 GS islands increased STAT3 activation (Extended Data Fig.5g). These findings further substantiate our conclusion that SHMT2-catalyzed hydroxymethylation leading to EGFR protein G>S or GS>SG editing is akin to an activating mutation of the GS-islands of EGFR gene in terms of EGFR signaling in cancer.

Gefitinib is employed as a targeted therapy for NSCLC patients with EGFR gene activating mutations, including the EGFR-G719S mutation. Given that
cells expressing either wild-type EGFR or EGFR with activating mutations can eventually develop drug resistance, we analyzed both gefitinib resistant (GR) H1299 cells that carry no EGFR gene mutation and the well-known HCC827 cells that carry an exon 19 (E746-A750) deletion mutation (Fig.6f-m). In H1299-GR cells, EGFR G>S editing became constitute on G719, G810 and G1103 sites correlating with an elevated SHMT2 expression in these cells (Fig.6f,g). H1299 cells were treated with the escalating amounts of gefitinib up to 10 µM and the IC-50 was reached at 2.5 µM (Fig.6h). Given that SHIN1 demonstrated a dose-dependent inhibition of colony formation in these lung cancer cell lines (Extended Data Fig.6a,b), we combined the IC-50 dose of gefitinib (2.5 µM) with increasing amounts of SHIN1 (2.5 - 10 µM) in H1299-GR cells. The IC50 of gefitinib shifted as apoptosis was increasingly induced by the combination of gefitinib and escalating amounts of SHIN1 (Fig.6i), whereas H1299-GR cells displayed high resistance to gefitinib-induced apoptosis (Fig.6h).

The gefitinib-resistant HCC827-GR cells that were evolved from HCC827 lung cancer cells, displayed similar trends. EGFR GS island G>S editing coupled with increased SHMT2 expression levels were observed in HCC827-DR cells as well (Fig.6j). Up to 20 µM, SHIN1 showed a linear but weak inhibitory effect on HCC827-GR cell viability (Fig.6k). Comparing with parental HCC827 cells, HCC827-GR cells exhibited significant resistance to apoptosis induction by gefitinib up to 10 µM (Fig.6l), SHIN1 at 5 µM or 10 µM enhanced the apoptosis-inducing capacity of gefitinib in HCC827-GR cells (Fig.6m). As anticipated, SHMT2-KO H1299 cells expressing SHMT2-K280E or SHMT2-5KR showed decreased sensitivity to gefitinib compared to SHMT2-KO H1299 cells reintroduced with wild-type SHMT2 for expression, in terms of both cell viability and STAT3 activation (Extended Data Fig.6c,d). Hence, SHMT2-mediated G>S editing might serve as an allosteric gene mutation factor contributing to drug resistance.

In lung adenocarcinoma tissue samples from NSCLC patients, immunostaining with antibodies that recognize G719>S, G810>S and G1103>S editing revealed that the EGFR protein carried hydroxymethylation-dependent G>S editing on these GS islands constitutively, even though no EGFR gene mutations were detected for these GS islands (Fig.6n-p). A strong correlation between constitutive EGFR protein G>S editing on these GS islands and elevated SHMT2 expression levels in these lung adenocarcinoma tissues compared to normal tissues (Fig.6o). The widespread overexpression of SHMT2 has been observed in various types of cancer, and an overall higher level of SHMT2 expression has been seen in lung squamous cell carcinoma.
compared to lung adenocarcinoma, according to the KEGG database (Extended Data Fig.6e-g)\textsuperscript{17}. This might elucidate why EGFR-targeted therapy with TKIs is less effective in lung squamous carcinoma than in lung adenocarcinoma\textsuperscript{48}. Our results suggest that SHMT2-mediated hydroxy-methylation of EGFR GS islands fosters lung cancer cell growth and aids the development of TKI resistance in these lung cancer cells.
Conclusion

Gene mutations, as well as similar amino acid substitution mutations that induce protein sequence alterations, generally arise due to errors at the gene replication/transcription or mRNA translation level\textsuperscript{3,49}. However, hydroxymethylation mediated G>S/GS>SG for protein sequence editing of the EGFR protein occur independently of \textit{EGFR} gene mutations or translational errors (Fig.6q). Comparable to tyrosine or serine phosphorylation, G>S or GS>SG residue change in orientation created from protein sequence editing is essential for EGFR to adopt an optimal conformation for signaling. Glycine hydroxymethylation, leading to G>S/GS>SG editing, allows a protein sequence to be edited post-translationally and therefore to provide a convenient means for signaling regulation in cytoplasm without bona fide gene mutation occurring in nuclei. Protein sequence editing simply indicates that residues of a protein can not only be post-translationally modified like phosphorylation and acetylation, but also can be edited in sequence. Such protein sequence editing can be push forward via accepting the sidechain groups from free amino acids or other free compounds in cell but also can also be proceeded via exchange between two adjacent residues of the protein. G>S/GS>SG editing changes protein sequence, which can be either reversed back or push forward to another amino acid residue.

Among the amino acid repertoire, glycine is the smallest and one of the most frequently utilized residues during translation, offering a keel for protein construction\textsuperscript{50}. There are some uncommon post-translational modifications such as myristoylation have been reported for glycine residues of protein\textsuperscript{51,52}. However, amino acid glycine at large has long been noted for its conversion into serine as well as other amino acids actively. We believe that glycine residues in protein can be converted into other translatable residues in addition to serine residue as we noted here including cysteine, threonine, and alanine by unknown mechanisms. An enzyme of anabolic metabolism such as SHMT2 here can be recruited by EGFR or other cytosolic proteins for their protein sequence editing or functional modification regulation\textsuperscript{18,23,28}, other amino acid metabolic enzymes should not be ruled out for such purpose. Moreover, the hydroxymethyl group, cleaved by SHMT2 from the serine residue of a protein can serve as a source of one-carbon unit for metabolism as well. Consequently, cellular transmembrane proteins like EGFR and perhaps other cytokine receptors are subject to a cellular environment favoring reprogramming, leading to dynamic alterations in their sequences\textsuperscript{21}. Therefore, protein sequence editing challenges the central dogma, which posits that a protein sequence faithfully reflects its gene sequence\textsuperscript{53}.  

Although the comparative distribution of protein editing in cells remains undefined, the dynamic protein G>S/ GS>SG editing via hydroxymethylation stands for a biochemical reaction in cells in response to changes such as EGF in the cellular microenvironment. It adds a new layer to the landscape of post-translational modifications and gene mutation-based protein sequence changes observed in both normal and cancer cells, as a significant enrichment of G>S/ GS>SG editing has been noted in cancer cells. Therefore, protein sequence editing is likely to play crucial roles in cellular functions and cancer development and may provide common therapeutic targets across different tumor types.
Legends
Fig. 1: EGFR protein carries EGFR gene-mutation independent G>S and GS>SG editing in lung cancer.

a. EGFR proteins were immunoprecipitated with a monoclonal antibody against EGFR from the whole cell lysates (WCL) of lung cancer tissues obtained from NSCLC patients. The immunoprecipitated EGFR proteins were trypsinized and analyzed with LC-MS/MS (AB Sciex TripleTOF 6600, Bruker TimsTof Pro, or Thermo Scientific Orbitrap Exploris 480). The heat map shows Log2 differences in signal intensity of G>S and/or GS>SG protein sequence changes appearance in EGFR protein. This heat map is based on lung cancer tissues of 45 NSCLC patients and 7 lung cancer cell lines (H1299, A549, 1975, H-522, Cula-3, H-520, and Beas-B2). We recovered peptides of EGFR protein with G>S protein sequence changes. Our trypsin digestion-mass spectrometry approach routinely covered 50-80% of the EGFR protein sequence.

b. EGFR protein changes and GS Islands. The full-length EGFR protein contains 85 glycine residues, out of which 35 are sporadically distributed within the cytoplasmic domain. Using mass spectrometry analysis, G>S protein sequence changes were detected in lung cancer tissues and cell lines, with G719, G810, and G1103 sites having the highest frequency of G>S and/or GS>SG protein sequence changes. The cytoplasmic domain of EGFR has 7 "GS islands" that align with the human and mouse origins.

c. Heat map and Venn diagram analysis. A heat map was generated to depict the efficiency of G>X, S>X, T>X, or L>X detected in lung cancer tissues and cell lines of the samples analyzed using mass spectrometry analysis. Peptides were detected in two replicates (n=2) of every sample. Venn diagram analysis of 52 samples in a with G>S, GS>SG, and S>G protein sequence changes of these GS-islands to reveal that 10 samples bear all three types of protein sequence changes, 33 samples bear GS>SG and G>S, 1 sample bears G>S and S>G whereas no sample bears GS>SG and S>G.

d. The representative LC-MS/MS spectra of GS island G>S and GS>SG protein sequence changes. Peptides with G>S single protein sequence changes (G719>S, G810>S, G1103>S) and GS>SG switch protein sequence changes (G719S720>SG, G810S811>SG, and G1103S1104>SG) of EGFR protein were detected. EGFR proteins were immunoprecipitated from lung cancer tissues of NSCLC patients, and AB Sciex TripleTOF 6600 mass spectrometry was used for analysis. The y and b series of ions were indicated along the sequence of the peptide derived, which was shown at the top of the panel.

e. EGFR gene sequencing. Genome DNA and mRNA extracted from the lung cancer tissues and cell lines used above were sequenced to determine whether there were EGFR mutations with fragment spans of G719S720,
G810S811, and G1103S1104 islands. The sequencing revealed that EGFR DNA or mRNA carried no mutations in the samples analyzed.

Fig. 2| EGFR recruits SHMT2 for GS island G>S and GS>SG editing induction.

a. EGFR proteins were immunoprecipitated from the H1299 and A549 cells, and mass spectrometry was performed to identify EGFR and binding proteins. As marked, such immunoprecipitated EGFR protein and binding proteins were separated in silver-staining gel.

b. SHMT2 was identified as an EGFR binding protein from the mass spectrometry analysis.

c. Coimmunoprecipitation analysis confirmed the EGFR and SHMT2 association in the H1299 and A549 cells, respectively.

d. Purified eGFP-EGFR protein and SHMT2 (WT, K280E, or 5KR variant) proteins were used for a microscale thermophoresis binding assay. The *Kd* of eGFP-EGFR and SHMT2 association were calculated using NanoTemper software. The *Kd* for EGFR and SHMT2-K280E mutant association was undetectable.

e. Secondary structural alignment analysis between human SHMT1 and SHMT2 showed that K280 is conserved between the two while N-terminal 5Ks (K459, K461, K464, K469, and K474) are specific to SHMT2.

f. In HEK293T cells, EGFR was transfected with empty vector (EV) or along with SHMT2. EGFR proteins, immunoprecipitated from the transfectants were analyzed with LC-MS/MS (AB6600). The LC-MS/MS spectra of G719 island were shown. While wild type G719 island-peptides were obtained from the cells transfected with EGFR alone, G719>S and G719>SS720>G mutated peptides were recovered from HEK293T cells transfected with EGFR and SHMT2 together.

g. DNA sequencing result of EGFR G719 island motif of above samples in f.

h. SHMT2 variants (K280E or 5KR) were transiently transfected to compare with EGFR WT on EGFR G>S or GS>SG mutation induction as in f. LC-MS/MS analysis revealed that SHMT2 with K280E or 5KR mutation failed to induce G>S and GS>SG mutations within G719, G810, and G1103 GS islands.

i. A time course of EGFR protein glycine hydroxymethylation G>S was induced by EGF in H1299 cells. Whole cell lysates (WCL) were then prepared and separated by 10% SDS-PAGE and subjected to immunoblotting analysis with antibodies recognizing EGFR proteins with G810>S and G1103>S editing respectively.
j. SHMT2-KO H1299 cells were generated using the CRSPRA/CAS9 approach. In H1299 cells with SHMT2-KO background, EV, SHMT2 variants as indicated were stably introduced. Equal amounts of whole cell lysates were separated by 10% SDS-PAGE and subjected to immunoblotting analysis with antibodies recognizing EGFR proteins with G719>S, G810>S and G1103>S respectively. EGFR, Flag-SHMT2, and GAPDH expression levels in blot were included.

k. EGFR proteins were immuno-stained with an antibody recognizing G719>S, G810>S, and G1103>S in both parental (WT) and SHMT2-KO H1299 cells. SHMT2 immuno-staining and DAPI staining were included in the experiment, and the cells were visualized using fluorescent confocal microscopy. The scale bar was set at 10 μm.

Fig. 3I Using free amino acid serine as hydroxymethyl doner for G>S and S>G editing induction.

a. Parental H1299 cells expressing wild type SHMT2 (WT), H1299 cells with SHMT2-KO, and SHMT2-KO H1299 cells with SHMT2-Flag reintroduction in different doses (1 ug and 2 ug) were grown in DMEM medium containing free [13C]-serine for 24 hours. The LC-MS spectrometry (adduct ionic peaks) results (AB6600) showed that both the G719 GS-island sequence (VLG719S AF GTV YK) and G810 GS-island sequence (DIG810 SQY LLN WCV QIA K) of EGFR protein were detected in all four groups of the cells. For G719 GS-island, while the wild type GSG-peptide and GGG-peptide ionic peaks were recovered from the cells under all 4 conditions, the ionic peaks of [13C]-SGG-peptide, [13C]-SSG-peptide, and [13C]-SSS-peptide were recovered only from the cells expressing SHMT2. For G810 GS-island, wild type GS-peptide and GG-peptide ionic peaks were recovered from all the cell types. However, [13C]-SG-peptide and [13C]-SS-peptide were recovered only from cells expressing SHMT2. The mass spectrum for peptide quantitation was provided for both G719 GS island and G810 GS-island for the cells expressing SHMT2-Flag ++ (right, next to the adduct ionic peak spectra).

b. Signal intensities of the results obtained in a.

c. EGFR proteins were immunoprecipitated from H1299 cells growing in the DMEM medium containing free [13C]-serine. Four peptides with different G>S editing mutations were recovered from the EGFR proteins, and their mass-mass spectra were analyzed. The G719 motif (VLG719S AF GTV YK) included two mono-editing residues: G719G720G721 and [13C]-S719S720G721
(“GGG” and “SSG”) and two double-editing residues: $[^{13}\text{C}]-\text{S}_{719}\text{G}_{720}\text{G}_{721}$ and $[^{13}\text{C}]-\text{S}_{719}\text{S}_{720}[^{13}\text{C}]-\text{S}_{721}$ (“SGG” and “SSS”).

d. It depicts how SHMT2 catalyzed hydroxymethylation reaction, with the $[^{13}\text{C}]-\text{hydroxymethyl}$ group of free $[^{13}\text{C},^{15}\text{N}]-\text{serine}$ being transferred onto the protein’s glycine residue as a side chain, resulting in G>$[^{13}\text{C}]-\text{S}$ editing.

e. Dot blot analysis of the catalytic activity of different forms of rSHMT2 as indicated on G>S editing induction in vitro by using G719-peptide, G810-peptide, and G1103-peptide as substrates (left panel). The LC-MS spectra of these in vitro reaction products were included (right panel).

f. The 5N,10N-$[^{13}\text{C}]-\text{CH}2-\text{THF}$ formation was detected with LC-MS analysis of the in vitro reaction with $[^{13}\text{C}]-\text{serine}$ catalyzed by rSHMT2 at different doses as indicated. THF was included in the reaction buffer. The mass spectrum of 5N,10N-$[^{13}\text{C}]-\text{CH}2-\text{THF}$ quantitation was given in the right panel.

g. The ADP level was detected with LC-MS analysis of the reaction. LC-MS spectrum of the G>S editing reaction catalyzed by SHMT2 were analyzed in the presence of PLP, ATP, and 5N,10N-MTHF.

h. The LC-MS spectrum of the G>S editing with the G810 peptide catalyzed by rSHMT2 in vitro assay. Peptides with G810>S point editing and S811>G point editing were recovered.

i. The above catalytic reaction in i based upon the G810-peptide at different concentrations (50, 100, 150, 200, and 250 mM), initiated by rSHMT2 protein in the presence of 10 mM THF, 40 mM PLP, 100 mM ATP, 5 mM MnCl$_2$ and 5 mM KCl in the in vitro reaction buffer (PBS 1 ml, pH7.2), 37°C, 15 minutes. The kinetic parameters $K_{cat}$ and $K_m$ were then calculated.

j. H1299 cells grew in the DMEM medium containing isotope labeled amino acids ($[^{13}\text{C}]-\text{Ser}$, $[^{13}\text{C}]-\text{Ala}$, $[^{13}\text{C}]-\text{Gln}$, $[^{13}\text{C}]-\text{Phe}$, $[^{13}\text{C}]-\text{Thr}$, and $[^{13}\text{C}]-\text{Cys}$) for 24 hours. Immunoprecipitated EGFR proteins from these cells were then subjected to LC-MS for analysis, mean ± SD, n = 3.

k. G719 peptide was incubated with rSHMT2 in the presence of isotope labeled amino acid ($[^{13}\text{C}]-\text{Ser}$, $[^{13}\text{C}]-\text{Ala}$, $[^{13}\text{C}]-\text{Thr}$, or $[^{13}\text{C}]-\text{Cys}$) in the reaction buffer for 1 hour in the in vitro assay. The peptides were then subjected to LC-MS for analysis, mean ± SD, n = 3.

Fig. 4| SENA editing in vitro.

a. Wild type or S720A-mutated EGFR cDNA was introduced for expression in EGFR-KO background H1299 cells. Immunoprecipitated EGFR proteins were subjected to trypsin digestion and mass-mass spectral analysis of EGFR protein G719>S editing efficacy. The intensity of EGFR protein
b. EGFR-KO background H1299 cells were stably introduced with wild type EGFR or A722S-mutated EGFR for expression. Immunoprecipitated EGFR proteins were subjected to trypsin digestion and LC-MS/MS spectral analysis of G721>S editing efficacy. The intensity of EGFR protein G721>S editing signal from LC-MS spectra was plotted.

c. SENA editing denotes the SHMT2 catalyzed hydroxymethylation for GS>SG switch editing reaction. The [13C]-hydroxymethyl group of [13C, 15N]-serine residue of the GS island can be transferred onto the glycine residue adjacent to it, resulting in G>[13C]-S switch editing.

d. Isotope-labeled peptides G719[13C]-S720G721 peptide (upper panel) and G810[13C]-S811 peptide (lower panel) were incubated with rSHMT2 protein at indicated concentrations, and LC-MS spectrometry analysis was performed. For G719 peptide, GGG peptide (i.e., G719G720G721-peptide) in upper panel and for G810-peptide, GG peptide (G810G811-peptide) in lower panel were recovered by incubating the isotope-labeled peptide with rSHMT2 protein. The peak in red color was a mixture of GS-peptide and SG peptide in both G719 and G810 peptides when 500 ng rSHMT2 protein was used, as revealed by the quantitative mass analysis included on the right.

e. LC-MS/MS spectra showed that rSHMT2 at a high dose (500 ng/ml) catalyzed GS>SG switch editing mutation. When incubated with rSHMT2 protein at 500 ng/ml for 30 min at 37°C, [13C]-S719G720G721 peptide was generated from G719[13C]-S720G721 peptide (VLG719SGAFGTGYK) or 13C-S810G811 peptide was generated from G810[13C]-S811-peptide (YVREHKDNIG810SQYLLNW). The serine residue's [13C]-hydroxymethyl group within the GS island can be transferred onto the adjacent glycine residue as the side chain, leading to SENA GS>SG editing mutation.

f. The protein editing reaction was based on the G719-peptide at different concentrations (100, 200, 300, 400, and 500 mM), initiated by rSHMT2 protein in the presence of 40 mM PLP, 100 nM ATP, 5 mM MnCl2 and 5 mM KCl in the in vitro reaction buffer (PBS 1 ml, pH7.2) at 37°C for 15 minutes. THF was not included. For G719SG>GSG and GSG722>GGS editing catalyzed by rSHMT2, the kinetic parameters (Kcat and Km) of SHMT2 on G719-peptide SENA-L (GSG>GSG) and SENA-R (GSG>GGS) were determined (upper panels). For G810 peptide (50, 100, 150, 200, and 250 mM) initiated by rSHMT2 protein in the above reaction conditions. The kinetic parameters (Kcat and Km) of SHMT2 were compared with and without THF (10 mM) in the editing mutation induction s were compared for their Kcat and Km (left panel). LC-MS/MS of SGG and GGS were included.
g. LC-MS/MS spectra of SGG and GGS peptides recovered from G719-peptide (left) and SG peptide recovered from G810-peptide catalyzed by rSHMT2 in vitro for $K_{cat}$ and $K_m$ calculation in f.

h. The synthesized G1054 peptide (CIDRNLQ$^{[13]C}$-SCPIK), G694 peptide (ELVEPLTP$^{[13]C}$-SGEPNQALLR), G719-peptide (IKVLG$^{[13]C}$-SGAFGTVVYK), and G810-peptide (YVREHDIN$^{[13]C}$-SQYLLNW) were compared for their ability to induce SENA-L editing by rSHMT2 in LC-MS spectra.

i. An equal amount of the isotope-labeled G719-peptide ($[^{13}C]$:S720) was mixed with the unlabeled G719-peptide (S720) and incubated with rSHMT2 under in vitro reaction conditions in the presence or absence of THF. Four types of peptides with G$>$S editing were recovered from the LC-MS spectra.

Fig. 5] Glycine hydroxymethylation dependent GS island G$>$S editing are activating one for EGFR signaling.

a. SHMT2-KO H1299 cells were stably transfected with cDNA of EV and SHMT2 variants (WT, K280E, and 5KR). After EGF treatment for 30 minutes, whole-cell lysates were collected for analysis of EGFR tyrosine autophosphorylation (pY1068) and ERK1/2 activation (pT202/pY204) via Western blotting. EGFR, ERK1/2, SHMT2, and GAPDH levels were blotted as indicated.

b. The signal intensities of EGFR-pY1068 (upper) and ERK1/2-pT202/pY204 (lower) of the samples in a were quantitated, mean ± SD, n = 4.

c. H1299 cells were pretreated with 5 mM SHIN1 for 24 hours, followed by EGF treatment for an additional 30 minutes. EGFR phosphorylation and ERK1/2 activation were analyzed via Western blotting with anti-EGFR-pY1068 and anti-ERK1/2-pT202/pY204, respectively. EGFR, ERK1/2, SHMT2, and GAPDH levels were blotted as indicated.

d. EV or SHMT2 cDNA was transiently overexpressed in H1299 cells. EGFR G1103$>$S editing peptides were recovered from mass spectrometric analysis of these cells and SHMT2-overexpressing H1299 cells treated with 10 mM SHIN1 for 24 hours. The relative signal intensities of EGFR G1103$>$S editing detected in LC-MS spectra were quantitated, mean ± SD, n = 3.

e. LC-MS spectrometry analysis was performed to determine the inhibitory effect of SHIN1 on glycine hydroxymethylation in vitro. G1103-peptide was incubated with purified rSHMT2 protein in an in vitro reaction buffer, and DMSO (CTRL) or SHIN1 (0 - 10 mM) in DMSO was included. The ionic
peak signal intensities obtained from the mass results of G1103>S conversion were plotted, mean ± SD, n = 3.

f. The inhibitory effect of SHIN1 on hydroxymethylation in vitro was evaluated via dot blot assay. G719-peptide and 1103-peptide were incubated with or without GST and rSHMT2 proteins in an in vitro reaction buffer, and DMSO (-) or 5 mM SHIN1 in DMSO was included.

g. EGFR proteins were immunoprecipitated from EGF-treated HEK293T cells with EGFR overexpression, trypsinized, and analyzed with AB6600 mass spectrometry. Phospho-G1103>S peptide of EGFR protein (G1103>pS) was detected.

h. EGFR phosphorylation on G1103>S was detected via immunostaining with anti-EGFR-G1103>pS in SHMT2-KO H1299 cells and parental H1299 cells treated with EGF or left untreated.

i. EGFR was knocked out in H1299 cells via CRSPRA/CAS9 approach. EV or EGFR variants (WT, G810>S, G1103>S, G810S811>SG, G1103S1104>SG) were stably transfected into these H1299 cells with EGFR-KO background. ERK1/2 activation (pT202/pY204) in response to EGF treatment (100 ng/mL, 30 minutes) was analyzed via Western blotting. GAPDH was used as a loading control.

j. The signal intensities of ERK1/2-pT202/pY204 of the samples in i were quantitated, mean ± SD, n = 4.

Fig. 6I EGFR protein GS island G>S and GS>SG editing develop EGFR targeting therapy resistance.

a. In a xenograft mouse model, the tumor formation ability of A549 cells with SHMT2 knockout (KO-1 and KO-2) was compared to parental A549 cells (WT) by implanting them subcutaneously for tumor formation in nude mice for 41 days. The xenograft tumors were then pictured and compared. The result showed that SHMT2 knockout had a significant effect on tumor cell growth, n = 6.

b. The xenograft tumors in a were weighed, and the result showed that the tumors with SHMT2 knockout were significantly smaller compared to the control group, n = 6.

c. In H1299 cells with EGFR-KO background, EV, SHMT2-WT, SHMT2-K280E, or SHMT2-5KR mutant was reintroduced stably. EDU staining was performed, and EDU positive cells were quantitated in percentage, mean ± SD, n = 6.

d. The above cell lines in c were implanted subcutaneously in nude mice for 40 days. Xenograft tumors were weighed, mean ± SD, n = 6.
e. H1299 cells with EGFR-KO background, cDNA of EV, EGFR-WT and EGFR mutation variants including G719>S, G719S720>SG, G810>S, G810S811>SG, G110>S or G1103S1104>SG were reintroduced stably. These cells were then implanted subcutaneously in nude mice for 35 days. All mice were sacrificed. Xenograft tumors were weighed, mean ± SD, n = 5.

f. Parental (CTRL) and gefitinib-resistant (GR) H1299 cells were compared in EGFR protein G>S editing mutation (i.e., G719>S, G810>S or G1103>S) induction by EGF treatment. Whole cell lysates were subjected to 10% SDS-PAGE and indicated antibodies for EGFR protein specific G>S sites were applied in Western blot.

g. The cell viability of above CTRL and GR H1299 cells, mean ± SD, n = 12.

h. The cell viability of H1299 cells was tested after treatment with different concentrations of gefitinib. IC50 was reached by a dose of 2.5 μM gefitinib.

i. H1299 cells were treated with 2.5 μM gefitinib followed by SHIN1 at different concentrations at the same time and cell viability was analyzed.

j. The G1103>S editing and SHMT2 expression were compared between PRT and GR HCC827 cells by analyzing the whole cell lysates in Western blot with indicated antibodies.

k. Violin plots were presented, which showed the effect of gefitinib and SHIN1 on cell viability in GR H1299 and GR HCC827 cells under different concentrations of SHIN1, respectively.

l. The cell viability of PRT and GR HCC827 cells were compared in response to gefitinib treatment.

m. GR HCC827 cells were treated with gefitinib alone and gefitinib at various concentrations plus SHIN1 to test the cell viabilities.

n. Representative adenocarcinoma tissues versus proximal normal tissues in pairs from NSCLC patients were stained with antibodies against EGFR G719>S, EGFR G810>S, EGFR G1103>S editing, or SHMT2 in IHC.

o. Lung adenocarcinoma tissues versus proximal normal tissues in pairs from 10 NSCLC patients were stained with IHC, and the IHC staining score was evaluated with violin plots.

p. Gene sequencing results of EGFR motifs covering G719S720, G810S811, and G1103S1104 GS islands of the samples in m were presented, revealing that none of the 20 patients carried EGFR gene mutations on these GS islands.

q. A schematic diagram was presented, which depicted the mechanism by which SHMT2 catalyzes glycine hydroxymethylation of EGFR protein GS islands, leading to G>S editing (G>S point, GS>SG switch, or GS>SG
SENA), independent of *EGFR* gene mutation. The diagram also showed that G>S editing sites can be phosphorylated in response to EGF treatment.
Extended Data Fig. 1| The GS Island editing alters the EGFR protein conformation.

a. Seven GS islands are mapped along the cytoplasmic region of the EGFR protein. While G719, G810, and G911 GS islands are located within the TK domain, G1103, G1161, G1054, and G1189 are located within the N-terminal regulatory domain. The local residues for G719 and S720 residues of G719 GS island contacting and for G810 and S811 residues of G810 island contacting are shown to have been altered due to GS>SG editing or gene mutation corresponding to the specific GS islands.

b. It presents mass-mass spectra of peptides with G>S editing recovered from EGFR proteins prepared from a lung cancer adenocarcinoma sample. A peptide with G901>S and G911>S editing sites was recovered (upper), while a G911S912>SG switch editing peptide was recovered with an untypical trypsinization pattern (middle). A switch editing peptide of G1161S1162>SG was also recovered from EGFR protein with an untypical trypsinization pattern.

c. The DNA sequence analysis of the local G901 motif of EGFR gene in the above sample used in b. It revealed no EGFR gene mutations corresponding to G901 and G911S912 motifs from the patient sample used for mass spectrometry analysis.

Extended Data Fig. 2| SHMT2 associates with EGFR via its N-terminal 5K cluster for hydroxymethylation initiation.

a. Immunoprecipitated EGFR and its associated proteins were subjected to mass spectrometry for identification, and SHMT2 protein was identified as binding with EGFR protein within H1299, A549, and H460 cell lines.

b. H1299 and BEAS2B cells were fluorescently stained with anti-SHMT2 as well as with Mitotracker and DAPI.

c. In BEAS2B, A549, and H1299 cells, cytosolic and mitochondrial fractions were analyzed in Western blot for SHMT2 localization. GAPDH and VDAC were included as the markers of cytosolic and mitochondrial fractions respectively (left panel). SHMT2 in full length (SHMT2 FL) and SHMT2 N-terminal 1-21Δ (SHMT2α) were transiently overexpressed in HEK293T cells. WCL were prepared from these HEK293T transfectants as well as A549 and H1299 cells were compared for SHMT2 FL and SHMT2α expression (right panel).

d. Structural analysis revealed that the N-terminal 5K cluster of SHMT2 protein is involved in the interaction with EGFR cytoplasmic domain. The N-terminal 5Ks of SHMT2 protein form hydrophilic bonds with the charged
residues of the EGFR protein. Meanwhile, K280 of SHMT2 resides within the center for dimer or tetramer of SHMT2 proteins, and K280 forms bonds with other residues locally.

e. Mass-mass spectra obtained for the peptides of EGFR protein bearing triple G>S mutations (G719>S, G721>S, G724>S) and G720S721>SG mutation. EGFR proteins were purified from HEK293T cells transfected with EGFR along with SHMT2.

f. DNA motif covered the cluster of glycine and serine residues (G719, G721, G724) was sequenced.

g. The dot blot results of polyclonal antibodies prepared for G719>S, G810>S, and G1103>S, respectively. Control and G>S peptides with the sequence of G719 (IKVLG719SGAFGTVYK), G810 (REHKDNIS810SQYL), and G1103 (SVPKRPS1103SVQNP) were also spotted on the nitrocellulose membrane with indicated concentrations. Antibody dilution with 1:1000 fold was applied.

h. The time course to depict H1299 cells treated with EGF (100 ng/ml) up to 240 minutes. EGFR protein G>S hydroxymethylation induction was immunostained with anti-EGFR G1103>S along with DAPI staining.

Extended Data Fig. 3| SHMT2 catalyzes hydroxymethylation in both MFTH-dependent and independent manners.

a. LC-MS spectra of in vitro-derived G1103>S peptide. SHMT2 and GST control proteins were used for the in vitro assay with G1103-peptide (RPAG1103SVQNPVYH).

b. Purified GST-SHMT2, rSHMT2 and GST were separated in 10% SDS-PAGE gel and stained with Coomassie Brilliant blue.

c. LC-MS spectra obtained from (i) SHMT2-catalyzed G1103-peptide hydroxymethylation reaction in vitro, (ii) synthetic S1103 peptide spectrum, and (iii) mixture of (i) and (ii) overlapping spectrum.

d. The mass-mass spectra of G1103>S peptide were obtained from the in vitro reaction (i), synthesized S1103 peptide (ii), and the mixture of the two (iii).

e. A synthesized G719-peptide (IKVLG719SGAFGTVYK) was incubated with purified rSHMT2 in an in vitro assay condition, which included free isotope labeled amino acid $^{13}$C-serine, PLP, THF, and ATP. The mass-mass spectra (AB6600 mass spectrometry) were obtained to show G719 peptides with G719(GSG)->[13C]-SSG, G719(GSG)>[13C]SGG, and G719(GSG)>[13C]-SS[13C]-S.

f. Synthesized G810-peptide or G1103 peptide was incubated with a high dose (500 ng/ml) of purified rSHMT2 (GST-SHMT2) in an in vitro assay.
condition, which included free amino acid $^{13}$C-serine, PLP, and MTHF. The reaction was analyzed with AB6600 mass spectrometry for GS>SG SENA editing induction.

g. Synthesized wild type G719 peptide (IKVLAG719SGAFGTVYK) and S720>A G719 peptide (IKVLAG719AGAFGTVYK) were compared for their response to SHMT2 in G719>S editing mutation induction under *in vitro* reaction condition (upper panel). The above synthesized wild type G719-peptide and A722>S mutated G719 peptide (IKVLAGSG721SFGGTVYK) were compared for their response to SHMT2 in G721>S editing induction under *in vitro* reaction condition (low panel).

**Extended Data Fig.4| Cytosolic SHMT2 affects EGFR protein signaling with a moderate effect on cytoplasmic metabolism.**

a. WCL prepared from HEK293T cells transfected with EV, Flag-EGFR, SHMT2-HA or Flag-EGFR/SHMT2-HA were analyzed in Western blot. EGFR auto-phosphorylation and ERK activation were detected using anti-EGFR-pY1068 and anti-ERK1/2-pT202/pY204 respectively. The protein expression levels of Flag-EGFR (anti-Flag), SHMT2-HA (anti-HA), ERK, and GAPDH were included.

b. Parental (PRT) and SHMT2 knocked out (KO1 and KO2) of H1299 and A549 cells were analyzed in Western blot for SHMT2 expression.

c. PRT and SHMT2 KO cells were treated with EGF (100 ng/ml) or left no treatment for 30 minutes, EGFR-pY1068 and ERK1/2-pT202/pY204 were analyzed in Western blot.

d. A549 cells were treated with DMSO (CTRL) or 5 mM SHIN1 in DMSO pretreatment for 24 hours followed by EGF (100 ng/ml) treatment or no treatment for an additional 30 minutes; EGFR phosphorylation and ERK1/2 activation were analyzed in Western blot as above.

e. Dot blot to show poly-clonal antibody against pG1103>S. G1103>S control peptide and pG1103>S peptide were spotted on the nitrocellulose membrane with indicated amounts.

f. In H1299 cells were incubated with or without 5 mM SHIN1 for 24 hours followed by EGF treatment, glycine hydroxymethylation G>S was detected via fluorescent immunostaining with specific antibody against EGFR-G1103>S. Nuclear DAPI staining was included.

g. In H1299 cells that received 5 mM SHIN1 incubation followed by EGF treatment, EGFR phosphorylation on S1103 (G1103>pS) was detected with specific antibody recognizing EGFR-pS1103. Nuclear staining with DAPI was included.
h. The cytoplasmic ROS level was analyzed in WT or SHMT2-KO A549 cells; WT and SHMT2-KO A549 cells were incubated with 10μM 2',7'-DCFH-DA, and intracellular ROS production was determined using flow cytometry (Ex 488 nm, Em 525 nm). Each ROS analysis was performed in triplicates, and a representative result was shown here.

i. The ROS levels in cytoplasm and mitochondria were analyzed in indicated H1299 cells.

j. Cytosolic NAD+ and NADH levels were examined in H1299 cells with OD450nm.

k. Cytosolic NAD+/NADH ratio was analyzed in SHMT2-KO H1299 cells expressing SHMT2 variants (EV, SHMT2-WT, K280E and 5KR) and treated with EGF (100 ng/ml, 30 minutes).

l. The JC-1 probe was applied to analyze mitochondrial membrane potential by flow cytometry in parental H1299 cells as well as in H1299 cells with SHMT2-KO background expressing SHMT2-WT, SHMT2-K280E, and SHMT2-5KR.

m. Free amino acid glycine was converted into serine by rSHMT2 in vitro. Purified rSHMT2 proteins including WT, K280E, and 5KR forms as well as the control protein GST were incubated with glycine and 5N,10N-MTHF in the cell free reaction buffer. DL-β-phenylserine content was then detected and quantified.

n. Free amino acid serine was converted into glycine in by rSHMT2 in vitro. Purified rSHMT2 proteins the control protein GST as above in m were incubated with serine and THF in the cell free reaction buffer. Glycine content was then detected and quantified.

o. Absolute quantitative capillary electrophoresis-mass (CE-MS) of amino acid profiling to compare levels between WT and SHMT2-KO H1299 cells that were not subjected to media change for 96 hours. Amino acids were in units of pg per ml from 1x10⁶ cells.

Extended Data Fig.5 G>S obtained from hydroxymethylation editing or from EGFR gene mutation effects EGFR signaling similarly.

a. Xenograft tumors of SHMT2-KO (KO1, KO2) A549 cells and parental A549 cells (WT) implanted subcutaneously in mice and stained them with Ki67 and H&E.

b. Colony formation abilities in mice by comparing SHMT2-KO1 and SHMT2- KO2 H1299 cells to parental H1299 cells (WT), mean ± SD, n = 3 (left panel). Colony formation abilities in mice by comparing SHMT2-KO1 and -KO2 A549 cells to parental A549 cells (WT), mean ± SD, n = 3 (right panel).
c. Colony formation ability was applied to test the cell viability of A549 cells expressing EV, EGFR-WT or EGFR-G1103S cDNA (upper panel). Flow cytometry analysis of apoptotic cells in A549 cells using Annexin-V and 7-AAD staining (bottom panel).

d. Flow cytometry was performed using PI staining to analyze SHMT2 KO H1299 cells and SHMT2 KO H1299 cells with SHMT2α or SHMT2 full length reintroduction. Cells in G2 and S phases were quantitated (left panel). EGFR proteins immunoprecipitated from the above H1299 cells were analyzed with mass spectrometry for G719>S editing analysis and peptides with G719>S were quantitated (right panel). The peptide ionic peaks obtained from mass spectrometry were analyzed using the MS software analysis. The G719>S and G719 signal intensity (peak area) were statistically analyzed using Prism 8.0.2 software for the ratio calculation.

e. A549 xenograft tumors expressing EV, EGFR-WT, or EGFR-G1103S were implanted subcutaneously in nude mice and analyzed for H&E staining and IHC with anti-Ki67.

f. In HEK293T cells, the SIE-luciferase reporter was transfected along with EV or SHMT2 variants followed by EGF treatment or no treatment for an additional 6 hours. STAT3 transcription activity based on SIE-luciferase reporter activation was then analyzed.

g. In HEK293T cells, the SIE-luciferase reporter was transfected along with EV or EGFR variants for EGF-activated STAT3 transcription activity analysis as above.

Extended Data Fig.6|SHIN1 inhibits lung cancer cell growth.

a. To assess the inhibitory effect of SHIN1 on lung cancer cell growth, we treated H1299 and A549 cells with various doses of SHIN1 and analyzed cell growth rates with the CCK8 assay, mean ± SD, n = 3.

b. Colony formation analysis on H1299 cells (left) and A549 cells (right) treated with 2.5 mM or 5 mM SHIN1 for 24 hours, mean ± SD, n = 3.

c. In SHMT2-KO H1299 cells reintroduced with SHMT2, cell viability was analyzed in response to 2.5 mM Gefitinib treatment for 24 hours.

d. In HEK293T cells, EGFR and SIE-luciferase reporter were co-transfected for 36 hours, followed by treatment with EGF, no EGF, or EGF plus SHIN1 for an additional 6 hours. These transfectants were subjected to STAT3 SIE-luciferase reporter activation analysis.

e. Overall survival in lung adenocarcinoma (LUAD) patients with low SHMT2 expression levels (n=239) was compared to those with high SHMT2 expression levels (n=239) (left), and in lung squamous cell carcinoma
(LUSC) patients with low SHMT2 expression levels (n=241) compared to those with high SHMT2 expression levels (n=241). We obtained these results from the TCGA database analysis.

f. Utilizing the TCGA database analysis to create a box plot of SHMT2 expression in patients with LUAD (T) (n=483) and unaffected controls (N) (n=347), as well as patients with LUSC (T) (n=486) and unaffected controls (N) (n=338).

g. Through a database analysis of TCGA (The Cancer Genome Atlas http://cancergenome.nih.gov/), SHMT2 gene expression levels in tumor tissues were compared with normal tissues from patients with various types of cancers.
References

14. Wan, R. *et al.* Comprehensive analysis of the discordance of EGFR mutation status between tumor tissues and matched circulating tumor


doi:10.1038/227561a0 (1970).


Acknowledgements
We thank John Slate and John Marshall (Brown University) and Tao O. Chen (University of Jena) for critically reviewing the manuscript. We appreciate Zhijie Chang (Tsinghua University), Jinhua Yang (Boston University), Zhong-ping Xia (The First Affiliated Hospital of Zhengzhou University), Jiong Deng (Binzhou Hospital of Binzhou Medical University) for their helpful and critical discussions during the revision of this manuscript. We thank Jinke Cheng, Bing Su, and Ming Lei of Shanghai Jiao Tong University for their input and helpful discussions. Additionally, we would like to thank Xianming Liu of Bruker Daltonics Corp and Ming Huang of ThermoFisher Scientific for their assistance with mass spectrometry analysis. We acknowledge Tongzheng Wei and Song Hu of the Health Science Institute of China Academy of Sciences at Shanghai, as well as the Shanghai Chest Hospital of Shanghai Jiao Tong University for their initial research assistance during the early stage of this project.

This research was supported by the following funding sources: National Natural Science Foundation of China (81820108023, 81530083, U1603284, 2016YFC1302402), Zhejiang Provincial People’s Hospital Clinical Research Funding, Wenzhou Medical University Yuying Hospital Funding, and Shandong Provincial Initiation Funding (YDZX2 021006). The authors of this paper have no conflicts of interest to disclose.

Author contributions
Y.E.C. supervised the research. Y.E.C., J.K.D., Y.Y.Z., T.C.Z. and J.Y.L. wrote and revised the manuscript. J.K.D., Y.Y.Z., R.D.Z. and J.Y.L. designed and performed the animal, mass spectrometry analysis, and drug-resistance experiments. J.K.D., Y.Y.Z., and R.D.Z. performed all the in vitro assays, Western blot, COIP analysis cDNA construction and recombinant protein purification. X.S.L., L.Y.L., F.F.H., and T.C.Z. performed fluorescent immunostaining, protein-protein interaction, DNA/mRNA sequencing, tissue sample collection and processing, and clinical data analysis. C.C.Z. analyzed protein structure and EGFR-SHMT2 protein interaction assays. Y.Y.Z. and X.M.L. performed data analysis. Y.E.C. provided advice and technical support for the entire project. All authors contributed to the manuscript's review and provided feedback.
Methods

Cell culture, treatment and transfection

Human immortalized normal epithelial cell line BEAS-2B and lung cancer cell lines A549, H1299, H460, HCC827 and H838 were all purchased from the Cell Bank of the Chinese Academy of Sciences (Shanghai). These cell lines were seeded at a density of $3.0 \times 10^4$ cells per well in 6-well plates containing complete RPMI-1640 or DMEM medium (HyClone, Logan, Utah, USA), supplemented with 10% fetal bovine serum (Gibco, Carlsbad, California, USA) and 1% penicillin/streptomycin (Sangon Biotech, Shanghai, China) at 37°C and 5% CO$_2$ condition. After 48 hours, cells were digested with trypsin, suspended within the cell supernatant, washed with cold PBS twice, and stained with binding buffer containing Annexin V-PE and 7-AAD-Cy5.5 for 15 minutes at room temperature in the dark. Flow cytometry was performed with a BD FACSCanto II system from BD Biosciences (San Jose, California, USA), and data were analyzed using FlowJo software of Tree Star (Ashland, Oregon, USA). For EGFR activation, recombinant human EGF protein (236-EG) (R&D Systems, Minneapolis, Minnesota, USA) was added to treat the cells with <15 minutes after starvation for 12 hours. SHMT2 inhibitor SHIN1 (RZ-2994) was purchased from MedChem Express (New Jersey, USA) for A549 and H1299 cells treatment. All plasmids were transfected into HEK293T cells by using PEI from Polysciences (Chicago, Illinois, USA) according to the manufacturer’s protocol.

Tissue collection and ethics approval

Fresh lung cancer and adjacent noncancerous tissues were collected from Zhejiang Provincial People’s Hospital and Yantaishan Hospital of Binzhou Medical University, with the approval of the Ethics Committees of the respective hospitals. The tissue samples were collected during surgery and stored at −80°C until use.

Western blotting and antibodies

To extract proteins, cell lysates were collected and lysed in lysis buffer containing 20 mM Tris-HCl (pH = 8.0), 100 mM NaCl, 1 mM EDTA, and 1% Triton X-100. Primary antibodies used for Western blotting included anti-SHMT2, anti-EGFR, anti-p44/42 MAPK, anti-phospho-p44/42 MAPK, anti-AKT, anti-phospho-AKT (Ser473), and anti-GAPDH (Cell Signaling Technology, Inc., Beverly, MA, USA).

Construction of stable cell lines using crispr-cas9 knockout and plasmids

The sgRNA sequences targeting SHMT2 were designed by a CRISPR designer.
at https://www.atum.bio/eCommerce/cas9/input. All targeting sequences were synthesized and ligased into LentiCRISPRv2 plasmid (52961) from Addgene (Watertown, Massachusetts, USA) to generate SHMT2KO plasmids. HEK293T cells were co-transfected with these plasmids and viral packaging plasmids (VSVG, REV, and MDL), and the medium was changed after transfection for 6 h. After 48 hours, the virus was collected from the cell supernatant using a 0.22 μm strainer and used to infect A549 and H1299 cells. Targeted cells were then selected by exposure to puromycin for 3 weeks. The guide sequences targeting Exon 2 of SHMT2 in Homo sapiens are as follows: SHMT2-gRNA-F1: 5'-CACCGTCTCACACGAC, SHMT2-gRNA-R1: 5'-AAACGTCGGACAGTGATCCTGAGAC; SHMT2-gRNA-F2: 5'-CACCG GACAGGCAGTGTCGTGGCC, SHMT2-gRNA-R2: 5'-AAACCG CCAGCAGCGACAC TGCCTGTCC.

Full-length cDNA cloning and site-directed mutagenesis
The full-length CDS region of EGFR and SHMT2 was amplified with the KOD Plus Mutagenesis Kit (Toyobo Co. Ltd., Osaka, Japan) and cloned into pCDH-CMV-MCS-EF1-copGFP lentiviral vector with 3XFLAG or HA tags in the C-terminal region. Site-directed mutagenesis was performed to construct G-to-S mutated EGFR (G1103S, G810S) and SHMT2-K280E mutants with Takara PrimerSTAR Max DNA Polymerase (Takara, Japan) based on the wild-type EGFR and SHMT2 plasmids. A series of SHMT2 constructs (K280E and other individual KR mutants) were provided by Jianyuan Luo25. GST-SHMT2 was subcloned using the pGEX-4T-1 plasmid. All DNA constructs were sequenced by Genewiz Biotech (Suzhou, Jiangsu, China) prior to use.

Colony formation assay and apoptosis analysis
To perform the colony formation assay, lung cancer cells were seeded in 6-cm dishes at a density of 4x103 cells/well and continuously cultured with complete RPMI-1640 medium for 7-10 days for the formation of colonies. Apoptosis analysis was carried out using flow cytometry as described previously (Xiong et al., 2017).

Hematoxylin and eosin (H&E) staining, immunofluorescent staining, and immunohistochemical (IHC) assay
H&E and IHC staining were conducted as described in previous studies (PMID: 28158983 and 27279345). The IHC staining score was evaluated by considering staining intensity and percentages of positively stained cells. IHC assay antibodies used were diluted with antibody dilution buffer (ZSGB-BIO,
Beijing, China) as follows: EGFR-G1103>S antibody in a 1:100 dilution, EGFR-
G810>S antibody in a 1:100 dilution, SHMT2 antibody for a 1:200 dilution,
ERK1/2-pT202/Y204 antibody with a 1:500 dilution, and Ki67 (9449) of Cell
Signaling Technology with a 1:400 dilution. For immunofluorescence staining,
lung cancer cells were seeded into glass slides set in 24-well plates with the
appropriate cell density. After 24 hours, cultured cells were washed with cold
PBS, fixed with 4% paraformaldehyde, and permeabilized with 0.5% Triton X-
100. The subsequent processes followed standard protocols.

**Immunoprecipitation and SHMT2 protein purification**
Whole-cell lysates from lung cancer cell lines were incubated with 1μg of mouse
EGFR monoclonal antibody overnight at 4°C. Protein A/G agarose-beads
(20421) of Thermo Pierce were added for 4 additional hours. The beads were
washed thrice with lysis buffer and boiled in SDS protein loading buffer for five
minutes, followed by Western blot analysis. GST-SHMT2 and GST proteins
were expressed in B21 E. coli strain and induced with 0.4 mM IPTG at 18°C for
16 hours. Bacterial cells were lysed in a buffer containing 140 mM NaCl, 2.7
mM KCl, 10 mM NaHPO4, 1.8 mM KH2PO4, 100 mM PLP, and 0.5mM PMSF
and then sonicated on ice. The lysate was obtained after centrifugation
(10,000xg, 30 min, 4°C) and incubated with GST-tag beads (Beyotime,
Shanghai) at 4°C with rotation for 4 hours. Finally, beads were gradient eluted
with elution buffer (50 mM Tris-HCl (pH 8.0), 50 mM GSH, 100 mM PLP, and
5 % glycerol).

**SHMT2 G>S and GS>SG SENA editing (hydroxymethylation) in vitro assay**
Recombinant SHMT2 protein obtained from the bacterially expressed GST-
SHMT2 proteins was used in the in vitro assay by using synthetic peptides as
the substrates. The reaction assay buffer contained 50 mM Tris-HCl (pH=8.0),
100 mM NaCl, 2 mM mTHF, 0.5 mM PLP, and the enzyme rSHMT2, incubated
with or without peptide (G719-peptide, G810-peptide, G1103-peptide, G1054-
peptide and S694-peptide). While unlabeled peptides of EGFR sequences
were synthesized by Chinapeptides Biotech Inc (Wuhan, Hubei, China), isotope
(13C and 15N) labeled peptides as well as amino acid 13C-serine were
provided by ChemSrc (Shanghai, China). The reactions were continued at 37°C
for 60 minutes and stopped by adding 10%.

**Microscale thermophoresis binding assay**
EGFR and SHMT2 protein-protein interaction was measured using microscale
thermophoresis. A range of concentrations of the ligands (from 0.07 to 2250
nM) were incubated with 1x10⁶ eGFP-EGFR overexpression HEK293T cells
lysis in the assay buffer (20mM Tris-HCl, 500 mM NaCl, 1mM MgCl₂
supplemented with 0.2% Tween, pH 7.4) for 10 min. Samples were then loaded
into NanoTemper glass capillaries and microscale thermophoresis was carried
out using 80% light-emitting diode power and 80% MST power on a
NanoTemper monolith NT.115 61. All Kd values were calculated using the
NanoTemper software.

Identification of EGFR protein mutations with LC-MS/MS analysis
The key measure in protein sequencing using mass spectrometry is to enhance
the recovery rate of trypsinized peptides from the analyzed proteins⁵⁴. Cell
lysates were collected from tissues of lung cancer patients, cell lines, and
HEK293T cells that were transfected with a Flag-EGFR plasmid, and were
immunoprecipitated using monoclonal EGFR antibody or anti-Flag M2 affinity
agarose beads. The boiled products were run on SDS-PAGE gels and
subjected to Coomassie blue staining. The bands corresponding to EGFR
protein were excised from the gel and treated with 100% Acetonitrile, 10 mM
dithiothreitol, 25 mM ammonium bicarbonate, 55 mM chloroacetamide and then
digested with trypsin (Thermo Fisher Scientific) for 16-24 hours. Digested
peptides were analyzed conventionally by using AB6600 (AB Sciex TripleTOF
6600) or Orbitrap Fusion Lumos Mass Spectrometer (Thermo Fisher Scientific)
according to our previous published protocol⁵⁵. Peptides were also analyzed
on a timsTOF pro mass spectrometer from Bruker Daltonics (Bremen, Germany)
with PASEF mode acquisition. Peptides were separated using a 90 min on a
C18 analytical column (Aurora Series Emitter Column, 25 cmx75 μmx1.6 μm,
IonOpticks). The column temperature was maintained at 50°C using an
integrated Toaster column oven. Mobile phases A and B were water and
acetonitrile with 0.1 vol% FA, respectively. The %B was linearly increased from
2 to 22% within 72 min, followed by an increase to 35% within 8 min and a
further increase to 80% within 5 min and sustained for 5 min. Mass
spectrometry raw files were analyzed by Peaks Studio Version 10.6
(Bioinformatics Solutions Inc.), and MS spectra were searched against the
Swiss-Prot Human reference library (Downloaded May 31, 2019) contains
20,420 sequences; carbamidomethyl cysteine (+57.0215 Da) as a fixed
modification; oxidized methionine (+15.9949), hydroxymethyl glycine
(+30.0106 Da), de-hydroxymethyl serine (-30.0106 Da) acetylation (+42.0106
Da) of protein n-terminal as variable modification; 20 ppm mass tolerance on
precursor ions, 0.05 Da on fragment ions. The false discovery rate (FDR) was
set at less than 1% for peptide sequences.

DNA and mRNA isolation, cDNA synthesis, and PCR
To isolate total DNA from tissues and cell lines, a universal genomic DNA
extraction kit (code No. 9767; Takara Biotech, Beijing, China) was utilized,
whereas total RNA was extracted using a universal RNA extraction kit (code No. 9767; Takara Biotech) as per the manufacturer’s instructions. The amount of DNA and RNA was measured using a NanoDrop One (Thermo Fisher Scientific, Waltham, MA). A PrimeScript™ IV 1st strand cDNA synthesis mix kit (code No. 6215A; Takara Biotech) was used for reverse transcriptase-mediated cDNA synthesis. Specific primers were synthesized to amplify regions of DNA and cDNA sequences containing mutant sites of

EGFR: EGFR-G810-F: 5’-CTCCAGGAAGCCTACGTGAT-3’; EGFR-G810-R: 5’-TTATCTCCCCTCCCCGTATC-3’; EGFR-G1103-F: 5’-TGTTGAGGACATTACGAGGTT-3’; EGFR-G1103-R: 5’-GGGCTGGACAGTGGGATAGA-3’; EGFR-G810-F: 5’-AAATCCTCGATGAAGCCTACG-3’; EGFR-G810-R: 5’-ATCCAGCACTTGACTACCGTG-3’; EGFR-G1103-F: 5’-CAGACTCCAACCTTCTACGTG-3’; EGFR-G1103-R: 5’-GTCGAATGTGCTGGACACA-3’.

PCR analysis was conducted using super-fidelity DNA polymerase kits of Vazyme (Nanjing, Jingsu, China) and PCR products were further sequenced.

**Mouse xenografts**

For mouse xenografts, female BALB/c nude mice (6 to 8 weeks old) were acquired from the Experimental Animal Center of Soochow University. A549 stable cells were trypsinized, counted, and mixed with PBS and Matrigel (BD Biosciences) before injecting into one side flank of a nude mouse subcutaneously, at a concentration of 3x10^6 cells per group. Tumor size was assessed every 2-3 days and tumor volume was calculated using the formula V = L (tumor length) × W (tumor width)²/2. After six weeks, mice were sacrificed, and xenograft tumors were removed, weighed and analyzed. All animal experiments were approved by the Ethics Committees of Zhejiang Provincial People’s Hospital, Soochow University, and Wenzhou Medical University.

**ROS, NAD/NADH, Mitochondrial membrane potential, amino acid analysis, and SHMT2 catalytic activity in vitro**

To analyze ROS, cells were seeded in a 6-well plate. After 24 hours, the cells were washed with PBS, and then incubated with 10 μM 2’,7’-Dichlorodihydrofluorescein diacetate (DCFH-DA) (Beyotime) in FBS-free RPMI 1640 medium (HyClone) at 37ºC for 30 minutes to load the fluorescent dye. The unloaded fluorescent dye was removed after washing the cells twice with PBS. Flow cytometry analysis (Ex 488 nm, Em 525 nm) was performed to determine the intracellular ROS production.
For detecting cellular NAD+ and NADH levels, the NAD+/NADH test kit (S0175) (Beyotime) was used as per the manufacturer’s instructions. Briefly, whole cell lysates were prepared as described before and centrifuged at 12,000 g for 5 minutes at 4°C, after which the supernatant was collected. Each sample was divided into three portions, one of which was incubated in a 60°C water bath for 30 minutes to break down NAD+. The ratio of NAD+ to NADH was computed by determining the total content of NAD+ and NADH after removing NAD+.

Mitochondrial membrane potential assay: The mitochondrial membrane potential was measured using a mitochondrial membrane potential assay kit with JC-1 according to the manufacturer’s instructions. Briefly, cells were cultured in serum-free DMEM containing JC-1 staining working solution at 37°C for 30 minutes followed by two washes with JC-1 buffer. DMEM (2ml) was then added, and the cells were photographed with a fluorescence microscope.

SHMT2 catalytic activity: SHMT2 catalytic activity in glycine-serine conversion in vitro was assayed by following the method described in a previous study (PMID: 30367038). In brief, the reaction mixture consists of rSHMT2 protein, 50mM DL-β-phenylserine (Sigma-Aldrich), 25 mM sodium sulfate, 50μM PLP, 1 mM EDTA, and 50 mM HEPES buffer (pH = 8.0). Absorbance of the product was measured every minute for 15 minutes at 279 nm by the microplate reader of BioTek Instruments Inc (Winooski, Vermont, USA) at 25°C.

Statistical analysis:
All statistical analyses were performed using Graphpad Prism 8.0 software from GraphPad Software Inc (San Diego, California, USA). All data are presented as mean ± SD.
EGFR protein carries EGFR gene-mutation independent G>S and GS>SG editing in lung cancer. a. EGFR proteins were immunoprecipitated with a monoclonal antibody against EGFR from the whole cell lysates (WCL) of lung cancer tissues obtained from NSCLC patients. The immunoprecipitated EGFR proteins
were trypsinized and analyzed with LC-MS/MS (AB Sciex TripleTOF 6600, Bruker TimsTof Pro, or Thermo Scientific Orbitrap Exploris 480). The heat map shows Log2 differences in signal intensity of G>S and/or GS>SG protein sequence changes appearance in EGFR protein. This heat map is based on lung cancer tissues of 45 NSCLC patients and 7 lung cancer cell lines (H1299, A549, 1975, H-522, Cula-3, H-520, and Beas-B2). We recovered peptides of EGFR protein with G>S protein sequence changes. Our trypsin digestion-mass spectrometry approach routinely covered 50-80% of the EGFR protein sequence. b. EGFR protein changes and GS Islands. The full-length EGFR protein contains 85 glycine residues, out of which 35 are sporadically distributed within the cytoplasmic domain. Using mass spectrometry analysis, G>S protein sequence changes were detected in lung cancer tissues and cell lines, with G719, G810, and G1103 sites having the highest frequency of G>S and/or GS>SG protein sequence changes. The cytoplasmic domain of EGFR has 7 "GS islands" that align with the human and mouse origins. c. Heat map and Venn diagram analysis. A heat map was generated to depict the efficiency of G>X, S>X, T>X, or L>X detected in lung cancer tissues and cell lines of the samples analyzed using mass spectrometry analysis. Peptides were detected in two replicates (n=2) of every sample. Venn diagram analysis of 52 samples in a with G>S, GS>SG, and S>G protein sequence changes of these GS-islands to reveal that 10 samples bear all three types of protein sequence changes, 33 samples bear GS>SG and G>S, 1 sample bears G>S and S>G whereas no sample bears GS>SG and S>G. d. The representative LC-MS/MS spectra of GS island G>S and GS>SG protein sequence changes. Peptides with G>S single protein sequence changes (G719>S, G810>S, G1103>S) and GS>SG switch protein sequence changes (G719S720>SG, G810S811>SG, and G1103S1104>SG) of EGFR protein were detected. EGFR proteins were immunoprecipitated from lung cancer tissues of NSCLC patients, and AB Sciex TrippleTOF 6600 mass spectrometry was used for analysis. The y and b series of ions were indicated along the sequence of the peptide derived, which was shown at the top of the panel. e. EGFR gene sequencing. Genome DNA and mRNA extracted from the lung cancer tissues and cell lines used above were sequenced to determine whether there were EGFR mutations with fragment spans of G719S720, G810S811, and G1103S1104 islands. The sequencing revealed that EGFR DNA or mRNA carried no mutations in the samples analyzed.
Figure 2

EGFR recruits SHMT2 for GS island G>S and GS>SG editing induction. a. EGFR proteins were immunoprecipitated from the H1299 and A549 cells, and mass spectrometry was performed to identify EGFR and binding proteins. As marked, such immunoprecipitated EGFR protein and binding proteins were separated in silver-staining gel. b. SHMT2 was identified as an EGFR binding protein from the mass spectrometry analysis. c. Coimmunoprecipitation analysis confirmed the EGFR and SHMT2 association.
in the H1299 and A549 cells, respectively. d. Purified eGFP-EGFR protein and SHMT2 (WT, K280E, or 5KR variant) proteins were used for a microscale thermophoresis binding assay. The Kds of eGFP-EGFR and SHMT2 association were calculated using NanoTemper software. The Kd for EGFR and SHMT2-K280E mutant association was undetectable. e. Secondary structural alignment analysis between human SHMT1 and SHMT2 showed that K280 is conserved between the two while N-terminal 5Ks (K459, K461, K464, K469, and K474) are specific to SHMT2. f. In HEK293T cells, EGFR was transfected with empty vector (EV) or along with SHMT2. EGFR proteins, immunoprecipitated from the transfectants were analyzed with LC-MS/MS (AB6600). The LC-MS/MS spectra of G719 island were shown. While wild type G719 island-peptides were obtained from the cells transfected with EGFR alone, G719>S and G719>SS720>G mutated peptides were recovered from HEK293T cells transfected with EGFR and SHMT2 together. g. DNA sequencing result of EGFR G719 island motif of above samples in f. h. SHMT2 variants (K280E or 5KR) were transiently transfected to compare with EGFR WT on EGFR G>S or GS>SG mutation induction as in f. LC-MS/MS analysis revealed that SHMT2 with K280E or 5KR mutation failed to induce G>S and GS>SG mutations within G719, G810, and G1103 GS islands. i. A time course of EGFR protein glycine hydroxymethylation G>S was induced by EGF in H1299 cells. Whole cell lysates (WCL) were then prepared and separated by 10% SDS-PAGE and subjected to immunoblotting analysis with antibodies recognizing EGFR proteins with G810>S and G1103>S editing respectively. j. SHMT2-KO H1299 cells were generated using the CRISPR/CAS9 approach. In H1299 cells with SHMT2-KO background, EV, SHMT2 variants as indicated were stably introduced. Equal amounts of whole cell lysates were separated by 10% SDS-PAGE and subjected to immunoblotting analysis with antibodies recognizing EGFR proteins with G719>S, G810>S and G1103>S respectively. EGFR, Flag-SHMT2, and GAPDH expression levels in blot were included. k. EGFR proteins were immuno-stained with an antibody recognizing G719>S, G810>S, and G1103>S in both parental (WT) and SHMT2-KO H1299 cells. SHMT2 immuno-staining and DAPI staining were included in the experiment, and the cells were visualized using fluorescent confocal microscopy. The scale bar was set at 10 μm.
Using free amino acid serine as hydroxymethyl donor for G>S and S>G editing induction. a. Parental H1299 cells expressing wild type SHMT2 (WT), H1299 cells with SHMT2-KO, and SHMT2-KO H1299 cells with SHMT2-Flag reintroduction in different doses (1 ug and 2 ug) were grown in DMEM medium containing free [13C]-serine for 24 hours. The LC-MS spectrometry (adduct ionic peaks) results (AB6600) showed that both the G719 GS-island sequence (VLG719SGAFGTVYK) and G810 GS-island sequence (DIG810SQYLLNWCVQIAK) of EGFR protein were detected in all four groups of the cells. For G719 GS-island, while the wild type GSG-peptide and GGG-peptide ionic peaks were recovered from the cells under all 4 conditions, the ionic peaks of [13C]-SGG-peptide, [13C]-SSG-peptide, and [13C]-SSS-peptide were recovered only from the cells expressing SHMT2. For G810 GS-island, wild type GS-peptide and GG-peptide ionic peaks were recovered from all the cell types. However, [13C]-SG-peptide and [13C]-SS-peptide were recovered only from cells expressing SHMT2. The mass spectrum for peptide quantitation was provided for both G719 GS island and G810 GS-island for the cells expressing SHMT2-Flag ++ (right, next to the adduct ionic peak spectra). b. Signal intensities of the results obtained in a. c. EGFR proteins were immunoprecipitated from H1299 cells growing in the DMEM medium containing free [13C]-serine. Four peptides with different G>S editing mutations were recovered from the EGFR proteins, and their mass-mass spectra were analyzed. The G719 motif (VLG719SGAFGTVYK) included two mono-editing residues: G719G720G721 and [13C]-S719S720G721 (“GGG” and “SSG”) and two double-editing residues: [13C]-S719G720G721 and [13C]-S719S720[13C]-S721 (“SGG” and “SSS”). d. It depicts how SHMT2 catalyzed hydroxymethylation reaction, with the [13C]-hydroxymethyl group of free [13C,15N]-serine being transferred onto the protein's glycine residue as a side chain, resulting in G>[13C]-S editing. e. Dot blot analysis of the catalytic activity of different forms of rSHMT2 as indicated on G>S editing induction in
in vitro by using G719-peptide, G810-peptide, and G1103-peptide as substrates (left panel). The LC-MS spectra of these in vitro reaction products were included (right panel). f. The 5N,10N-[13C]-CH2-THF formation was detected with LC-MS analysis of the in vitro reaction with [13C]-serine catalyzed by rSHMT2 at different doses as indicated. THF was included in the reaction buffer. The mass spectrum of 5N,10N-[13C]-CH2-THF quantitation was given in the right panel. g. The ADP level was detected with LC-MS analysis of the reaction. LC-MS spectrum of the G>S editing reaction catalyzed by SHMT2 were analyzed in the presence of PLP, ATP, and 5N,10N-MTHF. h. The LC-MS spectrum of the G>S editing with the G810 peptide catalyzed by rSHMT2 in vitro assay. Peptides with G810>S point editing and S811>G point editing were recovered. i. The above catalytic reaction in i based upon the G810-peptide at different concentrations (50, 100, 150, 200, and 250 mM), initiated by rSHMT2 protein in the presence of 10 mM THF, 40 mM PLP, 100 mM ATP, 5 mM MnCl2 and 5 mM KCl in the in vitro reaction buffer (PBS 1 ml, pH7.2), 37°C, 15 minutes. The kinetic parameters Kcat and Km were then calculated. j. H1299 cells grew in the DMEM medium containing isotope labeled amino acids ([13C]-Ser, [13C]-Ala, [13C]-Gln, [13C]-Phe, [13C]-Thr, and [13C]-Cys) for 24 hours. Immunoprecipitated EGFR proteins from these cells were then subjected to LC-MS for analysis, mean ± SD, n = 3. k. G719 peptide was incubated with rSHMT2 in the presence of isotope labeled amino acid ([13C]-Ser, [13C]-Ala, [13C]-Thr, or [13C]-Cys) in the reaction buffer for 1 hour in the in vitro assay. The peptides were then subjected to LC-MS for analysis, mean ± SD, n = 3.

Figure 4

SENA editing in vitro. a. Wild type or S720A-mutated EGFR cDNA was introduced for expression in EGFR-KO background H1299 cells. Immunoprecipitated EGFR proteins were subjected to trypsin digestion and
mass-mass spectral analysis of EGFR protein G719>S editing efficacy. The intensity of EGFR protein G719>S editing signal from LC-MS spectra was plotted. b. EGFR-KO background H1299 cells were stably introduced with wild type EGFR or A722S-mutated EGFR for expression. Immunoprecipitated EGFR proteins were subjected to trypsin digestion and LC-MS/MS spectral analysis of G721>S editing efficacy. The intensity of EGFR protein G721>S editing signal from LC-MS spectra was plotted. c. SENA editing denotes the SHMT2 catalyzed hydroxymethylation for GS>SG switch editing reaction. The [13C]-hydroxymethyl group of [13C,15N]-serine residue of the GS island can be transferred onto the glycine residue adjacent to it, resulting in G>[13C]-S switch editing. d. Isotope-labeled peptides G719[13C]-S720G721 peptide (upper panel) and G810[13C]-S811 peptide (lower panel) were incubated with rSHMT2 protein at indicated concentrations, and LC-MS spectrometry analysis was performed. For G719 peptide, GGG peptide (i.e., G719G720G721-peptide) in upper panel and for G810-peptide, GG peptide (G810G811-peptide) in lower panel were recovered by incubating the isotope-labeled peptide with rSHMT2 protein. The peak in red color was a mixture of GS-peptide and SG peptide in both G719 and G810 peptides when 500 ng rSHMT2 protein was used, as revealed by the quantitative mass analysis included on the right. e. LC-MS/MS spectra showed that rSHMT2 at a high dose (500 ng/ml) catalyzed GS>SG switch editing mutation. When incubated with rSHMT2 protein at 500 ng/ml for 30 min at 37°C, [13C]-S719G720G721 peptide was generated from G719[13C]-S720G721 peptide (VLG719SGAFGTLYK) or 13C- S810G811 peptide was generated from G810[13C]-S811-peptide (YVREHKDNIG810SQYLLNW). The serine residue's [13C]-hydroxymethyl group within the GS island can be transferred onto the adjacent glycine residue as the side chain, leading to SENA GS>SG editing mutation. f. The protein editing reaction was based on the G719-peptide at different concentrations (100, 200, 300, 400, and 500 mM), initiated by rSHMT2 protein in the presence of 40 mM PLP, 100 nM ATP, 5 mM MnCl2 and 5 mM KCl in the in vitro reaction buffer (PBS 1 ml, pH7.2) at 37°C for 15 minutes. THF was not included. For G719SG>SGG and GSG722>GGS editing catalyzed by rSHMT2, the kinetic parameters (Kcat and Km) of SHMT2 on G719-peptide SENA-L (GSG>SGG) and SENA-R (GSG>GGS) were determined (upper panels). For G810 peptide (50, 100, 150, 200, and 250 mM) initiated by rSHMT2 protein in the above reaction conditions. The kinetic parameters (Kcat and Km) of SHMT2 were compared with and without THF (10 mM) in the editing mutation induction s were compared for their Kcat and Km (left panel). LC-MS/MS of SGG and GGS were included (lower panels). g. LC-MS/MS spectra of SGG and GGS peptides recovered from G719- peptide (left) and SG peptide recovered from G810-peptide catalyzed by rSHMT2 in vitro for Kcat and Km calculation in f. h. The synthesized G1054 peptide (CIDRNGLQ[13C]-SCPIK), G694 peptide (ELVEPLTP[13C]-SGEAPNQALLR), G719-peptide (IKVLG[13C]- SGAFGTLYK), and G810-peptide (YVREHKDNIG[13C]-SQYLLNW) were compared for their ability to induce SENA-L editing by rSHMT2 in LC-MS spectra. i. An equal amount of the isotope-labeled G719-peptide ([13C]-S720) was mixed with the unlabeled G719-peptide (S720) and incubated with rSHMT2 under in vitro reaction conditions in the presence or absence of THF. Four types of peptides with G>S editing were recovered from the LC-MS spectra.
Figure 5

Glycine hydroxymethylation dependent GS island G>S editing are activating one for EGFR signaling. a. SHMT2-KO H1299 cells were stably transfected with cDNA of EV and SHMT2 variants (WT, K280E, and 5KR). After EGF treatment for 30 minutes, whole-cell lysates were collected for analysis of EGFR tyrosine autophosphorylation (pY1068) and ERK1/2 activation (pT202/pY204) via Western blotting. EGFR, ERK1/2, SHMT2, and GAPDH levels were blotted as indicated. b. The signal intensities of EGFR-pY1068

---

1. **Panel a**: Western blot showing expression levels of SHMT2 (SHMT2-Flag), EGFR, and pERK1/2, with treatment conditions indicated (EV, WT, K280E, 5KR, SHMT2, GAPDH).
2. **Panel b**: Bar graph showing relative intensities of EGFR-pY1068 with p-values (CTRL, SHMT2).
3. **Panel c**: Western blot showing expression levels of SHMT2, ERK1/2, and GAPDH.
4. **Panel d**: Bar graph showing relative intensities of SHMT2 and SHMT2/SHMT2 with p-values (CTRL, SHMT2).
5. **Panel e**: Graph showing relative intensity of EGFR-pY1068 with p-values (CTRL, 0, 0.1, 1, 10 μM SHMT2).
6. **Panel f**: Peptide analysis showing G719, G810, G1103, G719>S, G810>S, G1103>S.
7. **Panel g**: Mass spectrometry showing m/z distribution for SHMT2 variants.
8. **Panel h**: Fluorescence images comparing EGF treatment in SHMT2-KO H1299 and parental H1299.
9. **Panel i**: Western blot showing expression levels of EGFR, ERK1/2, and GAPDH.
10. **Panel j**: Bar graph showing relative intensities of pERK1/2 with p-values (CTRL, EGF).
(upper) and ERK1/2-pT202/pY204 (lower) of the samples in a were quantitated, mean ± SD, n = 4. c. H1299 cells were pretreated with 5 mM SHIN1 for 24 hours, followed by EGF treatment for an additional 30 minutes. EGFR phosphorylation and ERK1/2 activation were analyzed via Western blotting with anti-EGFR- pY1068 and anti-ERK1/2-pT202/pY204, respectively. EGFR, ERK1/2, SHMT2, and GAPDH levels were blotted as indicated. d. EV or SHMT2 cDNA was transiently overexpressed in H1299 cells. EGFR G1103>S editing peptides were recovered from mass spectrometric analysis of these cells and SHMT2-overexpressing H1299 cells treated with 10 mM SHIN1 for 24 hours. The relative signal intensities of EGFR G1103>S editing detected in LC-MS spectra were quantitated, mean ± SD, n = 3. e. LC-MS spectrometry analysis was performed to determine the inhibitory effect of SHIN1 on glycine hydroxymethylation in vitro. G1103-peptide was incubated with purified rSHMT2 protein in an in vitro reaction buffer, and DMSO (CTRL) or SHIN1 (0 - 10 mM) in DMSO was included. The ionic peak signal intensities obtained from the mass results of G1103>S conversion were plotted, mean ± SD, n = 3. f. The inhibitory effect of SHIN1 on hydroxymethylation in vitro was evaluated via dot blot assay. G719-peptide and 1103-peptide were incubated with or without GST and rSHMT2 proteins in an in vitro reaction buffer, and DMSO (-) or 5 mM SHIN1 in DMSO was included. g. EGFR proteins were immunoprecipitated from EGF-treated HEK293T cells with EGFR overexpression, trypsinized, and analyzed with AB6600 mass spectrometry. Phospho-G1103>S peptide of EGFR protein (G1103>pS) was detected. h. EGFR phosphorylation on G1103>S was detected via immunostaining with anti-EGFR-G1103>pS in SHMT2-KO H1299 cells and parental H1299 cells treated with EGF or left untreated. i. EGFR was knocked out in H1299 cells via CRSPRA/CAS9 approach. EV or EGFR variants (WT, G810>S, G1103>S, G810S811>SG, G1103S1104>SG) were stably transfected into these H1299 cells with EGFR-KO background. ERK1/2 activation (pT202/pY204) in response to EGF treatment (100 ng/mL, 30 minutes) was analyzed via Western blotting. GAPDH was used as a loading control. j. The signal intensities of ERK1/2-pT202/pY204 of the samples in i were quantitated, mean ± SD, n = 4.
Figure 6

EGFR protein GS island G>S and GS>SG editing develop EGFR targeting therapy resistance. a. In a xenograft mouse model, the tumor formation ability of A549 cells with SHMT2 knockout (KO-1 and KO-2) was compared to parental A549 cells (WT) by implanting them subcutaneously for tumor formation in nude mice for 41 days. The xenograft tumors were then pictured and compared. The result showed that SHMT2 knockout had a significant effect on tumor cell growth, n = 6. b. The xenograft tumors in a were...
weighed, and the result showed that the tumors with SHMT2 knockout were significantly smaller compared to the control group, n = 6. c. In H1299 cells with EGFR-KO background, EV, SHMT2-WT, SHMT2-K280E, or SHMT2-5KR mutant was reintroduced stably. EDU staining was performed, and EDU positive cells were quantitated in percentage, mean ± SD, n = 6. d. The above cell lines in c were implanted subcutaneously in nude mice for 40 days. Xenograft tumors were weighed, mean ± SD, n = 6. e. H1299 cells with EGFR-KO background, cDNA of EV, EGFR-WT and EGFR mutation variants including G719>S, G719S720>SG, G810>S, G810S811>SG, G110>S or G1103S1104>SG were reintroduced stably. These cells were then implanted subcutaneously in nude mice for 35 days. All mice were sacrificed. Xenograft tumors were weighed, mean ± SD, n = 5. f. Parental (CTRL) and gefitinib-resistant (GR) H1299 cells were compared in EGFR protein G>S editing mutation (i.e., G719>S, G810>S or G1103>S) induction by EGF treatment. Whole cell lysates were subjected to 10% SDS-PAGE and indicated antibodies for EGFR protein specific G>S sites were applied in Western blot. g. The cell viability of above CTRL and GR H1299 cells, mean ± SD, n = 12. h. The cell viability of H1299 cells was tested after treatment with different concentrations of gefitinib. IC50 was reached by a dose of 2.5 μM gefitinib. i. H1299 cells were treated with 2.5 μM gefitinib followed by SHIN1 at different concentrations at the same time and cell viability was analyzed. j. The G1103>S editing and SHMT2 expression were compared between PRT and GR HCC827 cells by analyzing the whole cell lysates in Western blot with indicated antibodies. k. Violin plots were presented, which showed the effect of gefitinib and SHIN1 on cell viability in GR H1299 and GR HCC827 cells under different concentrations of SHIN1, respectively. l. The cell viability of PRT and GR HCC827 cells were compared in response to gefitinib treatment. m. GR HCC827 cells were treated with gefitinib alone and gefitinib at various concentrations plus SHIN1 to test the cell viabilities. n. Representative adenocarcinoma tissues versus proximal normal tissues in pairs from NSCLC patients were stained with antibodies against EGFR G719>S, EGFR G810>S, EGFR G1103>S editing, or SHMT2 in IHC. o. Lung adenocarcinoma tissues versus proximal normal tissues in pairs from 10 NSCLC patients were stained with IHC, and the IHC staining score was evaluated with violin plots. p. Gene sequencing results of EGFR motifs covering G719S720, G810S811, and G1103S1104 GS islands of the samples in m were presented, revealing that none of the 20 patients carried EGFR gene mutations on these GS islands. q. A schematic diagram was presented, which depicted the mechanism by which SHMT2 catalyzes glycine hydroxymethylation of EGFR protein GS islands, leading to G>S editing (G>S point, GS>SG switch, or GS>SG SENA), independent of EGFR gene mutation. The diagram also showed that G>S editing sites can be phosphorylated in response to EGF treatment.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [EGFREditingExtF123456.pdf](#)