Expanding the Genetic Code of *Escherichia coli* with Phosphoserine

Hee-Sung Park, *et al.*

Science **333**, 1151 (2011);
DOI: 10.1126/science.1207203

Expanding the Genetic Code of *Escherichia coli* with Phosphoserine

Hee-Sung Park,[1*†] Michael J. Hohn,[1*] Takuya Umehara,[1] Li-Tao Guo,[1] Edith M. Osborne,[2‡] Jack Benner,[2] Christopher J. Noren,[2§] Jesse Rinehart,[3,4§] Dieter Söll[1,5§]

O-Phosphoserine (Sep), the most abundant phosphoamino acid in the eukaryotic phosphoproteome, is not encoded in the genetic code, but synthesized posttranslationally. Here, we present an engineered system for specific cotranslational Sep incorporation (directed by UAG) into any desired position in a protein by an *Escherichia coli* strain that harbors a Sep-accepting transfer RNA (tRNA^{Sep}), its cognate Sep–tRNA synthetase (SepRS), and an engineered EF-Tu (EF-Sep). Expanding the genetic code rested on reengineering EF-Tu to relax its quality-control function and permit Sep-tRNA^{Sep} binding. To test our system, we synthesized the activated form of human mitogen-activated ERK activating kinase 1 (MEK1) with either one or two Sep residues cotranslationally inserted in their canonical positions (Sep^{218}, Sep^{222}). This system has general utility in protein engineering, molecular biology, and disease research.

O-Phosphoserine (Sep) was identified 80 years ago as a constituent of phosphoproteins from egg yolk (*1*). Since then, the extent and importance of the eukaryotic phosphoproteome has been realized and has provided insight into large interconnected networks of kinases and phosphatases (*2*). Protein kinases represent one of the largest eukaryotic gene families, composing nearly 2% of all human genes (*3*). Sep is the most abundant phosphoamino acid; based on an analysis of >2000 HeLa cell phosphoproteins, the relative abundance of Sep is 7.3 and 48 times higher than that of phosphothreonine and phosphotyrosine, respectively (*2*). A major research limitation is the inability to biosynthesize these phosphoproteins for detailed studies of their enzyme or substrate properties.

The discovery of Sep–tRNA synthetase (SepRS), a unique aminoacyl–tRNA synthetase devoted to Sep-tRNA^{Cys} formation in methanogenic archaea, provided an opportunity to develop our Sep-insertion strategy. The natural role of SepRS is the formation of Sep-tRNA^{Cys}, which is then converted to Cys-tRNA^{Cys} by the enzyme SepCysS in the presence of a sulfur donor (*4*) (Fig. 1A). Given the high specificity of SepRS for Sep and for tRNA^{Cys} and our knowledge of the identity elements in this tRNA (*5*), and based on our understanding of the structure of this enzyme and its catalytic site (*6*, *7*), we devised a system to incorporate Sep into proteins directed by the UAG (amber) codon. For this, we chose *Methanocaldococcus jannaschii* (Mj) tRNA^{Cys} and the mesophilic *Methanococcus maripaludis* (Mmp) SepRS as the orthogonal pair [reviewed in (*8*)] for the synthesis of phosphoserylated amber suppressor tRNA.

We first designed tRNA^{Sep} (Fig. 1B), an amber suppressor tRNA derived from Mj tRNA^{Cys} by two mutations in the anticodon, and an additional C20U change that improves aminoacylation by SepRS (*5*). In vitro aminoacylation by Mmp SepRS showed (Fig. 1C) that the anticodon change lowered (to about 40%) the ability of tRNA^{Sep} to be aminoacylated when compared to tRNA^{Cys}. In agreement with earlier data (*5*), total *Escherichia coli* tRNA could not be acylated with Sep (Fig. 1C). On the basis of these in vitro data, Mj tRNA^{Sep} and Mmp SepRS appear to be an orthogonal pair.

Efficient and selective addition of Sep to the *E. coli* genetic repertoire requires exclusive interaction of SepRS with tRNA^{Sep} for Sep-tRNA^{Sep} formation without interfering in the host translation system, as well as a sufficient intracellular concentration of Sep. Because *E. coli* has a Sep-compatible transporter (*9*), Sep (2 mM) was added to the LB growth medium, and the endogenous *serB* gene encoding phosphoserine phosphatase was deleted in the *E. coli* test strain without affecting growth. To assess whether the Mj tRNA^{Sep}–Mmp SepRS pair is functional and orthogonal in *E. coli* in vivo, we performed a suppression assay that used a gene encoding chloramphenicol acetyltransferase (CAT) with a UAG stop codon at the permissive position 112 (wild-type amino acid: Asp) to produce the CAT enzyme. Cell survival was measured in the presence of Sep and varying amounts of chloramphenicol (Cm) where the different half-maximal inhibitory concentration (IC_{50}) values and the

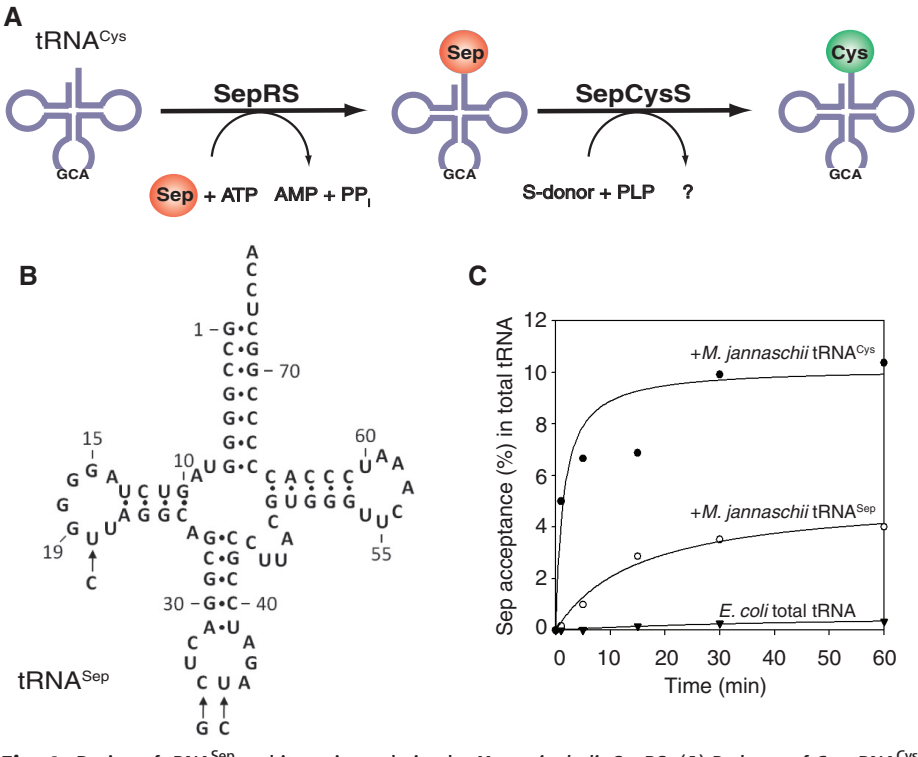

Fig. 1. Design of tRNA^{Sep} and its aminoacylation by *M. maripaludis* SepRS. (**A**) Pathway of Cys-tRNA^{Cys} formation in *M. maripaludis*. ATP, adenosine 5′-triphosphate; AMP, adenosine 5′-monophosphate; PPi, inorganic pyrophosphate; PLP, pyridoxal phosphate. (**B**) Cloverleaf structure of tRNA^{Sep}. Arrows indicate the three nucleotide changes compared to *M. jannaschii* tRNA^{Cys}. (**C**) Acylation with Sep of *M. jannaschii* tRNA^{Cys} and tRNA^{Sep} catalyzed by *M. maripaludis* SepRS. Total *E. coli* tRNA (▼), or tRNA from *E. coli* strains expressing the *M. jannaschii* tRNA^{Cys} (●) or the tRNA^{Sep} (○) gene was acylated by *M. maripaludis* SepRS with [^{14}C]Sep (0.1 mM) in the presence of ATP (10 mM).

[1]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. [2]New England BioLabs, Ipswich, MA 01938, USA. [3]Department of Cellular and Molecular Physiology, Yale University, New Haven, CT 06520, USA. [4]Systems Biology Institute, Yale University, West Haven, CT 06516, USA. [5]Department of Chemistry, Yale University, New Haven, CT 06520, USA.

*These authors contributed equally to this work.
†Present address: Department of Chemistry, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea.
‡Present address: Department of Chemistry, Angelo State University, San Angelo, TX 76909, USA.
§To whom correspondence should be addressed. E-mail: dieter.soll@yale.edu (D.S.); noren@neb.com (C.J.N.); jesse.rinehart@yale.edu (J.R.)

tRNA^{Sep}-dependent CAT synthesis correlate with suppression efficiency (Fig. 2). When only tRNA^{Sep} was expressed (Fig. 2, column B), Cm resistance increased about 3.3-fold over background (Fig. 2, column A). Thus, tRNA^{Sep} can be aminoacylated to a certain degree by an unknown *E. coli* aminoacyl–tRNA synthetase (we later found that Gln is being incorporated at the amber stop codon). In contrast, simultaneous expression of tRNA^{Sep} and SepRS did not provide Cm resistance (Fig. 2, column C). This may indicate that SepRS can outcompete any endogenous aminoacyl–tRNA synthetase and form Sep-tRNA^{Sep}; however, this aminoacyl-tRNA is neither delivered to the ribosome nor accommodated on it. Providing additional EF-Tu did not improve the result (Fig. 2, column D). Coexpression of tRNA^{Sep}, SepRS, and SepCysS should result in formation of Sep-tRNA^{Sep} and subsequent SepCysS-mediated conversion to Cys-tRNA^{Sep} (*4*). Indeed, a 2.3-fold increase in Cm resistance was observed (Fig. 2, column E). This further supports the notion that although Sep-tRNA^{Sep} is synthesized, it cannot be used properly by the *E. coli* protein biosynthesis machinery. By contrast, coexpression of tRNA^{Sep} and Mmp CysRS generated a 12.3-fold increase in Cm resistance (Fig 2, column F), demonstrating that Cys-tRNA^{Sep} can be readily used for amber codon suppression in the CAT gene.

Given that EF-Tu is a component of quality control in protein synthesis (*10*), it is plausible that Sep-tRNA^{Sep} may be rejected by EF-Tu. Chemically synthesized Sep-tRNA^{Gln} was a poor substrate for in vitro protein synthesis (*11*). tRNAs carrying negatively charged amino acids are bound poorly by EF-Tu (*12*), and molecular dynamics simulations suggested that Sep-tRNA^{Cys} may not be bound by EF-Tu (*13*). We tested this assumption in EF-Tu–mediated Sep-tRNA hydrolysis protection experiments (*14*). Although EF-Tu protected [^{35}S]Cys-tRNA^{Cys} from deacylation at pH 8.2 (fig. S1), [^{14}C]Sep-tRNA^{Cys} was significantly deacylated irrespective of the presence of EF-Tu (Fig. 3B and fig. S1). Thus, insufficient binding of Sep-tRNA^{Sep} to EF-Tu may explain the lack of Sep insertion into protein.

This observation required the generation of EF-Tu variants able to productively bind Sep-tRNA. We were encouraged by reports that EF-Tu variants allow binding of tRNAs charged with certain unusual amino acids (*15*, *16*). Guided by the structure of the *E. coli* EF-Tu:Phe-tRNA^{Phe} complex (*17*), we selected six residues (His^{67}, Asp^{216}, Glu^{217}, Phe^{219}, Thr^{229}, and Asn^{274}) in the amino acid binding pocket of EF-Tu (Fig. 3A) for complete randomization in order to generate EF-Tu variants that bind Sep-tRNA. Variants that permitted SepRS and tRNA^{Sep}-dependent Sep incorporation were selected in vivo (see SOM). One clone, designated EF-Sep (amino acid variants shown in Fig. 3A), led to a 10-fold increase in Cm resistance (Fig. 2, column H), whereas the combination of SepRS and EF-Sep without tRNA^{Sep} was not active in the CAT suppression assay (Fig. 2, column G). Thus, it appeared that EF-Sep did bind Sep-tRNA^{Sep}, a conclusion that was confirmed in the hydrolysis protection assay (Fig. 3B). This assay also shows that EF-Sep still retained some ability to bind Cys-tRNA (fig. S1).

To prove that the observed suppression is due to Sep incorporation, we expressed myoglobin with an amber codon in the Asp^{127} position (fig. S2A). The expected full-length protein was synthesized only when EF-Sep, SepRS, and tRNA^{Sep} were coexpressed (fig. S2A). Mass spectrometry–time-of-flight (MS-TOF) and MS/MS analysis showed that Sep is present at the position specified by UAG in both the intact and trypsin-digested proteins (Fig. S2, B and C).

Final validation of our strategy was the synthesis of a Sep-containing human protein MEK1 (mitogen-activated ERK activating kinase 1). This key eukaryotic enzyme of the mitogen-activated

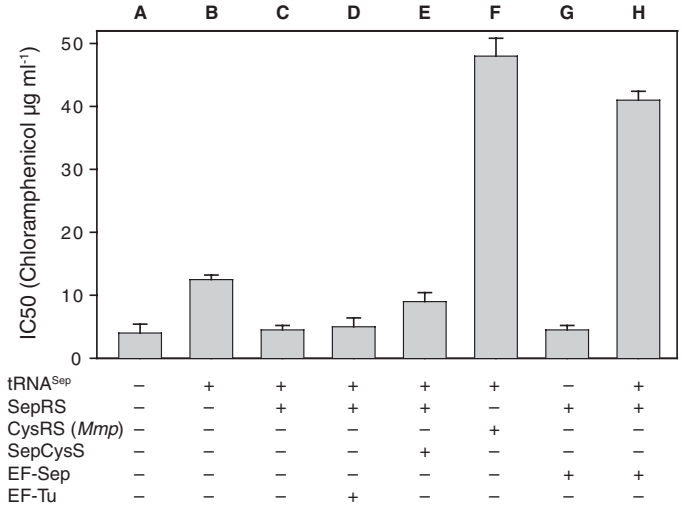

Fig. 2. In vivo synthesis of chloramphenicol acetyltransferase (measured by IC_{50} value) by tRNA^{Sep}-dependent read-through of an amber mutation in the CAT gene. tRNA^{Sep} was coexpressed in the *E. coli* Top10Δ*serB* strain with the proteins indicated in the figure (SepRS, *M. maripaludis* CysRS, SepCysS, EF-Sep, EF-Tu). Selection was carried out on LB agar plates containing 2 mM Sep and various concentrations of chloramphenicol. Error bars indicate SEM.

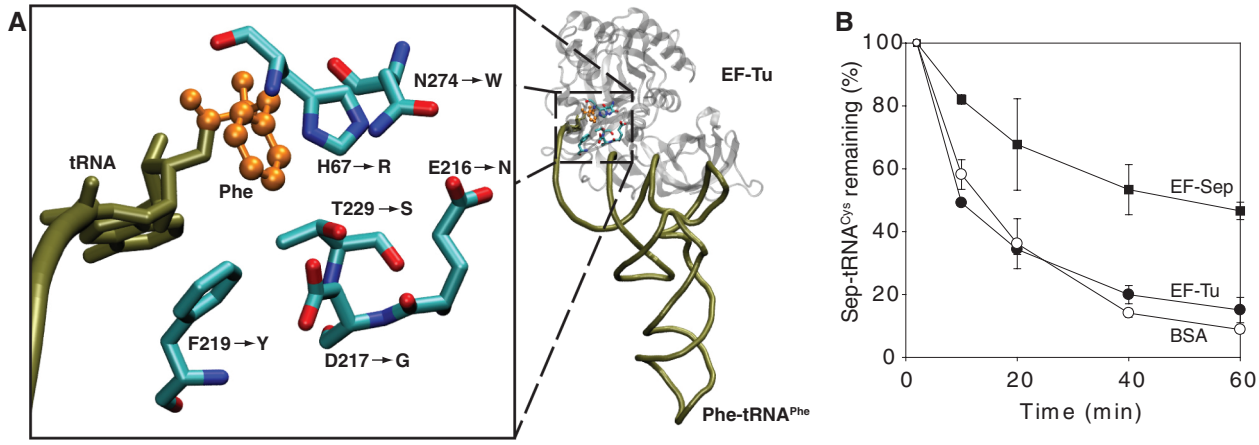

Fig. 3. Design of EF-Sep. (**A**) Model of the amino acid binding pocket of *E. coli* EF-Tu bound to Phe-tRNA (based on Protein Data Bank structure 1OB2). To accommodate Sep-tRNA, an *E. coli tufB* library was constructed that would allow the six highlighted amino acid residues to change to any of the 20 canonical amino acids. The six mutations in our final EF-Sep are indicated by an arrow. (**B**) EF-Sep protects Sep-tRNA^{Cys} from deacylation. *M. jannaschii* [^{14}C]Sep-tRNA^{Cys} was incubated at pH 8.2 and at room temperature in the presence or absence of wild-type EF-Tu or EF-Sep. Deacylation of Sep-tRNA^{Cys} was measured by acid precipitability. Error bars indicate SEM. Abbreviations for the amino acid residues are as follows: D, Asp; E, Glu; F, Phe; G, Gly; H, His; N, Asn; R, Arg; S, Ser; T, Thr; W, Trp; and Y, Tyr.

signaling cascade is crucial for cell proliferation, development, differentiation, cell cycle progression, and oncogenesis (*18*). Activation of MEK1 requires posttranslational phosphorylation of Ser^{218} and Ser^{222} by MEK activating kinases (e.g., Raf-1, MEKK, or MOS). Substitution of both Ser residues with Glu yields a constitutively active enzyme, albeit with lower activity (*19*). We generated a clone encoding a MEK1 fusion protein [with the maltose binding protein (MBP) at the N terminus and a His_6 tag at the C terminus] in which Ser^{222} was changed to Glu, and the Ser^{218} codon was replaced by UAG to encode Sep. After expression in the presence of SepRS, $tRNA^{Sep}$, and EF-Sep, 25 μg of full-length MBP-MEK1(Sep^{218},Glu^{222}) was isolated from 1 liter of culture. The presence of Sep in MBP-MEK1(Sep^{218},Glu^{222}) protein was demonstrated by its ability to phosphorylate ERK2, which then phosphorylates myelin basic protein. MBP-MEK1(Sep^{218},Glu^{222}) had a 2500-fold higher specific activity than MBP-MEK1(Ser^{218},Ser^{222}) and a 70-fold higher specific activity than the constitutively active MBP-MEK1(Glu^{218},Glu^{222}) (Fig. 4A). MS/MS analysis confirmed the incorporation of Sep at position 218 (Fig. 4B). To determine if our *E. coli* expression system would allow the simultaneous insertion of two Sep residues into the protein, we changed the Ser codons in positions 218 and 222 to UAG. As expected, the expression efficiency of MBP-MEK1(Sep^{218},Sep^{222}) was markedly reduced compared to that of wild-type MBP-MEK1 (only about 1 μg of full-length protein was obtained from 1 liter of culture). The presence of Sep at both active-site positions of MEK1 was tested by Western blot analysis with a monoclonal antibody specific to phosphorylation at these two residues (Fig 4C). Only MBP-MEK1(Sep^{218},Sep^{222}), and to a weaker extent MBP-MEK1(Sep^{218},Ser^{222}), were detected in this experiment, whereas neither MBP-MEK1(Ser^{218},Ser^{222}), MBP-MEK(Sep^{218},Glu^{222}), or MBP-MEK(Glu^{218},Glu^{222}) was recognized by the antibody. This demonstrates that the addition of SepRS, $tRNA^{Sep}$, and EF-Sep endows *E. coli* with the ability to read UAG as a phosphoserine codon.

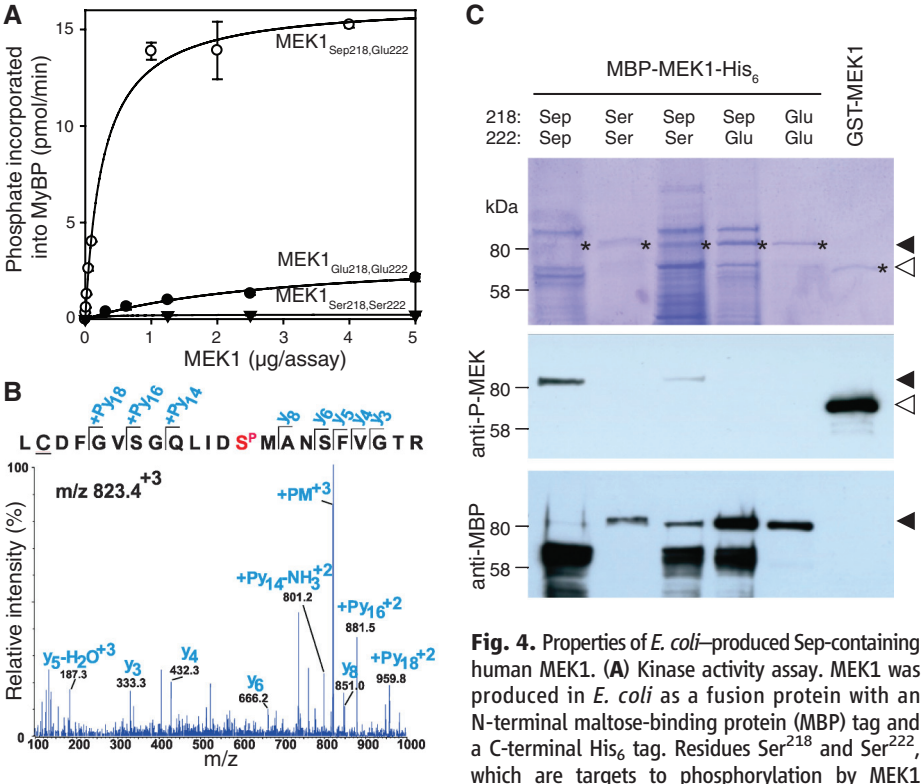

Fig. 4. Properties of *E. coli*–produced Sep-containing human MEK1. (**A**) Kinase activity assay. MEK1 was produced in *E. coli* as a fusion protein with an N-terminal maltose-binding protein (MBP) tag and a C-terminal His_6 tag. Residues Ser^{218} and Ser^{222}, which are targets to phosphorylation by MEK1 activators, were mutated to either Glu^{218}/Glu^{222} or Sep^{218}/Glu^{222} to produce active MEK1 variants. Various amounts of MBP-MEK1-His_6 with active-site residues Ser^{218}/Ser^{222} (▼), Glu^{218}/Glu^{222} (●), or Sep^{218}/Glu^{222} (○) were used to phosphorylate inactive ERK2 in vitro. ERK2 activity was then measured in a radiometric assay using γ-[^{32}P]ATP and myelin basic protein (MyBP) as substrates. Error bars indicate SEM. (**B**) MS/MS spectrum confirming the presence of Sep^{218} in MBP-MEK1(Sep^{218}/Ser^{222}). *m/z*, mass-to-charge ratio. (**C**) MBP-MEK1-His_6 variants with genetically encoded active-site residues Sep^{218}/Sep^{222} (lane 1), Ser^{218}/Ser^{222} (lane 2), Sep^{218}/Ser^{222} (lane 3), Sep^{218}/Glu^{222} (lane 4), and Glu^{218}/Glu^{222} (lane 5) were produced in *E. coli* and partially purified by Ni^{2+} affinity chromatography. Proteins were separated by SDS–polyacrylamide gel electrophoresis and either stained with Coomassie (top) or transferred to a nylon membrane and detected with monoclonal antibodies specific for the phosphorylated active site of human MEK (center) or the MBP tag (bottom). Purchased activated glutathione *S*-transferase (GST)–MEK was used as a control (lane 6). Dark and light arrowheads indicate the positions of MBP-MEK1-His_6 and GST-MEK1, respectively. The strong bands (~70 kD size) are probable truncation products caused by termination at UAG.

Our work underscores the key role of EF-Tu in quality control of protein synthesis by ensuring facile delivery of the correct cognate aminoacyl-tRNA to the ribosome (*10*, *20*). Generating an orthogonal aminoacyl–tRNA synthetase:tRNA pair was insufficient to genetically encode Sep. Expansion of the genetic code to include Sep depended critically on reengineering of EF-Tu to bind Sep-$tRNA^{Sep}$. This situation is precisely paralleled in the naturally evolved genetic encoding system for selenocysteine, which requires a specialized elongation factor [SelB in prokaryotes, and EFSec in eukaryotes; reviewed in (*21*)]. Inspired careful manipulation of components of the protein-synthesizing system will allow further expansion of the genetic code without sacrificing organismal fitness.

Orthogonal aminoacyl–tRNA synthetase:tRNA pairs (in our case, SepRS:$tRNA^{Sep}$) are critical elements for genetic code expansion, whether produced by evolutionary processes and found in nature [e.g., pyrrolysyl–tRNA synthetase:$tRNA^{Pyl}$ (*22*)] or designed in the laboratory by synthetic biologists [reviewed in (*8*)]. The tRNA of choice is typically an amber suppressor tRNA; however, a general limitation in product yield is associated with recoding the new amino acid by a stop codon (e.g., UAG), because peptide chain elongation by the designed aminoacylated suppressor tRNA competes with chain termination by the release factor (e.g., RF1). Yet this impediment may soon be removed by the advent of *E. coli* expression strains with genome-wide UAG codon reassignments and an RF1 deletion (*23*, *24*).

The ability to generate physiologically relevant active kinases and stoichiometrically phosphorylated protein domains could reveal new types of kinase inhibitors for drug development, allow systematic dissection of phosphorylation-dependent protein-protein interactions, and expose new structure-function relationships (*18*, *25*, *26*).

References and Notes

1. F. Lipmann, *Biochem. Z.* **262**, 3 (1933).
2. J. V. Olsen *et al.*, *Cell* **127**, 635 (2006).
3. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, *Science* **298**, 1912 (2002).
4. A. Sauerwald *et al.*, *Science* **307**, 1969 (2005).
5. M. J. Hohn, H. S. Park, P. O'Donoghue, M. Schnitzbauer, D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18095 (2006).
6. S. Kamtekar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2620 (2007).
7. R. Fukunaga, S. Yokoyama, *Nat. Struct. Mol. Biol.* **14**, 272 (2007).
8. C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **79**, 413 (2010).
9. B. L. Wanner, W. W. Metcalf, *FEMS Microbiol. Lett.* **79**, 133 (1992).
10. F. J. LaRiviere, A. D. Wolfson, O. C. Uhlenbeck, *Science* **294**, 165 (2001).
11. D. M. Rothman *et al.*, *J. Am. Chem. Soc.* **127**, 846 (2005).
12. T. Dale, L. E. Sanderson, O. C. Uhlenbeck, *Biochemistry* **43**, 6159 (2004).
13. J. Eargle, A. A. Black, A. Sethi, L. G. Trabuco, Z. Luthey-Schulten, *J. Mol. Biol.* **377**, 1382 (2008).
14. J. Ling *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15299 (2007).
15. Y. Doi, T. Ohtsuki, Y. Shimizu, T. Ueda, M. Sisido, *J. Am. Chem. Soc.* **129**, 14458 (2007).
16. T. Ohtsuki, H. Yamamoto, Y. Doi, M. Sisido, *J. Biochem.* **148**, 239 (2010).

17. P. Nissen *et al.*, *Science* **270**, 1464 (1995).
18. J. S. Sebolt-Leopold, R. Herrera, *Nat. Rev. Cancer* **4**, 937 (2004).
19. D. R. Alessi *et al.*, *EMBO J.* **13**, 1610 (1994).
20. J. M. Schrader, S. J. Chapman, O. C. Uhlenbeck, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5215 (2011).
21. S. Yoshizawa, A. Böck, *Biochim. Biophys. Acta* **1790**, 1404 (2009).
22. K. Nozawa *et al.*, *Nature* **457**, 1163 (2009).
23. T. Mukai *et al.*, *Nucleic Acids Res.* **38**, 8188 (2010).
24. F. J. Isaacs *et al.*, *Science* **333**, 348 (2011).
25. J. D. Scott, T. Pawson, *Science* **326**, 1220 (2009).
26. M. B. Yaffe *et al.*, *Cell* **91**, 961 (1997).

Acknowledgments: We thank P. Dennis, P. O'Donoghue, and J. Ling for enthusiastic discussions. M.J.H. was a Feodor Lynen Postdoctoral Fellow of the Alexander von Humboldt Foundation (Bonn, Germany). H.-S.P. held a postdoctoral fellowship of the Korean Science Foundation. This work was supported by grants from NSF (MCB-0645283 and MCB-0950474) (to D.S.), National Institute of General Medical Sciences (GM 22854) (to D.S.), National Research Foundation of Korea (ABC-20100029737) (to H.-S.P.), and National Institute of Diabetes and Digestive and Kidney Diseases (K01DK089006) (to J.R.). Yale University holds the U.S. Patent 7,723,069 B2: "Site Specific Incorporation of Phosphoserine into Polypeptides Using Phosphoseryl-tRNA Synthetase" by D. Soll and J. Rinehart. Yale University has applied for a patent that covers the engineered EF-Tu described in this manuscript. Reagents are available under a Yale Material Transfer Agreement.

Supporting Online Material
www.sciencemag.org/cgi/content/full/333/6046/1151/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References (*27*–*37*)

19 April 2011; accepted 7 July 2011
10.1126/science.1207203

Exome Sequencing of Head and Neck Squamous Cell Carcinoma Reveals Inactivating Mutations in *NOTCH1*

Nishant Agrawal,[1,2]*† Mitchell J. Frederick,[3]* Curtis R. Pickering,[3]* Chetan Bettegowda,[2,4]* Kyle Chang,[5] Ryan J. Li,[1] Carole Fakhry,[1] Tong-Xin Xie,[3] Jiexin Zhang,[6] Jing Wang,[6] Nianxiang Zhang,[6] Adel K. El-Naggar,[7] Samar A. Jasser,[3] John N. Weinstein,[6] Lisa Treviño,[5] Jennifer A. Drummond,[5] Donna M. Muzny,[5] Yuanqing Wu,[5] Laura D. Wood,[8] Ralph H. Hruban,[8] William H. Westra,[8] Wayne M. Koch,[1] Joseph A. Califano,[1,9] Richard A. Gibbs,[5,9] David Sidransky,[1] Bert Vogelstein,[2] Victor E. Velculescu,[2]† Nickolas Papadopoulos,[2] David A. Wheeler,[5] Kenneth W. Kinzler,[2]† Jeffrey N. Myers[3]†

Head and neck squamous cell carcinoma (HNSCC) is the sixth most common cancer worldwide. To explore the genetic origins of this cancer, we used whole-exome sequencing and gene copy number analyses to study 32 primary tumors. Tumors from patients with a history of tobacco use had more mutations than did tumors from patients who did not use tobacco, and tumors that were negative for human papillomavirus (HPV) had more mutations than did HPV-positive tumors. Six of the genes that were mutated in multiple tumors were assessed in up to 88 additional HNSCCs. In addition to previously described mutations in *TP53*, *CDKN2A*, *PIK3CA*, and *HRAS*, we identified mutations in *FBXW7* and *NOTCH1*. Nearly 40% of the 28 mutations identified in *NOTCH1* were predicted to truncate the gene product, suggesting that *NOTCH1* may function as a tumor suppressor gene rather than an oncogene in this tumor type.

More than half a million new cases of head and neck squamous cell carcinoma (HNSCC) will occur in 2011, including 50,000 cases in the United States, making it the sixth most common cancer in the world (*1*–*3*). HNSCC and its treatment can result in cosmetic deformity and functional impairment of vital functions, including breathing, swallowing, speech, phonation, taste, hearing, and smell. These cancers are frequently lethal, with a five-year survival of only ~50% (*4*). HNSCCs, like all solid tumors, are thought to be initiated and to progress through a series of genetic alterations. Indeed, several cellular signaling pathways are dysregulated in this tumor type through genetic and epigenetic alterations, such as those involving *TP53* and *CDK2NA* (*4*). HNSCCs also exhibit many chromosomal abnormalities, including amplifications of region 11q13, which contains the *cyclin D1* gene, and region 7p11, which encodes *epidermal growth factor receptor* (*EGFR*) (*5*). Tobacco use and excessive alcohol consumption are major risk factors for HNSCC in the United States (*6*). More recently, human papillomavirus (HPV) has emerged as an additional risk factor for the development of cancers of the oropharynx (*7*). Patients with HPV-associated cancers have an improved overall and disease-specific survival, suggesting that these tumors have distinct biological features (*8*).

To gain a comprehensive view of the genetic alterations underlying HNSCC, we sequenced ~18,000 protein-encoding genes in tumors from 32 patients. Thirty of the 32 patients had not been treated with chemotherapy or radiation before their tumor biopsy, so the spectrum of changes we observed largely reflects those of tumors in their naturally occurring state. Tumor samples were carefully selected or microdissected in order to achieve a neoplastic cellularity of >60%. DNA was purified from these tumors as well as matched nonneoplastic tissue and used to generate libraries suitable for massively parallel sequencing. After capture of the coding sequences with a SureSelect (Agilent, Santa Clara, California) or CCDS (NimbleGen, Madison, Wisconsin) Enrichment System, the DNA was sequenced by using an Illumina (San Diego, California) GAIIx/HiSEq (17 tumors) or SOLiD (Carlsbad, California) V3/V4 (15 tumors) instruments. The average coverage of each base in the targeted regions was 77-fold and 44-fold for the Illumina and SOLiD instruments, and 92.6% and 90% of targeted bases were represented by at least 10 reads in these platforms, respectively (table S1).

Using stringent criteria for analysis of these data (*9*), we identified 911 candidate somatic mutations in 725 genes among the 32 tumors. To ensure that our algorithms for identifying mutations were reliable, we evaluated the candidate mutations by Sanger sequencing or by 454 sequencing and confirmed 609 of them (67%) (table S2). One hundred and fifty-two (17%) mutations did not confirm, and 150 (16%) mutations could not be tested because of an unusually high GC content, difficulty in the design of unique primers, or other unknown factors preventing specific amplification and sequencing of the locus. The range of confirmed mutations per tumor was 2 to 78, with a mean and SD of 19 ± 16.5 mutations per tumor (table S1).

There were obvious differences in the genetic landscapes of HPV-associated and HPV-negative HNSCCs. First, far fewer genes were mutated per tumor in the HPV-associated tumors as compared with those tumors not epidemiologically related

[1]Department of Otolaryngology–Head and Neck Surgery, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287, USA. [2]Ludwig Center for Cancer Genetics and Howard Hughes Medical Institutions, Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. [3]Department of Head and Neck Surgery, University of Texas M.D. Anderson Cancer Center, 1515 Holcombe, Houston, TX 77030, USA. [4]Department of Neurosurgery, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287, USA. [5]Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. [6]Department of Bioinformatics and Computational Biology, University of Texas M.D. Anderson Cancer Center, 1515 Holcombe, Houston, TX 77030, USA. [7]Department of Pathology, University of Texas M.D. Anderson Cancer Center, 1515 Holcombe, Houston, TX 77030, USA. [8]Department of Pathology, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287, USA. [9]Milton J. Dance Head and Neck Center, Greater Baltimore Medical Center, Baltimore, MD 21204, USA. [10]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: nagrawal@jhmi.edu (N.A.); velculescu@jhmi.edu (V.E.V.); kinzlke@jhmi.edu (K.W.K.); jmyers@mdanderson.org (J.N.M.)