Title: Core cysteine residues in the PAN domain are critical for HGF/c-MET signaling

Authors: Debjani Pal[1†], Kuntal De[1†], Carly M. Shanks[1], Kai Feng[1], Timothy B. Yates[1,2], Jennifer Morrell-Falvey[1], Russell B. Davidson[1], Jerry M. Parks[1], Wellington Muchero[1,2*].

Affiliations:

[1] Bioscience Division, Oak Ridge National Laboratory; 1 Bethel Valley Rd, Oak Ridge, TN 37831, USA.

[2] Bredesen Center for Interdisciplinary Research; University of Tennessee, Knoxville, 37996, USA.

[†]These authors contributed equally

*Corresponding author. Email: mucherow@ornl.gov

Author contributions:

Conceptualization: DP, KD, CMS, WM

Methodology: DP, KD, KF, TBY, JM-F, RBD, JMP

Funding acquisition: WM

Writing – original draft: DP, KD, WM

Writing – review & editing: DP, KD, CMS, KF, TBY, JM-F, RBD, JMP, WM

Data and materials availability: All data are available in the main text or the supplementary materials.

Keywords: Signal transduction, Cancer, c-MET, HGF, Receptor tyrosine kinase

This PDF file includes:

Main Text

Materials and Methods

Acknowledgements

References

Legends for Figures 1 to 4

Figures 1 to 4

Notice: This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Abstract: The Plasminogen-Apple-Nematode (PAN) domain, with a core of four to six cysteine residues, is found in > 28,000 proteins across 959 genera but its role in protein function is not fully understood. The PAN domain was initially characterized to be present in numerous proteins including hepatocyte growth factor (HGF). Dysregulation of HGF-mediated signaling results in numerous deadly cancers. All biological impacts of HGF in cell proliferation are triggered by binding of HGF to its cell surface receptor, cellular mesenchymal-epidermal transition (c-MET). Here, we show that four PAN domain cysteine residues are essential for HGF/c-MET signaling. Mutating these residues resulted in retardation of perinuclear localization, cellular internalization of HGF and its receptor, c-MET, and c-MET ubiquitination. Our observations indicate that the PAN domain of HGF is required for the c-MET binding and subsequent c-MET autophosphorylation and phosphorylation of its downstream targets, protein kinase B (AKT), extracellular signal-regulated kinase (ERK), and signal transducer and activator of transcription 3 (STAT3). Furthermore, transcriptional activation of HGF/c-MET signaling-related genes including matrix metalloproteinase-9 (MMP9), ETS translocation variant 1, 4, and 5 (ETV1, ETV4, ETV5), and early growth response 1 (EGR1) was impaired and cell proliferation was attenuated. These results suggest that core cysteine residues in the PAN domain are critical for HGF/c-MET interaction, c-MET mediated signal transduction, and cell survival. Thus, targeting the PAN domain of HGF may represent a mechanism for selectively regulating the binding and activation of the c-MET pathway.

Significance: HGF/c-MET signaling induces multifunctional cellular responses. Dysregulation of HGF/c-MET signaling cascade can lead to tumorigenesis by transforming normal cells to tumor cells. This work defines the importance of core cysteine residues in the PAN domain of HGF in downstream activation of HGF/c-MET signaling. To understand the role of cysteines in the PAN

66 domain, PAN mutants of HGF were used to stimulate c-MET signaling in cells and the impact was
67 delineated by determining phosphorylation and transcription of downstream targets. Mutations in
68 core cysteines in the HGF-PAN domain completely blocked downstream phosphorylation and
69 perinuclear accumulation of c-MET. These results suggest an indispensable role for the cysteine-
70 rich PAN domain in HGF/c-MET interaction and could set the stage for future therapies that
71 selectively disrupt the MET signaling cascade with limited off-target effects in tumors
72 overexpressing HGF/c-MET.
73
74
75 **Main Text:**
76
77 **Introduction:**
78
79 Despite its prevalence in a wide variety of organisms, the role of the Plasminogen-Apple-
80 Nematode (PAN) domain in protein function has largely remained elusive. A key contributing
81 factor is that this domain is found in many phylogenetically unrelated proteins across divergent
82 organisms falling into categories including Alveolata, Archea, Amoebazoa, Bacteria, Cryptophyta,
83 Euglenozoa, Haptophyceae, Opisthokonta, Rhizaria, Rhodophyta, Stramenopilles, Viridplantae,
84 and Viruses. The PAN domain was first characterized by Tordai et al. (*1*) in which they noted that
85 the domain was shared by the plasminogen/hepatocyte growth factor protein family, the
86 prekallikrein/coagulation factor XI protein family, and nematode proteins. The domain possesses
87 the characteristic 4-6-cysteine residues in its core that are strictly conserved. These cysteine
88 residues have been proposed to engage in two or three disulfide bridges to form a hairpin loop
89 structure (*1, 2*). The PAN domain has been suggested to mediate protein-protein or carbohydrate-

90 protein interactions and facilitate receptor dimerization (*2-6*). HGF is secreted by mesenchymal
91 cells as a single-chain, biologically inert precursor and is converted into its bioactive form when
92 extracellular proteases cleave the bond between Arg494 and Val495 (*7*). The mature form of HGF
93 consists of an α- and β-chain, which are held together by a disulfide bond (*8-10*). The α-subunit of
94 HGF includes N-terminal hairpin loop structure and four kringle domains (K1-K4) whereas the β-
95 subunit consists of serine protease homology (SPH) domain (*11*).
96
97 To provide evidence for its role in signal transduction, we characterized the functional role of the
98 PAN domain in the heparin binding glycoprotein HGF, which functions as a ligand for the high-
99 affinity receptor, c-MET. HGF/c-MET signaling has been shown to mediate cellular processes
100 including angiogenesis, anti-apoptosis, mitogenesis, morphogenesis, mitogenesis and neurite
101 extension(*12*). Dysregulation of the HGF/c-MET signaling cascade can lead to tumorigenesis by
102 transforming normal cells into tumor cells and c-MET hyperactivation has been reported in cancers
103 including lung cancer, colorectal cancer, glioblastoma, and acute myeloid lymphoma among
104 others(*13, 14*).
105
106 An essential aspect of the HGF/c-MET signaling cascade is proteolytic processing of both HGF
107 and c-MET into their active forms, which can then form ligand-receptor complexes in the ECM.
108 Internalization and ubiquitin-dependent degradation of activated HGF and MET triggers signaling
109 cascades via a lysosomal degradative pathway that result in the biological cellular processes
110 described above. (*15, 16*). Previously, Okigaki et al. reported that deleting the N-terminal domain,
111 the first, second, or both kringle domains made HGF functionally inactive as the deletion mutants
112 were unable to trigger c-MET signaling (*17*).
113

114

115 **Results:**

116

117 **PAN domain-carrying proteins are enriched in cell recognition and cellular signaling**
118 **processes:**

119

120 PAN domain-carrying proteins do not share a clear evolutionary or phylogenetic trajectory
121 suggesting that this domain may have been co-opted independently by organisms to serve protein
122 functions that are yet to be revealed. To provide evidence of its putative function, we scanned the
123 InterPro (https://www.ebi.ac.uk/interpro/) and UniProt (https://G-LecRK.uniprot.org) databases
124 and found 28,300 proteins across 2,496 organisms falling into 959 genera (Table S1). Based on
125 predicted cellular localization, PAN domain proteins predominantly are present in the extracellular
126 matrix (ECM) either as cell surface receptors or ligands for cell surface receptors. Gene Ontology
127 (GO) enrichment analyses using the 28,300 proteins revealed 39 unique GO-terms that were
128 enriched at $p < 0.05$ (Table S2). These included terms such as cell recognition (*p-value* = 1E-30),
129 cell communication (*p-value* = 1E-30), proteolysis (*p-value* = 1E-30), pollen-pistil interaction (*p-*
130 *value* = 1E-30), reproduction (*p-value* = 1E-30), response to stimulus (*p-value* = 1E-30) and
131 response to stress (*p-value* = 1E-30). The predominant occurrence of these proteins in the ECM
132 and their enrichment in processes typically associated with cellular signal transduction suggested
133 that the PAN domain may play essential roles in immune response, albeit in different pathways for
134 divergent organisms. Based on these results, we hypothesized that the PAN domain would be
135 essential for HGF/c-MET signaling. To test this hypothesis, we designed experiments to mutate
136 core cysteine residues and assess implications on downstream signaling cascades.

137

138 **Core PAN domain cysteine residues modulate HGF overall-abundance:**

139

140 Alignment of the PAN domain of proteins from 14 model organisms revealed four strictly
141 conserved cysteine residues occurring at amino acid positions 70, 74, 84, and 96 of the HGF
142 protein (Fig. S1A&B; Fig. S2). To evaluate the functional significance of these residues, we
143 sequentially mutated single cysteines and also simultaneously mutated all four cysteine residues
144 and examined HGF stability in 293T cells. Mutant HGFs, in which cysteines 70 (C70A), 74
145 (C74A), 84 (C84A) and 96 (C96A) were substituted with alanine, led to a marked increase in the
146 protein expression of exogenously expressed HGFs in cells compared to their wild-type (WT)
147 counterpart (Fig. 1A-B). Furthermore, mutating all four core cysteines in the PAN domain
148 dramatically increased the abundance of exogenously expressed HGF-4Cys-4Ala in the cells (Fig.
149 1C). Moreover, the c-MET receptor exhibited a similar increase in abundance in response to
150 stimulation by the HGF 4Cys-4Ala mutant compared to WT HGF. These results suggest that
151 mutant HGF and its c-MET receptor did not undergo the expected lysosomal degradation when
152 the core cysteines were mutated.

153

154 The kringle domains of HGF have been reported to be crucial for protein-protein interactions. In
155 particular, the first and second kringle domains are especially important for the proper biological
156 function of the protein (*11*). Kringle domains are usually disulfide crosslinked domain and
157 previous studies have established that kringle domains in the α-subunit and the SPH domain in the
158 β-subunit provide c-MET binding sites on HGF (*18, 19*). Recently, Uchikawa et al. determined
159 structures of c-MET/HGF complex mimicking their active state at 4.8 Å resolution using cryo-
160 electron microscopy (*19*). They identified multiple distinctive c-MET binding sites on the HGF
161 protein including the N-terminal and kringle domains but did not implicate the PAN domain in the

162 HGF/c-MET interaction (*19*). Here, we exclusively investigated the role of the PAN domain of
163 HGF on the c-MET signaling cascade.
164
165 As we established a connection between the core cysteines in the PAN domain of HGF with HGF
166 abundance, we introduced four more mutations of additional cysteines located in the K1, K2, and
167 SPH domains of HGF (C128A, C149A, C271A, and C535A). None of these mutant HGFs showed
168 increased protein expression when expressed exogenously in cells as did the HGF 4Cys-4Ala (Fig.
169 1D and 1E). Thus, cysteine residues in the kringle and SPH domains are apparently not involved
170 in HGF activity. Rather, activity is specifically defined by the four strictly conserved cysteine
171 residues in PAN domain.
172
173 We hypothesized that the enhanced HGF expression of the 4Cys-4Ala mutant may results from an
174 intra-PAN domain structural change that prevents HGF degradation. To test this hypothesis, we
175 performed molecular dynamics (MD) simulations of the WT and 4Cys-4Ala PAN domains using
176 models generated with AlphaFold2 (*20*) and compared the resulting structural ensembles. Five
177 models were generated for each system and used as starting structures for the simulations. A 200-
178 ns simulation was then performed for each model. Thus, the cumulative simulation time was 1 μs
179 for each system. Analysis of the time evolution of the root-mean-square deviation (RMSD) of each
180 system revealed only minor structural deviations in both sets of simulations. Using the top WT
181 model as a reference structure, the average RMSDs and standard errors of the means were 1.61 +/-
182 0.02 Å for the WT and 1.71 +/- 0.01 Å for the 4Cys-4Ala system, indicating highly similar overall
183 structures despite the four Cys substitutions in the mutant (Figure S3). The structural analysis
184 indicates that no large-scale structural changes occurred in the 4Cys-4Ala mutant PAN domain
185 compared to the WT on this time scale (Figure S4).

186

187 Previous findings have suggested that, upon binding with its receptor, MET, HGF internalization
188 and degradation precedes activation of MET signaling. Translocation of MET-bound HGF toward
189 the perinuclear region is a key step for the degradation of HGF and recycling of the receptor (*7,*
190 *16, 21, 22*). The existence of a parallel pathway has also been reported in which HGF-activated
191 MET translocates to the nucleus to initiate calcium signaling (*23*). In agreement with a role for the
192 PAN domain in HGF abundance, we found that mutating the core cysteine residues reduced the
193 perinuclear signal for all mutated HGFs in HeLa cells (Fig. 1F). HeLa cells transfected with HGF
194 4Cys-4Ala showed less than 5% of the total perinuclear staining of recombinant HGF. However,
195 under the same transfection efficiency 80% of cells transfected with WT HGF showed perinuclear
196 staining for HGF under confocal microscopy (Fig. 1G). As such, both biochemical and
197 immunofluorescence results suggest a critical role for core PAN domain cysteine residues in HGF
198 stability and cellular uptake. Because all single cysteine mutants yielded the same results as the
199 4Cys-4Ala mutant, all subsequent studies were performed using only the recombinant HGF 4Cys-
200 4Ala variant.

201

202 **PAN domain cysteine residues are essential for c-MET, AKT and ERK phosphorylation:**

203

204 Following binding of HGF at the MET semaphoring homology (SEMA) domain, MET
205 homodimerizes and autophosphorylates at two tyrosine residues (Y1234 and Y1235), followed by
206 subsequent phosphorylation of two additional tyrosines in the carboxy-terminal tail (Y1349 and
207 Y1356). This series of events creates a multifunctional docking site for downstream adaptors and
208 effectors that lead to the activation of the Rat sarcoma (RAS)/ERK and phosphatidylinositol 3-
209 kinase (PI3K)/AKT axis via phosphorylation (*24, 25*). To determine the influence of core

210 cysteines on MET phosphorylation, both 293T and glioblastoma U-87 MG cells were stimulated
211 with purified wild-type HGF and HGF 4Cys-4Ala proteins. Western blot analyses revealed the
212 absence of phosphorylated MET, AKT, and ERK in unstimulated controls and in cells stimulated
213 with the HGF 4Cys-4Ala protein in both cell types (Fig. 2A).
214
215 To characterize the incompetence of PAN mutant HGF in turning on the phosphorylation cascade
216 of c-MET, we performed a direct interaction assay between endogenous c-MET and *in-vitro*
217 translated HGF proteins and showed that HGF 4Cys-4Ala was unable to interact with c-MET (Fig.
218 2B and 2C). Our results thus far indicate a central role for the PAN domain in its initial recognition
219 by c-MET. In addition, HGF 4Cys-4Ala, like the unstimulated controls, failed to promote MET
220 perinuclear translocation compared to wild-type HGF based on stimulation assays using HeLa
221 cells (Fig. 2D).
222
223 To gain additional insight into the impact of core cysteines in the HGF PAN domain on MET
224 degradation, we performed an *in-vivo* ubiquitination assay with c-MET. Ubiquitination was
225 inhibited with the alanine mutants of HGF compared to wild type (Fig. 2E). Based on these
226 observations, we propose that MET receptors are unable to become internalized by endocytosis,
227 suggesting that the mutated cysteine residues impose an overall retardation of MET activity and
228 its endocytic trafficking.
229
230
231 **Mutating core cysteine residues suppresses STAT3 phosphorylation and nuclear**
232 **translocation:**
233

234 STAT3 is a transcription factor that is present in the cytoplasm, forms dimers upon activation, and
235 functions as a downstream effector molecule of the HGF/c-MET signaling pathway (*26-28*).
236 STAT3 is reported to be constitutively active in several cancers, which leads to malignant
237 transformation by playing a critical role in stimulating cell proliferation and arresting apoptosis
238 (*28*). We tested the impact of the core cysteines in the HGF PAN domain on STAT3 activation by
239 following its phosphorylation in U-87 MG cells. STAT3 phosphorylation was significantly
240 reduced in cells post-stimulation with the HGF 4Cys-4Ala mutant compared to the wild type (Fig.
241 3A). Delayed time points were chosen for this assay because HGF induces delayed STAT3
242 phosphorylation (*29*). Confocal imaging confirmed that HGF 4Cys-4Ala, like the unstimulated
243 controls, was unable to promote STAT3 nuclear translocation, while wild-type HGF promoted
244 normal STAT3 nuclear localization (Fig, 3B). These results are consistent with the above
245 observations, suggesting that mutating the cysteine residues in the PAN domain leads to disrupted
246 HGF/c-MET signaling.
247
248 **PAN mutations downregulate HGF/c-MET-dependent cell proliferation and alters**
249 **expression of genes essential for diverse cellular responses:**
250
251 Dysregulated expression of HGF/c-MET acts as a catalyst in many cancers, with overexpression
252 of HGF often leading to aberrant cell proliferation and extracellular matrix invasion (*14*). Direct
253 evidence has been established connecting a primary role for HGF with increased expression of
254 MMP9, which is crucial for angiogenesis (*30*). HGF facilitates MMP9 expression via the
255 PI3K/AKT and p38 mitogen-activated protein kinases (MAPK) axis. Given the reported role of
256 HGF in modulating expression of MMP9, we evaluated the impact of mutated cysteine residues
257 on its transcriptional response. To evaluate the role of PAN domain mutation on the pattern of the

258 signature gene expression as well as cell proliferation, both 293T and U-87 MG cells were
259 stimulated with wild-type HGF or 4Cys-4Ala HGF for 24 hours following serum starvation. Cell
260 viabilities, as assessed by the MTT assay, were notably reduced by treatment with HGF 4Cys-
261 4Ala, whereas wild-type HGF could overcome serum starvation-mediated growth inhibition in
262 both cell types (Fig. S5A). 293T cells transiently transfected with Flag-WT HGF showed a
263 significant increase in cell proliferation over time compared to cells transfected with Flag-HGF
264 4Cys-4Ala (Fig. S6A). Quantitative real-time PCR (qRT-PCR) was performed to determine
265 whether MMP9 expression was similarly impacted. mRNA levels of MMP9 were markedly
266 decreased in HGF 4Cys-4Ala-stiumlated 293T and U-87 MG cells compared to the wild type in
267 both cell types (Fig. S5B). Similarly, we observed a significant decline in MET mRNA expression
268 in both cell types stimulated with HGF 4Cys-4Ala (Fig. S5B). Increased c-MET expression might
269 be a crucial determinant for the overall balance of the HGF/c-MET cascade in cells and could be
270 a trigger for the malignant transformation of normal cells. These data suggests that induction of
271 the MMP9 receptor is correlated with overall MET expression in an HGF-dependent manner and
272 could be regulated by minimally mutating the four core cysteines in the PAN domain of HGF.
273
274 Based on the apparent PAN domain-dependent transcriptional modulation of MMP9 expression
275 by HGF, we characterized the transcriptional responses of additional downstream genes in the
276 HGF/c-MET signaling pathway. Total RNA was extracted from 293T cells following a 24-hour
277 post treatment with wild-type HGF or 4Cys-4Ala HGF. Based on this analysis, we identified
278 significant differences in the expression of genes previously implicated in MET signaling, cell
279 cycle regulation, and cancer-related processes (Fig. 4; Fig. S6B). Specifically, we observed a
280 significant reduction in the expression levels of the downstream targets of the MET signaling
281 cascade, ETV1, ETV4 and ETV5, in cells stimulated with HGF 4Cys-4Ala compared to wild type

282 HGF. These transcription factors are members of the polyoma enhancer activator 3 (PEA3)
283 subgroup of the E–twenty six (ETS) family and confer resistance to early growth factor response
284 (EGFR) targeted therapy in lung cancer (*31*). Some known target genes of PEA3 are matrix
285 metalloproteinase-2 (MMP2), matrix metalloproteinase-7 (MMP7) and MMP9, which are well
286 recognized for their role in the invasiveness of cancer (*32*). Mutating the core cysteine residues in
287 the PAN domain led to reduced expression of EGR1, as well as the metalloproteinases, A
288 disintegrin and metalloprotease domain-containing protein 9 and 10 (ADAM9 and ADAM10), all
289 of which were previously linked to increased adhesion and cancer progression including,
290 hepatocellular carcinoma, and triple negative breast cancer via the AKT/NF-κB axis (*33-35*).
291 Hyperactivation of focal adhesion kinase (FAK) and PI3K/receptor for activated C kinase 1
292 (RAC1) are responsible for metastasis progression and mutations of the core cysteines in the HGF
293 PAN domain also resulted in their reduced expression. Apart from these examples, relative
294 expression of proteins involved in DNA damage repair (double-strand-break repair protein or
295 RAD21), chromosomal integrity and cohesion (structural maintenance of chromosomes protein
296 1A or SMC1A), cell cycle progression (kinesin family member 2A or KIF2A), posttranslational
297 modification, and RNA splicing (protein arginine N-methyltransferase 5 or PRTM5) were also
298 reduced. (Fig. 4) (*36-39*). Furthermore, this analysis revealed well-known cancer biomarkers
299 whose expression levels were affected by mutating the core cysteine residues. These biomarkers
300 included calponin-3 or CNN3 (marker for colorectal cancer), dolichyl-diphosphooligosaccharide-
301 protein glycosyltransferase subunit 2 or RPN2 (which inhibits autophagy and upregulates MMP9
302 expression), branched chain amino acid transaminase 1 or BCAT1 (which promotes hepatocellular
303 carcinoma and chemoresistance), ski interacting protein or SKIP, epithelial cell transforming
304 sequence 2 oncogene or ECT2 (overexpressed in different cancers including non-small cell lung

305 cancer), and elastin microfibril interfacer 2 or EMILIN2 (which promotes angiogenesis and
306 inflammation) (*40-45*).
307
308 **Discussion:**
309
310 In summary, this study provides evidence that the PAN domain of HGF provides the functional
311 catalytic core and plays a crucial role in c-MET binding. Further, GO enrichment analysis
312 predicted the strong association with cellular recognition and signaling processes of all the PAN
313 domain containing proteins, suggesting a role in immune signaling. To support these observations,
314 we demonstrated that mutating core cysteine residues in the HGF PAN domain results in a cascade
315 of negative regulation of HGF/c-MET signaling, which starts with retarded HGF degradation,
316 impaired c-MET interaction, disrupted perinuclear localization for both HGF and its receptor c-
317 MET, and is followed by a subsequent lack of phosphorylation for c-MET and its downstream
318 targets AKT, ERK and STAT3 and c-MET ubiquitination.
319
320 This disruption of upstream events was confirmed by the lack of transcriptional activation of
321 markers genes associated with cell migration and invasion including ADAM9/10, EGR1, ETV1,
322 ETV4, ETVE5 and MMP9 (Fig. S6B). Targeting the PAN domain in HGF could act as a multi-
323 pathway target as it entirely shuts down the c-MET signaling cascade and downstream expression
324 of relevant proteins known for their role in cancer prognosis and other diseases (Fig. S7). Taken
325 together, the PAN domain and its core cysteine residues are essential for HGF/c-MET signaling in
326 human cells. The observed enhanced stability of HGF with mutated cysteine residues is due to the
327 lack of its ability to bind with c-MET. These results are consistent with predictions based on GO
328 enrichment analysis in which the GO term *proteolysis* was highly enriched in the organismal

329 categories Alveolata, Amoebozoa, bacteria, opisthokonta, stramenopilles and viruses. This finding
330 suggests that the PAN domain may have been co-opted into different proteins along these lineages
331 to facilitate similar functions in the proteolytic processing of proteins involved in immune
332 signaling.
333
334 **Materials and Methods**
335
336 **PAN domain sequence alignment:** Proteins with PAN domains were identified from Uniprot and
337 were selected from 14 model organisms according to Bateman et al (*46*). The PAN domain
338 coordinates from UniProt were used to extract the corresponding sequences from the full-length
339 proteins, which were then aligned with MAFFT linsi(*47*). The alignment was visualized with
340 Geneious[41] and a phylogenetic tree accompanying the alignment was constructed via the neighbor-
341 joining method with Geneious (Table S3).
342
343 **GO enrichment of 14 PAN domain categories:** GO terms of each category were extracted from
344 InterPro2GO database from InterPro (*48*). GO enrichment analysis for each category against all
345 PAN domain genes was performed by Fisher's exact test via the TopGO package (*49*). Only the
346 biological process GO category was used for GO enrichment.
347
348 **Mammalian cell culture, transfection and drug treatment:** HeLa, HEK 293T and Glioblastoma
349 U87 cells were obtained from ATCC and maintained in a humidified atmosphere at 5% CO_2 in
350 Dulbecco's Modified Eagle's (DMEM) complete medium (Corning) supplemented with 10% fetal
351 bovine serum (FBS; Seradigm) in 37°C. Plasmid transfections were done with TransIT-LT1
352 (Mirus Bio) per the manufacturer's instructions.

353

354 **Plasmids and Recombinant Proteins:** Flag-HGF (GeneBankTM accession number
355 NM_000601.5) and various mutants of Flag-HGF, cloned into pcDNA 3.1 were obtained from
356 GenScript. Purified HGF protein and HGF 4Cys-4Ala protein were synthesized in GenScript. c-
357 MET-C-GFPSpark® Clone was obtained from SinoBiological.

358

359 **Immunofluorescence and confocal microscopy:** HeLa cells were seeded on coverslips in 24 well
360 plates. Where indicated, cells were transfected with Flag-HGF WT and different PAN domain
361 mutants for 36 hours, followed by fixation in 4% paraformaldehyde. Next, the cells were
362 permeabilized with 0.5% Triton X-100 in PBS, washed and then blocked for 30 minutes at room
363 temperature with 5% BSA in PBS. Cells were incubated with primary antibodies in 5% BSA in
364 1X-PBS with 0.5% Triton X-100 for 1 hour at room temperature. After washing the cells were
365 incubated with appropriate secondary antibodies in 5% BSA in PBST for 30 minutes at room
366 temperature. DNA was counterstained with 1 µg/mL Hoechst 33342 and mounted with
367 Fluorimount G (Southern Biotech). Cells were imaged using a Zeiss LSM 710 confocal
368 microscope. For MET and STAT3 localization assays, cells were serum starved for 24 hours before
369 the stimulations. Cells were then stimulated with 100 ng mL^{-1} of WT HGF and or 4Cys-4Ala HGF
370 purified proteins for the indicated amount of time at 37°C and treated as described above.

371

372 **Antibodies:** The following commercial antibodies and the indicated concentrations were used in
373 this study. HA antibody (HA.C5 #18181; 1:1000) was purchased from Abcam. Flag (#2368S;
374 1:1000), (Met (clone 25H2 #3127; 1:1000), Phospho-Met (Tyr 1234/1235) (clone D26 # 3077;
375 1:1000), Stat3 (clone 124H6 #9139; 1:1000), Phospho-Stat3 (Tyr 705) (clone D3H7 #9145;
376 1:1000), Akt (pan) (clone 40D4 #2920; 1:1000), Phospho Akt (Thr 308) (clone 244F9 #4056;

377 1:1000), p44/42 MAPK (Erk1/2) (#9102; 1:1000), Ubiquitin (clone E4I2J) (#43124; 1:1000), Met
378 (clone D1C2) (#8198; 1:1000) and Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (clone
379 D13.14.4E) (#4370; 1:1000) were purchased from Cell signaling. M2 anti Flag Mouse antibody
380 (#SLBT7654; 1:5000) and Actin (#087M4850; 1:10,000) were purchased from Sigma. HA
381 (#902302; 1:1000) antibody was purchased from Biolegend. Invitrogen GFP (clone A-11122)
382 (1:1000) was purchased from Thermo Fisher Scientific. Met Antibody (clone D-4) (#sc-514148;
383 1:500) was purchased from Santa Cruz Biotechnology. Chemiluminescence detection was
384 performed according to the manufacturer's instructions (Amersham ECL Western Blotting
385 Detection Reagent kit) followed by exposure using Chemidoc gel-documentation system
386 (BioRad). For imaging using Li-Cor, secondary antibodies for western blotting were purchased
387 from LI-COR Biosciences.
388
389 **Western Blotting and immunoprecipitation:** For in vivo stimulation experiments, cells were
390 grown for 36 hours and then stimulated with HGF WT and HGF 4Cys-4Ala where indicated (100
391 ng mL^{-1}), washed with PBS, and lysed. Briefly, cell extracts were generated on ice in EBC buffer,
392 50mM Tris (pH 8.0), 120mM NaCl, 0.5% NP40, 1mM DTT, and protease and phosphatase
393 inhibitors tablets (Thermo Fisher Scientific). Extracted proteins were quantified using the
394 PierceTM BCA Protein assay kit (Thermo Fisher). Proteins were separated by SDS acrylamide gel
395 electrophoresis and transferred to IMMOBILON-FL 26 PVDF membrane (Millipore) probed with
396 the indicated antibodies and visualized either by chemiluminescence (according to the
397 manufacturer's instructions) or using a LiCor Odyssey infrared imaging system.
398
399 For immunoprecipitation, endogenous c-MET was immunoprecipitated on MET antibody-bound
400 beads (Protein G from Thermo Fisher) and Flag-tagged HGF WT and HGF 4Cys-4Ala were *in-*

401 *vitro* translated (T_NT quick coupled Transcription/Translation system, Promega) and were
402 incubated with the bead bound c-MET for 4h at 4 ℃. Beads were then washed and proteins
403 resolved by SDS-PAGE and analyzed by western blotting as above.
404
405 **In vivo ubiquitination:**
406 293T cells were transfected with the construct encoding c-MET-C-GFPSpark. Cells were
407 stimulated with either HGF WT or HGF 4Cys-4Ala as indicated following serum starvation. Cells
408 were collected at indicated time points and washed with ice-cold PBS, lysed in ice-cold buffer
409 containing 10 mM Tris-HCl (pH 8), 150 mM NaCl, 0.1% SDS, 20mM NEM (*N*-ethylmaleimide)
410 supplemented with protease and phosphatase inhibitors tablets (Thermo Fisher Scientific). The
411 lysates were cleared by centrifugation at 10,000 × *g* at 4°C for 20 min, followed by preclearance
412 using protein G-beads (protein G from Thermo Fisher). 500 µg of each precleared lysate was
413 incubated with anti-MET antibody and protein-G beads for overnight at 4°C with rotation. The
414 samples were then washed three times with the lysis buffer and eluted with SDS-gel loading buffer
415 (with reducing agent added). Proteins were resolved by SDS-PAGE and immunoblotted with the
416 indicated antibodies.
417
418 **Cell Proliferation Assay:** 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT)
419 assay was used to determine cell viability. HEK293T and U-87-MG cells were plated in 96-well
420 plates with 500 cells per well in triplicate in serum-free medium for 24 hours prior to HGF
421 stimulations. 100 ng mL^{-1} HGF WT, HGF 4Cys-4Ala, or both were added to the cells and
422 incubated for 24 hours prior to the addition of MTT solution (Abcam, Inc # ab211091) and cell
423 viability was measured according to the manufacturer's instruction. The assays were performed in
424 triplicate and the experiment was repeated three times. Data were expressed as the mean ± SD.

425 Statistical analyses were performed. *P < 0.05 was considered to indicate a statistically significant
426 difference.
427
428 **Quantitative Real-Time PCR and RNA-Seq Analysis:** Total RNA was extracted from
429 glioblastoma U87 and HEK293T cell line 24 hours post HGF stimulation using TRIzol reagent
430 (Invitrogen), and reverse transcription was performed using the SuperScript II RT kit (Integrative
431 DNA Technologies) with total RNA (1 μg) according to the manufacturer's instructions. The MET
432 and MMP9 mRNA expression levels were detected by conventional RT-PCR with Taq DNA
433 Polymerase, Recombinant (Invitrogen, no. 10342-020). Glyceraldehyde-3-phosphate
434 dehydrogenase (GAPDH) was used as the internal control. The specific primers for MET, MMP9
435 and GAPDH were designed with Primer Premier software. The primers used were MET, forward:
436 5′-TTAAAGGAGACCTCACCATGTAATC-3′ and reverse: 5′-
437 CCTGATCGAGAAACCACAACCT -3′; MMP9, forward: 5′-
438 GATCCAAAACTACTCGGAAGACTTG-3′ and reverse: 5′-GAAGGCGCGGGCAAA-3′ and
439 GAPDH, forward: 5′-TTGCCATCAATGACCCCTTCA-3′ and reverse: 5′-
440 CGCCCCACTTGATTTTGGA-3′. The PCR reaction was performed according to the
441 manufacturer's instructions. The PCR conditions were as follows: amplification reaction protocol
442 was performed for 35 cycles consisting of 30 s at 94°C (denaturation), annealing 30 s at 45°C and
443 extension 30 s at 72°C.
444
445 For RNA-seq analysis, Total RNA was extracted from the HEK293T cell line 24 h post HGF
446 stimulation using TRIzol reagent (Invitrogen) according to the manufacturer’s instructions.
447 Quantification and quality control of isolated RNA was performed by measuring absorbance at
448 260 nm and 280 nm on a NANODROP ONEC spectrophotometer (Thermo Scientific, USA). The

449 RNA-seq run was performed with four biological replicates. Library prep and sequencing was
450 performed by BGI using the DNBSEQ-G400 platform which generated 100 bp paired-end reads.
451 The raw RNA-seq reads have been deposited at NCBI under BioProject ID PRJNA718097. Clean
452 reads were aligned to the human reference genome GRCh38. Reads were mapped with bowtie2
453 v2.2.5. (*50*) Expression levels for RNAs were calculated using fragments per kilobase per million
454 reads (FPKM) values with RSEM v1.2.8. (*51*) Differential expression analysis was performed with
455 DESeq2 and genes with an adjusted p-value less than 0.05 were considered differentially expressed
456 (*52*).
457
458 **Molecular Dynamics Simulations:** MD simulations were initiated from the top five models
459 generated with ColabFold (*20, 53*) of the WT and 4Cys-4Ala mutant PAN domain. The program
460 *tleap* from AmberTools20 (*54*) was used to prepare the parameter and coordinate files for each
461 structure. The ff14SB force field (*55*) and TIP3P water model (*56*) were used to describe the
462 protein and solvent, respectively. Energy minimization was performed using *sander* from
463 AmberTools20. At least a 12 Å solvent buffer between the protein and the periodic images. Sodium
464 and chloride ions were added to neutralize charge and maintain a 0.10 M ion concentration. The
465 simulations were performed with OpenMM version 7.5.1 (*57*)) on the Cuda platform (version
466 11.0.3) using Python 3.8.0. ParmEd was used to incorporate the force field parameters into the
467 OpenMM platform (*58*). The Langevin integrator and Monte Carlo barostat were used to maintain
468 the systems at 300 K and 1 bar, respectively. Direct non-bonded interactions were calculated up to
469 a 12-Å distance cutoff. All bonds involving hydrogen atom were constrained to their equilibrium
470 values. The particle mesh Ewald method was used to compute long-range Coulombic interactions
471 (*59*). A 2-fs integration time step was used with energies and positions written every 2 ps.
472

473 **Simulation Analysis:** Analyses of MD trajectories were performed using Python 3.8.0 and the
474 MDAnalysis version 1.0.1 (*60, 61*). Matplotlib was used to plot the data.
475
476 **Statistical Analysis:** Statistical analyses were performed on individual experiments, as indicated,
477 with GraphPad Prism 8 Software using an unpaired t-test, equal variance, for comparisons between
478 two groups. A P-value of *P < 0.05 was considered statistically significant.
479
480
481 **Acknowledgments:**
482
483 Bioinformatics analyses of PAN domain distribution and functional inference was supported by
484 the United States Department of Energy, Office of Science, Early Career Research Program under
485 the Biological and Environmental Research office. Biochemical, immunofluorescence and
486 transcriptome analyses in human cell lines, and protein modeling and simulation were supported
487 by the Lab Directed Research Development program at Oak Ridge National Laboratory (ORNL).
488 Part of this research used resources at the Oak Ridge Leadership Computing Facility (OLCF) and
489 the Compute and Data Environment for Science (CADES) at ORNL, which is managed by UT-
490 Battelle, LLC for the U.S. Department of Energy under Contract Number DE-AC05-00OR22725.
491
492 **Competing interests:** The authors declare that they have no competing interests.
493

494 **References:**

495

496 1. H. Tordai, L. Bányai, L. Patthy, The PAN module: the N-terminal domains of plasminogen
497 and hepatocyte growth factor are homologous with the apple domains of the prekallikrein
498 family and with a novel domain found in numerous nematode proteins. *FEBS letters* **461**,
499 63-67 (1999).
500 2. H. Zhou *et al.*, The solution structure of the N-terminal domain of hepatocyte growth factor
501 reveals a potential heparin-binding site. *Structure* **6**, 109-116 (1998).
502 3. H. Herwald *et al.*, Mapping of the discontinuous kininogen binding site of prekallikrein: a
503 distal binding segment is located in the heavy chain domain A4. *Journal of Biological*
504 *Chemistry* **271**, 13061-13067 (1996).
505 4. D. H. Ho, K. Badellino, F. A. Baglia, P. N. Walsh, A binding site for heparin in the apple
506 3 domain of factor XI. *Journal of Biological Chemistry* **273**, 16382-16390 (1998).
507 5. F. A. Baglia, K. O. Badellino, D. H. Ho, V. R. Dasari, P. N. Walsh, A binding site for the
508 kringle II domain of prothrombin in the apple 1 domain of factor XI. *Journal of Biological*
509 *Chemistry* **275**, 31954-31962 (2000).
510 6. S. Naithani, T. Chookajorn, D. R. Ripoll, J. B. Nasrallah, Structural modules for receptor
511 dimerization in the S-locus receptor kinase extracellular domain. *Proceedings of the*
512 *National Academy of Sciences* **104**, 12211-12216 (2007).
513 7. S. L. Organ, M. S. Tsao, An overview of the c-MET signaling pathway. *Ther Adv Med*
514 *Oncol* **3**, S7-s19 (2011).
515 8. E. Gohda *et al.*, Purification and partial characterization of hepatocyte growth factor from
516 plasma of a patient with fulminant hepatic failure. *J Clin Invest* **81**, 414-419 (1988).
517 9. T. Nakamura *et al.*, Molecular cloning and expression of human hepatocyte growth factor.
518 *Nature* **342**, 440-443 (1989).
519 10. R. Zarnegar, G. Michalopoulos, Purification and biological characterization of human
520 hepatopoietin A, a polypeptide growth factor for hepatocytes. *Cancer Res* **49**, 3314-3320
521 (1989).
522 11. K. Mizuno *et al.*, Hairpin loop and second kringle domain are essential sites for heparin
523 binding and biological activity of hepatocyte growth factor. *Journal of Biological*
524 *Chemistry* **269**, 1131-1136 (1994).
525 12. H. Funakoshi, T. Nakamura, Hepatocyte growth factor: from diagnosis to clinical
526 applications. *Clinica chimica acta* **327**, 1-23 (2003).
527 13. M. Jeffers *et al.*, Activating mutations for the Met tyrosine kinase receptor in
528 human cancer. *Proceedings of the National Academy of Sciences* **94**, 11445-11450 (1997).
529 14. G. Vande Woude *et al.*, in *Ciba Foundation Symposium 212-Plasminogen-Related Growth*
530 *Factors: Plasminogen-Related Growth Factors: Ciba Foundation Symposium 212*. (Wiley
531 Online Library, 2007), pp. 119-132.
532 15. M. J. Clague, Met receptor: a moving target. *Sci Signal* **4**, pe40 (2011).
533 16. D. Naka *et al.*, Internalization and degradation of hepatocyte growth factor in hepatocytes
534 with down-regulation of the receptor/c-Met. *FEBS Lett* **329**, 147-152 (1993).
535 17. M. Okigaki, M. Komada, Y. Uehara, K. Miyazawa, N. Kitamura, Functional
536 characterization of human hepatocyte growth factor mutants obtained by deletion of
537 structural domains. *Biochemistry* **31**, 9555-9561 (1992).

538 18. O. Holmes *et al.*, Insights into the structure/function of hepatocyte growth factor/scatter
539 factor from studies with individual domains. *J Mol Biol* **367**, 395-408 (2007).
540 19. E. Uchikawa, Z. Chen, G.-Y. Xiao, X. Zhang, X.-c. Bai, Structural basis of the activation
541 of c-MET receptor. *Nature Communications* **12**, 4074 (2021).
542 20. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
543 583-589 (2021).
544 21. M. Jafarnejad *et al.*, Mechanistically detailed systems biology modeling of the HGF/Met
545 pathway in hepatocellular carcinoma. *NPJ systems biology and applications* **5**, 1-14
546 (2019).
547 22. J. V. Abella *et al.*, Met/Hepatocyte growth factor receptor ubiquitination suppresses
548 transformation and is required for Hrs phosphorylation. *Mol Cell Biol* **25**, 9632-9645
549 (2005).
550 23. D. A. Gomes *et al.*, c-Met must translocate to the nucleus to initiate calcium signals. *J Biol*
551 *Chem* **283**, 4344-4351 (2008).
552 24. J. Lefebvre *et al.*, Met degradation: more than one stone to shoot a receptor down. *Faseb j*
553 **26**, 1387-1399 (2012).
554 25. S. Kermorgant, P. J. Parker, c-Met signalling: spatio-temporal decisions. *Cell Cycle* **4**, 352-
555 355 (2005).
556 26. Y. W. Zhang, L. M. Wang, R. Jove, G. F. Vande Woude, Requirement of Stat3 signaling
557 for HGF/SF-Met mediated tumorigenesis. *Oncogene* **21**, 217-226 (2002).
558 27. S. Kermorgant, P. J. Parker, Receptor trafficking controls weak signal delivery: a strategy
559 used by c-Met for STAT3 nuclear accumulation. *J Cell Biol* **182**, 855-863 (2008).
560 28. W.-H. Lin *et al.*, STAT3 phosphorylation at Ser727 and Tyr705 differentially regulates the
561 EMT–MET switch and cancer metastasis. *Oncogene* **40**, 791-805 (2021).
562 29. B.-S. Lee, M. Park, H.-Y. Cha, J.-H. Lee, Hepatocyte growth factor induces delayed
563 STAT3 phosphorylation through interleukin-6 expression. *Cellular signalling* **21**, 419-427
564 (2009).
565 30. S. Koh, K. H. Lee, HGF mediated upregulation of lipocalin 2 regulates MMP9 through
566 nuclear factor-κB activation. *Oncology reports* **34**, 2179-2187 (2015).
567 31. T. Qi *et al.*, Function and regulation of the PEA3 subfamily of ETS transcription factors in
568 cancer. *American journal of cancer research* **10**, 3083 (2020).
569 32. Z. Kherrouche *et al.*, PEA3 transcription factors are downstream effectors of Met signaling
570 involved in migration and invasiveness of Met-addicted tumor cells. *Molecular oncology*
571 **9**, 1852-1867 (2015).
572 33. O. Fischer, S. Hart, A. Gschwind, A. Ullrich, EGFR signal transactivation in cancer cells.
573 *Biochemical Society Transactions* **31**, 1203-1208 (2003).
574 34. M. J. Duffy *et al.*, The ADAMs family of proteases: new biomarkers and therapeutic targets
575 for cancer? *Clinical proteomics* **8**, 1-13 (2011).
576 35. R. Zhou *et al.*, ADAM9 Mediates Triple-Negative Breast Cancer Progression via AKT/NF-
577 κB Pathway. *Frontiers in Medicine* **7**, (2020).
578 36. N. A. Beauchene *et al.*, Rad21 is required for centrosome integrity in human cells
579 independently of its role in chromosome cohesion. *Cell Cycle* **9**, 1774-1780 (2010).
580 37. X. W. Pan *et al.*, SMC1A promotes growth and migration of prostate cancer in vitro and
581 in vivo. *Int J Oncol* **49**, 1963-1972 (2016).
582 38. T. Miyamoto *et al.*, The Microtubule-Depolymerizing Activity of a Mitotic Kinesin Protein
583 KIF2A Drives Primary Cilia Disassembly Coupled with Cell Proliferation. *Cell Rep* **10**,
584 664-673 (2015).

585 39. P. Andreu-Pérez *et al.*, Protein arginine methyltransferase 5 regulates ERK1/2 signal
586 transduction amplitude and cell fate through CRAF. *Sci Signal* **4**, ra58 (2011).
587 40. C. Nakarai *et al.*, Expression of AKR1C3 and CNN3 as markers for detection of lymph
588 node metastases in colorectal cancer. *Clin Exp Med* **15**, 333-341 (2015).
589 41. L. Huang *et al.*, RPN2 promotes metastasis of hepatocellular carcinoma cell and inhibits
590 autophagy via STAT3 and NF-κB pathways. *Aging (Albany NY)* **11**, 6674-6690 (2019).
591 42. Y. H. Zheng *et al.*, BCAT 1, a key prognostic predictor of hepatocellular carcinoma,
592 promotes cell proliferation and induces chemoresistance to cisplatin. *Liver International*
593 **36**, 1836-1847 (2016).
594 43. V. Justilien *et al.*, Ect2-dependent rRNA synthesis is required for KRAS-TRP53-driven
595 lung adenocarcinoma. *Cancer cell* **31**, 256-269 (2017).
596 44. Y. Q. Liu *et al.*, Skp1 in lung cancer: clinical significance and therapeutic efficacy of its
597 small molecule inhibitors. *Oncotarget* **6**, 34953-34967 (2015).
598 45. E. Andreuzzi *et al.*, Deregulated expression of Elastin Microfibril Interfacer 2 (EMILIN2)
599 in gastric cancer affects tumor growth and angiogenesis. *Matrix Biology Plus* **6**, 100029
600 (2020).
601 46. R. D. Finn *et al.*, HMMER web server: 2015 update. *Nucleic Acids Res* **43**, W30-38 (2015).
602 47. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version 7:
603 Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772-
604 780 (2013).
605 48. A. L. Mitchell *et al.*, InterPro in 2019: improving coverage, classification and access to
606 protein sequence annotations. *Nucleic Acids Res* **47**, D351-d360 (2019).
607 49. A. Alexa, J. Rahnenfuhrer, topGO: enrichment analysis for gene ontology.
608 50. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature methods*
609 **9**, 357 (2012).
610 51. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or
611 without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
612 52. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
613 RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
614 53. M. Mirdita, S. Ovchinnikov, M. Steinegger, ColabFold - Making protein folding accessible
615 to all. *bioRxiv*, 2021.2008.2015.456425 (2021).
616 54. D. A. Case *et al.* (University of California, San Francisco, 2020).
617 55. J. A. Maier *et al.*, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone
618 Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-3713 (2015).
619 56. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison
620 of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*
621 **79**, 926-935 (1983).
622 57. P. Eastman *et al.*, OpenMM 7: Rapid development of high performance algorithms for
623 molecular dynamics. *PLoS Comput Biol* **13**, e1005659 (2017).
624 58. C. Tian *et al.*, ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against
625 Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **16**, 528-552
626 (2020).
627 59. T. E. Cheatham, III, J. L. Miller, T. Fox, T. A. Darden, P. A. Kollman, Molecular Dynamics
628 Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads
629 to Stable Trajectories of DNA, RNA, and Proteins. *Journal of the American Chemical*
630 *Society* **117**, 4193-4194 (1995).

631 60. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, O. Beckstein, MDAnalysis: A toolkit
632 for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*
633 **32**, 2319-2327 (2011).
634 61. R. Gowers *et al.*, *MDAnalysis: A Python Package for the Rapid Analysis of Molecular*
635 *Dynamics Simulations*. (2016), pp. 98-105.
636
637

638 **Figure legends:**

639 **Fig. 1: PAN domain modulates HGF stability.** (A and B) Core cysteines in PAN domain are
640 crucial for HGF stability. Immunoblot analysis of whole cell lysates derived from 293T cells,
641 transfected with Flag-HGF WT and different single cysteine mutants of Flag-HGF constructs as
642 indicated. 30 h post-transfection, whole-cell lysates were prepared for immunoblot analysis.
643 Representative image of n=3 biological replicates. (B) Quantification of the band intensities in
644 (A). The intensities of Flag-HGF (WT and mutants) bands were normalized to actin and then
645 normalized to Flag-HGF WT. Data are represented as mean ± SD, n = 3, and *$p < 0.05$, **$p <$
646 0.005, *** $p < 0.0005$ (student's t-test). (C) Mutation in all four core cysteine residues in PAN
647 domain significantly alters HGF stability. Immunoblot analysis of whole cell lysates derived from
648 293T cells, transfected with Flag-HGF WT and Flag-HGF 4Cys-4Ala constructs as indicated. 30
649 h post-transfection, whole-cell lysates were prepared for immunoblot analysis. (D) Cysteines in
650 kringle domains and SPH domain of HGF have no impact on the protein abundance in cells.
651 Immunoblot analysis of whole cell lysates derived from 293T cells, transfected with Flag-HGF
652 WT, Flag-HGF 4Cys-4Ala and different single cysteine mutants of Flag-HGF constructs as
653 indicated. 30 hours post-transfection, whole-cell lysates were prepared for immunoblot analysis.
654 Representative image of n=3 biological replicates. (E) Quantification of the band intensities in (D).
655 The intensities of Flag-HGF (WT and mutants) bands were normalized to actin and then
656 normalized to Flag-HGF WT. Data are represented as mean ± SD, n = 3, and *$p < 0.05$, **$p <$

657 0.005, ***p < 0.0005 (student's t-test). (F) Localization of Flag-HGF WT and different Flag-HGF
658 mutants' expression by confocal immunofluorescence microscopy in HeLa cells. The cells were
659 transiently transfected with Flag-HGF WT and different mutants of Flag-HGFs as indicated. 30 h
660 post-transfection cells were fixed, mounted and protein expression patterns were visualized using
661 a Zeiss LSM 710 confocal microscope outfitted with a 63x objective. Scale bars represent 20 µm.
662 The images shown are representative from three independent biological experiments (average 100
663 cells were observed per experimental condition per replicate). (G) Percentage of transfected HeLa
664 cells showing perinuclear/nuclear staining for Flag-HGF WT and Flag-HGF 4Cys-4Ala were
665 quantified. Data are represented as mean ± SD, n = 3 (average 100 cells were observed for each
666 condition per experiment), and *p<0.05, **p < 0.005, *** p < 0.0005 (Student's t test).

667 **Fig. 2: HGF PAN domain regulates MET signaling cascade via four core cysteines.** (A)
668 Mutation of the core cysteines in HGF PAN domain blocks HGF induced c-MET signaling. Both
669 293T and U-87 MG cells were stimulated with HGF WT and HGF 4Cys-4Ala proteins for
670 indicated amount of time. Cells were harvested and immunoblot analysis shows the absence of
671 phosphorylation for MET, AKT and ERK in presence of HGF 4Cys-4Ala. Representative blot
672 images from n=2 experiments for individual cell line. (B) Immunoblot showing that the core
673 cysteines on the HGF PAN domain regulate its binding with c-MET. c-MET was
674 immunoprecipitated from 293T cells on anti-MET bound beads. *In-vitro* translated Flag-tagged
675 HGF WT and HGF 4Cys-4Ala were added to the beads as indicated to detect the interaction
676 between endogenous c-MET and Flag-HGF WT and Flag-HGF 4Cys-4Ala. (C)Right panel,
677 quantification of the band intensities ($n = 2$; *** p < 0.0005 (Student's t test)). Immunoprecipiated
678 Flag-HGFs band intensities were normalized to the respective c-MET IP bands and then further
679 normalized to HGF-WT. (D) HGF PAN domain defines perinuclear translocation of MET in cells.
680 HeLa cells were stimulated with HGF WT and HGF 4Cys-4Ala proteins for indicated amount of

681 time following serum starvation. Post stimulation cells were fixed, mounted and endogenous c-
682 MET localization pattern was visualized using Zeiss LSM 710 at 63x objective. Scale bars
683 represent 20 μm. The images shown are representative from three independent biological replicates
684 (average 100 cells were observed for each condition per replicate). (E) PAN domain regulates c-
685 MET ubiquitination. In vivo ubiquitination assay shows that HGF WT promotes c-MET
686 ubiquitination in a PAN-dependent manner. 293T cells were transfected with the construct c-MET-
687 C-GFPSpark. After serum starvation, cells were stimulated with HGF WT and HGF 4Cys-4Ala as
688 indicated. The lysates were collected at specific time points and incubated with anti-GFP protein
689 G beads. Ubiquitinated-MET proteins were eluted, resolved by SDS-PAGE and immunoblotted
690 with the indicated antibodies.

691 **Fig. 3: Core cysteines in HGF PAN domain is essential for STAT3 phosphorylation and**
692 **nuclear translocation.** (A) Impaired HGF PAN domain is unable to initiate STAT3
693 phosphorylation. U-87 MG cells were treated with HGF WT and HGF 4Cys-4Ala where indicated
694 for 1, 2 and 4 h. Cell extracts were prepared and probed for phosphor-STAT3 and total STAT3.
695 (B) STAT3 nuclear localization is suppressed by PAN mutant HGF. HeLa cells were stimulated
696 with HGF WT and HGF 4Cys-4Ala as indicated. Cells were fixed and STAT3 was immunostained
697 with a STAT3-specific antibody. The localization of STAT3 (green) and 4,6-diamidino-2-
698 phenylindole (DAPI) (blue) in U-87 MG cells. Images were visualized using Zeiss LSM 710 at
699 63x objective. Scale bars represent 20 μm. The images shown are representative of three
700 independent biological replicates (an average of 100 cells were observed for each condition per
701 replicate).

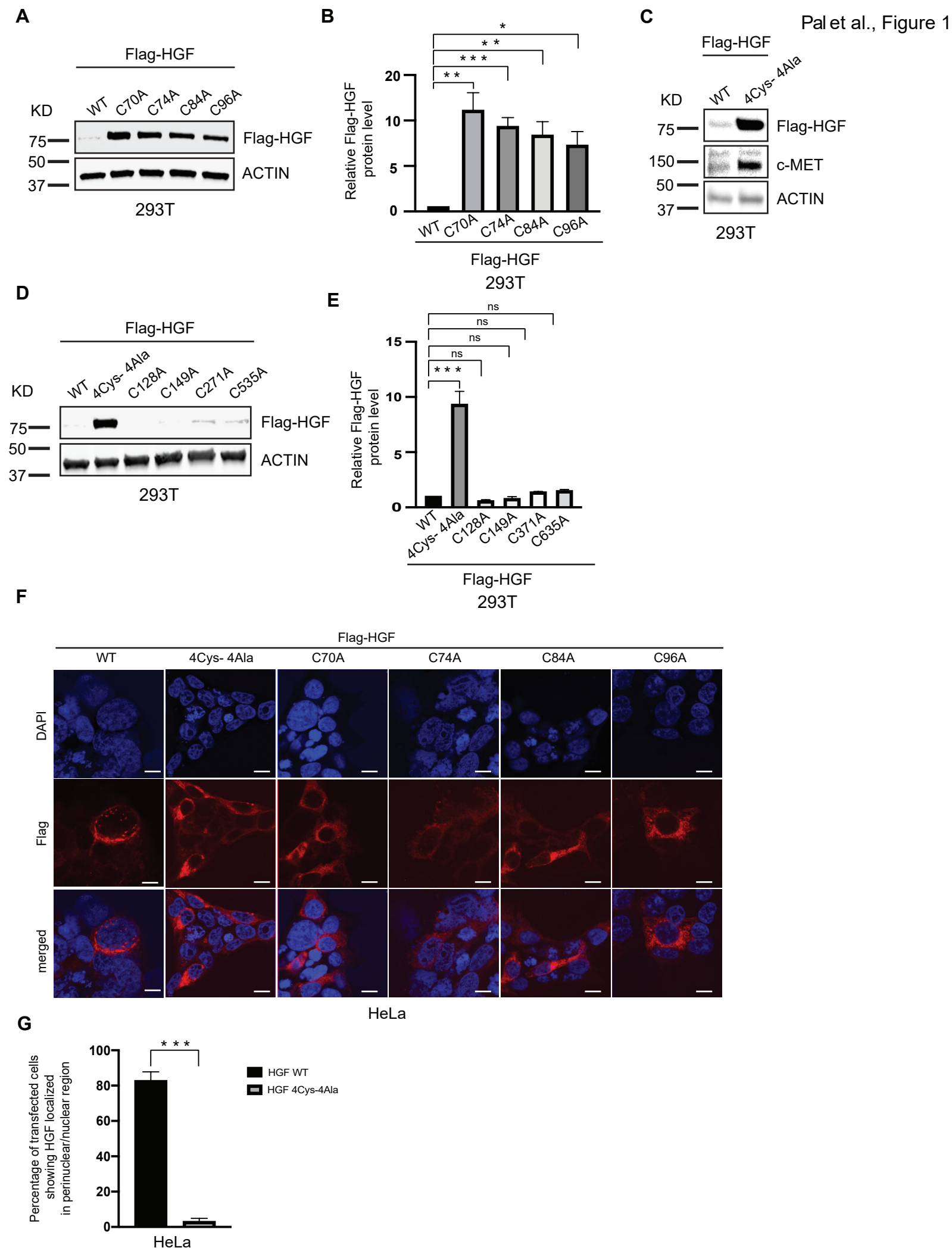
702 **Fig. 4: Transcriptome analysis post HGF stimulation in 293T cells.** Differential expression
703 analysis by RNA seq in 293T cells following HGF WT and HGF 4Cys-4Ala treatment as indicated

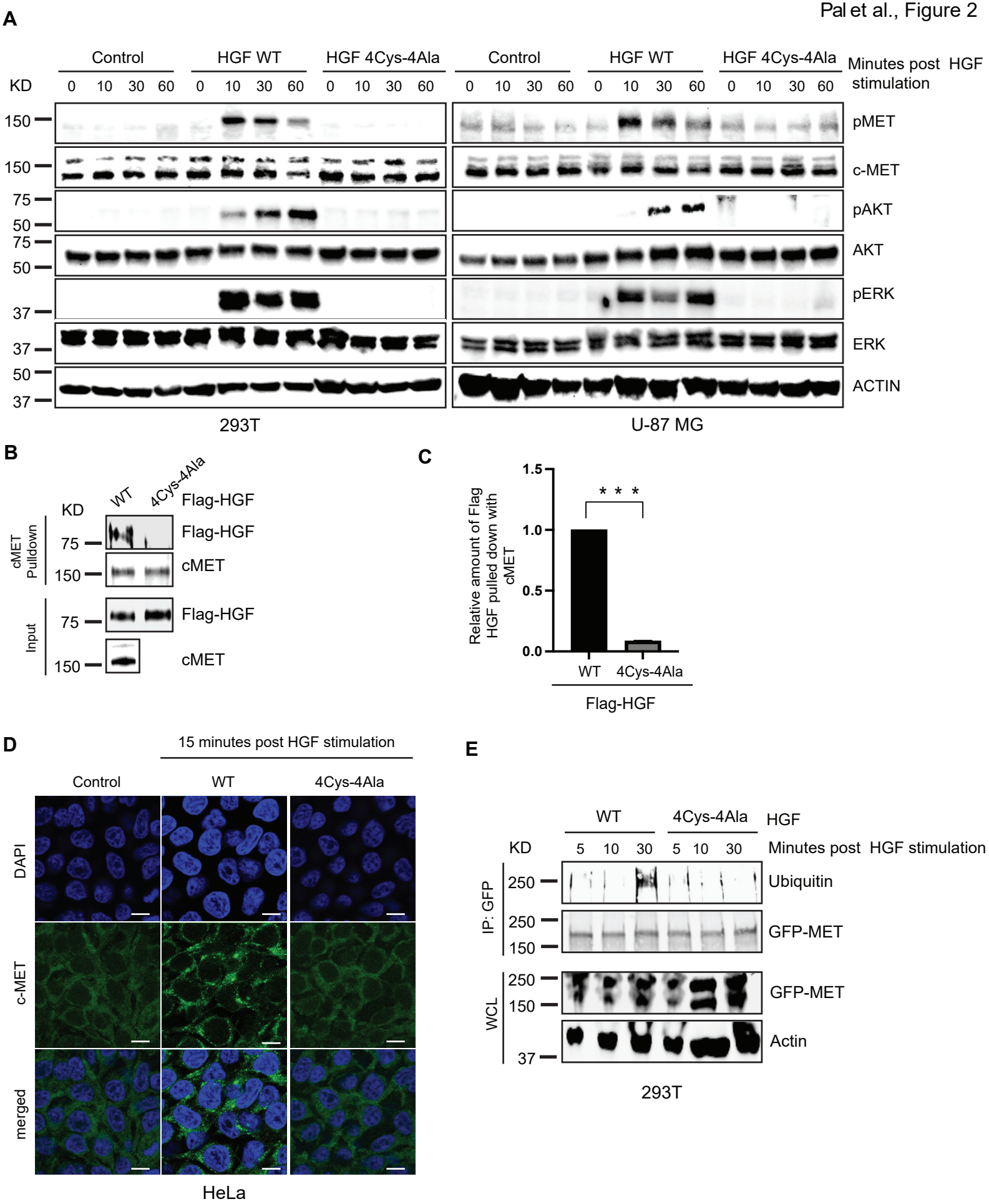
704 confirms that core cysteines in HGF PAN domain are necessary for the expression of a wide range
705 of genes. Responsive genes were normalized to FPKM value for non-treated cells and then
706 normalized to HGF WT treated cells. Data represents the average of three independent biological
707 replicates and *$p < 0.05$, **$p < 0.005$ and ***$p < 0.0005$ were calculated with a student's t-test.

708

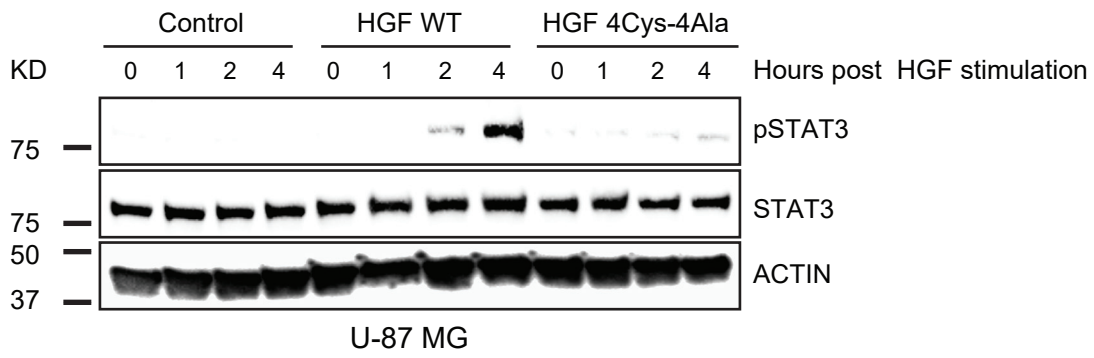
709

Pal et al., Figure 1

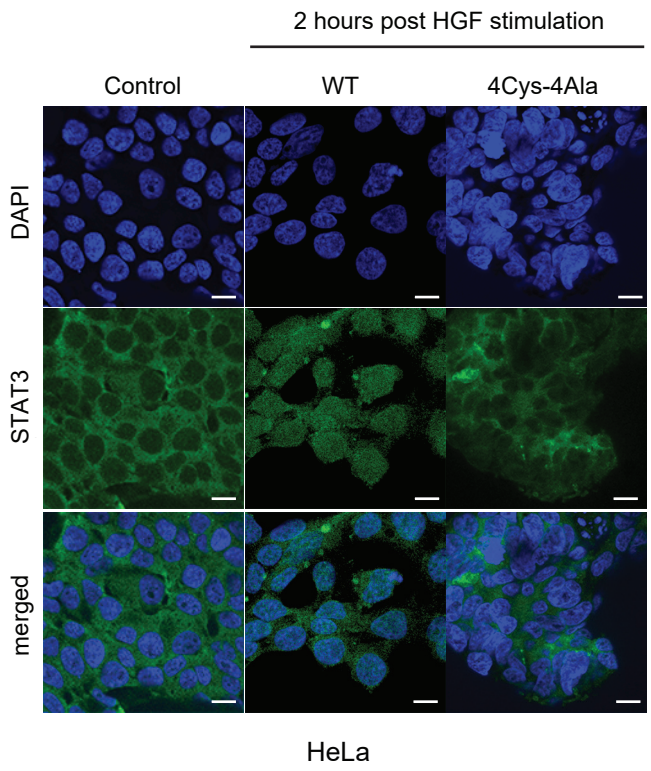


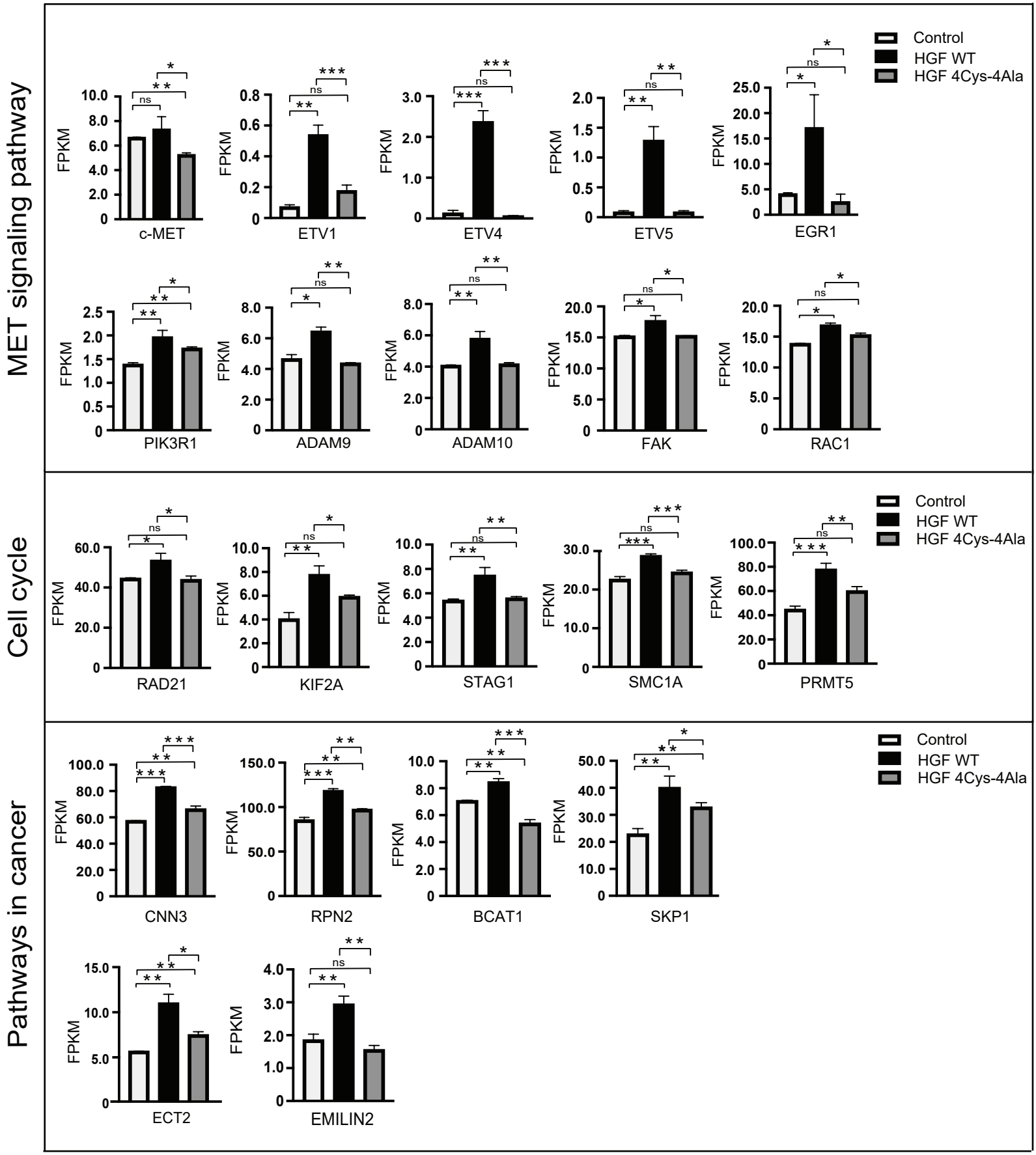

A
Control
HGF WT
HGF 4Cys-4Ala
Control
HGF WT
HGF 4Cys-4Ala
Minutes post HGF stimulation
KD
0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60
150
150
75
50
75
50
37
37
50
37
pMET
c-MET
pAKT
AKT
pERK
ERK
ACTIN
293T
U-87 MG
B
WT
4Cys-4Ala
Flag-HGF
KD
cMET Pulldown
75
Flag-HGF
150
cMET
Input
75
Flag-HGF
150
cMET
C
Relative amount of Flag HGF pulled down with cMET
1.5
1.0
0.5
0.0
* * *
WT
4Cys-4Ala
Flag-HGF
D
15 minutes post HGF stimulation
Control
WT
4Cys-4Ala
DAPI
c-MET
merged
HeLa
E
WT
4Cys-4Ala
HGF
KD
5 10 30 5 10 30
Minutes post HGF stimulation
IP: GFP
250
Ubiquitin
250
150
GFP-MET
WCL
250
150
GFP-MET
37
Actin
293T

A


Control
HGF WT
HGF 4Cys-4Ala
KD
0 1 2 4 0 1 2 4 0 1 2 4
Hours post HGF stimulation
75
pSTAT3
75
STAT3
50
37
ACTIN
U-87 MG

B


2 hours post HGF stimulation
Control
WT
4Cys-4Ala
DAPI
STAT3
merged
HeLa


MET signaling pathway
Cell cycle
Pathways in cancer
Control
HGF WT
HGF 4Cys-4Ala
FPKM
c-MET
ETV1
ETV4
ETV5
EGR1
PIK3R1
ADAM9
ADAM10
FAK
RAC1
RAD21
KIF2A
STAG1
SMC1A
PRMT5
CNN3
RPN2
BCAT1
SKP1
ECT2
EMILIN2
ns