THE
EMBO
JOURNAL

Elucidation of WW domain ligand binding specificities in the Hippo pathway reveals STXBP4 as YAP inhibitor

Rebecca Vargas, Vy Thuy Duong, Han Han, Albert Paul Ta, Yuxuan Chen, Shiji Zhao, Bing Yang, Gayoung Seo, Kimberly Chuc, Sunwoo Oh, Amal El Ali, Olga Razorenova, Junjie Chen, Ray Luo, Xu Li, and Wenqi Wang

Review timeline:

Submission date:	8th Oct 2019
Editorial Decision:	10th Jul 2019
Revision received:	17th Sep 2019
Editorial Decision:	5th Nov 2019
Revision received:	6th Nov 2019
Accepted:	11th Nov 2019

Editor: Ieva Gailite

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 10th Jul 2019

Thank you for submitting your manuscript for consideration by the EMBO Journal. I sincerely apologise for the unusual delay in the assessment of your work due to belated submission of referee reports. We have now received the full set of reviewers' reports on your manuscript, which are included below for your information.

As you will see from the comments, while reviewer #2 is more critical, reviewers #1 and #3 appreciate the work and the topic. However, all reviewers raise a number of substantial concerns that need to be addressed before they can support publication here. Based on the overall interest expressed in the reports of reviewers #1 and #3, I would like to invite you to submit a revised version of your manuscript in which you address the comments of all three referees. In particular, please address the concerns of reviewer #2 regarding the residues relevant for the WW domain protein interaction site and the points regarding the characterisation of the role of STXBP4 in YAP regulation raised by reviewers #1 and #3. Additionally, all reviewers indicate that the manuscript should be streamlined and indicate that Figure 3 is not sufficiently conclusive. I would therefore suggest to remove this figure in the absence of experimental validation of the data, or move it to the Appendix.

REFEREE REPORTS:

Referee #1:

The manuscript by Rebecca Vargas and colleagues from the laboratory of Wenqi Wang examines specificity of WW domain-mediated complexes in the Hippo pathway. Based on their detailed analysis of WW domain structures and on the results of Mass Spectrometry analyses of WW domain-assembled complexes, the authors have identified a new WW domain-containing protein,

STXBP4 that has ability to act as a suppressor of YAP. STXBP4 was shown to interact with alpha-catenin and this interaction was mapped in the current report. Importantly, the authors suggest that STXBP4 is a tumor suppressor and YAP-alpha-catenin-STXBP4 signaling is a part of the mechanism by which the Hippo pathway responds to mechanical cues.

The manuscript is very dense with data. It seems as if two manuscripts, one that has to do with structure-function analysis of WW domains and the other that is focused on the STXBP4 complex with alpha-catenin that attenuates YAP function in actin cytoskeleton-dependent manner. Generally, the presentation of the data is clear and considering that the Hippo-YAP pathway is still burgeoning in the field of cancer signaling, this work is of potential interest to a wide readership, well beyond regulars of the EMBO journal.

The following changes that could be addressed in the span of three months are suggested to improve the manuscript:

1. It is suggested to consider moving Figures 1 to 3 to supplementary data and to focus the report and the presented data in main figures on the STXBP4 complex with alpha catenin and regulation of YAP. The title could be also more specific to emphasize STXBP4 and cancer.

2. Figures 4J and 5J are of low quality and could benefit from multiple repetitions and quantitative/statistical evaluation of changes. Also the p-tag YAP assay needs careful description.

3. As above, whenever the authors state: phosphorylated YAP, which residues they have in mind. This is especially important when the statement is made as follows: "Interestingly, loss of STXBP4 specifically attenuated YAP phosphorylation when actin cytoskeleton was either depolymerized or its tension was inhibited (Figure 4J). In contrast, YAP was still fully phosphorylated under serum and glucose-deprived conditions (Figure 4J). These data suggest that STXBP4 is required for the actin cytoskeleton tension-mediated Hippo pathway regulation." Also, please explain in detail the serum and glucose deprived conditions and the rationale for this treatemnt. These stress conditions could lead to YAP to being phosphorylated by AMPK on Serine 94?

4. All blots would benefit from adding at least two molecular weight markers and tissues staining would benefit from the inclusion of "size bars".

5. Several references could be added to the discussion if the space allows:
*A very fine discussion of how Hippo-YAP pathway responds to mechanical cues, from the lab of Mike Sheetz: Low BC, et al.,.(2014) YAP/TAZ as mechanosensors and mechanotransducers in regulating organ size and tumor growth. FEBS Lett. 588 (16): 2663-70.

**A focused review on WW domains of the Hippo pathway, discussing also potential of homo- and hetero-dimer formations among WW domains: Sudol, M., and Harvey, K. (2010) The Modularity in the Hippo Signaling Pathway. TiBS, 35, 627-633.

***An important report on how YAP regulated actin polymerization via one of the regularors of RhoA: Qiao Y,et al.,. (2017) YAP regulates actin dynamics through ARHGAP29 and promotes metastasis. Cell Reports 19, 1495-1502.

Referee #2:

There are several well-established WW-PY interactions among Hippo components that are functionally important for signaling output. In this work, the authors explore the molecular basis for the Hippo-specific WW-PY recognition and discover that Hippo-related WW domains all contain a unique 8-residue sequence that is essential for pathway regulation. Although the 'concept' of Hippo-specific WW domain is quite appealing, the fact that there are 20-fold more PPxY-containing

peptides that WW domains in human proteome suggesting that other factors must contribute to binding specificity in cellular context.

Besides the disputable concept of Hippo-specific WW domain, my major concern with this manuscript is that although the overall biochemical data seem solid, the authors do not seem to understand the structural basis for protein-protein interactions. In particular, it is well-established that the binding site for proline-containing peptides is formed on the concave surface of 3-stranded beta-sheet of WW domain. The binding surface contains a Pro-binding site formed by Y18 and W29 (YAP-WW1 numbering and residue type from Figure 2 are used here) that recognizes the canonical proline in target peptides, and a specificity Y-binding site formed by residues L20, H22 and Q25 from beta-2 strand and the beta-2/3 loop, respectively. However, the 8-residue sequence identified in this work (..L3P4.G6W7E8.....F19..H22....T27.W29..P32..) is missing three key residues (Y18, L20 and Q25) that are required for canonical WW-PY recognition. More importantly, residues L3, P4, W7, F20 and P32 are required for WW-domain folding, but not directly involved in PY-binding, therefore should not be considered in binding specificity. The authors need to address these concerns.

If there is Hippo-specific sequence for WW domains, then the authors should comment on Hippo-specific sequence for PY motifs.

It is surprised that SAV1-WW is not considered as Hippo-specific WW domain by the authors. SAV1-WW binds to LATS-PY motifs with functional relevance in Hippo signaling. The authors should comment on this contradiction.

The structure models for YAP-WW1 mutants based on simulation and the distances to SMAD7-PY reported in Figure 3 have no validation and provide no further information to support their conclusion.

Referee #3:

In this manuscript, Vargas and colleagues report their investigation of STXBP4 in the Hippo pathway. The authors have reanalyzed data (their own published work) for WW-proteins/PY interactomes, added some additional purifications, and conducted their own validation experiments to reveal eight amino acids in the WW-domain sequence conferred specificity for binding to PY motifs present in Hippo pathway proteins and a few additional proteins (but not for SAV1). With this WW consensus sequence (unclear how), they performed a proteome-wide search and identified STXBP4, which they then implicate as a cytoskeletal-tension mediated regulator of Hippo. They establish a connection between alpha-catenin and LATS/AMOT that depends on STXBP4, and identify disease mutants that appear to be less functional in the Hippo pathway. A deregulation of STXBP4 was identified in kidney cancers, and overexpression of STXBP4 reduced xenograft tumor formation in one model of renal cancer.

This study would in principle be of high interest to readers of EMBO J as it helps defining a new protein that links mechanotransductive signals to the core Hippo pathway. Yet, the current version of the manuscript is somewhat lacking in critical details, much of which could be fixed by textual changes, and perhaps by either removing some of the modelling data or validating it by additional experiments. Some of the mechanistic details need further validation (i.e. How can a single WW protein interact with multiple PY proteins? Are the interactions detected observed with endogenous proteins? Etc.) and some of the results are less convincing than ideal. In addition, the presentation of the manuscript itself should be streamlined to make it more accessible to the readers.

Detailed comments

1. Introduction (p.4-5): The introduction could be shortened and streamlined, notably the lengthy description of all WW proteins associated with Hippo which would perhaps be better suited for a table in the main or supplementary text.
2. In the description of the Figure 1 (and other parts of the paper), it should be made clearer which of the proteomics analyses are previously published and which are new. New datasets should be

more clearly described in both the Results and Methods section (including replicate analysis, negative controls, etc.). MUSE scoring across the dataset requires description of all other mass spectrometry runs used for scoring, and links to the other data not specifically generated for this publication in order to enable external re-analysis. Similarly, whether the initial results from Wang 2014 were used as initially published or whether a computational reanalysis was performed should be more explicitly described.
3. Results (p6, ¶1): Although it is interesting that SAV1, a WW-containing Hippo pathway protein does not bind Hippo specific PY motifs, if the point of this section of results is that there is a WW sequence that is specific for Hippo-related PY motifs, focusing in on SAV1, an anomaly, detracts from this point. This section should be streamlined.
4. Figure 2 (C-D) - In the pulldown of AMOT with full-length TAZ (C), AMOT is still able to be recovered despite G/A and E/A mutations, but the TAZ-WW domain (D) only shows faint pulldown of AMOT with the G/A mutation. Does this mean that there are other portions of TAZ that are important for binding to AMOT outside of this Hippo-specific 8 AA WW-related region? Perhaps these mutations are less detrimental to the WW-PY interaction as evidenced by the low degree of conservation, but this is not discussed in the paper at all. It seems that the authors may be oversimplifying the interpretation of their results here.
5. Have the authors tried more conservative substitutions (e.g. Y to F, E to D, etc.) for their 8 amino acid "motif" to help refine the binding determinants and to help with the proteome-wide (of WW-wide) analysis?
6. Figure 3 does not add much to the paper in my opinion, especially after the mutational study performed in Figure 2. All results presented are from simulations (though they kind of read like an experimental mutagenesis scanning), based on only 2 structural scaffolds, and while the results appear in agreement with the conclusion of the 8 amino acids as being key for specificity, it is not currently tested experimentally and therefore remains fairly speculative. The authors also do not go outside of the mutants they have tested in Figure 2 (e.g. to expand to more conservative changes). Perhaps a brief mention at the end of the Figure 2 discussion and a supplementary figure or table would be more appropriate than a separate figure ? This part of the paper is not needed for the identification of STXBP4. [The next few comments specifically refer to technical issues with this figure, should it stay in the paper].
7. Results (p.10, ¶2): Although mentioned in the figure legend, it would have been better to be up front in the text that these alanine mutation scan data were derived from simulations. The authors conclude that the large change in RMSD given alanine mutations in the hydrophobic cluster confirm their hypothesis. Perhaps they can also include an actual experimental method to orthogonally show that these mutations result in a change in YAP-WW1 structure instead of relying solely on simulated data?
8. Figure 3 (C): Is this bond frequency graph calculated over the average of all of the NMR derived structures? What are the authors are trying to convey with this? At what point does a particular interaction become classified as one or the other? Does a lower frequency indicate that the other portion is of a different bond type or is unbonded?
9. Figure 3 (D): RMSD needs units. They also discuss that the RMSDs are relatively high - there is no discussion as to if this is actually biologically relevant (i.e. the RMSDs may be different, but does a difference of an RMSD of 1.4 A and 4.2 A actually make a biological difference?)
10. Figure 3 (F): With a stable interaction (YAP-SMAD7), the average distance is lower and with a lower magnitude of variation, but the mutants have a high value with higher variation. Are the authors saying that YAP with 7.8 +/- low variation mean binding and anything deviating from that is not binding? Could they have also introduced a negative control mutation either outside of the conserved 8 AA sequence or the WW domain and look at the change in average distance for their simulations? The methodologies and statistics were unclear from the text.
11. The transition to STXBP4 is very abrupt. I could not find the parameters for the search through the "proteome". I assume that the authors have restricted themselves to WW-domain containing proteins? Have they then manually scanned all the sequences and found only STXBP4 to contain it? Have they considered the possibility for conservative substitutions that may still retain binding? The authors need to expand their method section to include these critical components. They also need to more clearly explain the results presented in Table S6 that lists other
12. STXBP4 association with some Hippo components (including PTPN14 and LATS2) was previously reported in the literature outside of the authors' own studies (see, e.g. the BioGRID database) - these references should be cited here.
13. The conservation of the 8 amino acid sequence of STXBP4 only in mammals is intriguing, especially in light of the extensive conservation of Hippo signaling. The substitutions in frog and

fish appear fairly conservative: could these sequences bind Hippo PY? What could be the significance of a mammalian specific mechanotransducer? Could other proteins functionally replace it, e.g. in Drosophila? (i.e. are there other WW proteins in Drosophila that fits the 8 amino acid criteria established by the authors?)
14. Figure 4 (J): Although latrunculin B and blebbistatin have effects on the cytoskeleton, do they actually modulate cytoskeleton tension directly? If this is not firmly established it may be a large stretch to say that the changes in p-YAP are due to cytoskeletal tension changes and not just proper cytoskeleton structure.
15. Figure 5 (C, myc-LATS1 pulldown): Is this really faint band enough evidence to say that STXBP4 promotes the association of a-catenin with LATS1?
16. Have the CRISPR-generated clones been sequenced? If so, please add the results of the sequencing to the Methods section.
17. Presumably, each WW domain should be capable of interacting with a single PY motif: this should be considered by the authors in their models - as it is, we are left with the impression that STXBP4 can simultaneously interact with LATS and AMOT through its WW domain.
18. I am puzzled by Figure 5: first, the rescue of YAP phosphorylation as detected by the Phos-tag gel seems very moderate. How reproducible is this? Has this been quantified? The mutants also do not all have the same effects, with some of them resulting in apparent disappearance of YAP - have phosphatase treatments been performed to confirm that protein levels are the same for YAP across all these panels? This experiment needs to be repeated with clearer readout of what actually happens to YAP phosphorylation as it is certainly not as clear as what the authors infer in the text.
19. Have the authors analyzed the relationships between other Hippo pathway WW or PY components and STXBP4/YAP in the TCGA dataset? Note that much of the TCGA analysis could be moved to supplementary, and this section shortened.
20. In Figure 6F, what is the YAP and phospho-YAP status? Does this correlate with STXBP4 levels? In 6G/H, are the three STXBP4 constructs expressed at the same levels and uniformly in the xenografts? Importantly, how do the levels of expression of any of these proteins correspond to endogenous levels in normal or cancer cells? As presented, these potentially exciting findings seem preliminary. Note that the authors emphasize the potential connection of STXBP4 with kidney cancers, specifically, but it would be interesting to know whether the effects are broader when STXBP4 is overexpressed (of course, it would be great to see whether deletion of STXBP4 is sufficient to induce cancer formation in a relevant model, but I feel that this is beyond the scope of the current study).
21. The entire text would benefit from editing: at the moment, parts of the text are difficult to follow. A few examples are listed below, but this is not an exhaustive list.
22. Results (p7, ¶1): It is unclear what "To further test this hypothesis..." is referring to. The previous paragraph ends with the conclusion that the WW domain of SAV1 is different from YAP/TAZ/KIBRA. Thus, "this hypothesis" seems to refer to this difference, but the text paragraph actually refers to the hypothesis that YAP/TAZ/KIBRA have a binding specificity that is exclusive to Hippo PY containing proteins. I may be missing something, but was the proteomics data necessary for this analysis? Looking at the sequence alignments of the Hippo-PY-specific WW domains (Figure 2A) versus the control WW domains (Figure S2A), the consensus for the eight AAs in Figure 2A is very strong while the consensus for the same residues in the control WW domains in S2A are relatively weak. Thus, would it not be possible to come to the same conclusion that these residues may be Hippo specific?
23. Results (p8, ¶2): This paragraph begins with "To elucidate the mechanism underlying...specificity..." but the paragraph content concludes with the identification of eight conserved amino acids. This should be made more clear. Also, the inclusion of SAV1's anomalous sequence detracts from the main point of the paragraph.
(p8, ¶3): This introductory sentence is unclear. What did the authors subject the identified sequence to?
24. Results (p9, ¶3): The introductory paragraph would benefit from editing to make this more clear and concise.

1st Revision - authors' response 17th Sep 2019

Please see next page.

Referee #1:

The manuscript by Rebecca Vargas and colleagues from the laboratory of Wenqi Wang examines specificity of WW domain-mediated complexes in the Hippo pathway. Based on their detailed analysis of WW domain structures and on the results of Mass Spectrometry analyses of WW domain-assembled complexes, the authors have identified a new WW domain-containing protein, STXBP4 that has ability to act as a suppressor of YAP. STXBP4 was shown to interact with alpha-catenin and this interaction was mapped in the current report. Importantly, the authors suggest that STXBP4 is a tumor suppressor and YAP-alpha-catenin-STXBP4 signaling is a part of the mechanism by which the Hippo pathway responds to mechanical cues.

Thanks for the nice summary of our work!

The manuscript is very dense with data. It seems as if two manuscripts, one that has to do with structure-function analysis of WW domains and the other that is focused on the STXBP4 complex with alpha-catenin that attenuates YAP function in actin cytoskeleton-dependent manner. Generally, the presentation of the data is clear and considering that the Hippo-YAP pathway is still burgeoning in the field of cancer signaling, this work is of potential interest to a wide readership, well beyond regulars of the EMBO journal.

Thanks!

The following changes that could be addressed in the span of three months are suggested to improve the manuscript:

1. It is suggested to consider moving Figures 1 to 3 to supplementary data and to focus the report and the presented data in main figures on the STXBP4 complex with alpha catenin and regulation of YAP. The title could be also more specific to emphasize STXBP4 and cancer.

Thanks for the suggestions! We agree that the data presentation should be more concise and the STXBP4-related studies could be the focus. Since STXBP4 was identified through characterizing the Hippo WW domain binding specificity (another key finding in this study), we think that **Figures 1** and **2** may help the readers easily trace the discovery of STXBP4. Based on it, we hope to keep **Figures 1** and **2** as the main figures but have moved the previous **Figure 3** to the supplemental data as **Figure EV3**. In addition, the title has been updated based on the suggestion.

2. Figures 4J and 5J are of low quality and could benefit from multiple repetitions and quantitative/statistical evaluation of changes. Also the p-tag YAP assay needs careful description.

Thanks for pointing out this issue! To better interpret the band-shift of YAP in phospho-tag gel, additional labeling has been included to indicate the different levels of YAP phosphorylation in the revised **Figures 3J** and **4J**. Moreover, the description of the phospho-tag assay has been updated in the revised manuscript.

3. As above, whenever the authors state: phosphorylated YAP, which residues they have in mind. This is especially important when the statement is made as follows: "Interestingly, loss of STXBP4 specifically attenuated YAP phosphorylation when actin cytoskeleton was either depolymerized or its tension was inhibited (Figure 4J). In contrast, YAP was still fully phosphorylated under serum and glucose-deprived conditions (Figure 4J). These data suggest that STXBP4 is required for the actin cytoskeleton tension-mediated Hippo pathway regulation." Also, please explain in detail the serum and glucose deprived conditions and the rationale for this treatemnt. These stress conditions could lead to YAP to being phosphorylated by AMPK on Serine 94?

Thanks for pointing out this issue! Since YAP's band-shift in phospho-tag gel is largely controlled by LATS kinases (Plouffe SW et al. Mol Cell 2016), we took this approach to examine the LATS-mediated YAP phosphorylation in the STXBP4 KO cells. In this study, we did not examine other YAP's phosphorylation events (e.g. the phosphorylation at S94 by AMPK). The experimental details for serum starvation and glucose starvation have been included in the revised **Figure legends** section. To determine the signaling contexts involving STXBP4, we tested several Hippo-related signaling events and found that loss of STXBP4 specifically interfered with the actin cytoskeleton-mediated YAP regulation, where both serum and glucose-deprived conditions were taken as controls.

4. All blots would benefit from adding at least two molecular weight markers and tissues staining would benefit from the inclusion of "size bars".

Thanks for the suggestions! All the figures have been revised accordingly.

5. Several references could be added to the discussion if the space allows:
**A very fine discussion of how Hippo-YAP pathway responds to mechanical cues, from the lab of Mike Sheetz: Low BC, et al.,.(2014) YAP/TAZ as mechanosensors and mechanotransducers in regulating organ size and tumor growth. FEBS Lett. 588 (16): 2663-70.*

***A focused review on WW domains of the Hippo pathway, discussing also potential of homo- and hetero-dimer formations among WW domains: Sudol, M., and Harvey, K. (2010) The Modularity in the Hippo Signaling Pathway. TiBS, 35, 627-633.*

****An important report on how YAP regulated actin polymerization via one of the regularors of RhoA: Qiao Y,et al.,. (2017) YAP regulates actin dynamics through ARHGAP29 and promotes metastasis. Cell Reports 19, 1495-1502.*

Agree! Thanks for the suggestions! These references have been cited in the revised manuscript.

Referee #2:

There are several well-established WW-PY interactions among Hippo components that are functionally important for signaling output. In this work, the authors explore the molecular basis for the Hippo-specific WW-PY recognition and discover that Hippo-related WW domains all contain a unique 8-residue sequence that is essential for pathway regulation. Although the 'concept' of Hippo-specific WW domain is quite appealing, the fact that there are 20-fold more PPxY-containing peptides that WW domains in human proteome suggesting that other factors must contribute to binding specificity in cellular context.

Sorry for the misunderstanding! We totally agree that WW domains may not be the only determinant for the specific WW-PY recognition among the Hippo pathway proteins and this binding specificity could involve the PY motif. As a matter of fact, only a group of PY motif-containing proteins were consistently identified as the binding partners for the Hippo WW proteins based on our proteomic analyses (**Figures 1B** and **1C**). This finding suggested that the PY motif could also play a role here. However, as we discussed in the manuscript, the PY motif is relatively short, flexible and could be easily buried into a higher level of protein structure, making the assessment of its role difficult. Thus, we majorly focused on role of WW domains in contributing to this Hippo WW-PY binding specificity in current study.

Besides the disputable concept of Hippo-specific WW domain, my major concern with this manuscript is that although the overall biochemical data seem solid, the authors do not seem to understand the structural basis for protein-protein interactions. In particular, it is well-established that the binding site for proline-containing peptides is formed on the concave surface of 3-stranded beta-sheet of WW domain. The binding surface contains a Pro-binding site formed by Y18 and W29 (YAP-WW1 numbering and residue type from Figure 2 are used here) that recognizes the canonical proline in target peptides, and a specificity Y-binding site formed by residues L20, H22 and Q25 from beta-2 strand and the beta-2/3 loop, respectively. However, the 8-residue sequence identified in this work (..L3P4.G6W7E8.....F19..H22....T27.W29..P32..) is missing three key residues (Y18, L20 and Q25) that are required for canonical WW-PY recognition. More importantly, residues L3, P4, W7, F20 and P32 are required for WW-domain folding, but not directly involved in PY-binding, therefore should not be considered in binding specificity. The authors need to address these concerns.

Thanks for pointing out these issues!

First, these conserved amino acids were identified through the sequence alignment among the WW domains derived from the known Hippo pathway components or regulators. Actually, only the YAP-WW1 domain contains the L20 and Q25 residues (**Figure 2A**). These two residues were mostly replaced by other amino acids in the rest of the Hippo-related WW domains (**Figure 2A**), which are all capable of binding the Hippo PY motif-containing proteins (**Figures 1E** and **1F**). These evidences suggested that the L20 and Q25 residues could be specifically involved in the YAP-WW1 domain-based recognition of PY motif.

Second, the Y/F18 residue was actually found highly conserved among the Hippo-related WW domains (**Figure 2A**). To test its role in the Hippo WW-PY recognition, we mutated this Y/F18

residue to alanine and found that its mutation also abolished the interaction between TAZ/TAZ-WW/KIBRA/STXBP4 and AMOT (**Figures 2C-2E** and **3C**), confirming the Reviewer's point. This residue was missed in the previous study because the SMURF2-WW domain has this Y/F18 site replaced by glutamine (Q) (**Figure 2A**). This could be explained by the fact that SMURF2 contains three WW domains and the WW domain tandem may contribute to the Hippo PY motif association (Lin Z. et al. eLife 2019). However, our current study only focused on single WW domain-based determination for the Hippo WW-PY recognition. Based on these findings, we have included this Y/F18 residue in the previously identified 8-residue sequence and repeated all the related experiments and simulation analyses (please see the revised figures and manuscript).

Sorry for the misunderstanding! We agree that residues L3, P4, W7, F20 and P32 play an indirect role to support the specific Hippo WW-PY recognition, where they are involved in the WW-domain folding but not the direct binding with PY motif (**Figure EV3**). Since a structurally stable WW domain is required for the Hippo WW-PY recognition and changes in these residues (L3, P4, W7, F20 and P32) were frequently identified in many non-Hippo WW domains (**Figures 2B** and **EV2A**), we concluded that these residues are also involved in the specific Hippo WW-PY recognition. To clarify this point, we have included more discussion for this group of residues in the revised manuscript.

If there is Hippo-specific sequence for WW domains, then the authors should comment on Hippo-specific sequence for PY motifs.

We agree that the Hippo PY motifs could also determine the specific Hippo WW-PY recognition (please also see our previous response to this point in page 3). As we discussed in the manuscript, the PY motif is relatively short, flexible and could be easily buried into a higher level of protein structure, making the assessment of its role more difficult. Thus, we majorly focused on the role of WW domain in determining the Hippo WW-PY binding specificity in this study, but did not further address this question from the PY motif-based perspective. Please see the revised **Discussion** section.

It is surprised that SAV1-WW is not considered as Hippo-specific WW domain by the authors. SAV1-WW binds to LATS-PY motifs with functional relevance in Hippo signaling. The authors should comment on this contradiction.

Thanks for pointing out this issue! Actually, this is how we initiated this project. By analyzing our published proteomic data (Wang W. et al. Mol Cell Proteomics 2014), we noticed that SAV1 hardly formed as a complex with the known Hippo PY motif-containing proteins (e.g. AMOT, PTPN14, LATS1), which is quite different from the other three Hippo WW domain-containing proteins YAP, TAZ and KIBRA (**Figure 1B**). These findings were further confirmed by the pulldown assays (**Figures 1E** and **1F**). To further address the Reviewer's concern, we examined the interaction between SAV1 and LATS1, but hardly detected it as compared with TAZ (**Appendix Figure S1A**). This finding is also consistent with a previous study (Chan E HY, et al. Oncogene 2005).

By analyzing the SAV1-WW domain sequence, we noticed that the E8 residue within the identified conserved amino acid sequence is replaced by serine (**Figures 2A** and **2B**), and the

reverse S/E mutation recovered the interaction between SAV1 and AMOT (**Figure 2G**). These evidences could help explain why SAV1 shows such low binding affinity with the known Hippo PY proteins as compared with the other three Hippo components (i.e. YAP, TAZ and KIBRA).

In addition, based on the Reviewer 3's suggestion, we found that the E/D substitution at the position 8 did not affect the Hippo WW-PY recognition (**Appendix Figure S1B**), indicating that negative charge at the position 8 of the identified conserved amino acid sequence could be essential for the specific Hippo WW-PY interaction. Interestingly, the SAV1 S206 site at this position can be phosphorylated *in vivo* (https://www.phosphosite.org/proteinAction.action?id=1237281&showAllSites=true), suggesting that the interaction between SAV1 and the Hippo PY-motif containing proteins (e.g. AMOT, LATS1) could be regulated by phosphorylation.

Taken together, these evidences could provide a potential explanation for the discrepancy with previous finding, which deserves further characterization.

The structure models for YAP-WW1 mutants based on simulation and the distances to SMAD7-PY reported in Figure 3 have no validation and provide no further information to support their conclusion.

Thanks for pointing out this issue! The aim of this structure model is to present possible mechanisms governing residue-based behavior, specifically in regard to the conserved residues identified in the study. Indeed, our computational models identified two possible overarching functions for the identified amino acid sequence (binding and structure-stabilizing groups) and by presenting these possible groupings in this work, this information can be useful for readers to design future experiment and validation. In addition, we actually included the APBB3-WW domain as a negative control for the YAP-WW1 domain (**Figures 1E**, **1F** and **Appendix Figure S3B**). As shown in the **Figure EV3F**, the simulated APBB3-WW/SMAD7-PY distance is longer than that of the YAP-WW1/SMAD7-PY complex, but close to the YAP-WW1 mutant /SMAD7-PY complexes. By using these control structure models, we think that these simulation studies could provide some mechanistic insights into the identified conserved amino acid sequence that dictates a functional Hippo WW domain for the specific Hippo WW-PY recognition.

Referee #3:

In this manuscript, Vargas and colleagues report their investigation of STXBP4 in the Hippo pathway. The authors have reanalyzed data (their own published work) for WW-proteins/PY interactomes, added some additional purifications, and conducted their own validation experiments to reveal eight amino acids in the WW-domain sequence conferred specificity for binding to PY motifs present in Hippo pathway proteins and a few additional proteins (but not for SAV1). With this WW consensus sequence (unclear how), they performed a proteome-wide search and identified STXBP4, which they then implicate as a cytoskeletal-tension mediated regulator of Hippo. They establish a connection between alpha-catenin and LATS/AMOT that depends on STXBP4, and identify disease mutants that appear to be less functional in the Hippo pathway. A deregulation of STXBP4 was identified in kidney cancers, and overexpression of STXBP4 reduced xenograft tumor formation in one model of renal cancer.

Thanks for the nice summary of our work!

This study would in principle be of high interest to readers of EMBO J as it helps defining a new protein that links mechanotransductive signals to the core Hippo pathway. Yet, the current version of the manuscript is somewhat lacking in critical details, much of which could be fixed by textual changes, and perhaps by either removing some of the modelling data or validating it by additional experiments. Some of the mechanistic details need further validation (i.e. How can a single WW protein interact with multiple PY proteins? Are the interactions detected observed with endogenous proteins? Etc.) and some of the results are less convincing than ideal. In addition, the presentation of the manuscript itself should be streamlined to make it more accessible to the readers.

Thanks for the positive comments on our work! To address the Reviewer's concerns, we have performed additional experiments and revised our text accordingly. Please see details below.

Detailed comments

1. Introduction (p.4-5): The introduction could be shortened and streamlined, notably the lengthy description of all WW proteins associated with Hippo which would perhaps be better suited for a table in the main or supplementary text.

Thanks for the suggestions! The **Introduction** section has been revised accordingly. As for the description of the Hippo-related WW proteins, we have shortened this part by only citing the previously published studies.

2. In the description of the Figure 1 (and other parts of the paper), it should be made clearer which of the proteomics analyses are previously published and which are new. New datasets should be more clearly described in both the Results and Methods section (including replicate analysis, negative controls, etc.). MUSE scoring across the dataset requires description of all other mass spectrometry runs used for scoring, and links to the other data not specifically generated for this publication in order to enable external re-analysis. Similarly, whether the initial results from Wang 2014 were used as initially published or whether a computational

reanalysis was performed should be more explicitly described.

Thanks for pointing out this issue! The proteomic data for four full-length Hippo WW components (YAP, TAZ, SAV1 and KIBRA) were retrieved from our previously published dataset (Wang W, et al. Mol Cell Proteomics 2014). We performed 61 new TAP-MS analyses for a group of full-length WW domain-containing proteins randomly selected from human proteome and the WW domains isolated from these proteins as well as the four Hippo WW components (YAP, TAZ, SAV1 and KIBRA) (**Figures 1B** and **EV1B**). This newly generated TAP-MS dataset has been deposited in the ProteomeXchange Consortium database. We have revised the **Methods** section by including this information. As for scoring, both the published TAP-MS data (i.e. full-length YAP, TAZ, SAV1 and KIBRA) and the newly generated TAP-MS data (i.e. newly included WW domain-containing proteins, the isolated WW domains) were re-analyzed together by the MUSE program (Li X, et al. Mol Cell Proteomics 2016). Details about the MUSE scoring have been included in **Methods** section of the revised manuscript.

3. Results (p6, ¶1): Although it is interesting that SAV1, a WW-containing Hippo pathway protein does not bind Hippo specific PY motifs, if the point of this section of results is that there is a WW sequence that is specific for Hippo-related PY motifs, focusing in on SAV1, an anomaly, detracts from this point. This section should be streamlined.

Sorry for the confusion! Actually, this part of **Results** indicated that the SAV1-WW domain is different from the WW domain of YAP, TAZ or KIBRA in binding the known Hippo PY proteins. Based on these interesting findings, we further examined additional WW domain-containing proteins (**Figures 1C, 1E** and **1F**) and eventually drew the conclusion for the specific Hippo WW-PY recognition. This section has been revised to make it clearer.

4. Figure 2 (C-D) - In the pulldown of AMOT with full-length TAZ (C), AMOT is still able to be recovered despite G/A and E/A mutations, but the TAZ-WW domain (D) only shows faint pulldown of AMOT with the G/A mutation. Does this mean that there are other portions of TAZ that are important for binding to AMOT outside of this Hippo-specific 8 AA WW-related region? Perhaps these mutations are less detrimental to the WW-PY interaction as evidenced by the low degree of conservation, but this is not discussed in the paper at all. It seems that the authors may be oversimplifying the interpretation of their results here.

Thanks for pointing out this issue! Indeed, the G/A and E/A mutations in the identified conserved sequence are less detrimental to the Hippo WW-PY interaction as compared with other identified sites. With a longer exposure, the interaction between the TAZ-WW domain G/A or E/A mutation and AMOT can still be detected, but the binding was largely decreased as compared with wild-type TAZ-WW domain. We have included additional discussion in the revised manuscript to clarify this issue.

5. Have the authors tried more conservative substitutions (e.g. Y to F, E to D, etc.) for their 8 amino acid "motif" to help refine the binding determinants and to help with the proteome-wide (of WW-wide) analysis?

Thanks for the suggestion! By taking the Reviewer's advice, we generated three conservative

substitutions for "E8" (E/D), "Y18" (Y/F, a newly updated site for the conserved amino acid sequence; please also see the response to Reviewer 2) and "F19" (F/Y), respectively. The residue number indicated here is based on the its position within the TAZ-WW domain (**Figure 2A**). As shown in **Appendix Figure S1B**, none of the E/D, Y/F and F/Y substitutions affected the interaction between TAZ/KIBRA and AMOT. Actually, the substitution between "Y" and "F" happens frequently at the positions 18 and 19 of the Hippo WW domains (**Figure 2A**). These data suggest that unique amino acid structure for the positions 18 (Y/F) and 19 (Y/F) as well as negative charge at the position 8 (E/D) are critical for the specific Hippo WW-PY recognition. We have re-searched all the WW domain-containing proteins in human proteome by using this refined Hippo WW criterion (**Figure EV4**).

6. Figure 3 does not add much to the paper in my opinion, especially after the mutational study performed in Figure 2. All results presented are from simulations (though they kind of read like an experimental mutagenesis scanning), based on only 2 structural scaffolds, and while the results appear in agreement with the conclusion of the 8 amino acids as being key for specificity, it is not currently tested experimentally and therefore remains fairly speculative. The authors also do not go outside of the mutants they have tested in Figure 2 (e.g. to expand to more conservative changes). Perhaps a brief mention at the end of the Figure 2 discussion and a supplementary figure or table would be more appropriate than a separate figure ? This part of the paper is not needed for the identification of STXBP4. [The next few comments specifically refer to technical issues with this figure, should it stay in the paper].

We appreciate the Reviewer's comments on the simulation data! The purpose of our computational modeling is to extract potential insights at a molecular level regarding the mechanism of WW domain proteins, since such information may otherwise be inaccessible via current experimental methodologies. These insights could provide information that can be used in future investigation of the Hippo pathway and potential drug design, which is why we chose to include those explorations in this work. We have moved this figure into the Supplemental Information (**Figure EV3**).

7. Results (p.10, ¶2): Although mentioned in the figure legend, it would have been better to be up front in the text that these alanine mutation scan data were derived from simulations. The authors conclude that the large change in RMSD given alanine mutations in the hydrophobic cluster confirm their hypothesis. Perhaps they can also include an actual experimental method to orthogonally show that these mutations result in a change in YAP-WW1 structure instead of relying solely on simulated data?

Thanks for the suggestions! We have revised the related text to indicate that these alanine mutation scan data were generated from simulations.

We agree that an actual experiment method will further strengthen the conclusion. Notably, the isolated YAP-WW1 domain is only marginally stable and highly flexible, suggesting that such a determination may be infeasible with current experimental methodologies. In addition, this experimental validation work could be out of the scope for our current study that revealed the determinants for the Hippo WW domains and STXBP4 as a new regulator for the Hippo pathway. Based on these, we used a combination of the mutagenesis studies (**Figure 2**) and

simulation analyses (**Figure EV3** and **Appendix Figure S3**) to address the molecular basis underlying the Hippo WW domain binding specificity.

8. Figure 3 (C): Is this bond frequency graph calculated over the average of all of the NMR derived structures? What are the authors are trying to convey with this? At what point does a particular interaction become classified as one or the other? Does a lower frequency indicate that the other portion is of a different bond type or is unbonded?

Thank you for pointing out this issue! The bond frequencies are not an average but were generated by summing frames containing a specific interaction over all frames within the YAP-WW1/SMAD7-PY simulations. The purpose of this subfigure was to convey the possible molecular interactions that the conserved residues were likely to participate in by simulation. This information otherwise would be very difficult to obtain via experimental methods. In addition, this could provide useful information to readers who are interested in the specific mechanism of these identified conserved residues for future design of experimental work. The specifically defined metrics were outlined in the **Methods** section (e.g. Baker-Hubbard criteria for hydrogen bonding), where the detailed frequency calculation was stated. Specifically, the most common/dominant interaction for each of the conserved residues was indicated in **Figure EV3C** based on the distance and other metrics as outlined in the **Methods** section.

9. Figure 3 (D): RMSD needs units. They also discuss that the RMSDs are relatively high - there is no discussion as to if this is actually biologically relevant (i.e. the RMSDs may be different, but does a difference of an RMSD of 1.4 A and 4.2 A actually make a biological difference?)

Thanks for pointing out the issue with our RMSD analysis! The RMSD unit is Å, which has been added in the revised figure. Actually, this metric would be very difficult to observe biologically, but could provide insights based on the larger standard deviations and averages as observed for the indicated backside hydrophobic cluster mutants. This metric is to present the relative difference between the identified conserved residues within the Hippo WW domain (i.e. the binding interface group and the backside hydrophobic cluster group). As shown in **Figure EV3D**, the larger standard deviations indicate more movement and flexibility as simulated for the backside hydrophobic mutant residues. During the RMSD analysis, the starting NMR structure of wild-type YAP-WW1 provided a reference point that can make these calculations biologically relevant by means of comparison.

10. Figure 3 (F): With a stable interaction (YAP-SMAD7), the average distance is lower and with a lower magnitude of variation, but the mutants have a high value with higher variation. Are the authors saying that YAP with 7.8 +/- low variation mean binding and anything deviating from that is not binding? Could they have also introduced a negative control mutation either outside of the conserved 8 AA sequence or the WW domain and look at the change in average distance for their simulations? The methodologies and statistics were unclear from the text.

Thanks for the suggestions! Yes, the high values and high standard deviations indicate the substantial movement and flexibility, which may lead to dissociation between two proteins. However, in molecular dynamics models, simulating unbinding events can require up to millisecond timescales. It is close to impossible to simulate protein complex to such extensive

timescales. As for this analysis, we included the mutations for two "W" residues (W/A) as positive controls. As shown in **Figure EV3F**, both W/A mutations yielded similarly long average distance and high variation as compared to the identified conserved amino acid mutations. In addition, we included the APBB3-WW/SMAD7-PY complex as a control, since APBB3-WW contains two unmatched residues (as compared to the identified 9-amino acid sequence) locating in the PY motif binding interface (**Appendix Figure S3B**) and cannot bind Hippo PY protein AMOT (**Figures 1E** and **1F**). Based on our simulation study, the APBB3-WW/SMAD7-PY complex also displayed long average distance and high variation similarly to those of the YAP-WW1 mutant/SMAD7-PY complexes (**Figure EV3F**). Details regarding the methodology and data interpretation were included in the **Results** and **Methods** sections.

11. The transition to STXBP4 is very abrupt. I could not find the parameters for the search through the "proteome". I assume that the authors have restricted themselves to WW-domain containing proteins? Have they then manually scanned all the sequences and found only STXBP4 to contain it? Have they considered the possibility for conservative substitutions that may still retain binding? The authors need to expand their method section to include these critical components. They also need to more clearly explain the results presented in Table S6 that lists other

Thanks for pointing out this issue! During the revision, we actually identified an additional residue "Y18", which is also highly conserved among the Hippo-related WW domains and required for the Hippo WW-PY interaction (**Figure 2**). Moreover, based on the Reviewer's suggestions, we performed conservative substitution analysis and found that the E/D (position 8), Y/F (position 18) and Y/F (position 19) substitutions did not affect the interaction between TAZ/KIBRA and AMOT (**Appendix Figure 1B**). Based on these new findings, we have further improved our binding criterion for the Hippo WW domains and re-searched all the WW domain-containing proteins in human proteome. Since there are only ~50 WW domain-containing protein in human proteome, we completed the searching manually.

In addition, we have included a new **Figure EV4** to better illustrate the searching scheme and result. Additional information regarding the searching process has been included in the revised **Methods** section.

12. STXBP4 association with some Hippo components (including PTPN14 and LATS2) was previously reported in the literature outside of the authors' own studies (see, e.g. the BioGRID database) - these references should be cited here.

Done! The BioGrid references have been included in the revised manuscript. Actually, these large-scale proteomic studies also confirmed our current finding that STXBP4 can form complexes with several Hippo PY motif-containing proteins.

13. The conservation of the 8 amino acid sequence of STXBP4 only in mammals is intriguing, especially in light of the extensive conservation of Hippo signaling. The substitutions in frog and fish appear fairly conservative: could these sequences bind Hippo PY? What could be the significance of a mammalian specific mechanotransducer? Could other proteins functionally replace it, e.g. in Drosophila? (i.e. are there other WW proteins in Drosophila that fits the 8

amino acid criteria established by the authors?)

Thanks for the suggestions! Actually, we are also very interested in these questions. Our recent data showed that the E/D conservative substitution did not affect the Hippo WW-PY complex formation (**Appendix Figure S1B**), suggesting that the fish Stxbp4-WW domain could bind Hippo PY motif-containing proteins. Thus, we have tuned down our statement about STXBP4 in evolution in the revised manuscript.

Since no STXBP4 homolog protein is identified in *Drosophila*, it is possible that some other proteins could act similarly to STXBP4 in *Drosophila*. Interestingly, *Jub* can be recruited by α-catenin to regulate *Warts* activity upon the change of actin cytoskeleton tension (Rauskolb C, et al. Cell 2015), which is similar to our current finding about STXBP4 in mammals.

Based on the Reviewer's suggestion, we have analyzed several Hippo WW domain-containing proteins in *Drosophila*. Interestingly, the *Yorkie* and *Kibra*'s WW domains nicely fit the binding criterion identified in this study, while this is not the case for the *Salvador*'s WW domain (**Appendix Figure S2**).

14. Figure 4 (J): Although latrunculin B and blebbistatin have effects on the cytoskeleton, do they actually modulate cytoskeleton tension directly? If this is not firmly established it may be a large stretch to say that the changes in p-YAP are due to cytoskeletal tension changes and not just proper cytoskeleton structure.

Thanks for pointing out this issue! It is known that the major non-muscle myosin, Myosin II dominantly controls tension within the actin cytoskeleton and targeting the Myosin II activity by blebbistatin results in the loss of actin cytoskeleton tension. We agree that depolymerizing actin cytoskeleton by latrunculin B has more severe effect on the actin cytoskeleton but not only for its tension. Since both inhibitors were included in our current study, we concluded that STXBP4 is involved in the actin cytoskeleton tension-mediated YAP regulation.

15. Figure 5 (C, myc-LATS1 pulldown): Is this really faint band enough evidence to say that STXBP4 promotes the association of a-catenin with LATS1?

We have repeated this experiment. As shown in the updated **Figure 4C**, expression of STXBP4 largely increased the interaction between α-catenin and LATS1.

16. Have the CRISPR-generated clones been sequenced? If so, please add the results of the sequencing to the Methods section.

Yes, the genomic editing for all the STXBP4 CRISPR-KO clones were confirmed by sequencing and the sequencing results can be found in **Appendix Figure S5**.

17. Presumably, each WW domain should be capable of interacting with a single PY motif: this should be considered by the authors in their models - as it is, we are left with the impression that STXBP4 can simultaneously interact with LATS and AMOT through its WW domain.

Thanks for this interesting question! Actually, how STXBP4 functions as a scaffold protein to assemble the complex formation including α-catenin and Hippo PY motif-containing proteins (e.g. AMOT, LATS) is still unclear. Our data have demonstrated that the WW domain is not required for the association between STXBP4 and α-catenin, suggesting that STXBP4 can simultaneously bind α-catenin and at least one Hippo PY motif-containing protein. One possible model is that multiple α-catenin/STXBP4/Hippo-PY complexes could exist and can be spatially clustered together at cell adherens junction. In addition, LATS has been shown to interact with AMOT (Mana-Capelli S, et al. JBC 2018; Adler JJ, et al. PNAS 2013; Dai X, et al. JBC 2013), which makes the current model much more complicated.

18. I am puzzled by Figure 5: first, the rescue of YAP phosphorylation as detected by the Phos-tag gel seems very moderate. How reproducible is this? Has this been quantified? The mutants also do not all have the same effects, with some of them resulting in apparent disappearance of YAP - have phosphatase treatments been performed to confirm that protein levels are the same for YAP across all these panels? This experiment needs to be repeated with clearer readout of what actually happens to YAP phosphorylation as it is certainly not as clear as what the authors infer in the text.

Thanks for pointing out this issue! The phospho-tag gel data has been repeated. The YAP level is not changed in the mutants-expressed cells.

19. Have the authors analyzed the relationships between other Hippo pathway WW or PY components and STXBP4/YAP in the TCGA dataset? Note that much of the TCGA analysis could be moved to supplementary, and this section shortened.

Thanks for the suggestions! Actually, several published studies have already analyzed the Hippo pathway in TCGA (Wang Y, et al. Cell Reports 2018; Sanchez-Vega F, et al. Cell 2018), which include the Hippo WW and PY components. It is known that YAP is an oncoprotein, whose expression and activation are highly elevated in major types of cancers (e.g. breast, colon, lung, ovary, kidney). However, depletion and mutation of the Hippo pathway components in these cancers are rare, suggesting additional oncogenic alterations could exist for YAP activation. Our current findings revealed STXBP4 as a new YAP regulator and a potential tumor suppressor in ccRCC (**Figure 5**), which could provide a mechanism for YAP activation in ccRCC. Based on these reasons, we think that the current TCGA analyses can help readers better understand the significant role of STXBP4 in ccRCC by restricting YAP; thus hope to keep these TCGA data in the main figure.

20. In Figure 6F, what is the YAP and phospho-YAP status? Does this correlate with STXBP4 levels? In 6G/H, are the three STXBP4 constructs expressed at the same levels and uniformly in the xenografts? Importantly, how do the levels of expression of any of these proteins correspond to endogenous levels in normal or cancer cells? As presented, these potentially exciting findings seem preliminary. Note that the authors emphasize the potential connection of STXBP4 with kidney cancers, specifically, but it would be interesting to know whether the effects are broader when STXBP4 is overexpressed (of course, it would be great to see whether deletion of STXBP4 is sufficient to induce cancer formation in a relevant model, but I feel that this is beyond the scope of the current study).

Thanks for the suggestions! We did examine the expression of YAP and phospho-YAP in the panel of RCC cell lines. Unfortunately, these cell lines display large difference in YAP expression, making the comparison of YAP's phosphorylation difficult. Alternatively, we have examined the YAP cellular localization in these RCC cell lines. Interestingly, YAP showed dominant nuclear localization in all the tested RCC cell lines (**Appendix Figure S8A**). This data is consistent with our current model that loss of STXBP4 would promote YAP activation in ccRCC. In addition, we have examined the expression of STXBP4 and its mutants in the 786-O cells that were used for the xenograft assay. As shown in **Appendix Figure S8B**, all the 786-O stable cells showed similar level of exogenously expressed STXBP4 proteins, which is close to that in HEK293A cells.

We totally agree that examining the kidney tumorigenesis using a related *Stxbp4*-deficient mouse model will further strengthen our current conclusion; however, as the Reviewer indicated, it would be out of the scope of our current study.

21. The entire text would benefit from editing: at the moment, parts of the text are difficult to follow. A few examples are listed below, but this is not an exhaustive list.

Agree! We have further improved the languages in the revised manuscript to make it more concise and easier to be followed.

22. Results (p7, ¶1): It is unclear what "To further test this hypothesis..." is referring to. The previous paragraph ends with the conclusion that the WW domain of SAV1 is different from YAP/TAZ/KIBRA. Thus, "this hypothesis" seems to refer to this difference, but the text paragraph actually refers to the hypothesis that YAP/TAZ/KIBRA have a binding specificity that is exclusive to Hippo PY containing proteins. I may be missing something, but was the proteomics data necessary for this analysis? Looking at the sequence alignments of the Hippo-PY-specific WW domains (Figure 2A) versus the control WW domains (Figure S2A), the consensus for the eight AAs in Figure 2A is very strong while the consensus for the same residues in the control WW domains in S2A are relatively weak. Thus, would it not be possible to come to the same conclusion that these residues may be Hippo specific?

Sorry for the confusion! Based on our proteomic analyses and experimental validations, Hippo components YAP/TAZ/KIBRA can specifically bind a group of known Hippo PY motif-containing proteins, but this was not the case for SAV1 and other control WW domain-containing proteins (**Figures 1B, 1C, 1E** and **1F**). Based on these findings, we hypothesized that the Hippo pathway WW-PY recognition could be specific. Indeed, by comparing the WW domain sequence, we revealed such amino acid sequence, which is highly conserved in the Hippo-related WW domains (**Figure 2A**) but not the control ones (**Figure EV2A**). Our experimental studies and simulation analyses have demonstrated and elucidated that these identified residues are the determinants for such binding specificity underlying the Hippo WW-PY protein complexes. This paragraph has been revised accordingly.

23. Results (p8, ¶2): This paragraph begins with "To elucidate the mechanism underlying...specificity..." but the paragraph content concludes with the identification of eight

conserved amino acids. This should be made more clear. Also, the inclusion of SAV1's anomalous sequence detracts from the main point of the paragraph. (p8, ¶3): This introductory sentence is unclear. What did the authors subject the identified sequence to?

Sorry for the confusion! We have improved the languages for these two paragraphs to make them clearer.

24. Results (p9, ¶3): The introductory paragraph would benefit from editing to make this more clear and concise.

Thanks for the suggestion! We have revised this paragraph accordingly.

2nd Editorial Decision 5th Nov 2019

Thank you for submitting a revised version of your manuscript. I apologise for the delay in the processing of your manuscript due to delayed submission of referee reports. Your manuscript has now been re-assessed by two of the original referees, who find that their main concerns have been addressed and are now broadly in favour of publication of the manuscript. There remain only a few mainly editorial issues that have to be dealt with before I can extend formal acceptance of the manuscript.
1. Please address the remaining minor comments from reviewer #2. If you have the data on PIN1/AMOT interaction requested by the reviewer available, you are welcome to add it to the manuscript, but this will not be required for the acceptance of the manuscript.

REFEREE REPORTS:

Referee #1:

The authors responded well to all the points of my critique and the manuscript is significantly improved. No further comments.

Referee #2:

The revised version has addressed most of my concerns. However, I would like the authors to do one more experiment before publication if possible:

PIN1 does not bind to AMOT (Fig. 1E), but PIN1-WW has only one Thr to Ser alteration in the proposed 9-amino acid consensus sequence. This threonine residue is located in the backside of the WW domain and is important for protein folding. Does the PIN1-WW-S/T mutant bind to AMOT like that of SAV1-WW-S/E mutant (Fig. 2G)? If it does, this result would make the conclusion of Hippo-specific "WW-PY" recognition much more convincing.

Minor point:
1) SAV1 should have two WW domains in Fig. EV1.
2) P3, 2nd paragraph, line 3: 'The Hippo is...', missing 'pathway'?

2nd Revision - authors' response 6th Nov 2019

Response to Reviewers

Referee #1:

The authors responded well to all the points of my critique and the manuscript is significantly improved. No further comments.

Thanks!

Referee #2:

The revised version has addressed most of my concerns. However, I would like the authors to do one more experiment before publication if possible:

PIN1 does not bind to AMOT (Fig. 1E), but PIN1-WW has only one Thr to Ser alteration in the proposed 9-amino acid consensus sequence. This threonine residue is located in the backside of the WW domain and is important for protein folding. Does the PIN1-WW-S/T mutant bind to AMOT like that of SAV1-WW-S/E mutant (Fig. 2G)? If it does, this result would make the conclusion of Hippo-specific "WW-PY" recognition much more convincing.

Thanks for the suggestion! Actually, the Thr residue of the WW domain is located at the “WW-PY”

binding interface (**Figure EV3A**) and required for the PY motif peptide association (**Figures 2C-2E**). We agree that examining the ability of the PIN1-WW-S/T mutation to rescue the interaction between PIN1 and AMOT would further strengthen our current conclusion. However, since our current study majorly focused on the difference between SAV1 and the other three Hippo WW domain components (YAP, TAZ and KIBRA) in binding with the known Hippo PY motif-containing proteins (e.g. AMOT), we did not further test this point by using the non-Hippo WW domain-containing proteins.

Minor point:
1) SAV1 should have two WW domains in Fig. EV1.

Thanks for pointing out this issue! In this study, "WW domain" was defined by the presence of two signature tryptophan residues that are spaced 20-23 amino acids apart (please see SMART: WW domain annotation). Actually, the second "WW domain" in human SAV1 (residues 234~267 based on Uniprot) only contains one tryptophan. Therefore, we did not include this region as a "WW domain" in this study. This definition information can be found in the **Methods** section.

2) P3, 2nd paragraph, line 3: 'The Hippo is...', missing 'pathway'?

Thanks! We have corrected this mistake in the text.

3rd Editorial Decision 11th Nov 2019

Thank you for solving the remaining issues. I am now pleased to inform you that your manuscript has been accepted for publication in the EMBO Journal.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Wenqi Wang
Journal Submitted to: EMOB Journal
Manuscript Number: EMBOJ-2019-102406

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	The animal experiment was set up by using five animals per group to detect tumor size difference with the power of 80% and significance level of 0.05 by a two-sided test for a significant study.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	As for the mouse experiments, there was no statistical method used to predetermine sample size. Please see the statement in Methods section (page 35).
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	There were no samples or animals excluded for the analyses in this study. No criteria are pre-established.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	Upon arrival, the animals were randomly assigned into cages with five mice each by the animal facility. We assigned the animals randomly to different groups for injection.
For animal studies, include a statement about randomization even if no randomization was used.	Done! Please see it in Methods section (page 35).
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	A laboratory technician was blinded to the group allocation and tumor collections during the animal experiments as well as the data analyses.
4.b. For animal studies, include a statement about blinding even if no blinding was done	Done! Please see it in Methods section (page 35).
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	NA

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jij.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

Is there an estimate of variation within each group of data?	NA
Is the variance similar between the groups that are being statistically compared?	Yes

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	All the antibody information has been inlcuded in the Methods section (page 23).
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	All the cell line source information has been included in the Methods section (page 24-25) and they were authenticated/tested by the vendors and providers.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	Done! Please see Methods section (page 34).
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	Please see Methods section (page 34).
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	Yes, we confirmed.

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA. The kidney tissue array was purchased from US Biomax, Inc. but not directly obtained from patients.
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	Done! Please see Methods section (page 34). According to the Declaration of Specimen Collection provided by US Biomax, each specimen collected from any clinic was consented by both hospital and individual.
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Done! Please see the Methods section (page 28-29).
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	NA
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA