

Selective integrin endocytosis is driven by α chain:AP2 interactions

Nicola De Franceschi¹, Antti Arjonen¹, Nadia Elkhatib^{2,3}, Konstantin Denessiouk⁴, Antoni G Wrobel⁵, Thomas A Wilson⁵, Jeroen Pouwels¹, Guillaume Montagnac^{2,3}, David J Owen⁵ and Johanna Ivaska^{1,6}

¹Turku Centre for Biotechnology, University of Turku, Turku, Finland

²Institut Gustave Roussy, Villejuif, France

³Institut National de la Santé et de la Recherche Médicale (Inserm), U1170, Villejuif, France

⁴Department of Biosciences, Åbo Akademi University, Turku, Finland

⁵Department of Clinical Biochemistry, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

⁶Department of Biochemistry and Food Chemistry, University of Turku, Turku, Finland

Correspondence should be addressed to J.I. (Johanna.ivaska@utu.fi)

Abstract

Integrins are heterodimeric, cell-surface adhesion molecules comprised of one of 18 α -chains and one of 8 β -chains which control a range of cell functions in a matrix and ligand specific manner. Integrins can be internalised by clathrin-mediated endocytosis (CME) using β -subunit based motifs found in all integrin heterodimers. Whether selective CME also applies to specific integrin heterodimers was unknown. We found that a subset of α -subunits harbour an evolutionarily conserved and functional Yxx Φ motif directing integrins to selective internalisation by the most abundant endocytic clathrin adaptor AP2. We determined the structure of the human (*homo sapiens*) integrin α 4-tail motif in complex with AP2 C- μ 2 and confirmed the binding by ITC. Mutagenesis of the motif impaired selective heterodimer endocytosis and attenuated integrin-mediated cell migration. We propose that integrins evolved to enable selective integrin receptor turnover in response to changing matrix conditions.

Running title: novel integrin endocytosis motif

Introduction

Integrins are heterodimeric transmembrane receptors composed of an α - and a β -subunit. They mediate cell adhesion to the extracellular matrix (ECM) and link it to the cellular cytoskeleton and signalling apparatus. Integrins regulate many physiological events, such as cell motility, cell survival, migration, proliferation and gene expression¹. In mammals, 24 different integrin heterodimers are expressed in a cell type- and tissue-specific manner with each heterodimer fulfilling a specific biological role². The majority of matrix-binding integrins are composed of an α -subunit paired with the common β 1-subunit giving rise to a total of 12 integrin receptors. These bind to a repertoire of ligands, exhibiting a certain degree of overlap in terms of binding specificity³. However, the signalling pathways activated by specific integrin heterodimers are distinct even when triggered by the same extra-cellular ligands. Thus within subsets of integrins with overlapping ligand-binding specificity each receptor fulfils a distinct biological function. Integrin signalling occurs via the short, unstructured integrin cytoplasmic domains and signalling specificity has been assigned to the α -subunit cytoplasmic domains. This is exemplified by the fact that swapping the α -cytoplasmic domains in integrins is sufficient to switch the intra-cellular signalling pathway triggered by the receptor irrespective of the ECM ligand engaged⁴ and integrin-mediated adhesion to collagen via distinct α -subunits triggers either attenuation of receptor-tyrosine kinase signalling or activation of a MAPK signalling cascade^{5, 6} on the basis of distinct protein-protein interactions at the α -cytoplasmic domain.

Integrins are constantly endocytosed and recycled back to the plasma membrane (PM) by multiple routes⁷. Tight regulation of integrin turnover from the cell surface is pivotal to a number of biological processes, including cell migration and cytokinesis^{8, 9}, and is implicated in cancer cell invasion and metastasis¹⁰. Integrins are predominantly endocytosed via clathrin-mediated endocytosis (CME) and this is governed by the intracellular domains of both α - and β -subunits. On the α -subunit, binding of the small GTPase Rab21 and p120RasGAP (RASA1) to the conserved membrane-proximal GFFKR-sequence, shared by all α -subunits, regulates endocytosis and recycling of the receptor, respectively^{11, 12}. Conversely, monomeric clathrin adaptor components Dab2 and ARH associate with conserved segments shared by all β -integrin subunits and are implicated in integrin endocytosis, whereas binding of the small GTPase Rab25 to β -subunits regulates recycling in invasive protrusions¹³. Thus, substantial advances have been made in our understanding of the mechanisms regulating integrin traffic. Nevertheless, the existence of regulatory pathways that would trigger preferential internalisation of one integrin heterodimer over another was unknown. Should they exist, such mechanisms would explain how cells rapidly and efficiently respond to the changing extracellular environment to orchestrate complex biological processes such as cell migration. Therefore, we aimed to identify a mechanism that allows cells to

specifically regulate the traffic of a specific subset of integrins, which is one of the major outstanding issues in the field of integrin traffic.

The AP2 (assembly polypeptide 2) complex is a central player in CME and controls both clathrin-coated pit (CCP) formation and the recruitment of endocytic cargo. AP2 recruitment of receptors is mediated by the μ 2 and σ 2 subunits of the complex. The μ 2 subunit binds to "Yxx ϕ " motifs (Y denotes Tyr; x, any amino acid; and ϕ , a bulky hydrophobic residues) such as that found in the transferrin receptor¹⁴. Here, we identified a previously uncharacterized "Yxx ϕ " motif in a subset of integrin α -cytoplasmic domains and set out to investigate its evolutionary conservation, AP2 binding and functionality in integrin endocytosis on different matrices.

Results

A subset of integrin α chains carries a conserved Yxx ϕ motif

By aligning the membrane-proximal and cytoplasmic regions of all human integrin α -chains, we identified a putative Yxx ϕ internalization motif (where ϕ can be L, I, M, V) present in a subset of α -chains (Fig. 1a) and separating integrins with overlapping ligand specificity into motif-containing and motif-lacking subgroups (Fig. 1b). This suggested that the sequence might be important in differential integrin endocytosis on the same matrix. A related internalization motif, YxxG ϕ ¹⁵, is also present but it is embedded in the membrane and thus not accessible to cytoplasmic interactors. As evolutionary conservation can define the functional importance of a specific sequence, we analyzed the evolutionary distribution of the integrin Yxx ϕ motif. The mechanism of endocytosis is evolutionarily ancient and the AP2 complex is present in almost all eukaryotes¹⁶ while integrins have evolved later¹⁷. Our results showed that in addition to the membrane proximal GFFKR sequence, the most conserved region in all integrin α -chains, the Yxx ϕ motif was highly conserved. The GFFKR motif in the α -integrin cytoplasmic domain was present in all metazoans while the Yxx ϕ motif first emerges in *B. Floridae* (Cephalochordata) (Fig. 1c), suggesting a later appearance of this motif in evolution. Interestingly, the integrin I-domain, an important ligand-binding structural module present in a subset of integrin α -chains, appeared to have evolved only in Tunicata, thus later than the Yxx ϕ motif (Fig. 1c). These events predate the first diversification of the integrins from the more primitive receptor family found in Tunicata to the full set of human-like integrin α -subunits found in *Clupeocephala*, a class of the ray-finned fishes (*Actinopterygii*)¹⁷. Surprisingly, *Clupeocephala* not only shared similarities with humans in the number and types of integrins α -chains but also, with the exception of α 4 (ITA4) and α E (ITAE), in the distribution of the Yxx ϕ motif (Supplementary Fig. 1a). Laminin binding integrins α 3, α 6 and α 7 all harboured

the motif (Fig. 1b). However, a further element of complexity was unveiled by aligning the splice variants of these integrins. These three laminin-binding integrins clustered together based on sequence similarity and ligand specificity¹⁸. Interestingly, they all had splicing isoforms showing modulation of the intracellular domain, exhibiting both isoforms with and without the Yxx ϕ motif (Supplementary Fig. 1b).

Sequence analysis on several mammals¹⁹ revealed that the cytoplasmic regions of some integrins (α M ITAM, α E ITAE, α X ITAX, α D ITAD) exhibited clear sequence variation but the critical residues of the Yxx ϕ motif were strictly conserved, similarly to the GFFKR sequence (Supplementary Fig. 1c). In addition, even in lower organisms¹⁹ the Yxx ϕ motif exhibited stricter conservation compared to the C-terminal cytoplasmic domain in the α 2 (ITA2) and α 4 (ITA4) chains (Supplementary Fig. 1d). The conservation of the GFFKR motif parallels its functional importance in integrin activation²⁰ and integrin traffic^{8, 10}. A similar degree of conservation of the Yxx ϕ sequence suggests that this motif also has a critical role in integrin function.

Integrin α 2 and α 4 Yxx ϕ motifs interact with the AP2 complex

Prompted by the potential functional importance of the Yxx ϕ motif in the integrin α -subunit we tested whether this sequence could mediate binding of integrin α -chains to the μ subunit of the AP complex. *In vitro* His-tagged AP2 μ 2 C-terminal subdomain (C- μ 2) (Supplementary Fig. 2a) exhibited clear binding to the α 2 wild-type (WT) peptide (Fig. 1d). This interaction was significantly impaired upon mutation of the Y and ϕ residues (α 2-AxxA) (Fig. 1d, Supplementary Fig. 2b, c) or following incubation with a competing unlabeled α 2 peptide (Supplementary Fig. 2b, c; we were unable to test the motif-containing full-length biotinylated α 4-integrin cytoplasmic peptide in this assay due to insolubility issues). Integrin peptides lacking the Yxx ϕ motif (including another collagen binding integrin α 1) demonstrated weaker binding to C- μ 2 (Fig. 1d). The strength of α -integrin tail:C- μ 2 interactions were analyzed with Isothermal Titration Calorimetry (ITC) using an α 4 peptide (QYKSILQE) and a longer α 2-integrin (WKLGFFKRKYEKMTKNPDEIDETTELSS) (solubility issues precluded the use of a comparably short α 2 peptide (see later)). These peptides showed clear, robust and reproducible binding to C- μ 2 with a K_D of \sim 70 μ M for the α 4 peptide (Fig. 1e) and \sim 50 μ M for the α 2 peptide (Supplementary Fig. 2d) as compared with an α 5-integrin peptide (YKLGFFKRSLPYGTAMEKAQLKPPATSDA; Supplementary Fig. 2d), β 1-integrin peptide, and an unrelated sequence control peptide, none of which showed any obviously detectable binding i.e. $K_D > 300 \mu$ M. By way of comparison the mainly internally localized proteins with strong endocytic signals TGN38 (DYQRLN), Transferrin receptor

(SYTRFS) and epidermal growth factor receptor (FYRLMS) have K_{DS} for C- μ 2 of $\sim 4\mu\text{M}$, $\sim 15\mu\text{M}$ and $\sim 20\mu\text{M}$, respectively (primary data not shown). In all cases there was a large and similar enthalpic contribution to binding of between 35 and 40kJ/mol, as would be expected since all peptides bind using key Y and ϕ residues and β -augmentation backbone H-bonds, but differ in their entropic counterbalance, a factor which may result from different numbers of water molecules associated with the peptides and their flexibility. *In vivo* all full length transmembrane proteins are embedded in a membrane and there will be less entropy to lose as the tails will be essentially restricted to only lateral diffusion in the plane of the membrane. Therefore, whilst there may be a modest increase in affinity when the interaction is measured on a membrane rather than in solution, there should be a similar increase for all transmembrane cargo proteins. Liposome based SPR measurements using TGN38 immobilised in PtdIns4,5P₂-containing liposome membranes²¹ however show that there is a considerable increase in apparent affinity of a Yxx ϕ signal for holo AP2, which is of around an order of magnitude, for the strongest known endocytic signal, TGN38, due to the co-recognition of the cargo with multiple PtdIns4,5P₂ molecules (affinity for PtdIns4,5P₂ is $\sim 5\mu\text{M}$).

In line with the *in vitro* data, in cells endogenous collagen receptor α 2-integrin, but not α 1-integrin or α 5-integrin, significantly associated with AP2 (Proximity Ligation Assay (PLA) with anti-AP2 α -adaptin) (Fig. 1f, Supplementary Fig. S2e). Taken together, these data demonstrate that the Yxx ϕ motif in α 2- and α 4-integrin interacts with AP2 μ 2 subunit.

Co-crystal of C- μ 2 and α 4-integrin Yxx ϕ peptide

We solved the structure of C- μ 2 in complex with the Yxx ϕ motif (QYKSILQE) from α 4-integrin to 2.8Å resolution (data collection and refinement in Table 1). The details of the integrin motif binding closely resembled those of previously characterized Yxx ϕ endocytic motifs. C- μ 2 is comprised of two inter-linked β -sandwich subdomains (subdomain1 strands 1-6 and 15-17 and subdomain2 strands 7-14) (Fig. 2a). The peptide, for which there is good electron density for residues 2-7 (Fig. 2b,c), formed a β -augmentation with subdomain1, packing next to strand (β 16) making three backbone to backbone H-bonds. The tyrosine and isoleucine side chains sat in complementary pockets lined mainly with side chains from residues on strands 1 and 16 (Fig2 b,c) with all other visible peptide residue side chains pointing into solvent. The walls of the tyrosine side chain pocket were created by Phe174, Lys203, Trp421 and Arg423 with the tyrosine of the motif forming a π -cation bond with Arg423 whilst the tyrosine's OH group forms a hydrogen bond with Asp176 at the base of the pocket. The presence of this negatively charged residue and the size of the pocket would

be expected to preclude the binding of a phosphorylated tyrosine residue due to electrostatic repulsion and steric clashing respectively (Supplementary Fig. 3a). This was indeed the case for the $\alpha 4$ -integrin motif, as the tyrosine-phosphorylated peptide (QY(p)KSILQE) showed no detectable binding to C- $\mu 2$ using ITC (Fig. 1e), and for the TGN38 motif in its tyrosine-phosphorylated form (D.J.O data not shown). The inability to bind a phosphorylated motif could provide a switch to regulate integrin endocytosis, although to date no published evidence exists for tyrosine phosphorylation at this site. The leucine side chain pocket was lined with aliphatic residues and, overall, the interaction of the $\alpha 4$ -integrin:C- $\mu 2$ buried approximately 850\AA^2 of total solvent accessible surface area, compatible with its μM K_D .

In complexes between other proteins and juxtamembrane fragments of integrin α -chain tails, including the region of an α -integrin corresponding to that in the complex with C- $\mu 2$ described here, adopts either a helical^{22, 23} or a largely unstructured conformation²⁴⁻²⁶. Based on our structure, to bind C- $\mu 2$ the α -integrin tail must be in an extended conformation and be accessible from the cytosol. This is in line with a study showing that a serine-phosphorylated $\alpha 4$ -tail, which does contain a Yxx Φ motif, adopts an extended conformation when in complex with 14-3-3 ζ ²⁶. The serine phosphorylated $\alpha 4$ -integrin tail in this structure and the C- $\mu 2$ -bound unphosphorylated $\alpha 4$ peptide were conformationally very similar (Fig. 2d), even though their respective binding partners exhibited totally different folds. Given that the intervening 'x' residues of the Yxx Φ motif pointed away from the C- $\mu 2$ surface into solvent, one would predict that phosphorylation of the Y+2 serine (Ser1011) in the $\alpha 4$ -integrin sequence QYKSIL would have little effect on the affinity of the motif for AP2 (Fig. 2e). In conjunction with our observation that tyrosine phosphorylation of the Yxx Φ -motif in the $\alpha 4$ -tail inhibited the interaction with C- $\mu 2$ (Fig. 1e), these data suggest that phosphorylation events at different sites could potentially function as integrin switches regulating endocytosis and signalling in cells. However, this will need further extensive investigations in the context of the full length integrin in cells.

Yxx Φ motif controls integrin recruitment to CCPs

CME is a major pathway for integrin endocytosis^{8, 27} and cargo recruitment to CCPs on the PM can be visualized in live cells with expression of fluorescent-tagged clathrin light chain (CLC) or AP2 μ ²⁸. We found that endogenous $\alpha 2$ was transiently recruited to AP2 μ -mCherry puncta on the ventral side of the cells with TIRFM (Supplementary Fig. 3b and Movie 1). To investigate this further, we co-expressed wild-type or Yxx Φ motif mutated (Supplementary Fig. 3c) GFP-tagged

$\alpha 2$ - and $\alpha 4$ -integrins with AP2 μ -mCherry in cells. These two integrins were selected for investigation as they recognize different ligands and have biologically distinct functions and expression patterns. Wild-type and AxxA mutant were expressed at similar levels on the cell surface for both $\alpha 2$ and $\alpha 4$ constructs (Supplementary Fig. 3d, e). Interestingly, wild-type $\alpha 2$ - and $\alpha 4$ -integrins, but not their mutant counterparts, colocalized with AP2 μ -mCherry (Fig. 3a). $\alpha 4$ WT also localized to paxillin-positive focal adhesions, while such a localization was limited for $\alpha 2$ WT (Fig. 3a). However, the $\alpha 2$ AxxA mutant exhibited increased localization to FAs compared to the WT (Fig. 3b), suggesting that reduced integrin endocytosis could indirectly result in increased integrin localizing to focal adhesions. $\alpha 2$ was also markedly endosomal compared to the AxxA mutant (Fig. 3c) and overlapped partially with Rab5 and Rab7 and most prominently with Rab21 vesicles, in line with important role of Rab21 in integrin endocytosis²⁹. Overlap with Rab11 and endoplasmic reticulum were minimal (Supplementary Fig. 4). $\alpha 2$ -GFP, in particular the AxxA mutant, also localized to intracellular tubular structures, which partially co-localized with a Rab4-positive compartment (Supplementary Fig. 4). Integrins are endocytosed to vesicles that move along microtubules at an average velocity of 0.5-2 μ m/sec²⁹. $\alpha 2$ WT positive vesicles showed similar dynamics, with abrupt accelerations followed by long pauses and an average speed of 0.9-4.6 μ m/sec. In contrast, $\alpha 2$ AxxA was predominantly localized to the PM (Fig. 3c, Supplementary Movie S2, S3). In cells spreading on collagen, $\alpha 2$ WT showed dynamic vesicular-tubular movements supportive of cell spreading (Supplementary Movie 4), while $\alpha 2$ AxxA was detected mainly in the cell periphery on the PM (Supplementary Movie 5). Taken together, these data show that mutagenesis of the Yxx ϕ motif on $\alpha 2$ - and $\alpha 4$ -integrins alters receptor dynamics on the PM.

Yxx ϕ motif mutation impairs $\alpha 2$ and $\alpha 4$ integrin endocytosis

AP2-mediated CME was found to be important for $\alpha 2$ -integrin endocytosis. Specific siRNA targeting of AP2³⁰ resulted in the downregulation of the AP2 complex as a whole, as shown by α -adaptin staining (Supplementary Fig. 5a,b) and inhibited endocytosis of endogenous $\alpha 2$ -integrin significantly (Fig. 4a; transferrin receptor was used as a positive control⁷). Congruent with this, integrin $\alpha 2$ or $\alpha 4$ AxxA mutants displayed significantly impaired endocytosis when compared with the WT receptor using a well-established biotin-based endocytosis assay⁷ (Fig. 4b,c). The conserved Y in the motif was the most critical residue as mutating it alone was sufficient to disrupt the motif and to inhibit $\alpha 2$ endocytosis ($\alpha 2$ AxxM) to a similar extent than the AxxA mutation (Supplementary Fig. 5c). These effects were predominantly due to impaired endocytosis and not altered integrin recycling since inhibition of recycling with primaquine³¹, did not abolish the

difference between internalization of $\alpha 4$ AxxA mutant receptor compared to internalized $\alpha 4$ WT (Supplementary Fig. 5d). The fact that AP2 silencing or disruption of the motif did not fully abolish integrin endocytosis would be in line with the ability of $\alpha 2$ -integrin to be endocytosed via different pathways including AP-2 independent CME⁴²¹¹, Numb- and a Dab2-dependent pathway from the dorsal side of cells^{32, 33}. For $\alpha 4$ -integrin the internalization routes remain poorly investigated. Integrin activity influences the trafficking kinetics of $\beta 1$ -integrins⁷. In line with this, collagen engagement of $\alpha 2$ -integrin increased endocytosis of $\alpha 2$ -GFP by $40\% \pm 28\%$ when compared to $\alpha 2$ -GFP endocytosis on fibronectin (Fig. 4d). Interestingly, the difference between endocytosed wild-type and AxxA-mutant $\alpha 2$ -integrin levels remained unaltered by collagen adhesion, suggesting that the Yxx ϕ motif is critical especially for uptake of unengaged integrins in cells (Fig. 4d). Accordingly, silencing of AP2 reduced endocytosis of inactive $\alpha 2\beta 1$ -heterodimer by 60% without a significant effect on active $\alpha 2\beta 1$ endocytosis (Supplementary Fig. 5e). According to our data, $\alpha 5$ -tail does not interact with C- $\mu 2$ albeit it harbors a Yxx ϕ motif-like sequence (YxxAM). Accordingly, we found that mutagenesis of $\alpha 5$ -integrin (AxxxA mutant) had no effect on receptor endocytosis (Supplementary Fig. 5f). Taken together, C- $\mu 2$ mediated integrin endocytosis is highly selective to Yxx ϕ motif containing integrins and predominantly regulates endocytosis of inactive integrins.

Introduction of Yxx ϕ to αV integrin stimulates endocytosis

To further investigate the role of the C- $\mu 2$ binding Yxx ϕ motif in integrin endocytosis, we introduced the wild-type (YEKM) or mutant motif (AxxA) from $\alpha 2$ -integrin into αV -integrin, which lacks the Yxx ϕ motif, either by replacing four or two of the endogenous residues with the motif sequences (Fig. 5a). Strikingly, introduction of the wild-type motif induced αV -CFP endocytosis significantly compared to the non-mutated αV -CFP and insertion of the AxxA-motif fully abolished αV -CFP uptake (Fig. 5a). Interestingly, the gain-of-function (GOF) effect was even more prominent in the αV -chimera where two instead of four αV -residues were replaced (Fig. 5a). Given that the affinity status of these GOF-integrins was not tested we cannot rule out the possibility that their increased endocytosis could be linked to altered activity. Taken together, these data demonstrate that the Yxx ϕ motif in integrins is functionally important for AP2-binding and integrin endocytosis.

AP2 regulates integrin endocytosis via the Yxx ϕ motif

As the AP2 complex regulates CME both by direct cargo-recruitment and by mediating the formation of CCPs as a whole¹⁴, we investigated whether decreased integrin endocytosis was a result of a loss of the AP2 μ -integrin Yxx ϕ motif interaction or due to a general effect of AP2 depletion on CME. We generated stable cell lines expressing siRNA-resistant myc-tagged AP2 μ WT or AP2 μ F174A/D176S mutant (AP2 Mut; unable to bind the Yxx ϕ motif but fully functional in other aspects of CME³⁰) and assessed integrin endocytosis in these cells. As previously reported³⁰, endogenous AP2 (lower band) was downregulated upon stable expression of exogenous AP2 μ subunits (Supplementary Fig. 5g) and completely absent following siRNA silencing (Supplementary Fig. 5h). Importantly, expression of the AP2 μ F174A/D176S mutant failed to support α 2 WT and α 4 WT endocytosis while cells expressing AP2 μ -myc WT effectively endocytosed the wild-type integrins (Fig. 5b,c). As expected, the endocytosis of AxxA mutant integrins was not supported by the expression of either AP2 μ WT or AP2 μ F174A/D176S mutant (Fig. 5b,c). These data demonstrate that specific interaction between AP2 and the integrin Yxx ϕ motif is functionally important for the CME of α 2- and α 4-integrins.

Yxx ϕ mutagenesis impairs α 2 and α 4 integrin functions

Integrin traffic has been implicated in cell spreading^{29, 34} and is fundamentally important for efficient cell migration²⁷. In order to assess the effect of the Yxx ϕ motif mutation on cell migration, we expressed α 2- or α 4-GFP WT or their AxxA mutants in cells lacking both endogenous collagen receptors (including α 2-integrin) and α 4-integrin (GD25 β 1A cells)^{35, 36}, and monitored cell migration on collagen or the α 4-integrin specific ligand H120 (fibronectin fragment). Due to the presence of plasma fibronectin in serum (which would complicate the collagen or H120 dependence of the assay by providing a ligand for endogenous α 5 β 1- and α v β 3-integrins expressed in these cells³⁶) these assays were performed in the absence of serum and in the presence of 10 ng/ml HGF, which supports integrin trafficking and stimulates cell migration³⁷. Mutation of the Yxx ϕ motif in α 2- and α 4-integrins strongly influenced cell shape and motility, such that cells expressing the mutant integrins adhered with multiple prominent adhesive protrusions (Fig. 6a) and were severely impaired in cell migration on collagen and on H120, respectively (Fig. 6b). Importantly, these differences in migration were not due to different cell surface expression levels (Supplementary Fig. 5i,j). In more migratory cells, we also observed the presence of retraction fibers when adhesion was mediated via α 2 AxxA mutant on Collagen I (Supplementary Fig. 5k). These data are in line with reduced ability of AxxA mutant integrins to be endocytosed. These data suggest that integrin

endocytosis is linked to cell motility, potentially via indirect effects on adhesion site disassembly as previously demonstrated^{9, 38}, and that the ability of AP2 to recruit a subset of integrins as cargo is critical for the normal function of these integrins on their specific matrix ligands.

Discussion

Here we have identified a previously uncharacterized conserved Yxx ϕ motif present in a subset of integrin α -subunits and demonstrated its functional importance in $\alpha 2$ - and $\alpha 4$ -integrin endocytosis. We propose that the Yxx ϕ motif in the integrin superfamily is a crucial and ancient molecular code, aimed at modulating endocytosis and possibly other trafficking events within the subcellular trafficking machinery. Indeed, the AP-superfamily of adaptors includes at least five members mediating trafficking from different subcellular membrane compartments and they all recognize the same motif-sequence³⁹. Thus this motif may function as a regulator of integrin traffic at multiple distinct steps. Therefore, the discovery of this motif on integrin can potentially open a new perspective into selective regulation of integrin traffic.

In conjunction with the evolutionary diversification of integrins, our findings provide insight into how a mechanism occurred to fine-tune their endocytosis to be specific to only one heterodimer out of several that are expressed in the same cell types at the same time and bind to the same matrix ligand. The ability of a cell to employ the Yxx ϕ motif to discriminate between different integrin heterodimers at the level of endocytosis is likely to be fundamentally important since different α - $\beta 1$ -integrin heterodimers with overlapping ligand specificity have fundamentally different signaling and cytoskeleton regulating functions in cells^{5,6}.

Tetraspanin CD151 is an integrin-binding protein with a Yxx ϕ motif. Mutation of this motif, however, had no effect on the overall integrin endocytosis rate, although it decreased the rate of endocytosis of the CD151-associated integrin⁴⁰. Several proteins binding to the conserved GFFKR motif in the integrin α -subunit have been characterized in the last few years^{20, 41}. Two such proteins are Rab21 and RASA1, which bind the GFFKR motif a mutually exclusive fashion and mediate integrin endocytosis and recycling, respectively¹². These proteins would be in very close proximity to the Yxx ϕ motif when in complex with the integrin α -chain, and may have either synergistic or antagonistic effects on integrin function in adhesion and traffic. Integrins undergo complex cross-talk with multiple kinases^{42, 43}. We find here that tyrosine phosphorylation of the Yxx ϕ motif is incompatible in vitro with C- $\mu 2$ binding. This implies the possibility of an additional level of regulation of integrin endocytosis in cells whereby tyrosine kinases capable of phosphorylating the

motif-containing integrin α -tails could regulate an endocytosis switch and potentially induce retention of a phosphorylated subset of integrins at the plasma membrane. Given that focal adhesions are hubs for active tyrosine kinases, adhesion signaling might regulate endocytosis via such a phosphorylation dependent switch. This, however, remains to be investigated in cells.

Several trafficking regulators binding to the integrin β -tail have been described⁴¹. Numb and Dab2 interact with the membrane-proximal NPXY motif on the integrin β -subunit⁴⁴. Interestingly, both Numb and Dab2 interact with AP2^{45, 46}. Therefore, describing the nature of these interactions in relation to AP2 binding will be a stimulating avenue to explore and critically important to understand how integrin heterodimer specific endocytosis occurs and how it is regulated within the numerous overlapping protein-protein interactions in the adhesome.

Accession code

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 5FPI.

Acknowledgements

M. Humphries (University of Manchester), R. Horwitz (University of Virginia), A. Helenius (Swiss Federal Institute of Technology, ETH), M. Davidson (Florida State University) and M. Robinson (Cambridge University for Medical Research, CIMR) are acknowledged for plasmids and S. Miller (CIMR) for assistance with ITC. We thank J. Siivonen, and P. Laasola for excellent technical assistance, H. Baghirov for help with the image analysis and M. Laasola for graphics generation. Turku Centre for Biotechnology Imaging Core facility and M. Saari for their help with the imaging and H. Hamidi for scientific writing and editing of the manuscript. We gratefully acknowledge the following funding sources: N.d.F. FinPharma Doctoral Program, Instrumentarium Foundation, Orion Research Foundation, Liv och Halsas foundation, Finsk-Norska Medicinska Stiftelsen and the Magnus Ehrnrooth Foundation; J.I. Academy of Finland CoE, European Research Council Consolidator Grant, the Sigrid Juselius Foundation, The Finnish Heart Foundation and Finnish Cancer Organizations. DJO, AGW and TW are funded by Wellcome Trust fellowship 090909 (DJO).

Images reproduced with courtesy of Bob Goldstein (The University of North Carolina at Chapel Hill) and of Phiseksi, Khunaspix, Kamnuan, Apanachai and Anankml at FreeDigitalPhotos.net.

Author contributions

J.I. supervised the cell biological part of the study, carried out experiments, analysed the data and wrote the manuscript with the contribution of N.dF., G.M. and D.J.O. A.G.W, T.A.W and D.J.O carried out the ITC and crystallography. N.dF. conceived the study, designed, carried out and analysed most of the experiments with crucial help from A.A., N.E. and J.P. N.dF. and K.D. carried out the evolutionary studies.

References

1. Campbell, I.D. & Humphries, M.J. Integrin structure, activation, and interactions. *Cold Spring Harb Perspect. Biol.* **3**, 10.1101/cshperspect.a004994 (2011).
2. Wickstrom, S.A., Radovanac, K. & Fassler, R. Genetic analyses of integrin signaling. *Cold Spring Harb Perspect. Biol.* **3**, 10.1101/cshperspect.a005116 (2011).
3. Humphries, J.D., Byron, A. & Humphries, M.J. Integrin ligands at a glance. *J. Cell. Sci.* **119**, 3901-3903 (2006).
4. Hemler, M.E., Kassner, P.D. & Chan, B.M. Functional roles for integrin alpha subunit cytoplasmic domains. *Cold Spring Harb. Symp. Quant. Biol.* **57**, 213-220 (1992).
5. Ivaska, J. et al. Integrin alpha2beta1 mediates isoform-specific activation of p38 and upregulation of collagen gene transcription by a mechanism involving the alpha2 cytoplasmic tail. *J. Cell Biol.* **147**, 401-416 (1999).
6. Mattila, E. et al. Negative regulation of EGFR signalling through integrin-alpha1beta1-mediated activation of protein tyrosine phosphatase TCPTP. *Nat. Cell Biol.* **7**, 78-85 (2005).
7. Arjonen, A., Alanko, J., Veltel, S. & Ivaska, J. Distinct recycling of active and inactive beta1 integrins. *Traffic* **13**, 610-625 (2012).
8. De Franceschi, N., Hamidi, H., Alanko, J., Sahgal, P. & Ivaska, J. Integrin traffic - the update. *J. Cell. Sci.* **128**, 839-852 (2015).
9. Ezratty, E.J., Bertaux, C., Marcantonio, E.E. & Gundersen, G.G. Clathrin mediates integrin endocytosis for focal adhesion disassembly in migrating cells. *J. Cell Biol.* **187**, 733-747 (2009).
10. Caswell, P.T., Vadrevu, S. & Norman, J.C. Integrins: masters and slaves of endocytic transport. *Nat. Rev. Mol. Cell Biol.* **10**, 843-853 (2009).
11. Pellinen, T. & Ivaska, J. Integrin traffic. *J. Cell. Sci.* **119**, 3723-3731 (2006).
12. Mai, A. et al. Competitive binding of Rab21 and p120RasGAP to integrins regulates receptor traffic and migration. *J. Cell Biol.* **194**, 291-306 (2011).

13. Caswell, P.T. et al. Rab25 associates with alpha5beta1 integrin to promote invasive migration in 3D microenvironments. *Dev. Cell.* **13**, 496-510 (2007).
14. Traub, L.M. Tickets to ride: selecting cargo for clathrin-regulated internalization. *Nat. Rev. Mol. Cell Biol.* **10**, 583-596 (2009).
15. Royle, S.J. et al. Non-canonical YXXGPhi endocytic motifs: recognition by AP2 and preferential utilization in P2X4 receptors. *J. Cell. Sci.* **118**, 3073-3080 (2005).
16. De Franceschi, N. et al. Longin and GAF domains: structural evolution and adaptation to the subcellular trafficking machinery. *Traffic* **15**, 104-121 (2014).
17. Chouhan, B.S. et al. Early chordate origin of the vertebrate integrin alphaI domains. *PLoS One* **9**, e112064 (2014).
18. Johnson, M.S., Lu, N., Denessiouk, K., Heino, J. & Gullberg, D. Integrins during evolution: evolutionary trees and model organisms. *Biochim. Biophys. Acta* **1788**, 779-789 (2009).
19. Hedges, S.B. The origin and evolution of model organisms. *Nat. Rev. Genet.* **3**, 838-849 (2002).
20. Bouvard, D., Pouwels, J., De Franceschi, N. & Ivaska, J. Integrin inactivators: balancing cellular functions in vitro and in vivo. *Nat. Rev. Mol. Cell Biol.* **14**, 430-442 (2013).
21. Honing, S. et al. Phosphatidylinositol-(4,5)-bisphosphate regulates sorting signal recognition by the clathrin-associated adaptor complex AP2. *Mol. Cell* **18**, 519-531 (2005).
22. Bhunia, A., Tang, X.Y., Mohanram, H., Tan, S.M. & Bhattacharjya, S. NMR solution conformations and interactions of integrin alphaIbeta2 cytoplasmic tails. *J. Biol. Chem.* **284**, 3873-3884 (2009).
23. Liu, J. et al. Structural mechanism of integrin inactivation by filamin. *Nat. Struct. Mol. Biol.* **22**, 383-389 (2015).
24. Metcalf, D.G. et al. NMR analysis of the alphaIIb beta3 cytoplasmic interaction suggests a mechanism for integrin regulation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22481-22486 (2010).
25. Yang, J. et al. Structure of an integrin alphaIIb beta3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17729-17734 (2009).
26. Bonet, R., Vakonakis, I. & Campbell, I.D. Characterization of 14-3-3-zeta Interactions with integrin tails. *J. Mol. Biol.* **425**, 3060-3072 (2013).
27. Bridgewater, R.E., Norman, J.C. & Caswell, P.T. Integrin trafficking at a glance. *J. Cell. Sci.* **125**, 3695-3701 (2012).
28. Aguet, F., Antonescu, C.N., Mettlen, M., Schmid, S.L. & Danuser, G. Advances in analysis of low signal-to-noise images link dynamin and AP2 to the functions of an endocytic checkpoint. *Dev. Cell.* **26**, 279-291 (2013).

29. Pellinen, T. et al. Small GTPase Rab21 regulates cell adhesion and controls endosomal traffic of beta1-integrins. *J. Cell Biol.* **173**, 767-780 (2006).
30. Motley, A.M. et al. Functional analysis of AP-2 alpha and mu2 subunits. *Mol. Biol. Cell* **17**, 5298-5308 (2006).
31. Somasundaram, B., Norman, J.C. & Mahaut-Smith, M.P. Primaquine, an inhibitor of vesicular transport, blocks the calcium-release-activated current in rat megakaryocytes. *Biochem. J.* **309** (Pt **3**), 725-729 (1995).
32. Nishimura, T. & Kaibuchi, K. Numb controls integrin endocytosis for directional cell migration with aPKC and PAR-3. *Dev. Cell.* **13**, 15-28 (2007).
33. Teckchandani, A., Mulkearns, E.E., Randolph, T.W., Toida, N. & Cooper, J.A. The clathrin adaptor Dab2 recruits EH domain scaffold proteins to regulate integrin beta1 endocytosis. *Mol. Biol. Cell* **23**, 2905-2916 (2012).
34. Sandri, C. et al. The R-Ras/RIN2/Rab5 complex controls endothelial cell adhesion and morphogenesis via active integrin endocytosis and Rac signaling. *Cell Res.* **22**, 1479-1501 (2012).
35. Pellinen, T. et al. Integrin trafficking regulated by Rab21 is necessary for cytokinesis. *Dev. Cell.* **15**, 371-385 (2008).
36. Wennerberg, K. et al. Beta 1 integrin-dependent and -independent polymerization of fibronectin. *J. Cell Biol.* **132**, 227-238 (1996).
37. Mai, A. et al. Distinct c-Met activation mechanisms induce cell rounding or invasion through pathways involving integrins, RhoA and HIP1. *J. Cell. Sci.* **127**, 1938-1952 (2014).
38. Chao, W.T. & Kunz, J. Focal adhesion disassembly requires clathrin-dependent endocytosis of integrins. *FEBS Lett.* **583**, 1337-1343 (2009).
39. Hirst, J., Irving, C. & Borner, G.H. Adaptor protein complexes AP-4 and AP-5: new players in endosomal trafficking and progressive spastic paraplegia. *Traffic* **14**, 153-164 (2013).
40. Liu, L. et al. Tetraspanin CD151 promotes cell migration by regulating integrin trafficking. *J. Biol. Chem.* **282**, 31631-31642 (2007).
41. Morse, E.M., Brahme, N.N. & Calderwood, D.A. Integrin cytoplasmic tail interactions. *Biochemistry* **53**, 810-820 (2014).
42. Ivaska, J. & Heino, J. Cooperation between integrins and growth factor receptors in signaling and endocytosis. *Annu. Rev. Cell Dev. Biol.* **27**, 291-320 (2011).
43. Zaidel-Bar, R. & Geiger, B. The switchable integrin adhesome. *J. Cell. Sci.* **123**, 1385-1388 (2010).
44. Calderwood, D.A. et al. Integrin beta cytoplasmic domain interactions with phosphotyrosine-binding domains: a structural prototype for diversity in integrin signaling. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2272-2277 (2003).

45. Santolini, E. et al. Numb is an endocytic protein. *J. Cell Biol.* **151**, 1345-1352 (2000).
46. Mishra, S.K. et al. Dual engagement regulation of protein interactions with the AP-2 adaptor alpha appendage. *J. Biol. Chem.* **279**, 46191-46203 (2004)..

Table 1 Data collection and refinement statistics (molecular replacement)

	Crystal C μ 2+ α 4peptide
Data collection	
Space group	P6 ₄
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	42.08, 86.93, 86.24
α , β , γ (°)	118.11, 99.65, 95.47
Resolution (Å)	2.75(2.90 - 2.75) ^a
<i>R</i> _{meas}	0.078(0.415)
<i>I</i> / σ <i>I</i>	14.2(3.7)
<i>CC</i> _{1/2} ^b	0.997(0.776)
Completeness (%)	97.2(80.8)
Redundancy	4.9(4.6)
Refinement	
Resolution (Å)	2.77
No. reflections	16115
<i>R</i> _{work} / <i>R</i> _{free}	0.174 / 0.224
No. atoms	
Protein	2130
Water	45
<i>B</i> factors	
Protein	62
Water	48
r.m.s. deviations	
Bond lengths (Å)	0.019
Bond angles (°)	2.1
Data collected off one crystal	

^aValues in parentheses are for highest-resolution shell.

^b *CC*_{1/2} is correlation coefficient between random half-datasets

Figure1: Evolutionarily conserved YxxΦ motif defines a subset of integrin α chains

(a): Alignment of cytoplasmic regions of all α-integrins from *Homo Sapiens*. The conserved membrane-proximal GFFKR region is indicated in blue, the YxxΦ motif in red.

(b): Distribution of the YxxΦ motif in relation to ligand binding specificity of integrins. α-integrin subunits containing the YxxΦ motif are indicated in red.

(c): Taxonomic distribution of the YxxΦ motif across Metazoa. Distribution of the membrane-proximal GFFKR motif and of the I-domain are shown for comparison.

(d): Pulldown assay with recombinant C-μ2 and biotinylated α-integrin peptides. Neg. CTRL= beads alone. Quantification shows binding (Neg. CTRL subtracted) relative to α2 WT. Bars indicate mean±s.e.m.; n=7 (n=independent binding assays); ***, p<0.001 (unpaired Student's *t* test; 2-tails distribution). Uncropped blot presented in Supplementary data set1.

(e): Representative isothermal titration calorimetry of integrin α4 peptide binding to C-μ2 (red). The experiment was performed on six occasions using two different protein preps giving a final overall estimate for the K_D for the interaction of $72 \pm 6 \mu\text{M}$ with $n=1.0 \pm 0.1$. No detectable binding was seen for the same peptide where the tyrosine residue was phosphorylated (phospho α4 integrin) (blue).

(f): Proximity Ligation assay between endogenous AP2 and endogenous α1- and α2-integrins. Numbers indicate mean±s.e.m.; n=3 (n=biological replicates, each one being an independent cell culture); ***, p<0.001 (Mann Whitney test).

Figure2 Structural analysis of the integrin α4 - AP2 Cμ2 complex

(a): Structure of the α4-integrin cytoplasmic domain in complex with C-μ2. C-μ2 is shown with subdomain1 colored pale purple and subdomain2 colored dark purple. The residues in the YxxI motif critical for the interaction are indicated in orange within the rest of the α4 integrin peptide shown in yellow.

(b): Magnified view of α4-integrin peptide binding in the same view as (a) shown on a surface representation of C-μ2. Carbon atoms in the residues critical for binding are colored orange, those from other residues are colored yellow with oxygen and nitrogen atoms colored red and blue. The position where the YxxΦ motif of TGN38 (1BXX) would bind (white) is shown for comparison as determined by overlaying the C-μ2 domains of the two structures.

(c): Stereo representation of the binding of the α4-integrin YKSL (carbons in yellow and orange) to C-μ2 (carbons in lilac). Electron density shown (black) is an omit map contoured at $0.09 \text{ e}/\text{Å}^3$.

(d): Modelling of integrin α4 phosphopeptide (QYKpSIL) in the C-μ2 binding pocket.

Figure3: Yxx ϕ motif controls integrin recruitment to CCPs

(a): Co-localization between GFP-tagged α 2- or α 4-integrins (WT or AxxA mutants) and μ 2-adaptin-mCherry. TIRF plane is shown. Arrows indicate co-localized puncta. Scale bar is 2 μ M.

Quantification of percentage of μ 2-adaptin-mCherry CCPs positive for integrin-GFP. Each spot represents the measurement for one cell; bars indicate mean and s.d. n=2 (n=biological replicates, each one being an independent cell culture. For each n, 5 cells per condition were analyzed); **, p<0.01 (Mann Whitney test).

(b): Recruitment of GFP-tagged α 2-integrin WT or AxxA mutant to focal adhesions visualized by paxillin staining. TIRF plane is shown; scale bar is 2 μ M. Quantification of α 2-integrin GFP signal intensity in FAs compared to cytosol. Each dot represent the value for one cell; bars indicate mean and s.d. n=2 (n=biological replicates, each one being an independent cell culture. In total, 13 cells for each condition were analyzed); ***, p<0.001 (Mann Whitney test).

(c): Localization of GFP-tagged α 2-integrin WT or AxxA in HEK293 cells. Maximum intensity projection of z-stack from confocal microscopy is shown; (scale bar is 10 μ m). Quantification shows mean \pm s.e.m.; n=3 (n=biological replicates, each one being an independent cell culture); *, p<0.05 (unpaired Student's *t* test; 2-tails distribution).

Figure4: Yxx ϕ motif regulates integrin α 2 and α 4 endocytosis

(a): Endocytosis of antibody-labelled cell-surface endogenous α 2-integrin and transferrin receptor in HeLa cells silenced with Ctrl or AP2 μ siRNA. The signal of the internalized antibody was analyzed following 30 min endocytosis. Quantification shows mean \pm s.e.m.; n=3 (n=biological replicates, each one being an independent cell culture. In total, 120 cells were analyzed for transferrin receptor and 300 cells for α 2 integrin); *, p<0.05, ***, p<0.001 (unpaired Student's *t* test; 2-tails distribution).

(b-c): Biotin-based endocytosis assays in HEK293 cells expressing either α 2-GFP (b) or α 4-GFP (c) WT or AxxA mutant. Endocytosis was allowed for the indicated time points and biotin signal was normalized against total α 2 amounts measured from the GFP blot. Bars indicate mean \pm s.e.m. n=3 (α 2) and n=4 (α 4) (n=biological replicates, each one being an independent cell culture); *, p<0.05 (unpaired Student's *t* test; 2-tails distribution). Uncropped blots presented in Supplementary data set1.

(d): Biotin-based endocytosis assays in HEK293 cells expressing either GFP-tagged α 2-integrin WT or AxxA mutant plated on Collagen I or fibronectin (5 μ g/ml) coated dishes. Endocytosis was

allowed for 10 minutes and biotin signal was normalized against total $\alpha 2$ -GFP amount measured from the GFP blot. Bars indicate mean \pm s.e.m. n=3 (n=biological replicates, each one being an independent cell culture); **, p<0.01 (unpaired Student's *t* test; 2-tails distribution). Right hand panel shows that the difference between endocytosed wild-type and mutant integrin remains the same on collagen and fibronectin (endocytosed $\alpha 2$ WT subtracted by $\alpha 2$ AxxA). Uncropped blots presented in Supplementary data set 1.

Figure 5: Introduction of Yxx ϕ motif induces endocytosis of integrin αV

(a): Biotin-based endocytosis assays in HEK293 cells expressing CFP-tagged αV -integrin WT, YEKM gain-of-function mutant or AxxA loss-of-function mutant. Endocytosis was allowed for the indicated time points and biotin signal was normalized against total αV amount measured from the GFP blot. Bars indicate mean \pm s.e.m. n=3 (n=biological replicates, each one being an independent cell culture); **, p<0.01 (unpaired Student's *t* test; 2-tails distribution). Uncropped blots presented in Supplementary data set 1.

(b-c): Biotin-based endocytosis assays in HEK293 cell line stably expressing AP2 μ WT or F174A/D176S mutant, and transiently expressing $\alpha 2$ -GFP (d) or $\alpha 4$ -GFP (e) WT or AxxA mutant. Biotin signal was normalized against total $\alpha 2$ amount measured from the GFP blot. Time point: 15 minutes. Numbers indicate mean; n=2. (n=biological replicates, each one being an independent cell culture). Uncropped blots presented in Supplementary data set 1.

Figure 6: Yxx ϕ mutagenesis impairs $\alpha 2$ and $\alpha 4$ integrin mediated cell migration

(a): Cell shape analysis of phalloidin-stained GD25 $\beta 1 A$ cells expressing either $\alpha 2$ -GFP or $\alpha 4$ -GFP WT or AxxA mutant and adhering to collagen ($\alpha 2$) or fibronectin-fragment H120 ($\alpha 4$). Solidity was calculated as cell body outline (blue) divided by convex hull (yellow). Each dot represents one cell; bars indicate mean and s.e.m.. n=2 (n=biological replicates, each one being an independent cell culture. In total, 8 cells for each condition were analyzed); *, p<0.05; **, p<0.01 (Mann Whitney test).

(b): Quantification of migration of GD25 $\beta 1 A$ cells expressing either $\alpha 2$ -GFP or $\alpha 4$ -GFP WT or AxxA mutant and adhering to collagen ($\alpha 2$) or fibronectin-fragment H120 ($\alpha 4$). Cells were imaged 16 h and the path length of the tracks analyzed. n=3 (n=biological replicates, each one being an independent cell culture. In total, 35 cells for each condition were analyzed); ***, p<0.001 (Mann Whitney test).

Online Methods

Evolutionary analysis

Using the Conserved domain Database ⁴⁷, the sequence of the transmembrane-cytoplasmic region integrin α was found to be the Pfam ⁴⁸ superfamily member *Integrin_alpha* pfam00357 (InterPro entry: IPR018184). Initial analysis of the integrin α cytoplasmic regions was made using the NCBI conserved domain search, which identified 407 sequences from 77 species with the *Integrin_alpha* pfam00357 signature sequence, and 12 known structures. The taxonomy of these sequences was analyzed by hand. For the complete analysis, we have extracted all (3,856) sequences from the UniProt database that belong to the integrin α chain family, among which there were 1171 reviewed sequences, according to UniProtKB/Swiss-Prot (the manually annotated and reviewed section of the UniProt Knowledgebase (UniProtKB)), and 2685 unreviewed sequences, according to UniProtKB/TrEMBL. All incomplete sequences showing lack of the transmembrane-cytoplasmic regions were then discarded. All the remaining sequences were analyzed and existence of the Yxx Φ motif was identified in every sequence by analyzing the segment immediately adjacent to the GFFKR transmembrane motif. Existence of the putative I domain was identified by the subsequent analysis of every UniProtKB integrin α chain sequence through the NCBI Conserved Protein Domain analysis. Sequences potentially containing the I-domains in their structure did show Von Willebrand factor type A (vWA; cd00198) domain consensus signatures, while those that did not, showed only existence of separate FG-GAP (β -propeller; smart00191) repeats. The taxonomy of all identified sequences was analyzed both manually and using the UniProtKB database taxonomy tools. Additional missing annotation was done using the Ensemble automatic gene annotation system ⁴⁹. In Figure 1a, Uniprot entry codes are: P17301, P13612, Q13797, P11215, P20702, Q13349, P38570, P23229, Q13683, P26006, P56199, P08648, P20701, P06756, P53708, P08514, O75578, Q9UKX5.

Antibodies and labelled Transferrin

The following antibodies were used in this study: GFP (A11122, Molecular Probes), β -tubulin (ab6160, Abcam), integrin $\alpha 2$ (MCA2025, Abd Serotech), integrin $\alpha 2$ (EPR5788, Abcam), integrin $\alpha 2$ (AB1936, Millipore), AP2 μ (EP2695Y, Novus Biological), AP2 α -adaptin (MA1-064, Pierce). Secondary antibodies: AlexaFluor488 and AlexaFluor568-conjugated IgGs (A21202, A21206, A10042, A10037, Immunofluorescence; Invitrogen) and HRP-conjugated IgGs (NA931, NA934, immunoblotting; GE Healthcare). Transferrin AlexaFluor568 conjugated (T-23365, Invitrogen). Validations for all antibodies can be found on the manufacturer's websites. In addition, the AP2

adapting and AP2 μ antibodies are validated in this manuscript by silencing as shown in the figures. Original images of immunoblots used in this study can be found in Supplementary data set 1.

Plasmids

The following plasmids were used in this study: pMWH₆C- μ 2(160-435)⁵⁰, Myc- μ 2³⁰, Integrin α 2-GFP²⁹, Integrin α 4-GFP⁵¹, integrin α 5-GFP⁵², dsRed-Rab7⁵³, RFP-Rab5⁵³, dsRed-Rab21²⁹, pPS-CFP2- α V-Integrin (Addgene #57212), mCherry-Rab4a (Addgene #55125), mCherry-Rab11 (Addgene #55124) Point mutations were done by QuikChange II Mutagenesis kit (Agilent technologies) according to manufacturer's instructions. All constructs were checked for integrity and correctness with sequencing.

TIRF microscopy

For total internal reflection fluorescent microscopy (TIRF-M), HeLa cells seeded onto a collagen- (for α 2-GFP expressing cells) or fibronectin- (for α 4-GFP expressing cells) coated glass-bottom dish were transfected with the indicated constructs and imaged the next day through a 100x 1.49 NA TIRF objective on a Nikon TE2000 (Nikon France SAS, Champigny sur Marne, France) inverted microscope equipped with a QuantEM EMCCD camera (Roper Scientific SAS, Evry, France / Photometrics, AZ, USA), a dual output laser launch which included 491 and 561 nm 50 mW DPSS lasers (Roper Scientific), and driven by Metamorph 7 software (MDS Analytical Technologies). A DV2 beam-splitter system (Roper Scientific / Photometrics) mounted on the light path enabled the simultaneous acquisition of the two emission channels. A motorized device driven by Metamorph allowed the accurate positioning of the illumination light for evanescent wave excitation. Mann-Whitney test was used for statistical analysis.

Cells and Transfections

HEK293 cells were grown in DMEM supplemented with 1% penicillin-streptomycin, 10% fetal bovine serum, and 1% L-glutamine. HeLa cells were grown in DMEM supplemented with 1% penicillin-streptomycin, 10% fetal bovine serum, 1% Na Pyruvate, 2% HEPES, 1% non-essential amino acids and 1% L-glutamine. Plasmid and siRNA transfections were done using Lipofectamine 2000 (Life Technologies) and Hiperfect (Qiagen) respectively. HEK293 cells stably expressing AP2 μ rescue constructs were obtained by selection with 250 μ g/ml of G418 (Sigma Aldrich). MDA-231 stable cell line was grown in DMEM supplemented with 10% fetal bovine serum and 0.5% Geneticin. The GD25- β 1A cells were provided by S. Johansson (Uppsala University) and are

described in³⁶. They were grown in DMEM Glutamax supplemented with 10% FCS and 10 μ g/ml of Puromycin. Unless indicated otherwise cell lines are from American Type Culture Collection (ATCC). All cells were routinely tested for mycoplasma and found to be mycoplasma free.

FACS analysis

The FACS staining was performed as described earlier⁵⁴. Briefly, transfected HEK293 or HeLa cells were detached, fixed, washed with PBS and stained with primary antibodies against α 2-integrin (MCA2025, Serotec, 1:100) or α 4 (MAB16983, Millipore, 1:100) or with secondary antibody only in control cells for 1 h. Cells were then washed with PBS and stained with AlexaFluor647-conjugated secondary antibody (ab150107, Invitrogen, 1:400). After washing, cells were suspended in PBS and fluorescence was analysed with flow cytometry (FACScalibur, BD). Cell surface integrin levels (AlexaFluor647 signal) were analysed from the GFP-integrin positive cells.

Structure determination

Residues 160-435 of rat μ 2 (C- μ 2) purified as described earlier⁵⁰ and co-crystallised with the QYKSILQE peptide from integrin α 4 by hanging drop vapour diffusion against 2.2M NaCl, 0.4M NaKphosphate, 20% v/v glycerol, 0.1M MES pH6.5, 5mM DTT. Crystals were of space group P62 (unit cell dimensions 125.6 \AA , 125.6 \AA , 74.4 \AA) and diffracted to 2.8 \AA resolution. Data were collected at 100K on a RIGAKU rotating anode, processed and scaled as in⁵⁰. The structure was solved by Molecular replacement with PHASER and refined to final R and Rfree of 18.2% and 23.2% with sequential rounds of refinement in REFMAC and manual rebuilding.

Isothermal Titration Calorimetry

C- μ 2 was gel filtered into 100mM HEPES pH7.5, 500mM NaCl, 1mM TCEP. Peptides were dissolved in the same buffer. Experiments were performed using a Nano ITC from TA Instruments. C- μ 2 at 100 or 300 μ M was placed in the cell at 10 $^{\circ}$ C and peptides at concentrations from 1 to 3 mM depending on peptide were titrated in with 24 injections of 2 μ l with injections separated by 5 minutes. A buffer into protein blank was subtracted from all data and a minimum of at least four independent runs that showed clear saturation of binding were used to calculate the mean K_D of the binding reaction, its stoichiometry and their corresponding SEM values. Analysis of results and final figures were carried out using the NanoAnalyzeTM Software.

Synthetic Peptides and Recombinant Proteins

Synthetic, biotinylated peptides corresponding to the cytoplasmic domains of human ITGA1, ITGA2 WT, ITGA2 AxxA mutant, ITGA4, ITGA5 and ITGB1 or the unlabeled competing ITGA2 WT peptide were purchased from LifeProtein. Recombinant His-tagged AP2 μ C-terminal subunit (residues 158 to 435) used in pull-down assay was produced in *E. coli* strain Rosetta BL21DE3 and purified according to manufacturer's instructions (BD Biosciences).

For crystallization and ITC, the following unlabelled peptides were used: TGN38 (DYQRLN), Integrin α 4 (QYKSILQE), phosphorylated integrin α 4 (QpYKSILQE), α 2 integrin (WKLGFFKRKYEKMTKNPDEIDETTELSS), α 5-integrin peptide (YKLGFFKRSLPYGTAMEKAQLKPPATSDA), Transferrin receptor (SYTRFS) and EGFR (FYRLMS) and unrelated control peptide (RMSQKIRLLSE).

Pull-down assays

For pull-down assay, equimolar amount of biotinylated peptides were incubated with equal amount of recombinant C- μ 2 in TBSt buffer + 3% BSA at +4C for 2 hours. Streptavidin-Sepharose beads were added and incubated 1 hour at +4C. After washing, the bound protein was detected by western blot analysis analysis. In competition experiments, 10 fold molar excess of unlabelled peptide compared to the biotinylated peptide was added

Immunofluorescence-based and PLA endocytosis assay

Silenced cells were plated on μ -Slide 8 well (Ibidi) coated with Collagen I 5 μ g/ml. Slides were cooled down on ice and cell surface α 2 integrins or Transferrin receptor were labelled either with indicated α 2 antibody and secondary Alexa Fluor 488-conjugated antibodies or 561-labelled transferrin, respectively, in serum-free cell culture medium. Staining medium was aspirated, washed and replaced with fresh culture medium at 37°C. After internalization, plates were lifted on ice. The Alexa Fluor 488 fluorescence on cell surface was quenched by adding anti-Alexa Fluor 488 antibody as described in⁷ and incubating for 1 h on ice or with an acid-wash protocol which carried-out as described in⁵⁵. Cells were fixed in 4% paraformaldehyde (PFAH) for 20 min at room temperature, and samples were imaged by confocal microscope. For the PLA endocytosis assays the fixed cells were permeabilized, stained with 12G10 or 4B4 anti β 1-integrin antibodies and the post endocytosis PLA reaction was carried out according manufacturer's standard protocol (Duolink, Sigma-Aldrich) with specific secondary antibodies recognizing the endocytosed α 2 antibody or the cellular β 1-integrins.

Biotin-based endocytosis assay

Biotin-based endocytosis assays were performed as described earlier⁷. In brief, cells transfected with the fluorescent protein-tagged integrins (wild-type or mutant) were placed on ice, washed once with ice-cold PBS, and surface proteins were then labelled with cleavable NHS-SS-biotin (Pierce) in PBS at 4 °C for 30 min. Labelled cells were subsequently transferred to pre-warmed serum-free medium, and internalisation was allowed at 37°C for the indicated times. Surface-remaining biotin was removed by MesNa cleavage at 4 °C for 30 min followed by quenching with iodoacetamide for 15 min on ice. Cells were lysed and subjected to immunoprecipitation with GFP-Trap beads according to the manufacturer's instructions (Chromotek). The immunoprecipitated integrin released from the beads by boiling in non-reducing Laemmli-buffer. Internalized integrin was blotted with anti-biotin-HRP, and the total immunoprecipitated integrin was blotted with anti-GFP antibody and used for normalization.

Statistical analysis

Statistical significance was analysed using Student's t-test with normal distribution and unequal variance or Mann Whitney test. $P < 0.05 = *$, $p < 0.01 = **$, $p < 0.005 = ***$. No statistical method was used to predetermine sample size.

Online Methods references

47. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222-6 (2015).
48. Finn, R.D. et al. Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222-30 (2014).
49. Curwen, V. et al. The Ensembl automatic gene annotation system. *Genome Res.* **14**, 942-950 (2004).
50. Owen, D.J. & Evans, P.R. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* **282**, 1327-1332 (1998).
51. Parsons, M., Messent, A.J., Humphries, J.D., Deakin, N.O. & Humphries, M.J. Quantification of integrin receptor agonism by fluorescence lifetime imaging. *J. Cell. Sci.* **121**, 265-271 (2008).
52. Laukaitis, C.M., Webb, D.J., Donais, K. & Horwitz, A.F. Differential dynamics of alpha 5 integrin, paxillin, and alpha-actinin during formation and disassembly of adhesions in migrating cells. *J. Cell Biol.* **153**, 1427-1440 (2001).
53. Vonderheit, A. & Helenius, A. Rab7 associates with early endosomes to mediate sorting and transport of Semliki forest virus to late endosomes. *PLoS Biol.* **3**, e233 (2005).

54. Virtakoivu, R. et al. Vimentin-ERK Signaling Uncouples Slug Gene Regulatory Function. *Cancer Res.* **75**, 2349-2362 (2015).

55. Gu, Z., Noss, E.H., Hsu, V.W. & Brenner, M.B. Integrins traffic rapidly via circular dorsal ruffles and macropinocytosis during stimulated cell migration. *J. Cell Biol.* **193**, 61-70 (2011).

Figure 1 De Franceschi et al.

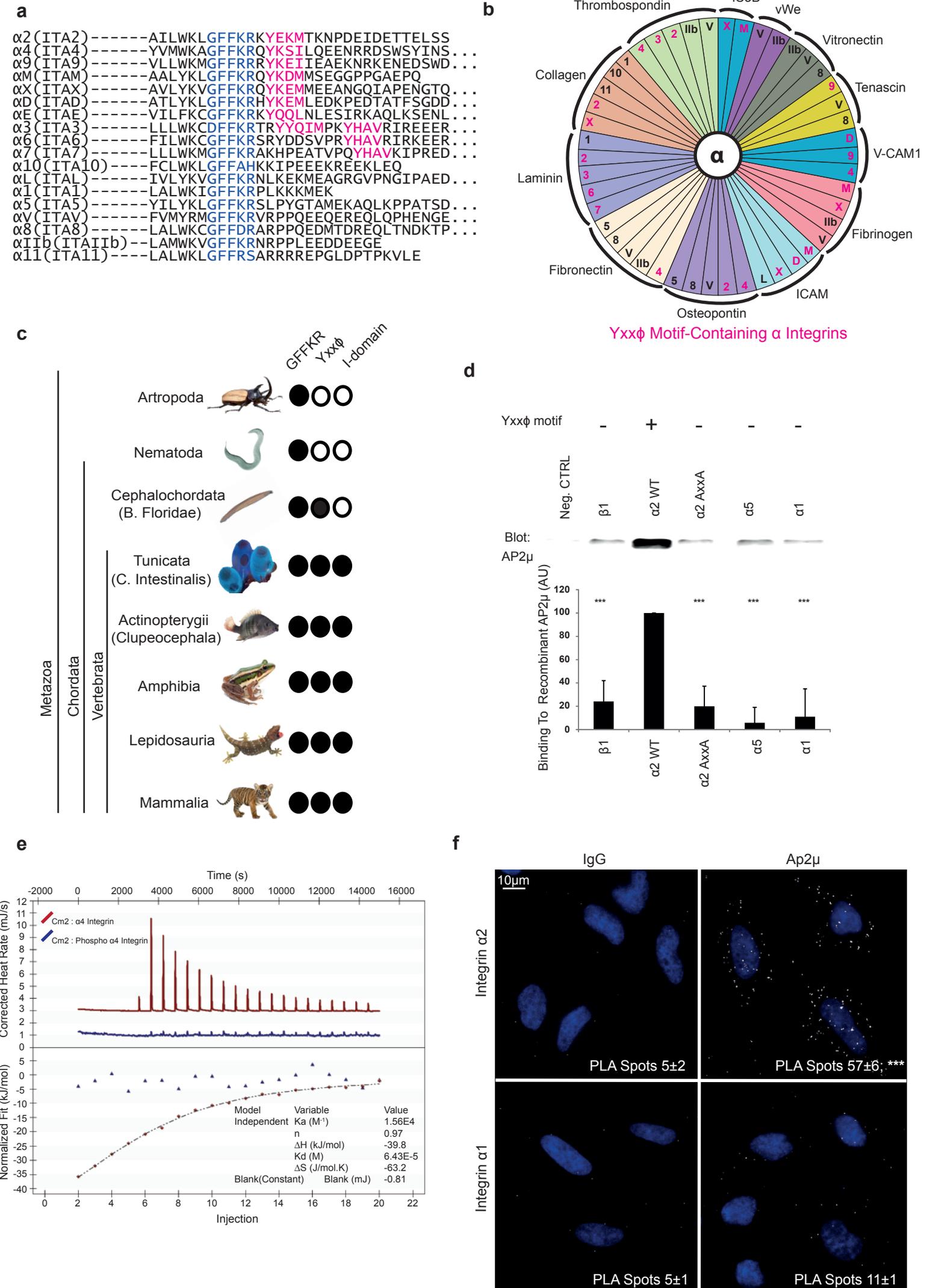
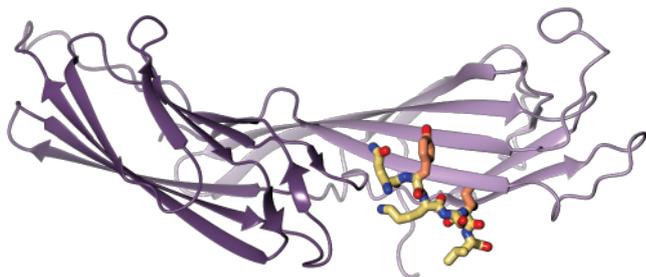
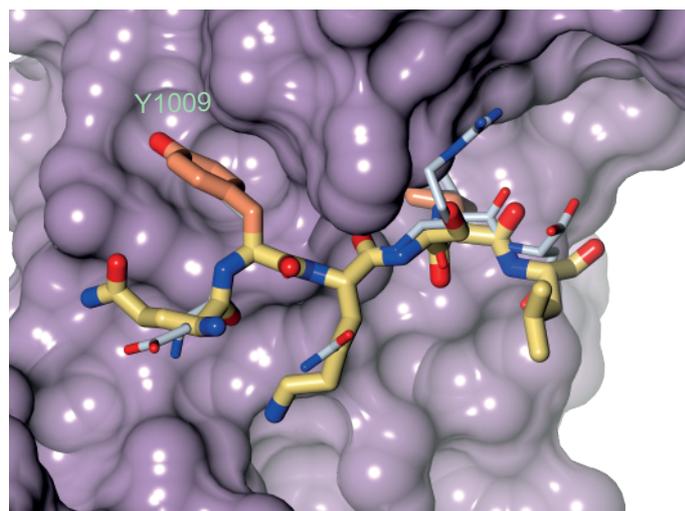


Figure 2 De Franceschi et al.

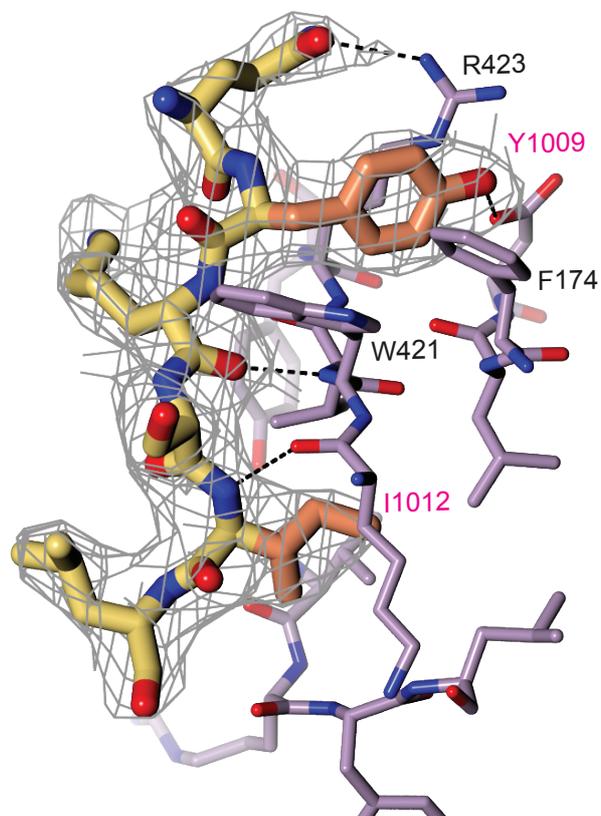
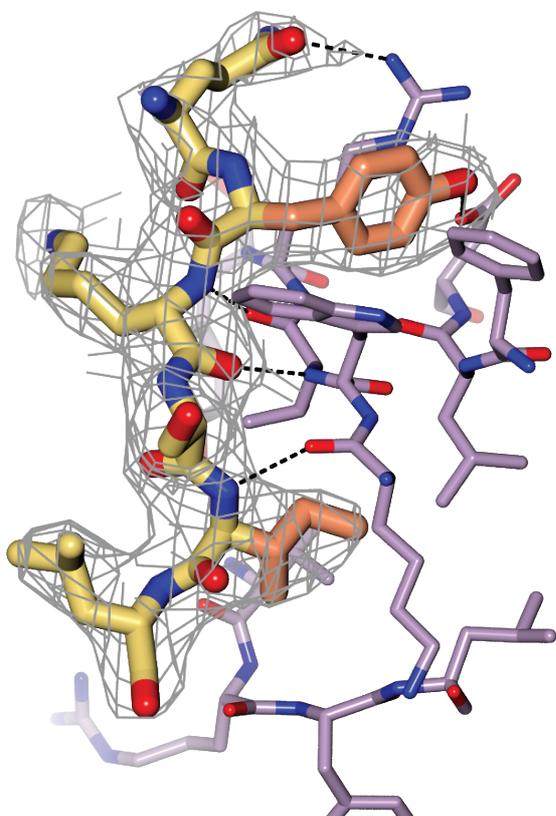
a



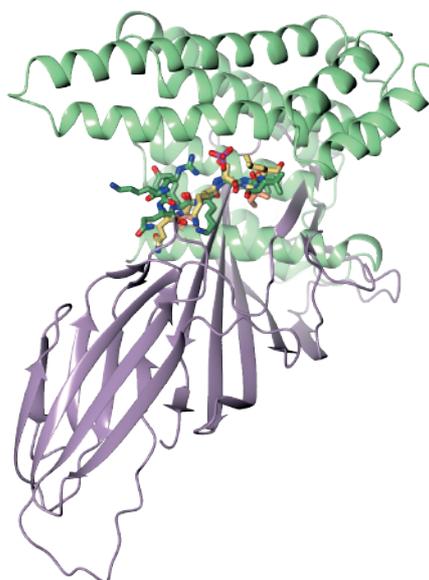
b



c



d



e

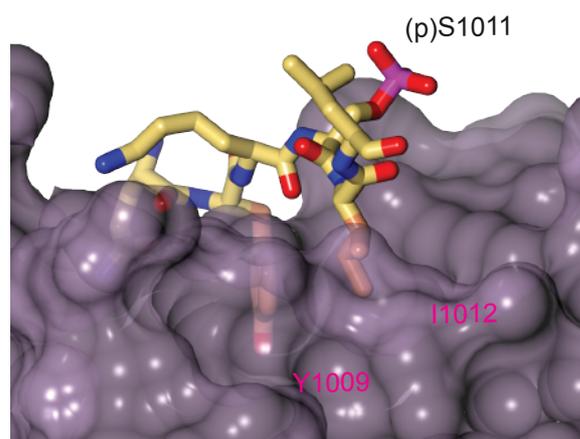


Figure 3 De Franceschi et al.

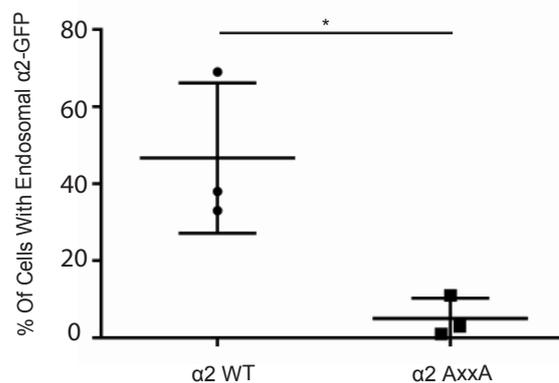
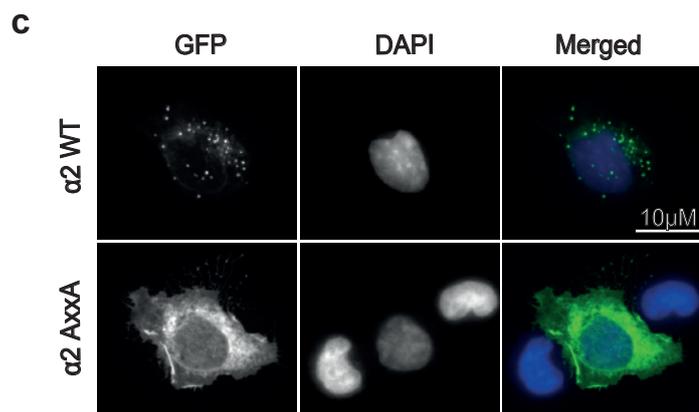
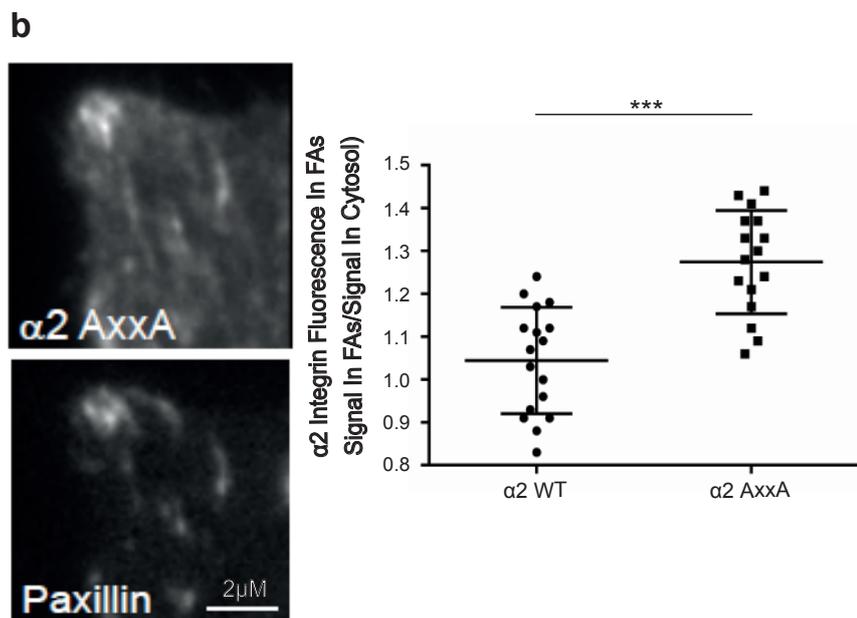
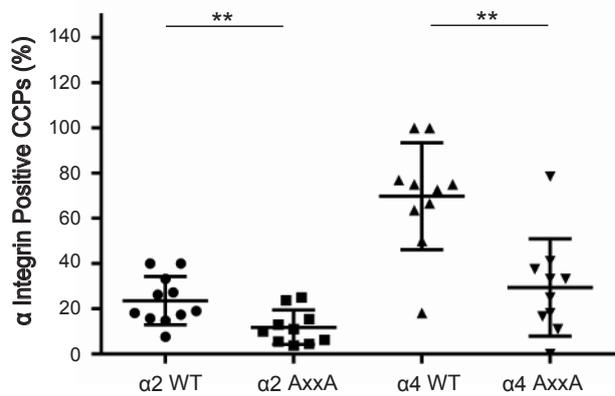
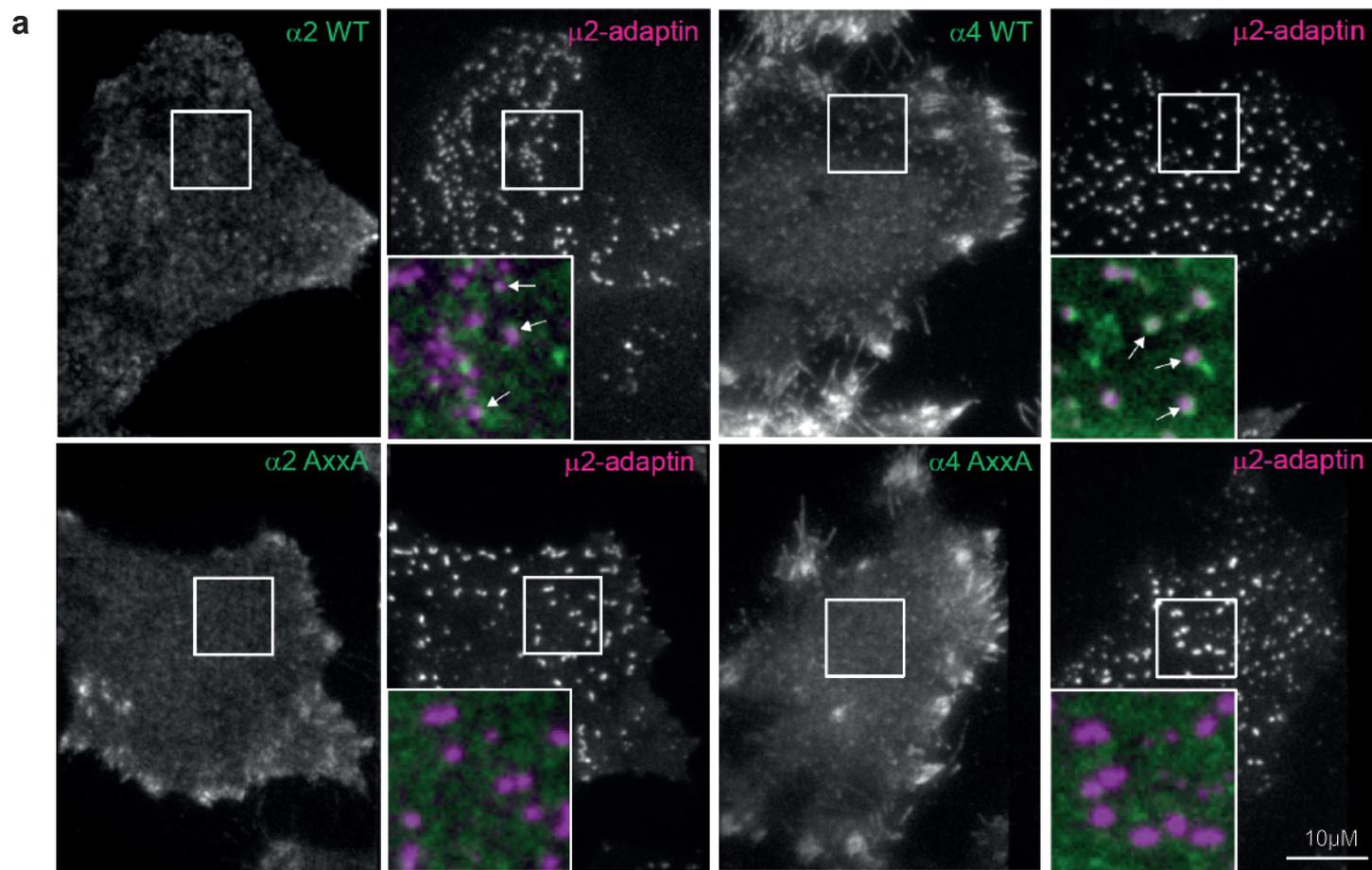


Figure 4 De Franceschi et al.

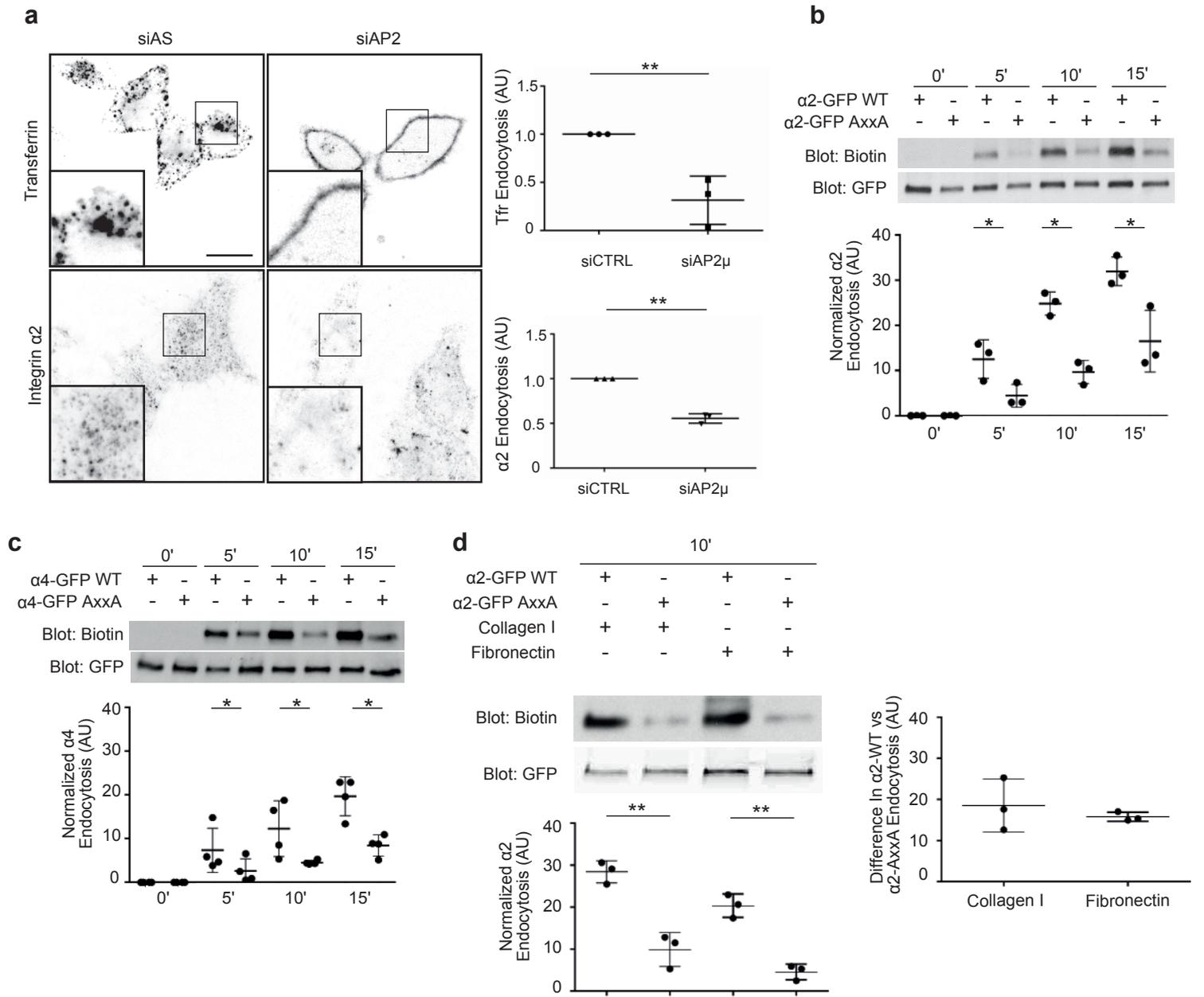


Figure 5 De Franceschi et al.

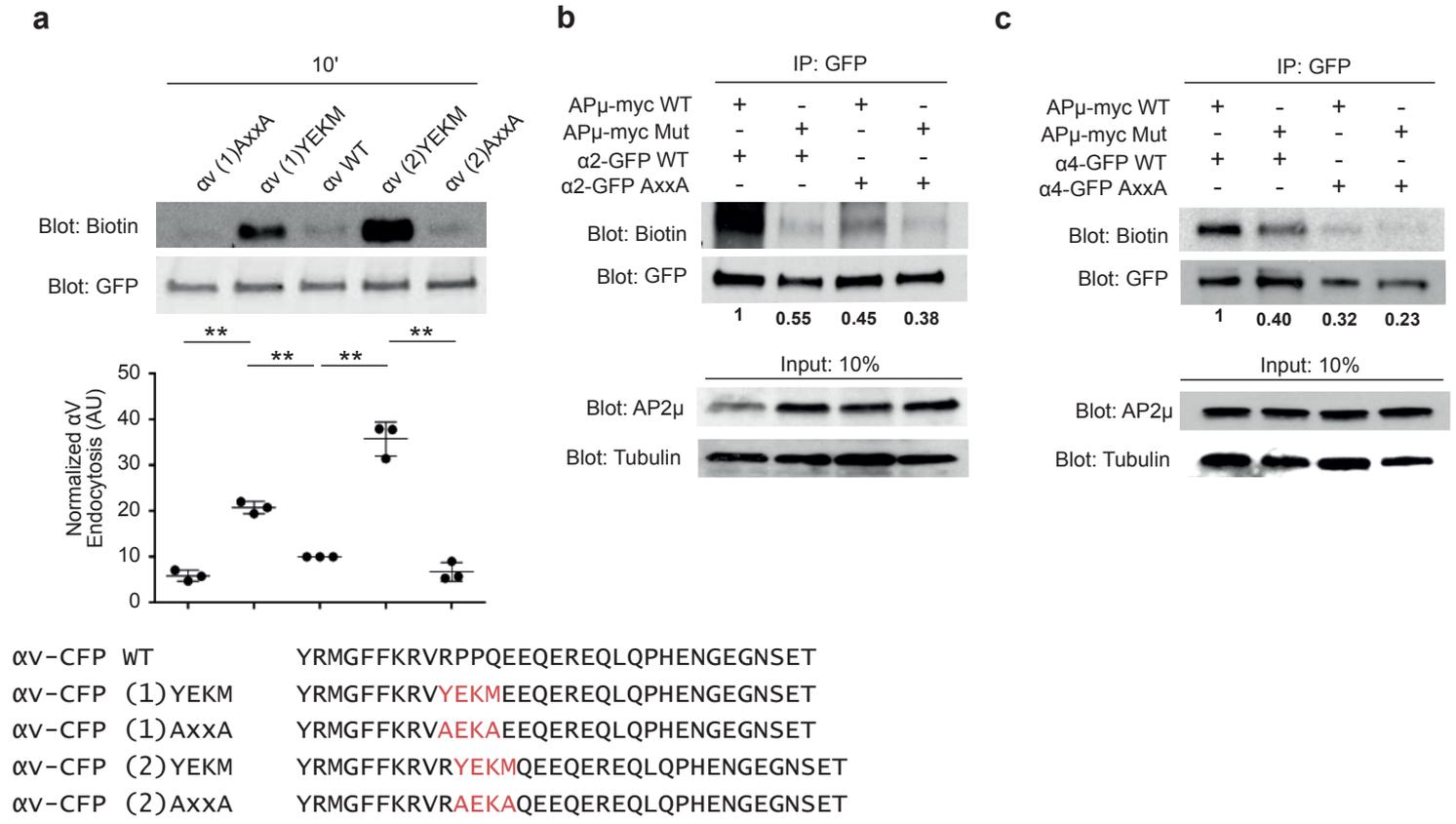
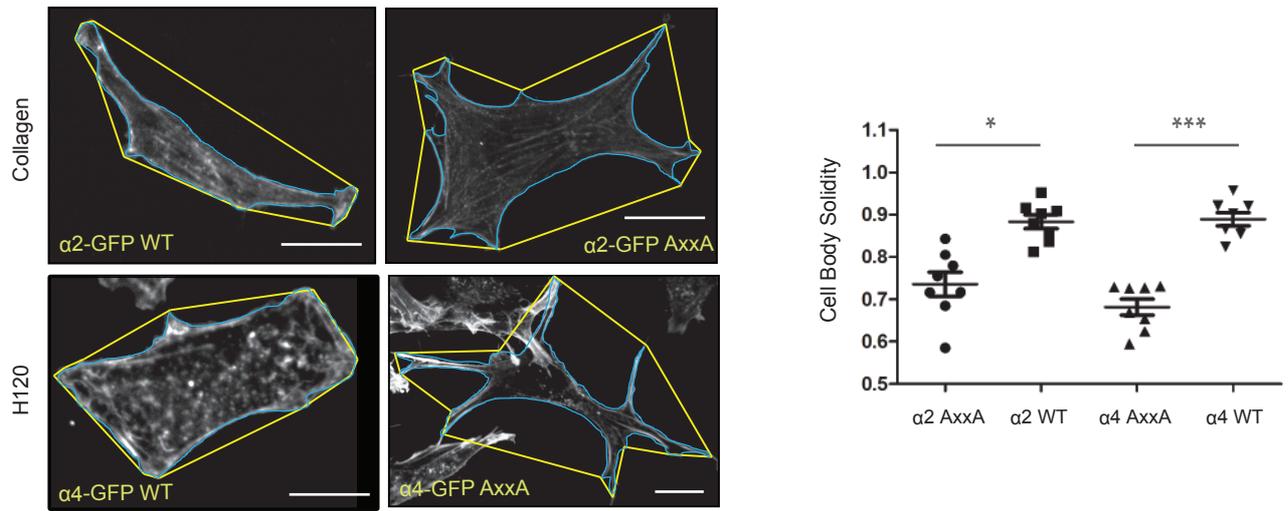


Figure 6 De Franceschi et al.

a



b

