THE UNIVERSITY
of EDINBURGH

Understanding the adhesome network in primary and metastatic cutaneous squamous cell carcinoma

Frederic Li Mow Chee

Doctorate in Molecular and Clinical Medicine

The University of Edinburgh

2018

Declaration

I, Frederic Li Mow Chee, declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree except as specified. The work presented is entirely my own except where stated otherwise by reference or acknowledgment.

Abstract

Understanding how cancer cells survive, invade and migrate is of fundamental importance to the development of approaches to inhibit invasion and metastasis in patients. Proteins recruited at adhesion complexes, known as the adhesome, are involved in multiple mechanisms which control cancer cell behaviour. Here, we used a proteomic and network analysis approach to perform a global assessment of functional molecular units associated with cancer cell progression in the context of the adhesome. We present the functional modules of the adhesome at different stages of human cutaneous squamous cell carcinoma (SCC) progression. From the network analysis, we found that exportin-1, a mediator of protein export from the nucleus to the cytoplasm, is an important hub during cancer progression. Interestingly, one of the interactors of exportin-1 is the actin-regulator Mena. We found that Mena has a nuclear function in metastatic SCC cells and that the nucleocytoplasmic shuttling of Mena is regulated by integrin activation and focal adhesion kinase (FAK). Moreover, our preliminary results suggest that Mena may act as a molecular clutch for the mechanosensing function of Nesprin-2 with actin. The putative molecular clutch activity of Mena regulates the

phosphorylation of EMERIN and possibly affects histone methylation. Here, we propose a novel mechanism by which Mena may regulate metastasis during SCC progression.

Lay Summary

Cancer cells have the ability to move to other parts of the body in a process called metastasis, which is responsible for 90% of cancer deaths. Understanding how cancer cells undergo metastasis is fundamental for the development of drugs which would stop metastasis. Cells are surrounded by a scaffold in the body named the extracellular matrix. The interaction between cancer cells and the extracellular matrix leads to the recruitment and assembly of small parts of a bigger machine called the adhesome. These small parts are protein molecules that together form the adhesome machine that regulates metastasis. We used patient-derived cutaneous squamous cell carcinoma cells to understand the molecular detail of the adhesome. Here, we found that a small piece of the machinery, called Mena, goes to the nucleus of metastatic cancer cells and acts as a clutch to connect the skeleton of the cells to the nucleus. Our results show that this clutch function of Mena is important in enabling the nucleus to sense its surrounding mechanical cues. This allows the nucleus to activate its own machinery, part of which is called EMERIN. The nucleus also responds by inhibiting genes though a chemical

modification, called methylation, of a protein around which DNA coils itself.

Acknowledgements

This thesis is dedicated to my mum, my two siblings and most especially to my father who sadly passed away. I am eternally grateful to my relatives who have been standing by our side during difficult times. This thesis would not have been possible without Michelle from whom I learned so much every day. Finally, this thesis would not have been possible without the support and guidance from my supervisors, Prof. Margaret Frame and Dr. Adam Byron.

Contents

CHAPTER 1

Introduction

■ Introduction

Cell-matrix adhesions are crucial for the invasion and metastasis of malignant cancer. The cell-matrix adhesions are mediated by adhesion receptors such as integrins, which initiate and modulate signalling pathways via multiprotein complexes at adhesion sites. These consist of scaffolding, signalling and cytoskeletal proteins, which are recruited to the cell-matrix interaction site thereby making up the adhesion complex. It is known that adhesome signalling pathways control proliferation, survival, differentiation, migration[1] and drug tolerance[2]. However, despite numerous advances in the field of metastasis and invasion, changes in the adhesome during cancer progression are poorly characterised. How are adhesome proteins precisely controlling cellular functions during cancer progression? Which adhesome proteins are important and what are their roles during cancer progression? To address those questions, the adhesome was isolated from malignant keratinocytes of primary and metastatic squamous cell carcinoma and analysed using a proteomic and network analysis approach to define functional molecular units associated with cancer cell progression. It has previously been reported that oncogenic stress results in the shuttling of adhesion proteins such as focal adhesion kinase (FAK) to the nucleus[3,4]. Moreover, globular actin has also been reported in to be in the perinuclear region[5], though the precise function of perinuclear actin is yet to be defined. Interestingly, we found that actin regulating protein Mena is at the outer nuclear membrane in metastatic keratinocytes though further work is required to define its role there.

■ Adhesion Proteins

Focal Adhesion Activation

Focal adhesions are sites of integrin-mediated cell attachment to the extracellular matrix (ECM), linking the latter to the actin cytoskeleton. Integrin receptors are $\alpha\beta$ heterodimers capable of undergoing conformational change allowing bidirectional signalling events, termed 'inside-out' and 'outside-in' signalling, across the plasma membrane[6]. Integrin attachment to ECM has been suggested to be initiated by an 'inside-out' signalling event whereby talin binds to the integrin cytoplasmic tail resulting in a conformational change of the integrin from an 'inactive' state to a fully 'active' state[6]. In the 'active' state, integrins have a high affinity for ECM ligand. Ligand binding triggers an 'outside-in' signalling event initiating integrin clustering and recruitment of focal adhesion proteins consisting of scaffolding proteins (e.g. VASP, p130Cas), kinases (e.g. Src, Focal Adhesion Kinase (FAK)), cytoskeletal proteins (e.g. actin), phosphatases (e.g. PTP-PEST, PTEN), GTPases (e.g. RhoA, Rac1), proteases (e.g. calpain) and ligases (e.g. E3 ubiquitin ligase Cbl). This recruitment produces multiple protein-protein interactions at the focal adhesion sites thereby allowing the cell to construct signalling complexes regulating diverse cell behaviours[7]. In cancer cells, aberrant activation of integrins and adhesion proteins triggers multiple events such as tyrosine (Tyr) phosphorylation, serine threonine (Ser/Thr) phosphorylation, actin re-organization, mechanosensing and

mechanotransduction leading to tumour cell proliferation, invasion and metastasis as discussed below.

Tyrosine Phosphorylation at Focal Adhesions Triggers Multiple Cell Signalling Pathways

Tyrosine phosphorylation is a key event at focal adhesions, mediated by a range of non-receptor tyrosine kinases such as FAK, Src family kinases (SFKs), c-Abl and Pyk2.

Integrin activation allows the cytoplasmic tail of the integrin β-subunit to relieve FAK of its autoinhibitory FERM domain, resulting in FAK autophosphorylation at Tyr-397[8]. The binding of Src SH2 domain to the phosphorylated Tyr-397 residue (pTyr-397) of FAK, and the binding of Src SH3 domain to FAK, relieves Src of its autoinhibitory fragment. This results in the activation of Src which further phosphorylates FAK at Tyr-576 and Tyr-577 to fully activate FAK. The FAK-Src complex recruits and phosphorylates the adaptor protein p130Cas at the focal adhesions[9,10], which results in its subsequent interactions with members of the Crk scaffolding protein family[10]. Src, with its tight association with FAK, also phosphorylates paxillin at Tyr-31 and Tyr-118, creating binding sites for Crk[11]. Crk interactions with phosphorylated paxillin and phosphorylated p130Cas have been implicated in integrin-mediated cell motility[11]. However, whether paxillin and p130Cas synergize to mediate cell motility is unclear, as it has been reported that both

exert opposite effects on haptotactic cell motility when either is overexpressed in certain cells[12].

The p130Cas-Crk and the FAK-Src complexes interacts with the Rho-family GTPases and guanine nucleotide exchange factor (GEF)- family members to mediate cell migration and cell invasion. Downstream effectors of the p130Cas-Crk complex include the Rho-family GTPase Rac1 and GEF-family member DOCK180, whose activation leads to increased expression of metallopeptidase 2 (MMP2) and metallopeptidse 9 (MMP9) via Jun kinase (JNK) activation[13] and lamellipodia formation[14], thereby promoting cell motility and cell invasion (Figure 1). FAK activates p190RhoGEF and PDZRhoGEF (Figure 1) leading to Rho GTP loading[7], resulting in downstream activation of the formin family protein mDia which serves as an actin nucleator and facilitates actin polymerization required for cell migration[7,15,16]. Moreover, the p85 subunit of phosphatidylinositide 3-kinase (PI3K) interacts with FAK at the phosphorylated Try-397 residue to promote cell migration[17], possibly via members of the Rho family GTPases[18].

Integrin-induced mitogenic signals are transduced by the FAK-Src complex. Src phosphorylation of FAK at Tyr-861 and Tyr-925 allows FAK interaction with Grb2 (Figure 1), thereby linking FAK to the mitogenic Ras-MAPK pathway[7]. Moreover, the p130Cas-Crk complex induced by FAK activates

the mitogenic Rac1-Ras-MAPK signalling pathway (Figure 1), which activates JNK, leading to cell cycle progression and promoting tumour growth[10,19]. Furthermore, FAK activation of the PI3K/Akt pathway has been reported to inhibit apoptosis[20] (Figure 1).

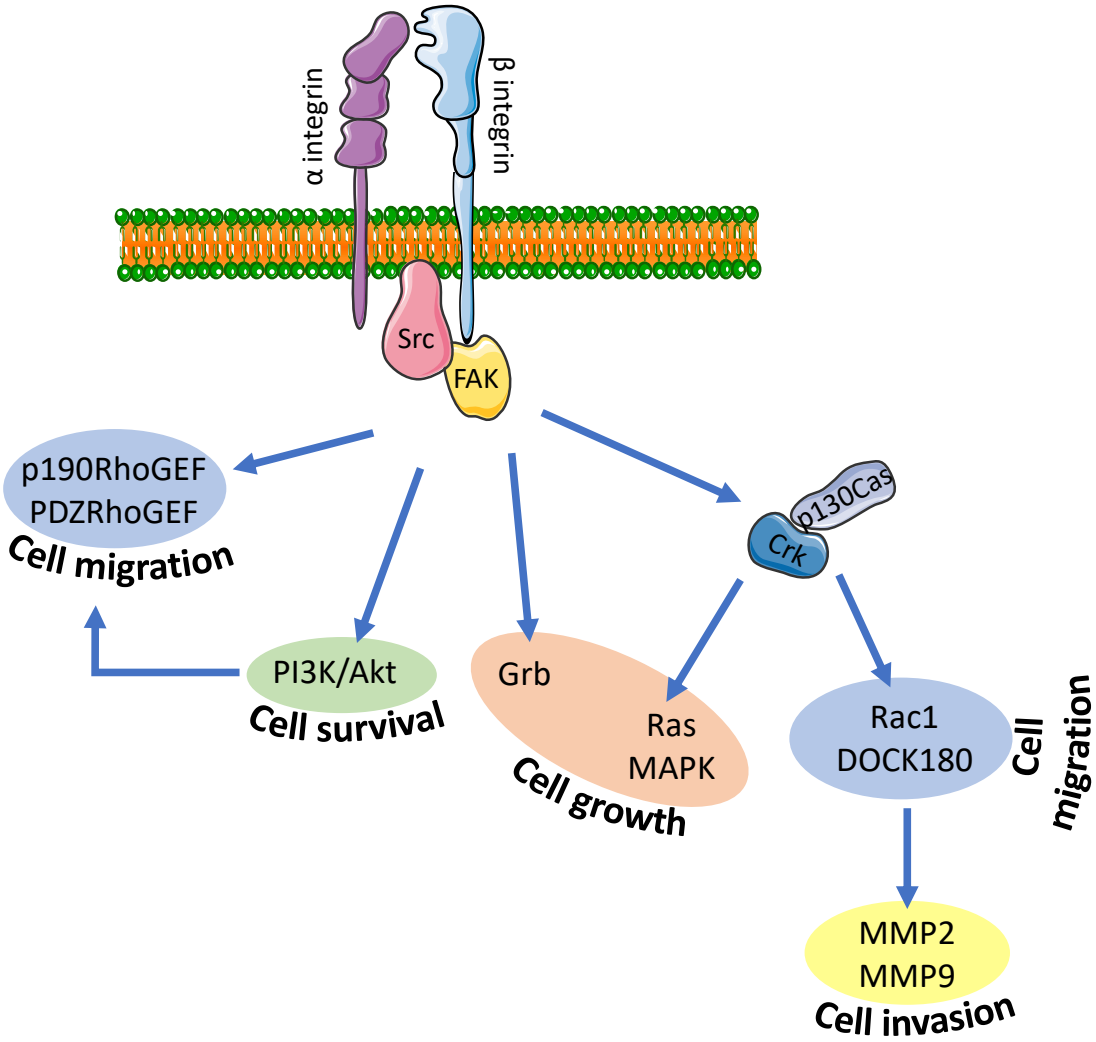


Figure 1 Schematic diagram of FAK-Src and p130Cas/Crk complexes at the focal adhesion with the downstream effectors to regulate molecular processes such as cell migration, cell survival, cell growth and cell invasion.

As discussed above, the FAK-Src and the p130Cas/Crk complexes are recruitment sites for focal adhesion protein tyrosine phosphorylation. Unsurprisingly, dephosphorylation by means of phosphatases is an essential regulatory mechanism at focal adhesions. The deactivation of phosphorylated p130Cas is mediated by PTP-PEST, which is recruited to focal adhesions by the LIM domain of paxillin[11]. Overexpression of PTP-PEST leads to reduced level of phosphorylated p130Cas at focal adhesions and reduced cell migration[21]. The dephosphorylation of p130Cas by PTP-PEST has also been implicated in focal adhesion disassembly. PTP-PEST-ablated cells also manifest a reduced cell migration which is associated with an increase in the number of focal adhesions and hyperphosphorylation of p130Cas[22]. Taken together, this suggests that the correct regulation of phosphorylated p130Cas by PTP-PEST is important for focal adhesion dynamics and turnover, both of which are important for cell migration. FAK is dephosphorylated by PTEN, leading to reduced cell migration, reduced cell spreading and a decrease in the number of focal adhesions formed[23].

The FAK-Src and p130Cas/Crk complexes are also regulated through paxillin, consistent with its role in focal adhesion dynamics. Non-receptor tyrosine kinase Csk interacts with the SH2-binding sites of phosphorylated paxillin in close proximity to the Src binding site. Csk negatively regulates Src activity by phosphorylating the Tyr-527 (pTyr-527) residue of Src. When this residue is phosphorylated, Src undergoes a conformational change

whereby pTyr-527 interacts with its own SH2 domain, rendering the Src protein catalytically inactive[18]. Moreover, upon cell attachment to ECM, c-Abl translocates transiently from the nucleus to focal adhesions, and presumably associates with paxillin[11,24]. c-Abl phosphorylates Crk at Tyr-221, resulting in the uncoupling of p130Cas/Crk complex[25].

Serine/Threonine Phosphorylation at Focal Adhesions

While Tyr phosphorylation is a key element of focal adhesion functions, much less is known about Ser/Thr phosphorylation at focal adhesions. Ser/Thr phosphorylation at focal adhesions is mediated by components of the focal adhesions Ser/Thr kinases such as protein kinase C (PKC) and PAK. PKCs are involved in the assembly and disassembly of, and in the activation of the mitogenic signalling cascade at, the focal adhesions[26]. Classical PKC consists of the α, βI, βII and γ isoforms. PKCα is activated by the integrin-β co-receptor syndecan-4 complexed with phosphatidylinositol 4,5-bisphosphate (PtdIns4,5P2), the latter being essential for the recruitment of PKCα and vinculin to focal adhesions[26]. PtdIns4,5P2 exposes a docking site for PKCα to phosphorylate vinculin, allowing the latter to bind to F-actin and hence connect focal adhesions with the actin cytoskeleton[26,27]. PKCα also mediates the phosphorylation of paxillin at Ser-178 which is required for cell migration. Furthermore, paxillin is also phosphorylated at Ser-188 and Ser-190, which may involve PKC and PAK kinases[11]. Paxillin Ser/Thr

phosphorylation was reported to be required for activation of the MAPK pathway[11].

Mechanosensing at Focal Adhesions

Being at the interface between the cell membrane and ECM, it is not surprising that some components of focal adhesions are involved as mechanosensors, i.e. they respond to mechanical cues and initiate a biochemical response. Integrin-ECM interactions strengthen in response to external forces which in turn leads to the integrin clustering[15,28].

The tension transmitted through the integrins promotes the binding of talin to integrins, thereby linking integrins with the actin cytoskeleton[15]. The globular head domain of talin contains a FERM domain, which binds to the cytoplasmic tail of integrins, while the tail of talin interacts with the actin cytoskeleton via the C-terminal rod domain. Increased actomyosin contractility results in the force-dependent separation of talin's head and tail domains, exposing vinculin binding sites[29]. Recruitment of vinculin at these sites allows p130Cas binding to vinculin via its SH3 domain. As mentioned above, the SH3 domain of p130Cas also binds to FAK. Since FAK and vinculin localise at different layers of the focal adhesions, the FAK-p130Cas or vinculin-p130Cas complexes could perform different roles[15,30]. The FAK-p130Cas complex mediates a strictly signalling complex coupled with Src whereas vinculin-p130Cas could be involved in the early response to force[15].

However, it is also possible that p130Cas may be crosslinking FAK and vinculin in a *trans*- fashion by forming p130Cas homodimers, which upon mechanical stretching, exposes otherwise hidden tyrosine residues that are phosphorylated by Src family kinases[31], thereby acting as a mechanosensing and signalling complex

In addition to talin, other actin binding proteins such as zyxin and VASP also respond to mechanical cues. Indeed, when fibroblasts undergo cyclic stretch, zyxin and VASP relocate from the focal adhesions to sites of force application along the actin stress fibres[32,33]. Actomyosin contraction is required for the localisation of zyxin at these force-bearing sites where it recruits VASP[32,34]. Zyxin interaction with VASP and α-actinin is required for stress fibres maintenance and repair presumably at these force bearing sites[32,34]. Moreover, zyxin and VASP were required for actin stress fibre thickening, thereby reinforcing the stress fibres after cyclic stretching[33]. Zyxin was also observed to undergo nuclear shuttling after cyclic stretching in vascular smooth muscle cells and endothelial cells[32].

The tyrosine phosphorylation of certain focal adhesion proteins in addition to p130Cas described above are also susceptible to mechanical cues. For instance, FAK is phosphorylated at Tyr-397 when mechanical strain is applied[7]. Moreover, phosphorylation of paxillin at Tyr-31 and Tyr-118 is reduced after actomyosin inhibition by bebblistatin treatment[32]. These

observations highlight that phosphorylation events at focal adhesions are intricately linked to mechanosensing.

Moreover, mechanical tension activates Rho GEFs such as Vav2, GEF-H1 and LARG, subsequently inducing GTP loading of Rho[15]. As mentioned previously, activation of Rho family members leads to actin polymerization through the activation of mDia. Another downstream effector of Rho is ROCK which activates the myosin light chain, thereby increasing the myosin-II activity and actomyosin contractility on the actin stress fibres.

Mechanical sensing is not only restricted to focal adhesions but is also present at the nucleus. Mechanical stress can be transmitted from the focal adhesions to the nucleus via the cytoskeleton[35]. The latter is connected to the nucleus via the LINC (linker of nucleoskeleton and cytoskeleton) complex which consists of the nesprins and sun proteins. Nesprin-1 and nesprin-2 are outer nuclear membrane proteins containing a calponin-like actin binding domain at the N-terminal, a central spectrin repeat domain and a KASH domain at the C-terminal. Nesprin-3, which lacks the actin binding domain, attach to the intermediate filaments through interaction with plectin 1. Nesprin-1 and Nesprin-2 connect the actin filament, while nesprin-3 connects the intermediate filament, to inner nuclear membrane proteins SUN1 and SUN2 which in turn bind to the nuclear lamina. The latter is composed of intermediate filaments, namely lamin A, B and C, and membrane associated proteins such as the LEM (lap 2β, EMERIN and MAN1) domain proteins.

The nuclear lamina anchors regions of chromatin known as lamina associated domains (LADs) and could thus be involved with regulation of gene transcription[36,37].

The physical connections between the cytoskeleton, LINC complex and nuclear lamina are required to couple nuclear structure and nuclear biochemistry to mechanical stress sensing[35]. Nesprin-1 stretching results in the phosphorylation of EMERIN on Tyr-74 and Tyr-95 as well as stiffening of the nucleus[38]. This force transmission to the nucleus is reported to be mediated by integrin associated cytoskeletal proteins since tensional force application on integrin also resulted in phosphorylation of EMERIN[38]. Moreover, strain exerted on keratinocytes resulted in a methylation switch presumably regulated by the polycomb repressive complex. Moreover, It was previously reported that force is also transmitted to the nuclear pore complex presumably via the LINC complex, thus regulating the Yap nuclear import/export[39].

Though much is known about the requirement of actin cytoskeleton and actomyosin for mechanical stress transduction to the nucleus, the requirement of actin regulator proteins for mechanotransduction is yet to be determined.

■ Mena: An Actin Regulator in Cancer Progression

Mena is a member of the actin regulator Ena/VASP protein family[40]. Other members of the Ena/VASP proteins family are Evl and VASP, Mena being the mammalian orthologue Ena[40]. The members of the Ena/VASP protein family share three conserved domains: EVH1 domain at the N-terminal, EVH2 domain at the C-terminal and a proline-rich region between the EVH1 and EVH2 domains (Figure 2). An additional domain consisting of LERER repeats, located between the EVH1 domain and proline-rich region, is unique to Mena[40]. The EVH1 domain binds to zyxin and vinculin, which mediate the recruitment of Mena to focal adhesions[40]. The proline-rich domain mediates binding to the SH3 domains of c-Abl and Src, with the former mediating the phosphorylation of Mena on residue Tyr-296[41]. The LERER repeat domain has been reported to interact with α5-integrin. The resulting Mena-α5-integrin complex is required for the 'outside-in' signalling function of $\alpha5\beta1$[42]. The EVH2 domain mediates the interaction of Mena with G- and F-actin as well as mediating Mena oligomerisation. Thus, the EVH2 domain presumably leads to F-actin bundling[40,43].

Ena/VASP proteins localise at the leading edge of the lamellipodia, where they associate with the barbed ends of F-actin and presumably antagonise the capping activity of CapZ[40]. However, the anti-capping activity remains controversial[44]. Moreover, Ena/VASP proteins have also been reported to inhibit Arp2/3-mediated actin filament branching[40]. The expression of

FPPPP-CAAX, which recruits Ena/VASP proteins at the lamellipodial membrane, results in longer and less branched actin filaments there[45]. This leads to slow but persistent protrusion of the lamellipodia[45]. The recruitment of Ena/VASP proteins to the lamellipodia was found to be mediated via the EVH2 domain since Mena lacking the actin-binding site poorly localises at the lamellipodia[43]. Furthermore, the protein Abi and WASP, which also localise at the lamellipodia, interacts with the EVH1 domain and proline-rich region of Ena/VASP proteins, respectively[43]. However, whether these proteins are required for Ena/VASP localisation to the lamellipodia still needs to be determined.

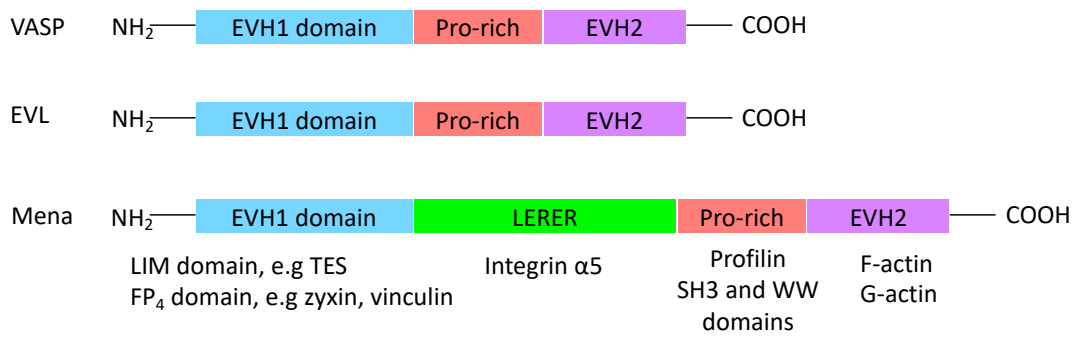


Figure 2 Schematic diagram showing the domains shared by members of the Ena/VASP protein family. They all share the EVH1, proline-rich (pro-rich) and EVH2 domains. The LERER repeat domain is unique to Mena.

Mena Isoforms

Murine cell lines express four splice variants of Mena, namely Mena+, Mena++, Mena(s) and MenaINV (Figure 3). Mena+ (also known as Mena-140) is a neuronal-specific splice variant of 140kDa resulting from the

addition of a proline-rich region after the LERER repeat domain[40,46]. Mena(s) is a small variant of 75kDa missing the proline-rich region. MenaINV is a murine isoform of Mena containing an INV exon (previously known as the +++ exon), which confers a 19 amino acid residue insertion just after the EVH1 domain. MenaINV was found to be upregulated in invasive murine adenocarcinoma cells *in vivo*[47], and promoted transendothelial migration[48].

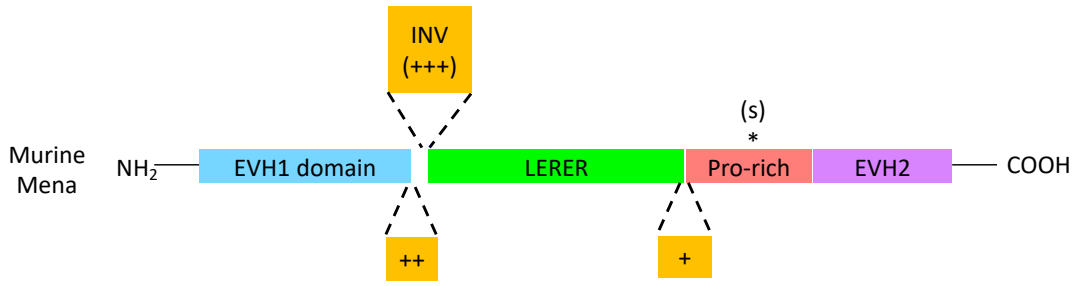


Figure 3 Schematic diagram showing the location of alternately-included (yellow box) or -excluded (denoted by *) exon of murine Mena.

cDNA transcripts of human Mena have been found to contain the +, ++ and INV exon[49]. However, their expression has yet to be confirmed in human cells with the exception of the INV exon, which was recently found to be expressed in human invasive ductal carcinomas[50]. The most characterised splice variants of Mena in human cancer cells are the hMena[51,52], hMena11a[51] and hMenaΔv6[52] (Figure 4). hMena is the 'consensus' human homologue of murine Mena, sharing 87% identity with the latter. The hMena11a variant has an extra exon of 63 nucleotides between exon 11 and exon 12, resulting in the addition of 21 amino acids in the EVH2 domain. hMena11a is associated with epithelial phenotypes[52] and p42/p44 mitogenic

activity[51]. However, hMena11a suppresses invasion *in vivo*[48]. The hMenaΔv6 variant lacks exon 6, which encodes for 37 amino acids between the LERER repeat domain and the proline-rich domain. hMenaΔv6 expression is associated with mesenchymal markers, such as N-cadherion and vimentin expression[52]. Cells transfected with hMenaΔv6 have an increased invasiveness through Matrigel[52].

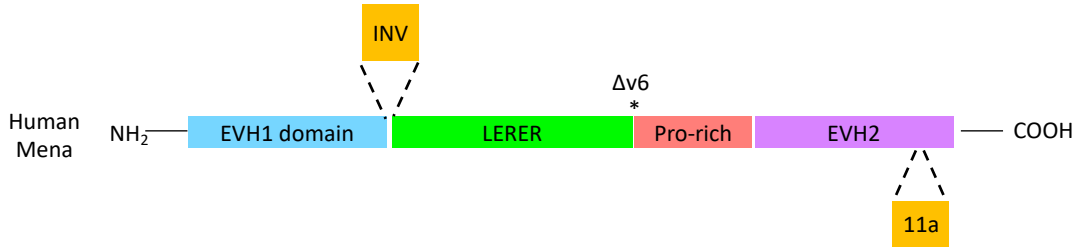


Figure 4 Schematic diagram showing the location of alternately-included (yellow box) or -excluded (denoted by ∗) exon of human Mena.

Mena and Metastasis

Mena expression is clinically associated with invasive cancers, regardless of the different hMena isoforms. Immunohistological staining for pan-hMena was positive in 70% of invasive breast cancer (infiltrating ductal carcinoma)[52] and in 91% of metastatic breast cancer, compared to 9% in low-risk benign diseases and is undetectable in normal breast biopsies[53]. Moreover, pan-hMena expression was positively correlated with tumour size, proliferation index as evaluated by Ki67 staining, Her2 overexpression and cancer stage in breast cancer[53]. Mena overexpression was also reported in synchronous (found at the time of primary diagnosis) and metachronous (found at the time

of recurrence) metastases of breast cancer, namely in lymph nodes, lung, bone, brain and pleura[53]. Mena was overexpressed in colorectal adenomatous polyps with high dysplasia but was not expressed in either normal colorectal mucosa or adenomatous polyps without dysplasia[54]. In colorectal carcinoma, Mena expression correlated with vascular invasion, tumour budding, tumour stage, Her-2 and p53 expression[54,55]. Likewise, Mena expression was correlated with tumour grade of cervical intraepithelial lesion[56]. Moreover, all invasive cervical squamous cell carcinomas overexpressed Mena[56]. Despite high Mena expression is correlated with metastasis, the overexpression of Mena has no effect on cancer survival (data not shown).

Numerous studies showed that MenaINV confers metastatic and motility properties *in vivo* and *in vitro*. Rat mammary adenocarcinoma cells expressing MenaINV were reported to sensitize cells to epidermal growth factors (EGF), thereby promoting chemotaxis at extremely low concentration of EGF. MenaINV also promoted multicellular streaming *in vivo* which is dependent on a paracrine signalling loop with macrophages[48]. Moreover, MenaINV-expressing cells have a higher ability to intravasate and the presence of macrophages potentiates these cells to undergo transendothelial migration[48,57]. Consistent with these findings, sections of human invasive ductal carcinomas of the breast with high immunohistological Mena staining were associated with perivascular macrophages[48].

Mena has an important function in metastasis in mediating cell motility and cell invasion of metastatic cells. However, it is possible that Mena has other unknown functions within metastatic cells. Intriguingly, Mena was recently reported to regulate local translation of mRNAs in the axon of neurons through the interactions with RNA-binding proteins heterogeneous nuclear ribonucleoprotein K (HnrnpK) and Poly(rC)-binding protein 1 (PCBP1)[58]. This highlights the possibility that Mena functions beyond focal adhesions, actin stress fibres and lamellipodia.

■ An Integrative View of the Adhesome in Cancer Progression

The Adhesome

In 2001, an early literature-curated list of cell-matrix adhesion proteins consisting of approximately 50 proteins including ECM receptors, scaffolding proteins, kinases, cytoskeletal proteins, phosphatases, GTPase modulators and proteases, was published[59,60]. Six years later, another literature-curated list of 156 proteins associated with integrin-mediated adhesion, termed the 'adhesome', was published[61]. After a further three years, the inventory of the literature-curated adhesome grew to 180 proteins[62]. Despite the progress, it was acknowledged early on that the literature-curated list does not cover the

full repertoire of adhesion proteins, since the search for new adhesion proteins had, until recently, been based on candidate-based immunolabeling fluorescence microscopy and biochemical or immunochemical binding assays[59]. Recently, multiple laboratories have described the systematic isolation and identification of components of the adhesome by mass spectrometry, allowing unprecedented depths of adhesome analysis[63]. These developments led to the identification of novel components of the adhesome[64], provided new insights in the mechanosensing regulatory pathways of focal adhesions[65,66] , and defined the role of integrin activation in regulating microtubule networks[67]. However, a global assessment of the adhesion signalling during cancer progression is currently lacking.

A meta-analysis of fibronectin-induced adhesome identified by mass-spectrometry from different laboratories led to the construction of the 'meta-adhesome'[68,69]. This consists of 2000 proteins, highlighting the complexity of the integrin-mediated adhesion nexus. How does the adhesome contain so many proteins? The adhesome isolation methods capture the transient interactions of multiple proteins with the adhesion complexes which were not previously appreciated by using low-throughput immunofluorescence microscopy. Moreover, the focal adhesion proteins have multiple binding sites that can assemble in numerous ways with different proteins.

The execution of metastasis requires continuous interactions between the cell surface receptors integrins and the extracellular matrices (ECM) such as fibronectin, collagen IV, laminin and vitronectin[70,71]. Moreover, the analysis of the adhesome network have previously revealed proximal and distal functions of the adhesion proteins to cell matrix-contacts in cancer[72]. Indeed, there is mounting evidence that focal adhesion proteins have activities beyond the adhesion sites in cancer cells. FAK for instance was reported to shuttle to the nucleus during tumourigenesis to mediate immune evasion[4]. LIM domain protein, paxillin and zyxin were also found to shuttle though their nuclear functions remain elusive.

The Adhesome Proteomics

Over the past several decades, cancer biologists have been seeking alterations in the cancer proteome, be they at cellular, tissue or serum level, in order to understand the mechanisms of cancer initiation, progression and metastasis[73]. Proteomic analysis has been challenging due to the high diversity of proteins present in cells, not least due to alternative splicing producing a range of protein isoforms[74]. Moreover, the proteomic complexity is further amplified with post-translational protein modifications (e.g. phosphorylation and ubiquitination) which affect protein dynamics and proteins interconnectivity in signalling networks[74]. However, with the advancements of mass spectrometry instrumentation, sample preparation

and computational analysis have allowed for unprecedented depths of analysis in cancer proteomics[73,74].

The first step in any proteomic experiment is the protein extraction from the cells. Recent advancement was made in the extraction of the adhesome which allowed for subsequent mass-spectrometry analysis[75]. The adhesome was isolated from malignant keratinocytes derived from cutaneous squamous cell carcinoma (SCC) at the primary site and at secondary metastatic sites. These cell lines are referred to as the Met series and they reflect the clinical progression of SCC[76]. The Met series is composed of Met1, which was derived from primary SCC, Met2, which was derived from recurrent SCC at the primary site and Met4, which was derived from distant metastatic SCC within left axillary lymph nodes. The Met series is a particularly good model to compare primary versus metastatic SCC. Indeed, multiplex fluorescence *in situ* hybridization (M-FISH) studies of the Met series showed that Met4 cells contained marker chromosomes of Met1 cells, suggesting direct evolution of Met4 cells from Met1 cells [77]. Thus, Met1 and Met4 cell lines were used here to investigate potential signalling crosstalk in the adhesome that may control metastasis. In order to obtain a reliable adhesome proteomics dataset, an adhesome isolation protocol was optimised for these cells.

For proteomic analysis of adhesomes, isolated adhesion complex proteins undergo proteolysis into peptides, which are then analysed by liquid-chromatography tandem mass-spectrometry (LC-MS/MS)[75]. These peptides

are separated by liquid-chromatography (LC) and subsequently ionised and analysed by the first mass spectrometer (MS1)[74,78]. The MS1 spectra obtained gives information about the mass/charge ratio (m/z) and intensity of the peptides eluting from the LC column[78]. The MS1 then isolates and fragments each peptide species inside a tandem mass spectrometer (MS2) to obtain an MS2 spectra[78]. The MS2 spectra and the masses of the peptides from MS1 are used as inputs to identify the peptides *in silico* using a peptide sequence database search engine[74]. The peptide fragments are assembled into proteins which are statistically validated to estimate the false positive identifications[74]. Quantification of the proteins through spectral counting or signal intensity of the detected peptide is termed label-free protein quantification (LFQ). LFQ provides robust and precise relative protein expression while being the least cumbersome way to obtain protein quantification[74].

Proteomic Data Pre-processing

Peptide identification and quantification by label-free quantification mass-spectrometry has facilitated high-throughput differential protein abundance analysis. Putative changes in relative protein abundance in the adhesome are not only caused by biological change from primary tumour to metastatic cancer, but are also caused by systematic error and variability introduced during experimental procedure[79,80].

Data normalisation aims to reduce extraneous variability and eliminates systematic error, both of which could introduce bias in proteomics analysis[79,81]. Extraneous variability is introduced for instance when the experimental parameters differ during biological or technical replicates. Systematic error results for instance from unequal loading of peptide masses. It is thus essential to assess multiple normalisation techniques that best reduces intragroup variation and eliminates systematic error.

Another major challenge for differential protein abundance analysis is imputing missing values in label-free quantification proteomics. Ignoring missing values results in loss of statistical power and introduces bias[82] (especially when missing values are not random), and hence is detrimental to differential protein abundance analysis. There are multiple reasons for missing values, ranging from biophysical characteristics of the peptides (e.g.: ion suppression of the peptides, ionization competition, etc..) to bioinformatics mechanisms (e.g. misidentification of peptides)[83] to peptide censoring, whereby the peptide's abundance is below the limit of detection and/or limit of quantification of the mass-spectrometry instrument[80]. Missing values are classified into three categories:

1. Missing Completely At Random (MCAR); these are missing values which are independent of peptide intensity. These are caused, for

instance, by stochastic fluctuations leading to misidentification of peptides[83] or due to random sampling of peptides for identification in mass spectrometry[84].

2. Missing At Random (MAR); these missing values are dependent on the observed values[82].
3. Missing Not At Random (MNAR); these are missing values which are dependent of peptide intensity. These missing values are referred to as censored values.

MCAR and MAR are also known as 'ignorable' missing values[82]. However, for differential abundance analysis, ignoring MCAR and MAR has a worse performance than when missing values are imputed[85]. Different tools have been developed to impute missing values[83,86]. However, not all imputation tools have similar estimation of the missing values as they are all designed using different mathematical concepts. Thus, it is important to evaluate the accuracy of the different imputation methods in estimating missing values. Webb-Robertson *et al*[86] evaluated a range of imputation methods and found that the least square adaptive (LSA) method of the LSimpute[87] package worked best for proteomics dataset in general. However, the evaluation did not take into consideration the types of missing values. Thus, *a priori* assumption that LSA is the best imputation method for proteomics dataset may be detrimental to downstream analysis. A multiple imputation (MI) method, which considers the different types of missing values in proteomics data, are known to perform well. However, in proteomic experiments, it is

very difficult to know the cause of missingness for proteins with missing values and apply the correct imputation method accordingly.

Network Biology

A Functional Module and Community Detection

Proteins can interact with one another to form signalling modules that drive cellular functions. The adhesome is one such signalling module of multiple protein-protein interactions that mediates mechanosensing and signal transduction. Efforts to map the entire protein-protein interactions generated the interactome, with most interactions yet to be mapped[74].

It is becoming evident that disease-associated proteins (such as those differentially regulated between disease phenotypes) interact with each other to form modules in the interactome map, the so-called 'disease module'[88-90]. Proteins within the disease modules (disease-specific functional modules) are postulated to act collaboratively to drive disease phenotypes[88]. In a similar way, we wish to identify adhesome proteins that drive cancer progression and interrogate their interconnectivity. This would lead to the identification of adhesome proteins driving cancer progression which may be targets for drug development. Thus, using differential protein abundance analysis alone is not sufficient as it ignores the interaction between individual proteins. Computational development in mathematical optimisation and network theory (also known as graph theory) has allowed new algorithms

which integrate differential protein abundance analysis within a protein-protein interaction network context [91]. Such algorithms produce a network of differentially abundant proteins (active proteins) and their interactions with one another, a so-called 'active module' or 'functional module'[91]. We applied an exact approach on the adhesome network to extract an active module that may putatively drive cancer progression.

The protein-protein interaction network is noisy and incomplete. Thus, the functional module may lack an interconnected 'clique'[92] or at least a multipartite[93] network topology expected from signalling pathways or from protein complexes. Assuming that the protein-protein interactions network was complete, one would expect protein complexes to form a highly connected 'clique', *i.e* all nodes are connected to the other nodes (mathematically, if *n* nodes form a clique, each node will have *n-1* edges) or multipartite graph. On the other extreme, in a non-complete fragmented network (network where most nodes are not connected), no discernible network module will be observed[89]. In its current state, the interactome allows large module to be observed but is too fragmented for small module to be extracted[89]. Thus, different community structure algorithms were used in order to obtain a mesoscale view and hence the functionality of the proteins in the active module[94,95]. The use of community structure algorithms is further supported by the fact that, as mentioned above, disease-associated

proteins are modular in nature (in other words, they tend to form community structures) and hence may function together to drive a disease phenotype[88].

Hub

As well as understanding how proteins act collaboratively to drive a disease phenotype, it is equally important to identify the most essential proteins in a signalling complex. In the context of a protein-protein interaction network, essential proteins are involved in a large number of interactions, the so called 'hubs'[96]. The functional importance of hubs were highlighted in *Saccharomyces cerevisiae (S. cerevisiae)* whereby deletion of highly connected hub proteins were found to be lethal[97].

Han *et al.*[98] reported two classes of hubs (hubs being defined as nodes with degree greater than five) based the level of coexpression between the hub and its interacting partners. The level of coexpression follow a binomial distribution, giving rise to the dichotomy of 'party' and 'date' hubs[98]. Party hubs are highly coexpressed with their neighbours and preferentially link proteins within the same community (intra-module hub)[99]. Date hubs, which have relatively low coexpression with their neighbours, are considered as inter-module hubs that connect communities to each other[99-101]. However, the dichotomy of 'party' and 'date' hubs is controversial[99,102] with Chang *et al.*[99] and Taylor *et al.*[100] showing evidence for a more complex dynamic architecture. Taylor *et al.*[100] reported a multimodal distribution of hubs

coexpression in human protein-protein interaction network as opposed to a bimodal distribution as observed by Han *et al.*[98] for the *S. cerevisiae* interactome. However, human protein hubs which are conserved in *S. cerevisiae* produce a binomial distribution of coexpression[100], suggesting that the human interactome has a higher dynamic complexity than that of yeast. Despite the multimodal distribution in the human interactome, hubs were still classified into two distinct classes of 'party' and 'date' hubs.

Guimera and Amaral[94,103] classified hubs in terms of their structural roles within their respective communities. Communities are clusters of interacting nodes within a network. In protein-protein interaction network, proteins within a community are postulated to participate in the same molecular processes[104]. Nodes with within-community degree (z) ≥ 2.5 are considered as hubs. Those hubs can be further categorised into provincial hubs, connector hubs and kinless hubs according to their participation coefficient (P). Provincial hubs have $P \leq 0.30$. These hubs have a majority of their links within their communities.

Connector hubs have $0.30 < P \leq 0.75$. The hubs have a majority of their links with other communities. Kinless hubs have $P > 0.75$. These hubs have links homogenously distributed among all communities within the network.

The classification of hubs such as date and party hubs[98], connector, kinless, and provincial hubs[94,103], inter-module and intra-module hubs[100,101] are not mutually exclusive. Indeed, connector and kinless hubs have been reported to behave as date hubs, and provincial hubs as party hubs, though the provincial and party hub similarity was quite modest[99].

■ Aims

A global assessment of the functional molecular units in adhesion complexes that mediate cancer progression is currently lacking. With the development of new methods to isolate adhesion complexes for mass-spectrometry and network analysis, we aim to isolate the adhesion complexes and identify by mass-spectrometry the adhesion proteins in patient-derived primary and metastatic SCC. By performing network analysis on the isolated SCC adhesion proteins, we aim to better understand the molecular processes mediated by groups of differentially abundant adhesion proteins in primary and metastatic SCC cells. We also aim to identify key adhesion hub(s) involved in the progression of SCC. Together, these results will allow us to form hypotheses onto the role(s) of adhesion proteins during cancer progression and identify adhesion proteins of interest. We aim to understand the functional relevance of these adhesion proteins of interest by knocking-out their expression in patient-derived primary and metastatic SCC cells using CRISPR/Cas9 system. This is complemented by immunofluorescence and subcellular fractionation analysis to reveal their subcellular localization.

Our aim is to obtain a mechanistic insight into novel roles of adhesion

proteins during SCC cancer progression.

CHAPTER 2

Methods

■ Adhesome Isolation Optimisation

Cell Culture

Human cutaneous SCC cell lines Met1, Met2 and Met4 were a gift from Professor Charlotte Proby (University of Dundee). They were cultured as previously described: Dulbecco's modified Eagle's medium (DMEM; Sigma-Aldrich, D 6546) and Ham's F12 nutrient mixture (Life Technologies, 21765-037) (3:1 ratio) supplemented with 10% fetal bovine serum (FBS; Life Technologies, Ref. 10270-106), 3mM L-glutamine (Sigma-Aldrich, G7513), and RM+ media (5 μg/ml transferrin, 0.4 μg/ml hydrocortisone, 1x10-10M cholera toxin, 10ng/ml epidermal growth factor (EGF), 5 μg/ml insulin, 2x10-11M liothyronine).

Adhesome Isolation

2D global focal adhesion isolation was carried out as previously described[75]. Briefly, Met series cell lines were seeded on plates coated with 10 μg/ml human fibronectin (BD Biosciences, 354008) and 20 μg/ml negative control ligand (e.g. Apo-transferrin (gift from Dr. Adam Byron) or Poly-D-Lysine (Sigma Life Sciences, P6407)) and incubated for 4 hours at 37°C in humidified 5% (v/v) CO_2. 3mM of reversible and cell permeable crosslinker DTBP (dimethyl-3-3'-dithiobispropioimidate; Thermo Scientific, 20665) in DMEM/F12 is added and incubated at 37°C for 5 minutes followed by

quenching with 150 ul 1M Tris-HCl for 3 minutes at room temperature. Cells are lysed after which the cells can either be frozen at -70°C overnight prior to washing procedure or washed just after cell lysis. In the wash step, hydrodynamic pressure is applied with a Waterpik device (Waterpik Ultra Water Flosser WP120). The adhesome proteins are subsequently isolated using 300 ul Adhesion Recovery buffer (125mM Tris.HCl pH6.8, 1% (w/v) SDS, 150mM dithiothreitol) and constituents of the isolated protein complexes are analysed by western blot and by Coomassie staining (Expedeon InstantBlue ISB1L) as per manufacturer's protocol.

The isolated adhesome complex fractions were checked for co-purifying nuclear (lamin A/C, Histone H4 (His H4), minichromosome maintenance complex component 2 (MCM2)), mitochondrial (complex IV), endoplasmic reticulum (calnexin) and cytoplasmic (Hsp90) contaminants. Cell lysate is included as a positive control only.

Lysis buffers evaluation.

Lysis buffer evaluation was performed on using Met1, Met2 and Met4. Three types of lysis buffer were evaluated; a hypotonic solution of 2.5mM triethanolamine (TEA) in PBS; a detergent containing lysis buffer, RIPA (50mM Tris.HCl pH 7.5, 150mM NaCl, 5mM EDTA, 0.5% (w/v) SDS, 1% (w/v) sodium deoxycholate, 1% (w/v) Triton X100); and a high pH lysis buffer, FA extraction buffer (0.05% (w/v) NH4OH, 0.5% (w/v) Triton X100, in PBS).

Negative control evaluation.

Negative control evaluation was performed as outlined in the protocol above except that both poly-D-lysine and apo-transferrin were used as negative control.

Crosslinking incubation time evaluation.

Crosslinking incubation time evaluation was performed as outlined in the protocol above except that crosslinking time was increased to 30 minutes while maintaining crosslinker concentration at 3mM.

Crosslinking concentration evaluation.

Crosslinking concentration evaluation was performed as outlined in the protocol above except that DTBP concentration was increased to 10mM while maintaining crosslinking time for 5 minutes.

Choice of crosslinker evaluation.

This evaluation was performed as outlined as above except that dithiobis[succinimidyl propionate] (DSP) and 1,4-di-[3′-(2′-pyridyldithio)-propionamido] butane (DPDPB) were used as crosslinker instead of DTBP. DSP and DPDPB were dissolved in 1% DMSO in PBS. Concentrations of 3mM DSP with 0.2mM DPDPB and 3mM DSP with 0.05mM DPDPB were used.

■ Data Pre-processing

Dataset

Adhesome isolation from human cutaneous squamous cell carcinoma was analysed by tandem liquid chromatography mass spectrometry (LC-MS/MS). The raw files were analysed by MaxQuant. Peptides were identified against the UniProt human database (UP000005640_9606) using the Andromeda search engine. Oxidation of methionine and acetylation of the N-terminal was allowed as variable modification and carbamidomethylation of cysteine as a fixed medication. The dataset consists of protein intensities for 3 replicates per condition (primary (Met1) and metastatic (Met4)). The 3 replicates were numbered 1-3 (e.g. Met1 Replicate 1, Met1 Replicate 2, …, Met4 Replicate 1, … and so on).

Filtering

Proteins with reverse sequences and potential contaminants were removed. Proteins which have been detected in only one replicate or only present once in either primary or metastatic adhesome proteomics dataset were removed.

Normalisation

Normalyzer[105] was used to evaluate different normalisation methods. Before normalisation, all LFQ intensities were log2 transformed. From this point onwards, log2 transformed LFQ intensities are referred to as simply LFQ intensities unless described otherwise. Normalisation by total intensity (TI), median intensity (MedI), average intensity (AI), quantile, Variance stabilising normalisation (VSN), robust linear regression (RLR) and cyclic LOESS normalisation were assessed. These normalisation methods were applied globally (denoted by G), *i.e* the whole dataset was normalised at once. VSN, RLR and LOESS were also applied locally (denoted by R), whereby the condition group are normalised separately.

Total Intensity, Median Intensity and Average Intensity Normalisation

Total intensity, median intensity and average intensity normalisation centre the protein abundance around the sum, median and average of the population of protein intensities from each replicate. The intensities of each protein are multiplied by a scaling factor such that total protein intensity, median protein intensity and average protein intensity for each replicate are equal. These normalisation techniques adjust for systematic error which occur due to unequal loading of peptide mass into the LC-MS/MS system[79]. However, these normalisation techniques assume that *most* protein abundances do not change[80].

Quantile Normalisation

Quantile normalisation transforms the protein intensities distribution in each replicate into similar distribution[79,81]. Briefly, the quantile normalisation approach is as follows[106]:

1) In each replicate, sort by protein abundance value (from highest LFQ intensity first to lowest LFQ intensity in each replicate, irrespective of the protein).
2) Find the mean value of the highest protein intensities.
3) The highest protein intensities in each replicate is substituted for that mean value.
4) Find the next highest protein intensities and substitute their value by their mean. This is repeated until all protein intensity values have been substituted.

This approach allows the condition groups to have the same quantile[81].

Variance Stabilising Normalisation

The variance in intensity measurements are associated with measurement error. It is important to model intensity variance in order to properly assign statistical significance to changes in protein abundance at different conditions[107]. It has been suggested that in LC-MS, the variance follows a quadratic function of the mean intensity[108]. Thus, variance stabilising normalisation (VSN), which was designed for microarray analysis whereby a

quadratic variance model was assumed[107], was applied to raw protein intensity from LC-MS/MS.

VSN transforms the raw protein intensities with a set of parametric estimations such that the transformed protein intensities are approximately independent of the variance.

Robust Linear Regression

Robust linear regression (RLR) assumes that the measurement variation (systematic error) is linearly dependent on the magnitude of the protein abundance[79,109]. The RLR approach corrects for the systematic error by centring the protein intensity in each replicate around the median protein intensity across condition groups[109].

RLR was applied locally and globally, denoted by -R and -G respectively. Global RLR fits the protein intensity value of the samples around the median protein intensity across the whole dataset. Local RLR fits the protein intensities of the replicates within a condition around the median protein intensities obtained within that condition.

Cyclic LOESS Normalisation

Cyclic LOESS normalisation (CL) uses the MA plot, M being the difference in protein intensity between two replicates and A is the average protein intensity, for normalisation[81,110]. A regression curve is fitted in the MA plot using loess regression (a local regression technique which fits a curve instead of a line). Ideally normalised data points between two sample runs would have the loess smoother line along the A axis[110]. CL normalises the two sample runs at a time by applying a correction factor to the loess curve such that the latter lies along the A-axis. For more than two sample runs, CL iteratively normalises all distinct pairwise combinations[81,109].

CL was applied locally and globally. Global CL iteratively normalises all distinct pairwise combinations across the whole dataset. Local CL iteratively normalises all distinct pairwise combinations across a condition group.

Imputation

LS_Adaptive

LS_Adaptive (LSA) is a function of the LSImpute package[87]. LSA is a local similarity-based imputation approach[86] whereby it is assumed that proteins which are co-regulated will have highly correlated expression profiles. Briefly, given a target protein with missing value, LSA finds the 10 most correlated proteins to the target protein. The missing value is first estimated by a weighted average of 10 single regressions between the target protein and the correlated proteins. Once an initial estimate is obtained for the

missing value, a multiple regression is then performed on the different sample replicates and a new re-estimate of the missing value is obtained. The missing value is finally imputed by a weighted average of the re-estimated missing value and the initial estimate.

Multiple Imputation

The multiple imputation (MI) method here involves a first pass single value imputation (SVI) followed by a second pass of imputation by Predictive Mean Matching (PMM).

SVI uses the local minimum log-intensity value across a replicate run to impute missing value[111]. SVI was performed on peptides which had missing values on *all* replicates per condition group. In such case, it is assumed those peptides are below the limit of detection of the mass spectrometry and thus the missing values are MNAR. SVI was reported to perform well under MNAR[83].

The remaining missing values are assumed not to be censored and thus imputed using PMM algorithm which produces multiple datasets with imputed values. Briefly, PMM uses a Bayesian method whereby a prior distribution of the regression weight between the variables is specified[112,113]. The posterior predictive distribution of the missing values is then derived from the prior distribution. Finally, imputations are drawn from this posterior density

distribution[112,113]. To model multivariate missing values, missing values are imputed iteratively in a process termed as regression switching[112]. The iteration is stopped when there is convergence to a multivariate posterior predictive distribution[112-114]. This process is carried out *m* times to produce *m* complete datasets. Each complete dataset has different estimates for the missing values. The multiple complete datasets are pooled together to obtain the average of a quantity of interest[114] (e.g. regression coefficient). However, since the primary goal here is to obtain the imputed values, only one complete dataset (out of *m* complete datasets) is selected.

■ Network Analysis

Active Module

BioNet

Identification of a functional module was performed by means an exact approach using BioNet package[115,116]. Briefly, differential abundance was performed on each protein between the Met1 (primary) and Met4 (metastatic) condition groups using Student's t-test. The *p*-values were converted into a score based on a signal-noise decomposition. Each node is assigned a score corrected for FDR and thus becomes a weighted network. Here, an FDR of 0.05 is used. The problem of finding a maximum-weight connected subgraph (MWCS) is an NP-hard problem. To circumvent this, the MWCS is transformed into a prize-collecting Steiner tree (PCST) problem. The exact solution to the PCST problem is computed using integer linear programming. BioNet was implemented as part of an R package.

Community Detection

Weighted Active Module

The edges of the functional module were weighted using the Spearman rank correlation. The community detection algorithms used here only use positive weights. Thus, the range of Spearman rank correlation coefficient was shifted by one integer so as to obtain positive values only. The positive weights thus ranged from 0 to 2.

Louvain Method

Louvain method[117] makes use of the Girvan-Newman modularity[118] as a quality function to optimise. Girvan-Newman modularity refers to the fraction of edges falling within a community minus the expected fraction if the edges were randomly distributed[118]. The Louvain method is an agglomerative approach consisting of two phases. The first phase of the algorithm assigns a community to each node. Therefore, initially, there are as many communities as nodes. The algorithm then computes the gain in modularity by placing a given node into the community of the neighbouring node. Each node is then placed in the community for which the gain in modularity is maximum. In the second phase, each node in the derived community is agglomerated into a single node. Therefore, each node in the new supernetwork represents a community derived from the first phase. The initial phase of the algorithm is then reapplied to this supernetwork, *i.e*. each new node is clustered which results in maximal increase in modularity. This algorithm is stopped when no increase in modularity is observed. Hence, community structures with maximal modularity are obtained. The Louvain method was implemented as part of the igraph R package[119].

Spectral Modularity

The Girvan-Newman modularity can also be optimised by using the spectral properties of the modularity matrix[120]. The modularity matrix is expressed in terms of its eigenvalues and eigenvectors and the network is recursively

partitioned. Spectral modularity was implemented as described by Mclean *et al*[121].

Infomap

Infomap finds community structures within a network by tracking the information flow through the network. A random walker is used as a proxy for the flow. The flow is encoded in such a way that if the walker steps out of a community, the codeword gets longer. Similarly, the longer the random walker stays within a group of nodes, the more likely those nodes are within the same community. Hence, a 'map equation', which accounts for the frequency the random walker exits the community and the fraction of time the random walker stays within a community, was derived as an objective function to minimise. A two-level algorithm similar to the Louvain method was used to minimise the map equation. Initially, each node is assigned a community and hence, there are as many nodes as communities. Each node is then placed in its neighbouring community and the map equation computed. The nodes are placed in the communities which result in the largest decrease of the map equation. A supernetwork is constructed whereby the derived communities form the new nodes. To improve accuracy and avoid local minima, two additional algorithms, namely submodule movements and single-node movements, are added at the end of the core algorithm. A submodule movements algorithm deconstructs each community into submodules. Each submodule is then free to move between the communities derived from the core algorithm and placed in a new neighbouring community provided this results in reduction in the map

equation. Single node movement works in a similar way to submodule movement except the community is deconstructed into single nodes which are then re-assigned to the neighbouring community only if a reduction in the map equation is observed. The network was deconstructed and reconstructed by submodule and single node movements 1000 times.

Constant Potts Model with Surprise Maximization

The Potts model is a statistical mechanics concept which describes the spin glass on a magnet. Reichardt and Bornholdt[122] reformulated the problem of community detection as a problem of finding the ground state of Potts spin glass model. Traag *et al*.[123,124] furthered this work and derived a Potts model with a tuneable resolution parameter γ. For an unweighted network, setting the $\gamma = 0$ results in all nodes being in a single community. Conversely, setting $\gamma = 1$ results in each node being its own community. Thus, it is essential to find the resolution parameter γ for which 'meaningful' communities are obtained. To this end, γ was scanned for which the quality function termed 'Surprise' was maximum. Surprise (S) measures of how unlikely a given community structure is found in a random network. Therefore, maximum S values lead to accurate community structure[125]. Constant Potts model with Surprise maximisation was implemented as part of the louvain package (https://pypi.org/project/louvain/).

Node Centrality Analysis

z-P Parameter Space

Guimera and Amaral[94,103] used the concept that nodes with similar role should have similar topological properties with respect to the community structure. The authors used two topological properties, namely, the within-community degree z-score (*z*) and the participation coefficient (P) to define the role of a node with respect to the community structure. Suppose a node i in community S, $k_i$is the number of links node i has to other nodes within the same community S, $\overline{k_S}$ and $\sigma_{k_S}$is the average and standard deviation respectively of the number of links over all nodes within the community S, then the within-community degree z-score of node i is given as $z_i = \frac{k_i - \overline{k_s}}{\sigma_{k_S}}$

The within-community degree z-score of node i, z_i, measures how interconnected node I is to other nodes within the same community.

The participation coefficient P_i of node i is defined as

$$P_i = 1 - \sum_{M=1}^{M} \left(\frac{l_{i,M}}{L_i}\right)^2$$

where $M$is the total number of communities in the network, L_i is the total number of links of node I (degree of node I), and lim is the number of links of node I to module M. Pi converges to one if the node I is homogeneously linked to other nodes within other communities in the network and to zero if node I is only connected to other nodes within its own community.

The nodes are classified as follows:

1. For nodes with $z < 2.5$ (non-hubs):
    a. Ultraperipheral nodes (R1) have $P \leq 0.005$. These nodes have all their links within their communities.
    b. Peripheral nodes (R2) have $0.05 < P \leq 0.625$. These nodes have at least 60% of their links within their communities.
    c. Non-hub connector nodes (R3) have $0.625 < P \leq 0.80$. These nodes have a high proportion of their links with other communities.
    d. Non-hub kinless nodes (R4) have $P > 0.80$. These nodes have links homogeneously distributed among all communities within the network.

2. For nodes with $z \geq 2.5$ (hubs):
    a. Provincial hubs (R5) have $P \leq 0.30$. These hubs have a majority of their links within their communities.
    b. Connector hubs (R6) have $0.30 < P \leq 0.75$. The hubs have a majority of their links with other communities.
    c. Kinless hubs (R7) have $P > 0.75$. These hubs have links homogenously distributed among all communities in within the network.

The z and P calculations were implemented using the BrainGraph R package.

PageRank

The PageRank algorithm ranks the nodes according to their importance in the network and assigns a numerical weighting to each node[126]. The importance of a node is determined by the quantity and quality of neighbouring nodes linked to it and the node is assigned a PageRank (PR) value. PageRank algorithm was used on the active module with Spearman Rank correlation coefficients as edge weights (negative values with range [-1,1]). A jumping factor (*s*) of 0.85 was used and sum of the PR values was set to 1. PageRank was implemented using the igraph package[119].

■ Mena Validation

CRISPR/Cas9

FAK and Mena gRNA constructs

CRISPR-Cas was carried out as previously described[127,128]. Guide oligonucleotides for FAK, targeting exon 2 of FAK (FAK h2: forward sequence:

TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGAGGTTCA CTGGCTTCACG,

reverse sequence:

GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCGTGAAGCCA GTGAACCTCC) and for Mena, targeting exon 2 of Mena (Mena h3: forward sequence: TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG GAGATGCTAGACAGGTGTA,

reverse sequence:

GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTACACCTGTCT AGCATCTCC), contain an AflII restriction site and protospacer adjacent motif (PAM). The guide oligonucleotides were annealed by PCR. gRNA cloning vector (Addgene plasmid no. 41824) was linearised by AflII digestion. The guide oligonucleotides were annealed to the linearised gRNA cloning vector by Gibson Assembly reaction. The constructs were transformed into chemically competent DH5α and selected on agar plates containing kanamycin. Positive gRNA guides were identified by sequencing.

FAK and Mena CRISPR/Cas9 nucleofection

Met1 and Met4 cells were grown to 60% confluency. Cells were co-nucleofected with 3μg gRNA (either FAK gRNA or Mena gRNA) and 3μg hCas9 (Addgene, plasmid ID 41824) using Amaxa Human Keratinocyte Nucleofector Kit (Lonza, #VPD-1002) according to manufacturer's protocol. Briefly, cells were trypsinised into cell suspensions which were then centrifuged at 210 x g for 5 min. The supernatant was aspirated and the cells were resuspended into 100 μl of Nucleofector solution containing 3μg gRNA and 3μg hCas9 and transferred to the nucleofection cuvette. Nucleofector was set to T-024 program. After nucleofection, the cells were resuspended into RM+ media. Three rounds of nucleofection with gRNA and hCas9 were carried out onto the cells. Cells were allowed to recover and grow to 60% confluency in between nucleofection rounds.

Treatment

Mena11a re-expression

pcDNA-Mena11a was a gift from Dr. Francesca Di Modugno. Cells were transiently transfected with pcDNA-Mena11a using the X-tremeGene HP DNA transfection reagent (Roche) according to manufacturer's protocol. Cells were transfected at a ratio transfection reagent:DNA of 2:1. Cells were incubated for 60 hours prior to phosphotyrosine immunoprecipitation experiment.

Leptomycin-B

Met4 cells were treated with 10nM and 20nM Leptomycin-B (Cell Signalling Technology, # 9676S) and ethanol as control for 4 hours. Cells were washed with ice-cold PBS three times prior to fractionation. Cells were harvested in Cyto buffer (10mM Tris-HCl, pH 7.5, 0.05% NP-40, 3mM $MgCl_2$, 100mM NaCl, 1mM EGTA) supplemented with protease inhibitors (1.25mM phenylmethylsulfonyl fluoride, 0.5mM NaF, 0.1mM sodium orthovanadate, 0.1%(v/v) Aprotinin). The lysate is centrifuged at 1000 x g for 10 min at 4°C. The supernatant (cytosolic fraction) is collected and kept for analysis. Pellet is homogenised in modified RIPA (50mM Tris-HCl, 150mM NaCl, 5mM EDTA, 0.5% (w/v) SDS, 1% (w/v) sodium deoxycholate, 1% Triton X100) supplemented with protease inhibitors and centrifuged overnight at 16,000 x g, 4°C.

Whole cell lysate (WCL) was obtained by harvesting the cells in modified RIPA buffer supplemented with protease inhibitors and centrifuged overnight at 16,000 x g, 4°C. The supernatant (WCL) is collected and the pellet discarded.

Sucrose Cushion Fractionation

Cells were seeded on fibronectin coated plates at a density of $3x10^6$ cells / 15cm dish and incubated overnight at 37°C prior to fractionation. For cell attachment on different extracellular matrices, cells were seeded on Matrigel,

fibronectin, poly-D-lysine coated plates at a density of $3x10^6$ cells / 15cm dish and incubated for 4 hours at 37°C prior to fractionation.

Nuclei Pure Prep Nuclei Isolation Kit (Sigma Aldrich, #NUC-201) was used to obtain nuclear and cytoplasmic fraction as per manufacturer's protocol. It is recommended to use at least four 15cm dish for fractionation. Cells were washed in 10ml ice-cold PBS per 15cm dish. 500µl of Lysis solution (Nuclei PURE lysis buffer, 0.1% Triton X-100, 1mM DTT) supplemented with protease and phosphatase inhibitors (1.25mM phenylmethylsulfonyl fluoride, 0.5mM NaF, 0.1mM sodium orthovanadate, 0.1%(v/v) Aprotinin, PhoSTOP tablet ()) was used per 15cm dish. An aliquot of 700ul of the cell lysate was set aside for total cell lysate and cytosolic fraction. The rest of the cell lysate was topped up to 10ml with the remaining Lysis solution. To each 10ml sample lysate, 18ml of 1.8M Sucrose Cushion solution containing 1mM DTT was added, making 28ml of lysate solution. The latter was layered on top of a sucrose cushion consisting of 10ml 1.8M Sucrose Cushion solution containing 1mM DTT in an ultracentrifuge tube (Beckman, #344058). The lysate was centrifuged at 30,000 x g for 45 minutes at 4°C. The supernatant is aspirated without disturbing the pellet. The pellet is vortexed in 500ul ice-cold PBS and transferred to a 1.5ml microcentrifuge tubes and centrifuged at 500 x g for 5 minutes. The supernatant is aspirated. The pellet is homogenised in modified RIPA buffer supplemented with protease inhibitors and centrifuged overnight at 16,000 x g, 4°C. The supernatant is kept for western blot analysis. The pellet is discarded.

To obtain the cytosolic fraction, the aliquoted cell lysate mentioned above was centrifuged at 16,000 x g for 10min. The supernatant was saved for western blot analysis and the pellet discarded.

To obtain the total cell lysate, SDS and sodium deoxycholate was added to the aliquoted cell lysate mentioned above to a final concentration of 0.5% (w/v) SDS and 1%(w/v) deoxycholate. The mixture is thoroughly mixed and centrifuged overnight at 16,000 x g, 4°C. The supernatant was saved for western blot analysis and the pellet discarded.

Immunoprecipitation

For Nesprin-2 immunoprecipitation, cells were harvested in NETN lysis buffer (100mM NaCl, 20mM Tris-HCl, pH 7.5-8.0, 0.5mM EDTA, 0.5% (v/v) NP-40) supplemented with protease inhibitors. For phosphotyrosine (PY20) immunoprecipitation, cells were harvested in RIPA buffer (50mM Tris-HCl, pH 7.5, 150mM NaCl, 1% Triton X100, 0.5% sodium deoxycholate) supplemented with protease and phosphatase inhibitors.

Cells were homogenised by 15 strokes through a 15G syringe needle. Homogenate was centrifuged at 16,000 x g, 4°C for 10 min. Supernatant (2mg protein, exceeding 400μl) was incubated with 2ug antibody (Nesprin-2 antibody, Abcam #ab217057; PY20 antibody; BD Transduction # 610000) and 7μl of Dynabeads Protein G (Invitrogen, #10003D) overnight with rotation at 4°C. Beads are washed three times in 500 μl of the respective lysis buffer. Immunoprecipitants are eluted by boiling in 20μl Laemmli sample buffer at 95°C for 10 min.

Western blot

Protein concentration is measured by Pierce BCA Protein Assay kit as per manufacturer's protocol (ThermoFisher, # 23227). Protein was supplemented with 5x Laemmli sample buffer to a final concentration of 0.5 – 1.0 mg/ml and boiled at 95°C for 10 min. Samples were separated by SDS-PAGE gel electrophoresis using Mini-PROTEAN gels (BioRad). Proteins were transferred to nitrocellulose membrane (Trans-Blot, BioRad) and blocked with 5% (w/v) milk in Tris-buffered saline. Membranes were probed with anti-Mena (Atlas, # HPA028448), anti-p53 (Santa Cruz, # SC-126), anti-Histone H3 (Cell Signaling Technology, #4499), anti-calnexin (Cell Signaling Technology, 2679), anti-lamin A/C (Cell Signaling Technology # 2032), anti-FAK (Cell Signaling Technology, # 3285), phospho-FAK (Tyr-397, Cell Signaling Technology, #3283), anti-calnexin (Cell Signaling Technology,

#2679), anti-GAPDH (Cell Signaling Technology, #5174), anti-YAP (Cell Signaling Technology, # 4912). Membranes were incubated in secondary anti-rabbit or anti-mouse streptavidin-conjugated horseradish peroxidase secondary Ab (Cell Signaling Technology) and visualized using the Bio-Rad ChemiDoc MP Imaging System.

Immunofluorescence Microscopy

Immunofluorescence was carried as follows: 16% paraformaldehyde (Alfa Aesar) was added to cells still in their medium to a final concentration of 4%. The cells were incubated for, 15 min, at room temperature and washed twice with PBS. Cells were then incubated in 0.1M glycine for 5 min. Cells permeabilised with 0.2% Triton X-100 in PBS, 5 min, room temperature followed by washing twice with PBS. Cells blocked with 2% BSA in PBS for 1 hour at room temperature. Cells were incubated with 100 µl anti-Mena (Atlas, # HPA028448), anti-Nesprin2 (Abcam), anti-lamin B1 (Abcam, ab90169), phospho-FAK (Tyr-397, Cell Signaling Technology, #3283), anti-H3K27Me2 (Cell Signaling Technology, # 9728), anti-H3K9Me2 (Cell Signaling Technology, # 4658), in 2% BSA in PBS overnight at 5°C

Incubated coverslips with 100 µl $F(ab)_2$ secondary antibody or phalloidin conjugated to Alexa Fluor 488, Alexa Fluor 594 (Thermo Fisher) or Atto647N (Rockland) in 0.05% Tween 20 in PBS for 45 min in the dark. Primary and secondary antibodies were fixed again with 4% paraformaldehyde, washed

with PBS, followed by incubation with 0.1M Glycine for 10 minutes.

Coverslips were mounted onto anti-fade mounting medium fluoroshield containing DAPI (Sigma Aldrich). Cells were imaged by confocal laser scanning microscope (Olympus FV1000), spinning-disk confocal microscopy (Dragonfly Andor) and N-Structured Illumination Microscopy (N-SIM, Nikon).

CHAPTER 3

Results: Adhesome Isolation Optimisation & Proteomics Data Pre-processing

■ Introduction

Metastasis is the cause of 90% of cancer mortality and, so far, drugs against metastatic cancers at secondary sites have failed to improve cancer progression-free survival or overall survival in clinical trials[129]. Understanding the molecular mechanisms that promote metastasis to secondary sites and maintain cellular metastatic growth is crucial for the development of new therapeutics against metastasis. The execution of metastasis requires continuous interactions between the cell-surface receptors integrins and extracellular matrix (ECM) proteins such as fibronectin, collagen IV, laminin and vitronectin[70,71]. These integrins-ECM interactions initiate the recruitment of a series of intracellular proteins to the plasma membrane to form adhesion complexes (the adhesion nexus), which regulate biological processes such as cell proliferation, survival, migration, tumourigenesis and cell invasion, all of which are necessary for metastasis[71]. This adhesion nexus, of which the sum of protein components is termed the adhesome, consists mainly of integrin receptors and scaffolding, signalling and cytoskeletal proteins[61,68,75]. Initially, the adhseome was analysed in a candidate-based manner, mainly by immunoprecipitation and immunofluorescence microscopy of the individual components[59,75]. Recently, a new methodology making use of liquid-chromatography tandem mass-spectrometry (LC-MS/MS) was developed to globally assess and identify components of the adhesome[75]. This and related methodologies have successfully been applied in numerous studies, and they have led to the identification of novel components of the adhesome[64], provided new insights into the mechanosensing regulatory

pathways of focal adhesions[65,66] , and defined the role of integrin activation in regulating microtubule network[67].

A global assessment of the adhesion signalling during cancer progression is currently lacking. Thus, to understand the molecular changes of the adhesome in a cancer progression setting, we isolated adhesion complexes from malignant keratinocytes derived from human cutaneous squamous cell carcinoma (SCC) at the primary site and at secondary metastatic sites (more specifically the lymph nodes) from the same patient. In order to obtain a reliable adhesome proteomics dataset from these patient-derived cells, the adhesome isolation protocol was first optimised and validated. Once optimised, the isolated SCC adhesion complex samples were then analysed by LC-MS/MS.

Progress in peptide identification and label-free quantification (LFQ) by mass spectrometry has facilitated high-throughput differential protein abundance analysis which in turn can be interpreted using biological network analysis[74]. Changes in relative protein abundance in the adhesome of cancer cells are not only caused by biological changes from primary tumour to metastatic cancer, but can also be caused by systematic error and noise (variability) introduced during experimental procedures[79,80]. For example, extraneous variability is introduced when the experimental parameters differ between

adhesome isolations in biological replicates. Systematic error is introduced, for instance, from unequal loading of peptide mass into LC-MS/MS. It is thus essential to assess multiple normalisation techniques to identify one that best reduces intragroup variation and eliminates systematic error prior to differential abundance analysis of adhesome proteomics data.

Imputation of missing protein abundance values is another challenge in LC-MS/MS proteomics. Ignoring missing values is detrimental for differential abundance analysis[85] and they should therefore be imputed in the dataset. It was reported that the accuracy (the number of true positives) of differentially abundant proteins increases when missing values were imputed[85]. However, the mechanisms resulting in missing values should be ideally taken into consideration. There are broadly two types of missingness: 1) missing completely at random (MCAR) which arises due to, for instance, stochastic fluctuations during peptide identification and quantification[83], and (2) missing not at random (MNAR) which arises, for instance, due to the peptide's abundance being below the limit of quantification and/or below the limit of detection of the mass-spectrometer[80]. MNAR values are also referred to as censored values.

Webb-Robertson *et al.*[86] evaluated a range of imputation methods and found that the least square adaptive (LSA) method of the LSimpute[87] package

worked best for proteomics datasets in general. However, the evaluation did not take into consideration the types of missing values in the datasets and thus an *a priori* assumption that LSA will work best in the SCC adhesome proteomics dataset may be detrimental to downstream analysis. Therefore, a multiple imputation (MI) method was used here which considers the different mechanisms of missing values in the adhesome proteomics data. The MI method was evaluated against the LSA imputation method for estimating log fold changes, which is used for differential protein abundance analysis of quantitative MS data.

In this chapter, we describe the optimisation of the adhesome isolation in human SCC cells and the subsequent normalisation and imputation of the adhesome proteomics dataset. The aims of data normalisation were twofold: i) reducing the extraneous variation between biological replicates due to for example, variations in conditions and technical random error when repeating the experiments, and ii) reducing bias between condition groups (*i.e.* primary vs metastatic human SCC cells). The aim of imputation was to accurately capture the fold change of proteins with missing abundance values.

■ Results

Adhesome Isolation Optimisation

We used cutaneous SCC cells referred to as the Met series which reflect the clinical progression of the disease[76]. The Met series is composed of Met1 cells, which were derived from primary SCC, Met2 cells, which were derived from recurrent SCC at the primary site and Met4 cells, which were derived from distant metastatic SCC within left axillary lymph nodes[76]. Multiplex fluorescence in situ hybridization (M-FISH) studies of the Met series showed direct evolution of Met4 cells from Met1 cells[77]. Hence, the Met series is a particularly good model to compare primary versus metastatic SCC, allowing us to investigate the potential signalling nexus that may control metastasis at the adhesome level.

Previously described methods for the isolation of adhesome proteins[65,67,75,130,131] were carried out on non-keratinocyte cells. Hence, the adhesome isolation[75] (outlined in Figure 5) had to be optimised for the Met cell lines with the aims of:

i) obtaining an efficient yield of adhesion complexes from keratinocytes for proteomics study (using LC-MS/MS);

ii) enriching known adhesome proteins (e.g. FAK, β1 integrin); and

iii) optimising the specificity of the adhesome isolation method for keratinocytes, *e.g.* no known adhesome proteins should be observed in the negative control.

Since fibronectin is essential for the metastatic cascade[132,133] and allows for the formation of integrin-mediated adhesion complexes, we seeded the Met cell lines on fibronectin to isolate the adhesome. Briefly, the adhesome isolation protocol (Figure 5), with the initial conditions used in the optimisation process, were as follows:

1. Met cell lines were allowed to spread on fibronectin-coated plates for 4 hours.
2. Adhesion complexes were stabilised in the presence of 3mM dimethyl 3,3'-dithiobispropionimidate (DTBP) crosslinker for 5 minutes. Excess crosslinker was quenched.
3. Cells were lysed and a hydrodynamic pressure applied to remove cell bodies.
4. The adhesome proteins were isolated and analysed by western blot and LC-MS/MS. The isolated adhesion complex fractions were checked for co-purifying contaminants such as nuclear (lamin A/C, Histone H3, minichromosome maintenance complex component 2 (MCM2)), mitochondrial (complex IV), endoplasmic reticulum (calnexin) and cytoplasmic (Hsp90, GAPDH, tubulin) proteins.

Key experimental conditions to be optimised in the adhesion complex isolation protocol, according to the authors, are the following[75]:

1. The choice of the negative control.
2. Crosslinker incubation concentration and duration.
3. The cell disruption technique; pH vs detergent lysis buffer.
4. Cell spreading times on extracellular substrate.

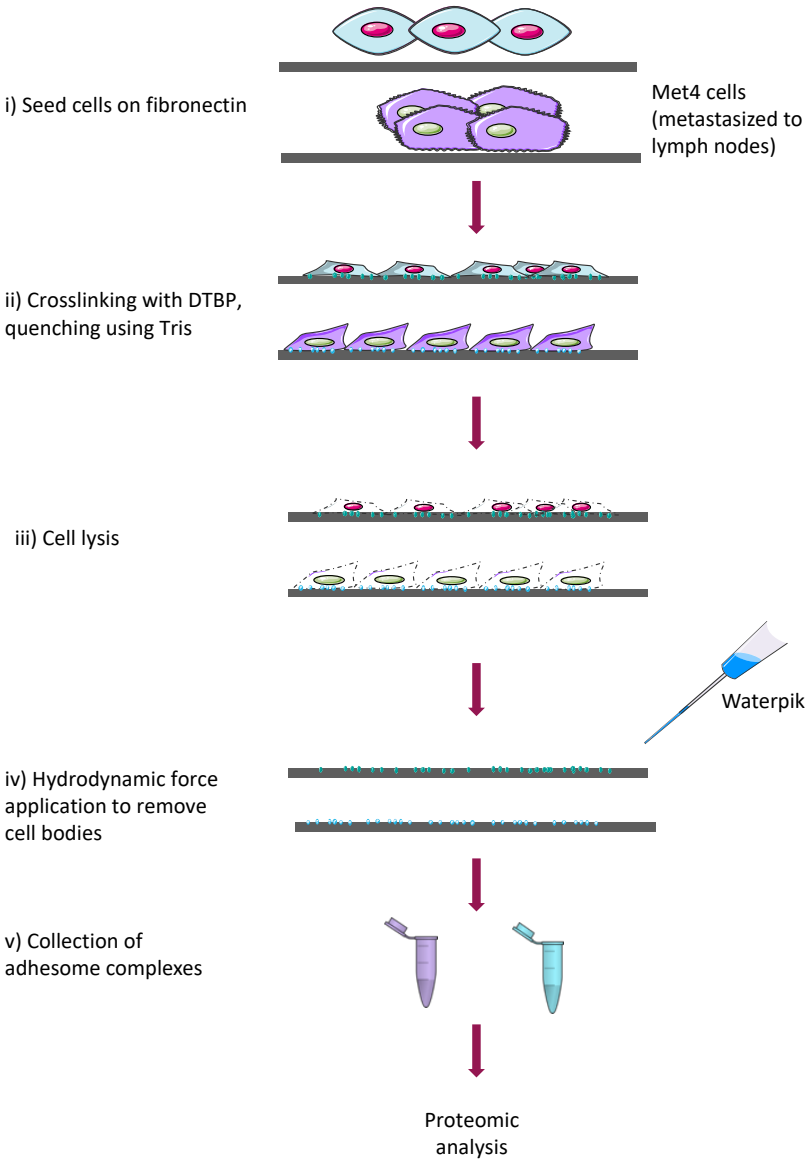


Figure 5 Schematic summary diagram of the adhesion complex isolation procedure. Cells are seeded on plates coated with extracellular ligand (e.g. fibronectin or poly-D-

lysine as a negative control). Cell-permeable crosslinker is added to stabilise the adhesion complexes. Crosslinker was quenched, and cells were lysed and a hydrodynamic pressure applied to remove cell bodies. The adhesome proteins were subsequently isolated and analysed by western blot and LC-MS/MS.

Lysis Buffer Evaluation

SCC cell lysis evaluation was carried out on three types of lysis buffer used in previous adhesion complex isolation studies: a hypotonic solution of 2.5mM triethanolamine (TEA)[65]; a high-detergent lysis buffer, RIPA[131]; and a high pH lysis buffer, FA extraction buffer[65].

Evaluation of the lysis buffers was carried out on Met1, Met2 and Met4 cells seeded on fibronectin (FN) and allowed to spread for 4 hours. Adhesion complexes were stabilised in the presence of 3mM DTBP crosslinker for 5 minutes[64,75].

Isolation of adhesion complexes from cells lysed with TEA buffer recovered the adhesion protein FAK in Met1 and Met4 cells but to a lesser extent in Met2 cells (Figure 6(a)). The non-adhesion-associated protein lamin A/C was also recovered in samples from Met4 cells (Figure 6(a)). Comparison between RIPA and FA extraction buffer showed the latter produced a higher recovery yield of the adhesion protein FAK in Met1 and Met4 cells (Figure 6(b)). The non-adhesion-associated protein lamin A/C was present at low levels in samples from Met4 cells but otherwise not observed in Met1 cells (Figure 6(b)).

This evaluation suggests that FA extraction buffer was most suitable for the chemical disruption of human SCC cell bodies under the experimental conditions tested. We subseqeuntly used FA extraction buffer for the rest of the adhesion complex isoaltion optimisation process.

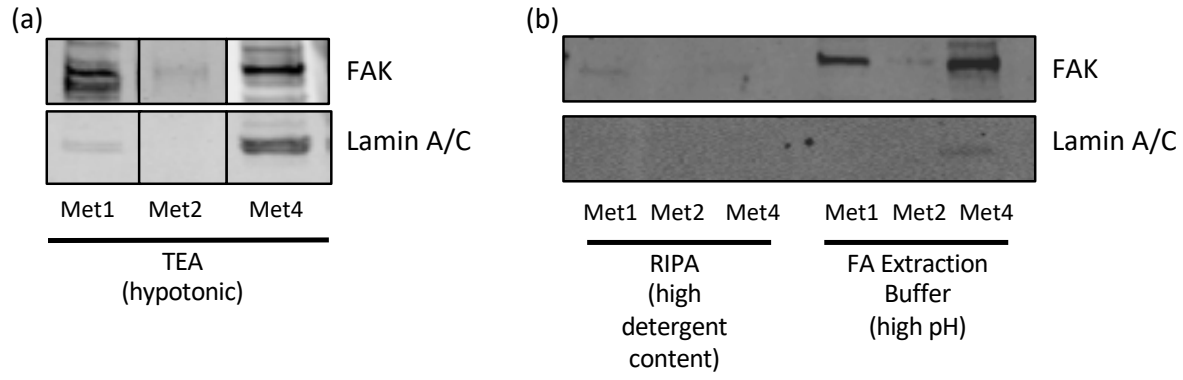


Figure 6 Western blot comparison of isolated adhesion complexes obtained from cell lysis with different buffers. (a) TEA hypotonic buffer; (b) RIPA (left) and FA extraction buffer (right). All Western blot results were from one experiment.

Evaluation of Extracellular Matrix Substrate as Negative Control

The use of a negative control ligand (e.g. apo-transferrin or poly-D-lysine) allows for the identification of specific fibronectin-induced adhesion complexes. Cells seeded on a negative control ligand do not form integrin-mediated cell attachments and thus the isolated proteins were considered as background. We evaluated the use of poly-D-lysine (PDL) and apo-transferrin (Apo-Tf) as negative control ligands as they have previously been used successfully[67,131]. To this aim, the adhesion complex isolation was performed on Met2 and Met4 cells seeded on plates coated with fibronectin

(FN), PDL and Apo-Tf for 4 hours. Adhesion complexes were stabilised in the presence of 3mM DTBP for 5 minutes.

There was no enrichment of the adhesome proteins FAK and β1-integrin when PDL was used as a substrate (Figure 7). Conversely, FAK and β1-integrin were isolated to a similar level when FN and Apo-Tf were used (Figure 7). Thus, PDL is more suitable than Apo-Tf as a negative control ligand for human SCC adhesion complex isolation. Cells seeded on PDL are only attached by electrostatic interactions and thus does not involve the activation of integrin.

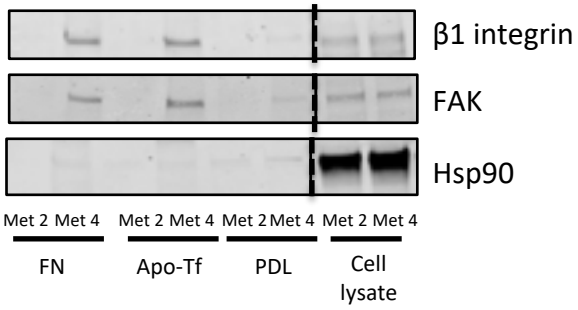


Figure 7 Western blot analysis of adhesion fraction from Met2 and Met4 cells seeded on fibronectin (FN), apo-transferrin (Apo-Tf) and poly-D-lysine (PDL) for adhesome proteins (β1-integrin and FAK) and non-adhesion associated protein (Hsp90). Cell lysate (right) served as a positive control. Western blot result was from one single experiment.

Crosslinking Concentration Evaluation and Crosslinking Time Evaluation

The effects of crosslinking time and crosslinker concentration on the specificity of the adhesion complex isolation were assessed. Met2 and Met4 cells were seeded on fibronectin for 4 hours after which the cells were crosslinked for 30 minutes in the presence of 3mM DTBP or for 5 minutes in the presence of 10mM DTBP.

While increasing crosslinking time to 30 minutes or increasing crosslinker concentration to 10mM isolated adhesion proteins FAK, phosphorylated FAK at Tyr397 (P-FAK), β1-integrin and vinculin for Met2 cells, these conditions also increased the amount of co-purifying nuclear contaminants MCM2 (Figure 8). Moreover, adhesion proteins were also isolated in the PDL control as well as co-purifying nuclear contaminants in Met4 cells (Figure 8). This indicated that, while these conditions isolated adhesome proteins in Met2 cells, neither condition was suitable for specific isolation of adhesome proteins in Met4 cells. Thus, we used the initial approach of stabilising the adhesion complexes in the presence of 3mM DTBP for 5 minutes.

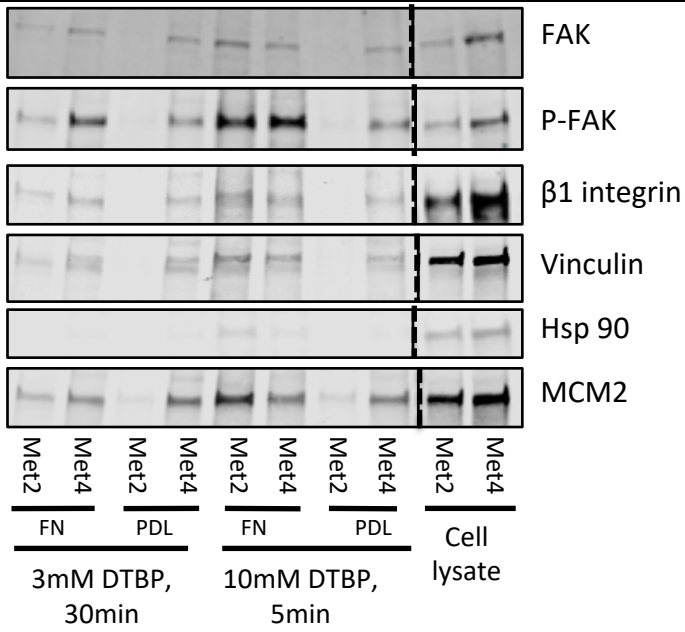


Figure 8 Western blot analysis of adhesome proteins isolated from Met2 and Met4 cells using 3mM DTBP and incubated for 30 minutes (left) or 10mM DTBP and incubated for 5 minutes (centre). Cell lysate (right) served as a positive control. Western blot result was from one experiment.

Crosslinker Evaluation

We next evaluated the use of different crosslinkers for the adhesion complex isolation in Met cells. Crosslinkers dithiobis[succinimidyl propionate] (DSP) and 1,4-di-[3′-(2′-pyridyldithio)-propionamido] butane (DPDPB) were assessed for the isolation of adhesion complexes from human SCC cells plated on both fibronectin and poly-D-lysine. Cell-permeable and reversible crosslinkers DSP and DPDPB, which crosslink primary amines and cysteine, residues respectively, were used in combinations at different concentrations. The combinations of DSP and DPDPB has been used successfully before for isolating adhesion complexes[131]. In this evaluation, concentrations of 0.5mM

DSP with 0.05mM DPDPB and 3mM DSP with 0.2mM DPDPB were assessed.

When a combination of 0.5mM DSP and 0.05mM DPDPB was used to isolate adhesion complexes, there was a low yield of enrichment of the adhesome proteins P-FAK and β1-integrin from Met2 cells and Met4 cells except for β1-integrin in Met4 cells seeded on fibronectin (Figure 9(a)). However, this was non-specific since non-adhesion associated proteins calnexin, tubulin and MCM2 were also isolated (Figure 9(a)). The isolation yield increased for Met4 cells when 3mM DSP and 0.2mM DPDPB were used (Figure 9(b)). However, there was loss of specificity as indicated by the presence of the adhesome proteins P-FAK and β1-integrin in the PDL negative control (Figure 9(b)).

The use of DSP and DPDPB resulted in non-specific isolation of adhesome proteins. Moreover, since DMSO was used to solubilise non-water-soluble cross-linkers DSP and DPDPB, this can introduce organic solvent artefacts[134]. We thus concluded that DSP and DPDPB were not suitable for isolation of adhesome-specific proteins. Hence, we maintained the use of DTBP as crosslinker.

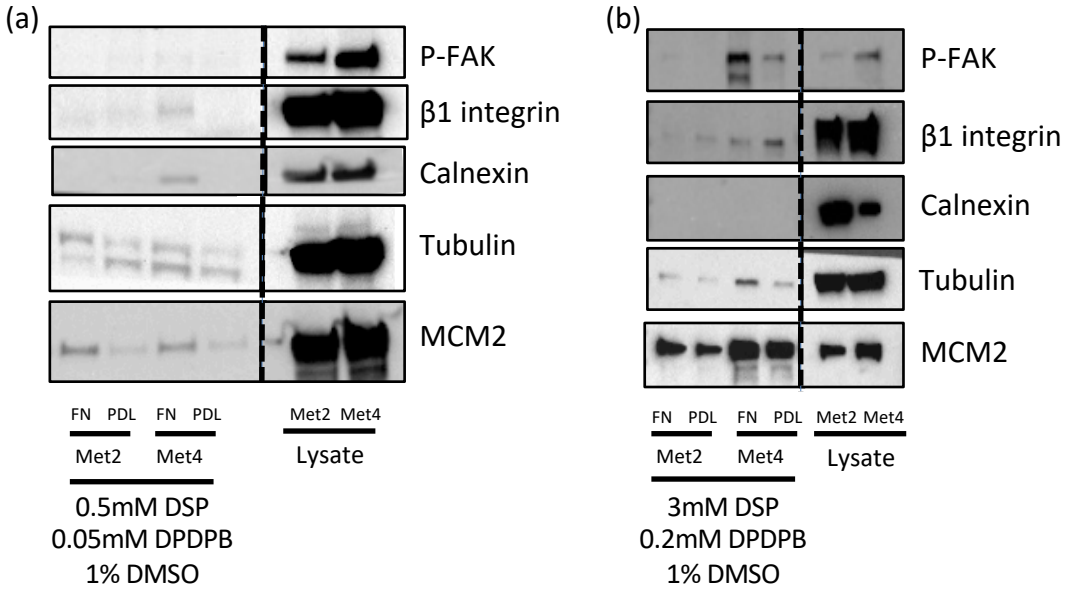


Figure 9 Western blot analysis of adhesion complexes isolated using different concentrations of DSP and DPDPB as crosslinkers. (a) At low crosslinker concentration, 0.5mM DSP and 0.05 mM DPDPB, no substantial enrichment of adhesion proteins phosphorylated FAK (P-FAK) and β1-integrin was observed. However, non-adhesion proteins MCM2 and tubulin were observed in the adhesion fraction of both Met2 and Met4 cells seeded on FN and PDL. Moreover, non-adhesion associated protein calnexin was also isolated in Met4 seeded on FN. (b) At high crosslinker concentrations, 3mM DSP and 0.2mM DPDPB, an enrichment in adhesome proteins P-FAK and β_1-integrin was observed in Met 4 cells but also an increase in the isolation of non-adhesome proteins MCM2 and tubulin. Western blot results were from one experiment.

Incubation period of cells prior to isolation of adhesome

We evaluated the maturity of adhesion complexes at different time periods to assess the optimum spreading time for the Met1 and Met4 cells to form mature focal adhesions on fibronectin. To this aim, Met1 and Met4 cells were incubated for 4 hours and overnight on fibronectin and immunostained for P-FAK. Actin filaments were co-stained by phalloidin. Cells were imaged by laser-scanning confocal microscopy.

After 4 hours incubation, Met1 and Met4 cells developed nascent focal complexes as indicated by P-FAK localised along cortical actin bundles (Figure 10(a)). Met1 and Met4 cells had a high level of P-FAK at the cells' apexes but no mature focal adhesions were observed (Figure 10(a)). We concluded that 4 hours spreading time was not enough for either Met1 or Met4 cells to spread on fibronectin and form mature focal adhesions.

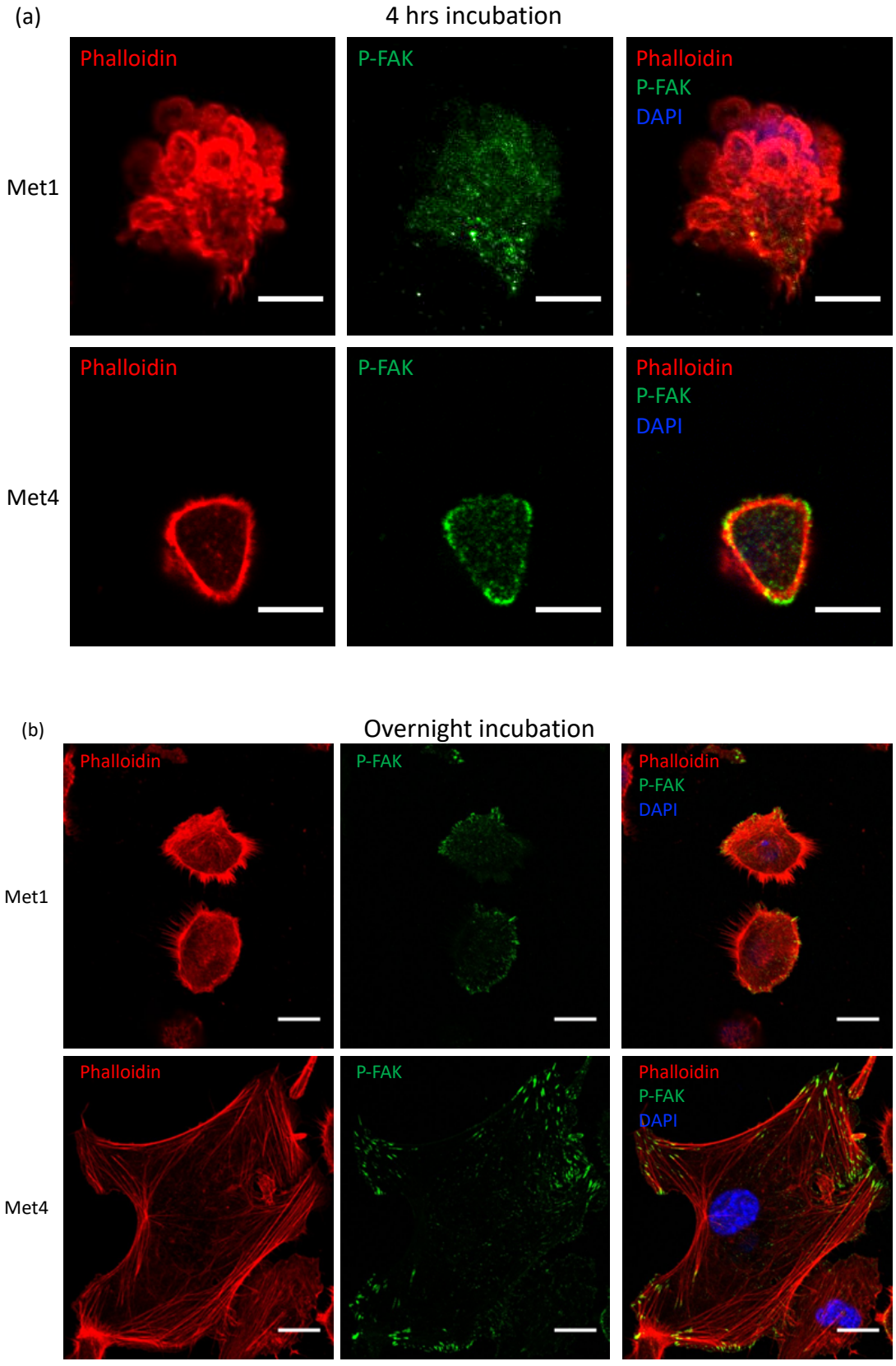

(a)
4 hrs incubation
Met1
Met4
Phalloidin
P-FAK
Phalloidin
P-FAK
DAPI
(b)
Overnight incubation
Met1
Met4
Phalloidin
P-FAK
Phalloidin
P-FAK
DAPI

Figure 10 Confocal image of Met1 and Met4 cells after (a) 4 hours incubation on fibronectin-coated cover slips. Actin (red) was co-stained by phalloidin, nucleus was co-stained with DAPI, and P-FAK (green) was immunolabelled and (b) after overnight incubation on fibronectin-coated cover slips. Actin (red) was co-stained by phalloidin, nucleus was co-stained with DAPI, and P-FAK (green) was immunolabelled. Images were representative of two independent replicates of the experiment.

Conversely, when cells were incubated overnight, focal adhesions were found at the end of actin filaments (presumably actin stress fibres) in both Met1 and Met4 cells (Figure 10(b)), thereby indicating formation of more mature focal adhesions. In Met1 cells, the focal adhesions were mostly found at the lamellipodial leading edge while in Met4, P-FAK was found at the end of actin stress fibres (Figure 10(b)).

Taken together, we concluded that Met1 and Met4 cells should be incubated overnight to allow for the formation of mature focal adhesions prior to adhesion complex isolation. Since cells secrete their own extracellular matrix when incubated for a long period of time[75], we reasoned that cells incubated overnight on PDL would not serve as an appropriate negative control. Moreover, the use of a negative control is not essential when capturing changes in the adhesome under different conditions[65,75].

Optimised Adhesome Isolation Conditions

From the evaluations discussed above, we concluded that specific isolation of the adhesome from Met1 and Met4 cells should be carried out using the following protocol. Met1 and Met4 cells were seeded on 15-cm dish plates

coated with 10 μg/ml fibronectin at a density of $3x10^6$ cells per plate and incubated overnight at 37°C in humidified 5% (v/v) CO_2 . The cells were crosslinked using 3mM DTBP in the presence of DMEM/F12 for 5 minutes at 37°C, after which the crosslinker was quenched with 20 mM Tris.HCl for 2 minutes at room temperature. Cells were washed with an additional 200mM Tris-HCl prior to lysis with ice-cold 5ml FA extraction buffer for 30 minutes. Cell bodies were removed by hydrodynamic force using a Waterpik device with ice-cold PBS. These conditions enriched for adhesion-associated proteins with minimal co-purifying contaminants (Figure 11).

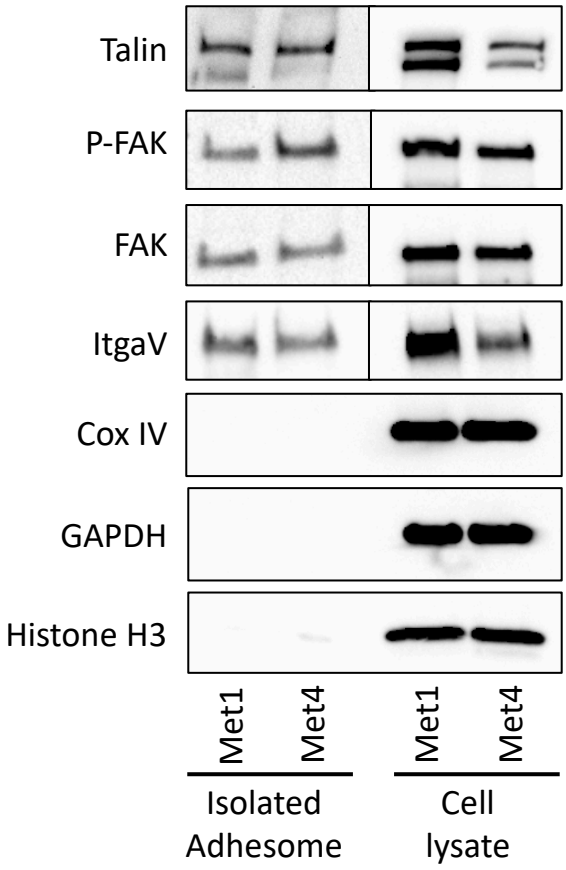


Figure 11 Western blot of optimised adhesion complex isolation for Met1 and Met4 cells. Focal adhesion proteins Talin, P-FAK, FAK and αV-integrin (ItgaV) were isolated whereas non-adhesion proteins such as mitochondrial protein Complex IV (Cox IV),

cytoplasmic protein GAPDH and nuclear proteins Histone H3 (HisH3) were not isolated. Western blot result was representative of two independent experimental replicates.

Adhesome Proteomics Dataset

For proteomic analysis of the human SCC adhesome, adhesion complexes were isolated in three biological replicates for each cell line, Met1 and Met4. The isolated adhesion complexes were analysed by LC-MS/MS, producing an adhesome proteomics dataset consisting of 1941 proteins identified by at least two peptides. The adhesome proteins were quantified by a label-free quantification (LFQ) approach whereby the LFQ of proteins are determined from the quantification of the MS1 spectra of the peptides from which the proteins were inferred to have been derived. MS1 spectra are the mass spectra derived from all the peptides eluted from the liquid-chromatography column at any given time into the first stage of the mass-spectrometer, yielding the m/z and intensity of the peptides[78].

All replicates had a minimal variation in total LFQ intensity (missing values ignored) between the replicates (coefficient of variation = 3.16%, Figure 12(a)). The number of missing values were approximately consistent between replicates with the exception of two biological replicates of Met1 (Rep 1 and Rep 3), both of which had more missing values than any other replicates (Figure 12(b)). Nevertheless, multidimensional scaling (MDS) of the Euclidean distance between the replicates shows that Met1 and Met4 triplicates clustered together as separate condition groups (*i.e.* the cluster of

red spots (Met1) is distinct from the cluster of green spots (Met4) (Figure 12(c)). These results showed that there were no substantial outliers in the biological replicates.

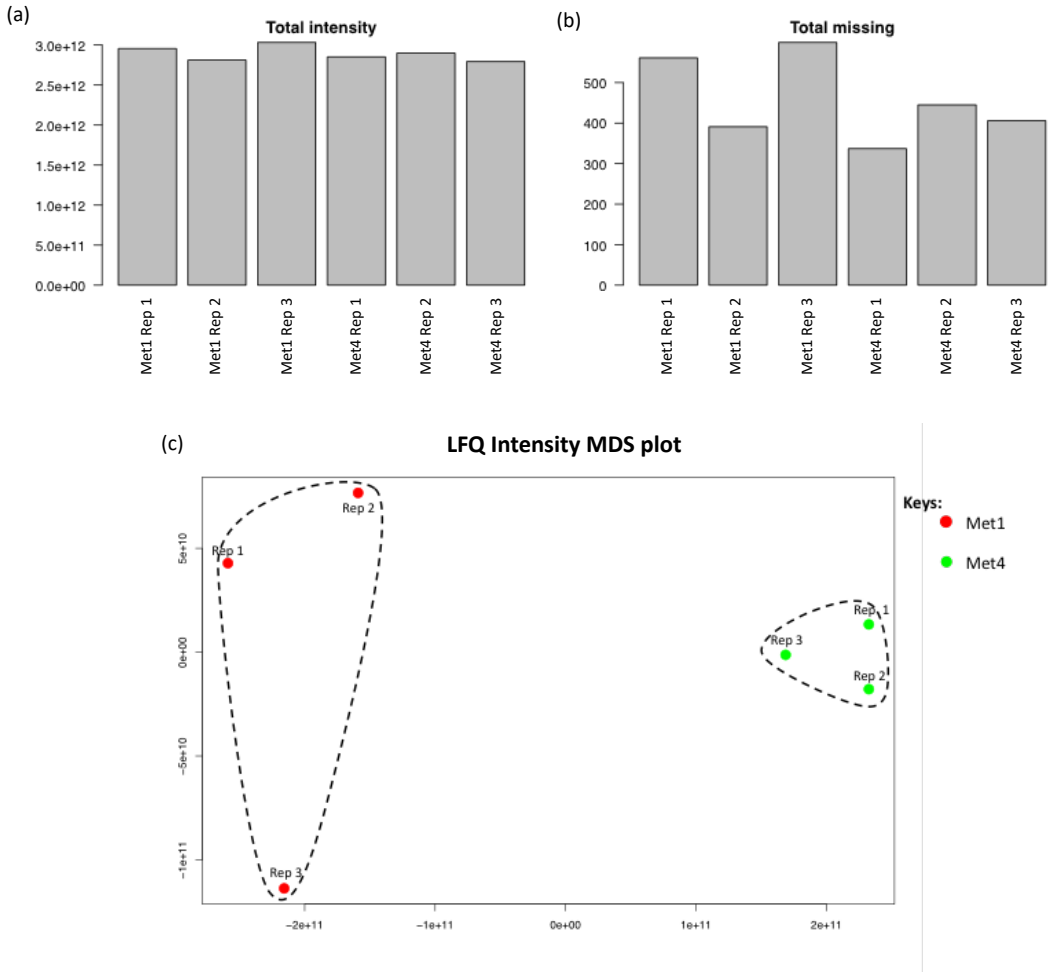


Figure 12 Bar chart of (a) total LFQ intensity of identified proteins and (b) total number of missing values in each replicate of the adhesome mass-spectrometry proteomics dataset. (c) LFQ intensity MDS plot applied to each replicate of the adhesome mass-spectrometry proteomics dataset.

Adhesome Proteomics Dataset Pre-processing

LC-MS/MS experiments can generate high-dimensional data: the abundance level of thousands of proteins measured simultaneously across many biological samples. Unfortunately, the proteomics data generated in such experiments is often affected by extraneous variation and systematic error, both of which are caused by technical factors[135]. Both extraneous variation and systematic error can introduce bias, thereby confounding the analysis of differential proteins abundance level due to biological change. To reduce extraneous variations and correct for systematic error, normalisation of the proteomics dataset is required. Karpievitch *et al*[80] suggested that normalisation of proteomics dataset be performed first, followed by imputation of missing values. Hence, we evaluated different normalisation approaches on the adhesome proteomics first, followed by evaluation of imputation methods on the normalised adhesome proteomics dataset.

Different normalisation methods —total intensity normalisation (TI), median intensity normalisation (MedI), average intensity normalisation (AI), quantile normalisation, variance stabilisation normalisation (VSN), robust linear regression normalisation (RLR) and cyclic loess normalisation (Loess) — were assessed for their ability to reduce intragroup variation between biological replicates of the adhesome proteomics dataset. These normalisation methods were applied globally (denoted by G), *i.e.* the whole dataset was normalised as one. VSN, RLR and Loess were also applied

locally (denoted by R), whereby the condition groups (adhesomes derived from Met1 and Met4 cells) were normalised separately. Prior to normalisation, the LFQ intensities were log_2 transformed to approximate the LFQ intensities into a normal distribution. Intragroup variation was measured using pooled coefficient of variation (PCV), pooled median absolute deviation (PMAD) and pooled estimate of variance (PEV). PCV, PMAD and PEV measure the variation in log_2-transformed LFQ intensities within the protein population across the Met1 and Met4 condition groups. Surprisingly, quantile, TI, AI and MedI normalisation increased the intragroup variation in the replicates compared to log_2 transformation alone (Figure 13). Therefore, these approaches were unsuitable for the normalisation of the isolated adhesome proteomics dataset. Loess, RLR and VSN applied both locally and globally performed better at reducing average PCV, PMAD and PEV than the rest of the normalisation methods (Figure 13). The ranges of PCV, PMAD and PEV were also reduced when Loess, RLR and VSN were applied globally compared to when applied locally. Globally applied RLR (RLR-G) normalisation produced the smallest range of PCV, PMAD and PEV (Figure 13). Thus, normalisation by RLR-G performed the best in reducing the intragroup variation in the isolated adhesome proteomics dataset and was used for the normalisation of this dataset.

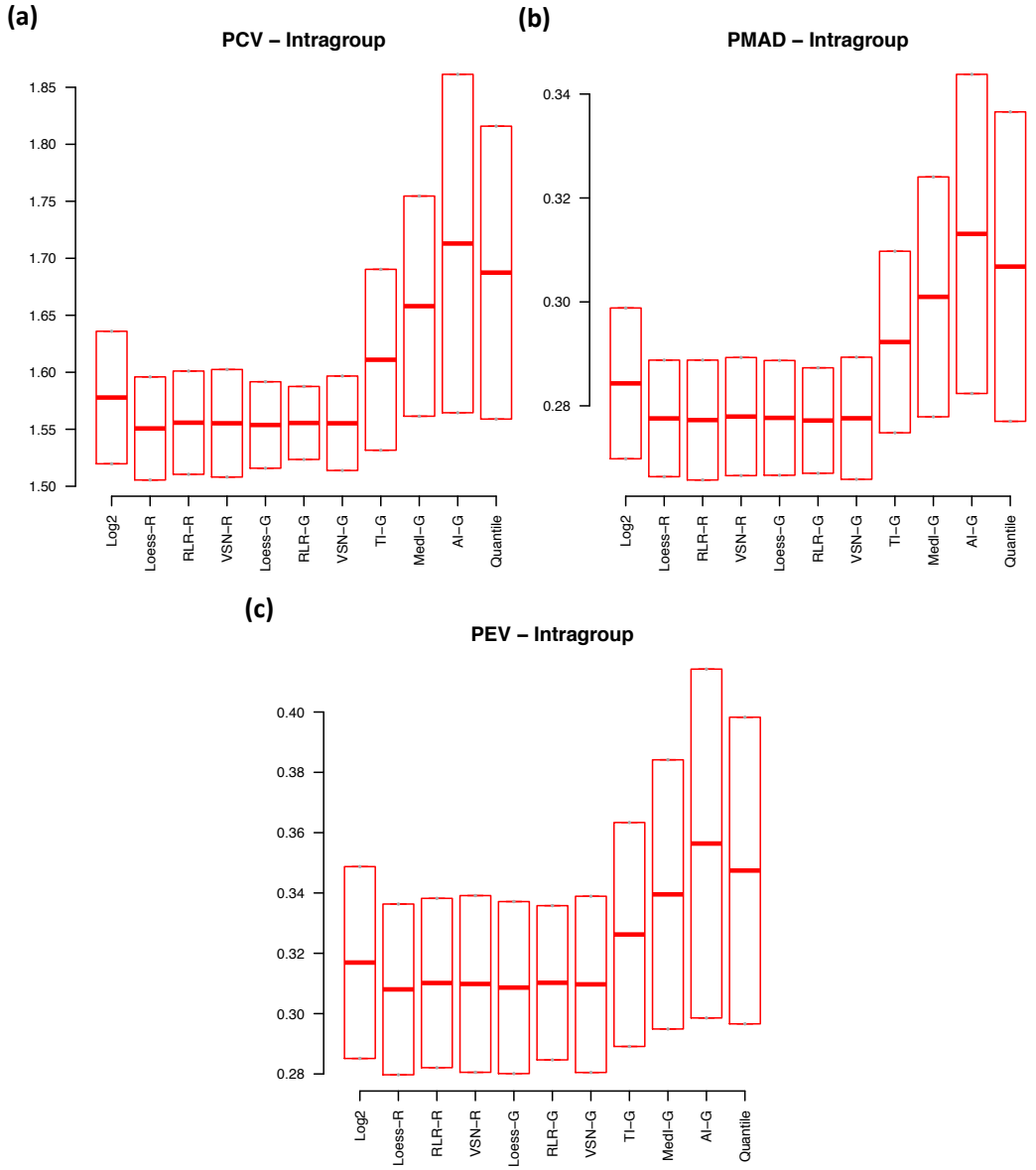


Figure 13 Comparison of the effects of normalisation methods on the intragroup variation measured by* (a) *PCV, (b) PMAD and (c) PEV. Loess, RLR and VSN applied both globally (-G) and locally (-R) produced the lowest average PCV, PMAD and PEV. RLR-G produced the smallest range of PCV, PMAD and PEV.

MA plots visualise the differences between the protein abundance measurements, by transforming the data onto the difference between biological replicates of (M) and average of (A) $\log_2$-transformed LFQ values for each protein. MA plots were used to explore the bias in the different biological replicates. The MA plot for ideally normalised data would have a

point cloud scattered along the x-axis (M=0)[110] with the loess smoother line (non-parametric line of best fit) lying along the x-axis. Identical biological replicates are not essential for this property to be observed. Rather, the samples only have to be biologically related such that most proteins are assumed not to be differentially abundant. Since Met1 and Met4 cell lines are both squamous cell carcinoma (SCC) cells derived from the same patient, we hold this assumption to be true. Figure 14 shows an MA plot whereby LFQ intensities of proteins quantified in Met1 Replicate 1 are set as the reference against which the LFQ intensities of proteins quantified in the other biological replicates are subtracted, and A is the average LFQ intensity for each protein among all biological replicates. Prior to normalisation, the loess smoothers noticeably deviated from the x-axis at the dense region of the point cloud (Figure 14(a)). Moreover, the loess smoothers were 'banana' shaped (Figure 14(a)), as has been previously described[79,109,136]. This suggests that the biological replicates were biased prior to normalisation, *i.e.* most proteins were inherently differentially abundant as a result of extraneous variations and systematic error and not due to biological change (primary vs metastatic cells). Normalisation by RLR-G reduced the deviation at the dense region of the point cloud though it did not correct for the deviation at the higher average LFQ intensity values and lower average LFQ intensities. The loess smoothers appear to be modified at the extremes by only a few data points. Thus, it is possible that the observed deviation at the high and low abundance proteins is due to biological change between adhesomes derived from Met1 and Met4 cells.

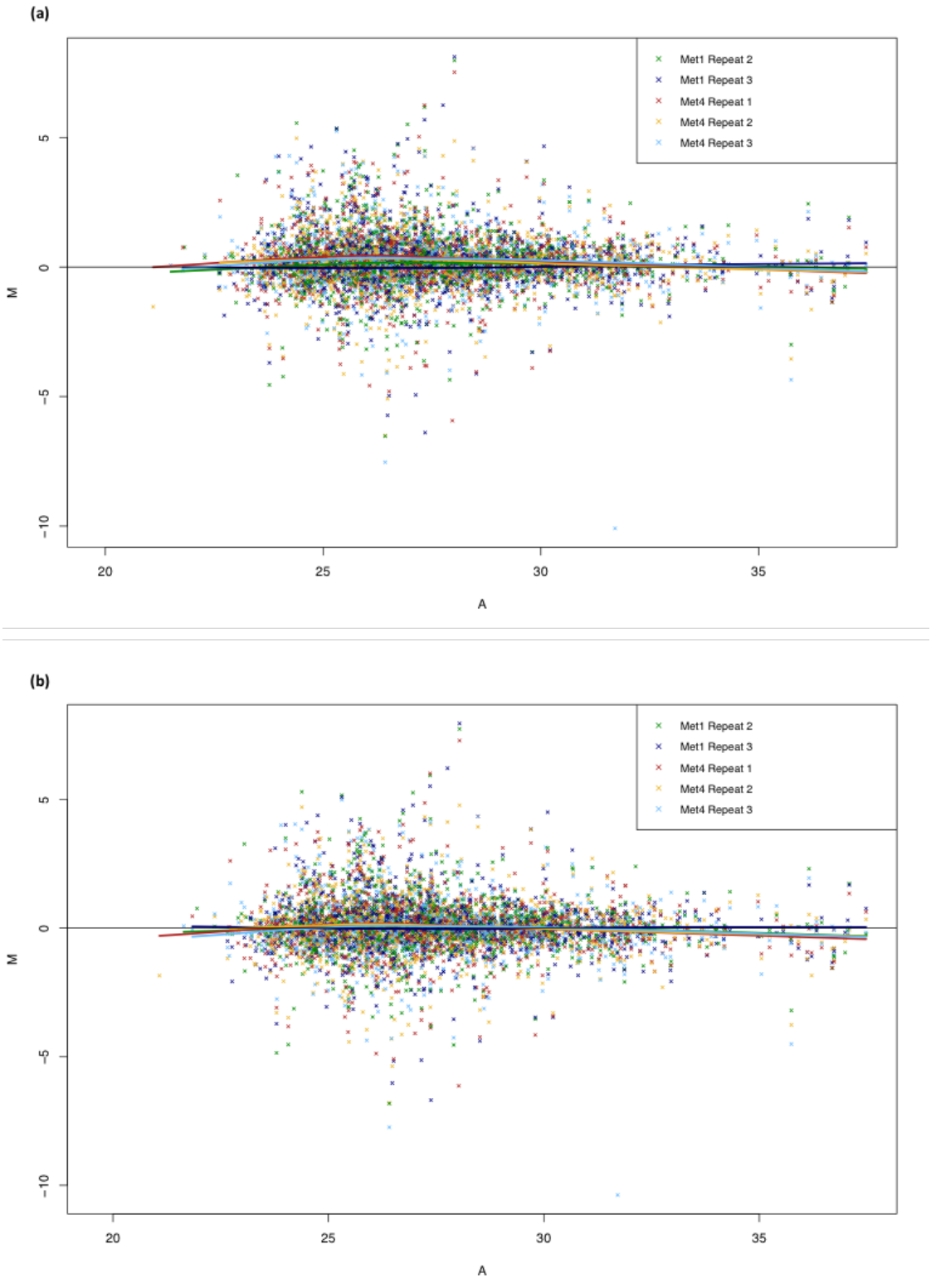


Figure 14 MA plot (a) before and (b) after RLR-G normalisation. The scatter points represent the difference versus the average protein intensities. The solid lines are

loess smoothers. Different sets of replicates for Met1 and Met4 are represented by different colours.

Imputation

Following RLR-G normalisation, the patient-derived SCC adhesome dataset was filtered for proteins which contained at least two log_2-transformed LFQ intensity values in either condition group. This resulted in 1727 adhesome proteins which have been identified in at least two biological replicates in the adhesome dataset. The missing values were imputed using MI and LSA.

LSA is a function of the LSImpute package[87]. LSA is a local similarity-based imputation approach[86] whereby it is assumed that proteins which are co-regulated will have highly correlated expression profiles. The missing values are estimated by multiple regressions between the proteins with missing values and their respective most correlated proteins as well as multiple regressions between the biological replicates.

The MI method used here involved a first-pass single value imputation (SVI) dependent on satisfaction of an assumption criterion followed by a second-pass imputation by predictive mean matching (PMM). PMM has the advantage of being a non-parametric imputation method; hence, the data need not follow a normal distribution at this stage. The first-pass imputation criterion is as follows: if and only if a protein had missing LFQ intensity values in all biological replicates in the same condition group, we assumed it

was missing-not-at-random (MNAR) and therefore the missing values in each replicate were imputed by the minimum LFQ intensity value (of all proteins) in that particular replicate.

Prior to imputation, 16.5% of the entire dataset were missing values. After the first-pass imputation, 8.7% of the entire dataset were missing values. These remaining missing values were treated as not MNAR and therefore imputed by PMM. Suppose a dataset Y, consisting of both observed values (Y_{obs}) and missing values (Y_{mis}). Using Bayesian methods, a posterior predictive distribution for Y_{mis} using Y_{obs} is generated and imputations are drawn from this posterior distribution. Using all imputed values and observed values, another posterior predictive distribution for Y_{mis} is generated and imputation values are again drawn from this second pass posterior predictive distribution. This process is repeated multiple times, refining the posterior predictive distribution for Y_{mis} until the distribution stabilises. At this stage, the predicted values for the observed values in the imputed dataset will be close to the initial observed value (Y_{obs}) of the corresponding dataset with missing values. Because PMM uses a Bayesian approach to 'predict' the missing values, it is useful to test whether the imputed values by PMM follow the same distribution as the original dataset. Figure 15 shows scatter plots of Met1 Replicate 1 against all the other replicates. The imputed values (magenta) fell within the same point spread as the original observed values (blue), indicating the imputed values were indeed 'plausible' values. Most of

the imputed values were low LFQ intensities, which is in agreement with the inverse correlation between missing values and protein abundance[137].

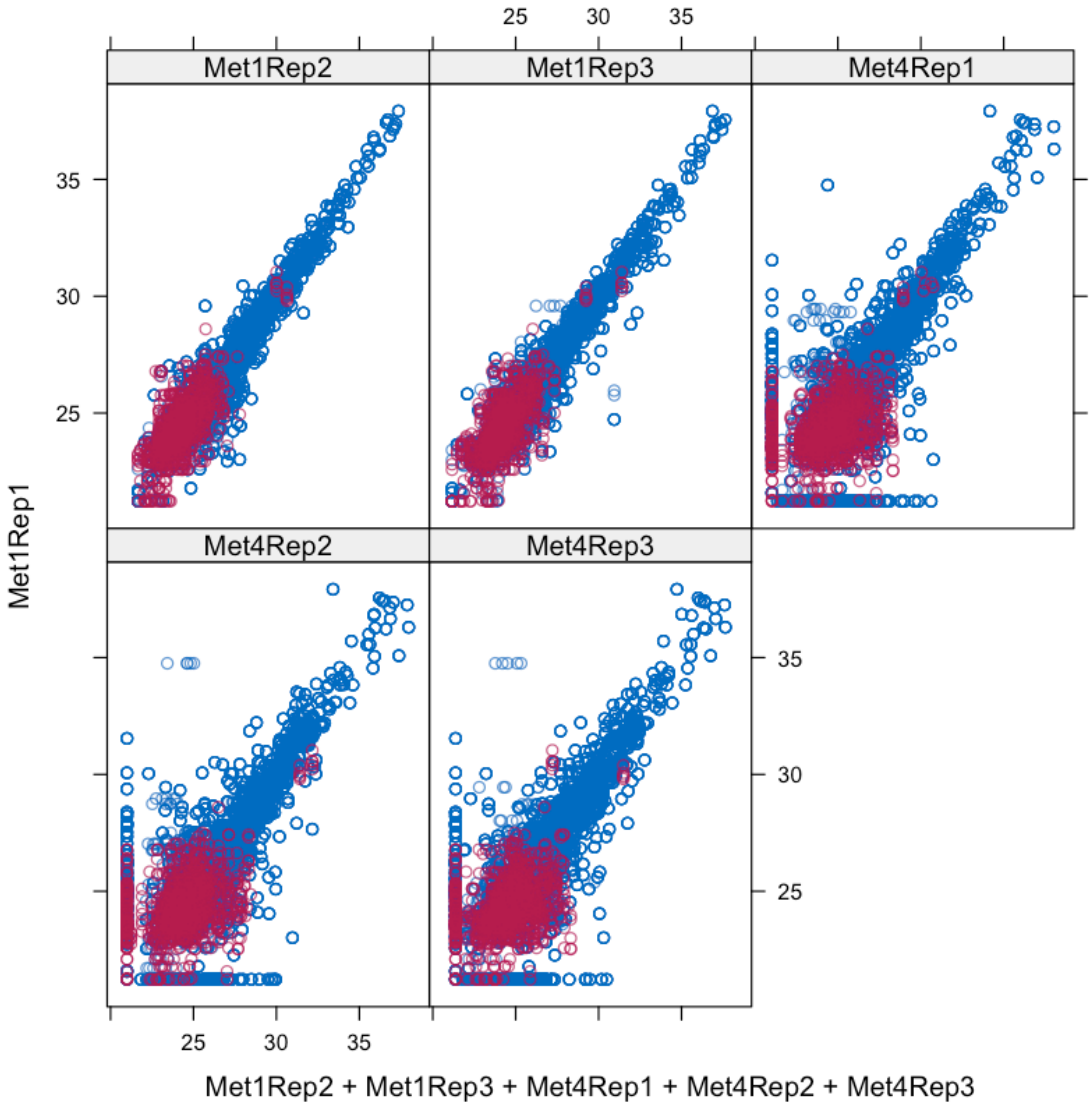


Figure 15 Scatter plot of log_2 transformed LFQ intensities of Met1 Replicate 1 versus log_2 transformed LFQ intensities for all other replicates. The original observed values are coloured blue, while the imputed values are in magenta.

Assessing the accuracy of the imputed values is not feasible in this experimental setting as the true LFQ intensity values are unknown.

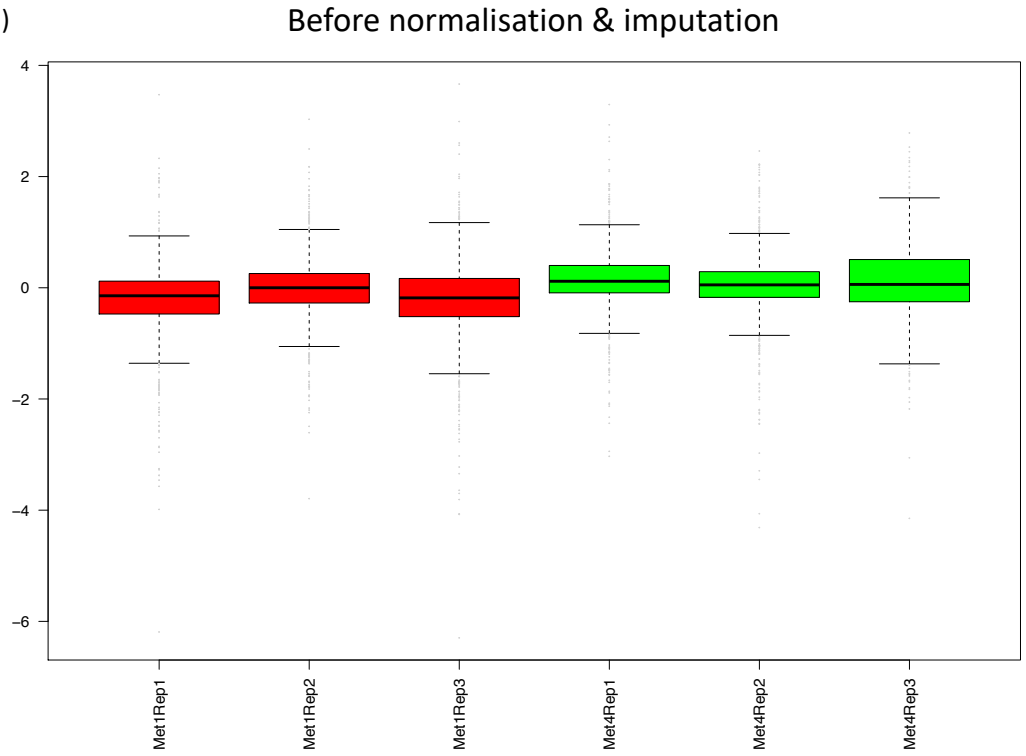
However, the aim is to obtain an accurate estimate of the log fold changes and not an accurate estimate of the LFQ intensities. Thus, the log fold change (logFC) of LFQ intensity in the imputed datasets was compared to the relative protein abundance (RPA) obtained from western blot validation of the isolated adhesome. Western blotting is a low-throughput method and thus the relative protein abundance of only a few proteins could be assessed. Moreover, only two western blot replicates were produced and hence the RPA estimates were only preliminary. Nevertheless, the RPA and logFC of three proteins namely Mena, αV-integrin and FAK, were compared. The accuracy of the fold change was assessed by the root mean square deviation (RMSD) between the RPA and the logFC for those three proteins. The lower the RMSD, the more accurate the estimate is to the true value. Mena and αV-integrin were both imputed by SVI as both had replicates missing completely in the isolated adhesome from Met1 and Met4 cells, respectively. FAK was imputed by PMM. MI outperformed LSA at accurately estimating differential protein abundance with RMSD of 7.33 for MI compared to RMSD of 31.29 for LSA. Hence, the adhesome proteomics dataset was imputed using MI.

We used a relative log expression (RLE) plot to check for bias[138] in the isolated adhesome proteomics biological replicates before and after normalisation and imputation. The RLE plot is a powerful visualisation tool to check for sample heterogeneity by computing the deviations between, in this

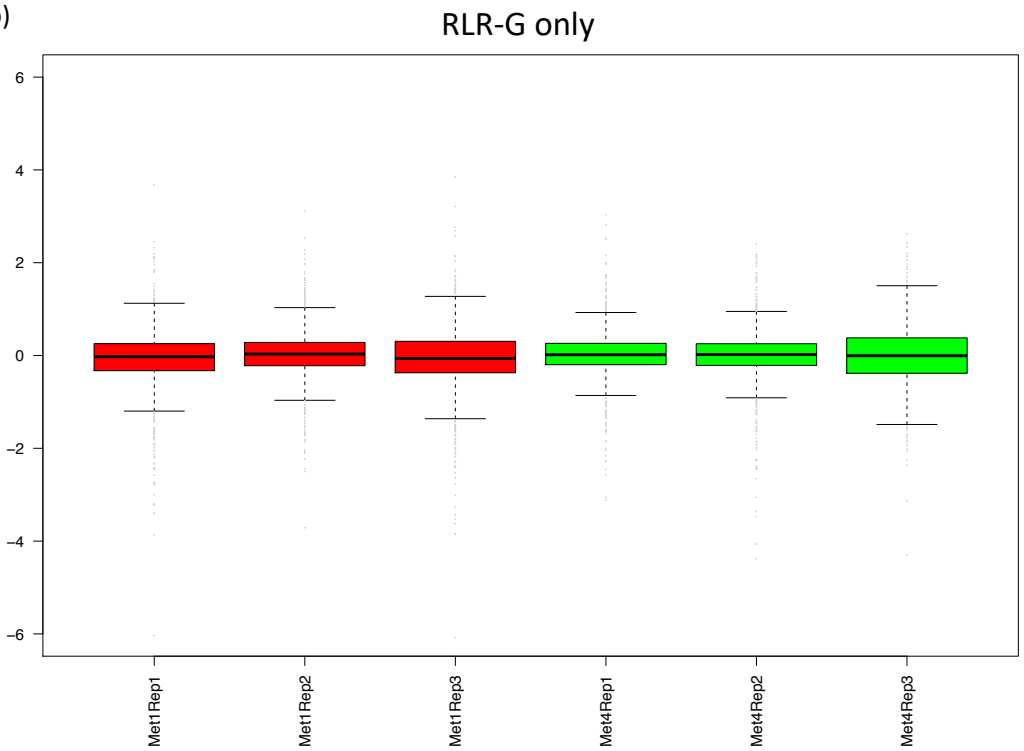
case here, the log_2-transformed LFQ intensity for each protein in each biological replicate and the median log_2-transformed LFQ intensity for each protein across the replicates. The median log_2-transformed LFQ intensity is used to protect against outliers. The computed deviations of all proteins are represented by a boxplot for each biological replicate.

Prior to normalisation and imputation, the biological replicates were biased as indicated by the varying positions of the boxplots (Figure 16(a)). This agrees with the earlier MA plot analysis indicating a similar conclusion (Figure 14(a)). Conversely, upon RLR-G normalisation, the mean deviations of the biological replicates centred along 0 in the RLE plot (Figure 16(b)), indicating RLR-G reduced the inherent bias present in the biological replicates prior to normalisation. Moreover, imputation by MI did not affect the distribution of the mean deviations, indicating that no bias was introduced by MI imputation (Figure 16(c)). We are therefore confident that the normalised and imputed adhesome proteomics dataset could be used for further downstream analysis.

(a)


Before normalisation & imputation
4
2
0
−2
−4
−6
Met1Rep1
Met1Rep2
Met1Rep3
Met4Rep1
Met4Rep2
Met4Rep3

(b)


RLR-G only
6
4
2
0
−2
−4
−6
Met1Rep1
Met1Rep2
Met1Rep3
Met4Rep1
Met4Rep2
Met4Rep3

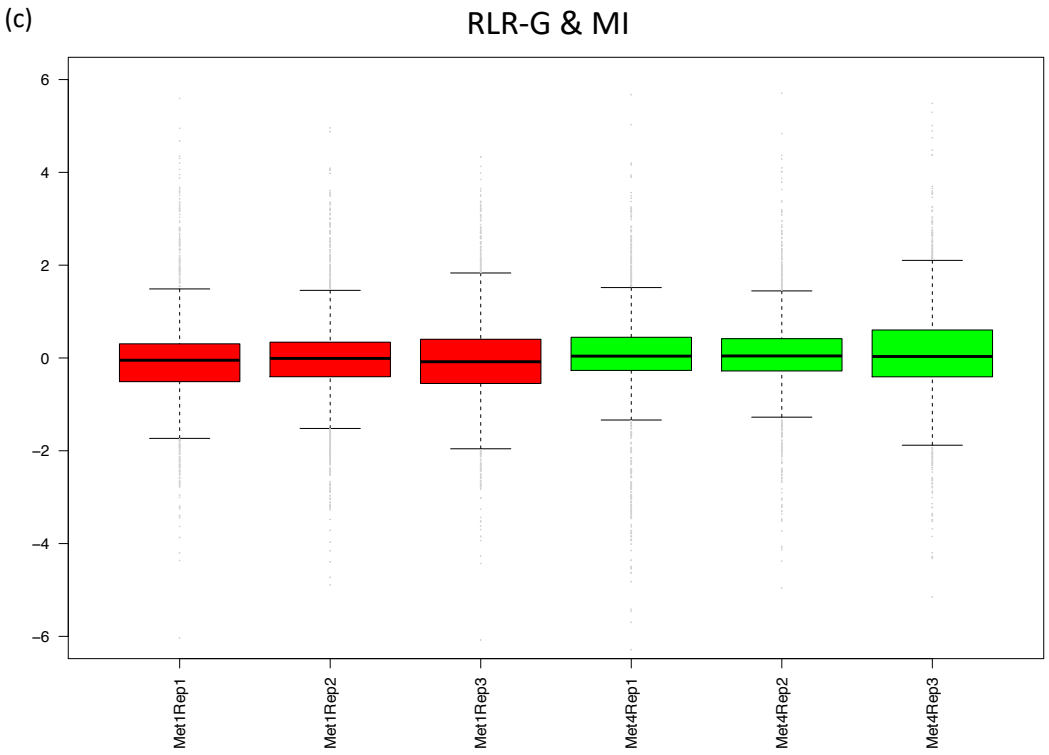


Figure 16 RLE plot of the log_2 transformed LFQ of the isolated adhesome for each biological replicate (a) prior to normalisation (b) after RLR-G normalisation and (c) after RLR-G followed by MI imputation.

■ Discussion

In order to study the adhesome in a cancer progression setting, we isolated adhesion complexes from malignant cutaneous SCC cells derived from the primary site (Met1) and secondary metastatic site (Met4) of the same patient. The adesome isolation methodology was optimised for Met1 and Met4 cells to enrich for adhesome proteins — such as FAK, phosphorylated FAK, talin and αV-integrin — whereas co-purifying contaminants, such as nuclear protein Histone H3, mitochondrial protein complex IV and cytosolic protein GAPDH, were not isolated. However, the optimisation experiments were

carried out once. Three biological replicates for each Met1 and Met4 adhesome fraction were analysed by LC-MS/MS to produce an adhesome proteomics dataset.

Data pre-processing was performed on the adhesome dataset with the aim of reducing bias and variability, as well as improving accuracy for protein abundance analysis downstream in the analysis pipeline. Different normalisation methods were evaluated for best reducing intragroup variability and bias. RLR-G showed the best performance at reducing intragroup variability as measured by PCV, PMAD and PEV. Moreover, both the MA plot and RLE plot showed that RLR-G normalisation reduced bias in the LFQ intensities of the biological replicates.

A multiple imputation method (MI) was used which accounted for different types of missingness. Here, performance of the imputation method was assessed for accuracy in estimating the log fold change rather than estimating the LFQ intensity. MI outperformed LSA in accurately (as assessed by RMSD) estimating the fold change. This agrees with the finding that multiple imputation strategies are known to produce the best results[83]. Moreover, imputation by MI did not introduce bias in the LFQ intensities as shown by RLE plot analysis.

Both MI and LSA were carried out at the protein level. It has been reported that MCAR imputation performs best at the peptide level, before peptide aggregation into proteins, while MNAR imputation performs best at the protein level[83]. However, this has not been implemented here since the peptide aggregation software used (MaxQuant) did not, at the time of analysis, offer imputation at the peptide level. Nevertheless, the evaluation here provides us with confidence that the normalised and imputed adhesome proteomics dataset can be used for differential protein abundance analysis in the context of a protein-protein interaction network (chapter 4).

CHAPTER 4

Results: Network Analysis of the Adhesome in a Cancer Progression Setting

■ Introduction

In the chapter 3, adhesion complexes isolated from patient-derived squamous cell carcinoma (SCC) cells were analysed by LC-MS/MS and the resulting proteomic dataset normalised and missing data imputed. Using differential protein abundance analysis alone treats the proteomics data as sets and ignores information such as the interaction between the individual proteins. The adhesome proteins interact with one another to form an adhesome network[61]. Therefore, we analysed the dataset for differential protein abundance using a network approach. .

It is becoming evident that disease-associated proteins (such as those differentially regulated between disease phenotypes) interact with each other to form modules in the interactome map[88,89]. Those modules (disease-specific functional modules) are postulated to drive disease phenotypes[88]. In a similar way, differentially abundant adhesome proteins between the patient-derived Met1 (primary) and Met4 (metastatic) SCC cells may interact with one another in the adhesome network to form a module (or modules) responsible for cancer progression.

An important feature of protein-protein interaction networks is the presence of highly connected proteins (in other words, proteins which interacted with multiple proteins) called hubs. During cancer progression, some hubs rewire

within the signalling network to drive phenotypic alterations but still maintain the robustness of the signalling network[100]. It is therefore important to study such network feature as hubs may be functional targets for inhibiting metastasis.

Han *et al.*[98] reported two classes of hubs based on the average Pearson correlation coefficient (avPCC) of expression levels between the hub and its interacting partners. 'Party' hubs were highly coexpressed with their neighbours (high avPCC) and thus coordinate proteins within the same community (intra-module hub)[99]. 'Date' hubs, which had relatively low coexpression (low avPCC), were considered as inter-module hubs. The latter interact with different binding partners under different conditions and may thus perform different functions[99-101]. Hence, the identification of 'date' hubs in the adhesome network could identify a 'switch' protein whose aberrant interaction with different binding partners may drive a metastatic phenotype.

Guimera and Amaral[94,103] classified hubs in terms of their structural roles within their respective communities. Communities are clusters of interacting nodes which are highly interconnected to one another. In protein-protein interaction networks, proteins within a community are postulated to participate in the same molecular processes[104]. Nodes with within-

community degree (z) $\geq$ 2.5 are considered as hubs. Those hubs can be further categorised into provincial hubs, connector hubs and kinless hubs according to their participation coefficient (see Methods).

The classification of hubs such as date and party hubs[98], connector, kinless, and provincial hubs[94,103], inter-module and intra-module hubs[100,101] is not mutually exclusive. Indeed, connector and kinless hubs have been reported to behave as date hub, and provincial hubs have been reported to behave as party hubs, though the similarity between provincial and party hubs was modest[99].

The aim here is to understand how the adhesome changes during cancer progression and subsequently identify adhesome hubs which may be crucial to cancer progression. Moreover, we wish to identify kinless and connector hubs (inter-module hubs, date hubs), as these hubs have been reported to integrate different molecular processes and perform different functions under different conditions[100] and hence may be sites of rewiring in the adhesome network driving metastasis.

■ Results

Active Module

The first step in the network analysis is to integrate changes in adhesome proteins abundance between Met1 and Met4 cells into the network system under study. Here, the network under study is the adhesome SCCs network, which consists of patient-derived SCC adhesome proteins isolated from Met1 and Met4 cells and their putative protein-protein interactions obtained from BioGrid[139] — a curated repository database for protein-protein interactions. Once the changes in adhesome proteins abundance are integrated in the adhesome SCCs network, we identify a connected module within the network that contains statistically significant differentially abundant adhesome proteins between the Met1 and Met4 as these proteins may interact with one another to drive a metastatic phenotype. Such a connected module is called the active module. To this aim, a comprehensive R package for finding active modules, BioNet[115,116], was used.

The patient-derived adhesome dataset consists of 1727 proteins identified by LC-MS/MS and isolated from at least two biological replicates of either Met1 or Met4 cells. The *p*-values of the $\log_2$-transformed fold changes between the Met1 and Met4 adhesome were converted into scores using a signal-noise decomposition scoring function. The signal component of the scoring function is modelled by a beta distribution function B(*a*,1), where *a* is the

shape parameter of the signal component. The noise component is modelled by a uniform beta distribution B(1,1). Taken together, the signal and noise component are fitted to a beta uniform mixture model. The scoring function, $S(x)$, was thus defined as the log-likelihood ratio of the beta distribution to the uniform distribution:

$$S(x) = log\left(\frac{B(a,1)(x)}{B(1,1)(x)}\right) = \log(a) + (a-1)\log(x)$$

where x is the empirical *p*-value and a is the shape parameter of the signal component.

When adjusted for false discovery rate, the scoring function can be simplified to:

$$S(x) = (a-1)(\log(x) - \log(\tau(FDR))$$

where x is the *p*-value, a is the shape parameter and $\tau(FDR)$ is the false discovery rate threshold.

The *p*-values determined from the adhesome dataset using Student's t-test were fitted to a beta uniform mixture model (Figure 17(a)) and the shape parameter of the signal component, a, was determined using the L-BFGS-B algorithm (called for by the BioNet package) which resulted in a = 0.3275. The Q-Q plot shows that the distribution of the observed *p*-values from the adhesome dataset does follow the fitted beta distribution (Figure 17(b)). The

score for each node was thus computed using an FDR of 0.05 and $a = 0.3275$.

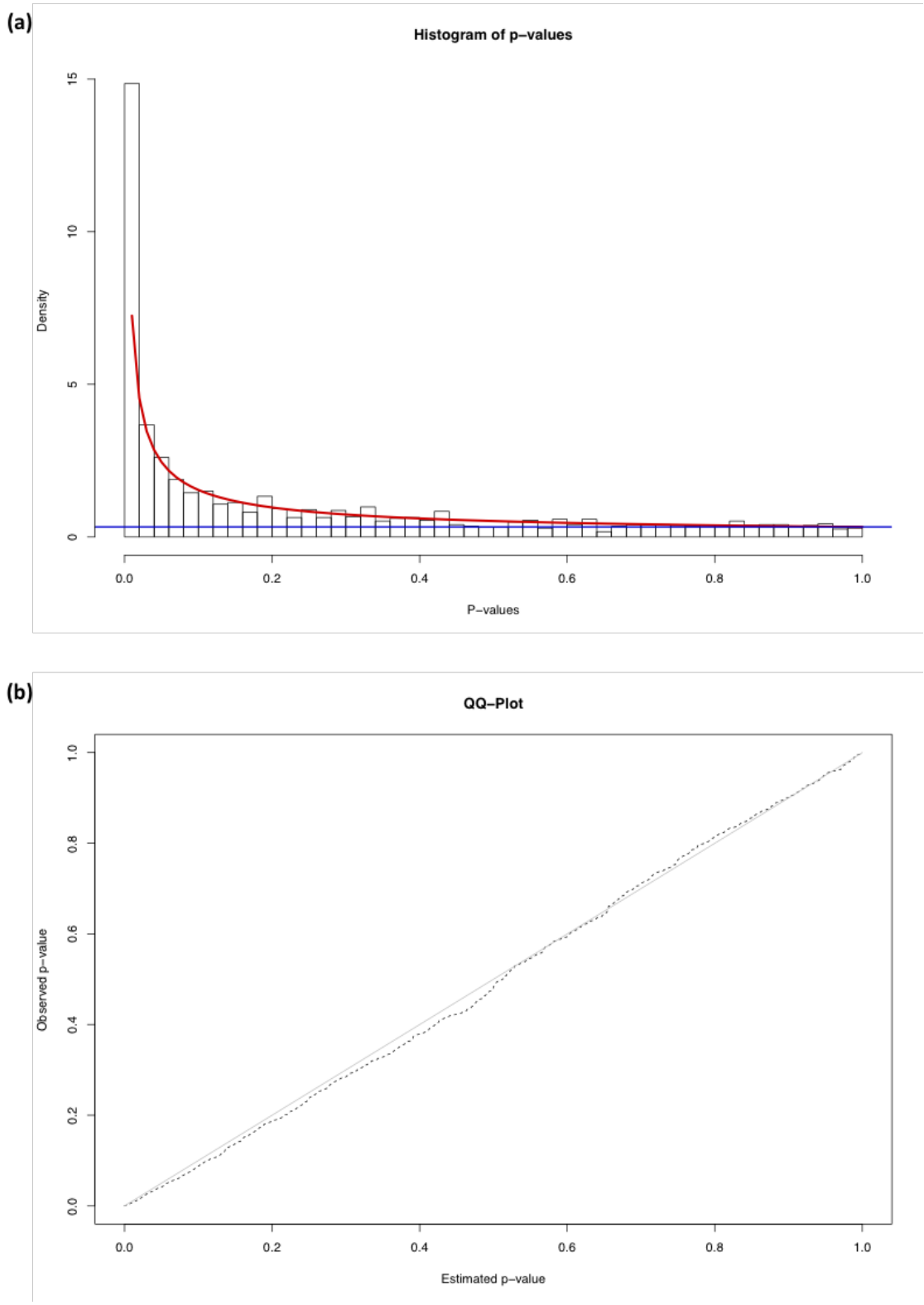


Figure 17 (a) Histogram of the p-values from the adhesome dataset fits the beta uniform mixture model. The signal component is denoted by the red line. The noise component is denoted by the blue line. The parameter for the mixture model is a = 0.3275. (b) The

observed p-values form the adhesome dataset fits the beta distribution as shown by the Q-Q plot.

Of the 1727 isolated adhesome protein, 1708 proteins were mapped onto the BioGrid physical protein-protein interaction network. The 19 unmapped isolated adhesome proteins were missing in the BioGrid physical protein-protein interaction network[139] at the time of the analysis. Of the 1708 mapped proteins in the BioGrid protein-protein interaction network, 1668 proteins formed a connected network. This connected network is referred to as the adhesome SCCs network. Around 60% of the proteins in the adhesome SCCs network also occur in the meta-adhesome (Figure 18). The latter is a database of fibronectin-induced adhesome proteins identified by mass spectrometry[68] (Figure 18).

Figure 18 Venn diagram illustrating the number of proteins in the adhesome SCCs which also occur in the meta-adhesome database.

The scores derived from the beta-uniform mixture model were mapped onto the adhesome SCCs network. The active module is the maximal-scoring connected subnetwork[116,140,141]. However, it has been mathematically proved that extracting the best-scoring connected subnetwork is an NP-hard problem[140] and thus only suboptimal solutions are obtained using heuristic approaches[141,142]. The advantage of using BioNet is that the problem of finding the maximal-scoring connected subnetwork is converted into a prize-collecting Steiner tree problem (PCST). The latter is then solved using integer linear programming, which produces a subnetwork considered as the

optimal best-scoring connected subnetwork. Using this approach on the scored adhesome SCCs network, an active module consisting of 720 proteins was obtained (Figure 19). The active module represents the statistically significant differentially abundant adhesome proteins between the Met1 and Met4 and how they are connected to one another (Figure 19). The active module shares around 59% of proteins with the meta-adhesome dataset (Figure 20).

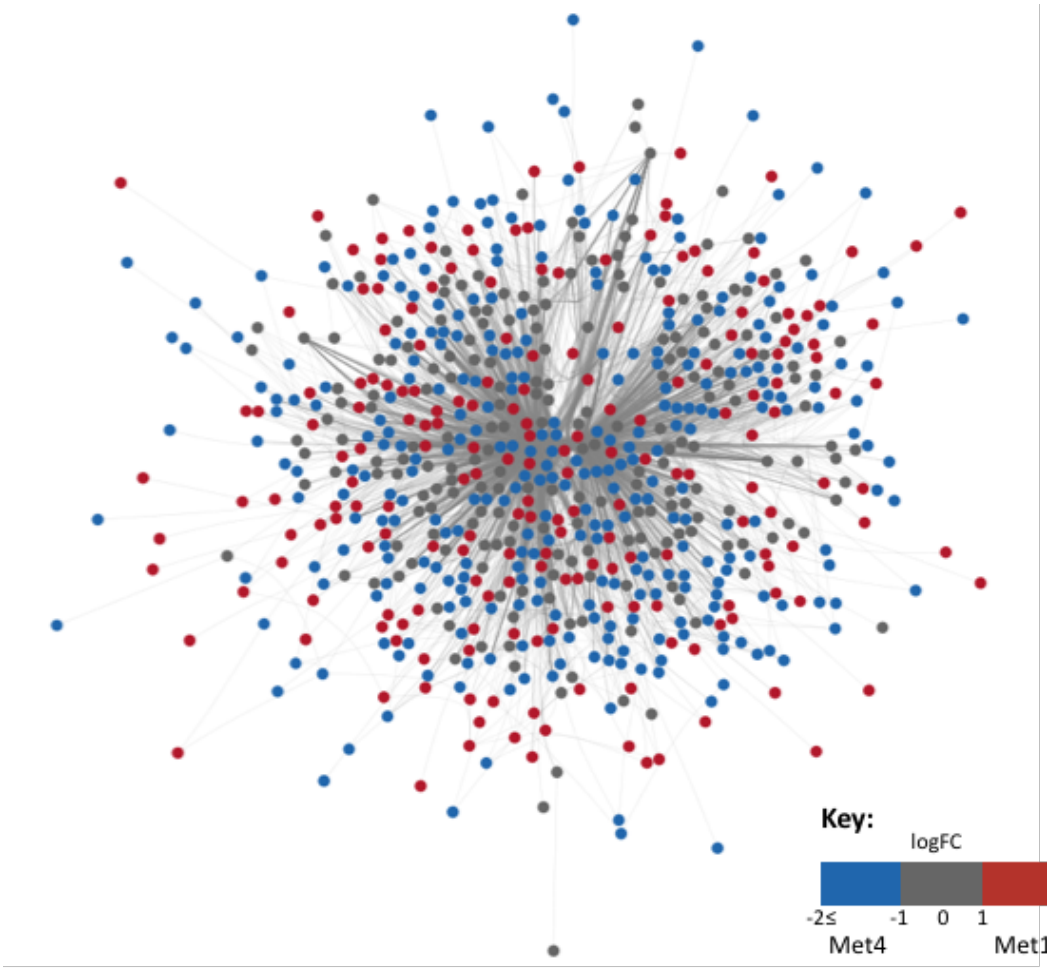


Figure 19 The active module identified by the scores based on the p-values obtained using Student's t-test and an FDR of 0.05. Blue nodes represent adhesome proteins which have at least two-fold upregulation in Met4 (metastatic). Red nodes represent adhesome proteins which have at least two-fold upregulation in Met1 (primary). Grey

nodes represent adhesome proteins which are not differentially regulated but are essential for the connectivity between the differentially regulated proteins.

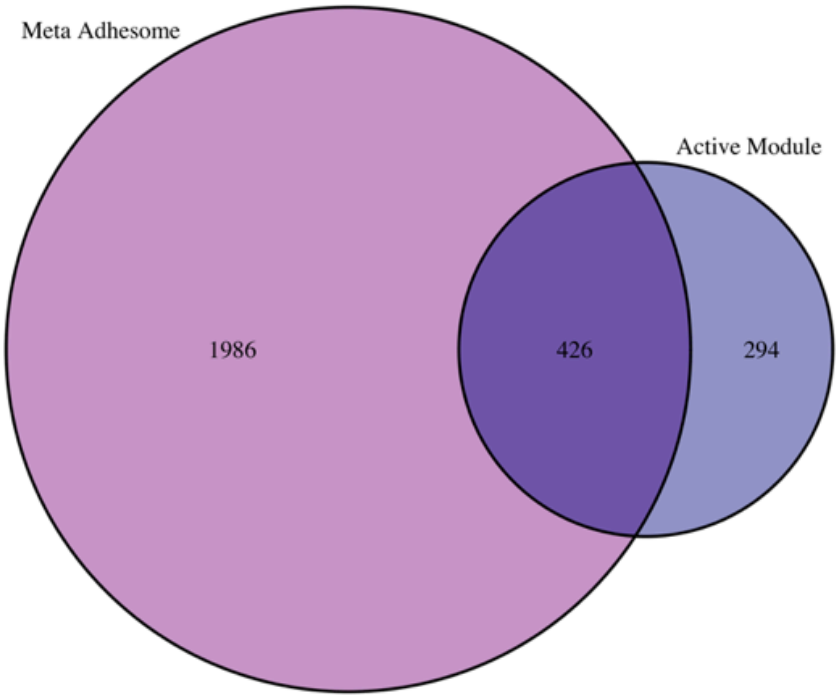


Figure 20 Venn diagram illustrating the number of common adhesome proteins in the active module to that of the meta-adhesome dataset.

Cartography of active module

To understand the organisation of the active module and identify the kinless and connector hubs, it is essential to obtain a cartographic picture[94,103] of the active module. To this aim, the active module was partitioned into community structures and the within-module degree and participation coefficient of each node were computed. Community structure is a property

common to many networks, including biological networks[143], and gives a mesoscale view of a biological system. Nodes within the same community in a protein-protein interaction network have been postulated to participate in the same molecular processes. Indeed, knocking out the nodes within the same community resulted in the same phenotype in yeast, suggesting that nodes within the same community participate in the same molecular processes[104]. A plethora of community structure algorithms[95,144] have been developed, each using different definitions of a community. The community structure algorithms used are the Louvain, spectral optimisation of modularity, Infomap and Constant Potts model (CPM). The Louvain method[117] and spectral optimisation of modularity approach[120,121] both define communities as clusters within the network with high links density compared to the rest of the network. Infomap considers communities as regions within the network where information flows quickly and easily[145]. CPM as applied by Traag *et al.*[123] is a resolution-limit-free approach that uses a tuneable resolution parameter. To find the best resolution parameter for CPM, a quality function called Surprise (*S*)[125] — a measure of how unlikely a given community structure is found in a random network — was used. Thus, the best resolution parameter for CPM will produce maximum Surprise (*i.e.* the community structure is most unlikely to occur in a random network). These four community detection algorithms were used to obtain a consensus role of the individual nodes with respect to the community structure.

The extracted active module is static and does not reflect the dynamics of protein-protein interactions under different biological conditions[98,146]. Thus, to capture the dynamics of the protein-protein interactions, the edges were weighted using the Pearson rank correlation coefficient which was previously found to be the most effective measure of functional links[147]. The Pearson rank correlation coefficient was increased by a value of one integer so that only positive values were obtained. Thus, edges between positively correlated interacting proteins have higher weights than those of negatively correlated interacting proteins.

The Louvain method[117] partitioned the active module into 12 communities (Figure 24(a)) resulting in a Girvan-Newman modularity value of 0.475. To test the significance of the partition (i.e. that the partition is not due to randomness), 1000 random networks with identical degree distribution and edge weights (also called a configuration model) as the active module[148] were generated and the modularity computed. This resulted in modularity of 0.255 ± 0.003. Hence, the active module has a significant modularity (z-test, *p*-value <0.0001, z-value 2435.0) suggesting that the community structure obtained is a relevant feature of the network and not a by-product of randomness and the degree sequence.

The Girvan-Newman modularity obtained by spectral optimisation was 0.553, with 57 communities detected (Figure 24(b)). Some of the communities were single nodes. Spectral optimisation seemed to result in a better partition (higher modularity value) compared to the Louvain method; however, the significance of the partition was not tested.

Modularity-based methods suffer from resolution limit[149], whereby the size of the communities depends on the size of the network. CPM is a resolution-limit-free approach, which uses a tuneable resolution parameter γ in the range of 0 to 1. Hence, the size of the communities depends on γ. The active module was partitioned using a resolution parameter, γ, of 0.0479, which produced maximum *S* (Figure 21). The latter parameter measures how unlikely a given community structure is found in a random network. Using maximum *S* has been found to give accurate community structures[125]. This approach identified 177 communities, many of which were single nodes.

Infomap finds community structures within a network by tracking the information flow through the network[145,150]. Infomap partitioned the active module into 50 communities, some of which are single nodes. The significance of the partition was not tested.

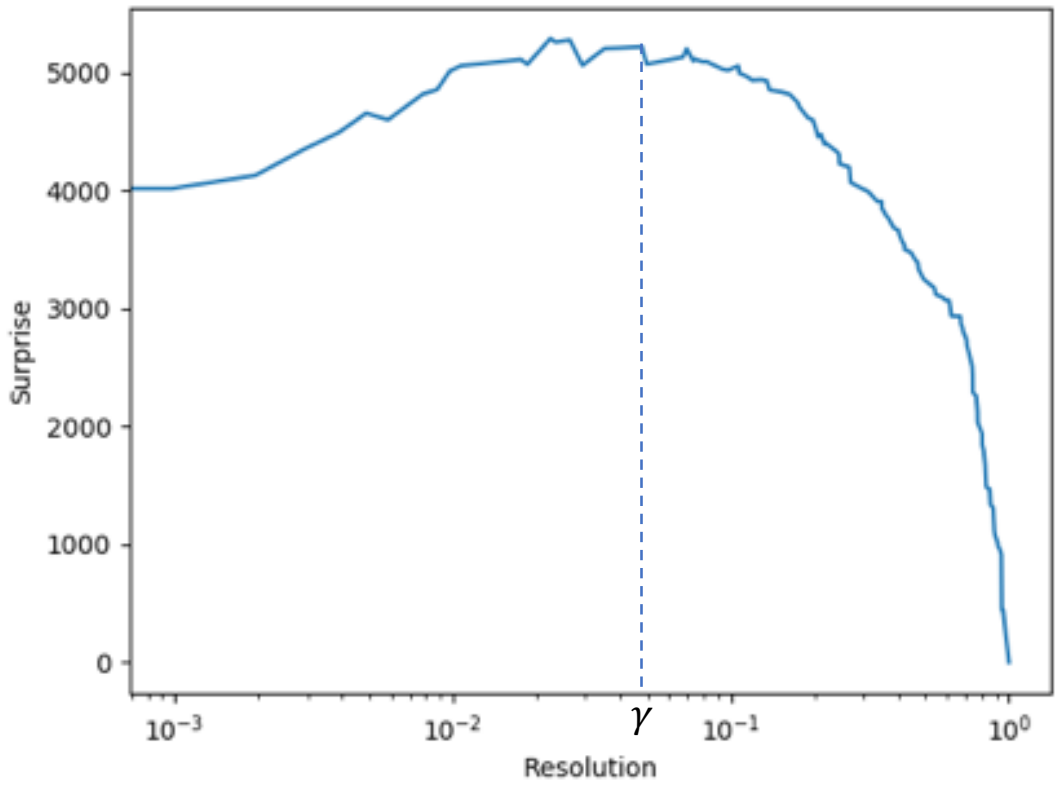


Figure 21 A resolution profile of the active module. The resolution parameter was scanned in the range of 0 to 1 for which Surprise is maximum for the active module. This resulted in γ = 0.0479. The resolution axis (x-axis) is in log scale.

The observations from the analysed community structures were twofold:

1) In general, proteins which are upregulated in the Met1 (primary) adhesome network tend to be within the same community, while proteins which are upregulated in the Met4 (metastatic) adhesome network tend to be in separate communities (red nodes cluster with red nodes, blue nodes cluster with blue nodes, Figure 24). Co-regulated proteins within the same community have been postulated to participate in the same molecular pathway. Thus, this observation

is indicative of differentially regulated molecular processes at the adhesome level in the SCC cells.

2) Different community detection algorithms produced diverse communities of varying sizes. This makes it hard to determine the true community structure. We could not find a consensus community in the SCC adhesome network, which hampered the subsequent identification of the differentially regulated molecular processes by gene ontology enrichment analysis.

Once the community structures are defined in a network, the nodes can be categorised in the z-P parameter space (Figure 22) In the z-P parameter space, the nodes are classified by virtue of the number of binding partners within their communities and the number of binding partners outside their communities. It has been postulated that nodes with similar role within the community would share the same region in the z-P parameter space[94,103].

Of interest is the identification of kinless (R7 in Figure 22) and connector (R6 in Figure 22) hubs. These hubs interact with nodes in other communities. Connector hubs have many links to most other communities, while kinless hubs have links homogeneously distributed among all other communities in the network. In protein-protein interaction networks, connector and kinless hub proteins link different molecular processes and perform different functions under different conditions[99-101]. Therefore, the kinless and

connector hubs in the adhesome SCCs network are proteins that may 'switch' interactions to proteins regulating different molecular processes, and thus perform different functions under primary and metastatic conditions. The connector and kinless hubs are labelled by their gene name in the z-P parameter space (Figure 23) and in the active module with respect to their communities (Figure 24).

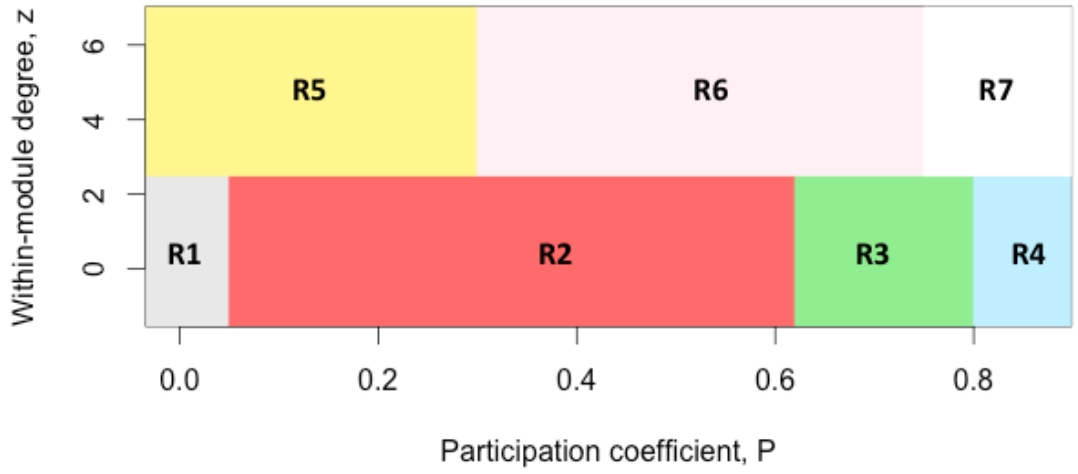
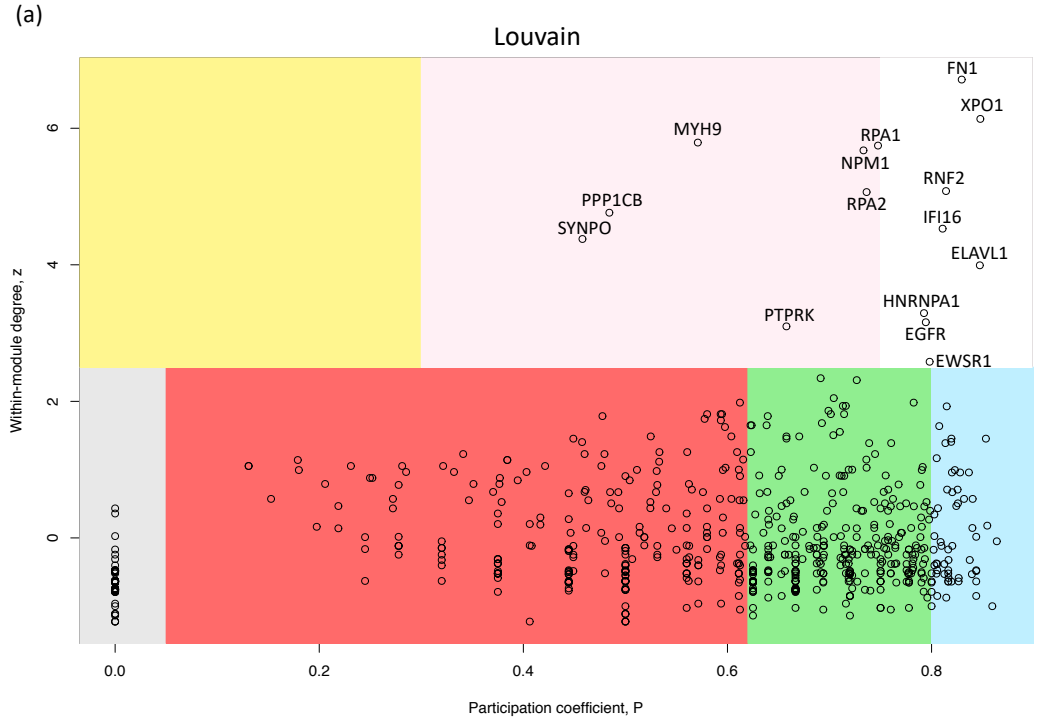
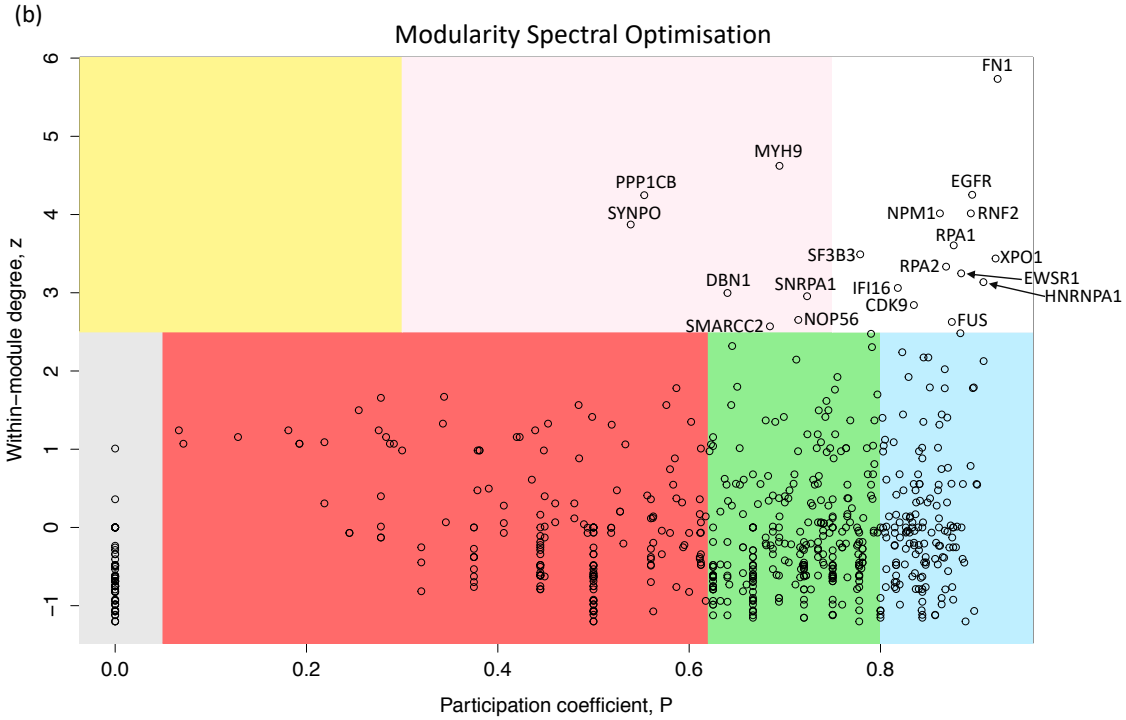


Figure 22 The z-P parameter space divided into regions defining the roles of nodes with respect to the community structures. Non-hubs are categorised into ultraperipheral (R1), peripheral (R2), connector (R3) and kinless (R4) nodes. Hubs are categorised into provincial (R5), connector (R6) and kinless (R7) hubs.

(a)


Louvain
FN1
XPO1
MYH9
RPA1
NPM1
RNF2
PPP1CB
RPA2
IFI16
SYNPO
ELAVL1
HNRNPA1
PTPRK
EGFR
EWSR1
Within-module degree, z
Participation coefficient, P
0.0
0.2
0.4
0.6
0.8
0
2
4
6

(b)


Modularity Spectral Optimisation
FN1
MYH9
PPP1CB
EGFR
SYNPO
NPM1
RNF2
RPA1
SF3B3
XPO1
RPA2
EWSR1
DBN1
SNRPA1
IFI16
HNRNPA1
CDK9
SMARCC2
NOP56
FUS
Within−module degree, z
Participation coefficient, P
0.0
0.2
0.4
0.6
0.8
−1
0
1
2
3
4
5
6

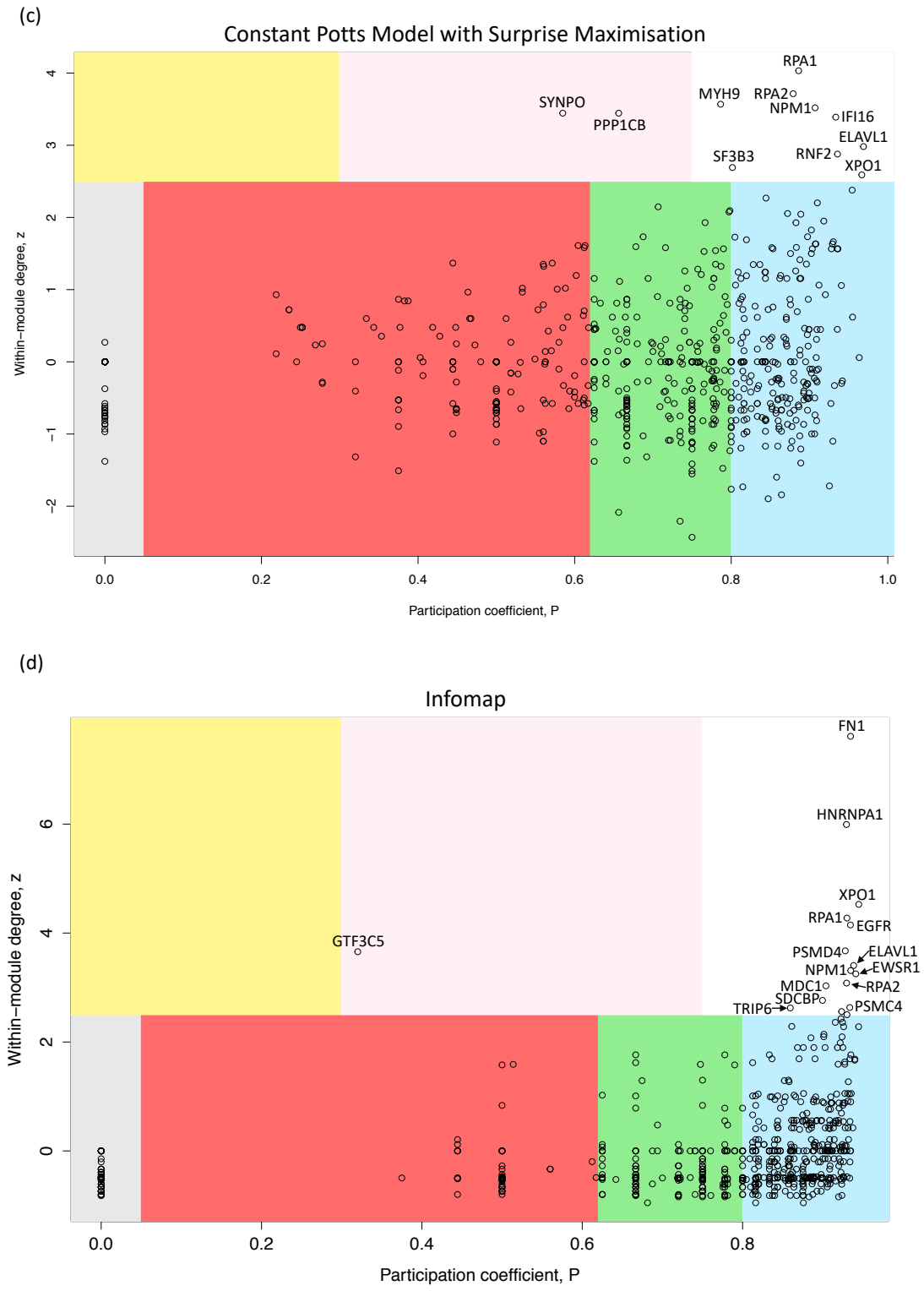
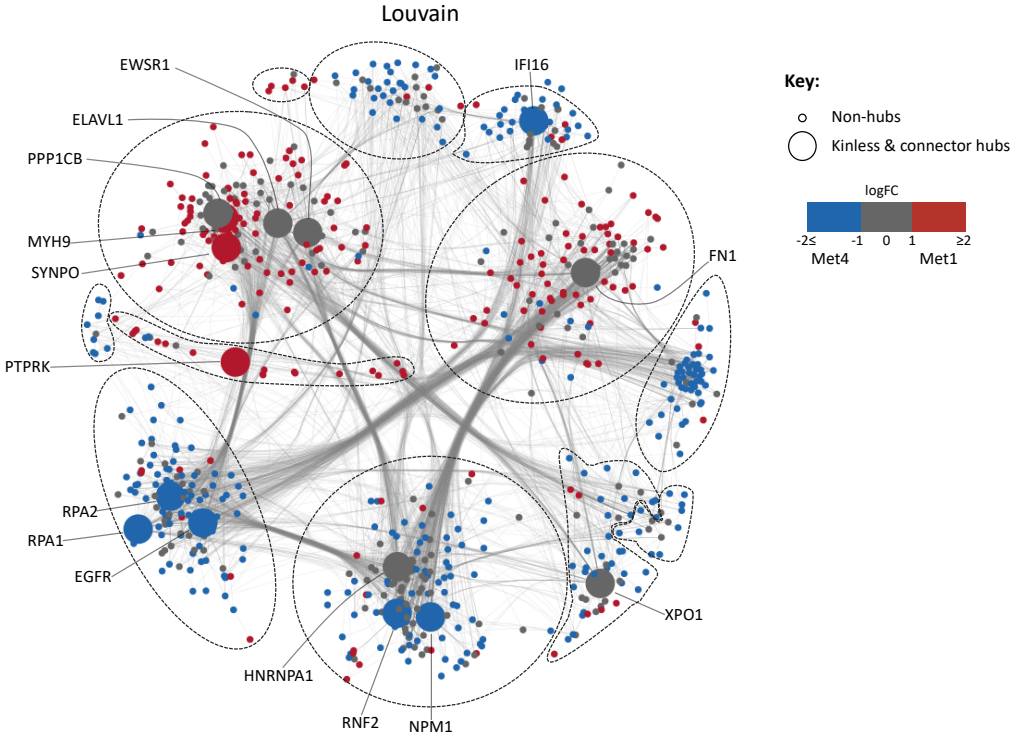
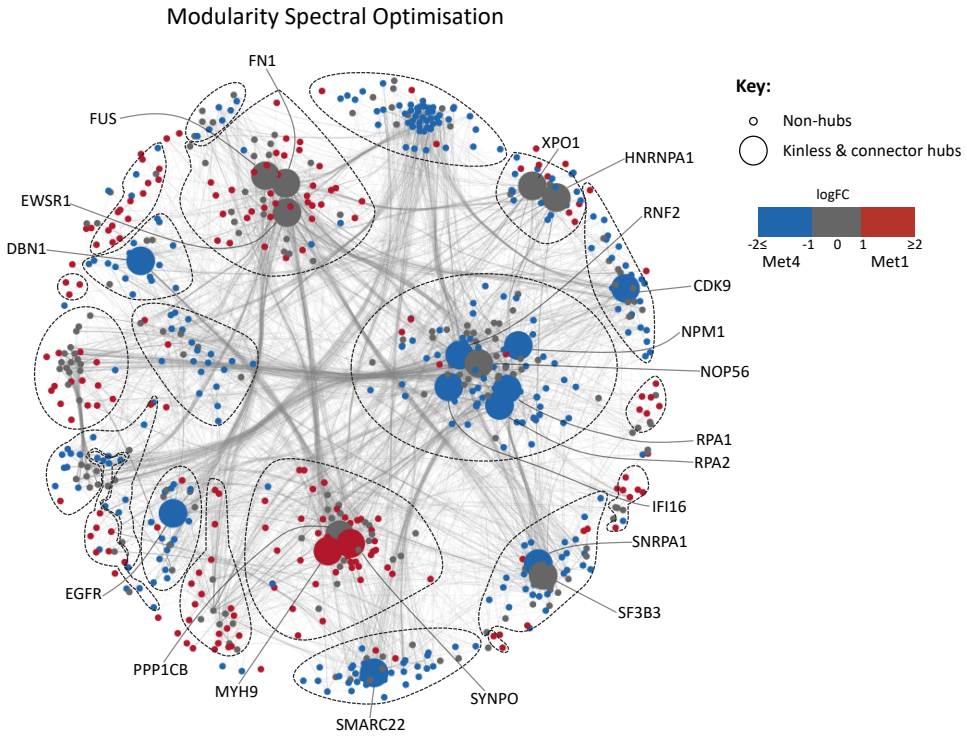


Figure 23. Location of nodes within the zP parameter space when the active module is partitioned into communities by (a) Louvain method, (b) modularity spectral optimisation, (c) Constant Potts model with Surprise maximisation and, (d) Infomap. The hubs are denoted by their gene name.

(a)


Louvain
EWSR1
IFI16
ELAVL1
PPP1CB
MYH9
SYNPO
FN1
PTPRK
RPA2
RPA1
EGFR
XPO1
HNRNPA1
RNF2
NPM1
Key:
Non-hubs
Kinless & connector hubs
logFC
-2≤ -1 0 1 ≥2
Met4
Met1

(b)


Modularity Spectral Optimisation
FN1
FUS
XPO1
HNRNPA1
EWSR1
RNF2
DBN1
CDK9
NPM1
NOP56
RPA1
RPA2
IFI16
SNRPA1
EGFR
SF3B3
PPP1CB
MYH9
SYNPO
SMARC22
Key:
Non-hubs
Kinless & connector hubs
logFC
-2≤ -1 0 1 ≥2
Met4
Met1

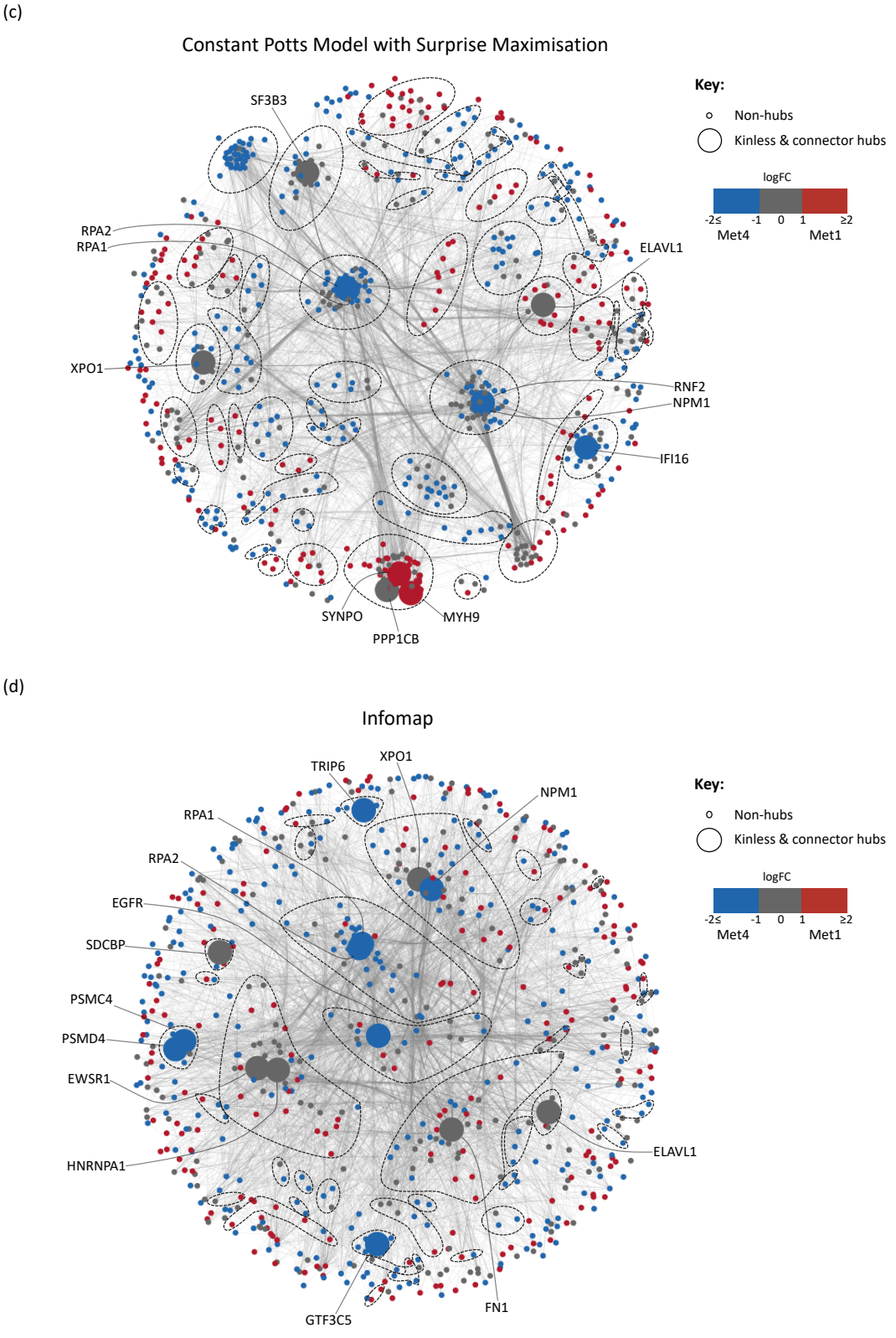


Figure 24 Active module partitioned by (a) Louvain method, (b) modularity spectral optimisation approach, (c) Constant Potts model with Surprise maximisation approach and (d) Infomap method. The detected communities are enclosed within the dashed

lines. Kinless and connector hubs are denoted by their gene name. Blue nodes represent adhesome proteins which have at least two-fold upregulation in Met4 (metastatic). Red nodes represent adhesome proteins which have at least two-fold upregulation in Met1 (primary). Grey nodes represent adhesome proteins which are not differentially regulated but are essential for the connectivity between the differentially regulated proteins.

Date and party hubs dichotomy

As mentioned above, Han *et al.*[98] reported two classes of hubs based on the average expression levels between the hub and its interacting partners. Party hubs were highly co-expressed with their neighbours and thus coordinate proteins within the same community, whereas date hubs have low co-expression with their interacting partners and may therefore interact with different binding partners to perform different functions under different conditions and at different locations[99-101].

To test whether the connector and kinless hubs in the active module are date hubs, the average Pearson correlation coefficient (avPCC) of the protein abundance between the hubs and their interacting partners were computed. The avPCC followed a multimodal distribution with three distinct population of hubs with increasing avPCC values (Figure 25). Nevertheless, the dichotomy of party and date hubs was maintained here with an avPCC of 0.5 used as the boundary (Figure 25, black arrow) to distinguish between date hubs and party hubs. A high proportion (85.7%) of kinless and connector hubs had low avPCC and hence were assigned as date hubs. A low

proportion (14.3%) of kinless and connector hubs had high avPCC and hence were assigned as party hubs.

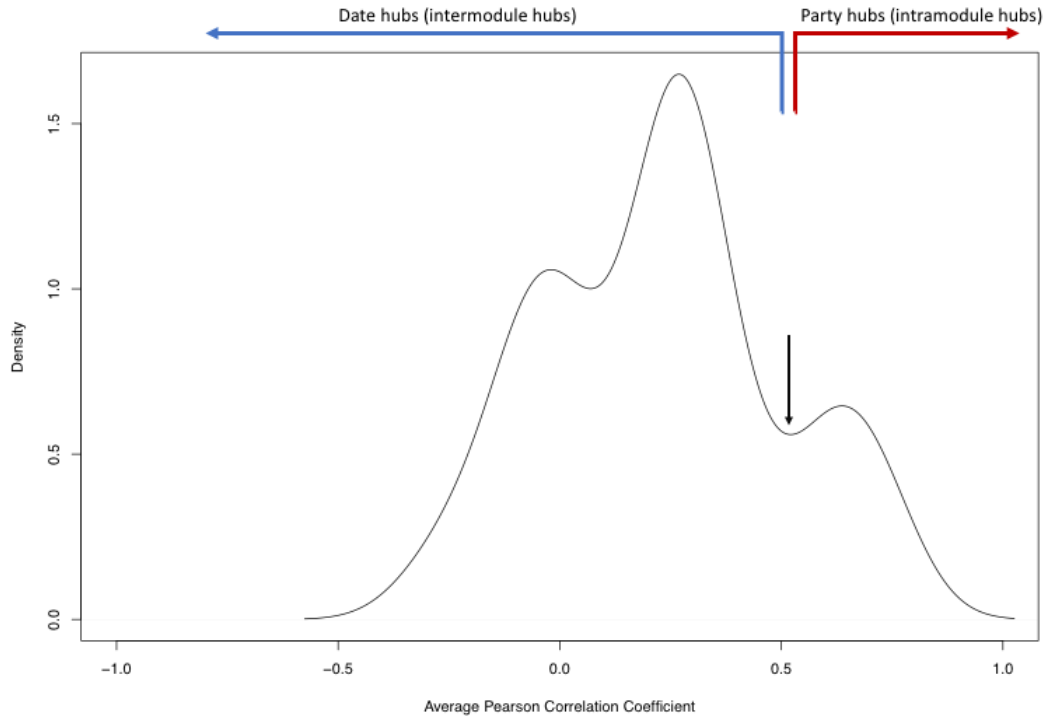


Figure 25 Distribution density of the avPCC between the kinless and connector hubs and their interacting partners in the active module. avPCC shows a multimodal distribution with avPCC = 0.5 (black arrow) used to separate the date (blue arrow) and party (red arrow) hubs.

The classification of connector and kinless hubs relies on the boundaries of the z-P parameters. Though these boundaries are based on mathematical arguments and validated empirically using real networks [94,103], it is plausible that these boundaries may not apply to the active module above. Hence, in order to validate those hubs, another class of method, namely, a direct method approach, was used to determine the importance of the nodes in the adhesome SCCs network. Here, the importance of a node not only depends

on its structural role within the network but also by its 'influence' or 'support' from its neighbours. One analogy is the citation of scientific papers. Influential scientific articles would be cited numerous times or given 'support' by other articles. However, an article with few citations may also be influential if it is cited by other influential articles. Using the same principles, we used the PageRank algorithm[151] to rank the importance of each protein in the active module. PageRank is the algorithm used by the Google search engine to determine the relevance of webpages and has been applied elsewhere to rank proteins[151]. PageRank determines the importance of a node not only by the quantity of links (degree centrality) but also by the quality of the neighbouring nodes[126]. In protein-protein interaction networks, influential nodes linking highly interconnected proteins will be highly scored.

As opposed to using positive weighted edges for the community detection algorithms in the z-P analysis, PageRank is compatible with negative weighted links. To rank the adhesome proteins by their influence and connectivity, the PageRank score was plotted against the degree connectivity. This approach is similar to that used by Reuter *et al.*[152]

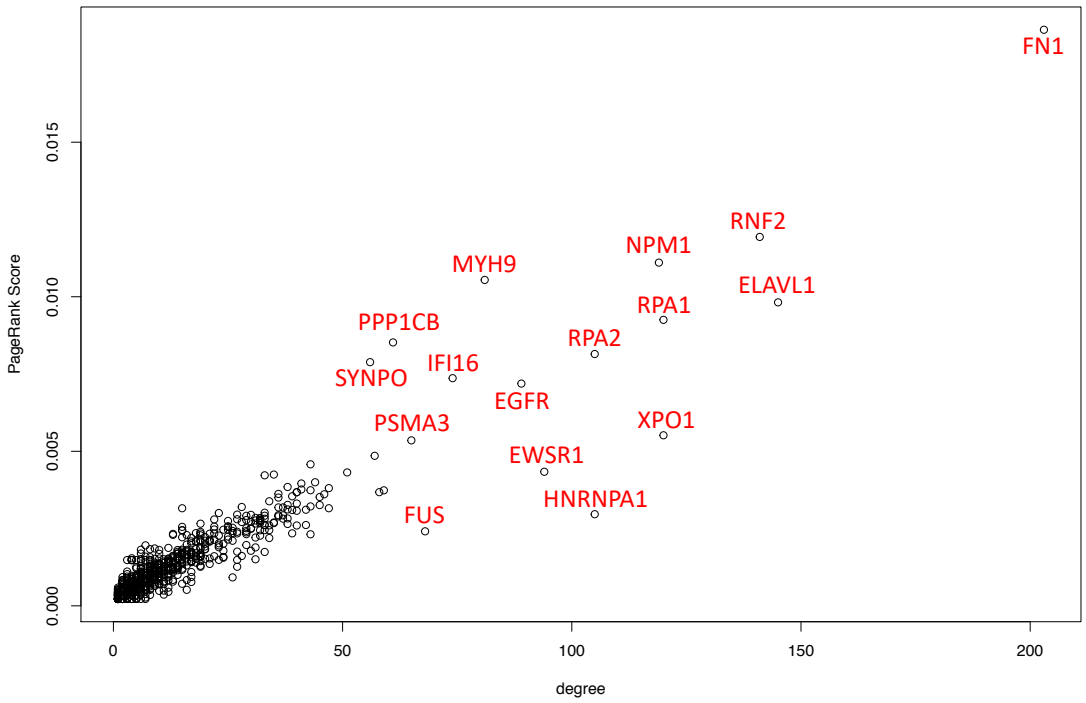


Figure 26 Influence and connectivity plot of the 720 adhesome proteins in the active module.

Fibronectin (FN1), an extracellular matrix protein, was ranked top — as presented in Figure 26 — out of the 720 nodes in the active module. Moreover, fibronectin was also found to be a kinless hub in the community-assisted approaches (Figure 23(a), (b), (d) & Figure 24(a), (b), (d)). Being a kinless hub indicates that fibronectin has links homogeneously distributed to other communities in the active module, thus suggesting that fibronectin may induce different molecular processes in Met1 and Met4 cells. Being highly scored by the PageRank-degree analysis confirms that fibronectin links highly interconnected protein clusters with other highly interconnected protein clusters. This is expected since Met1 and Met4 cells were seeded on

fibronectin, an extracellular matrix protein, to induce the formation of the adhesome. Taken altogether, it is plausible that fibronectin may induce different molecular pathways during cancer progression. This hypothesis has yet to be experimentally validated.

Consensus vital hubs.

Approximately half (54%) of the kinless and connector hubs determined in the z-P analysis were also highly ranked by PageRank-degree as shown in Figure 26. The consensus adhesome SCCs hubs (CASH) are hubs which have been determined by both the z-P analysis (community-assisted approach) and PageRank-degree analysis (direct approach) and which occurred in at least three datasets in the meta-adhesome database. These criteria resulted in the identification of eight CASH hubs, namely exportin-1 (XPO1), replication protein A1 (RPA1), fibronectin (FN1), ELAV-like protein 1 (ELAVL1), heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1), Ewing sarcoma breakpoint region 1 protein (EWSR1), myosin-9 (MYH9) and RNA-binding protein fused in sarcoma (FUS). Figure 27 shows the frequency by which the CASH proteins were identified by the different community detection algorithms. All CASH behaved as date hubs, having low avPCC (avPCC < 0.5) with their neighbouring proteins.

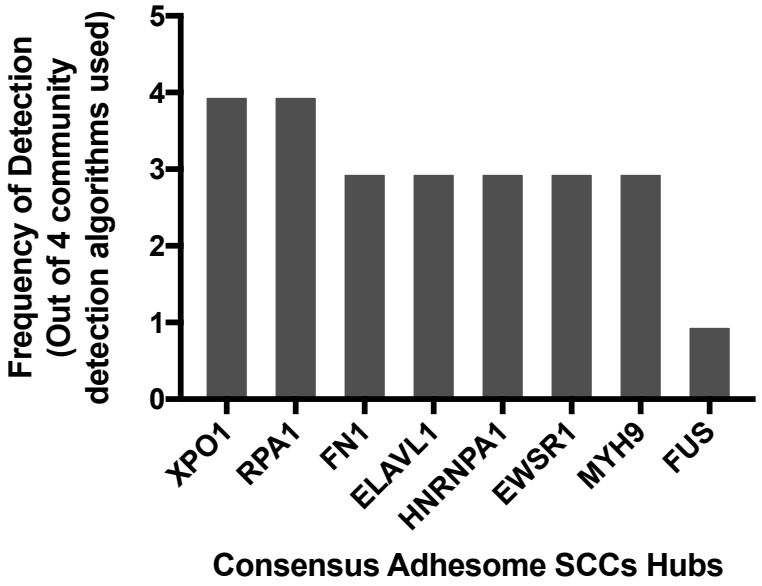


Figure 27 Bar chart showing the number of times the consensus adhesome SCCs hubs (CASH, x-axis) were defined as hubs using the community-assisted approach. XPO1 and RPA1 were the most frequently detected as hubs by the different community detection algorithms.

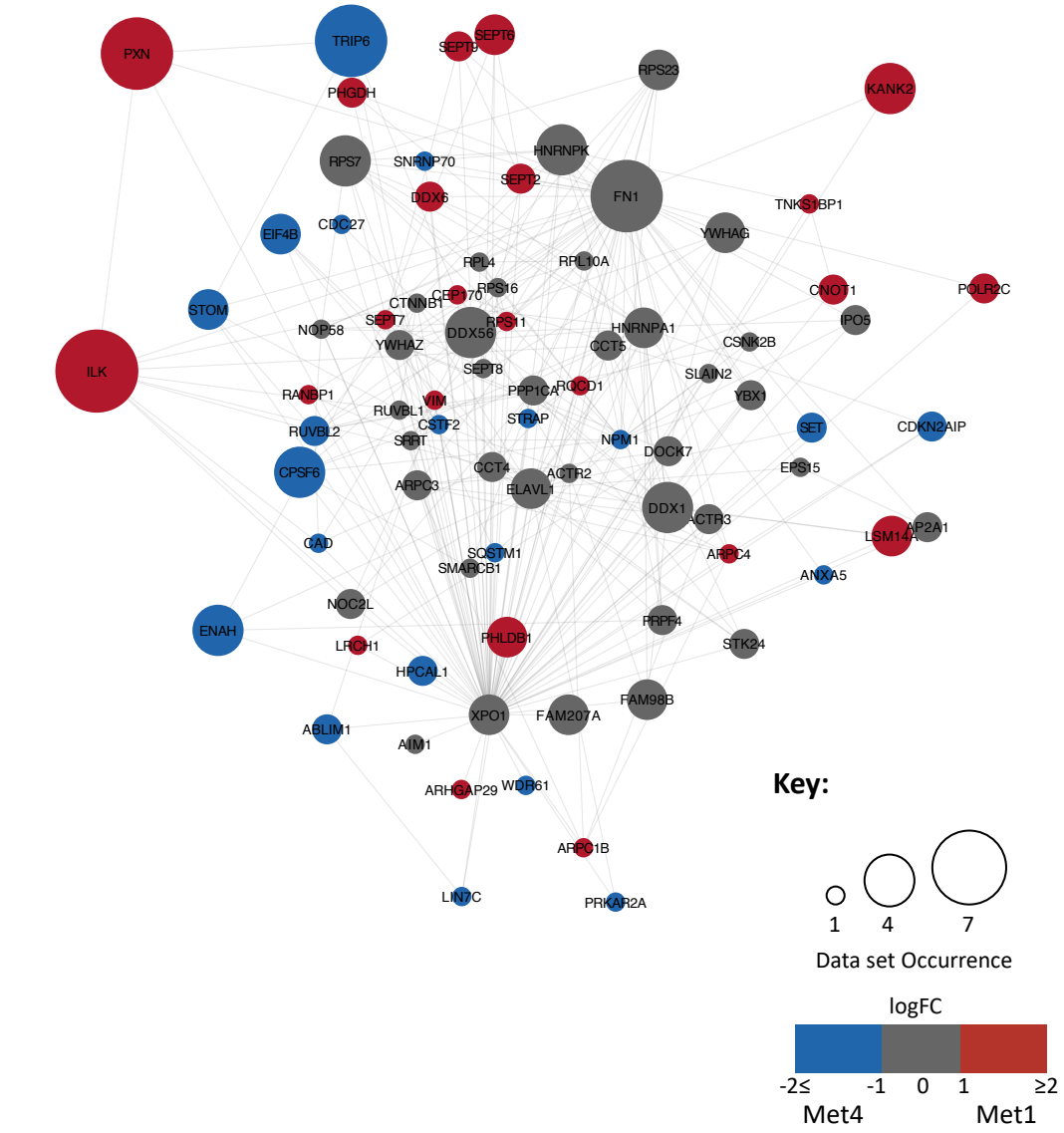
RPA1 and XPO1 were the most frequently identified hubs by the community detections algorithms (Figure 27). Replication protein A1 (RPA1) binds to single strand DNA to mediate DNA repair, DNA replication and homologous recombination[153,154]. Exportin-1 (XPO1 also known as CRM1) binds to protein cargoes with a nuclear export sequence (NES) and mediates their nuclear export from the nucleus to the cytoplasm[155].

Both RPA1 (avPCC = 0.240) and XPO1 (avPCC = 0.034) behave as date hubs which suggests that XPO1 and RPA1 may bind to different proteins in primary and metastatic cells (Met1 and Met4 respectively). RPA1 and XPO1 putatively interact with adhesome proteins which are upregulated in Met1

and in Met4 as well as with adheseome proteins whose abundance did not change. It has been reported that interactors of date hubs tend to be localised in multiple subcellular compartments within the cell[98]. Since both XPO1 and RPA1 behave as date hubs, we expect interactors of XPO1 (Figure 28(a)) and RPA1 (Figure 28(b)) to be in multiple subcellular compartments within the cell. The calculation of the localization entropy[98] of the interactors is required to test this hypothesis. This will be carried out in future work.

The primary function of exportin-1 is to mediate the nuclear export of proteins to the cytoplasm. Interestingly, interactors of XPO1 in the active module consist of canonical focal adhesion components, such as paxillin (PXN), integrin-linked kinase (ILK), mammalian enabled homologue (Mena, gene name: ENAH) and zyxin-related protein 1 (TRIP6). This suggests that those proteins may shuttle to the nucleus and indeed paxillin, ILK and TRIP6 were previously reported to shuttle to the nucleus[156-162], although this not been reported for ENAH.

(a)


PXN
TRIP6
SEPT6
SEPT9
RPS23
KANK2
PHGDH
HNRNPK
RPS7
SNRNP70
SEPT2
FN1
TNKS1BP1
DDX6
EIF4B
CDC27
YWHAG
RPL4
RPL10A
CNOT1
POLR2C
RPS16
CEP170
STOM
CTNNB1
IPO5
SEPT7
RPS11
HNRNPA1
NCP58
DDX56
CCT5
CSNK2B
YWHAZ
ILK
SEPT8
SLAIN2
RANBP1
PPP1CA
RQCD1
YBX1
VIM
RUVBL1
SET
CDKN2AIP
CSTF2
STRAP
RUVBL2
SFRT
NPM1
DOCK7
CPSF6
CCT4
ACTR2
EPS15
ARPC3
ELAVL1
DDX1
ACTR3
AP2A1
LSM14A
CAD
SQSTM1
ARPC4
SMARCB1
ANXA5
NOC2L
PRPF4
ENAH
PHLDB1
LRCH1
STK24
HPCAL1
FAM98B
XPO1
FAM207A
ABLIM1
AIM1
ARHGAP29
WDR61
ARPC1B
LIN7C
PRKAR2A
Key:
1
4
7
Data set Occurrence
logFC
-2≤
-1
0
1
≥2
Met4
Met1

(b)

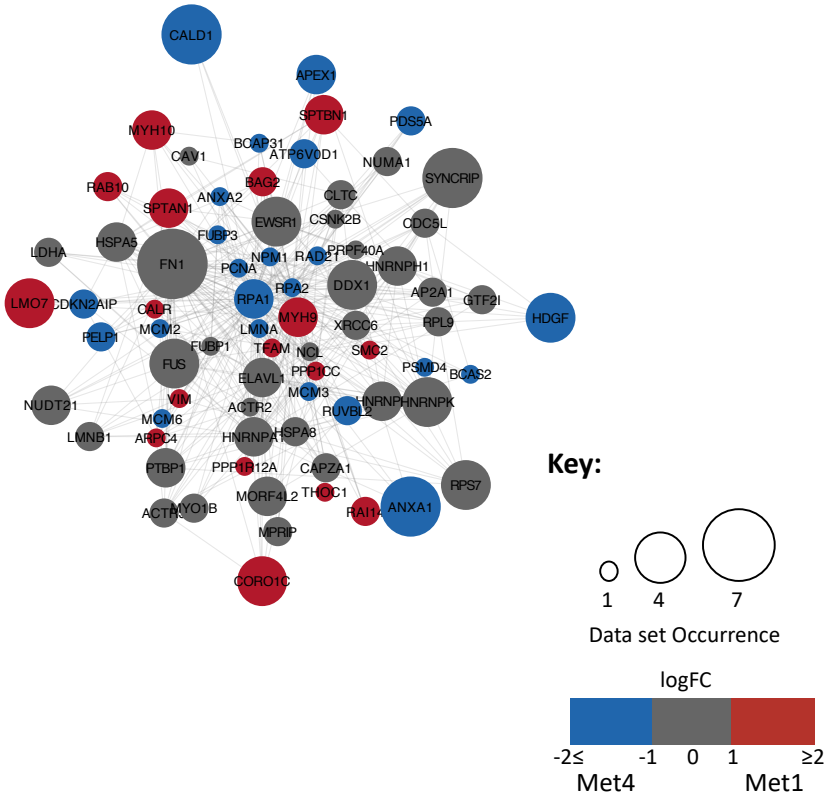


Figure 28 Interactors of (a) XPO1 and (b) RPA1 in the active module filtered for meta-adhesome proteins. The size of the node represents the occurrence of the protein in datasets of the meta-adhesome. Red nodes represent adhesome proteins that have at least two-fold upregulation in Met1 (primary). Blue nodes represent adhesome proteins that have at least two-fold upregulation in Met4 (metastatic). Grey nodes represent adhesome proteins that are not differentially regulated.

■ Discussion

In chapter 3, we previously isolated adhesome proteins from primary (Met1) and metastatic (Met4) patient-derived cutaneous SCC cells. Three biological replicates of the isolated adhesome for Met1 and Met4 were analysed by LC-MS/MS. This resulted in the identification of 1727 proteins which was interrogated using a network approach. The aim was to understand how the

adhesome changes during cancer progression and subsequently identify adhesome proteins that may be crucial to cancer progression.

The changes in the adhesome between primary (Met1) and metastatic (Met4) SCC cells were captured by the active module, which is a network of differentially regulated adhesome proteins and how they are connected to one another. Interestingly, the active module has a community structure (modular structure) with coregulated adhesome proteins clustering among themselves. Indeed, four separate community detection algorithms (Louvain, spectral modularity optimisation, CPM with Surprise maximisation and Infomap methods) produced clusters of upregulated adhesome proteins in Met1 which were distinct from clusters of upregulated adhesome proteins in Met4. This community structure is indicative of differentially regulated molecular processes between the Met1 and Met4 at the adhesome level.

However, different community detection algorithms all produced different communities with different sizes (a resolution problem), which hampered the subsequent gene ontology enrichment analysis to identify the differentially regulated molecular processes in Met1 and Met4. Even for algorithms which seek communities based on similar ideas, the difference is the way they do it[144]. For instance, both the Louvain method and the spectral modularity optimisation method seek to optimise the same quality function (i.e Girvan

Newman modularity) but they do it differently. The Louvain method uses an agglomerative approach, while the spectral modularity optimisation method uses linear algebra to find community structures. The specific procedures affect the reliability of the results as exemplified by the resolution problem. It is true that one can assess the accuracy of the communities using an artificial network models with known community structures (artificial benchmark), however, artificial benchmarks do not properly describe real networks[144] (the active module is an example of a real network). Thus, community detection methods that perform well on artificial benchmarks are not guaranteed to provide reliable solutions when applied to real networks. In addition to the resolution problem, real networks have specific features for which the methods used cannot accommodate. For instance, the active module may have overlapping communities, but the methods used here will only partition the active module into non-overlapping communities. However, we circumvent this by looking at the structural roles of the proteins in the communities, which may indicate the proteins (kinless proteins) that are likely to be in overlapping communities. Moreover, the community detection methods used here have the advantage of allowing for communities to be observed despite the interaction network being incomplete. This is because the methods used integrate the dynamics of the adhesome (e.g. the correlation between protein abundance) with the topology of the network.

Once community structures were defined, we proceeded with the identification of hubs in the active module. We hypothesize that the kinless and connector hubs in the active module rewire to different proteins, leading to metastasis. Moreover, those hubs may be involved in the different molecular processes involved during cancer progression. Here, we used the community-assisted approach as well as the direct approach to find hubs. Despite the identification of hubs being method dependent, we reasoned that hubs identified by multiple algorithms and by different approaches are more robust. The community-assisted approach finds hubs according to their structural role with respect to the community structure. The direct approach finds hubs based on properties from neighbouring nodes (e.g. PageRank). Thus, hubs which were determined by both the direct and community-assisted approaches and occurred at least three times in the meta-adhesome dataset were termed as consensus adhesome SCCs hubs (CASH).

One such CASH was fibronectin, which was ranked first in the active module using the direct approach suggesting that it links highly connected protein clusters together. This is unsurprising since Met1 and Met4 were seeded on fibronectin to induce the formation of the adhesome. Being a kinless and a date hub, we speculate that fibronectin may be inducing different molecular pathways in Met1 and Met4.

RPA1 and XPO1 were the most frequent CASH to be identified by the different community detection algorithms. Being kinless hubs suggests that RPA1 and XPO1 may be connecting different yet to be defined molecular processes during cancer progression. Moreover, RPA1 and XPO1 were date hubs suggesting that they may be binding different proteins in primary and metastatic SCC cells. We show that RPA1 and XPO1 have putative interactions with adhesome proteins that are upregulated in Met1 as well as with adhesome proteins that are upregulated in Met4. Interestingly, XPO1 putatively interacts with canonical focal adhesion components paxillin, Mena (ENAH), integrin linked kinase (ILK) and TRIP6. Whether those putative interactions change during cancer progression still needs to be determined. It is tempting to speculate that XPO1 interacts with paxillin and ILK in primary cells but rewires to Mena and TRIP6 during metastasis. Co-immunoprecipitation of XPO1 with Mena, ILK, TRIP6 and paxillin in Met1 and Met4 cells should validate this hypothesis. This will be carried out in future work.

It has been reported that interactors of date hubs are diverse in spatial localisation within the cell[98]. This is true for interactors of XPO1, which mediates the export of proteins from the nucleus to the cytoplasm. Thus, interactors of XPO1 would be in at least two compartments within the cell namely the nucleus and the adhesome (or at the very least the cytoplasm). Interestingly, canonical focal adhesion components paxillin, ILK and TRIP6

were previously described to be in the nucleus[156-162] , whereas Mena was not. Since proteins that are close to one another in the interactome network have been shown to share similar functions[163-165], the association of Mena with XPO1 suggests that Mena may also shuttle to the nucleus. This hypothesis was investigated in chapter 5.

CHAPTER 5

Results: From Bioinformatics to Bench: Mena has a nuclear function in Metastatic cells

■ Introduction

In chapter 4, we identified exportin-1 (XPO1) as a hub in the adhesome network of patient-derived squamous cell carcinoma (SCC) cells that may interact with different molecular processes during cancer progression. Interestingly, network analysis of the SCC adhesome revealed a putative interaction between XPO1 and Mena, which is part of the actin-regulatory Ena/VASP proteins family. The Ena/VASP proteins family consists of Evl, VASP and Drosophila Enabled (Ena; Mena being the mammalian orthologue of Ena). The members of the Ena/VASP protein family share three conserved domains: an EVH1 domain at the N-terminus, an EVH2 domain at the C-terminus and a proline-rich region in-between the EVH1 and EVH2 domains. An additional domain consisting of LERER repeats is unique to Mena.

Three isoforms, namely hMena, hMena11a and hMenaΔv6 have been characterised in human cancer cells. hMena is the 'consensus' human homologue of murine Mena and shares 87% identity with the latter[49]. hMena consists of 570 amino acids encoded by 14 exons[49]. The hMena11a variant has an extra exon of 63 nucleotides between exon 11 and exon 12, resulting in the addition of 21 amino acids in the EVH2 domain. hMena11a is associated with epithelial phenotypes[52] and p42/p44 mitogenic activity[51]. However, hMena11a suppresses invasion *in vivo*[48]. The hMenaΔv6 variant

lacks exon 6, which encodes for 37 amino acids between the LERER repeat domain and the proline-rich domain. hMenaΔv6 expression is associated with mesenchymal markers such as N-cadherin and vimentin expression[52]. Cells transfected with hMenaΔv6 has an increased invasiveness through Matrigel[52]. In addition, MenaINV is a murine isoform of Mena containing an INV exon (previously known as the +++ exon), which confers a 19-amino acid residue insertion just after the EVH1 domain. MenaINV was found to be expressed in human invasive ductal carcinomas[50], but all of MenaINV functional studies were performed on murine models.

Mena expression is clinically associated with invasive cancers, regardless of the different hMena isoforms. Immunohistological staining for pan-hMena was positive in infiltrating ductal carcinomas of the breast[52], and in metastatic breast cancer[53]. Mena overexpression was also reported in metastases of breast cancer (lymph nodes, lung, bone, brain, pleura)[53]. hMena was overexpressed in colorectal adenomatous polyps with high dysplasia[54] and invasive cervical squamous cell carcinomas[56]. Moreover, pan-hMena expression was positively correlated with tumour size, proliferation index as evaluated by Ki67 staining, Her2 overexpression and cancer stage in breast cancer[53]. In colorectal carcinoma, Mena expression correlated with vascular invasion, tumour budding, tumour stage, Her-2 and p53 expression[54,55,166]. Likewise, Mena expression was correlated with tumour grade of cervical intraepithelial lesion[56].

Numerous studies have shown that MenaINV confers metastatic and motility properties *in vivo* and *in vitro*. Rat mammary adenocarcinoma cells expressing MenaINV were reported to be more sensitized to epidermal growth factors (EGF), thereby promoting chemotaxis at extremely low concentration of EGF[167]. MenaINV also promoted multicellular streaming *in vivo,* which is dependent on a paracrine signalling loop with macrophages[48]. Moreover, MenaINV-expressing cells have a higher ability to intravasate, and the presence of macrophages potentiates these cells to undergo transendothelial migration[48,57]. Consistent with these findings, sections of human invasive ductal carcinomas of the breast with high immunohistological Mena staining were associated with perivascular macrophages[48].

Here, we report a novel function of Mena at the perinuclear region of metastatic human SCC cells.

■ Results

Mena is highly abundant in the nucleus of patient-derived Met4 SCC cells

XPO1 is a date hub (a hub that interacts with different binding partners under different conditions), which implies that interactors of XPO1 are diverse in spatial distribution within the cell. Since XPO1 is a mediator of protein

nuclear export, interactors of XPO1 in the active module of the patient-derived SCC cells adhesome network (the active module is the subnetwork of the adhesome network consisting of connected, differentially abundant adhesome proteins that may drive a metastatic phenotype) are expected to be in at least the nucleus and adhesion complexes. Interestingly, XPO1 interactors in the active module include the adhesome components paxillin, ILK, TRIP6 and Mena. Paxillin, ILK and TRIP6 were previously reported to be in the nucleus, whereas Mena (ENAH) was not. The association of Mena with XPO1 suggests Mena may also localise to the nucleus. To test this hypothesis, subcellular fractionation using a sucrose cushion was performed on patient-derived Met1 and Met4 SCC cells, and the nuclear and cytoplasmic fractions were probed for Mena (Figure 29). The abundance of Mena in the nucleus was higher in Met4 cells than Met1 cells. Furthermore, it was observed that Met4 cells express mainly a doublet Mena band, while Met1 cells express one Mena band. The observed doublet in Met4 cells could arise from serine/threonine de-phosphorylation of Mena[168] or from expression of different Mena isoforms as observed in metastatic cells[47,52]. The nature of the Mena doublet in Met4 cells is yet to be determined.

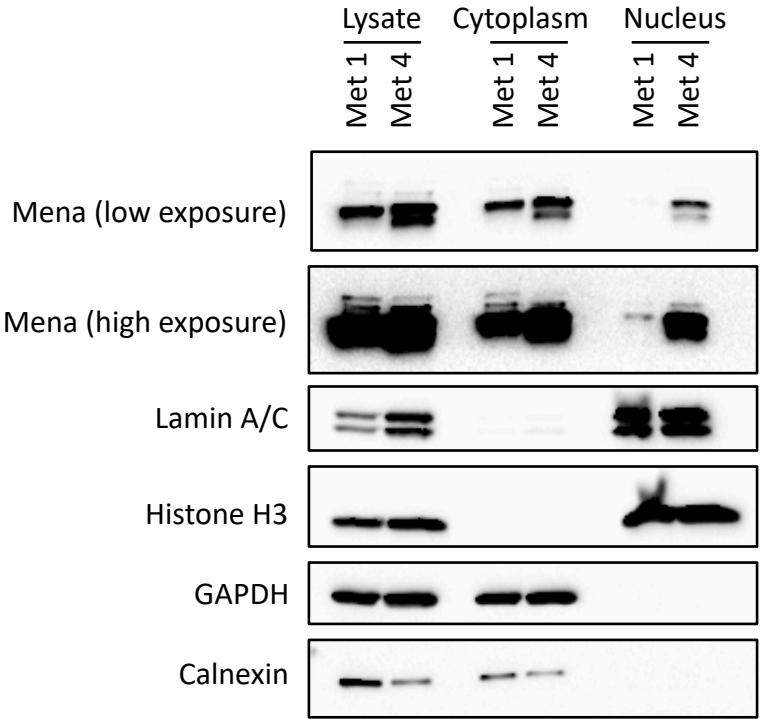


Figure 29 Western blot showing the total cell lysate, nuclear and cytoplasmic fractions of Met1 and Met4 after sucrose cushion fractionation. Mena was highly abundant in the nuclear fraction of Met4 compared to Met1. Lamin A/C and Histone H3 are nuclear markers, GAPDH is a cytosolic marker and calnexin is an endoplasmic reticulum marker. Western blot is representative of two experimental replicates.

Mena nuclear translocation is matrix dependent and FAK mediated

To induce adhesion complex formation for the analysis of the SCC adhesome (see Optimised Adhesome Isolation Conditions), Met1 and Met4 were seeded onto the extracellular matrix (ECM) protein fibronectin. To test whether Mena nuclear translocation was mediated by fibronectin-dependent adhesion, Met1 and Met4 cells were seeded on Matrigel-, fibronectin- and poly-D-lysine-coated plates for 4 hours followed by sucrose cushion fractionation. Matrigel is a mixture of ECM proteins, consisting mainly of

collagen IV and laminin. Matrigel and fibronectin both activate integrins thereby triggering a series of signalling events, though Matrigel and fibronectin might activate different sets of integrins. Cells seeded on PDL are attached by electrostatic interaction and thus integrin-mediated adhesion signalling is minimal.

In Met1 cells, the ECM composition had no effect on Mena nuclear abundance (Figure 30(a-b)). By contrast, nuclear Mena abundance in Met4 cells was reduced by approximately twofold when seeded on PDL (Figure 30(c-d)). This suggests that Mena nuclear translocation is dependent on integrin-mediated adhesion in Met4 cells but not in Met1 cells.

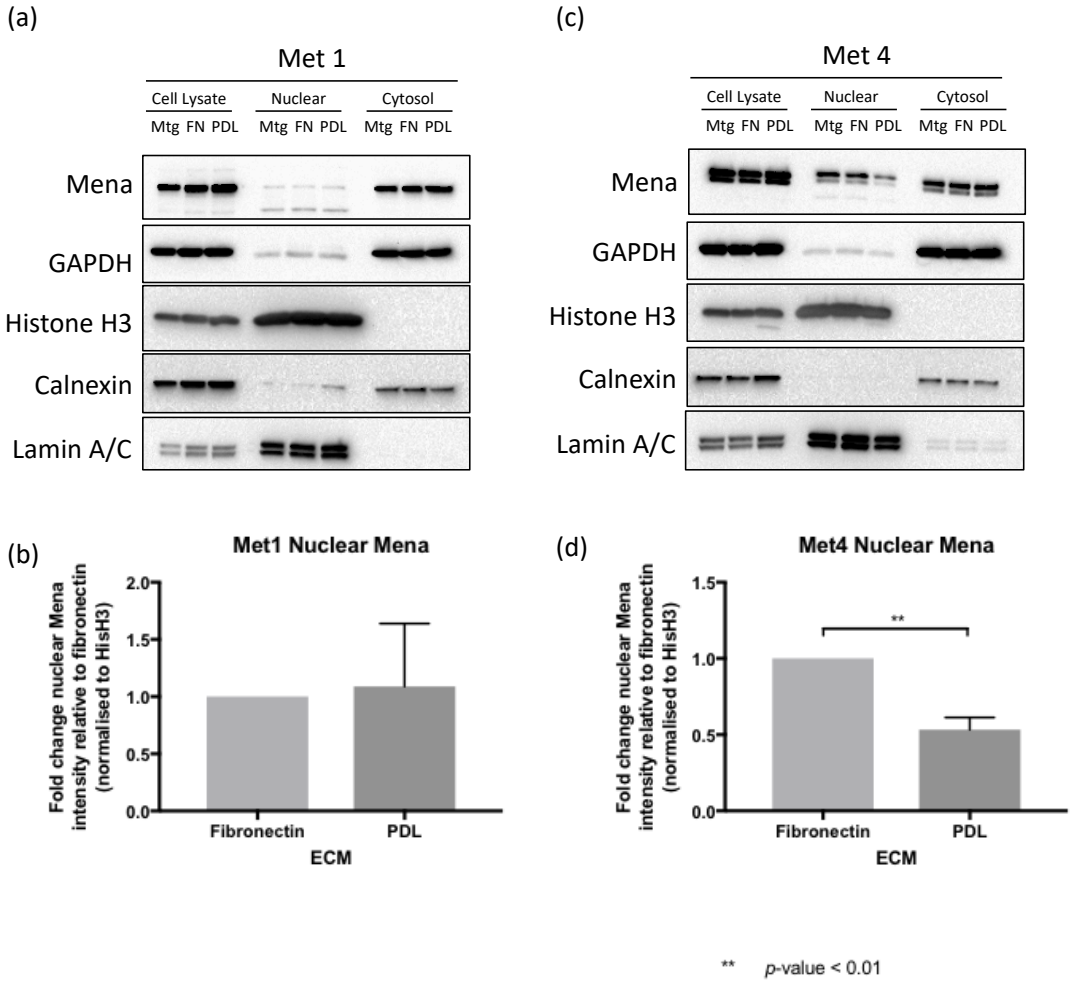


Figure 30 (a,c) Western blots showing the total cell lysate, nuclear and cytoplasmic fractions of (a) Met1 and (c) Met4 after seeding cells on Matrigel (Mtg), fibronectin (FN) and poly-D-lysine (PDL) followed by sucrose cushion fractionation. Lamin A/C and Histone H3 are nuclear markers, GAPDH is a cytosolic marker and calnexin is an endoplasmic reticulum marker. Western blots are representative of three independent experimental replicates. (b,d) Bar charts representing the quantification of the fractionation Western blot for nuclear Mena. (b) No statistically significant change was observed in the abundance of nuclear Mena in Met1 cells (n=3, Welch's t-test). (d) A two-fold reduction in nuclear Mena was observed when Met4 cells were seeded on poly-D-lysine as compared to fibronectin (n=3, Welch's t-test).

To understand how integrin-mediated adhesion is involved in nuclear Mena translocation, FAK was knocked out using CRISPR-Cas9 in Met1 and Met4 cells. FAK is activated upon integrin-mediated adhesion and mediates the assembly of focal adhesions[169]. Sequencing of the genomic FAK exon-2

region where the FAK guide RNA (gRNA) was targeted shows that hCas9 nuclease activity induced deletion three nucleotides upstream of the PAM sequence (currently, sequencing data is only available for Met1) (Figure 31(g)).

To test whether FAK could mediate nuclear translocation of Mena, Mena was probed in the nuclear fraction of Met1 and Met4 cells expressing endogenous FAK or with FAK knocked out. In Met1 cells, FAK knock-out increased nuclear Mena abundance, though the increase was not statistically significant (Figure 31(a,b)). Conversely, FAK knock-out in Met4 cells reduced nuclear Mena abundance by half (Figure 31(c,d)). This suggests that FAK mediates Mena nuclear localisation, albeit differently in Met1 and Met4 cells.

It was previously reported that FAK could mediate mechanosignalling through the FAK-Yap-mTor axis[170] and the FAK-Src-PI3K pathway[171]. To test whether FAK could mediate a similar process in the Met4 cells, we probed for nuclear Yap as a proxy for the mechanosignaling pathways in the nuclear fraction of Met1 and Met4 cells expressing endogenous FAK or with FAK knocked-out. No change in nuclear Yap was observed upon FAK knock-out in Met4 cells (Figure 31(f)). Conversely, in Met1 cells, FAK knock-out did result in a reduction of nuclear Yap (Figure 31(e)). This suggests that

Mena nuclear shuttling is unlikely to be mediated by the mechanosignalling axis of FAK.

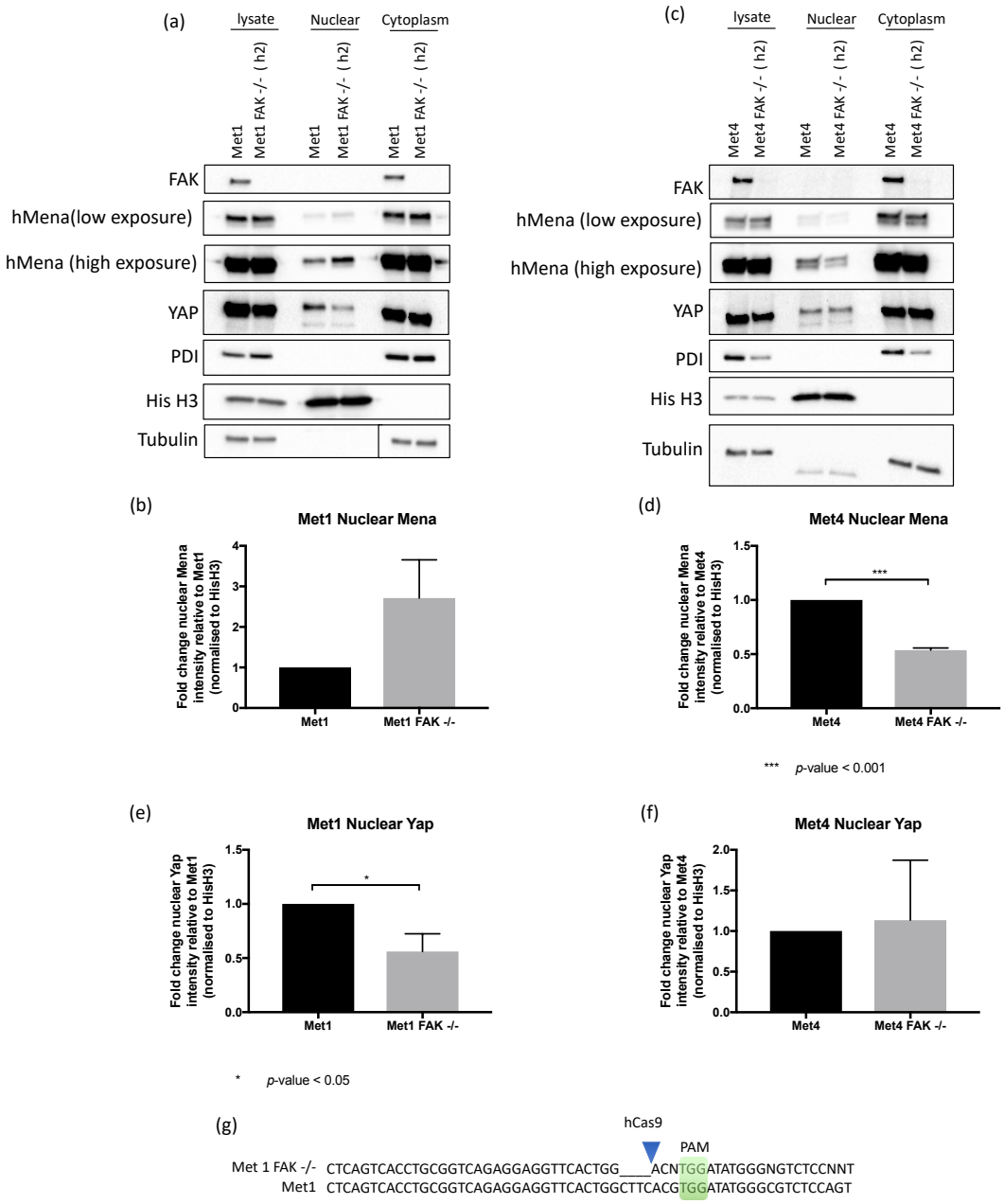


Figure 31 (a,c) Western blots showing the total cell lysate, nuclear and cytoplasmic fractions of (a) Met1 and Met1 FAK knock-out (Met1 FAK-/-) cells, and (c) Met4 and Met4 FAK knock-out (FAK -/-) cells. Histone H3 (His H3) is a nuclear marker, tubulin is a cytosolic marker and PDI is an endoplasmic reticulum marker. Western blots are

representative of three independent experimental replicates. (b,d,e,f) Bar charts representing the quantification of the fractionation western blots for (b,d) nuclear Mena and (e,f) nuclear Yap. (b) No statistically significant change was observed in the abundance of nuclear Mena in Met1 and Met1 FAK knock-out cells. (n=3, Welch's t-test). (e) However, FAK knock-out resulted in a two-fold reduction in nuclear Yap. (n=3, Welch's t-test). (d) A two-fold reduction in nuclear Mena was observed when FAK was depleted in Met4 cells (n=3, Welch's t-test). (f) No change in nuclear Yap was observed upon FAK knock out in Met4 cells (n=3, Welch's t-test). (g) Genomic sequence of FAK exon 2 showing the specificity of FAK gRNA-hCas9-mediated nuclease activity. As expected, the genome was cut three nucleotides upstream of the PAM sequence.

FAK knock-out cells were obtained using one FAK gRNA only. Thus, there is a possibility that the change in nuclear Mena could be due to non-specific targeting of the FAK gRNA to other proteins. Thus, further validation needs to be carried out by re-expressing FAK in the FAK knock-out cells and test whether the observed change in nuclear Mena abundance is reversed. Moreover, the mechanism by which FAK mediates Mena nuclear translocation is yet to be determined, though it is unlikely to be mediated by the mechanosignalling axis of FAK.

Mena export is mediated by XPO1

Since Mena was highly enriched in the nuclear fraction of Met4 cells and XPO1 was isolated as part of the SCC adhesome complex, we asked whether the nuclear export of Mena was XPO1 dependent. To test this, Met4 cells were treated with increasing concentrations (10nM and 20nM) of leptomycin-b (LMB), an inhibitor of XPO1, for 4 hours. This was followed by subcellular fractionation using a mild detergent. The nuclear and cytoplasmic

fraction were probed for Mena and p53, the latter being a positive control for LMB treatment[172].

Subcellular fractionation experiments showed increasing accumulation of nuclear Mena and p53 with increasing concentration of LMB (Figure 32). The increase in both nuclear p53 and Mena was statistically significant only when treated with 20nM LMB. This suggests that XPO1 mediates the export of Mena from the nucleus.

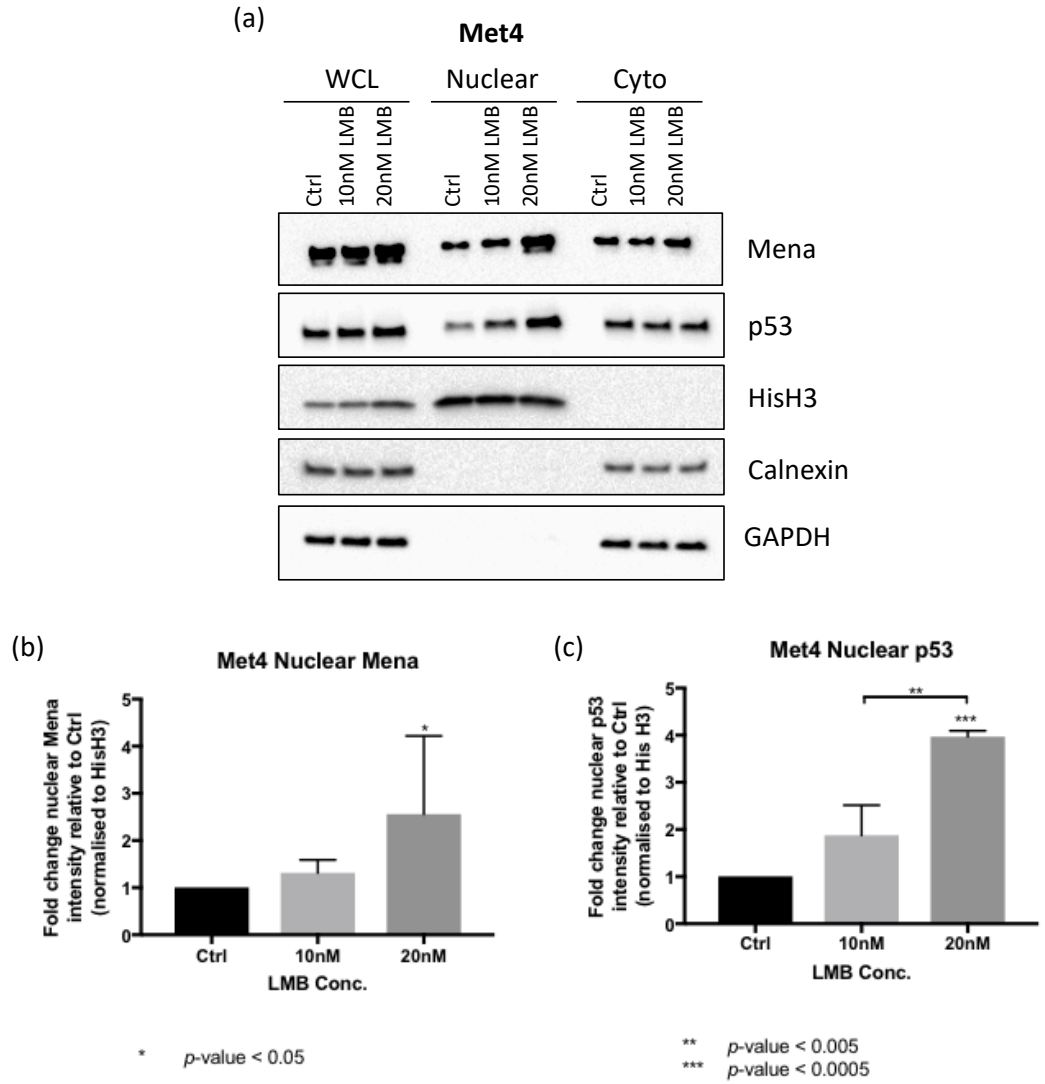


Figure 32 Western blot showing the whole cell lysate (WCL), nuclear and cytoplasmic (Cyto) fractions of (a) Met4 cells treated with ethanol (control), 10nM and 20nM leptomycin-b (LMB) for 4 hours followed by sucrose cushion fractionation. Histone H3 (HisH3) is a nuclear marker, GAPDH is a cytosolic marker and calnexin is an endoplasmic reticulum marker. Western blots are representative of three independent experimental replicates. (b,c) Bar charts representing the quantification of (b) nuclear Mena and (c) nuclear p53. (b) No statistically significant change was observed in the abundance of nuclear Mena when treated with 10nM LMB but a ≈2.5-fold increase in nuclear Mena was observed when Met4 cells were treated with 20nM LMB (n=3, one-way ANOVA). (e) As expected, nuclear p53 increases with increasing concentration of LMB (n=3, one-way ANOVA).

In addition to the western blot analysis above, LMB-treated and non-treated

Met4 cells were immunostained for Mena, paxillin and laminB1 and imaged

using spinning-disck confocal microscopy (SDCM). Paxillin served as a positive control since it was previously observed that LMB treatment results in paxillin accumulation in the nucleus of mouse embryonic fibroblast cells[173].

An enrichment of Mena and paxillin (the latter as a positive control) in the nucleus was observed by immunofluorescence when Met4 cells were treated with 20nM LMB for 4 hours (preliminary result of a single experiment) (Figure 33(a)). The increase in the amount of nuclear Mena was modest compared to paxillin (Figure 33(b)). Taken together, these results suggest that the nuclear export of Mena from the nucleus is XPO1 mediated.

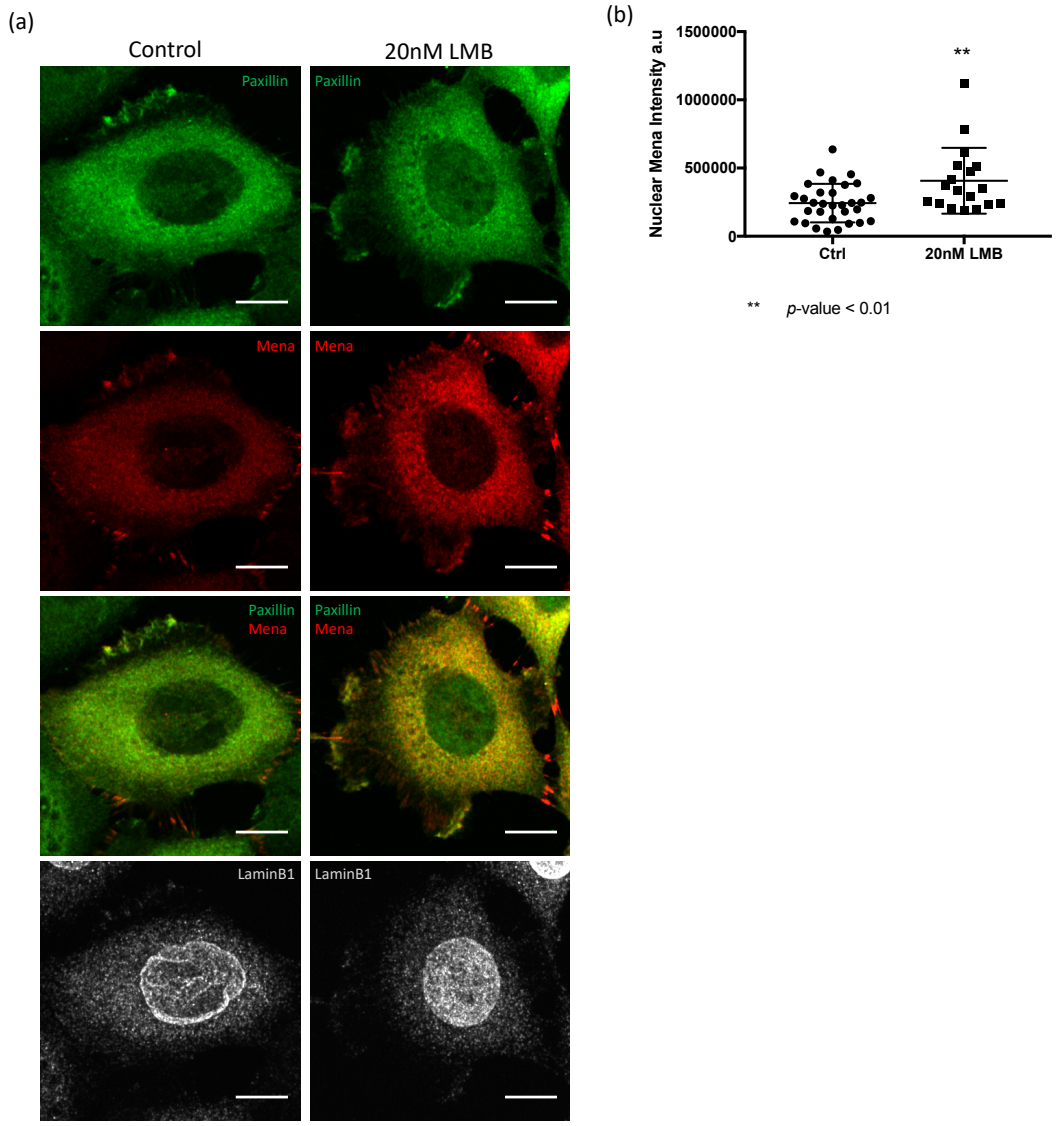


Figure 33 (a) Spinning-disk confocal image of control Met4 cells and Met4 cells treated with 20nM leptomycin-b (LMB). Cells were stained for paxillin (green) as a positive control for leptomycin-b treatment, Mena (red) and laminB1 (grey). Image is representative of multiple images from one single experiment. Scale bar = 10μm. (b) A modest increase in the amount of nuclear Mena was observed (Control: n = 31 nuclei; 20nM LMB: n = 18 nuclei; single experiment, Mann-Whitney test).

Mena is at the nuclear periphery and complexes with Nesprin-2 and actin

Mena's function as an actin regulator is well characterised[40,43]. Since Mena was more abundant in the nuclear fractions of Met4 cells, we used these cells to gain insight into the biology of nuclear Mena. In the SCC adhesome network analyses, Mena was in the same community as Nesprin-2, an actin-binding protein of the LINC (Linker of Nucleoskeleton to the Cytoskeleton) complex, suggesting that Mena and Nesprin-2 may participate in the same molecular process. To examine the localisation of Mena and Nesprin-2, they were immunostained in Met4 cells and filamentous actin (F-actin) was co-stained by phalloidin. Mena, Nesprin-2 and phalloidin were imaged by immunofluorescence using spinning-disk confocal microscopy (SDCM) and structured illumination microscopy (SIM).

Interestingly, confocal microscopy revealed that Mena colocalised with Nesprin2 at the nuclear periphery (Figure 34(a)). To determine whether Mena form complexes with Nesprin-2, the latter was immunoprecipitated from Met4 cell lysates and probed for Mena by western blot. Mena co-immunoprecipitated with Nesprin-2 (Figure 34(b)), indicating that Nesprin-2 and Mena form a complex. Orthogonal sections through a high-resolution SIM image revealed that Mena colocalises at the apical region of Nesprin-2 staining, where actin filaments also colocalise (Figure 34(c)).

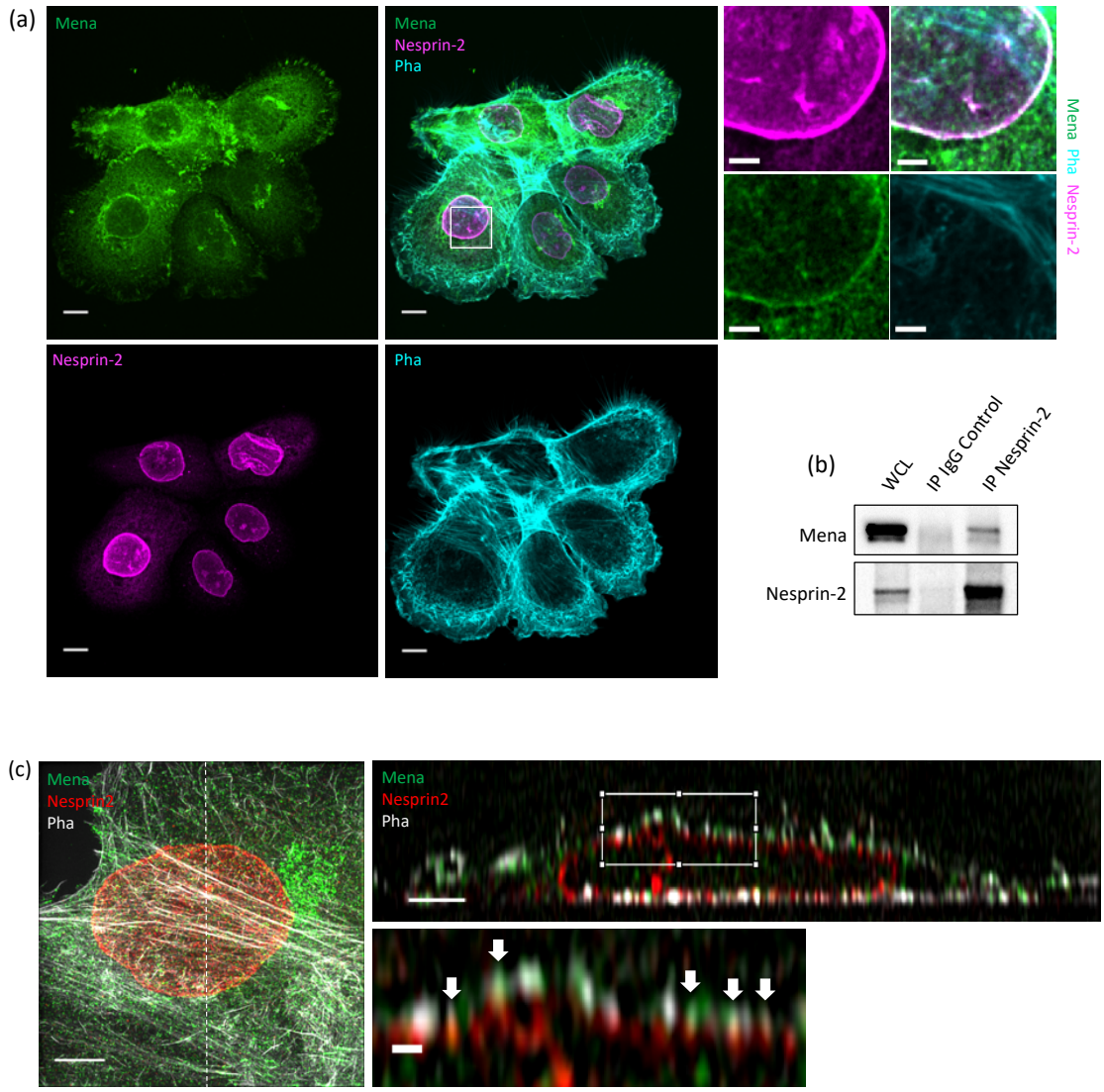


Figure 34 (a) Spinning disk confocal image of Met4 immunostained for Mena (green) and Nesprin-2 (magenta); F-actin is co-stained by phalloidin (cyan). Scale bar = 10μm. Boxed region is shown on the right at 4x magnification. Scale bar = 3μm. Image is representative of two experimental replicates. (b) Western blot analysis of shows Mena immunoprecipitates with Nesprin-2. Western blot is representative of five independent replicates. (c) Structured illumination microscopy immunostained for Mena (green) and Nesprin-2 (red); F-actin is co-stained with phalloidin (grey). The orthogonal section through the white dashed line is shown on the right. Scale bar = 2.5μm. Boxed region is shown below. Mena and F-actin was found to co-localise at the apical tip of Nesprin-2 (white arrows).

It was previously reported that actin filaments form stress fibres which bind to Nesprin-1 and Nesprin-2 and transmit tension to the nuclear lamina[35]. Since Mena was observed to complex with F-actin and Nesprin-2, we asked whether Mena could mediate the interaction between actin and Nesprin-2 and hence mediate the LINC complex interaction with lamin A/C. To test this hypothesis, we generated a Mena knocked-out Met4 cell line using CRISPR-Cas9. Nesprin-2 was immunoprecipitated in Met4 cells expressing endogenous Mena and with Mena knocked-out, and the immunoprecipitates probed for actin and lamin A/C by western blot. Preliminary results suggest that Mena depletion may result in reduced interaction between Nesprin-2 and actin and disrupts the interaction between Nesprin-2 and lamin A/C (Figure 35(a)).

Since Mena is involved in F-actin polymerisation, we estimated the effect of Mena depletion on actin filaments formation by staining for actin filaments using fluorophore phalloidin and measured the fluorescence intensity of phalloidin in Mena-expressing Met4 and Mena knock-out Met4 cells. No change was observed in phalloidin staining in Mena knock-out cells compared to Mena-expressing cells (Figure 35(b,c)), suggesting that Mena actin polymerisation activity is redundant in Met4.

We also probed for phospho-myosin light chain but failed to obtain a positive band in neither Met4 expressing endogeneous Mena nor in Met4 with Mena knocked-out.

These results suggest that Mena may play a role in force transmission through Nesprin-2 that may not involve its function in regulating actin filament formation and elongation. However, further experiments are required to confirm this. Mena needs to be re-expressed in the Mena-depleted cells to test whether the interaction between Nesprin-2 and actin can be rescued. An F-actin/G-actin ratio assay will be performed to assess more accurately the effect of Mena on actin filament formation. We also still need to assess the tension in the stress fibres in Mena depleted cells by using a phos-tag based analysis of phospho-myosin light chain as previously described[174].

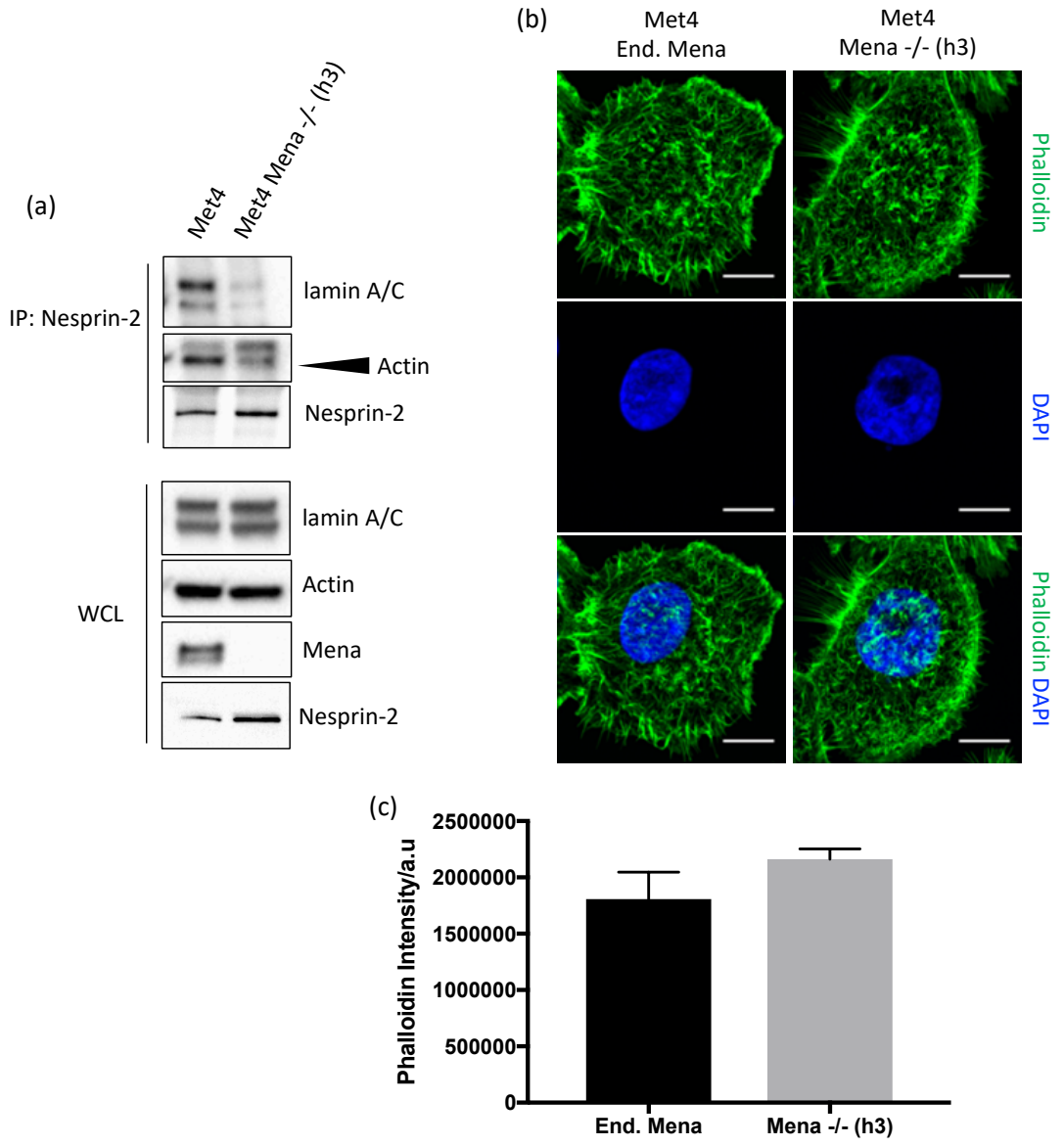


Figure 35 (a) Preliminary Western blot result of a single experiment showing reduction in the amount of actin and lamin A/C co-immunoprecipitated with Nesprin-2 upon Mena depletion in Met4 cells. (b) Spinning-disk confocal image of Met4 cells expressing endogeneous Mena (Met4 End. Mena) and Met4 Mena knock-out cells (Met4 Mena -/- (h3)) stained for F-actin (phalloidin, green) and nucleus (DAPI, blue). (c) Bar chart representing the intensity of phalloidin staining in Met4 End. Mena (n = 8 cells) and Met4 Mena knock-out cells (n = 19 cells). Error bars represent the standard deviation (no statistical significance, Mann-Whitney test). All results were from a single experiment.

Mena regulates global H3K27Me2 and EMERIN phosphorylation.

It was previously reported that force transmission through the LINC complex results in the phosphorylation of its LEM- (Lap2, EMERIN, Man1) domain binding partner EMERIN[38]. Since Mena is hypothesized to act as a clutch between Nesprin-2 and actin and thus mediates force transmission, we asked whether Mena may be involved in the phosphorylation of EMERIN. To test this, EMERIN was probed in the phosphor-tyrosine (PY20) immunoprecipitates of Met4 cells expressing endogeneous Mena, Met4 cells with Mena knocked-out (transiently transfected with an empty cDNA vector as control) and Met4 Mena knock-out cells with transiently transfected Mena11a cDNA.

Interestingly, Mena knock-out resulted in a reduction in phosphorylated EMERIN (pEMERIN) which was partially recovered upon Mena11a re-expression (Figure 36(a,b)). We did not observe complete recovery of pEMERIN despite an over-expression of Mena11a. This is possibly due to the low efficiency of transient transfection whereby only a few cells were transfected with Mena11a cDNA plasmids. It could also be due to the type of isoform being re-expressed. Despite Mena11a seemingly migrating to a similar degree to the endogenous Mena in Met4 cells, hMena (11a exon missing) has previously been shown to co-migrate with Mena11a[51]. Since

we do not know which isoform of Mena is being expressed in Met4 cells, it is possible that hMena may also be expressed. Thus, it will be interesting to see whether a complete recovery of pEMERIN will be observed upon hMena re-expression as well as upon stable re-expression of Mena11a.

EMERIN has previously been reported to anchor chromatin regions to the nuclear lamina (lamina-associated domains; LADs)[175] which are associated with di-methylation of histone 3 on lysine 9 (H3K9Me2)[37]. Moreover, it was previously reported that strain-induced perinuclear actin filament remodelling by EMERIN provides a switch from global di/tri-methylation of histone 3 on lysine 9 (H3K9Me2,3) to tri-methylation of histone 3 on lysine 27 (H3K27Me3)[5]. We thus asked whether the loss of Mena could affect histone methylation. Interestingly, preliminary data suggests that Mena depletion results in a global increase in di-methylation of histone 3 on lysine 27 (H3K27Me2) (Figure 36(c,d)). Moreover, Met4 cells expressing endogenous Mena have strong H3K27Me2 staining along the edge of the nucleus (Figure 36(c)). Whether this region corresponds to LADs still remains to be determined. Conversely, Mena-depleted Met4 cells have a more diffuse H3K27Me2 staining (Figure 36(c)). The function of H3K27Me2 is not well understood but it has been reported that H3K27Me2 protects histone from H3K27 acetylation and thus prevents non-cell type specific enhancers from being activated[176].

Strain induces a global decrease in H3K9Me2,3[5]. Here, preliminary results suggest that Mena depletion induces a decrease in H3K9Me2 (Figure 36(c, e)). H3K9Me2 is associated with LADs as well as heterochromatin (densely packed chromatin). This result therefore suggests that Mena depletion may result in less densely packed chromatin. Since both H3K27Me2 and H3K9Me2 are associated with transcription repression, we hypothesize that the increase in H3K27Me2 is a compensatory mechanism for loss in H3K9Me2 upon Mena depletion. This hypothesis still needs to be tested by ChIP-Seq.

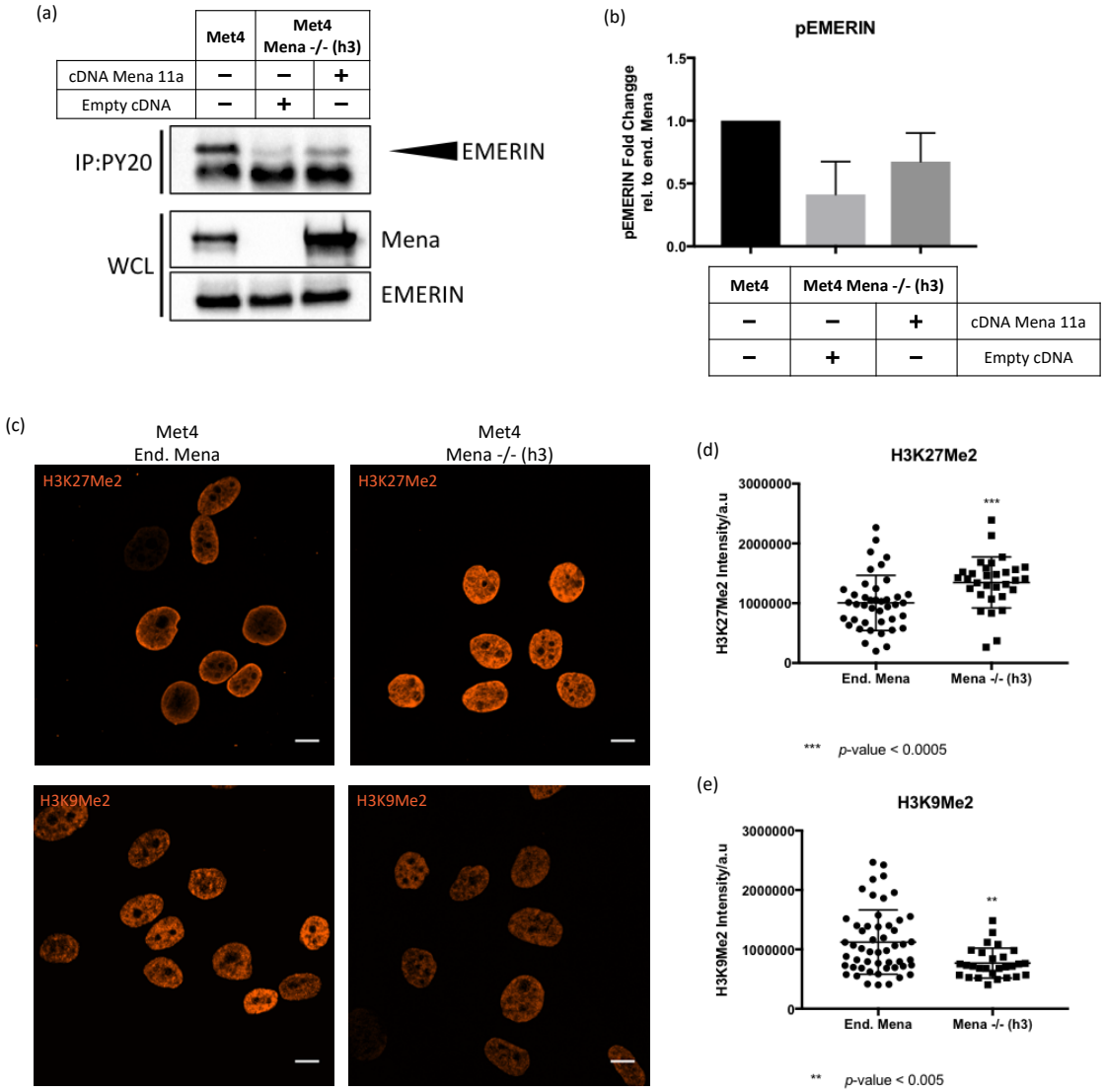


Figure 36 (a) Tyrosine phosphorylation of EMERIN is reduced upon Mena knock-out in Met4 cells and is partially rescued by transient transfection with Mena11a. Western blots are representative of two independent replicates. (b) Relative quantification of the EMERIN intensity in the western blots. Error bars represent the standard deviation from the two independent replicates. (c). Spinning-disk confocal image of Met4 cells expressing endogenous Mena (Met4 End. Mena) and Met4 cells with Mena knock-out (Met4 Mena -/- (h3)), both immunostained for H3K27Me2 and H3K9Me2. Representative image of a single experiment. Scale bar = 10 μm. (d) Mena depletion increased the intensity of H3K27Me2 staining (Met4 End. Mena, n = 41 nuclei; Met4 Mena -/-, n = 31 nuclei; Mann-Whitney test, single experiment). (e) Conversely, Mena depletion resulted in reduced H3K9Me2 staining (Met4 End. Mena, n = 52 nuclei; Met4 Mena -/-, n = 28 nuclei; Mann-Whitney test, single experiment).

■ Discussion

Here, we report a novel function of Mena in patient-derived SCC cells. Our preliminary results suggest that Mena may act as a mechanosensor and/or as a scaffold at the perinuclear region.

XPO1 was found to be an SCC adhesome hub which may connect different molecular processes during cancer progression. Putative interactors of XPO1 from the SCC adhesion complex are canonical adhesion proteins, namely paxillin, ILK and TRIP6, and Mena. We still need to validate the interactions of those canonical focal adhesion proteins with XPO1 in Met1 and Met4 cells. Paxillin, ILK and TRIP6 have previously been reported to have nuclear functions but Mena has not. Thus, based on the prediction that interactors within network proximity have similar functions[164], we asked whether Mena was also present in the nucleus of SCC cells. Interestingly, Mena was indeed present in the nucleus and was more abundant in the nucleus of Met4 than in the nucleus of Met1. As Met1 cells were derived from a primary SCC tumour and Met4 cells were derived from SCC metastasis from the same patient, this suggests that nuclear Mena may play a role in cancer progression.

Mena nuclear translocation was found to be modulated by integrin-mediated cell adhesion in Met4 but not in Met1 cells. Furthermore, FAK knock-out

reduced nuclear Mena abundance in Met4 but surprisingly increased nuclear Mena abundance in Met1 cells. FAK has yet to be re-expressed in those FAK-depleted cell lines to confirm that nuclear Mena shuttling is FAK dependent and how FAK regulates Mena nuclear shuttling is yet to be determined. We hypothesize that FAK may regulate Mena nuclear translocation by stabilising Mena at the focal adhesions.

Inhibition of XPO1 in Met4 cells resulted in nuclear accumulation of Mena, confirming that Mena nuclear export is XPO1 dependent. XPO1 has previously been isolated as part of the adhesome, though its function in the adhesome is currently unknown.

In Met4 cells, Mena was associated with the perinuclear membrane, forming a complex with the outer-nuclear membrane protein Nesprin-2. Mena depletion reduced the interaction of Nesprin-2 with actin as well as with lamin A/C. Preliminary phalloidin staining results showed no change in F-actin formation between Mena-depleted Met4 and in Mena-expressing Met4 cells. Taken together, these results suggest that Mena may act as a molecular clutch with Nesprin-2 and actin, thereby mediating the force transmission through the LINC complex. Further experiments are still required to validate this hypothesis: 1) determine the F-actin/G-actin ratio in Mena-expressing and Mena-depleted cells to assess more accurately the effect of Mena

depletion on F-actin formation, 2) rescue the interaction between Nesprin-2 and actin by re-expressing Mena in Mena-depleted cells, 3) assess levels of phospho-myosin light chain to assess the effect of Mena depletion on actin stress fibre tension.

It has previously been reported that tension applied to the LINC complex resulted in the phosphorylation of EMERIN[38]. Mena depletion also resulted in a reduction of pEMERIN which was partially rescued upon transient Mena11a re-expression, consistent with the hypothesis that Mena mediate the force transmission through Nesprin-2. However, we did not assess whether Mena depletion caused changes in tension along the stress fibres since we could not detect a positive band for phospho-myosin light chain in neither Mena-expressing Met4 nor in Mena-depleted Met4 cells.

It was previously reported that EMERIN is mostly found at the outer nuclear membrane when the cells were under stress[5]. Mena interacts with c-Abl, Src, PKA[40,41], all of which have been reported to phosphorylate LEM domain proteins such as EMERIN[177]. We hypothesize that Mena may also recruit those kinases to phosphorylate EMERIN present at the outer nuclear membrane. It is plausible that the reduction of pEMERIN is due to a reduction in recruitment of those kinases at the nuclear periphery. However,

immunoprecipitation of EMERIN in Mena-expressing and Mena-depleted Met4 cells is required to test this hypothesis.

Depletion of Mena reduced global H3K9Me2 but conversely increased global H3K27Me2 levels. We hypothesize that the global increase in H3K27Me2 is a compensatory mechanism for the loss of H3K9Me2 upon Mena depletion. This needs to be validated by ChIP-Seq experiments to test whether gene enhancers or promoters occupied by H3K9Me2 has been replaced by H3K27Me2 upon Mena depletion. A reduction in H3K9Me2 suggests decondensation of heterochromatin upon Mena depletion in Met4.

Figure 37 gives a brief overview of the working model for Mena localisation and function in Met4 cells.

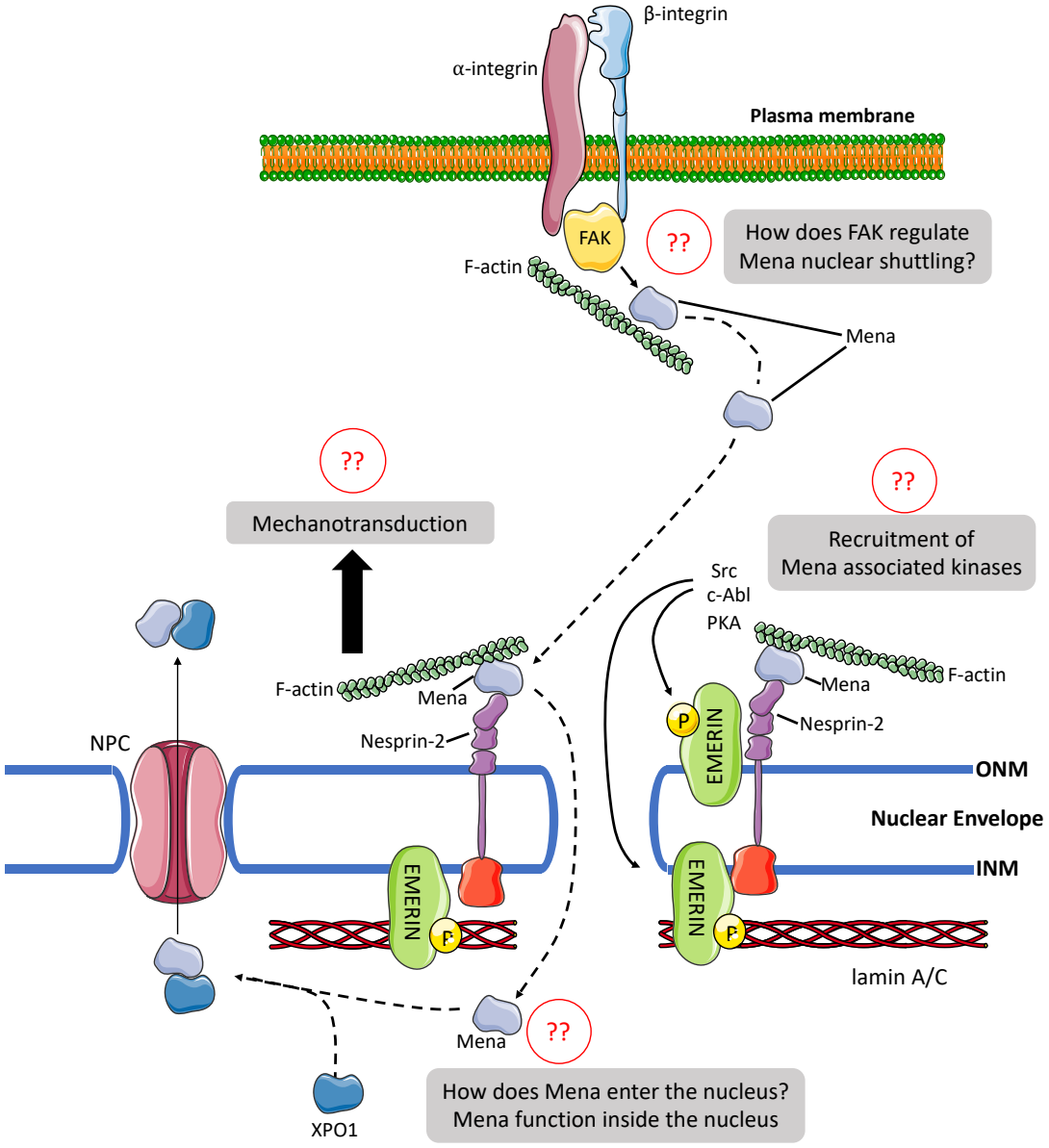


Figure 37 Summary diagram of the functions of Mena in Met4. We hypothesize that Mena nuclear shuttling is regulated by FAK at the focal adhesion. How FAK regulates Mena nuclear shuttling still needs to be determined. Mena shuttles to the perinuclear region and complex with Nesprin-2 where Mena may act as a molecular clutch between the actin filaments and Nesprin-2. This molecular clutch may transmit force through Nesprin-2 which results in the phosphorylation of EMERIN. Furthermore, since EMERIN is mostly found at the outer nuclear membrane under tension, we also hypothesize that Mena may be recruiting kinases to phosphorylate EMERIN at the outer nuclear membrane. However, we still need to test this hypothesis. Mena may enter the nucleus where it interacts with XPO1 and is exported back to the cytosol via the nuclear pore complex (NPC).

CHAPTER 6

Conclusions

Interaction between cancer cells and the extracellular matrix is required for cancer progression. Here, we aimed to study the adhesome in a cancer progression setting. We isolated the adhesome from patient-derived squamous cell carcinoma (SCC) from the primary site (Met1) and secondary metastatic site (Met4). We optimised the adesome isolation methodology for Met1 and Met4 and showed that adhesome proteins such as focal adhesion kinase (FAK), phosphorylated FAK, talin and αV-integrin were enriched while co-purifying contaminants such as nuclear proteins Histone H3, mitochondrial protein complex IV and cytosolic protein GAPDH were not isolated. Fibronectin-induced adhesome consisting of 1941 proteins from Met1 and Met4 cells were identified by liquid-chromatography tandem mass-spectrometry (LC-MS/MS).

Changes in the adhesome between primary (Met1) and metastatic (Met4) SCC cells were captured by a protein-protein interaction network of differentially regulated adhesome proteins, called the active module. Interestingly, the active module has a modular structure (community structure) with coregulated adhesome proteins clustering among themselves. This modular structure is indicative of differentially regulated molecular processes between the Met1 and Met4 at the adhesome level. These putative molecular processes could potentially be driving cancer progression at the adhesome level. Future work involves the identification of these molecular processes by gene ontology enrichment.

Hubs identification using direct and community-assisted approaches were used to identify and prioritise essential proteins in those putative molecular processes. Exportin-1 (XPO1) was recurrently identified as one of the essential hubs that may be driving cancer progression at the adhesome level. Interestingly, XPO1 was also isolated in other fibronectin-induced adhesome though its function there is not yet known[68]. XPO1 had putative interactions with canonical focal adhesion components paxillin, Mena, integrin linked kinase (ILK) and TRIP6. Whether those putative interactions change during cancer progression is yet to be determined. It is tempting to speculate that XPO1 interacts with paxillin and ILK in primary cells but rewires to Mena and TRIP6 during metastasis. Co-immunoprecipitation of XPO1 with Mena, ILK, TRIP6 and paxillin in Met1 and Met4 cells should validate this hypothesis.

FAK knock-out in Met1 resulted in an increase in Mena nuclear localisation. Interestingly, Sathe *et al.*[173] described a similar observation whereby FAK or vinculin knock-out resulted in increased paxillin nuclear translocation. They thus postulated that FAK and vinculin stabilise paxillin at focal adhesions, thereby decreasing paxillin translocation to the nucleus. We hypothesize that a similar mechanism may be involved for the Mena translocation to the nucleus in Met1. Conversely, integrin activation and FAK are required for Mena nuclear localisation in Met4. Here, we hypothesize that activation of integrin and presumably FAK allow for the formation of a Mena complex,

thereby enabling Mena nuclear transport. This could provide a novel function for both integrin and FAK in metastasis.

There are increasing evidence that that the nuclear localisation of focal adhesion components (e.g. paxillin, zyxin, FAK) have a role in carcinogenesis[4,158]. Our preliminary results showed that Mena has a nuclear function in metastatic cells and may act as a molecular clutch with Nesprin-2 and actin, thereby mediating the force transmission through the LINC complex. Two human Mena isoforms namely hMena11a and hMenaΔv6 were previously characterised in human cancers, with hMenaΔv6 conferring an invasive potential to metastatic cells while hMena11a was associated with epithelial phenotypes[52] and MAPK mitogenic activity[51]. Moreover, hMena11a suppresses invasion *in vivo*[48]. The Mena isoforms expressed in Met1 and Met4 has yet to be determined. Nevertheless, hMena11a seems to migrate to a similar degree to endogenous nuclear Mena in Met4, suggesting that Mena11a may be present in the nucleus of metastatic cell line. We speculate that Mena11a may confer cell growth and/or cell survival to Met4 in the lymph node.

It has previously been reported that tension applied to the LINC complex resulted in the phosphorylation of EMERIN[38]. Mena depletion also resulted in a reduction of phosphorylated EMERIN which was partially rescued upon

transient Mena11a re-expression, consistent with the hypothesis that Mena mediate the force transmission through Nesprin-2. This force transmission could be tested using Nesprin-2 FRET probes as previously described[178]. However, we have yet to assess whether Mena depletion caused changes in tension along the stress fibres.

It was previously reported that EMERIN is mostly found at the outer nuclear membrane when the cells were under stress[5]. Mena interacts with c-Abl, Src, PKA[40,41], all of which have been reported to phosphorylate LEM domain proteins such as EMERIN[177]. We hypothesize that Mena may recruit those kinases to phosphorylate EMERIN present at the outer nuclear membrane. Future works would involve immunoprecipitation of EMERIN and assess the changes, if any, in the interaction between EMERIN and its kinases in Mena-expressing and Mena-depleted Met4 cells.

Our preliminary results suggest that the nucleus of Met4 cells may regulate histone modification by the mechanosensing activity of Mena at the nuclear periphery. Interestingly, it was previously reported that nuclear actin mediates a polycomb-mediated histone methylation switch[5]. Therefore, it has been proposed that regulation of nuclear intake of G-actin by mechanosensing complexes at the nuclear periphery controls the transcriptional activity and levels of H3K27Me3. Since Mena binds to G-actin

at its EVH2 domain, we therefore speculate that Mena may regulate the level of G-actin in the nucleus. This will be tested by probing for actin in the nuclear fraction of Mena-expressing and Mena-depleted Met4 cells. Moreover, it has been suggested that the switching off of gene expression by histone modification drives cancer progression. It is thus tempting to speculate that Mena mediates cancer progression by maintaining H3K9Me2 which prevents transcription activity[179].

Force transmission from the focal adhesion to the nucleus regulates gene transcription via genome reorganisation, transcription factors nuclear translocation and histone modification, Although force transmission to the nucleus is known to require the use of the actin cytoskeleton, focal adhesion proteins and strain-maintaining proteins such as actomyosin, the role of actin regulating proteins in regulating force transmission to the nucleus was unknown. Moreover, focal adhesion proteins are also observed at the nucleus but their roles at the nucleus are yet to be defined. However, it is tempting to speculate that adhesion proteins can form complexes with the cytoskeleton at the nuclear membrane. Moreover, the fact that nuclear confinement requires actin bundles suggest that in addition to filamentous actin[180], actin binding proteins may also be required for filamentous actin coupling to the nucleus. Our findings here provide a foundation for further studies in the role of actin regulating proteins at the nuclear periphery controlling metastasis.

References

1 Schwartz, M. A. Integrin signaling revisited. *Trends Cell Biol* **11**, 466-470, doi:Doi 10.1016/S0962-8924(01)02152-3 (2001).

2 Hirata, E. *et al.* Intravital Imaging Reveals How BRAF Inhibition Generates Drug-Tolerant Microenvironments with High Integrin beta1/FAK Signaling. *Cancer cell* **27**, 574-588, doi:10.1016/j.ccell.2015.03.008 (2015).

3 Serrels, B. *et al.* IL-33 and ST2 mediate FAK-dependent antitumor immune evasion through transcriptional networks. *Sci Signal* **10** (2017).

4 Serrels, A. *et al.* Nuclear FAK controls chemokine transcription, Tregs, and evasion of anti-tumor immunity. *Cell* **163**, 160-173, doi:10.1016/j.cell.2015.09.001 (2015).

5 Le, H. Q. *et al.* Mechanical regulation of transcription controls Polycomb-mediated gene silencing during lineage commitment. *Nature cell biology* **18**, 864-875, doi:10.1038/ncb3387 (2016).

6 Askari, J. A., Buckley, P. A., Mould, A. P. & Humphries, M. J. Linking integrin conformation to function. *J Cell Sci* **122**, 165-170, doi:10.1242/jcs.018556 (2009).

7 Wozniak, M. A., Modzelewska, K., Kwong, L. & Keely, P. J. Focal adhesion regulation of cell behavior. *Biochim Biophys Acta* **1692**, 103-119, doi:10.1016/j.bbamcr.2004.04.007 (2004).

8 Frame, M. C., Patel, H., Serrels, B., Lietha, D. & Eck, M. J. The FERM domain: organizing the structure and function of FAK. *Nature reviews. Molecular cell biology* **11**, 802-814, doi:10.1038/nrm2996 (2010).

9 Donato, D. M., Ryzhova, L. M., Meenderink, L. M., Kaverina, I. & Hanks, S. K. Dynamics and mechanism of p130Cas localization to focal adhesions. *J Biol Chem* **285**, 20769-20779, doi:10.1074/jbc.M109.091207 (2010).

10 Bouton, A. H., Riggins, R. B. & Bruce-Staskal, P. J. Functions of the adapter protein Cas: signal convergence and the determination of cellular responses. *Oncogene* **20**, 6448-6458, doi:10.1038/sj.onc.1204785 (2001).

11 Turner, C. E. Paxillin and focal adhesion signalling. *Nature cell biology* **2**, E231-236, doi:10.1038/35046659 (2000).

12 Yano, H. *et al.* Paxillin alpha and Crk-associated substrate exert opposing effects on cell migration and contact inhibition of growth through tyrosine phosphorylation. *Proc Natl Acad Sci U S A* **97**, 9076-9081 (2000).

13 McLean, G. W. *et al.* The role of focal-adhesion kinase in cancer - a new therapeutic opportunity. *Nat Rev Cancer* **5**, 505-515, doi:10.1038/nrc1647 (2005).

14 Mitra, S. K., Hanson, D. A. & Schlaepfer, D. D. Focal adhesion kinase: in command and control of cell motility. *Nature reviews. Molecular cell biology* **6**, 56-68, doi:10.1038/nrm1549 (2005).

15 Janostiak, R., Pataki, A. C., Brabek, J. & Rosel, D. Mechanosensors in integrin signaling: the emerging role of p130Cas. *Eur J Cell Biol* **93**, 445-454, doi:10.1016/j.ejcb.2014.07.002 (2014).

16 Sit, S. T. & Manser, E. Rho GTPases and their role in organizing the actin cytoskeleton. *J Cell Sci* **124**, 679-683, doi:10.1242/jcs.064964 (2011).
17 Reiske, H. R. *et al.* Requirement of phosphatidylinositol 3-kinase in focal adhesion kinase-promoted cell migration. *J Biol Chem* **274**, 12361-12366 (1999).
18 Jones, R. J., Brunton, V. G. & Frame, M. C. Adhesion-linked kinases in cancer; emphasis on src, focal adhesion kinase and PI 3-kinase. *Eur J Cancer* **36**, 1595-1606 (2000).
19 Almeida, E. A. *et al.* Matrix survival signaling: from fibronectin via focal adhesion kinase to c-Jun NH(2)-terminal kinase. *The Journal of cell biology* **149**, 741-754 (2000).
20 Zhao, J. & Guan, J. L. Signal transduction by focal adhesion kinase in cancer. *Cancer Metastasis Rev* **28**, 35-49, doi:10.1007/s10555-008-9165-4 (2009).
21 Garton, A. J. & Tonks, N. K. Regulation of fibroblast motility by the protein tyrosine phosphatase PTP-PEST. *J Biol Chem* **274**, 3811-3818 (1999).
22 Angers-Loustau, A. *et al.* Protein tyrosine phosphatase-PEST regulates focal adhesion disassembly, migration, and cytokinesis in fibroblasts. *The Journal of cell biology* **144**, 1019-1031 (1999).
23 Tamura, M. *et al.* Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN. *Science* **280**, 1614-1617 (1998).
24 Lewis, J. M., Baskaran, R., Taagepera, S., Schwartz, M. A. & Wang, J. Y. Integrin regulation of c-Abl tyrosine kinase activity and cytoplasmic-nuclear transport. *Proc Natl Acad Sci U S A* **93**, 15174-15179 (1996).
25 Chodniewicz, D. & Klemke, R. L. Regulation of integrin-mediated cellular responses through assembly of a CAS/Crk scaffold. *Bba-Mol Cell Res* **1692**, 63-76, doi:10.1016/j.bbamcr.2004.03.006 (2004).
26 Fogh, B. S., Multhaupt, H. A. & Couchman, J. R. Protein kinase C, focal adhesions and the regulation of cell migration. *J Histochem Cytochem* **62**, 172-184, doi:10.1369/0022155413517701 (2014).
27 Ziegler, W. H., Tigges, U., Zieseniss, A. & Jockusch, B. M. A lipid-regulated docking site on vinculin for protein kinase C. *J Biol Chem* **277**, 7396-7404, doi:10.1074/jbc.M110008200 (2002).
28 Fedorchak, G. R., Kaminski, A. & Lammerding, J. Cellular mechanosensing: Getting to the nucleus of it all. *Prog Biophys Mol Bio* **115**, 76-92, doi:10.1016/j.pbiomolbio.2014.06.009 (2014).
29 del Rio, A. *et al.* Stretching single talin rod molecules activates vinculin binding. *Science* **323**, 638-641, doi:10.1126/science.1162912 (2009).
30 Kanchanawong, P. *et al.* Nanoscale architecture of integrin-based cell adhesions. *Nature* **468**, 580-584, doi:10.1038/nature09621 (2010).
31 Sawada, Y. *et al.* Force sensing by mechanical extension of the Src family kinase substrate p130Cas. *Cell* **127**, 1015-1026, doi:10.1016/j.cell.2006.09.044 (2006).
32 Horton, E. R., Astudillo, P., Humphries, M. J. & Humphries, J. D. Mechanosensitivity of integrin adhesion complexes: role of the consensus adhesome. *Exp Cell Res* **343**, 7-13, doi:10.1016/j.yexcr.2015.10.025 (2016).

33 Yoshigi, M., Hoffman, L. M., Jensen, C. C., Yost, H. J. & Beckerle, M. C. Mechanical force mobilizes zyxin from focal adhesions to actin filaments and regulates cytoskeletal reinforcement. *The Journal of cell biology* **171**, 209-215, doi:10.1083/jcb.200505018 (2005).

34 Smith, M. A. *et al.* A zyxin-mediated mechanism for actin stress fiber maintenance and repair. *Dev Cell* **19**, 365-376, doi:10.1016/j.devcel.2010.08.008 (2010).

35 Wang, N., Tytell, J. D. & Ingber, D. E. Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus. *Nat Rev Mol Cell Bio* **10**, 75-82, doi:10.1038/nrm2594 (2009).

36 van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791, doi:10.1016/j.cell.2017.04.022 (2017).

37 Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-192, doi:10.1016/j.cell.2013.02.028 (2013).

38 Guilluy, C. *et al.* Isolated nuclei adapt to force and reveal a mechanotransduction pathway in the nucleus. *Nature cell biology* **16**, 376-381, doi:10.1038/ncb2927 (2014).

39 Elosegui-Artola, A. *et al.* Force Triggers YAP Nuclear Entry by Regulating Transport across Nuclear Pores. *Cell* **171**, 1397-+, doi:10.1016/j.cell.2017.10.008 (2017).

40 Krause, M., Dent, E. W., Bear, J. E., Loureiro, J. J. & Gertler, F. B. Ena/VASP proteins: regulators of the actin cytoskeleton and cell migration. *Annu Rev Cell Dev Biol* **19**, 541-564, doi:10.1146/annurev.cellbio.19.050103.103356 (2003).

41 Tani, K. *et al.* Abl interactor 1 promotes tyrosine 296 phosphorylation of mammalian enabled (Mena) by c-Abl kinase. *J Biol Chem* **278**, 21685-21692, doi:10.1074/jbc.M301447200 (2003).

42 Gupton, S. L. *et al.* Mena binds alpha5 integrin directly and modulates alpha5beta1 function. *The Journal of cell biology* **198**, 657-676, doi:10.1083/jcb.201202079 (2012).

43 Sechi, A. S. & Wehland, J. ENA/VASP proteins: multifunctional regulators of actin cytoskeleton dynamics. *Front Biosci* **9**, 1294-1310 (2004).

44 Bear, J. E. & Gertler, F. B. Ena/VASP: towards resolving a pointed controversy at the barbed end. *J Cell Sci* **122**, 1947-1953, doi:10.1242/jcs.038125 (2009).

45 Bear, J. E. *et al.* Antagonism between Ena/VASP proteins and actin filament capping regulates fibroblast motility. *Cell* **109**, 509-521, doi:Doi 10.1016/S0092-8674(02)00731-6 (2002).

46 Lanier, L. M. *et al.* Mena is required for neurulation and commissure formation. *Neuron* **22**, 313-325, doi:Doi 10.1016/S0896-6273(00)81092-2 (1999).

47 Goswami, S. *et al.* Identification of invasion specific splice variants of the cytoskeletal protein Mena present in mammary tumor cells during invasion in vivo. *Clin Exp Metastasis* **26**, 153-159, doi:10.1007/s10585-008-9225-8 (2009).

48 Roussos, E. T. *et al.* Mena invasive (Mena(INV)) promotes multicellular streaming motility and transendothelial migration in a mouse model of breast cancer. *J Cell Sci* **124**, 2120-2131, doi:10.1242/jcs.086231 (2011).
49 Urbanelli, L. *et al.* Characterization of human Enah gene. *Biochim Biophys Acta* **1759**, 99-107, doi:10.1016/j.bbaexp.2006.01.001 (2006).
50 Oudin, M. J. *et al.* Characterization of the expression of the pro-metastatic Mena(INV) isoform during breast tumor progression. *Clin Exp Metastasis* **33**, 249-261, doi:10.1007/s10585-015-9775-5 (2016).
51 Di Modugno, F. *et al.* Molecular cloning of hMena (ENAH) and its splice variant hMena(+11a): Epidermal growth factor increases their expression and stimulates hMena(+11a) phosphorylation in breast cancer cell lines. *Cancer Res* **67**, 2657-2665, doi:10.1158/0008-5472.Can-06-1997 (2007).
52 Di Modugno, F. *et al.* Splicing program of human MENA produces a previously undescribed isoform associated with invasive, mesenchymal-like breast tumors. *Proc Natl Acad Sci U S A* **109**, 19280-19285, doi:10.1073/pnas.1214394109 (2012).
53 Di Modugno, F. *et al.* The cytoskeleton regulatory protein hMena (ENAH) is overexpressed in human benign breast lesions with high risk of transformation and human epidermal growth factor receptor-2-positive/hormonal receptor-negative tumors. *Clin Cancer Res* **12**, 1470-1478, doi:10.1158/1078-0432.Ccr-05-2027 (2006).
54 Gurzu, S. *et al.* The expression of cytoskeleton regulatory protein Mena in colorectal lesions. *Rom J Morphol Embryol* **49**, 345-349 (2008).
55 Toyoda, A. *et al.* Aberrant expression of human ortholog of mammalian enabled (hMena) in human colorectal carcinomas: implications for its role in tumor progression. *Int J Oncol* **34**, 53-60 (2009).
56 Gurzu, S., Jung, I., Prantner, I., Chira, L. & Ember, I. The immunohistochemical aspects of protein Mena in cervical lesions. *Rom J Morphol Embryol* **50**, 213-216 (2009).
57 Pignatelli, J. *et al.* Invasive breast carcinoma cells from patients exhibit MenaINV- and macrophage-dependent transendothelial migration. *Sci Signal* **7**, ra112, doi:10.1126/scisignal.2005329 (2014).
58 Vidaki, M. *et al.* A Requirement for Mena, an Actin Regulator, in Local mRNA Translation in Developing Neurons. *Neuron* **95**, 608-622 e605, doi:10.1016/j.neuron.2017.06.048 (2017).
59 Zamir, E. & Geiger, B. Molecular complexity and dynamics of cell-matrix adhesions. *J Cell Sci* **114**, 3583-3590 (2001).
60 Zamir, E. & Geiger, B. Components of cell-matrix adhesions. *J Cell Sci* **114**, 3577-3579 (2001).
61 Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R. & Geiger, B. Functional atlas of the integrin adhesome. *Nature cell biology* **9**, 858-867, doi:10.1038/ncb0807-858 (2007).
62 Zaidel-Bar, R. & Geiger, B. The switchable integrin adhesome. *J Cell Sci* **123**, 1385-1388, doi:10.1242/jcs.066183 (2010).

63 Geiger, T. & Zaidel-Bar, R. Opening the floodgates: proteomics and the integrin adhesome. *Curr Opin Cell Biol* **24**, 562-568, doi:10.1016/j.ceb.2012.05.004 (2012).

64 Humphries, J. D. *et al.* Proteomic Analysis of Integrin-Associated Complexes Identifies RCC2 as a Dual Regulator of Rac1 and Arf6. *Sci Signal* **2**, doi:ARTN ra51

DOI 10.1126/scisignal.2000396 (2009).

65 Kuo, J. C., Han, X., Hsiao, C. T., Yates, J. R., 3rd & Waterman, C. M. Analysis of the myosin-II-responsive focal adhesion proteome reveals a role for beta-Pix in negative regulation of focal adhesion maturation. *Nature cell biology* **13**, 383-393, doi:10.1038/ncb2216 (2011).

66 Schiller, H. B. *et al.* beta1- and alphav-class integrins cooperate to regulate myosin II during rigidity sensing of fibronectin-based microenvironments. *Nature cell biology* **15**, 625-636, doi:10.1038/ncb2747 (2013).

67 Byron, A. *et al.* A proteomic approach reveals integrin activation state-dependent control of microtubule cortical targeting. *Nature communications* **6**, 6135, doi:10.1038/ncomms7135 (2015).

68 Horton, E. R. *et al.* Definition of a consensus integrin adhesome and its dynamics during adhesion complex assembly and disassembly. *Nature cell biology* **17**, 1577-1587, doi:10.1038/ncb3257 (2015).

69 Horton, E. R. *et al.* The integrin adhesome network at a glance. *J Cell Sci* **129**, 4159-4163, doi:10.1242/jcs.192054 (2016).

70 Berrier, A. L. & Yamada, K. M. Cell-matrix adhesion. *J Cell Physiol* **213**, 565-573, doi:10.1002/jcp.21237 (2007).

71 Ganguly, K. K., Pal, S., Moulik, S. & Chatterjee, A. Integrins and metastasis. *Cell Adh Migr* **7**, 251-261, doi:10.4161/cam.23840 (2013).

72 Byron, A. & Frame, M. C. Adhesion protein networks reveal functions proximal and distal to cell-matrix contacts. *Curr Opin Cell Biol* **39**, 93-100, doi:10.1016/j.ceb.2016.02.013 (2016).

73 Hanash, S. & Taguchi, A. The grand challenge to decipher the cancer proteome. *Nat Rev Cancer* **10**, 652-660, doi:10.1038/nrc2918 (2010).

74 Altelaar, A. F., Munoz, J. & Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **14**, 35-48, doi:10.1038/nrg3356 (2013).

75 Jones, M. C. *et al.* Isolation of integrin-based adhesion complexes. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **66**, 9 8 1-9 8 15, doi:10.1002/0471143030.cb0908s66 (2015).

76 Proby, C. M. *et al.* Spontaneous keratinocyte cell lines representing early and advanced stages of malignant transformation of the epidermis. *Experimental dermatology* **9**, 104-117 (2000).

77 Popp, S. *et al.* Genetic characterization of a human skin carcinoma progression model: from primary tumor to metastasis. *The Journal of investigative dermatology* **115**, 1095-1103, doi:10.1046/j.1523-1747.2000.00173.x (2000).

78 Choudhary, C. & Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nature reviews. Molecular cell biology* **11**, 427-439, doi:10.1038/nrm2900 (2010).
79 Callister, S. J. *et al.* Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J Proteome Res* **5**, 277-286, doi:10.1021/pr050300l (2006).
80 Karpievitch, Y. V., Dabney, A. R. & Smith, R. D. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13 Suppl 16**, S5, doi:10.1186/1471-2105-13-S16-S5 (2012).
81 Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185-193 (2003).
82 Haukoos, J. S. & Newgard, C. D. Advanced statistics: missing data in clinical research--part 1: an introduction and conceptual framework. *Acad Emerg Med* **14**, 662-668, doi:10.1197/j.aem.2006.11.037 (2007).
83 Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J Proteome Res* **15**, 1116-1125, doi:10.1021/acs.jproteome.5b00981 (2016).
84 Wang, J. *et al.* In-depth method assessments of differentially expressed protein detection for shotgun proteomics data with missing values. *Sci Rep* **7**, 3367, doi:10.1038/s41598-017-03650-8 (2017).
85 Koopmans, F., Cornelisse, L. N., Heskes, T. & Dijkstra, T. M. Empirical Bayesian random censoring threshold model improves detection of differentially abundant proteins. *J Proteome Res* **13**, 3871-3880, doi:10.1021/pr500171u (2014).
86 Webb-Robertson, B. J. *et al.* Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics. *J Proteome Res* **14**, 1993-2001, doi:10.1021/pr501138h (2015).
87 Bo, T. H., Dysvik, J. & Jonassen, I. LSimpute: accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Res* **32**, doi:ARTN e34
10.1093/nar/gnh026 (2004).
88 Goh, K. I. *et al.* The human disease network. *Proc Natl Acad Sci U S A* **104**, 8685-8690, doi:10.1073/pnas.0701361104 (2007).
89 Menche, J. *et al.* Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 841-+, doi:Artn 1257601
Doi 10.1126/Science.1257601 (2015).
90 Barabasi, A. L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* **12**, 56-68, doi:10.1038/nrg2918 (2011).
91 Mitra, K., Carvunis, A. R., Ramesh, S. K. & Ideker, T. Integrative approaches for finding modular structure in biological networks. *Nature Reviews Genetics* **14**, 719-732, doi:10.1038/nrg3552 (2013).

92 Adamcsek, B., Palla, G., Farkas, I. J., Derenyi, I. & Vicsek, T. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* **22**, 1021-1023, doi:10.1093/bioinformatics/btl039 (2006).
93 Liu, Q., Chen, Y. P. & Li, J. k-Partite cliques of protein interactions: A novel subgraph topology for functional coherence analysis on PPI networks. *J Theor Biol* **340**, 146-154, doi:10.1016/j.jtbi.2013.09.013 (2014).
94 Guimera, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895-900, doi:10.1038/nature03288 (2005).
95 Fortunato, S. Community detection in graphs. *Phys. Rep.-Rev. Sec. Phys. Lett.* **486**, 75-174, doi:10.1016/j.physrep.2009.11.002 (2010).
96 Mason, O. & Verwoerd, M. Graph theory and networks in Biology. *IET Syst Biol* **1**, 89-119 (2007).
97 Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41-42, doi:10.1038/35075138 (2001).
98 Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88-93, doi:10.1038/nature02555 (2004).
99 Chang, X., Xu, T., Li, Y. & Wang, K. Dynamic modular architecture of protein-protein interaction networks beyond the dichotomy of 'date' and 'party' hubs. *Sci Rep* **3**, 1691, doi:10.1038/srep01691 (2013).
100 Taylor, I. W. *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol* **27**, 199-204, doi:10.1038/nbt.1522 (2009).
101 Fraser, H. B. Modularity and evolutionary constraint on proteins. *Nat Genet* **37**, 351-352, doi:10.1038/ng1530 (2005).
102 Batada, N. N. *et al.* Still stratus not altocumulus: Further evidence against the date/party hub distinction. *Plos Biol* **5**, 1202-1206, doi:ARTN e154 10.1371/journal.pbio.0050154 (2007).
103 Guimera, R. & Amaral, L. A. Cartography of complex networks: modules and universal roles. *J Stat Mech* **2005**, nihpa35573, doi:10.1088/1742-5468/2005/02/P02001 (2005).
104 Chen, J. & Yuan, B. Detecting functional modules in the yeast protein-protein interaction network. *Bioinformatics* **22**, 2283-2290, doi:10.1093/bioinformatics/btl370 (2006).
105 Chawade, A., Alexandersson, E. & Levander, F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J Proteome Res* **13**, 3114-3120, doi:10.1021/pr401264n (2014).
106 Josh Starmer. *Statquest: Quantile normalization*, <https://www.youtube.com/watch?v=ecjN6Xpv6SE> (2017).
107 Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A. & Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96-104 (2002).
108 Anderle, M., Roy, S., Lin, H., Becker, C. & Joho, K. Quantifying reproducibility for differential proteomics: noise analysis for protein liquid chromatography-

mass spectrometry of human serum. *Bioinformatics* **20**, 3575-3582, doi:10.1093/bioinformatics/bth446 (2004).

109 Valikangas, T., Suomi, T. & Elo, L. L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform* **19**, 1-11, doi:10.1093/bib/bbw095 (2018).

110 Ballman, K. V., Grill, D. E., Oberg, A. L. & Therneau, T. M. Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* **20**, 2778-2786, doi:10.1093/bioinformatics/bth327 (2004).

111 Clough, T., Thaminy, S., Ragg, S., Aebersold, R. & Vitek, O. Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. *BMC Bioinformatics* **13 Suppl 16**, S6, doi:10.1186/1471-2105-13-S16-S6 (2012).

112 van Buuren, S., Boshuizen Hc Fau - Knook, D. L. & Knook, D. L. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* **18**, 681-694 (1999).

113 Newgard, C. D. & Haukoos, J. S. Advanced statistics: missing data in clinical research--part 2: multiple imputation. *Acad Emerg Med* **14**, 669-678, doi:10.1197/j.aem.2006.11.038 (2007).

114 van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw* **45**, 1-67 (2011).

115 Beisser, D., Klau, G. W., Dandekar, T., Muller, T. & Dittrich, M. T. BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics* **26**, 1129-1130, doi:10.1093/bioinformatics/btq089 (2010).

116 Dittrich, M. T., Klau, G. W., Rosenwald, A., Dandekar, T. & Muller, T. Identifying functional modules in protein-protein interaction networks: an integrated exact approach. *Bioinformatics* **24**, I223-I231, doi:10.1093/bioinformatics/btn161 (2008).

117 Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J Stat Mech-Theory E*, doi:Artn P10008 10.1088/1742-5468/2008/10/P10008 (2008).

118 Newman, M. E. & Girvan, M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **69**, 026113, doi:10.1103/PhysRevE.69.026113 (2004).

119 Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems* (2006).

120 Newman, M. E. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* **103**, 8577-8582, doi:10.1073/pnas.0601602103 (2006).

121 McLean, C., He, X., Simpson, I. & Armstrong, J. Improved Functional Enrichment Analysis of Biological Networks using Scalable Modularity Based Clustering. *Journal of Proteomics & Bioinformatics* **9**, doi:10.4172/jpb.1000383 (2016).

122 Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Physical Review E* **74**, doi:ARTN 016110 10.1103/PhysRevE.74.016110 (2006).

123 Traag, V. A., Krings, G. & Van Dooren, P. Significant scales in community structure. *Sci Rep* **3**, 2930, doi:10.1038/srep02930 (2013).
124 Traag, V. A. & Van Dooren, P. Narrow scope for resolution-limit-free community detection. *Physical Review E* **84**, doi:ARTN 016114 10.1103/PhysRevE.84.016114 (2011).
125 Aldecoa, R. & Marin, I. Surprise maximization reveals the community structure of complex networks. *Sci Rep* **3**, 1060, doi:10.1038/srep01060 (2013).
126 Lü, L. *et al.* Vital nodes identification in complex networks. *Physics Reports* **650**, 1-63, doi:https://doi.org/10.1016/j.physrep.2016.06.007 (2016).
127 Yang, L. H., Mali, P., Kim-Kiselak, C. & Church, G. CRISPR-Cas-Mediated Targeted Genome Editing in Human Cells. *Gene Correction: Methods and Protocols* **1114**, 245-267, doi:10.1007/978-1-62703-761-7_16 (2014).
128 Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
129 Steeg, P. S. Targeting metastasis. *Nat Rev Cancer* **16**, 201-218, doi:10.1038/nrc.2016.25 (2016).
130 Robertson, J. *et al.* Defining the phospho-adhesome through the phosphoproteomic analysis of integrin signalling. *Nature communications* **6**, 6265, doi:10.1038/ncomms7265 (2015).
131 Schiller, H. B., Friedel, C. C., Boulegue, C. & Fassler, R. Quantitative proteomics of the integrin adhesome show a myosin II-dependent recruitment of LIM domain proteins. *EMBO reports* **12**, 259-266, doi:10.1038/embor.2011.5 (2011).
132 Humphries, M. J., Obara, M., Olden, K. & Yamada, K. M. Role of fibronectin in adhesion, migration, and metastasis. *Cancer Invest* **7**, 373-393 (1989).
133 Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* **19**, 1423-1437, doi:10.1038/nm.3394 (2013).
134 Byron, A. Proteomic Profiling of Integrin Adhesion Complex Assembly. *Methods in molecular biology* **1764**, 193-236, doi:10.1007/978-1-4939-7759-8_13 (2018).
135 Piehowski, P. D. *et al.* Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J Proteome Res* **12**, 2128-2137, doi:10.1021/pr301146m (2013).
136 Mayer, C.-D. & Glasbey, C. A. in *Probabilistic Modeling in Bioinformatics and Medical Informatics* (eds Dirk Husmeier, Richard Dybowski, & Stephen Roberts) 211-238 (Springer London, 2005).
137 Karpievitch, Y. *et al.* A statistical framework for protein quantitation in bottom-up MS-based proteomics. *Bioinformatics* **25**, 2028-2034, doi:10.1093/bioinformatics/btp362 (2009).
138 Gandolfo, L. C. & Speed, T. P. RLE plots: Visualizing unwanted variation in high dimensional data. *PLoS One* **13**, e0191629, doi:10.1371/journal.pone.0191629 (2018).

139 Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535-539, doi:10.1093/nar/gkj109 (2006).
140 Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18 Suppl 1**, S233-240 (2002).
141 Muller, T. & Dittrich, M. Functional Modules in Protein-Protein Interaction Networks. *Syst Bio* **1**, 353-369, doi:10.1007/978-1-4419-5797-9_14 (2010).
142 Wu, Z., Zhao, X. & Chen, L. Identifying responsive functional modules from protein-protein interaction network. *Molecules and Cells* **27**, 271-277, doi:10.1007/s10059-009-0035-x (2009).
143 Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *P Natl Acad Sci USA* **99**, 7821-7826, doi:10.1073/pnas.122653799 (2002).
144 Fortunato, S. & Hric, D. Community detection in networks: A user guide. *Phys. Rep.-Rev. Sec. Phys. Lett.* **659**, 1-44, doi:10.1016/j.physrep.2016.09.002 (2016).
145 Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci U S A* **105**, 1118-1123, doi:10.1073/pnas.0706851105 (2008).
146 Ideker, T. & Krogan, N. J. Differential network biology. *Mol Syst Biol* **8**, 565, doi:10.1038/msb.2011.99 (2012).
147 Yona, G., Dirks, W., Rahman, S. & Lin, D. M. Effective similarity measures for expression profiles. *Bioinformatics* **22**, 1616-1622, doi:10.1093/bioinformatics/btl127 (2006).
148 Lancichinetti, A. & Fortunato, S. Community detection algorithms: A comparative analysis. *Physical Review E* **80**, doi:ARTN 056117 10.1103/PhysRevE.80.056117 (2009).
149 Fortunato, S. & Barthelemy, M. Resolution limit in community detection. *P Natl Acad Sci USA* **104**, 36-41, doi:10.1073/pnas.0605965104 (2007).
150 Rosvall, M., Axelsson, D. & Bergstrom, C. T. The map equation. *The European Physical Journal Special Topics* **178**, 13-23, doi:10.1140/epjst/e2010-01179-1 (2009).
151 Brin, S. & Page, L. The anatomy of a large-scale hypertextual Web search engine. *Comput Networks Isdn* **30**, 107-117, doi:Doi 10.1016/S0169-7552(98)00110-X (1998).
152 Reuter, J. A. *et al.* Modeling inducible human tissue neoplasia identifies an extracellular matrix interaction network involved in cancer progression. *Cancer cell* **15**, 477-488, doi:10.1016/j.ccr.2009.04.002 (2009).
153 Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* **66**, 61-92, doi:10.1146/annurev.biochem.66.1.61 (1997).
154 Sakaguchi, K., Ishibashi, T., Uchiyama, Y. & Iwabata, K. The multi-replication protein A (RPA) system--a new perspective. *FEBS J* **276**, 943-963, doi:10.1111/j.1742-4658.2008.06841.x (2009).

155 Cook, A., Bono, F., Jinek, M. & Conti, E. Structural biology of nucleocytoplasmic transport. *Annu Rev Biochem* **76**, 647-671, doi:10.1146/annurev.biochem.76.052705.161529 (2007).
156 Dong, J. M., Lau, L. S., Ng, Y. W., Lim, L. & Manser, E. Paxillin nuclear-cytoplasmic localization is regulated by phosphorylation of the LD4 motif: evidence that nuclear paxillin promotes cell proliferation. *Biochem J* **418**, 173-184, doi:10.1042/Bj20080170 (2009).
157 Ma, X. T. & Hammes, S. R. Paxillin actions in the nucleus. *Steroids* **133**, 87-92, doi:10.1016/j.steroids.2017.10.012 (2018).
158 Wang, Y. & Gilmore, T. D. Zyxin and paxillin proteins: focal adhesion plaque LIM domain proteins go nuclear. *Biochim Biophys Acta* **1593**, 115-120 (2003).
159 Woods, A. J. *et al.* Paxillin associates with poly(A)-binding protein 1 at the dense endoplasmic reticulum and the leading edge of migrating cells. *J Biol Chem* **277**, 6428-6437, doi:10.1074/jbc.M109446200 (2002).
160 Acconcia, F., Barnes, C. J., Singh, R. R., Talukder, A. H. & Kumar, R. Phosphorylation-dependent regulation of nuclear localization and functions of integrin-linked kinase. *Proc Natl Acad Sci U S A* **104**, 6782-6787, doi:10.1073/pnas.0701999104 (2007).
161 Nakrieko, K. A. *et al.* Modulation of integrin-linked kinase nucleo-cytoplasmic shuttling by ILKAP and CRM1. *Cell Cycle* **7**, 2157-2166, doi:10.4161/cc.7.14.6241 (2008).
162 Kadrmas, J. L. & Beckerle, M. C. The LIM domain: from the cytoskeleton to the nucleus. *Nature reviews. Molecular cell biology* **5**, 920-931, doi:10.1038/nrm1499 (2004).
163 Freschi, V. Protein function prediction from interaction networks using a random walk ranking algorithm. *Proceedings of the 7th Ieee International Symposium on Bioinformatics and Bioengineering, Vols I and Ii*, 42-48 (2007).
164 Sharan, R., Ulitsky, I. & Shamir, R. Network-based prediction of protein function. *Molecular Systems Biology* **3**, doi:ARTN 88
10.1038/msb4100129 (2007).
165 Vazquez, A., Flammini, A., Maritan, A. & Vespignani, A. Global protein function prediction from protein-protein interaction networks. *Nature Biotechnology* **21**, 697-700, doi:10.1038/nbt825 (2003).
166 Gurzu, S., Ciortea, D., Ember, I. & Jung, I. The possible role of Mena protein and its splicing-derived variants in embryogenesis, carcinogenesis, and tumor invasion: a systematic review of the literature. *Biomed Res Int* **2013**, 365192, doi:10.1155/2013/365192 (2013).
167 Philippar, U. *et al.* A Mena invasion isoform potentiates EGF-induced carcinoma cell invasion and metastasis. *Dev Cell* **15**, 813-828, doi:10.1016/j.devcel.2008.09.003 (2008).
168 Gertler, F. B., Niebuhr, K., Reinhard, M., Wehland, J. & Soriano, P. Mena, a relative of VASP and Drosophila enabled, is implicated in the control of microfilament dynamics. *Cell* **87**, 227-239, doi:Doi 10.1016/S0092-8674(00)81341-0 (1996).

169 Burridge, K., Chrzanowska-Wodnicka, M. & Zhong, C. Focal adhesion assembly. *Trends Cell Biol* **7**, 342-347, doi:10.1016/S0962-8924(97)01127-6 (1997).

170 Hu, J. K. *et al.* An FAK-YAP-mTOR Signaling Axis Regulates Stem Cell-Based Tissue Renewal in Mice. *Cell Stem Cell* **21**, 91-106 e106, doi:10.1016/j.stem.2017.03.023 (2017).

171 Kim, N. G. & Gumbiner, B. M. Adhesion to fibronectin regulates Hippo signaling via the FAK-Src-PI3K pathway. *The Journal of cell biology* **210**, 503-515, doi:10.1083/jcb.201501025 (2015).

172 Turner, J. G. *et al.* CRM1 Inhibition Sensitizes Drug Resistant Human Myeloma Cells to Topoisomerase II and Proteasome Inhibitors both In Vitro and Ex Vivo. *J Cancer* **4**, 614-625, doi:10.7150/jca.7080 (2013).

173 Sathe, A. R., Shivashankar, G. V. & Sheetz, M. P. Nuclear transport of paxillin depends on focal adhesion dynamics and FAT domains. *J Cell Sci* **129**, 1981-1988, doi:10.1242/jcs.172643 (2016).

174 Aguilar, H. N., Tracey, C. N., Tsang, S. C., McGinnis, J. M. & Mitchell, B. F. Phos-tag-based analysis of myosin regulatory light chain phosphorylation in human uterine myocytes. *PLoS One* **6**, e20903, doi:10.1371/journal.pone.0020903 (2011).

175 Berk, J. M., Tifft, K. E. & Wilson, K. L. The nuclear envelope LEM-domain protein emerin. *Nucleus* **4**, 298-314, doi:10.4161/nucl.25751 (2013).

176 Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol Cell* **53**, 49-62, doi:10.1016/j.molcel.2013.10.030 (2014).

177 Tifft, K. E., Bradbury, K. A. & Wilson, K. L. Tyrosine phosphorylation of nuclear-membrane protein emerin by Src, Abl and other kinases. *J Cell Sci* **122**, 3780-3790, doi:10.1242/jcs.048397 (2009).

178 Arsenovic, P. T. *et al.* Nesprin-2G, a Component of the Nuclear LINC Complex, Is Subject to Myosin-Dependent Tension. *Biophys J* **110**, 34-43, doi:10.1016/j.bpj.2015.11.014 (2016).

179 Baxter, E., Windloch, K., Gannon, F. & Lee, J. S. Epigenetic regulation in cancer progression. *Cell Biosci* **4**, 45, doi:10.1186/2045-3701-4-45 (2014).

180 Hatch, E. M. & Hetzer, M. W. Nuclear envelope rupture is induced by actin-based nucleus confinement. *Journal of Cell Biology* **215**, 27-36, doi:10.1083/jcb.201603053 (2016).