Available online at www.sciencedirect.com



www.elsevier.com/locate/jprot

SILAC-iPAC: A quantitative method for distinguishing genuine from non-specific components of protein complexes by parallel affinity capture



PROTFOMICS

EuPA

Johanna S. Rees^{a,b,*}, Kathryn S. Lilley^b, Antony P. Jackson^{a,**}

^aDepartment of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK ^bCambridge Centre for Proteomics, University of Cambridge, Cambridge CB2 1QR, UK

ARTICLE INFO

Article history: Received 14 October 2014 Accepted 11 December 2014 Available online 20 December 2014

Keywords: SILAC–iPAC Protein–protein interactions Beadomes Phosphatidyl-inositol-5phosphate-4-kinase-2β Fanconi anaemia

ABSTRACT

Pull-down assays can identify members of protein complexes but suffer from co-isolation of contaminants. The problem is particularly acute when the specifically interacting partners are of low-abundance and/or bind transiently with low affinity. To differentiate true interacting partners from contaminants, we have combined SILAC labelling with a proteomic method called "Interactomes by Parallel Affinity Capture" (iPAC). In our method, a cell-line stably expressing a doubly tagged target endogenous protein and its tag-less control cell-line are differentially SILAC labelled. Lysates from the two cell-lines are mixed and the tagged protein is independently purified for MS analysis using multiple affinity resins in parallel. This allows the quantitative identification of tagged proteins and their binding partners. SILAC-iPAC provides a rigorous and sensitive approach that can discriminate between genuine binding partners and contaminants, even when the contaminants in the pull-down are in large excess. We employed our method to examine the interacting partners of phosphatidyl inositol 5-phosphate 4-kinase 2β subunit (PI5P4K2_β) and the Fanconi anaemia core complex in the chicken pre-B cell-line DT40. We confirmed known components of these two complexes, and we have identified new potential binding partners. Combining the iPAC approach with SILAC labelling provides a sensitive and fully quantitative method for the discrimination of specific interactions under conditions where low signal to noise ratios are unavoidable. In addition, our work provides the first characterisation of the most abundant proteins within the DT40 proteome and the non-specific DT40 'beadomes' (non-specific proteins binding to beads) for common epitope tags. Given the importance and widespread use of the DT40 cell-line, these will be important resources for the cell biology and immunology communities.

** Corresponding author.

E-mail addresses: jsr31@cam.ac.uk (J.S. Rees), apj10@cam.ac.uk (A.P. Jackson).

http://dx.doi.org/10.1016/j.jprot.2014.12.006

1874-3919/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Abbreviations: AP-MS, Affinity purification mass spectrometry; CS, Chicken serum; FA, Fanconi anaemia; FDR, False discovery rate; LTQ, Linear trap quadrupole; PI5P4K2β/α, Phosphatidyl inositol 5-phosphate 4-kinase 2 β/α subunits; PPI, Protein protein interactions; SILAC– iPAC, Stable isotope labelling of amino acids in culture-interactomes by parallel affinity capture; TAP, Tandem-affinity purification; UPLC, Ultra performance liquid chromatography.

^{*} Correspondence to: J. Rees, Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK.

Biological significance

SILAC-iPAC provides an improved method for the analysis of low-affinity and/or low abundance protein-protein interactions. We use it to clarify two examples where the nature of the protein complexes are known, or are currently unclear. The method is simple and quantitative and will be applicable to many problems in cell and molecular biology. We also report the first chicken beadomes.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Proteins often functionally interact as part of complex and dynamic networks [1,2]. An important goal of current proteomics research is to develop techniques that can accurately identify such protein 'interactomes' [3]. A popular method uses epitope-tagged proteins in pull-down assays. For example, tandem-affinity purification (TAP-tagging) methodology uses a 'bait' protein containing two separate epitope tags in series, separated by a protease-cleavable site. This enables sequential purification, with cleavage at the protease site used to bring about elution from the first affinity matrix column [4]. In principle, the serial nature of the purification should provide higher stringency and lower false positive rates compared to pull-downs of single tagged proteins. Unfortunately, this serial approach often compromises recovery, making yield a limiting factor and selecting for only the most stable complexes. Furthermore, the extended time course inevitably associated with TAP-tagging may lead to proteins being more vulnerable to proteolytic degradation [2,5]. These technical limitations are becoming more apparent as experimentalists increasingly examine the functional interactions of low abundance proteins, or proteins that bind transiently and with low affinity. Although cross-linking can be used to stabilise transient interactors, the shortcomings are that non-specific interactors would also cross-link, and perhaps in different proportions as the controls. Many proteins that play critical roles in cell signalling, DNA replication and other key pathways fall into these categories. Hence there is a clear need for improved methods of proteomic analysis that can be applied in such cases.

The assignment of proteins as genuine 'bait' partners is often qualitative. But as instrument sensitivity improves, it is becoming increasingly difficult to discriminate between specific and non-specific proteins from pull-downs by purely qualitative means or by simple presence/absence criteria [6]. To provide quantitation, label free methods such as spectral counting using SAINT or emPAI scoring using APEX have been adopted [7]. As an alternative, metabolic labelling of samples as early as possible offers significant improvement in precision, because it can minimise technical variability [8]. A more fundamental problem is that pull-down experiments typically generate long lists of proteins, most of which are likely to be non-specific contaminants [2,9]. Moreover, different affinity matrix resins can bind different subsets of proteins non-specifically and identifying contaminants that are selectively associated with particular resins is a problem that has only recently begun to be appreciated and addressed [2,10].

We have previously described the 'interactomes by parallel affinity capture' (iPAC) method [2,5]. Here, a doubly-tagged protein is affinity purified using each tag in separate parallel experiments. Genuine binding partners are identified by their consistent presence in both independent pull-downs [2,5]. We now combine iPAC with stable isotope labelling of amino acids in cell culture (SILAC) [11], to provide a fully quantitative characterisation of specific and non-specifically-bound proteins (SILACiPAC). Whilst AP-MS coupled with SILAC has been performed successfully [12], SILAC-iPAC offers additional precision with its combination of quantitative proteomics and multiple tag approaches, and is particularly designed for low-abundance interactors. For our proof of concept experiments, we have chosen two challenging examples where the target proteins are present in cells at only low levels, and where previous proteomic analysis has been limited or ambiguous. These are the β-subunit of phosphatidyl inositol 5-phosphate 4-kinase (PI5P4K2_B) [13], and the Fanconi anaemia core complex member FANCC [14], both in the chicken pre-B cell-line DT40.

The DT40 cell-line exhibits an extraordinarily high rate of homologous recombination. It is therefore possible to generate cell-lines containing specific gene targeted constructs. Since these constructs are regulated by their natural promoters, the epitope-tagged proteins are expressed at their physiologically normal levels [15]. The DT40 cell line JPR3 contains one PI5P4K2 β allele modified to encode a tandem C-terminal 3× FLAG tag and two ($6 \times$ His) tags [13]. PI5P4K2 β can thus be isolated using immobilised anti-(FLAG epitope) or metal affinity resin. The enzyme PI5P4K2 β is localised in the nuclear membrane, and co-assembles with its partner PI5P4K2 α [16]. The enzyme catalyses the 4 phosphorylation of PtdIns5 to generate Ptdins (4,5) P2, a lipid regulator of proteins. Based on semi-quantitative affinity purifications it is likely that there is less than 1.5 pmole of PI5P4K2 α per mg of total cell extract. This corresponds to about 9×10^4 molecules of enzyme per cell, and the level of PI5P4K2 β is probably at least 5 fold lower [16,17]. As well as the presence of homodimers, these authors report a new heterodimeric PI5P4K2 α : PI5P4K2 β association and no other significant interacting proteins as it is likely that binding of this enzyme to the nuclear membrane occurs solely via lipid interactions. We therefore use this established interaction as a positive control to test our SILAC-iPAC approach.

The Fanconi anaemia protein FANCC is a component of a multi-protein core complex that promotes homologous recombination and mutational repair of DNA interstrand crosslinks. Mutational disruption of this complex is responsible for the human chromosome breakage syndrome Fanconi anaemia (FA) [18]. The FANC complex typically exists as a low-abundance complex in growing cells. In the DT40 FANCC cell-line, the FANCC allele is modified to encode a C-terminal tandem calmodulin binding protein (CBP) and a Protein-A tag [14]. FANCC can thus be isolated using immobilised Calmodulin or IgG resin. In both of these examples, proteomic analysis represents a significant technical challenge and provides experimental opportunities to test the potential of SILAC–iPAC. Here we demonstrate that by combining stable isotope labelling with the iPAC protocol we can distinguish between true interacting partners from co-isolating contaminants.

2. Materials and methods

2.1. Cell lines, cell maintenance and harvesting

Wild type DT40, JPR3 and FANCC cells were all grown in RPMI media with L-glutamine supplemented with 10% FBS and 1% chicken serum (CS) (all GIBCO) and maintained at 37 °C at 5% CO_{2} , as described previously [19]. Cells were counted using an C6 Accuri flow cytometer and lysed in lysis buffer comprising 1× PBS (pH7.4) 1% Triton X-100, 1 mM PMSF and 1.5× EDTA free protease inhibitor cocktail (Roche).

2.2. SILAC labelling

Exponential cells were transferred to SILAC K6R6 or K0R0 RPMI media (Dundee Cell Products, UK) containing 10% dialysed FBS and 1% dialysed CS. Incorporation of ¹³C was measured by LCMS of crude lysates over 10 days and found to be approximately 98% by day 8 based on spectra from 5 proteins differing in abundance, turnover, cellular region including the tagged bait.

2.3. Affinity purifications

Initial studies to identify abundant DT40 proteins, three replicates of 10⁶ total cells were lysed in 10 ml lysis buffer on ice and the cleared supernatants separated by reducing SDS-PAGE and stained. All bands were excised and the sample prepared and analysed by MS as described in 'Sample preparation for MS/MS' and 'Mass spectrometry' below. Data was processed using Mascot (Matrix Science) and xml outputs processed using ProteinCenter Version 3.13.2003 (Thermo).

For DT40 beadomes and SILAC-iPAC experiments all cell lines were grown in parallel to 108 total cells maintained at 10⁶ cells/ml. Cells were lysed in 10 ml lysis buffer on ice and the cleared supernatants quantified, then added to 100 μ l prewashed resins. For SILAC-iPAC experiments quantified lysates were mixed 1:1 (typically 5 mg) then added to 100 μ l prewashed resins. JPR3 and wt DT40 cell extracts were purified using EZview ANTI- FLAG M2 Affinity Gel (Sigma) and TALON Metal Affinity Resin (Clontech). FANCC and wild-type DT40 cell extracts were purified with Calmodulin Sepharose 4B and IgG Sepharose 6 (both GE Healthcare). Native protein complexes were allowed to bind for 1 h then non-bound proteins removed by centrifugation at 2000 ×q, the resins washed 3 times for 15 minutes in 1 ml lysis buffer and the bound proteins eluted in 100 μ l of elution buffer. Elution buffers were lysis buffer pH 7.4 containing either 100 µg/ml 3× FLAG peptide for FLAG resin, 150 mM imidazole for TALON resins, or 200 mM EDTA for calmodulin resins. 100 mM glycine pH 2.5 was used for IgG resins. The experimental outline is depicted in Fig. 1A. Western blots were also performed using anti-tag antibodies (Sigma) or

monoclonal antibodies to FANCC, FANC-G and FANC-M (all gifts from E. Rajendra).

2.4. Sample preparation for MS/MS

Eluates were acetone precipitated overnight at -80 °C, resuspended in LDS sample buffer and resolved for ~2 cm on 10% reducing SDS-PAGE gels (Invitrogen). Analysis gels were Coomassie stained and four equal sized portions of the stained area were excised, washed, reduced in 2 mM DTT for 1 h at RT, alkylated in 10 mM iodoacetamide for 30 min at RT, and in-gel digested with 2 µg sequencing-grade porcine trypsin (Promega) overnight at 37 °C. Digests were concentrated using a speedvac and resuspended in 20 µl 0.1% formic acid.

2.5. Mass spectrometry

All LC-MS/MS experiments were performed using a nanoAcquity UPLC (Waters Corp., Milford, MA) system coupled to an LTQ Orbitrap Velos hybrid ion trap mass spectrometer (Thermo Scientific, Waltham, MA). Separation of peptides was performed by reverse-phase chromatography using a Waters reverse-phase nano column (BEH C18, 75 μ m i.d. \times 250 mm, 1.7 µm particle size) at flow rate of 300 nl/min. 10 µl peptide sample solution was initially loaded onto a pre-column (Waters UPLC Trap Symmetry C18, 180 μ m i.d. \times 20 mm, 5 μ m particle size) from the nanoAcquity sample manager with 0.1% formic acid for 3 min at a flow rate of 10 µl/min. After this period, the column valve was switched to allow the elution of peptides from the pre-column onto the analytical column. Solvent A was water + 0.1% formic acid and solvent B was acetonitrile + 0.1% formic acid. The linear gradient employed was 5-35% B in 60 min. Washes with 0.1% formic acid were performed in between each biological sample type to minimise carryover. The Velos was operated in data-dependent mode with a dynamic exclusion of 0.3 Da m/z.

The LC eluant was sprayed into the mass spectrometer by means of a nanospray source (Thermo). All *m*/*z* values of eluting ions were measured in the Orbitrap Velos mass analyser, set at a resolution of 30,000. Data dependent scans (Top 10) were employed to automatically isolate and generate fragment ions by collision-induced dissociation in the linear ion trap, resulting in the generation of MS/MS spectra. Ions with charge states of 2+ and above were selected for fragmentation.

2.6. Protein identification

Post-run, the data was processed using Protein Discoverer (version 1.2., ThermoFisher). Briefly, all MS/MS data were converted to .mgf files and these files were then submitted to the Mascot search algorithm (v 2.3, Matrix Science, London UK) and searched against the UniprotKB *Gallus gallus* database (31,529 protein entries, 2012), using a fixed modification of carbamidomethyl (C) and a variable modification of oxidation (M), allowing 2 missed cleavages, a peptide mass tolerance of 25 ppm and a fragment mass tolerance of 0.8 Da. Significance threshold was set to P > 0.01 and ion score cut off was set to 20. FDR was calculated using the reverse *Gallus* database and only proteins with an FDR of <5% were considered. In addition Mascot Percolator was used for scoring and to compare protein



Fig. 1 - SILAC-iPAC experimental outline using tagged PI5P4K2βas an example.

lists. For quantitation we used MaxQuant v 1.3 [20]. Default parameters were used with FDR of 0.01%, and a minimum of two unique and razor peptides were selected for quantifying proteins. We used the MaxQuant generated normalised data for quantitative comparison across replicates and reciprocal labelling experiments and used Microsoft excel to plot the Log₂ ratio distributions. We set a significance threshold of a ratio > +/- 1 SD of the median of the entire normalised data set. Data processing and filtering are outlined in supplemental Fig. S1. Venn diagrams were produced using http://bioinfogp.cnb.csic. es/tools/venny/index.html [21].We used STRING v 1.9 to search for known interactors with emphasis on experimental evidence [22].

3. Results

3.1. SILAC-iPAC

The SILAC-iPAC schema is summarised in Fig. 1. Here a 'bait' protein contains two different epitope tags (1 and 2) in series. However, unlike traditional TAP-tagging, there is no need to separate the epitope tags with a protease-cleavable site. Rather, a cell line expressing the double-tagged protein is grown in light (L) SILAC media, and control cells lacking tagged protein are grown in heavy (H) SILAC media. Cells are lysed, and equal amounts of proteins from the two samples mixed together. In parallel experiments, the mixed lysates are separately incubated with anti-(tag 1) and anti-(tag 2) resins. Proteins are eluted from the columns, peptides are prepared, then analysed by standard LC-MS. Peptides corresponding to

individual proteins are identified and their H/L ratios are quantified. The experiment is repeated but with reciprocal labelling, such that cells expressing doubly-tagged bait are grown in heavy SILAC media and control cells in light SILAC media. The 'bait' protein and its genuine binding partners will exhibit isotope ratios that deviate significantly from 1 in both parallel affinity purifications.

In a clean pull-down, a highly specific interaction between tag and resin will yield the tagged protein with the isotope signature corresponding to the sample labelling with no or few corresponding peptides from the control cells. In such a case, the H/L ratios cannot be reported. However with the increased sensitivity of MS instrumentation, complete presence/absence is rarely seen. In contrast, any non-specific matrix binding proteins will appear in both samples, since these proteins are expected to bind the matrix equally from tagged and untagged cells. These proteins will display H/L ratios close to 1. High abundance proteins that specifically bind to the bait may be present with isotopic peptides corresponding to both tagged and control cells. However, these proteins will display H/L ratios that deviate significantly from 1 (Fig. 1). Thus, SILAC quantitation will discriminate the genuine from non-specific binders.

It is possible that different affinity resins could potentially bind distinct patterns of non-specific proteins. However, the peptides from such proteins will still exhibit H/L isotope ratios of ~1, although in this case the proteins will be more prominently or exclusively associated with individual affinity resins. In addition, proteins that bind to the tag itself could interfere with specific binding by reducing association with the resin. Alternatively, where a low affinity non-specific protein binds the tag, a higher affinity resin-binding competitor peptide will compete off the tagged protein for that corresponding resin but not for the alternate resin, giving no ratio for the first case and a deviation of 1 in the alternate resin (depicted by the 'tag specific artefact' in Fig. 1). An outline of the data processing workflow and acceptance criteria is depicted in supplemental Fig. S1 (Fig. 3 in [23]).

3.2. The DT40 'beadome': Interaction of DT40 proteins with different affinity resins

In preliminary experiments to gauge the abundant proteins in a crude lysate and the composition of non-specific binders, we performed an MS analysis on proteins from replicate DT40 cell lysates, and identified those that were captured by each of the four different affinity resins: FLAG, TALON, IgG and calmodulin sepharose. We report the top 150 abundant proteins as ranked by emPAI score (supplemental Table S1A, Table 1 in [23]). A GO annotation was performed and identified 53 over-represented proteins that included ribosomes, viral proteins, vesicular and mitochondrial proteins. The under-represented proteins included membrane and extracellular matrix proteins (supplemental Fig. S2, Fig. 1 in [23]). Neither our cytoplasmic bait proteins PISP4K2 β and FANCC, nor any other cytoplasmic or nuclear

FANC proteins were present in this list, confirming their lower abundance.

The DT40 beadomes identified a total of 367 proteins (1.16% of the *Gallus* proteome) of which 150 (41%) bound to two or more different affinity resins (supplemental Table S1B, Table 2 in [23]). These proteins were predominantly cytoplasmic and ribosomal as classified by Gene Ontology and Kegg pathway annotations (supplemental Fig. S3, Fig. 2 in [23]). Almost half (45%) of these proteins were present in the top 150 abundant DT40 protein list highlighting the fact that abundant proteins can non-specifically bind any affinity resin and mask genuine interactors.

These initial studies provide data on the nature of potential contaminants, and further highlight the problem of discriminating specific interactors from non-specific proteins in any type of affinity purification of native complexes.

3.3. PI5P4K2*β*

To test our SILAC–iPAC method we performed the procedure on the known phosphatidyl inositol 5-phosphate 4-kinase 2 heterodimeric complex comprising PI5P4K2 α and PI5P4K2 β . Previous qualitative and semi-quantitative MS analysis of this low-abundance enzyme identified both subunits, but could not



Fig. 2 – (A) Identification of the bait proteinsPI5P4K2 β (JPR3 cell line) using FLAG MAb (i) and coomassie (ii) and FANCC using α Protein A (iii) and rabbit α FANCG (iv) and coomassie (v) in parallel pull-downs compared to control DT40 eluates (two right hand lanes). (B) PI5P4K2 peptides and (C) Fanconi peptides identified by mass spectrometry. (see Supplemental Tables S2A and B for all peptide sequences).



Fig. 3 – (A&B) Scatter plots to show the correlation of Log₂ protein ratios from a representative replicate for the PI5P4K2β pull-downs with control <u>DT40 Heavy SILAC</u> labelling, JPR3 Light labelling (DHJL) and the reciprocal labelling experiments (JHDL) respectively for FLAG (A) and TALON (B). (C) List of significant (>1SD of median) proteins from FLAG and TALON pull-downs of 4 combined replicates with green, orange and red being high, medium and low confidence respectively. (D) STRING interaction map of the significant candidates. (See Supplemental Tables S4 for full list of identified proteins and ratios.)

С			D	SCD
				EEFIAI
UniprotKB ID	GENE	Description	and the second sec	NIJC2
Q5F356;F1NKL5	PIP4K2A	Phosphatidylinositol-5-phosphate 4-kinase type-2 alpha		
F1NLK4	PIP4K2B	Phosphatidylinositol-5-phosphate 4-kinase type-2 beta	PGK1	
F1NCR3;Q5ZIZ4	NT5C2	Cytosolic purine 5-nucleotidase	101	
F1P4H4;Q5ZIM7	TXNDC5	Thioredoxin domain-containing protein 5	1491	HDLBP
F1NWB7;P08110	HSP90B1	Endoplasmin		
P00508;F1P180	GOT2	Aspartate aminotransferase, mitochondrial		
Q90593	HSPA5	78 kDa glucose-regulated protein	Dennis	EIF5A2
F1N8F0;Q9YGM2	SCD	fatty acid biosynthetic process	PGAMI	
E1C3A9;F1NNX1;P81021	HDLBP	Vigilin		(F07)
F1P3F9	GLUD1	Glutamate dehydrogenase (Fragment)	Cher Cher Cher Cher Cher Cher Cher Cher	-1420 V
Q5ZLN1;F1NHM9;F1NQ41	PGAM1	Phosphoglycerate mutase 1		\smile
E1BV44	CSE1L	Uncharacterized protein		
P00337	LDHB	L-lactate dehydrogenase B chain		HSP90B1
F1NU17;P51903	PGK1	Phosphoglycerate kinase	LDHB	GLUDI
P18359	DSTN	Destrin	G	
F1NDY9;Q5ZK20	PDIA4	Uncharacterized protein	A A A A A A A A A A A A A A A A A A A	
Q5ZKM2;Q90835	EEF1A	Elongation factor 1-alpha		PDIA4
F1NK96;F1N966;Q5F472	PDIA6	Uncharacterized protein	\smile	HSPA5
Q09121;Q07460;G1K322	EIF5A1	Eukaryotic translation initiation factor 5A-1	CBLB	TXNDC5
			1 2 2 2 2	(SXP
)	PDIA6
			Kov	- 44
			Key	
			🛰 Neighborhood	
			Sene Fusion	
			Cooccurrence	
			Coexpression	
			Experiments	
			Databases	
			Textmining	
			[Homology]	

С

Fig. 3 (continued).



Fig. 4 – (A) STRING interaction map of the known protein candidates present only associated with tagged FANCC using the Gallus gallus database. (B) Venn diagrams to show replicate reproducibility of all quantified identifications from the two parallel affinity pull-down approaches. (C) Scatter plots to show the distribution of protein ratios for the FANCC calmodulin pull-downs for DT40 Heavy SILAC labelling, FANC Light labelling (DHFL) and the reciprocal labelling experiments (FHDL) respectively and (D) IgG pull-downs that showed ratios +/- >1SD from the median ratio. (E) STRING interaction map of the newly identified protein candidates in both calmodulin IgG pull-downs. (See Supplemental Fig. S9 for tag specific interactors).



Fig. 4 (continued).

resolve the issue of other potential interactors [16]. Four biological replicates were performed for the PI5P4K2 β as well as reciprocal labellings for each, totalling 8 replicates per affinity pull-down. We were able to detect both PI5P4K2 β and its known partner PI5P4K2 α with sequence coverages of 35% and 38% respectively for FLAG pull-downs. TALON pull-downs generated fewer peptides, with maximum of 9% and 25% sequence coverages respectively, most of which were reproducible with FLAG pull-downs demonstrating the utility and importance of parallel affinity purifications. The presence of bait protein and resulting peptides by MS is shown in Fig. 2, supplemental Fig. S4 and supplemental Table S2A.

Consistent with its low abundance, we could not quantify PI5P4K2 β in the first two replicates. However upon increasing cell numbers, we were able to quantify both isoforms. We therefore used datasets 3 and 4 for mining interactors. We found good correlation of SILAC ratios when comparing the replicates and reciprocals (Fig. 3A, B) and for FLAG pull-downs both PI5P4K2 β and PI5P4K2 α had statistically significant ratios (>2 SD from the median) as well as one other protein NT5C2, a cytosolic purine 5-nucleotidase. The TALON data showed alternate proteins that were of greatest significance with the PI5P4K2 being only 1SD from the median ratio. Of the 218 FLAG and 348 TALON purified proteins 84 proteins were identified using the two different affinity tags. We took all proteins with >1 SD of median from each of the four FLAG and TALON datasets (Supplemental Table 3A&B, Table 3 in [23]) and found that only 18 were in common for both pull-downs. These are listed in Fig. 3C and STRING interaction mapping was performed to analyse the validity of these (Fig. 3D). STRING clearly showed both the PI5P4K2 β and PI5P4K2 α isoforms as well as NT5C2, TXNDC5 and SCD as non-interaction proteins distinct

from a large cluster of mostly metabolic related proteins and a smaller cluster of protein processing proteins (PDIAs and chaperones). To check the validity of these non-interacting proteins we expanded the string network to include increments of ten additional known interactors (Supplemental Fig. S6) and still did not find a connecting protein for PI5P4K2 β and PI5P4K2 α thus concluding that these isoforms do not interact with any other proteins. For proteins NT5C2, TXNDC5 and SCD we have ranked these as high, medium or low confidence based on the identification and ratio reproducibility amongst replicated, reciprocal and tags (Fig. 3D). NT5C2, a cytosolic protein may have a critical role in the maintenance of a constant composition of intracellular purine/pyrimidine nucleotides in cooperation with other nucleotidases, and preferentially hydrolyses inosine 5'-monophosphate (IMP) and other purine nucleotides. TXNDC5 is an ER membrane protein that has protein disulfide isomerase activity. This was significant across all experiments and will need further validating but is unlikely to be biologically meaningful. SCD, an ER protein, is involved in fatty acid biosynthesis but the ratios were not consistent so was ranked as low.

The enzyme PI5P4K2 acts on lipid substrates, and previous experimental findings indicate no other interactors in this complex [16]. Hence for the case of PI5P4K2, our exhaustive SILAC-iPAC data has successfully identified the one known and biologically relevant binding partner of PI5P4K2 β subunit against a large excess of non-specific interactors. Moreover, our data strongly implies that PI5P4K2 does not interact with any other protein. Hence, the enzyme is most likely is localised to the nuclear membrane *via* protein–lipid interactions. When comparing our data with the traditional non quantitative iPAC approach (by searching Mascot with no quantitation) the PI5P4K2 proteins were not ranked highly (120 of 197 hits by Mascot) in identification because of the mass of non-specific proteins but with quantitation we were able to successfully enrich for these and classify most co-purifying proteins as non-specific contaminants. In conclusion there is no evidence that PI5P4K2 proteins interact with other proteins in this lipid environment, consistent with previous reports.

3.4. FANCC

Here we carried out two reciprocal labellings, totalling 4 replicates per affinity tag. The bait FANCC and known Fanconi core complex members, FANC-A, B, E, F, G, L, M and FAAP100 were identified in combined replicate datasets for both parallel pull-downs (Fig. 2, supplemental Table S2B). The two different affinity purifications resulted in some different peptides being observed, with IgG pull-downs yielding higher percentage sequence coverage of 8.4% for the bait FANCC (Supplemental Fig. S5).

In these experiments, we looked at the FANC complex present in rapidly and exponentially growing cells in which DNA damage had not been artificially initiated. We searched our lists for proteins known to be involved in DNA repair [24,25] and found good representation of known proteins including FANCD2, FANCI, FANCJ, (with 7%, 6.5% and 5.8% sequence coverage respectively). A comprehensive list of known FANC interactors is presented in supplemental Table S3C as well as a STRING interaction map to indicate all experimental and predicted FANC interactions in *G. gallus* (Fig. 4A). Not all were quantified due to low peptide hits in the individual replicates, and those that were, CHEK1; CDK5; UBE2V2, were not statistically significantly enriched for.

To search for new interactors we analysed the 837 of 1126 quantified proteins that were common to calmodulin and IgG resins (Fig. 4B) and present the three largest datasets for each affinity resin (Fig. 4C&D). As expected the majority of proteins cluster around the 1:1 ratio, corresponding to non-specific interactors, but there were also some notable proteins common to both tags which appeared as highly significant that lay >1SD from the median ratio (Supplemental Table S4A, Table 4 in [23]). Of particular interest were eight nuclear proteins involved in DNA damage, including RUVBL1 (44% sequence coverage by MS). We used STRING to assess the validity of these proteins and showed co-expression evidence of RUVBL1 with WBSC22, RCC, NPM1 and NPM3 (Fig. 4E). The peptide count and % sequence coverage of these proteins was significant (>2 unique and >10% respectively) confirming their existence. These are indeed all nuclear localised proteins and may be newly identified interactors, not previously identified in Gallus. We also searched these using the equivalent human identifiers and found experimental evidence of NPM1 binding both FANCA and FANCC (Supplemental Fig. 9A). RUVBL1 has been shown to interact with RUVBL2 in other species so we searched our MS data against a bovine database and found up to 19% protein coverage in all replicates with evidence of both heavy and light peptide pairs that would have resulted in quantitation had RUVBL2 not been absent from the Gallus database. With our new quantitative evidence of NPM1 and RUVBL1 interacting with FANCC we have identified a new family of interactors involved in DNA repair in Gallus (Fig. 4E).

In addition to comparing the reproducibility of all identifications across the four replicates for both Calmodulin and IgG pull-downs (Fig. 4D) we also looked at statistically significant proteins that were observed with only one or other resin (supplemental Tables S4B&C). Although some nuclear proteins were present, most were cytosolic, endomembrane or mitochondrial, and thus unlikely to be biologically relevant. STRING mapping showed experimental evidence of only a few proteins directly interacting, via FANC-A: brg1/SMARCA4a, a SWI/SNF-related matrix-associated actin-dependent regulator of chromatin proteins for Calmodulin resin (Supplemental Fig. S9B) and RPA1 (Ras related) that in turn binds SSRP1, a recombination signal sequence recognition protein 1 for IgG resin (Supplemental Fig. S9C). These were reproducible and statistically significant and show that not all resins can capture genuine interactors demonstrating that multiple AP-MS approaches are required to capture all complex members.

Comparing these significant lists with those obtained using traditional iPAC (by processing data by Mascot with no quantitation) only NPM1 of the newly identified proteins was found and it was also present in the negative control so would have been excluded by iPAC due to the similar numbers of peptides demonstrating that quantitation is necessary.

4. Discussion

A major technical challenge in proteomics is how to detect low-abundance binding interactions amid a large excess of non-specific proteins. The SILAC-iPAC method is designed to address this question. Firstly, we use multiple affinity capture carried out in parallel. This approach offers the advantages of reduced incubation times and potentially higher recoveries compared to serial methods such as TAP tagging. Secondly, as technical variability can often be high with AP-MS experiments [26] we use the quantitation provided by SILAC labelling to provide confident discrimination between specific and nonspecific interactors. In an ideal and totally clean pull-down, a highly specific interaction between tag and resin will yield the tagged protein that is labelled with either only light or heavy isotopes from non-labelled or SILAC labelled cells respectively. It was in these lists where we found some of our more interesting and potentially biologically relevant proteins for the Fanconi complex. However, high abundance proteins may appear in both samples, and here quantitation will be essential to discriminate the genuine from non-specific binders.

During the course of the work, we became aware of an additional, and under-appreciated source of error in traditional SILAC AP-MS methods [27]. Not all affinity resins bind to the same sets of non-specific proteins. This can make it hard for genuine low abundant binders to compete for an immobilised ligand, thus making quantitation difficult. The severity of the problem varies from resin to resin. Our approach has been to compare proteins from multiple parallel affinity purifications so that we can be sure that we are investigating genuine interactors. In our experiments, we detected some protein with a statistically significant ratio from one resin but not from an alternate resin. We even found statistically significant proteins that were only associated with tagged bait from one affinity resin, but present in both specific and control pull-downs with the second resin with non-significant ratios (depicted in Fig. 1). Such proteins would seem genuine if only a single tag AP experiment is performed. Our data confirm that they are really non-specific interactors and of significant concern. They represent a class of non-specific artefacts that have not been previously emphasised in traditional pull-down experiments.

We deliberately chose two contrasting examples whose interactomes are currently known or uncertain: PIP5P4K2p and FANCC. In the case of PI5P4K2 β , other than its homodimeric association in the nucleus, the only other established partner is PI5P4K2 α . Under these conditions, PI5P4K2 β and its known binding partner PI5P4K2α were successfully identified. The PI5P4K2^β pull-downs were highly specific, so that initial experiments failed to detect any PI5P4K2^B peptides in the control sample. Only when using two fold higher cell preparations were some non-tagged PI5P4K2^B peptides detectable in the mixed SILAC samples. The ratio of PI5P4K2_B: PI5P4K2 α , calculated at 1.28 was consistent with 1.36 using an internal labelled standard approach [16]. Importantly we were able to detect these interactions using only 1×10^8 cells where previous attempts had to use 2×10^{10} cells [16]. No other proteins in the pull-downs were reproducibly detected, and none had ratios that were statistically different from 1:1. We have confirmed this interaction using two independent tags and quantitatively proven that the remaining candidate proteins identified by MS across 16 replicates are, in fact, non-specific contaminants and/or non-reproducible or perhaps transient interactors. Thus there still remains no experimental evidence that PI5P4K2 β stably binds to any protein other than PI5P4K2 α in DT40 cells. The interaction of PI5P4K2 with the nuclear membrane is therefore likely to occur via protein-lipid interactions, rather than protein-protein interactions.

The Fanconi complex was more challenging to address. The complex is prominently associated with stalled replication forks, so isolating enough bait with intact native complexes can be limiting. The FA core is composed of a large multi-protein nuclear complex comprised of the proteins mentioned [18]. Within the core complex are sub-complexes, A&G, C, E & F and B, L and FAAP100 proteins, and together with M, FAAP proteins 10, 16, 20, and 24 this core complex ubiquitinates D2 and FANCI. From multiple replicates we were able to identify not only the Fanconi core groups A, B, C, E, F, G, J, L and M but also their substrates D2 and I, and downstream players FANCD1, N, O, P & Q. A key objective has been to identify new interacting partners for this complex. These interactors may in themselves be new FA proteins and may provide further mechanistic insight into how the pathway works. We have identified proteins involved in the DNA repair pathway, namely kinases ATM, ATR, PRKDC, CHK1&2; BRCA1&2 and known interactors FAM175, BRE, PALB2 (FANCN), and BLM; endonucleases ERCC4, EME1 and SLX4 (FANCP); CDK5 and its regulatory proteins CDKRAP1&2; ubiquitin conjugating enzyme E2 family Ube2T; and FAN1. Many of these are transient and/or low abundance proteins, and present at different stages of the pathway, and were all enriched for in our pull-downs.

Addressing abundant proteins using SILAC again allowed us to distinguish the genuine binders from the mass of non-specific proteins. Interesting DNA-binding proteins that were common in both affinity methods were RUVBL1, a TATA box-binding protein-interacting protein, WBSCR22 involved in DNA methylation, RCC2, a telophase disk protein and NMP1 (Nucleophosmin), a nucleolar phosphoprotein B23 involved in genomic stability and DNA repair amongst other roles. STRING showed evidence of RUVBL1 interacting with all these proteins (Fig. 4). Importantly RUVBL1 has recently been independently characterised to interact with the FA complex, serving a role in DNA crosslink repair using functional readouts [28]. It should be emphasised that our method detected RUVBL1 with up to 10³ fold lower cell preparations. Therefore, SILAC-iPAC offers significant practical advantages in terms of sensitivity and quantitative analysis. We also found equivalent proportions of the RUVBL1 interacting partner RUVBL2 when searching our data using a bovine database. This was also functionally confirmed by Rajendra et al. [28]. We also identified ATPase subunits which according to STRING are indirect interactors of RUVBL1-FA through a cluster of ribosomal proteins, all of which were abundant proteins in our DT40 parts lists. When searching our list against human interactions we found evidence of both FANCA and C binding NPM1 that co-purifies with EIF2S2, which interacts with many ribosomal proteins that also associate with ATP synthesis machinery. Other enriched proteins such as CAPRIN1, a cell cycle associated protein, and G3BP1, a DNA-unwinding enzyme, are of potential biological interest and will have to be independently confirmed.

With two different complex models being studied here we were interested in seeing if any identified and significant proteins appeared in each other's pull-down lists (Fig. 5A and B respectively). Using data from four replicates each for FLAG, TALON, Calmodulin and IgG pull-downs, of all identified and quantified proteins 181 (11.2%) were present in all four different affinity pull-downs with a further 229 (14.2%) present in three resins and 68 (4.2%) from two resins that were not from the same bait. In comparison 30.9% of all proteins were from parallel datasets with the same tag (Fig. 5A). Taking only the proteins that had at least one significant ratio in any one replicate, the proportion of significant proteins in both the parallel pull-downs was vastly increased by comparing only statistically significant ratios and the overlap across all resins was greatly reduced (Fig. 5B) and comprised only PDIA in all four resins and ATP5H, NPM3 and RUVBL1 in three of four, and in these three cases the proportion of significant ratios was higher in the FANCC pull-downs where we found published experimental evidence of interactions. It is possible that chaperones such as PDIA are recruited to endogenously tagged proteins to assist in folding. This highlights the fact that even stringent filtering measures can result in promiscuous proteins supposedly interacting and hence the need for published beadomes [2,10]. Without suitable antibodies for such a large number of candidates it would be unfeasible to prove if they are indeed genuine.

Lastly we compared the statistically significant proteins from both complex models to the abundant DT40 protein list and the lists of sticky proteins observed with two or more different resins (beadome) (Fig. 5C) and reassuringly no proteins were common to all sets. Of the significant FANCC candidates 85% were unique to FANCC pull-downs and not identified in any abundant or



Fig. 5 – Venn diagrams to show the comparison of interaction proteins from two different complexes, PI5P4K2 and FANC, (A) total hits, (B) statistically significant, and (C) a comparison with the DT40 abundant proteins and previously identified non-specific binders to \geq 2 of the four resins used (beadome).

beadome lists. The significant RUVBL1 was absent from abundant or beadome lists but did co-purify with PI5P4K2 β from FLAG resins. NPM1 pulled down with FANCC was present in the abundant and beadome lists but they were reassuringly absent from the PI5P4K2 β lists. Most other proteins in common to both complexes and the beadome were protein processing proteins, such as chaperones, that could be genuine transient interactors whilst the bait proteins are synthesised and trafficked.

Our initial survey to ascertain abundant DT40 proteins from a crude lysate to establish likely non-specific contaminants showed that almost half were reproducibly identified in the four DT40 'beadomes'. These individual beadomes will be useful resources for the community in planning targeted PPI studies and have been deposited in the forthcoming release of the CRAPome repository. These comprise the first reports of chicken data and findings are consistent with other beadome lists from human yeast and Drosophila.

Although the chicken DT40 cell line has been extensively used in performing homologous recombination to endogenously tag proteins of choice, one major disadvantage is that the chicken genome and proteome databases are poorly maintained and annotated, and there is an unacceptable level of redundancy. This makes the identification of unique peptides unnecessarily problematic. We had to use the UniprotKB database with TrEMBL entries as some of our bait and known interactors did not have SwissProt entries and this significantly increases proteome redundancy. For example, the UniprotKB 2012 release contained 31,529 protein entries and the 2013 release had a 7% increase to 33,804 entries for only 16,294 genes. Approximately 20,000 (~60%) of these proteins were computationally predicted. Similarly the 'reference proteome' with 23,395 Gallus only entries comprise only 9.6% SwissProt entries and the remainder are TrEMBL of which most are predicted. In addition a significant proportion of protein entries in the UniprotKB database are represented by multiple accession numbers with one example having 20 that included fragments of the same protein. Whilst various quantitation software packages have the option of grouping similar proteins according to homology it was difficult to quantify, for example, the distinct PI5P4K2 β and α isoforms due to their high sequence

similarity, thereby making reliable quantitation challenging. In addition when investigating genuine binding partners that did not appear in controls, protein scoring can be misleading with a redundant database. In addition to redundancy there is evidence of incompleteness such as the RUVBL2 protein that we identified with 25% coverage when searching other organisms' databases. With an increasing number of researchers using the DT40 cell line and chicken as a model organism for immunology and embryology and in analysis of human disease [9,29], efforts must be made to consolidate and update these databases.

Compared to other methods SILAC is generally the method of choice for quantitative AP-MS experiments [3] and allows proteins with ratios deviating from 1:1 to be pursued. Whilst many other quantitative interactome studies using AP combined with other isobaric tags such as iTRAQ and TMT can discriminate between specific interacting partners and non-specific binding contaminants, the addition of parallel APs combined with SILAC provides another level of assurance to our quantitative method. Our results indicate that SILAC-iPAC will be useful for determining genuine complex binding partners as well as cataloguing non-specific proteins that bind to specific or multiple affinity resins. It can be used for cultured cells, or organisms that can be metabolically labelled that have proteins tagged either endogenously or exogenously. This will make the method widely applicable to the proteomics community.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.12.006.

Conflict of interest

None

Acknowledgements

We thanks Prof R. Irvine for providing the JPR3 cell line, to Dr E. Rajendra for providing the FANCC cell line, FANC antibodies and helpful discussions, to Dr M. Deery and J. Howard for assistance with MS. This work was funded by the Biotechnology and Biological Sciences Research Council (UK) grant BB/H024085/1.

REFERENCES

- [1] Figeys D. Mapping the human protein interactome. Cell Res 2008;18:716–24.
- [2] Rees JS, Lowe N, Armean IM, Roote J, Johnson G, Drummond E, et al. In vivo analysis of proteomes and interactomes using Parallel Affinity Capture (iPAC) coupled to mass spectrometry. Mol Cell Proteomics 2011;10(M110):002386.
- [3] Paul FE, Hosp F, Selbach M. Analyzing protein–protein interactions by quantitative mass spectrometry. Methods 2011;54:387–95.
- [4] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17:1030–2.

- [5] Lowe N, Rees JS, Roote J, Ryder E, Armean IM, Johnson G, et al. Analysis of the expression patterns, subcellular localisations and interaction partners of Drosophila proteins using a pigP protein trap library. Development 2014; 141:3994–4005.
- [6] Armean IM, Lilley KS, Trotter MW. Popular computational methods to assess multiprotein complexes derived from label-free affinity purification and mass spectrometry (AP-MS) experiments. Mol Cell Proteomics 2013;12:1–13.
- [7] Nahnsen S, Bielow C, Reinert K, Kohlbacher O. Tools for label-free peptide quantification. Mol Cell Proteomics 2013; 12:549–56.
- [8] Russell MR, Lilley KS. Pipeline to assess the greatest source of technical variance in quantitative proteomics using metabolic labelling. J Proteomics 2012;77:441–54.
- [9] Trinkle-Mulcahy L, Boulon S, Lam YW, Urcia R, Boisvert FM, Vandermoere F, et al. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. J Cell Biol 2008;183:223–39.
- [10] Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. Nat Methods 2013;10: 730–6.
- [11] Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002;1:376–86.
- [12] Altelaar AF, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HT, et al. Benchmarking stable isotope labeling based quantitative proteomics. J Proteomics 2013; 88:14–26.
- [13] Richardson JP, Wang M, Clarke JH, Patel KJ, Irvine RF. Genomic tagging of endogenous type IIbeta phosphatidylinositol
 5-phosphate 4-kinase in DT40 cells reveals a nuclear localisation. Cell Signal 2007;19:1309–14.
- [14] Mosedale G, Niedzwiedz W, Alpi A, Perrina F, Pereira-Leal JB, Johnson M, et al. The vertebrate Hef ortholog is a component of the Fanconi anemia tumor-suppressor pathway. Nat Struct Mol Biol 2005;12:763–71.
- [15] Buerstedde JM, Takeda S. Increased ratio of targeted to random integration after transfection of chicken B cell lines. Cell 1991;67:179–88.
- [16] Wang M, Bond NJ, Letcher AJ, Richardson JP, Lilley KS, Irvine RF, et al. Genomic tagging reveals a random association of endogenous PtdIns5P 4-kinases IIalpha and IIbeta and a partial nuclear localization of the IIalpha isoform. Biochem J 2010;430:215–21.
- [17] Irvine RF. Nuclear lipid signalling. Nat Rev Mol Cell Biol 2003; 4:349–60.
- [18] Joenje H, Patel KJ. The emerging genetic and molecular basis of Fanconi anaemia. Nat Rev Genet 2001;2:446–57.
- [19] Hall SL, Hester S, Griffin JL, Lilley KS, Jackson AP. The organelle proteome of the DT40 lymphocyte cell line. Mol Cell Proteomics 2009;8:1295–305.
- [20] Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, et al. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. Nat Protoc 2009;4:698–705.
- [21] Tamames J, Clark D, Herrero J, Dopazo J, Blaschke C, Fernandez JM, et al. Bioinformatics methods for the analysis of expression arrays: data clustering and information extraction. J Biotechnol 2002;98:269–83.
- [22] Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein–protein interaction networks, with increased coverage and integration. Nucleic Acids Res 2013;41:D808–15.
- [23] Rees JS, Lilley KS, Jackson AP. The Chicken B cell line DT40 proteome, beadomes and interactomes. Data in Brief; 2015[in press].

- [24] Alpi AF, Pace PE, Babu MM, Patel KJ. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. Mol Cell 2008;32:767–77.
- [25] MacKay C, Declais AC, Lundin C, Agostinho A, Deans AJ, MacArtney TJ, et al. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. Cell 2010;142:65–76.
- [26] Dazard JE, Saha S, Ewing RM. ROCS: a reproducibility index and confidence score for interaction proteomics studies. BMC Bioinforma 2012;13:128.
- [27] Hubner NC, Bird AW, Cox J, Splettstoesser B, Bandilla P, Poser I, et al. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. J Cell Biol 2010;189:739–54.
- [28] Rajendra E, Garaycoechea JI, Patel KJ, Passmore LA. Abundance of the Fanconi Anaemia core complex is regulated by the RuvBL1 and RuvBL2 AAA+ ATPases. Nucleic Acids Res 2014;42:13736–48.
- [29] Kohonen P, Nera KP, Lassila O. Avian model for B-cell immunology—new genomes and phylotranscriptomics. Scand J Immunol 2007;66:113–21.