Protein context shapes the specificity of SH3 domain-mediated interactions in vivo

Ugo Dionne1,2,3,4,5, Émilie Bourgault3,4,5,6,7,13, Alexandre K. Dubé3,4,5,6,7,13, David Bradley3,4,5,6,7,13, François J. M. Chartier1,2,3,13, Rohan Dandage3,4,5,6,7, Soham Dibyachintan3,4,6,7,8, Philippe C. Després3,4,5,6, Gerald D. Gish9, N. T. Hang Pham3,10, Myriam Létourneau3,10, Jean-Philippe Lambert1,2,11, Nicolas Doucet3,10, Nicolas Bisson1,2,3,12✉ & Christian R. Landry3,4,5,6,7✉

Protein–protein interactions (PPIs) between modular binding domains and their target peptide motifs are thought to largely depend on the intrinsic binding specificities of the domains. The large family of SRC Homology 3 (SH3) domains contribute to cellular processes via their ability to support such PPIs. While the intrinsic binding specificities of SH3 domains have been studied in vitro, whether each domain is necessary and sufficient to define PPI specificity in vivo is largely unknown. Here, by combining deletion, mutation, swapping and shuffling of SH3 domains and measurements of their impact on protein interactions in yeast, we find that most SH3s do not dictate PPI specificity independently from their host protein in vivo. We show that the identity of the host protein and the position of the SH3 domains within their host are critical for PPI specificity, for cellular functions and for key biophysical processes such as phase separation. Our work demonstrates the importance of the interplay between a modular PPI domain such as SH3 and its host protein in establishing specificity to wire PPI networks. These findings will aid understanding how protein networks are rewired during evolution and in the context of mutation-driven diseases such as cancer.

1 Centre de Recherche du Centre Hospitalier Universitaire (CHU) de Québec-Université Laval, Québec, QC, Canada. 2 Centre de Recherche sur le Cancer de l’Université Laval, Québec, QC, Canada. 3 PROTEO-Regroupement Québécois de Recherche sur la Fonction, l’Ingénierie et les Applications des Protéines, Québec, QC, Canada. 4 Centre de Recherche en Données Massives de l’Université Laval, Université Laval, Québec, QC, Canada. 5 Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada. 6 Département de Biochimie, de Microbiologie et de Bio-Informatique, Université Laval, Québec, QC, Canada. 7 Département de Biologie, Université Laval, Québec, QC, Canada. 8 Département de Chimie, Université Laval, Québec, QC, Canada. 9 Département de Chimie, Université Laval, Québec, QC, Canada. 10 Department of Chemical Engineering, Indian Institute of Technology Bombay (IIT), Powai, Mumbai, Maharashtra, India. 11 Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Joseph and Wolf Lebovic Health Complex, Toronto, ON, Canada. 12 Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique (INRS), Université du Québec, Laval, QC, Canada. 13 These authors contributed equally: Émilie Bourgault, Alexandre K. Dubé, David Bradley, François J.M. Chartier. ✉ email: Nick.Bisson@crchudequebec.ulaval.ca; christian.landry@bio.ulaval.ca
Proteins often display a modular architecture defined by folded domains that bind short linear peptide motifs on their interaction partners. Modular domains are generally considered to act as “beads on a string” by virtue of their ability to independently fold and bind target peptides with high intrinsic specificity in vitro. However, binding domains are often part of larger proteins that can comprise many functional elements. Whether/how PPI domain binding specificity is modulated by positioning within their host protein and/or intramolecular interactions (collectively defined here as protein context) remains poorly defined. Such regulation of interaction specificity would imply that during evolution and in disease states, mutations occurring either within or outside a modular binding domain could alter its protein interaction specificity. We examined this question by studying PPIs of proteins containing SRC Homology 3 domains (SH3s) in vivo. SH3s are one of the most prevalent families of modular binding domains, having expanded in number throughout evolution with 27 in yeast (on 23 proteins) and nearly 300 in humans. These ~60 amino acid domains are present on signaling proteins, regulating functions such as endocytosis and actin cytoskeleton remodeling. SH3s typically bind to Pro/Arg-rich peptide motifs on their target partners with an archetypical PXXP motif (where X represents any amino acid).

In this work, we combine genome editing, cellular phenotyping, and proteomics to determine the in vivo contribution of protein context to SH3 domain specificity and functions. We find that these PPI modules rarely mediate interactions independently and that their position within their host is important for biological processes such as endocytosis and phase separation. Our results contribute to the current understanding of how PPI networks achieve specificity and how it is altered by mutations, domain gains, and losses.

Results and discussion

SH3s contribute to PPI networks complexity and protein function in vivo. To assess the requirement of SH3 domains for PPIs in vivo, we first measured binary interactions between 22 WT SH3-containing proteins (Supplementary Fig. 1A) as baits and 575 putative partners from their interconnected signaling networks using the dihydrofolate reductase protein–fragment complementation assay (DHFR-PCA) in the budding yeast Saccharomyces cerevisiae. About 33% (202/607) of the PPIs detected were described before, mostly by direct methods (Supplementary Fig. 1B, C). We repeated the experiments with baits in which the SH3s were individually replaced with a flexible linker by genome editing (SH3 domain deletion/stuffing, Fig. 1A). The stuffer DNA sequence (GGCCGAAGTCCTGGAGGTGGTGGT) codes for a flexible poly-Gly with Ser peptide (GGSSGGGG) that is inspired by previous experiments. About a third of the SH3-containing protein interactome is functional and nearly 300 in human, whereas 27 in yeast (on 23 proteins) have the highest sequence identity (human orthologs of SH3s (CTTN, HCLS1, and DBNL, Fig. 2B). The reintroduction of its own SH3 (Abp1SH3 in Abp1) reconstitutes almost perfectly Abp1’s interaction profile (Kendall’s τ = 0.93, Fig. 2B). For most cases where Abp1 loses many of its PPIs, Abp1 expression level was not significantly affected by SH3 swapping (Supplementary Fig. 2A, B). A notable exception is for instance Bem1SH3,2 in Abp1, which leads to the loss of a large number of interactions and shows reduced abundance. This suggests a complex interplay between SH3 domains and their host proteins. No homologous SH3 domain fully re-establishes the normal Abp1 PPI profile (Fig. 2B and Supplementary Fig. 2C). However, we observe a significant correlation (sophisticated correlation = 0.20, p value = 0.005, Fig. 2C) between the similarity of PPI profiles (Fig. 2B) and the sequence similarity of the SH3s. For instance, Abp1 swapped with SH3s from its human orthologs, which have the highest sequence identity (human orthologs 46–52%, other yeast SH3s 19–41%), displays PPI profiles most strongly correlated with WT Abp1 (Kendall’s τ: CTTN = 0.89, DBNL = 0.89, and HCLS1 = 0.86, Fig. 2B and Supplementary Fig. 2C). Nonetheless, a subset of SH3-dependent PPIs is only observed with the endogenous domain. For instance, Hua2 and App1, both well-characterized partners of Abp1, are only detected when Abp1 contains its own SH3 (WT or swapped via editing). A given SH3 may therefore not be fully replaceable with other paralogous or orthologous domains. This pattern is confirmed by cellular growth phenotypic analyses (Supplementary Fig. 2D, E). PPIs, SH3 sequence similarity, and growth profiles under stress conditions are correlated (Supplementary Fig. 2D, E). This relationship is particularly clear when analyzing growth on media supplemented with hygromycin. Indeed, Abp1 deletion

SH3s are rarely sufficient to dictate the breadth of PPIs driven by their host. Having determined that multiple PPIs require SH3s, we asked whether the latter establish PPI specificity independently from their host proteins by swapping domains (Fig. 2A). Using the second round of genome editing, we individually replaced Abp1 SH3 with the 27 yeast SH3s and 3 human Abp1 orthologs’ SH3s (CTTN, HCLS1, and DBNL, Fig. 2B). The reintroduction of its own SH3 (Abp1SH3 in Abp1) reconstitutes almost perfectly Abp1’s interaction profile (Kendall’s τ = 0.93, Fig. 2B). For most cases where Abp1 loses many of its PPIs, Abp1 expression level was not significantly affected by SH3 swapping (Supplementary Fig. 2A, B). A notable exception is for instance Bem1SH3,2 in Abp1, which leads to the loss of a large number of interactions and shows reduced abundance. This suggests a complex interplay between SH3 domains and their host proteins. No homologous SH3 domain fully re-establishes the normal Abp1 PPI profile (Fig. 2B and Supplementary Fig. 2C). However, we observe a significant correlation (sophisticated correlation = 0.20, p value = 0.005, Fig. 2C) between the similarity of PPI profiles (Fig. 2B) and the sequence similarity of the SH3s. For instance, Abp1 swapped with SH3s from its human orthologs, which have the highest sequence identity (human orthologs 46–52%, other yeast SH3s 19–41%), displays PPI profiles most strongly correlated with WT Abp1 (Kendall’s τ: CTTN = 0.89, DBNL = 0.89, and HCLS1 = 0.86, Fig. 2B and Supplementary Fig. 2C). Nonetheless, a subset of SH3-dependent PPIs is only observed with the endogenous domain. For instance, Hua2 and App1, both well-characterized partners of Abp1, are only detected when Abp1 contains its own SH3 (WT or swapped via editing). A given SH3 may therefore not be fully replaceable with other paralogous or orthologous domains. This pattern is confirmed by cellular growth phenotypic analyses (Supplementary Fig. 2D, E). PPIs, SH3 sequence similarity, and growth profiles under stress conditions are correlated (Supplementary Fig. 2D, E). This relationship is particularly clear when analyzing growth on media supplemented with hygromycin. Indeed, Abp1 deletion
leads to hygromycin resistance\(^2\) and this phenotype is dependent on Abp1 SH3 (Supplementary Fig. 2E). None of the swapped SH3s, except for the Abp1SH3 in Abp1, fully reproduces WT sensitivity, confirming our observation with PPI patterns (Fig. 2B and Supplementary Fig. 2E). However, the orthologous human SH3s CTTN\(_{SH3}\), HCLS1\(_{SH3}\), and DBNL\(_{SH3}\) in Abp1 display intermediate phenotypes; this is consistent with PPI profiles and sequence similarity clusters (Fig. 2B, C and Supplementary Fig. 2E).

Several SH3 swaps lead to an inhibition of >50% of Abp1 PPIs, including SH3-independent interactions (Fig. 2B and Supplementary Fig. 2C). This suggests that SH3s can affect binding that is mediated by other regions of the protein, most likely through allosteric effects. In addition, two-thirds (21/30) of SH3 swaps lead to gains of PPIs that were not detected with WT Abp1 (Fig. 2B and Supplementary Fig. 2C). As expected, some SH3s can bring a subset of their SH3-dependent PPIs to the Abp1 protein context; for example, Sho1\(_{SH3}\) and Nbp2\(_{SH3}\) in Abp1 promote the interaction with Pbs2\(^{21-23}\) (Fig. 2D and Supplementary Fig. 2F). However, the majority of the gained PPIs were not identified in our SH3 deletion screen as being SH3-dependent (Fig. 2D). Thus, most SH3s are not sufficient to establish their endogenous specificity into a new protein context (Supplementary Fig. 2F). Consistent with these results, partners gained by Abp1 SH3-swapped proteins are not enriched for SH3-specific binding motifs relative to unaffected Abp1 PPIs (\(p = 0.52\), Mann-Whitney test, one-sided, Supplementary Fig. 2G).
We also examined whether the ability of a given SH3 to mediate PPIs depends on its position within the same host protein. We individually swapped Abp1 SH3 into each of Sla1’s three SH3 positions (Fig. 2E). Inserting Abp1 SH3 at either of the first two positions only slightly affects Sla1 PPIs (Fig. 2E). SH3 swapping at the third position led to the detection of seven new PPIs despite our observation that only two Sla1 partners depend on this third SH3 (Fig. 1B, E). None of the PPIs gained by Sla1 following Abp1 SH3 swapping were originally found to be dependent on Abp1 SH3, further supporting our finding that the ability of a domain to dictate its host PPI partners is highly dependent on the identity of the host and the position of the SH3. Overall, the observation that SH3s are rarely dictating their host PPI partners by themselves, but rather alter PPIs in a manner that cannot be predicted from their intrinsic specificity suggests the presence of complex interactions between an SH3 domain and its host protein.

Allosteric effects of SH3 domains on SH3-independent interactions. The above analysis revealed that domain swapping impacts PPIs in a sequence-dependent manner, with divergent SH3 domains having the strongest effects. Swapping SH3s into a protein also affects PPIs that were not previously found to be SH3-dependent, suggesting that SH3s can alter PPIs allosterically. To systematically investigate the distinction between sequence-dependent effects on SH3-dependent and SH3-independent PPIs, we measured the binding of Abp1 to an SH3-independent (Lsb3) and to an SH3-dependent partner (Hua2) for all possible single mutants of Abp1 SH314,19 (Fig. 3A–D and Supplementary Fig. 3A–E). We validated that mutating the Abp1 SH3 domain has little effect on its abundance (Supplementary Fig. 3C, D). Mutation sensitivity profiles for the two targets are overall highly correlated (Kendall’s τ = 0.59, p value = 1.0 × 10⁻²⁰¹), but also show significant differences (Fig. 3B–D). These results are reproducible on a small scale (Supplementary Fig. 3F), and confirm that SH3-independent PPIs (e.g., Abp1-Lsb3) are affected by changing the sequence of Abp1 SH3 (Fig. 3C).

Some positions are sensitive to any mutation for both SH3-dependent and SH3-independent PPIs (e.g., L17), while others are specific to only one (e.g., D32). Mutations affecting both PPI types correspond to buried residues that are distal from the Abp1-bound peptide and likely affect protein folding rather than amino acid-amino acid interactions24 (Fig. 3D–F and Supplementary Fig. 3G). Positions that are only destabilizing Abp1-Hua2 lie at the peptide interface, which is in clear contrast with the sensitive
SH3 positions are not interchangeable in multi-SH3 proteins. Proteins containing multiple SH3s can mediate the formation of multivalent interactions, bringing an additional level of complexity to PPI regulation. The position of SH3s in their host is highly conserved and is generally independent of the extent of their amino acid sequence conservation (Pearson’s correlation = 0.22, p = 0.28, Supplementary Fig. 4A–C), suggesting that SH3 positioning iskey to function. To quantify the importance at the network level of SH3 position in their host, we focused on Sla1, a cytoskeleton binding protein that has three SH3s with low sequence identity (Fig. 4A, B and Supplementary Fig. 4A).

Significant differences in PPIs and growth phenotypes are dependent on Sla1 SH3 domains (Fig. 1B, F). We constructed all possible domain-position permutations within Sla1 (i.e., domain shuffling, Fig. 4A). Sla1 PPIs are highly dependent on SH3 positions (Fig. 4B and Supplementary Fig. 5A, B), which weakly correlates with Sla1 level of expression (Supplementary Fig. 5C, D). As Sla1 SH3-1 and SH3-2 bind to the same peptide motifs in vitro\[^{14}\], we expected little impact from exchanging their position if peptide recognition was the sole determinant of SH3 specificity in vivo. Surprisingly, shuffling the first two domains (2[1]3) results in the loss of ~80% of Sla1 PPIs (Supplementary Fig. 5A). Many PPIs are lost when the SH3-1 position is occupied by SH3-2 (2[2]3) but maintained with SH3-3 (3[2]3), even if there are few PPIs that depend on SH3-3 when its positions that specifically destabilize Abp1-Lsb3 that are distant from the SH3 binding peptide (Fig. 3D–F and Supplementary Fig. 3H). Few positions specifically affecting the Abp1-Hua2 SH3-dependent interaction are conserved among Abp1 SH3 and other yeast SH3s; this could explain why none of the other SH3s can complement its loss upon domain swapping (Figs. 2B and 3B). A subset of mutations at positions 15 and 52, both predicted to be in contact with a target peptide, specifically strengthen the Abp1-Hua2 interaction, and could in principle lead to a higher interaction strength, red: increased strength. The secondary structure of the SH3 is shown above. The black bars on the top represent the level of destabilizing both PPIs, (2) specifically destabilizing Hua2 PPI or (3) Lsb3 PPI, and (4) neutral or slightly increasing binding. The minimum distance relative to a binding peptide is shown in blue (Ark1 peptide*24). e Minimum distance to the Abp1-bound peptide and (f) the relative solvent accessibility (RSA) for each category. Dest is for destabilizing mutations. For e, f, n = 349 neutral amino acid mutations, n = 98 amino acid mutations that destabilize Lsb3 only, n = 126 amino acid mutations that destabilize Hua2 only, n = 528 amino acid mutations that destabilize both Lsb3 and Hua2. For the boxplots, the median is represented as a bold center line and hinges are for the 25th and 75th percentiles (first and third quartiles). Whiskers extend from the hinges to a maximum of 1.5 times the Q3–Q1 interquartile range and outliers are represented as black dots. Source data are provided as a Source Data file.
two first endogenous SH3 are present (Figs. 1B and 4B). Shuffling two or all three Sla1 SH3s leads to the loss of most of the Sla1 PPIs, despite all three SH3s still being present albeit in a different order (Supplementary Fig. 5A). In general, the similarity of PPI profiles of the mutants is significantly correlated with their growth phenotypes (cophenetic correlation = 0.28, p value = 0.00002, from permutation, Supplementary Fig. 5B).

During clathrin-mediated endocytosis, Sla1 is an adaptor linking cargos to clathrin and recruits members of the actin machinery via SH3-dependent PPIs (Fig. 4C)25,26. Sla1 compartmentalizes into foci at internalization sites on the plasma membrane where it is present until vesicles are fully internalized (Fig. 4C); its deletion was shown to alter endocytosis dynamics25–27. The shuffling of Sla1 SH3s alters PPIs with

---

**Fig. 4** Shuffling SH3 positions alters Sla1 interactome in vivo and impacts its function in endocytosis. a Domains are moved from one position to another in proteins containing multiple SH3s (domain shuffling). Possible outcomes on PPIs are presented. b Number of PPIs detected relative to WT Sla1 for SH3-shuffling or -deletion per SH3 position (measured in quadruplicates). c Function of Sla1 as an adaptor linking early proteins with the actin machinery in clathrin-mediated endocytosis. d PPIs of Sla1 SH3-shuffled or -deleted with clathrin-mediated endocytosis-related partners. The color code represents the strength of the PPIs (PCA score). Sla1 1|2|3 represents the WT protein. e Schematic of Sla1 foci assembled at the cell membrane and their movement toward the center of the cell during internalization before disassembly, in time. f Representative microscopy images of cells expressing different Sla1-GFP proteins at multiple time points. Foci from WT Sla1-GFP, the negative control D|D|D, and a Sla1-GFP shuffling or -deletion per SH3 position (measured in quadruplicates). g Average effective distance traveled by Sla1-GFP particles towards the centroid of the cell for the different proteins. Color transparency represents the proportion of events that are not completed yet. h Average number of complete or incomplete events, i.e., that were not completed before the end of the image acquisition period. For panels f-h, the numbers of Sla1-GFP foci that were analyzed are n = 3286 for Sla1 1|2|3, n = 473 for Sla1 D|D|D, n = 686 for Sla1 2|1|3, n = 963 for Sla1 1|3|2, n = 652 for Sla1 3|2|1, n = 745 for Sla1 2|3|1 and n = 994 Sla1 3|1|2). Source data are provided as a Source Data file. See also Supplementary Data 1.
partners involved in different phases of endocytosis (Fig. 4D). We followed mutant Slal alleles by measuring the distance traveled by Slal-GFP-labeled vesicles from the cell periphery toward the cell center (Fig. 4E and Supplementary Fig. 5E). Slal-GFP foci movement is mostly regulated by Slal two N-terminal SH3s, as previously reported (Supplementary Fig. 5F, G)\(^{26,28}\). Shuffling Slal SH3s in any combination results in a drastic decrease in the effective distance traveled by Slal-positive foci, as well as their persistence in time, likely due to the inability to complete the process (Fig. 4F, G and Supplementary Fig. 5H, I). This observation is consistent with a >50% decrease in the number of Slal-GFP-labeled particle internalizations per cell and an increase in the number of incomplete events (Fig. 4H). The linearity of Slal-GFP foci trajectories is also drastically perturbed in all Slal SH3-shuffled strains, suggesting inefficient vesicle internalization (Supplementary Fig. 5H). The identity and position of each of Slal SH3 domains, therefore, regulate its PPIs and quantitatively impact the cell’s ability to tolerate stress, and to perform dynamic processes such as endocytosis.

**Protein context influences human SH3 PPIs.** We extended to humans the study of multi-SH3 proteins, as they are more prevalent (62/216 SH3-containing proteins) than in yeast (3/23). NCK adaptors (NCK1 and NCK2) connect growth factor receptors to the actin machinery via their single SH2 and three SH3 domains\(^{29}\), which bind to similar types of peptide motifs but are known to interact with different partners\(^{6,30}\) (Fig. 5A). We combined affinity purification with sequential window acquisition of all theoretical mass spectra (AP-SWATH) quantitative proteomics to test whether SH3 shuffling impairs NCK2 PPIs in human cells\(^{31,32}\).

We generated a high confidence WT NCK2 interactome (56 PPIs, false discovery rate (FDR) < 1%) that significantly overlaps with previously reported PPIs (71%), displaying about a third of all reported NCK2 interactors\(^1\). We detect significant changes in the SH3-shuffled NCK2 complexes including the complete loss of up to 17 PPIs (30% of WT NCK2 PPIs) and the gain of up to 13 new partners (Fig. 5A). The SH3-3 position appears to be more critical as NCK2 PPI profiles cluster based on the identity of the SH3 at this location (Fig. 5A). Well-characterized SH3-dependent PPIs of NCK2 depend on the position of each target cognate SH3 (Fig. 5B). In particular, NCK2 SH3-2 association with PAK1\(^33\) is significantly impaired when the domain is at the third position (Fig. 5B and Supplementary Fig. 6A, B). In contrast, NCK2 interaction with WASL, which is mediated by multiple SH3-PXXP interactions likely involving all three NCK2 SH3s\(^34\), is only slightly altered by shuffling (Fig. 5C). SH3 shuffling also leads to longer-distance disruptions of NCK2 SH2-dependent PPIs. For example, GIT1 and P130CAS/BCAR1 associations with NCK2 SH2 are significantly modulated in several SH3-shuffled mutants\(^{35,36}\) (Fig. 5C and Supplementary Fig. 6C-D). These results indicate that SH3 positioning can affect PPIs that are mediated by modular domains other than SH3s. Such indirect regulation of non-SH3 domains by adjacent SH3 positioning may also explain why many SH3-independent PPIs were altered by domain shuffling and swapping in yeast experiments.

To test whether shuffling affects the intrinsic SH3 affinity to targets via a direct effect on peptide binding, we performed in vitro fluorescence polarization binding assays using purified full-length recombinant NCK2 and peptides from direct SH3 partners CD3ε (SH3-1)\(^{37}\) and PAK1 (SH3-2). SH3-1 shuffling leads to modest variations for NCK2 interaction with CD3ε, with a maximum of approximately twofold decrease when NCK2 SH3-1 is moved to the third position (Fig. 5D and Supplementary Fig. 6E). Similarly, NCK2 binding to its PAK1 target peptide decreases by approximately twofold when its SH3-2 is inserted at the SH3-3 site (Fig. 5D and Supplementary Fig. 6F). Using circular dichroism (CD) spectropolarimetry, we determined that shuffling NCK2 SH3s does not significantly alter the overall composition of its secondary structure (Supplementary Fig. 6G). These results indicate that shuffling SH3s in their host protein might only slightly affect the position and/or accessibility of their binding pocket even though changing the positions of NCK2 SH3s resulted in the complete loss of a subset of PPIs in cells (Fig. 5A). The consequences of SH3 shuffling on the NCK2 interactome are thus possibly enhanced by factors other than binding pocket availability in vivo.

The ability of NCK2 to phase separate in vitro depends on the position of its SH3s. The separation of phases in the cytosol is an important feature of multiple cellular processes. The ability of NCK2’s close paralog NCK1 to phase separate depends on both its number of functional domains and its capacity to self-associate\(^38\). We examined the ability of NCK2 SH3-shuffled proteins to undergo phase separation in vitro via self-association. Remarkably, the different NCK2 SH3-shuffled mutants initiate varying levels of phase separation (Fig. 5E). The most striking differences occur when the SH3-2 is shuffled at the third position, which results in a significant approximately twofold increase in phase separation relative to WT (Fig. 5E). Phase separation is inhibited with the addition of a PAK1 peptide, presumably by competing for homotypic binding to NCK2 SH3-2 (Supplementary Fig. 7A). Based on this and previous observations on NCK1, we hypothesized that phase separation is dependent on NCK2 SH3-2 electrostatic interaction in trans with the first interdomain region of another NCK2 molecule\(^38\), which would depend on salt concentration. The propensity of NCK2 SH3-shuffled proteins to phase-separate is altered by increasing NaCl concentration; constructs bearing the SH3-2 at the third position are the most resilient (Supplementary Fig. 7B). In contrast, the addition of 1.6-hexanediol, a small aliphatic alcohol thought to impact the weak hydrophobic interactions\(^39\), disrupts all NCK2 SH3-shuffled proteins in a similar manner (Supplementary Fig. 7C). Taken together, these results demonstrate that NCK2 phase separation is highly dependent on the position of its SH3 domains. Shuffling NCK2 SH3-2 at the third position severely disrupts NCK2 PPIs in cells (Fig. 5A) while it stabilizes NCK2 condensates in vitro (Fig. 5E), suggesting that NCK2’s capacity to phase separate might impact its potential to nucleate PPIs in vivo.

PPI networks dictate how signaling pathways and biological processes are physically organized\(^40\). In human cells, it has been estimated that around half a million binary PPIs occur\(^41\) and that a given individual might express nearly one million distinct proteoforms\(^42\). It is therefore an outstanding challenge to understand how proteins find their cognate PPI partners while avoiding spurious interactions. Here, we demonstrate that most SH3s mediate specific interactions in combination with their host proteins. How SH3 domains regulate the PPIs of their host, for example via intramolecular interactions or other allosteric effects, remains to be determined and may be dependent on the SH3 domain itself and/or the host protein structure. Our results suggest that the description of protein domains as “beads on a string” does not capture their behavior in living cells and that the protein context of structured domains is also of great importance. We assert that our findings may be extended to other families of protein binding modules, and therefore contribute to a better understanding of complex signaling networks in normal and disease states. Our observations also suggest that the evolution of protein domains by gains and losses can have complex effects that will depend on the sequential ordering of the domains.
themselves, and this beyond motif recognition residues and individual domain–peptide interactions.

**Methods**

*S. cerevisiae* strains. For the DHFR-PCA experiments, bait strains were constructed and preys were either constructed or retrieved from the yeast protein interactome collection (Horizon)43. BY4741 (MATa, his3Δ leu2Δ met15Δ ura3Δ) strains with a specific gene of interest fused at its 3’ to the DHFR F[1,2] with a nourseothricin-resistance marker (NAT 100 μg/ml, Werner BioAgents) were used as baits. BY4742 (MATα, his3Δ leu2Δ lys2Δ ura3Δ) prey strains are each expressing a gene of interest fused at its 3′ end to the DHFR F[3] and are resistant to hygromycin B (HPH 250 μg/ml, Bioshop Canada). The same BY4741 strains were used in the growth assays with the addition of knockout (KO) strains from the yeast knockout collection (Horizon)16. For the fluorescence microscopy experiments, the different Sla1 engineered proteins were constructed using the BY4741 Sla1-GFP strain from the GFP collection (Thermo Fisher Scientific)44,45. See Supplementary Table 2 for the complete list of all strains used for this study.

Fig. 5 Human NCK2 SH3 domain shuffling alters its interactome in cells and its ability to phase separate. a–c AP-SWATH quantitative MS data of NCK2 PPIs (duplicates, PPIs with a SAINT analysis FDR < 1%, see "Methods"). a Scaled cartoon representation of NCK2 with its domains. Log2 of the average spectral counts (SWATH score) for the PPIs are shown. Baits were clustered based on the similarity of their PPI profiles. The color code represents the PPI strength (SWATH score). The bait nomenclature 1|2|3 is for WT NCK2 and D|D|D is for the triple SH3-inactive negative control (W38K/W148K/W234K). b, c NCK2 PPIs for partners with known binding sites on NCK2. The spectral count from the two biological replicates was compared to the WT NCK2 score (Ratio NCK2mut/NCK2WT). The log2 average ratio is shown (WT NCK2 ratio = 0). Error bar represents the mean plus and minus one standard deviation (n = 2 independent biological experiments). d Fluorescence polarization dissociation constants (K_D) for NCK2 full-length recombinant proteins with an SH3-1 direct (CD3Ε) or SH3-2 direct (PAK1) partner (triplicates, K_D error values represent plus and minus one SE). e NCK2 phase separation after 24 h of incubation. Soluble and phase-separated (pellet) proteins were quantified via Coomassie staining (typical replicate shown above). The proportion of proteins in the pellet compared to the total protein content (soluble + pellet) is shown for the seven replicates. Proportions of phase-separated proteins were compared to WT NCK2 (pairwise one-way ANOVA, p values: 2|1|3 = 0.90, 1|3|2 = 0.00067, 3|2|1 = 0.41, 2|3|1 = 0.60 and 3|1|2 = 0.0054). For the boxplot, the median is represented as a bold center line and hinges are for the 25th and 75th percentiles (first and third quartiles). Whiskers extend from the hinges to a maximum of 1.5 times the Q3–Q1 interquartile range. The black dots represent the different data points. Source data are provided as a Source Data file. See also Supplementary Data 3.
S. cerevisiae growth conditions. Cells were grown with their specific selection antibiotics in YPD medium (1% yeast extract, 2% tryptone, 2% glucose, and 2% agar). DMSO (Biosearch Canada) was used as the control for diploid selection. Synthetic medium (0.67% yeast nitrogen base without amino acids and without ammonium sulfate, 2% glucose, 2.5% noble agar, drop-out without adenine, methionine, and lysine, and 200 µg/ml methotrexate (MTX, Biosearch Canada) diluted in dimethyl sulfoxide (DMSO, Biosearch Canada)) was used for the DHFR-PCA experiments (PCA selection). For the WR experiments, cells were grown in a synthetic complete medium (0.174% yeast nitrogen base without amino acids, without ammonium sulfate, 2% glucose, 0.134% amino acid dropout complete, 0.5% ammonium sulfate). Sla1-GFP strains were grown in a synthetic complete medium supplemented with 10% fetal calf serum under 5% CO2 at 37 °C. TRANSFECTAM was used for this study as the assays require crosses. For genes containing two SH3s, stuffer sequences were targeted by specific Cas9 editing step with the same procedures. The correct insertion of the stuffer was validated by PCR and Sanger sequencing. The SH3-swapped, -shuffled, or digested from a plasmid. The stuffer DNA sequence (GGCGGAAGTTCTGGATACG) was used, which allowed targeting each SH3 position in a synthetic complete medium with monosodium glutamate (0.1% MSG) instead of ammonium sulfate in hygromycin B growth experiments. See Supplementary Table 3 for all the growth media used.

Cells. Human embryonic kidney 293T cells (HEK293T, ATCC-CRL-3216) were grown in Dulbecco’s Modified Eagle’s medium (DMEM) high glucose supplemented with 10% fetal calf serum under 5% CO2 at 37 °C. Transient expression of cDNA in cells was performed by polyethylenimine (PEI, Millipore Sigma) transfection, and Supplementary Table 1 for primers used was specified. Mutated SH3 domains for Deep Mutational Scanning (DMS) were amplified in-frame with a 6×His-GST TEV-cleavable tag in a synthetic complete medium with monosodium glutamate (0.1% MSG) instead of ammonium sulfate in hygromycin B growth experiments. See Supplementary Table 3 for all the growth media used.

Strain construction. The complete list of strains, primers, and gRNAs used in this study can be found in Supplementary Tables 1–3. All constructed yeast strains were validated by polymerase chain reaction (PCR) and confirmed by Sanger sequencing of the knockouts of the SH3 domains for the DHFR-PCA experiments were constructed as follows. The genes of interest were fused into the genome of the BY4741 strain. The correct insertion of the stuffer was validated by PCR and Sanger sequencing. The SH3-swapped, -shuffled, or digested from a plasmid. The stuffer DNA sequence (GGCGGAAGTTCTGGATACG) was used, which allowed targeting each SH3 position in a synthetic complete medium with monosodium glutamate (0.1% MSG) instead of ammonium sulfate in hygromycin B growth experiments. See Supplementary Table 3 for all the growth media used.

PCA experiments and analysis. The overall DHFR-PCA pipeline is based on Tarasov et al. and the procedures have also been published as a visualized protocol. The screens were performed with robotically manipulated pin tools (BMS-SC1, S&R Robotics Inc.). First, the preys were cherry-picked from 96-well plates to omnitrays in a 384 format on YPD + HPH solid media. Plates were then combined into the 1536 format on YPD + HPH. The array was constituted of 575 preys with DHFR F[3] preys were done on YPD followed by incubation at 30 °C for 48 h. Two diploid selection steps were performed by replicating the plates on YPD + NAT and then transferred on solid YPD + NAT omnitrays. They were then replicated on YPD + NAT omnitrays in 1536 array format. Mating of DHFR F[1,2] baits with DHFR F[3] preys was done on YPD followed by incubation at 30 °C for 48 h. Two diploid selection steps were performed by replicating the plates on YPD + NAT + HPH and incubating them for 48 h at 30 °C. Pictures were taken at the end of the second diploid selection for quality control. All images were acquired with a EOS Rebel T5i camera (Canon). Finally, the diploid cells were replicated on omnitrays containing solid PCA selection media and incubated for 4 days as a first selection step in a splmager custom robotic platform (S&R Robotics Inc.). Cells were replicated for a second PCA selection step and incubated in the same manner. The last time point of the second PCA selection step was used for the data analysis as in previous studies using this method.
Liquid DHFR-PCA low-throughput experiments were performed to validate the PPIs identified as affected by the SH3 domain deletions in the high-throughput solid phase interaction screen (SPIS). 16 PPIs were scored and quantified using following domain deletions. Mating and the diploid selection were performed as described for the large-scale experiment, with the exception that the plates were in 384 formats. Following the second diploid selection, diploid cells were inoculated in a 96 v-shaped well plate using a robotically manipulated 96 pin tool in a synthetic complete medium with MSG at pH 6.0 with NAT and HPH. After 48 h of growth, optical density (OD) was read using a TECAN Infinite F200 Pro, and dilutions to 1 OD in sterile nanopure water were prepared. Cells were then diluted to 0.1 OD by combining 25 µl of cell suspension at OD = 1 and 225 µl of PCA selection media without dilution. Plates were then incubated at 30°C in a TECAN Infinite M Nano and OD was monitored every 15 min for 72 h.

A separate DHFR-PCA screen was performed for the Sla1 SH3 shuffling experiment, with the same analysis as previously described. Three plates had no detectable level of background, which correspond to the second plate of the three Sla1 SH3 WT re-insertion controls. They were attributed the average of the detectable level of background, which correspond to the second plate of the three in 96 wells plate (synthetic complete medium with MSG or with MSG and Hygromycin B 120 µg/mL removed. The colony sizes after 74 h of incubation at 37 °C were log2 transformed was manually inspected. Positions with an abnormality in colony circularity were from plate pictures were also measured with the R package platform with images acquired every two hours for four days. The colony sizes were detected by sequencing as described below.

**Growth assays in stress conditions.** All WT, SH3-deleted, KO (except for CDC25 which is the only essential SH3-containing gene in yeast), ABP1 SH3-swapped, and Sla1 SH3-shuffled strains were grown on omnitrips in 1536 formats and robotically manipulated with a pin tool (the DHFR F[1,2] bait strains were used except for the KO strains). Each strain was randomly positioned in 6 replicates per plate and two plates per condition were performed, for 12 replicates per strain in total. A border of two rows and two columns of WT strains (BY4741) around the plate was used to remove border effects. Strains were grown on YPD media for two days at 30°C. They were then replicated on 51 different media that are described in the Supplementary Table 3 and were incubated at 37 °C in a splmager custom platform with images acquired every two hours for four days. The colony sizes from plate pictures were also measured with the R package “Gitter” and the output was manually inspected. Positions with an abnormality in colony circularity were removed. After 74 h of incubation, the colonies were measured and used for the analyses. First, the differences between the two plate replicates were corrected to remove plate effects. The median colony size was calculated for each strain per plate. Strains with a difference in their median colony size of more than two between the two plate replicates were removed. Finally, the average of the median colony size was calculated for each strain per condition (growth score). For the hygromycin B resistance assays in liquid cultures, the ABP1 SH3 swapped strains including the ABP1 SH3-deleted and ABP1 KO strain were grown overnight at 30 °C as precultures in synthetic complete medium with MSG. The next day, OD was measured with a TECAN Infinite F200 Pro plate reader, and the cultures were diluted to OD of 1.0 in sterile nanopure water. The different strains were further diluted at 0.1 OD by adding 25 µl of cultures to 225 µl of synthetic complete medium with MSG or with MSG and Hygromycin B 120 µg/mL in 96 wells plate (final concentration of 108 µg/mL). Cells were incubated in a TECAN Infinite M Nano plate reader at 37°C for 24 h and OD was measured every 15 min.

Deep mutational scanning. Single site mutation libraries were generated by a PCR-based saturation mutagenesis method as described below. The mutagenesis was carried out on the pUC19 plasmid containing the SH3 domain sequence of ABP1 flanked by its homology arms subsequently used in the CRISPR mediated genomic recombination. We used oligos containing degenerate nucleotides (NNN) to carry out the mutagenesis at each codon of the domain sequence. In the first step of the two-step PCR procedure (short PCR step), the plasmid was amplified using an oligo containing the degenerate codon which is positioned within the domain sequence, and another oligo lying outside Abp1 sequence in the plasmid. Short PCR step was carried out with the following settings: 5 min at 95 °C, 35 cycles: 20 s at 98 °C, 15 s at 60 °C, 30 s at 72 °C, and a final extension of 1 min at 72 °C. The amplified product can be used as a template in the second step (long PCR step) to amplify the whole plasmid. Long PCR step was carried out with the following settings: 5 min at 95 °C, 22 cycles: 20 s at 98 °C, 15 s at 68 °C, 3 min 30 s at 72 °C, and a final extension of 5 min at 72°C. The long PCR product was digested for 2 h at 37°C with DpnI. The digestion product was then transformed into E. coli MC1061 competent cells. In order to obtain most of the 64 possible codons, we recovered ~1000 colonies from each plate to retrieve the plasmid single mutant libraries. Saturation mutagenesis was carried out in a codon-position-wise manner. Mutants at each position were stored separately. Library quality control was assessed by amplifying the purified plasmid DNA preparations by PCR followed by Illumina sequencing as described below. The genomic insertion of the mutant libraries (one library per amino acid position) was performed by targeting the stuffer sequence in the ABP1 SH3-deleted strain as described in the strain construction section. ABP1 SH3-deleted yeast competent cells were co-transformed with pCAS containing the stuffer-specific gRNA and PCR amplified ABP1 SH3 mutant libraries were transformed by electroporation into the yeast library generation, ~1000 colonies from each plate were retrieved. Glycerol stocks of yeast cells were kept for each position (58 different stocks corresponding to the 58 amino acid positions of Abp1 SH3). The stocks were also validated by Illumina sequencing (see below).

Liquid DHFR-PCA with libraries of mutants. Liquid DHFR-PCA experiments were performed following the generation of the ABP1 SH3 single position mutated yeast libraries using a biological independent number of cells from each yeast mutant strain (master pool). The Liquid DHFR-PCA experiment, first PCA screening and amplification were performed using the DHFR F[3] or Lab3-DHFR F[3] preys strains were mixed in YPD (2:1) for mating and incubated for 8 h at 30 °C. Diploid cells were selected the first time by transferring the mixture in a YPD + NAT + HPH (QD5) plate for 16 h at 30°C. The next day, the first diploid selection was transferred in SC complete pH 6.0 + NAT + HPH at 0.5 OD and the culture was grown for 24 h at 30 °C. The validation of the yeast mutant libraries was done the next day with a fraction of the diploid cells (first DHFR-PCA experiment time point without PCA selection, reference condition or amplification, XQD = 0) were collected for 16 h at 30°C. The next day, the first DHFR-PCA experiment step using the same procedures. Finally, the genomic DNA of the cells (SU OD) after two cycles of PCA selection was extracted and the surviving ABP1 SH3 mutant strains were detected by sequencing as described below.

Single mutant libraries sequencing. For the validation of the plasmid single mutant libraries, the plasmid DNA was extracted from bacteria following transformation using a mini-prep plasmid extraction kit (ProGene). Yeast genomic DNA extraction of the library of ABP1 SH3 mutants was performed using a standard phenol/chloroform protocol. Sequencing of yeast genomic DNA was performed at three different time points during the liquid DHFR-PCA experiments. After diploid selection of the bait-prey strains (single mutant library validation, XQD = 0) the genomic DNA was extracted using a NanoDrop sequencing as described in the section below. Following the selection, diploid cells OD was read using a TECAN Infinite F200 Pro, and dilutions to 0.1 OD in PCA selection media without agar in a 50 ml tube was done. Tubes were incubated at 30°C for 72 h (second-time point of the DHFR-PCA experiment, first PCA selection and amplification, XQD = 1) of the selected pool from the second PCR amplification was collected for three preamplification cycles. The cells from the first PCA step were grown for a second PCA selection step using the same procedures. Finally, the genomic DNA of the cells (SU OD) after two cycles of PCA selection was extracted and the surviving ABP1 SH3 mutant strains were detected by sequencing as described below.

**Analysis of DMS data.** The raw data generated from deep sequencing (fastq format) was first demultiplexed based on the barcode sequences using custom scripts. Forward and reverse reads corresponding to each sample were merged using a proprietary tool (the merged fastq files were filtered to remove reads with an average Phred score of less than 30 and the nucleotides with Phred score of less than 30 using fastp). The reads were aligned to their corresponding reference sequence using bowtie2. The global alignment was carried out using the command: "bowtie2 -p 6 -end-to-end --very-sensitive --no-discard --no-mixed -x Reference -U SamFastq -S Sampath, where Reference is the path to the bowtie2 built reference, SamFastq is the path to the unaligned fastq file and Sampath is the path to the aligned file in sam format. The aligned sam file was used to count the number of mutations using samtools and pysam (https://github.com/pysam-developers/pysam). In order to normalize the counts by the depth of sequencing, the number of mutations were divided by the PCR fold change at the position of the mutation. The normalized counts are referred to as frequencies. Next, the log2 ratios (pseudolog with pseudo count of 0.5) of frequencies in the test condition
The centroids of the segmented cells were used to identify the locations of each cell in the image using Scikit-Image. The location of each cell was used to isolate individual cell time-frames of fluorescence (GFP) images. Each cell time frame of the fluorescence images was processed through a particle tracking tool—Trackpy. Trackpy detected the locations of Shl-1 GFP positive foci in each frame and linked them together to provide particle-wise trajectories. Trajectories detected in less than ten frames were considered spurious and were filtered out. Segments of the trajectories spuriously indicating movement of the particles back to the cell membranes were trimmed. The preprocessed trajectories of the particles were then used for the calculation of distances using the spatial distance module of SciPy. The python-based code used for the analysis of microscopy images is included in the endocytosis module of "htsimaging" package.

**Protein purifications.** E. coli BL21 (DE3) cells were transformed with mNCK2 constructs in the pET-30b vector and grown overnight at 37 °C in LB with 100 μg/ml kanamycin. The next day, cultures were expanded in LB with kanamycin until they reached their exponential growth phase (OD600 of around 1.0). Protein production was then induced by the addition of 1 mM IPTG (Bio Basic) and incubation at 16 °C overnight with agitation. Cultures were pelleted with a centrifugation step (3200g at 4 °C, 30 min), washed with ice-cold PBS, and pellets were kept at −80 °C. Bacteria were thawed in GST buffer (20 mM Tris-HCL pH 7.5, 0.5 M NaCl, 1 mM EDTA, and 1 mM DTT (Millipore Sigma)) with complete protease inhibitors (Millipore Sigma) and disrupted with sonication. Lysates were then centrifuged at 20,000 g for 20 min at 4 °C and the supernatant loaded to a GSTTrap FF column (GE Healthcare). Following a wash with 10 mL of GST-Buffer, the GST-Tagged proteins were eluted with 15 mM GSH (Bio-basic) and cleaved in a solution using purified recombinant His-tagged TEV protease (Thermo Fisher Scientific). Tag-free NCK2 proteins were further purified using a 1 mL HisTrap FF column (GE Healthcare) to remove free His-tagged GST and TEV and were then concentrated using Amicon 10 K filters (Millipore Sigma) and their concentration assessed using Coomassie and BCA assays (Thermo Fisher Scientific).

**Fluorescence polarization.** Increasing amounts of NCK2 full-length recombinant proteins in GST buffer (final concentrations of 0–150 μM for CD3ε and 0–175 μM for PAK1) were used for binding assays with a constant amount of fluorescent isothiocyanate-conjugated peptides (resuspended in 17 mM Tris-HCl pH 8, 100 mM NaCl and 0.5% Brij L23 (Millipore Sigma)), peptides final concentration of 40 nM). CD3ε (RGQNKERPPPVPNPDY) and PAK1 (DIQDKPPAPPMRNTST) peptides were used as NCK2 SH3-1 and NCK2 SH3-2 ligands, respectively. Binding assays were performed in the FP buffer (17 mM Tris-HCl pH 8 and 100 mM NaCl) and fluorescence polarization was measured on a Synergy H1 multimode plate reader (Bio Tek) at 535 nm with 485 nm excitation. Calculation of the dissociation constants was performed with a one-site total binding model and nonlinear regression in GraphPad Prism version 7.

**Phase separation.** Purified NCK2 proteins were diluted in GST buffer to identical concentrations. At 500 mM NaCl present in this buffer, all NCK2 constructs were completely soluble. Phase transition was initiated by dilution in imidazole buffer (10 mM Tris-HCl (Bio Basic) pH 7.0, 1 mM DTT, 1 mM EGTA (Millipore Sigma), 1 mM MgCl2, 2.5% Glycerol (Biosharp Canada), 3% Dextran (Millipore Sigma)) containing various amount of NaCl so that the final concentration of NaCl would correspond to those in Supplementary Fig. 7B. For all the other experiments, the phase transition of NCK2 protein preparations was performed in 60 mM NaCl. NCK2 at 8 μg/mL was added to 35 μL imidazole buffer in a plastic 96-well plate that had first been pretreated with a 0.2% bovine serum albumin solution to reduce nonspecific binding. The plate assay was then incubated overnight (24 h) at 4 °C before imaging. Phase-image contrast images of each well were acquired using a Zeiss Axios Vert.A1 inverted microscope with an LD A-Plan 40x/0.55 Ph1 objective (Zeiss) in Zen Blue software (version 2.3.69.1000). After imaging, the soluble protein fraction (supernatant) was recovered for sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis. Protein droplets attached at the bottom of the well (pellet) were washed once with imidazole buffer with 60 mM NaCl and then solubilized in Laemmli buffer. The pellets to supernatant fractions were used (100 μg) for future Western blotting and antibodies detection.

**Live imaging of endocytosis.** Overnight cultures of Slai-GFP cells were diluted to an OD600 of 0.15 and grown in synthetic complete medium without tryptophan until they reached an OD600 of ~0.5–0.5 at 30 °C. The cells were then seeded on an ammonavolin A 0.05 mg/mL (Millipore Sigma) coated 8-well glass-bottom chamber slides (Sarstedt). Imaging acquisition was performed using a Perkin Elmer UltraView confocal spinning disk unit attached to a Nikon Eclipse TE2000-U inverted microscope equipped with a Plan Achromat DIC H 100×/1.4 oil objective (Nikon), and a Hamamatsu Orca Flash 4.0 LT + camera. Imaging was done at 25 °C in a humidified chamber. The software package ImageJ was used for image capture. For each field, one brightfield and a series of fluorescence (GFP) images were taken. Cells were excited with a 488 nm laser and emission was filtered with a 530/630 nm filter. GFP time laps were acquired continuously at a rate of 1 frame/sec for 3 min.

**Analysis of microscopy images.** Bright-field images were used to segment cells using YeastSpotter6. Cells were filtered based on circularity, solidity, and the normalized difference between their minor and major axes to remove weakly detected cells. Out-of-focus cells were also filtered out based on size and brightness. The centroids of the segmented cells were used to identify the locations of each cell.
temperature was maintained by a Pelтиer-type JASCO CDF-4268/15 thermostatic controller. Raw spectral data were extracted using the Spectra Manager Suite (JASCO). All experiments were performed in triplicate. Mean values and standard error of the mean (SEM) for each normalized data point were plotted and analyzed using GraphPad Prism 9.0. To account for small experimental variations in protein concentration during sample preparation, molar ellipticity (θ, mdeg) was normalized using the following equation:

$$\text{Normalized ellipticity fraction} = \frac{\theta - \theta_{\text{min}}}{\theta_{\text{max}} - \theta_{\text{min}}}$$  (1)

**Experimental design for MS experiments.** NCK2s purified on beads from HEK293T cells (as previously described in the methods) were washed two additional times in 20 mM Tris-HCl pH 7.4 and eluted in 50 mM formic acid. Peptides were then digested with trypsin (Promega) on-beads as previously detailed. Briefly, eluted peptides were concentrated on ZipTip®-SCX (Millipore Sigma), reduced with TCEP 100 mM (Millipore Sigma) and alkylated during the tryptic digestion with iodoacetamide 10 mM (Millipore Sigma). Peptides were then eluted in ammonium bicarbonate 200 mM (Millipore Sigma) and desalted on Stage Tips C18 columns. The final eluted peptides were dried using an Eppendorf Vacufuge™ Concentrator. For each bait, two biological replicates were processed independently. These were analyzed alongside four replicates to avoid systematic bias. We validated that the same quantity of baits was used for each NCK2 chimeras by quantifying NCK2 peptides in each sample (see Supplementary Data 3).

**Proteins identification by MS.** The analyses were performed at the proteomic platform of the Quebec Genomics Center. Protein samples were separated by online reversed-phase (RP) nanoscale capillary liquid chromatography (nanoLC) and analyzed by electrospray MS (ESI MS/MS). The experiments were performed with a Dionex UltiMate 3000 nanoRSLC chromatography system (Thermo Fisher Scientific) connected to a TripleTOF 5600+ mass spectrometer (Sciex) equipped with a nanoelectrospray ion source. Peptides were trapped at 20 μl/min in loading solvent (2% acetonitrile, 0.05% TFA) on a 5 μm 300 mm C18 pepmap cartridge pre-column (Thermo Fisher Scientific) for 5 min. Then, the precolumn was switch online with a self-made 50 cm × 75 μm internal diameter separation column packed with ReproSil-Pur C18-AQ 3-μm resin (Dr. Maisch). The peptides were eluted with a linear gradient from 5 to 40% solvent B (A: 0.1% formic acid, B: 80% acetonitrile, 0.1% formic acid) in 120 min, at 300 nL/min. In DDA, the instrument method consisted of one 250 ms MS1 TOF survey scan from 400 to 1300 Da followed by twenty 100 ms MS2 candidate ion scans from 100 to 2000 Da in high sensitivity mode. Only ions with a charge of 2+ to 5+ that exceeded a threshold of 150 cps were selected for MS2, and former precursors were excluded for 20 s after one occurrence. In DIA, acquisition consisted of one 50 ms MS1 scan followed by 32 × 25 a.m.u. ion isolation windows covering the mass range of 400–1250 a.m.u. (cycle time of 2.52 s); an overlap of 1 Da between SWATH was preselected. The spectral library was then used for protein identification with the lowest MS-GFDB (Beta version 1.0072 (6/30/2014)) probability for each peptide identified by MSPLIT73 with mass tolerance ±50 ppm for fragment ions. Peptide length was limited to 8 amino acids. The results from each search engine were analyzed cation by MSPLIT73 with peptides identified by MSPLIT-DIA passing a 1% FDR subsequently matched to mass data generated by the triple quadrupole TOF/TOF and the Orbitrap Velos platforms using the Ensembl REST API79. The spectra were searched with the NCBI RefSeq database (version 57, January 30th, 2013) against a total of 72,482 human and adenovirus sequences using “common contaminants” from the Max Planck Institute (http://141.61.102.106:8080/share/cgi/ sid=0fg2fuB) and the GPM (http://www.thegpm.org/crap/index.html). Decoy spectra were appended using the decoy library command built into to MSPLIT-DIA, with a fragment mass tolerance of 0.05 Da. The pairwise similarity was then used for protein identification by MSPLIT in DDA mode while only two unique peptides were required in DIA mode.

**Data archival.** All MS files used in this study were deposited at MassIVE (http://massive.ucsd.edu). They were assigned the identifiers MassIVE MSV000085093 (DIA files) and MSV000085092 (DDA files). The files can be accessed using the following links: ftp://MSV000085093@massive.ucsd.edu and ftp://MSV000085092@massive.ucsd.edu and ftp://

**Position and sequence conservation of orthologs.** The UniProt database was used to retrieve full-length sequences of all S. cerevisiae SH3 proteins except for the Sdc25 (YLL017W) pseudogene. One-to-one orthologs for each protein were then retrieved from fungal species using Ensembl Compara. The process was automated for each SH3 protein using the Ensembl REST API. The hmmsearch function of HMMER v3.3 was then used for the domain annotation of all orthologs using default significance thresholds of 0.01 (E = 0.01 and donE = 0.01) and the Pfam-A HMM library. Any domain belonging to the Pfam SH3 clan (CL0010) was assigned as an SH3 domain. SH3 domain positions for each orthologous sequence were taken by dividing the domain start point (‘env’) by the full sequence length. This calculation was repeated for each S. cerevisiae homolog, and the one protein pair with the highest similarity was then selected. The sequences within an orthologous group were clustered using the MAFFT L-INS-i method. The pairwise similarity between domain sequences was determined from the seqidentity() function in the R package bioclust, and we took the mean pairwise sequence similarity between the S. cerevisiae sequence and all its orthologs as a measure of sequence conservation for each domain.

**Tree comparison analysis.** In Fig. 2C, the relationship between SH3 domain sequences and their PPI profiles in Abp1 was explored by constructing dendrograms for both features and then quantifying the similarity between dendrograms. For the interaction similarities (Fig. 2C, left), a Euclidean distance matrix of the PCA scores was first constructed using the dist() function in R, and then the distance matrix used to generate a dendrogram from the hclust() function in R with default parameters. For the sequence similarities (Fig. 2C, right), the SH3 domain sequences were first aligned using the MAFFT L-INS-i method. The pairwise similarity between domain sequences was determined from the seqidentity() function in the R package bioclust, and we took the mean pairwise sequence similarity between the S. cerevisiae sequence and all its orthologs as a measure of sequence conservation for each domain.

**DIA MS analysis with MSPLIT-DIA.** DIA MS data was analyzed using MSPLIT-DIA (version 1.0, 20) implemented in ProHits 4.019. To generate a sample-specific spectrum file for retrieving feature matches (FSMs) from matched DADA runs (18 runs) were pooled by retaining only the spectrum with the lowest MS-GFBDF (Beta version 1.0072 (6/30/2014)) probability for each unique peptide sequence and precursor charge state, and a peptide-level FDR of 1% was enforced using TDA. The MS-GFBDF parameters were set to search for true peptide matches, allowing no missed cleavage sites, 1 C11 atom per peptide with a mass tolerance of 50 ppm for precursors with charges of 2+ to 4+ and tolerance of ±50 ppm for fragment ions. Peptide length was limited to 8–30 amino acids. Variable modifications were deamidated asparagine and glutamine and oxidized methionine. The analyses were performed at the proteomic platform of the Quebec Genomics Center. Protein samples were separated by online reversed-phase (RP) nanoscale capillary liquid chromatography (nanoLC) and analyzed by electrospray MS (ESI MS/MS). The experiments were performed with a Dionex UltiMate 3000 nanoRSLC chromatography system (Thermo Fisher Scientific) connected to a TripleTOF 5600+ mass spectrometer (Sciex) equipped with a nanoelectrospray ion source. Peptides were trapped at 20 μl/min in loading solvent (2% acetonitrile, 0.05% TFA) on a 5 μm 300 mm C18 pepmap cartridge pre-column (Thermo Fisher Scientific) for 5 min. Then, the precolumn was switch online with a self-made 50 cm × 75 μm internal diameter separation column packed with ReproSil-Pur C18-AQ 3-μm resin (Dr. Maisch). The peptides were eluted with a linear gradient from 5 to 40% solvent B (A: 0.1% formic acid, B: 80% acetonitrile, 0.1% formic acid) in 120 min, at 300 nL/min. In DDA, the instrument method consisted of one 250 ms MS1 TOF survey scan from 400 to 1300 Da followed by twenty 100 ms MS2 candidate ion scans from 100 to 2000 Da in high sensitivity mode. Only ions with a charge of 2+ to 5+ that exceeded a threshold of 150 cps were selected for MS2, and former precursors were excluded for 20 s after one occurrence. In DIA, acquisition consisted of one 50 ms MS1 scan followed by 32 × 25 a.m.u. ion isolation windows covering the mass range of 400–1250 a.m.u. (cycle time of 2.52 s); an overlap of 1 Da between SWATH was preselected. The collision energy for each window was set independently as defined by CE $\text{m/z} = 0.06 \times m/z + 4$, where $m/z$ is the center of each window, with a spread of 15 eV performed linearly across the accumulation time.
was used to compare PPI profiles with growth profiles for Sla1 constructs generated from the domain shuffling experiments (Supplementary Fig. 5B). Both the DHFR-PCA interaction data and the Sla1 growth data were clustered as described above, by generating a Euclidean distance matrix and then performing hierarchical clustering.

**PWMs enrichment analysis.** All yeast PWMs were taken from a 2009 study of SH3 domain specificity. The similarity between SH3 PWMs and known inter-actors was assessed using a matrix similarity score (MSS) derived from the MATCH algorithm. This scoring method assigns a score of 1 to perfect sequence matches to the PWM and vice versa. For each interactor assigned to an SH3 domain, MSS scores for the corresponding PWM were calculated for all possible sequence k-mers (k = number of PWM columns) and then the maximum is taken (Max. MSS). This procedure was repeated for all SH3 domains with an assigned PWM and then the results pooled to generate Fig. 1C, D, and Supplementary Fig. 2G. Some SH3 domains were represented by more than one PWM from ref. 14, reflecting multiple specificities. In these cases, MSSs were calculated for both PWMs, and then the overall maximum score took forward for further analysis. In Fig. 1, maximum MSS scores were calculated for sequences belonging to the “Random”, “SH3-independent PPI”, “SH3-inhibited PPI”, and “SH3-dependent PPI” category of interaction sequences. For the “Random” category, random peptides were generated by sampling amino acids according to their background frequency in the S. cerevisiae proteome. For SH3-independent PPI, we included only those interactors that were found to be unaffected by any SH3 domain deletions given that in vitro-derived SH3 PWMs overlap strongly, which could lead to spurious MSS enrichment for SH3-independent PPIs. In Fig. 1D, maximum MSSs were also recalculated for the “SH3-dependent PPI” category of interaction sequences after randomly assigning PWMs to each SH3 domain; this procedure was repeated until 10,000 maximum MSS scores were sampled. For Supplementary Fig. 2G, the “PPI gained by Abp1” corresponds to PPIs gained by Abp1 after domain swapping, and the “PPI unaffected” corresponds to PPIs of Abp1 that do not change after domain swapping. The ‘Random’ category was generated using the same approach described directly above for Fig. 1.

**Comparison of interactome with literature.** To find previously reported PPIs, we parsed a recent release (v 3.5.16) of the BioGRID and searched for reports of physical interactions between baits and preys. PPIs were searched in the two directions possible (A–B and B–A) in the database.

**Quantification and statistical analysis.** All the statistical details of the different experiments can be found in the figure legends, figures, and results sections.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The MS datasets generated during this study are available at MassIVE: MSV00085092 (DSB SWATH MS) [doi:10.25345/C5CM55] and MSV00085093 (DDA MS) [doi:10.25345/C5CT3M]. The deep sequencing datasets generated during this study are available at NCBI as BioProjects: SAMN14758855 (Abp1-Hua2 DMS reference condition), SAMN14758856 (Abp1-Lab3 DMS reference condition), SAMN14758857 (Abp1-Hua2 DMS DHFR-PCA condition) and SAMN14758858 (Abp1-Lab3 DMS DHFR-PCA condition). Accession codes, unique identifiers, or web links for publicly available datasets: SMART V.8.0 (http://smart.embl-heidelberg.de), RRID:SCR_005026, PDB (PDB2RPN, 10.2210/pdb2rpn/pdb), BioGRID (version 3.5.16, https://thebiogrid.org/), RRID:SCR_007393, RefSeq (v 57 (01/30/2013)), NCBI, https://www.ncbi.nlm.nih.gov/ refseq/, Common mass spectrometry contaminants (Max Planck Institute, http://141.61.102.160:8080/share/cgi/sqlid=02flgub), Global Proteome Machine (https://www.theembl.org/crap/index.html), Ensembl Compare (https://www.ensembl.org/mfo/ genome/compara/index.html), NCBI (NP_059099.3) and Uniprot (Q55033). The large datasets with the raw data for all DHFR-PCA, growth, MS, and CD experiments are included as Supplementary Data 1–3 (Excel files). Source data are provided with this paper.

**Code availability**

The code generated during this study is available at DionneEtal2020 repository (github.com/Landrylab).

Received: 2 September 2020; Accepted: 17 February 2021; Published online: 12 March 2021

**References**

1. Scott, J. D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they’re apart. Science 326, 1220–1224 (2009).
2. Iverson, Y. & Jenuth, P. Affinity and specificity of motif-based protein-protein interactions. Curr. Opin. Struct. Biol. 54, 26–33 (2019).
3. Brown, T., Brown, N. & Stollar, E. J. Most yeast SH3 domains bind peptide targets with high intrinsic specificity. PLoS ONE 13, e0193128 (2018).
4. Mayer, B. J. The discovery of modular binding domains: building blocks of cell signalling. Nat. Rev. Mol. Cell Biol. 16, 691–698 (2015).
5. Cunningham, J. M., Kotigire, G., Sorger, P. K. & AlQuraishi, M. Biophysical prediction of protein-peptide interactions and signaling networks using machine learning. Nat. Methods 17, 175–183 (2020).
6. Dwayne, U. et al. Direct phosphorylation of SRO homology 3 domains by tyrosine kinase receptors disassembles ligand-induced signaling networks. Mol. Cell 70, 995–1007 (2018).
7. Xin, X. et al. SH3 interactome conserves general function over specific form. Mol. Syst. Biol. 9, 652 (2013).
8. Vogel, C. & Chothia, C. Protein family expansions and biological complexity. PLoS Comput. Biol. 2, e8 (2006).
9. Kell, A., Dubreuil, B., Levy, E. D. & Michnick, S. W. Exhaustive search of linear information encoding protein-peptide recognition. PLoS Comput. Biol. 13, e1005499 (2017).
10. Stark, C. et al. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 34, D535–D539 (2006).
11. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein–protein interactions. Protein Sci. 22, 153–167 (2013).
12. Li, G. et al. Construction of a linker library with widely controllable flexibility for fusion protein design. Appl. Microbiol. Biotechnol. 100, 215–225 (2016).
13. Tonkian, R. et al. Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol. 7, e1000218 (2009).
14. Merc, Y., Pond, M. P. & Roux, B. Tyrosine kinase activation and conformational flexibility: lessons from src-family tyrosine kinases. Acc. Chem. Res. 50, 1193–1201 (2017).
15. Giaever, G. et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418, 387–391 (2002).
16. Biot-Pelletier, D. & Martin, V. J. J. Seamless site-directed mutagenesis of the Saccharomyces cerevisiae genome using CRISPR-Cas9. J. Biol. Eng. 10, 6 (2016).
17. Tonkian, R. et al.yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol. 7, 1000218 (2009).
18. Barredo, L. et al. A genomewide screen for tolerance to cationic drugs reveals genes important for potentiation of ion channels in Saccharomyces cerevisiae. Eukaryot. Cell 10, 1241–1250 (2011).
19. Zarrinnar, A., Park, S.-H. & Lim, W. A. Optimization of specificity in a cellular protein interaction network by negative selection. Nature 426, 676–680 (2003).
20. Mapes, J. & Ota, I. M. Nbh2 targets the Ptc1-type 2C Ser/Thr phosphatase to the HOG MAPK pathway. EMBRYO J 23, 302−311 (2004).
21. Kell, A., Levy, E. D. & Michnick, S. W. Evolution of domain-peptide interactions to co-adapt specificity and affinity to functional diversity. Proc. Natl Acad. Sci. USA 113, E3862–E3871 (2016).
22. Stollar, E. J. et al. Structural, functional, and bioinformatic studies demonstrate the crucial role of an extended peptide binding site for the SH3 domain of yeast Abp1. J. Biol. Chem. 284, 26918–26927 (2009).
23. Tolsma, T. O., Cuevas, L. M. & Di Pietro, S. M. The Sla1 adaptor-clathrin interaction regulates coat formation and progression of endocytosis. Traffic 19, 446–462 (2018).
24. Feliciano, D. & Di Pietro, S. M. SLAC, a complex between Sla1 and Las17, regulates actin polymerization during clathrin-mediated endocytosis. Mol. Biol. Cell 23, 4526–4527 (2012).
25. Holtzman, D. A., Yang, S. & Drubin, D. G. Synthetic lethal interactions identify two novel genes, SLA1 and SLA2, that control membrane cytoskeleton assembly in Saccharomyces cerevisiae. J. Cell Biol. 122, 635–644 (1993).
26. Rodal, A. A., Manning, A. L., Goode, B. L. & Drubin, D. G. Negative regulation of yeast WASP by two SH3 domain-containing proteins. Curr. Biol. 13, 1000–1008 (2003).
27. Chaki, S. P. & Rivera, G. M. Integration of signaling and cytoskeletal remodeling by Nck in directional cell migration. Bioarchitecture 3, 57–63 (2013).
28. Curo, J., Li, M., Ran, X., Fan, J.-S. & Song, J. Structural insight into the binding diversity between the human Nck2 SH3 domains and proline-rich proteins. Biochemistry 45, 7171–7184 (2006).
31. Lambert, J.-P. et al. Mapping differential interacomymes by affinity purification coupled with data-independent mass spectrometry acquisition. Nat. Methods 10, 1239–1245 (2013).

32. Collins, B. C. et al. Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. Nat. Methods 10, 1246–1253 (2013).

33. Bisson, N., Poitras, L., Mikryukov, A., Tremblay, M. & Moss, T. EphA4 signaling regulates blastomere adhesion in the Xenopus embryo by recruiting Pak1 to suppress Cdc42 function. Mol. Biol. Cell. 18, 1030–1043 (2007).

34. Koh, T., G., Z., and B. N. et al. gitter: a robust and accurate method for quanti-

35. Frese, S. et al. The phosphotyrosine peptide binding specificity of Nck1 and Nck2 Src homology 2 domains. J. Biol. Chem. 281, 18236–18245 (2006).

36. Schlaepfer, D. D., Broome, M. A. & Hunter, T. Fibronectin-stimulated signaling from a focal adhesion kinase-c-Src complex: involvement of the Grb2, p130cas, and Nck adaptor proteins. Mol. Cell. Biol. 17, 1702–1713 (1997).

37. Gil, D., Schamel, W. W. A., Montoya, M., Sánchez-Madrid, F. & alarcón, B. Recruitment of Nck by CD3 epsilon reveals a ligand-induced conformational change essential for T cell receptor signaling and synapse formation. Cell 109, 901–912 (2002).

38. Banjade, S. et al. Conserved interdomain linker promotes phase separation of the multifunctional adaptor protein Nck. Proc. Natl Acad. Sci. USA 112, E6426–E6435 (2015).

39. Kroshwald, S. et al. Promiscuous interactions and protein disaggregates determine the material state of stress-inducible RNPs. Elife 4, e06807 (2015).

40. Landry, C. R., Levy, E. D., Abd Babbo, D., Tarassov, K. & Michnick, S. W. Emerging insight from noisy cellular networks. Cell 155, 983–989 (2013).

41. Stumpf, M. P. H. et al. Estimating the size of the human interactome. Proc. Natl Acad. Sci. USA 105, 6959–6964 (2008).

42. Aebersold, R. et al. How many human proteoforms are there? Mol. Cell. Proteom. 10, 1745–1758 (2011).

43. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate method for removing PCR and adaptor artifacts from paired-end reads. Bioinformatics 30, 1486–1492 (2014).

44. Scaffold, R. raad88f43: single particle tracking to monitor mitochondrial dynamics. Mol. Biol. Cell. 18, 1896–1906 (2007).

45. Liu, G. et al. Data independent acquisition analysis in ProHits 4.0. Mol. Cell. Proteomics 14, 69–68 (2015).

46. Kessner, D., Chambers, M., Burke, R., Agus, D. & Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2533–2536 (2008).

47. Eng, J. K., Jahn, T. A. & Hoopmann, M. R. Comet: an open-source MS/MS sequence database search tool. Proteomics 13, 22–24 (2013).

48. Beavis, R. C. Using the global proteome machine for protein identification. Methods Mol. Biol. 328, 217–228 (2006).

49. Deutsch, E. W. et al. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteom. Clin. Appl. 9, 755–754 (2015).

50. Shysterberg, D. et al. iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol. Cell. Proteomics 10, M111.007690 (2011).

51. Wang, J. et al. MSPLIT-DIA: sensitive peptide identification for data-dependent acquisition. Nat. Methods 12, 1106–1108 (2015).

52. Kim, S. et al. The generating function of CID, ETD, and CID/ETD pairs of tandem mass spectra: applications to database search. Mol. Cell. Proteom. 9, 2840–2852 (2010).

53. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 210–214 (2007).

54. The, G. et al. SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J. Proteom. 100, 37–43 (2014).

55. Uniprot Consortium. Uniprot: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515 (2019).

56. Herrero, J. et al. Ensemble comparative genomics resources. Database 2016, bav096 (2016). https://doi.org/10.1093/database/bav096.

57. Yates, A. et al. The ensemble REST API: ensemble data for any language. Bioinformatics 31, 143–145 (2015).

58. El-Gebali, S. et al. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432 (2019).

59. Grant, B. J., Rodrigues, A. P. C., Elsawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22, 2695–2696 (2006).

60. Galli, T. dendextend: a robust and accurate method for quantification of colony sizes from plate images. J. Vis. Exp. 105, e52255 (2015).

61. Hagenaars, G. and C. Gagne-Thivierge for critical reading of the paper. This work was supported by a Canadian Institutes of Health Research (CIHR) Foundation grant 387697 and a HFSP grant (RGP00034/2016) to CRJ, CIHR Project grant 162439 to N.D., Discovery Grants 1304616–2017 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to J.P.L., by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (under Award R01GM105978 to N.D.) and the National Sciences and Engineering Research Council of Canada (NSERC) (Discovery Grants RGPIN-2016-05557 to N.D.). CRJ holds the Canada Research Chair in Cellular Systems and Synthetic Biology. N.D. holds the Canada Research Chair in Cancer Proteomics. J.P.L. is supported by a Janie 1 salary award from the Fonds de Recherche du Québec-Santé (FRQ-S). N.D. holds a Fonds de Recherche Québec-Santé (FRQ-S) Research Scholar Senior Career Award (281993). U.D. holds an Alexander-Graham Bell Canada graduate scholarship and the Louis-Poirier scholarship of the Fondation du CHU de Québec-Université Laval. E.B. was supported by an NSERC graduate scholarship. R.D. is funded by an FRQ-S postdoctoral fellowship. S.D. was supported by a MTAcs globalink scholarship. P.D. is supported by a CIRI Vanier Canada Graduate Scholarship. NTPH is the recipient of an FRQs Doctoral Training Scholarship.

Acknowledgements
We thank Bisson and Landry lab members for discussions; Véronique L'italien for expert assistance with figure artwork; N. A. L. A. H. E. S. W. E. T. M. C. O. M. A. L. S. A. and Gagné-Thivierge for critical reading of the paper. This work was supported by a Canadian Institutes of Health Research (CIHR) Foundation grant 387697 and a HFSP grant (RGPI0034/2016) to CRJ, CIHR Project grant 162439 to N.D., Discovery Grants 1304616–2017 from the Natural Sciences and Engineering Research Council of Canada (NSERC) to J.P.L., by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (under Award R01GM105978 to N.D.) and the National Sciences and Engineering Research Council of Canada (NSERC) (Discovery Grants RGPIN-2016-05557 to N.D.). CRJ holds the Canada Research Chair in Cellular Systems and Synthetic Biology. N.D. holds the Canada Research Chair in Cancer Proteomics. J.P.L. is supported by a Janie 1 salary award from the Fonds de Recherche du Québec-Santé (FRQ-S). N.D. holds a Fonds de Recherche Québec-Santé (FRQ-S) Research Scholar Senior Career Award (281993). U.D. holds an Alexander-Graham Bell Canada graduate scholarship and the Louis-Poirier scholarship of the Fondation du CHU de Québec-Université Laval. E.B. was supported by an NSERC graduate scholarship. R.D. is funded by an FRQ-S postdoctoral fellowship. S.D. was supported by a MTAcs globalink scholarship. P.D. is supported by a CIRI Vanier Canada Graduate Scholarship. NTPH is the recipient of an FRQs Doctoral Training Scholarship.
**Author contributions**

U.D., A.K.D., F.C., R.D., N.B., and C.R.L. designed the experiments. U.D., E.B., A.K.D., R.D., S.D., F.C., G.D.G., N.T.H.P., and M.L. performed the experiments. D.B., P.D., R.D., S.D., and C.R.L. designed and performed the computational analyses. U.D., E.B., D.B., R.D., P.C.D., C.R.L., J.P.L., and N.D. analyzed the results. U.D., N.B., and C.R.L. wrote the paper with contributions from all authors. N.B. and C.R.L. supervised the research.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21873-2.

**Correspondence and requests for materials** should be addressed to N.B. or C.R.L.

**Peer review information** Nature Communications thanks Ben Collins and other, anonymous, reviewers for their contributions to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021