

Journal Pre-proof

A family-wide assessment of latent STAT transcription factor interactions reveals divergent dimer repertoires

Andreas Begitt, Sebastian Krause, James R. Cavey, Doratha E. Vinkemeier, Uwe Vinkemeier

PII: S0021-9258(23)00345-9

DOI: <https://doi.org/10.1016/j.jbc.2023.104703>

Reference: JBC 104703

To appear in: *Journal of Biological Chemistry*

Received Date: 17 January 2023

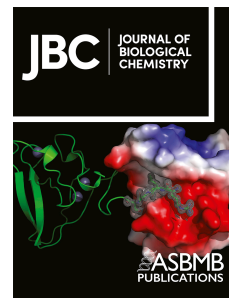
Revised Date: 4 April 2023

Accepted Date: 6 April 2023

Please cite this article as: Begitt A, Krause S, Cavey JR, Vinkemeier DE, Vinkemeier U, A family-wide assessment of latent STAT transcription factor interactions reveals divergent dimer repertoires, *Journal of Biological Chemistry* (2023), doi: <https://doi.org/10.1016/j.jbc.2023.104703>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.



A family-wide assessment of latent STAT transcription factor interactions reveals divergent dimer repertoires

Andreas Begitt¹, Sebastian Krause¹, James R. Cavey¹, Doratha E. Vinkemeier², and Uwe Vinkemeier^{1*}

1 The University of Nottingham, School of Life Sciences, Nottingham, UK

2 Nottingham Trent University, School of Computer Science, Nottingham, UK

* Corresponding author

Abstract

The conversion of STAT proteins from latent to active transcription factors is central to cytokine signalling. Triggered by their signal-induced tyrosine phosphorylation, it is the assembly of a range of cytokine-specific STAT homo- and heterodimers that marks a key step in the transition of hitherto latent proteins to transcription activators. In contrast, the constitutive self-assembly of latent STATs and how it relates to the functioning of activated STATs, is understood less well. To provide a more complete picture, we developed a co-localization-based assay and tested all 28 possible combinations of the seven unphosphorylated STAT (U-STAT) proteins in living cells. We identified five U-STAT homodimers —STAT1, STAT3, STAT4, STAT5A and STAT5B— and two heterodimers —STAT1:STAT2 and STAT5A:STAT5B— and performed semi-quantitative assessments of the forces and characterizations of binding interfaces that support them. One STAT protein —STAT6— was found to be monomeric. This comprehensive analysis of latent STAT self-assembly lays bare considerable structural and functional diversity in the ways that link STAT dimerization before and after activation.

Introduction

The STAT (signal transducer and activator of transcription) proteins are an evolutionarily conserved family of seven transcription factors in mammals, namely STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B and STAT6 [1]. These proteins are the substrate of both receptor and non-receptor tyrosine kinases such as the Janus kinases (JAK), which catalyse their phosphorylation of a single C-terminal tyrosine residue after ligand binding to cell surface receptors [2]. About 50 different growth factors and cytokines including interferons, interleukins and growth hormones are known to signal via STAT proteins [3]. The activated STATs accumulate in the nucleus and participate in the transcription of hundreds of genes [4]. This sequence of events, usually referred to as canonical JAK-STAT signalling, entails the dimerization of STAT proteins through mutual phosphotyrosine:SH2 domain interactions. Such dimers can bind short palindromic stretches of DNA called GAS elements in the promoter region of target genes [5]. In addition to homodimers, activated STATs can assemble heterodimers with the other family members [6].

While the phosphotyrosine-mediated functions of STATs were linked to their dimerization early on [7], much less is known to date regarding the constitutive self-assembly of latent, that is unphosphorylated STATs (U-STATs). Initially, U-STATs were believed to be monomeric [7], but further studies showed that STATs can assemble high molecular weight complexes already before tyrosine phosphorylation [8, 9]. A well-documented example is STAT1, which forms equally strong dimers before and after activation as shown by analytical ultracentrifugation [10]. Importantly, crystallographic evidence demonstrates

that unphosphorylated dimers of STAT1, STAT3 and STAT5 are stabilized not by interactions of the carboxy-terminal SH2 domains, but by interactions between amino terminal regions [11, 12, 13], resulting in antiparallel orientation of monomers as opposed to their parallel orientation in the phosphodimers [14]. It has been recognized that unphosphorylated STATs are involved in many biological events in both normal and pathological situations [15, 16, 17, 18, 19]. In some instances, the self-assembly of U-STATs has been linked to tyrosine phosphorylation and cytokine signalling. For example, the homodimerization of U-STAT4 is a prerequisite for cytokine-induced STAT4 activation [20], and heterodimerization of U-STAT1 with U-STAT2 has been shown to have positive or negative consequences for type 1 and type 2 interferon signalling, respectively [21, 22]. These and other findings indicate that U-STAT dimers fulfil roles critical for cytokine functioning, but the knowledge of dimerization before cytokine-induced phosphorylation remains incomplete. This is particularly true for heterotypic interactions across the STAT family and the *in-vivo* situation generally. To fill in knowledge gaps, we devised an assay to explore systematically the repertoires of homo- and heterotypic interactions among U-STATs in living cells. The assay is based on U-STATs being nucleocytoplasmic shuttling proteins that can freely cross the soft diffusion barrier posed by the nuclear pore, presumably via direct contacts with nuclear pore proteins [23, 24]. It results in generally pancellular distributions of U-STATs [25], which for STAT1 and STAT2 can be shifted to nuclear or cytoplasmic accumulation by tagging them with transferable carrier-dependent nuclear localization (NLS) or nuclear export (NES) signals [26]. We reasoned that such signal-tagged variants might function as baits that attract co-expressed test proteins if binding interactions occurred. This would be evident by their co-localization in the bait protein's compartment. After rigorous testing and verification, this approach was used to probe homo- and heterotypic binding interactions within cells across the entire STAT family for the first time.

Results

Co-localization as an assay to detect the dimerization of U-STATs within cells

Latent STATs are nucleocytoplasmic shuttling proteins that can freely cross the nuclear envelope by directly contacting nuclear pore proteins and additional carrier-mediated mechanisms [23]. Accordingly, they display near pancellular distributions in cells before cytokine-treatment (with the exception of STAT2, see below), which is preserved upon C-terminal fusion of fluorescent marker proteins such as mEGFP or mCherry (Supporting Fig. 1A, panels 1,12,16,20,24,28). A Western blot demonstrating expression of full-length STAT-fluorophore fusion proteins is shown in Supporting Fig. 1B. As mentioned above, STAT2 deviates from the near pancellular distribution of the other STAT proteins; it accumulates in the cytoplasm due to potent nuclear export activity in its C-terminal transactivation domain (Supporting Fig. 1A, panel 6). Consistent with previous observations [26, 27], STAT2 variants with truncated transactivation domain, referred to as STAT2 Δ C, showed strongly reduced cytoplasmic accumulation, thus adopting a distribution more like the other STAT family members (Supporting Fig. 1A, panel 9). Addition of transferable heterologous nuclear export (NES) or nuclear import signals (NLS) to wild-type STATs (or the C-terminally truncated STAT2) allowed us to direct bait proteins to the cytoplasm or the nucleus, respectively (Supporting Fig. 1A, panels 3,4,10,11,14,15,18,19,22,23,26,27,29). We used two well-characterized and highly active signals, namely, a NES derived from protein kinase A inhibitor and the NLS from simian virus 40 large T-antigen [28, 29]. These nuclear translocation-related features potentially allow STATs to alter the localization of proteins they interact with. Schematics of the assay design are shown in Figure 1A.

Another important aspect of the assay we considered was the strength of binding interactions, as this posed a main constraint on whether they are detectable as co-localization. Current data on U-STAT dimer assembly is largely qualitative, however, equilibrium sedimentation studies indicate that unphosphorylated STAT1 and STAT3 form high-affinity homodimers, both with a K_d in the low nanomolar range [10, 30]. Moreover, structural and mutational studies of the dimer interfaces have identified the conserved N-domain and key anchoring residues therein as critical for unphosphorylated dimer assembly. Unphosphorylated STATs with N-domain truncation or mutated dimerization hot spot residues retain nucleocytoplasmic shuttling and their subcellular distributions differ little from their wild-type counterparts [31, 32] (see also Supporting Fig. 1A, panels 2,8,13,17,21,25). Nonetheless, these mutations render them essentially monomeric, as quantitative analyses of N-domain-deleted or residue F77-mutated STAT1 showed a > 100-fold drop in binding affinities [10, 12]. For STAT3, mutation of residue L78 similarly results in unphosphorylated dimer dissociation in living cells as demonstrated by Förster resonance energy transfer [31]. These established results allowed us to use wild-type and mutant STAT1 and STAT3 as positive and negative controls to test if the co-localization of STAT proteins could be a reliable indicator for their dimerization inside cells. As shown in Figure 1B, co-expression of NES-fused STAT3 with wild-type STAT3 indeed resulted in their co-localization in the cytoplasm. Importantly, co-localization was lost if the monomeric STAT3-L78R mutant was used, which displayed pan-cellular distribution irrespective of the presence of STAT3-NES. This was the expected behaviour of non-interacting, monomeric STAT3, in accord with the aforementioned *in-vitro* and cellular studies. We then examined the homodimerization of STAT1 to validate further the co-localization assay. In the same way that STAT3 interacted with STAT3-NES, wild-type STAT1 co-localized with STAT1-NES in the cytoplasm, suggestive of homodimerization (Fig. 1B). To examine whether dimerization can also occur in the cell nucleus, we co-expressed STAT1 and nuclear-accumulated STAT1-NLS. As shown in Figure 1B, this likewise resulted in the co-localization of the two STAT1 variant proteins, albeit in the nucleus, demonstrating that dimerization was not limited to the cytoplasmic compartment. To test the consequences of a dimer-disrupting mutation for STAT1 homodimers, we used the mutant STAT1-F77A. In contrast to wild-type STAT1, the F77A mutant did not accumulate in the nucleus with co-expressed STAT1-NLS but retained pan-cellular distribution, similar to STAT3-L78R, thus again presenting the expected behaviour of monomeric STATs. Finally, it was important to ascertain that failure of a test protein to co-localize with a bait was correctly attributed to a lack of dimer formation, rather than a shortfall of bait protein expression. Therefore, the relative expression of test and bait proteins were determined using quantitative fluorescence imaging as described in Experimental Procedures, and cells were disregarded for inclusion in co-localization analyses if the expression of bait variants was below equimolar levels or exceeded the test protein's concentration by more than 4-fold.

In summary, co-localization in nuclear or cytoplasmic compartments was a reliable indicator to assess the homodimerization of unphosphorylated STAT1 and STAT3. We therefore expanded this approach to assess homodimer formation of all seven STATs as well as their ability to heterodimerize.

Latent STATs assemble five homodimers and two heterodimers

To similarly examine the homodimerization of the other five STATs, we co-expressed STAT4, STAT5A, STAT5B and STAT6 with their respective NES-tagged counterparts. As shown in Figure 2A, STAT6 failed to accumulate in the cytoplasm, in stark contrast to STAT4, STAT5A and STAT5B, which co-localized with their NES-tagged equivalents. We

quantified the extent of co-localization for each experiment by calculating Pearson's correlation coefficients (r_P) for 10-30 cells. Numerical values for homodimers are shown in Figure 2C and summarized in Figure 3. STAT1, STAT3, STAT4, STAT5A and STAT5B have Pearson's correlation coefficients of 0.94 or higher, indicative of near complete co-localization due to stable homodimerization. STAT6 diverged strongly as it showed no cytoplasmic co-localization, with an accordingly significantly lower r_P value of 0.41. To assess the homodimerization of STAT2, the C-terminally truncated variant STAT2 Δ C with strongly reduced intrinsic nuclear export activity was used, as described above (Supporting Fig. 1A, panel 9). Although the C-terminus is needed for efficient constitutive nuclear export of STAT2, this region generally appears to be dispensable for the dimerization of U-STATs [10, 33], including the heterodimerization of U-STAT2 Δ C and U-STAT1, which was demonstrated using STAT1-NES and STAT2 Δ C-NLS as baits (Supporting Fig. 2). The C-terminally truncated STAT2 mutant was therefore co-expressed with wild-type STAT2 to probe U-STAT2 homodimerization. We observed incomplete cytoplasmic co-localisation (Fig. 2A) and a correlation coefficient $r_P = 0.71$ that was significantly reduced compared to stable dimers of other STAT family members, yet higher than for non-interacting monomeric STAT6 (Fig. 2C). This could signify genuine homodimer assembly, albeit with lower affinity, or merely constitute an artefact reflecting co-localization due to residual cytoplasmic accumulation of STAT2 Δ C. To distinguish between these possibilities, STAT2 Δ C was directed to the nucleus through its tagging with an NLS (Supporting Fig. 1A, panel 11) to examine if this resulted in the nuclear localization of co-expressed STAT2 Δ C. This was not the case, however, as the slight cytoplasmic accumulation of STAT2 Δ C appeared unaltered, with no indication of increased nuclear translocation in the presence of the nuclear accumulated STAT2 bait protein. The correlation coefficient accordingly dropped sharply into the negative, $r_P = -0.45$ (Fig. 2C) which indicated opposite distributions and hence a lack of co-localization of the two STAT2 variants. We concluded that unphosphorylated STAT2, like STAT6, did not homodimerize, in contrast to STAT1, STAT3, STAT4, STAT5A and STAT5B, which formed stable homodimers in living cells. Next, we probed heterodimer assembly of the seven unphosphorylated STATs by co-expressing NES-tagged STAT proteins as baits (except STAT2, where wild-type was used) and untagged wild-type STATs as the test proteins. Of the 21 possible heterotypic pairings, only two showed co-localization of bait and test proteins, namely U-STAT1:STAT2 and U-STAT5A:STAT5B (Fig. 2B, 2C), with r_P values of 0.97 in both cases. All other combinations, including STAT1:STAT3 or STAT3:STAT4 (Fig. 2B), which readily assemble heterodimers upon their cytokine-induced tyrosine-phosphorylation [6], did not appear to heterodimerize in the absence of cytokine stimulation, and the corresponding r_P values were accordingly low (Fig. 3). Thus, unphosphorylated STATs were generally present as stable dimers, predominantly as homodimers. Heterodimers were formed only between STAT1 and STAT2 and the two very closely related STAT5 proteins. STAT6 was the only family member devoid of detectable dimerization activity.

Latent STAT dimers require N-domain interactions

Several lines of experimental inquiry indicate that unphosphorylated STAT dimers adopt an antiparallel conformation that is dependent on N-domain interactions [8]. Deletion of the N-domain, and even single N-domain point mutations, can therefore result in the dissociation of dimers. For STAT1, N-domain residues phenylalanine 77 and leucine 78 have been shown to be critical for assembly of unphosphorylated dimers [12], and likewise the leucine residues 78 for STAT3 [31, 33] and STAT4 [20], see also Figure 1. We therefore mutated the homologous N-domain residues of STAT2 (L82A), STAT4 (L78S), STAT5A (L82A) and STAT5B (L82A) to examine which of the unphosphorylated STAT dimers shared the dependence on N-domain interactions. As is shown in Figure 4A, all

seven homo- and heterodimers were destabilized upon mutation of the same single homologous side chain in the N-domain. We concluded that the U-STAT dimers adopted similar N-domain-mediated conformations. However, the dissociating effect on the U-STAT1:STAT2 heterodimer was comparatively weak, as indicated by the relatively small albeit statistically significant reduction in its r_p value (Fig. 4B). This could be an indication that this STAT dimer adopted an exceptionally stable conformation, which was tested next.

U-STAT1:STAT2 heterodimers are exceptionally stable

To compare the relative binding strengths of U-STAT dimers, we co-expressed STATs that harboured opposed localization signals, namely, PKI NES or SV40 NLS. The nuclear export activity conferred by the PKI NES was determined to dominate over import activity associated with the NLS of SV40, since proteins such as GFP or GST accumulate in the cytoplasm when both signals are appended simultaneously [35]. Of note, the same outcome, namely accumulation in the cytoplasm, was observed if STAT1 was used as the acceptor of the two opposed signals (Supporting Fig. 1A, panel 5). We reasoned that subjecting the different dimers to these same antipodal translocation activities might reveal differences in the forces driving U-STAT association (Fig. 5A). As shown in Figures 5B and 5D, the five U-STAT homodimers and the U-STAT5A:STAT5B heterodimer showed the same behaviour, that is, the dimer subunits localized to the nucleus or the cytoplasm in accordance with their respective localization signals. We inferred that the opposed translocation forces dissociated these dimers. In notable difference, STAT2 and the nuclear-targeted STAT1-NLS variant sustained their cytoplasmic co-localization (Fig. 5C, 5D), suggesting that U-STAT1:STAT2 heterodimers uniquely resisted dissociation. However, the co-localization of STAT2 and STAT1-NLS was lost upon the alanine mutation of STAT2 hot spot interface residue L82 (Fig. 5C, 5D). The weakening of U-STAT1:STAT2 interactions caused by this mutation (see Fig. 4A, 4B) evidently sufficed to reduce the binding affinity below the threshold required for continued co-localization, as complete separation of STAT1-NLS and STAT2 localizations was observed. To corroborate these observations we assessed the strength of heterodimerization using C-terminally-modified STAT2, such that its less-well characterized intrinsic NES activity was removed and replaced by the known dominant NES activity of PKI that was used also for the other family members. The outcome, however, was unchanged irrespective of the specific NES activities used, as STAT1-NLS again co-localized with the STAT2-NES variant in the cytoplasm (Fig. 5C, 5D). In agreement with this reasoning, the two STATs retained their co-localization, albeit in the nucleus (Fig. 5C), if STAT2's CRM1-mediated nuclear export was disabled by the inhibitor leptomycin B [26]. Of note, in the absence of co-expressed STAT1-NLS, treatment with leptomycin B resulted in pan-cellular STAT2 distribution but not its nuclear accumulation, in line with continued carrier-independent nuclear export likely being able to counter STAT2's constitutive intrinsic import activity (Supporting Fig. 1A, panel 7). We concluded that the heterodimers of unphosphorylated STAT1 and STAT2 are distinguished by exceptionally strong binding interactions.

Discussion

Activated STAT proteins assemble a rather well defined set of homo- and heterodimers that are crucial for cytokine signalling. The associations of latent STATs before activation, and how they relate to the cytokine-induced range, are less understood. Here, we report the first family-wide study on homo- and heterotypic interactions between unphosphorylated STAT proteins, which reveals that latent and activated STAT dimer repertoires overlap only

partially. These results were obtained by evaluating the co-localization of nucleus- or cytoplasm-targeted STAT proteins with co-expressed wild-type or mutated family members. Like previous translocation-based methods for detecting protein–protein interactions [35, 36], our assay uses fluorophore-fused proteins and transient transfections, which can result in substantially increased protein concentrations. Because the amount of protein complexes formed is very much linked to the protein's concentration, we compared experimental overexpression and natural STAT concentrations. STAT1 is the only family member for which the natural concentration in our HeLa cell model is known, where it is present at ~ 40 nM [10]. Such a low concentration is typical of transcription factors, however, STAT gene and protein expressions can be substantially increased in certain cell types, e.g. NK cells contain 10-20-fold more STAT4 protein than non-NK cells [37]. Additionally, cytokine-dependent positive feedback loops can cause STAT protein levels to rise several-fold [38, 39, 40, 41], for example in chronic inflammatory situations such as autoimmune diseases [42], infections [43, 44, 45] or cancers [46, 47, 48]. As shown in Supporting Figure 3, STAT1 expression was increased about 3- to 15-fold in the majority of transiently transfected HeLa cells in the co-localization assay but could go up to ~ 20 -fold. In light of the literature, this probably includes levels at the upper limit of what can be found in nature, but not necessarily beyond it. These considerations apply to all family members, since the different STATs were expressed at similar levels in our experiments, as indicated by the recording of similar fluorescence intensities. Moreover, we assume that the endogenous STATs add to the total cellular STAT concentration, but otherwise act like the transfected counterparts and do not change their behaviour. The indistinguishable binding of endogenous and transfected STAT1 to STAT2 baits is in line with this assumption (Supporting Fig. 3C). Although elevated protein concentrations promote self-assembly and thus may give rise to false-positive results, one STAT protein was exclusively monomeric in our hands, namely U-STAT6. We are aware of only a single report on the complex formation of unphosphorylated STAT6. Lackmann and colleagues probed the molecular size distribution of several STAT proteins in detergent-free HeLa cell extracts and found that they resided in cytoplasmic high-molecular weight complexes prior to any cytokine action, except for STAT6, which fractionated according to its monomeric molecular weight, consistent with our observations in living cells [49]. For the other STATs, the co-localization assay indicated that they formed dimers. This is in agreement with current biochemical, structural, and *in-vivo* results that demonstrate homodimerization of U-STAT1 [10, 12, 30, 50], U-STAT3 [30, 52, 53, 54, 55, 56], U-STAT4 [20] and U-STAT5A [13, 57]. STAT2, in contrast, did not assemble homodimers but heterodimers with STAT1. The heterodimerization of U-STAT1 and U-STAT2 has previously been documented by several independent lines of evidence [21, 22, 58, 59, 60]. We observed just one additional unphosphorylated heterodimer, which was assembled by the STAT5A and STAT5B proteins, which are 93.5% identical. Of the possible further 19 heterodimer combinations, none were observed with the co-localization assay. To our knowledge, these as well as the U-STAT5 heterodimer and absence of U-STAT2 homodimers have not previously been disproven or demonstrated, with the exception of unphosphorylated STAT1:STAT3 heterodimers, which have repeatedly been reported using invasive methods such as co-immunoprecipitations [57, 61, 62, 63, 64]. Several authors noted sensitivity of U-STAT1:STAT3 interactions to salts and detergents [59, 63], they observed them only in cytoplasmic extracts [62], or at low abundance [61], suggesting weak binding forces. A quantitative assessment of U-STAT1 dimers using *in situ* single cell pull-down of GFP-tagged STAT protein complexes reached this conclusion, too, as substantially weaker binding and rapid dissociation of U-STAT1:STAT3 compared to U-STAT1 homodimers and U-STAT1:STAT2 heterodimers was observed [60]. These data and our results indicate stability of U-STAT1:STAT3 heterodimers under certain immunoprecipitation conditions rather than constitutive complex formation within living cells.

Previous work has shown that different nuclear import and export signal sequences differ in the binding strength to their respective importin or exportin carrier molecules, whereby binding affinities measured *in vitro* and the transport activities observed in living cells generally correlate well [28, 29]. Although interactions with non-carrier proteins can also have a decisive role, the balance between the strengths of import and export signals is often critical for determining the steady-state localization of soluble proteins. We took advantage of this phenomenon and appended just one of the opposing signal sequences to dimer-forming STATs. Upon their co-expression they functioned as protomers of STAT dimers, and their antagonistic translocation activities were used to assess their strength of association, which to our knowledge is a method not previously reported. Since the opposing translocation activities of PKI and SV40 are among the strongest known, it is perhaps unsurprising that they precluded co-localization indicative of dimer assembly in all but one case, namely U-STAT1:STAT2 heterodimers. Notably, the heterodimers resided in the cytoplasm, mirroring the behaviour of reporters where the dominant PKI nuclear export signal and the SV40 import signal operate on the same molecule [35]. This outcome suggests exceptionally strong binding interactions between the unphosphorylated STAT1 and STAT2. To date, *in vitro* binding affinities for full-length STAT2 homodimers or heterodimers with STAT1 have not been determined. However, heterodimers formed by their N-domains are of very high affinity with a dissociation constant in the low nanomolar range, thus even exceeding the affinity of full-length U-STAT1 homodimers [10, 21]. Homotypic STAT2 N-domain interactions, in contrast, are several 1000-fold weaker [21], in line with the observed absence of U-STAT2 homodimers in cells.

For activated STATs, dimerization is essential to bind DNA and hence their function as transcription factors. A similar functional imperative for dimerization does not appear to exist for latent STATs. For STAT6, which was monomeric before cytokine stimulation in our assay, the recruitment to cytokine receptors, kinase interactions and tyrosine phosphorylation and subsequent assembly of activated dimers evidently do not require latent dimers. Likewise, the events associated with cytokine-induced activation of STAT1, STAT3 and STAT5A also proceed largely undisturbed in the presence of mutations that dissociate their unphosphorylated dimers [31, 32]. STAT4 is an outlier in this regard, as it requires unphosphorylated dimers to become activated and execute its cytokine-inducible activities [20]. Assembly of constitutive dimers hence does not appear to be a universal requirement for the subsequent activation of STAT proteins. The expansion of the heterodimer repertoire from just two before tyrosine phosphorylation to at least eight thereafter (STAT1:STAT2, STAT1:STAT3, STAT1:STAT4, STAT2:STAT6, STAT3:STAT4, STAT3:STAT5A/B, STAT5A:STAT5B) [6] also argues against a necessity for constitutive dimers to assemble activated counterparts. Nonetheless, the binding interfaces of unphosphorylated STAT dimers share highly conserved hotspot residues that contribute substantially to homo- and heterodimerization. This suggests that there must be an enormous selection pressure to maintain these interfaces and the assembly of latent STAT dimers. In fact, mutations in STAT1 and STAT3 that dissociate latent dimers give rise to rare genetic disorders of the immune system. Of note, the alteration of a single protomer sufficed in the co-localization assay to achieve dimer dissociation, which mimics the heterozygous germline mutations in patients with STAT1 or STAT3 gain-of-function disease [65, 66]. Remarkably, it is not lowered but heightened STAT activity that results from such mutations and causes disease [4]. Activated STAT dimers are thought to oscillate between parallel and antiparallel conformations, whereby the antiparallel conformation is similar or identical to the conformation of unphosphorylated latent dimers [10]. Importantly, antiparallel dimers of activated STATs are the substrate of tyrosine dephosphorylation [14], which is why mutations that dissociate unphosphorylated dimers can also affect activated dimers and cause resistance to dephosphorylation and heightened tyrosine phosphorylation *in vivo* [67, 68], which are defining features of STAT

gain-of-function diseases. In other words, latent dimers may not be maintained as the precursors of activated STAT dimers but rather as direct products of their inactivation. Moreover, interfaces that stabilize latent dimers can participate in additional vital activities after STAT activation, such as cooperative DNA binding mediated by the N-domains, which adds to evolutionary pressures to maintain them [32, 69, 70].

In conclusion, this comprehensive analysis of latent STAT self-assembly shows that dimerization of unphosphorylated STATs is linked to the regulation of STAT transcription activity in at least three distinct ways. One requires latent dimers for activation and (probably) inactivation and would apply to STAT4. A second way, where latent dimers seem to be dispensable for both activation and inactivation, and this applies to STAT6. The final way, where STAT activation occurs essentially normally in the absence of latent dimerization, but where the latent dimer conformation is necessary for inactivation and applies to the remaining members of the STAT protein family.

Experimental procedures

Cell culture and transfections

HeLa (ECACC 93021013) and HEK293T (ECACC 85120602) cells were grown in DMEM (Sigma D6429), supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS; Sigma F9665) and 1% (v/v) penicillin/streptomycin (Sigma P0781; with 10,000 units penicillin and 10 mg streptomycin per mL in 0.9% (w/v) NaCl), in a humidified incubator with 5% CO₂ at 37°C. Cells were transfected at ~80% confluence using Lipofectamine LTX according to the manufacturer's recommendations (Invitrogen) with the following modifications. For co-localization assays, HeLa cells were transfected in 24-well plates with 1.2 µg DNA, 2 µl lipofectamine and 2 µl PLUS reagent per well, whereby DNA encoding the bait protein was used in about 2-fold excess. Where indicated, cells were treated with 10 ng/ml leptomycin B (LMB; Cell Signaling Technology) to inactivate NES-mediated nuclear export.

Expression constructs

STAT proteins were expressed with mCherry or mEGFP (monomeric EGFP [71]) fused to the C-terminus through cloning into pmCherry-N1 (Clontech) and pmEGFP-N1 (derived from pEGFP-N1 (Clontech), encoding the A206K mutant of EGFP), respectively. Cytoplasmic or nuclear accumulation of fusion proteins was achieved by appending canonical nuclear export (NES) or nuclear import (NLS) signal sequences. We used the NES from protein kinase A inhibitor (PKI, ³⁴NSNELALAGLDINK⁴⁹) and the NLS from simian virus 40 large T-antigen (SV40, ¹²⁶PKKKRKV¹³²) [28, 29]. Where a single signal was appended, the respective cDNA sequence was placed between STAT and fluorophore, encoding additional 3-5 heterologous residues N- and C-terminal of the NES or NLS sequence. To express STAT1 with dual NES and NLS signals, plasmid pSTAT1-NES-mEGFP was used and the SV40 NLS (plus an additional C-terminal glutamate residue) was placed downstream of the fluorophore. STAT2ΔC is a C-terminally truncated variant expressing residues 1-703. N-domain mutations were as follows, STAT1-F77A, STAT2-L82A, STAT3-L78R, STAT4-L78S, STAT5A-L82A, STAT5B-L82A. Plasmid pmEGFP-mCherry encoded mEGFP fused to the N-terminus of mCherry. The two fluorophores were linked via an extended rigid helical linker (¹YSDLELAEAAAKEAAAKEAAAKEAAAKEAAAKAARDPPVAT⁴²) to minimize basal Förster resonance energy transfer [72]. The mutations were introduced using the Q5 site-directed mutagenesis kit (New England Biolabs). Sequences of all the plasmids were confirmed by DNA sequencing. Molecular cloning details are available upon request.

Fluorescence imaging

Twenty hours after transfection, cells were fixed with mild agitation in 2% (v/v) paraformaldehyde in PBS for 15 min at room temperature followed by staining of nuclei for 3 min with 2.5 µg/ml Hoechst 33258 (Sigma Aldrich) and mounting in fluorescence mounting medium (Dako S3023). A Zeiss AxioPlan 2 microscope with Zeiss Plan Apochromat 63x (NA=1.4) oil immersion objective and with FITC ("green channel") and Texas red ("red channel") filter sets for recording EGFP and Cherry / Cy3 emissions, respectively, and Zeiss' AxioVision 4.7 software were used for wide-field fluorescence imaging. Fourteen-bit black and white images were captured with a digital AxioCam CCD camera (Carl Zeiss Jena). Fluorescence quantitation and image analysis were performed using ImageJ [73]; total cellular EGFP and Cherry signals were measured by calculating the integrated pixel intensity. All pixel values were measured below saturation limits. Adobe Illustrator (Adobe) was used to present images in the figures.

Principle of co-localization assay and quantification of protein co-localization

The assay scores the extent of co-localization of a cytoplasmic or nuclear accumulated STAT protein (STAT-NES or STAT-NLS), which functions as the bait, with another STAT, the test protein, whereby the two proteins are fused to different fluorophores. We used mCherry and mEGFP, as they are monomeric, and their emission spectra display minimal overlap [71]. Co-localization is taken as a proxy for the degree of dimerization via homo- or heterotypic interactions. In the basic assay configuration, wild-type STAT proteins were tested. To assess dimer conformation, we tested single point mutants homologous to key interacting residues of STAT1 (F77 or L78), which are critical for the assembly of antiparallel dimers of the unphosphorylated protein [12, 14]. To assess dimer stability, bait and test proteins were furnished with opposed translocation signals (STAT-NES co-expressed with STAT-NLS) to provide forces that counteract the association of the two protomers in STAT dimers. For all assay configurations, knowledge of the relative cellular expression levels of the two co-expressed STAT proteins is necessary to ensure bait proteins are present in excess. As the STATs were expressed as fusions with mEGFP or mCherry, information about their relative expression levels could be obtained by comparison with a mEGFP-mCherry fusion protein, which expressed mEGFP and mCherry fluorophores in a known and fixed ratio of one-to-one. To facilitate quantification of the co-localization assay, all images were acquired using exposure times calibrated such that the mEGFP and mCherry fluorescence intensities were approximately equal (Supporting Fig. 4). We found that this acquisition setting reproducibly gave equivalent readings even when used on different days. Using these setting, cells were imaged, and single cells were manually segmented in ImageJ. Cells were eligible for quantitative co-localization analyses if the bait STAT protein was present from equimolar amounts to up to 4-fold molar excess compared to the co-expressed test variant. To obtain quantitative estimates of the degree of co-localization in the images, Pearson (r_p) and Spearman (r_s , given in the source data files) correlation coefficients were calculated for individual cells with the PSC co-localization ImageJ plug-in as described [74]. Both tests produce values in the range $[-1, 1]$, 0 indicating that there is no discernible correlation and -1 and $+1$ meaning strong negative and positive correlations, respectively. Cells where fewer than 1000 pixels had intensity values above that which might be considered image noise for at least one of the channels at that data point were also deemed ineligible and excluded from further calculations. The PSC programme's default intensity setting of 40 was used as the image noise threshold. Values for Pearson correlation coefficients were calculated for 10-30 eligible cells per experiment and are given as means \pm standard deviation (SD).

Statistical analyses

Statistical analyses were performed in GraphPad Prism Software Version 9.3.0. D'Agostino-Pearson's omnibus K2 was used to test for normal and lognormal distribution of individual variables. Kruskal-Wallis test in conjunction with Dunn's test to correct for multiple comparisons was used for hypothesis testing. The ROUT method with Q = 1% was utilized to identify outliers. Significance is designated as *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001.

Data availability

Fluorescence microscopy raw data from Figs. 2A-C, 3, 4A, B, 5B-D and Supporting Figs. 3A, B, and 4 are provided as publicly available source data files in the BioImage Archive (<https://www.ebi.ac.uk/bioimage-archive/>) under accession number S-BIAD669. All other remaining data are available within the article or supporting information.

Supporting information

This article contains supporting information.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Acknowledgments

We thank Seema Bagia and Robert Markus of School of Life Sciences Imaging for reagents and microscopy advice. This work was supported by BBSRC grant BB/V004824/1 (UV), MRC Confidence in Concept grant MC_PC_17173 (UV), and a fellowship from the Nottingham BBSRC Doctoral Training Partnership to SK.

References

1. Levy, D. E., and Darnell, J. E. Jr. (2002) STATs: transcriptional control and biological impact. *Nat. Rev. Mol. Cell Biol.* **3**, 651–662
2. Morris, R., Kershaw, N. J., and Babon, J. J. (2018) The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Sci.* **27**, 1984-2009
3. O'Shea, J. J., Schwartz, D. M., Villarino, A. V., Gadina, M., McInnes, I. B., and Laurence, A. (2015) The JAK-STAT pathway: impact on human disease and therapeutic intervention. *Annu. Rev. Med.* **66**, 311-328
4. Philips, R. L., Wang, Y., Cheon, H., Kanno, Y., Gadina, M., Sartorelli, V., *et al.* (2022) The JAK-STAT pathway at 30: Much learned, much more to do. *Cell* **185**, 3857-3876
5. Schindler, C., Levy, D. E., and Decker, T. (2007) JAK-STAT signaling: from interferons to cytokines. *J. Biol. Chem.* **282**, 20059-20063
6. Delgoffe, G. M., and Vignali, D. A. (2013) STAT heterodimers in immunity: A mixed message or a unique signal? *JAKSTAT* **2**, e23060

7. Shuai, K., Horvath, C. M., Huang, L. H., Qureshi, S. A., Cowburn, D., and Darnell, J. E. Jr. (1994) Interferon activation of the transcription factor Stat91 involves dimerization through SH2-phosphotyrosyl peptide interactions. *Cell* **76**, 821-828
8. Droscher, M., and Vinkemeier, U. (2012) Self-association of STAT proteins from monomers to paracrystals. In *Jak-Stat Signaling: From Basics to Disease*, Decker, T., Müller, M. (ed) pp 47-63, Vienna, Springer
9. Sehgal, P. B. (2008) Paradigm shifts in the cell biology of STAT signaling. *Semin. Cell Dev. Biol.* **19**, 329-340
10. Wenta, N., Strauss, H., Meyer, S., and Vinkemeier, U. (2008) Tyrosine phosphorylation regulates the partitioning of STAT1 between different dimer conformations. *Proc Natl Acad Sci. U. S. A.* **105**, 9238-9243
11. La Sala, G., Michiels, C., Kükenshöner, T., Brandstötter, T., Maurer, B., Koide, A., et al. (2020) Selective inhibition of STAT3 signaling using monobodies targeting the coiled-coil and N-terminal domains. *Nat. Commun.* **11**, 4115.
12. Mao, X., Ren, Z., Parker, G. N., Sondermann, H., Pastorello, M. A., Wang W., et al. (2005) Structural bases of unphosphorylated STAT1 association and receptor binding. *Mol. Cell* **17**, 761-777
13. Neculai D., Neculai, A. M., Verrier, S., Straub, K., Klumpp, K., Pfitzner, E., et al. (2005) Structure of the unphosphorylated STAT5a dimer. *J. Biol. Chem.* **280**, 40782-40787
14. Zhong, M., Henriksen, M. A., Takeuchi, K., Schaefer, O., Liu, B., ten Hoeve, J., et al. (2005) Implications of an antiparallel dimeric structure of nonphosphorylated STAT1 for the activation-inactivation cycle. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3966-3971
15. Chatterjee-Kishore, M., Wright, K. L., Ting J. P., and Stark, G. R. (2000) How Stat1 mediates constitutive gene expression: a complex of unphosphorylated Stat1 and IRF1 supports transcription of the LMP2 gene. *EMBO J.* **19**, 4111-4122
16. Cui, X., Zhang, L., Luo, J., Rajasekaran, A., Hazra, S., Cacalano, N., et al. (2007) Unphosphorylated STAT6 contributes to constitutive cyclooxygenase-2 expression in human non-small cell lung cancer. *Oncogene* **26**, 4253-4260
17. Majoros, A., Platanitis, E., Szappanos, D., Cheon, H., Vogl, C., Shukla, P., et al. (2016) Response to interferons and antibacterial innate immunity in the absence of tyrosine-phosphorylated STAT1. *EMBO Rep.* **17**, 367-382
18. Park, H. J., Li, J., Hannah, R., Biddie, S., Leal-Cervantes, A.I, Kirschner, K., et al. (2016) Cytokine-induced megakaryocytic differentiation is regulated by genome-wide loss of a uSTAT transcriptional program. *EMBO J.* **35**, 580-594
19. Stark, G. R., Cheon, H., and Wang, Y. (2018) Responses to cytokines and interferons that depend upon JAKs and STATs. *Cold Spring Harb. Perspect. Biol.* **10**, a02855
20. Ota, N., Brett, T. J, Murphy, T. L., Fremont, D. H., and Murphy, K. M. (2004) N-domain-dependent nonphosphorylated STAT4 dimers required for cytokine-driven activation. *Nat. Immunol.* **5**, 208-215
21. Ho, J., Pelzel, C., Begitt, A., Mee, M., Elsheikha, H. M., Scott, D. J., et al. (2016) STAT2 is a pervasive cytokine regulator due to its inhibition of STAT1 in multiple signaling pathways. *PLoS Biol.* **14**, e2000117
22. Wang, Y., Song, Q., Huang, W., Lin, Y., Wang, X., Wang, C., et al. (2021) A virus-induced conformational switch of STAT1-STAT2 dimers boosts antiviral defenses. *Cell Res.* **31**, 206-218
23. Vinkemeier, U. (2004) Getting the message across, STAT! Design principles of a molecular signaling circuit. *J. Cell. Biol.* **167**, 197-201
24. Popken, P., Ghavami, A., Onck, P. R., Poolman, B., and Veenhoff, L. M. (2015) Size-dependent leak of soluble and membrane proteins through the yeast nuclear pore complex. *Mol. Biol. Cell* **26**, 1386-1394

25. Meyer, T., Gavenis, K., and Vinkemeier, U. (2002) Cell type-specific and tyrosine phosphorylation-independent nuclear presence of STAT1 and STAT3. *Exp. Cell Res.* **272**, 45-55
26. Frahm, T., Hauser, H., and Köster, M. (2006) IFN-type-I-mediated signaling is regulated by modulation of STAT2 nuclear export. *J. Cell Sci.* **119**, 1092-1104
27. Banninger, G., and Reich, N. C. (2004) STAT2 nuclear trafficking. *J. Biol. Chem.* **279**, 39199-39206
28. Fu, S. C., Fung, H. Y. J., Cağatay, T., Baumhardt, J., and Chook, Y. M. (2018) Correlation of CRM1-NES affinity with nuclear export activity. *Mol. Biol. Cell* **29**, 2037-2044
29. Hodel, A. E., Harreman, M. T., Pulliam, K. F., Harben, M. E., Holmes, J. S., Hodel, M. R., *et al.* (2006) Nuclear localization signal receptor affinity correlates with in vivo localization in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **281**, 23545-23556
30. Braunstein, J., Brutsaert, S., Olson, R., and Schindler, C. (2003) STATs dimerize in the absence of phosphorylation. *J. Biol. Chem.* **278**, 34133-34140
31. Domszalai, T., Martincuks, A., Fahrenkamp, D., Schmitz-Van de Leur, H., Küster, A., and Müller-Newen, G. (2014) Consequences of the disease-related L78R mutation for dimerization and activity of STAT3. *J. Cell Sci.* **127**, 1899-1910
32. Meyer, T., Hendry, L., Begitt, A., John, S., and Vinkemeier, U. (2004) A single residue modulates tyrosine dephosphorylation, oligomerization, and nuclear accumulation of Stat transcription factors. *J. Biol. Chem.* **279**, 18998-19007
33. Lim, C. P., and Cao, X. (2006) Structure, function, and regulation of STAT proteins. *Mol. Biosyst.* **2**, 536-550
34. Sgrignani, J., Olsson, S., Ekonomiuk, D., Genini, D., Krause, R., Catapano, C. V., *et al.* (2015) Molecular determinants for unphosphorylated STAT3 dimerization determined by integrative modeling. *Biochemistry* **54**, 5489-5501
35. Knauer, S. K., Moodt, S., Berg, T., Liebel, U., Pepperkok, R., and Stauber, R. H. (2005) Translocation biosensors to study signal-specific nucleo-cytoplasmic transport, protease activity and protein-protein interactions. *Traffic* **6**, 594-606
36. Dixon, A. S., and Lim, C. S. (2010) The nuclear translocation assay for intracellular protein-protein interactions and its application to the Bcr coiled-coil domain. *Biotechniques* **49**, 519-24
37. Miyagi, T., Gil, M. P., Wang, X., Louten, J., Chu, W. M., and Biron, C. A. (2007) High basal STAT4 balanced by STAT1 induction to control type 1 interferon effects in natural killer cells. *J. Exp. Med.* **204**, 2383-2396
38. Cheon, H., Yang, J., and Stark, G. R. (2011) The functions of signal transducers and activators of transcriptions 1 and 3 as cytokine-inducible proteins. *J. Interferon Cytokine Res.* **31**, 33-40
39. Gough, D. J., Messina, N. L., Hii, L., Gould, J. A., Sabapathy, K., Robertson, A. P., *et al.* (2010) Functional crosstalk between type I and II interferon through the regulated expression of STAT1. *PLoS Biol.* **8**, e1000361
40. Lehtonen, A., Matikainen, S., and Julkunen, I. (1997) Interferons up-regulate STAT1, STAT2, and IRF family transcription factor gene expression in human peripheral blood mononuclear cells and macrophages. *J. Immunol.* **159**, 794-803
41. Michalska, A., Blaszczyk, K., Wesoly, J., and Bluysen, H. A. R. (2018) A positive feedback amplifier circuit that regulates interferon (IFN)-stimulated gene expression and controls type I and type II IFN responses. *Front Immunol.* **9**, 1135
42. Aue, A., Szelinski, F., Weißenberg, S. Y., Wiedemann, A., Rose, T., Lino, A. C., *et al.* (2020), Elevated STAT1 expression but not phosphorylation in lupus B cells correlates with disease activity and increased plasmablast susceptibility. *Rheumatology* **59**, 3435-3442

43. Cheon, H., Holvey-Bates, E. G., Schoggins, J. W., Forster, S., Hertzog, P., Imanaka, N., *et al.* (2012) IFN β -dependent increases in STAT1, STAT2, and IRF9 mediate resistance to viruses and DNA damage. *EMBO J.* **32**, 2751-2763
44. Gil, M. P., Ploquin, M. J., Watford, W. T., Lee, S. H., Kim, K., Wang, X., *et al.* (2012) Regulating type 1 IFN effects in CD8 T cells during viral infections: changing STAT4 and STAT1 expression for function. *Blood* **120**, 3718-3728
45. Gil, M. P., Salomon, R., Louten, J., and Biron, C. A. (2006) Modulation of STAT1 protein levels: a mechanism shaping CD8 T-cell responses in vivo. *Blood* **107**, 987-993
46. Lesinski, G. B., Valentino, D., Hade, E. M., Jones, S., Magro, C., Chaudhury, A. R., *et al.* (2005) Expression of STAT1 and STAT2 in malignant melanoma does not correlate with response to interferon-alpha adjuvant therapy. *Cancer Immunol. Immunother.* **54**, 815-825
47. Nan, J., Wang, Y., Yang, J., and Stark, G. R. (2018) IRF9 and unphosphorylated STAT2 cooperate with NF- κ B to drive IL6 expression. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3906-3911
48. Zhou, M., Zhang, P., Da, M., Yang, R., Ma, Y., Zhao, J., *et al.* (2022) Pan-cancer analysis of the expression of STAT family genes in tumors and their relationship to the tumor microenvironment. *Front. Oncol.* **12**, 92553
49. Lackmann, M., Harpur, A. G., Oates, A. C., Mann, R. J., Gabriel, A., Meutermans, W., *et al.* (1998) Biomolecular interaction analysis of IFN gamma-induced signaling events in whole-cell lysates: prevalence of latent STAT1 in high-molecular weight complexes. *Growth Factors* **16**, 39-51
50. Luker, K. E., Smith, M. C., Luker, G. D., Gammon, S. T., Piwnicka-Worms, H., and Piwnicka-Worms, D. (2004) Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proc. Natl. Acad. Sci. U. S. A.* **101**: 12288-12293
51. Kretzschmar, A. K., Dinger, M. C., Henze, C., Brocke-Heidrich, K., and Horn, F. (2004) Analysis of Stat3 (signal transducer and activator of transcription 3) dimerization by fluorescence resonance energy transfer in living cells. *Biochem. J.* **377**, 289-297
52. Li, L., and Shaw, P. E. (2004) A STAT3 dimer formed by inter-chain disulphide bridging during oxidative stress. *Biochem. Biophys. Res. Commun.* **322**, 1005-1011
53. Ren, Z., Mao, X., Mertens, C., Krishnaraj, R., Qin, J., Mandal, P. K., *et al.* (2008) Crystal structure of unphosphorylated STAT3 core fragment. *Biochem. Biophys. Res. Commun.* **374**, 1-5
54. Schröder, M., Kroeger, K. M., Volk, H. D., Eidne, K. A., and Grütz, G. (2004) Preassociation of nonactivated STAT3 molecules demonstrated in living cells using bioluminescence resonance energy transfer: a new model of STAT activation? *J. Leukoc. Biol.* **75**, 792-797
55. Vogt, M., Domszalai, T., Kleshchanok, D., Lehmann, S., Schmitt, A., Poli, V., *et al.* (2011) The role of the N-terminal domain in dimerization and nucleocytoplasmic shuttling of latent STAT3. *J. Cell Sci.* **124**, 900-909
56. Novak, U, Ji, H., Kanagasundaram, V., Simpson, R., and Paradiso, L. (1998) STAT3 forms stable homodimers in the presence of divalent cations prior to activation. *Biochem. Biophys. Res. Commun.* **247**, 558-563
57. Bernadó, P., Pérez, Y., Blobel, J., Fernández-Recio, J., Svergun, D. I., and Pons, M. (2009) Structural characterization of unphosphorylated STAT5a oligomerization equilibrium in solution by small-angle X-ray scattering. *Protein Sci.* **18**, 716-726
58. Platanitis, E., Demiroz, D., Schneller, A., Fischer, K., Capelle, C., Hartl, M., *et al.* (2019) A molecular switch from STAT2-IRF9 to ISGF3 underlies interferon-induced gene transcription. *Nat Commun.* **10**, 2921

59. Stancato, L. F., David, M., Carter-Su, C., Larner, A. C., and Pratt, W. B. (1996) Preassociation of STAT1 with STAT2 and STAT3 in separate signalling complexes prior to cytokine stimulation. *J. Biol. Chem.* **271**, 4134-4137
60. Wedeking, T., Löchte, S., Richter, C. P., Bhagawati, M., Piehler, J., and You, C. (2015) Single cell GFP-trap reveals stoichiometry and dynamics of cytosolic protein complexes. *Nano Lett.* **15**, 3610–3615
61. Blumert, C., Kalkhof, S., Brocke-Heidrich, K., Kohajda, T., von Bergen, M., and Horn, F. (2013) Analysis of the STAT3 interactome using in-situ biotinylation and SILAC. *J. Proteomics* **94**, 370-386
62. Haan, S., Kortylewski, M., Behrmann, I., Müller-Esterl, W., Heinrich, P. C., and Schaper, F. (2000) Cytoplasmic STAT proteins associate prior to activation. *Biochem. J.* **345**, 417-421
63. Nivarthi, H., Gordziel, C., Themanns, M., Kramer, N., Eberl, M., Rabe, B., *et al.* (2016) The ratio of STAT1 to STAT3 expression is a determinant of colorectal cancer growth. *Oncotarget* **7**, 51096-51106. Erratum in: *Oncotarget* **9**, 33865
64. Yao, K., Chen, Q., Wu, Y., Liu, F., Chen, X., and Zhang, Y. (2017) Unphosphorylated STAT1 represses apoptosis in macrophages during Mycobacterium tuberculosis infection. *J. Cell Sci.* **130**, 1740-1751
65. Okada, S., Asano, T., Moriya, K., Boisson-Dupuis, S., Kobayashi, M., Casanova, J. L., *et al.* (2020) Human STAT1 gain-of-function heterozygous mutations: Chronic mucocutaneous candidiasis and type I interferonopathy. *J. Clin. Immunol.* **40**, 1065-1081
66. Vogel, T, P., Milner, J. D., and Cooper, M. A. (2015) The ying and yang of STAT3 in human disease. *J. Clin. Immunol.* **35**, 615-623
67. Menon, P. R., Doudin, A., Gregus, A., Wirths, O., Staab, J., and Meyer, T. (2021) The anti-parallel dimer binding interface in STAT3 transcription factor is required for the inactivation of cytokine-mediated signal transduction. *Biochim. Biophys. Acta Mol. Cell Res.* **1868**,119118
68. Mertens, C., Zhong, M., Krishnaraj, R., Zou, W., Chen, X., and Darnell, J. E. Jr. (2006) Dephosphorylation of phosphotyrosine on STAT1 dimers requires extensive spatial reorientation of the monomers facilitated by the N-terminal domain. *Genes Dev.* **20**, 3372-3381
69. Lin, J. X., Li, P., Liu, D., Jin, H. T., He, J., Ata Ur Rasheed, M., *et al.* (2012) Critical role of STAT5 transcription factor tetramerization for cytokine responses and normal immune function. *Immunity* **36**, 586-599
70. Begitt, A., Droescher, M., Meyer, T., Schmid, C. D., Baker, M., Antunes, F., *et al.* (2014). STAT1-cooperative DNA binding distinguishes type 1 from type 2 interferon signaling. *Nat. Immunol.* **15**, 168-176. Erratum in: *Nat. Immunol.* **15**, 1090 (2014)
71. Cranfill, P. J., Sell, B. R., Baird, M. A., Allen, J. R., Lavagnino, Z., de Gruiter, M., *et al.* (2016) Quantitative assessment of fluorescent proteins. *Nat. Methods* **13**, 557–562
72. Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001) Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Eng.* **14**, 529-532
73. Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., *et al.* (2017) ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529
74. French, A. P., Mills, S., Swarup, R., Bennett, M. J., and Pridmore, T. P. (2008) Colocalization of fluorescent markers in confocal microscope images of plant cells. *Nat. Protoc.* **3**, 619-628

Legends

Figure 1. Outline and validation of co-localization as an assay to probe dimerization of latent STATs.

- A.** Schematics describing expected subcellular distributions of (top) interacting and (bottom) non-interacting bait (green fluorescence) and test proteins (red fluorescence). Shown are the anticipated outcomes obtained in green and red channels together with associated Pearson correlation coefficients, r_P , which quantify the degree of co-localization of the fluorophores, obtained with the co-localization assay. The bait protein is fused to a nuclear export signal (NES), directing it to the cytoplasm, whereas the test protein is without heterologous translocation signal. Arrow orientation and width signify the direction and relative efficiency of nucleocytoplasmic translocation of the different protein species.
- B.** Fluorescence micrographs of representative HeLa cells co-expressing bait and test protein combinations for STAT3 (top) and STAT1 (bottom). Baits were directed to the cytoplasm or nucleus through appended NES or NLS signal sequences as indicated in the panels. Homotypic STAT protein interactions were probed by co-expressing bait and wild-type (WT) STATs; mutant STAT1 (F77A) and STAT3 (L78R) were used to examine the consequences of known dimer-dissociating mutations on the co-localization of bait and test STAT proteins. Shown are green channel, red channel, and merged channels, which includes visualisation of nuclei using Hoechst dye (blue).

Figure 2. Latent STATs assemble five homodimers and two heterodimers via N-domain interactions.

- A, B.** Fluorescence micrographs of representative HeLa cells co-expressing the indicated homotypic (A) and heterotypic (B) bait and test STAT protein pairings. Shown are the green channel, red channel, and merged channels, which includes visualisation of nuclei using Hoechst dye (blue). The rightmost columns show the distribution of test proteins in cells that do not co-express the bait (w/o bait).
- C.** Bar diagram depicting corresponding Pearson correlation coefficients (r_P) for the experiments shown in (A) and (B) with the individual data points (black dots) superimposed. Co-localization was determined in eligible cells, i.e. cells that expressed bait and test proteins from equimolar amounts up to a 4-fold excess of bait protein. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$ using a Kruskal–Wallis test; ns, not significant. R_p numerical values \pm SD are given above the bars; the number of cells analysed in each experiment are shown in brackets below bars. Error bars, SD.

Figure 3. Homo- and heterotypic dimerization of the unphosphorylated STAT proteins.

Summary of Pearson correlation coefficients obtained with the translocation assay for the 28 possible pairings of unphosphorylated STAT proteins. Data are obtained with NES fusion proteins and wild-type STAT2 as the baits. STAT2 homodimer data are for U-STAT2:STAT2 Δ C (+) and U-STAT2 Δ C-NLS:STAT2 Δ C (++) . Given are means \pm standard deviation. Light and dark green highlighting marks stable homo- and heterodimers, respectively. Number of cells analysed in each experiment are shown in brackets. Source data are included as Supporting Information.

Figure 4. Dimer assembly of latent STATs requires N-domain interactions.

- A.** Fluorescence micrographs of representative HeLa cells co-expressing bait and test STAT proteins. Bait proteins were directed to the nucleus or cytoplasm by appending heterologous NLS or NES signals (endogenous NES activity was used for STAT2), and the effect of single, potentially dimer-disrupting N-domain residues was probed on the co-localization of bait and test proteins. Presented are the green channel, red channel, and merged channels, which includes visualisation of nuclei using Hoechst dye (blue).
- B.** Bar diagram depicting corresponding Pearson correlation coefficients (r_p) for the experiments shown in (A) with the individual data points (black dots) superimposed. Co-localization was determined as described in Figure 2C. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$ using a Kruskal–Wallis test. R_p numerical values \pm SD are given above the bars; the number of cells analysed in each experiment are shown in brackets below bars. White stars signify homo- or heterodimers containing the indicated N-domain mutations. Error bars, SD.

Figure 5. U-STAT1:STAT2 heterodimers are exceptionally stable compared to the other latent STAT dimers.

- A.** Schematics describing a co-localization assay for probing the interaction strength of dimer-forming STATs. Antagonistic heterologous translocation activities (NES or NLS) are appended to bait (green fluorescence) and test (red fluorescence) proteins, respectively, whereby NES activity is dominant. Arrow orientation and width signify the direction and relative efficiency of signal-mediated protein translocation.
- B.** Fluorescence micrographs of representative HeLa cells co-expressing bait and test STAT proteins. Shown are the consequences for the co-localization of bait and test STAT proteins if they each were fused to one of the two opposed translocation signals (NES or NLS) as indicated. Presented are the green channel, red channel, and merged channels, which includes the Hoechst-stained cell nuclei (blue).
- C.** Same as (B). Where indicated (+LMB), cells were treated for 4 h with leptomycin B before imaging to inhibit NES-mediated nuclear export.
- D.** Bar diagram depicting corresponding Pearson correlation coefficients (r_p) for experiments shown in (B) and (C) with the individual data points (black dots) superimposed. Co-localization was determined as described in Figure 2C. **** $p < 0.0001$ using a Kruskal–Wallis test, ns, not significant. White star signifies heterodimer containing the indicated N-domain mutation. R_p numerical values \pm SD are given above the bars; the number of cells analysed in each experiment are shown in brackets below bars. Error bars, SD.

Supporting Figure 1. Characterization of STAT variant proteins used.

- A.** Representative micrographs of HeLa cells expressing the indicated STAT variant proteins. The intracellular distributions of the following STAT-fluorophore fusion proteins are shown. Wild-type STAT1-6; N-domain mutated STATs (F77A, L78R/S, L82A); full-length STATs with heterologous signal sequence for either nuclear

export (NES) or nuclear import (NLS) appended; full-length STAT1 with both the export and import signal sequences appended (NES-NLS); truncated STAT2 devoid of the C-terminus (Δ C); C-terminally truncated STAT2 with heterologous signal sequence for either nuclear export (NES) or nuclear import (NLS) appended. Presented are the green or red channels, and the Hoechst-stained cell nuclei (blue). Where there was treatment with leptomycin B (+LMB) it was for four hours. The nuclei of interest are labelled with white asterisks to facilitate the matching of the GFP/Cherry signals and the Hoechst staining.

- B.** Western blotting results using whole cell lysates (30 μ l/lane) from HEK293T cells expressing mCherry fusion proteins of the full-length STATs or C-terminally truncated STAT2, as indicated. After blotting, total protein was stained with Ponceau S (bottom panel), followed by incubation with anti-mCherry antibody and immunodetection (top panel). In brief, experimental details were as follows. Twenty hours after transient transfection with vectors encoding STAT-mCherry fusion proteins as indicated, 4×10^6 HEK293T cells (transfection efficiency $\sim 70\%$) were lysed by boiling in 0.5 ml 2 x SDS sample buffer (42 mM Tris-HCl, pH 6.8, 1.3% SDS (w/v), 233 mM DTT, 8.3% glycerol (v/v), 0.04 % Bromphenol blue). Proteins (30 μ l/lane) were resolved by 7% SDS-PAGE followed by semi-dry protein transfer to a nitrocellulose membrane. Total protein loading was visualized by staining with 0.1% (w/v) Ponceau S in 1% (v/v) acetic acid for 1 min. Excess stain was removed with water (1 min), followed by blocking in 4% (w/v) dried skimmed milk powder in Tris-buffered saline containing 0.05% (v/v) Tween 20 (Sigma) and incubation with rabbit polyclonal anti-mCherry antibody (GTX128509, GeneTex). IRDye 800CW-coupled goat anti-rabbit secondary antibody (#926-32211) was from Li-Cor Bioscience. Blots were imaged by infrared fluorescence detection (Odyssey, Li-Cor Bioscience).

Supporting Figure 2. U-STAT1:STAT2 heterodimers do not require the C-terminal transactivation domain of STAT2.

Representative micrographs of HeLa cells expressing the indicated bait and test protein pairings of variants of STAT1 and C-terminally truncated STAT2. (S1) wild-type STAT1; (S1-NES) full-length STAT1 with heterologous nuclear export signal appended; (S2- Δ C) truncated STAT2 devoid of the C-terminus and its intrinsic nuclear export activity; (S2- Δ C-NLS) C-terminally truncated STAT2 with heterologous nuclear import signal appended. Presented are the green channel, red channel, and merged channels, which includes the Hoechst-stained cell nuclei (blue).

Supporting Figure 3. Assessing the distribution of endogenous STAT1 and the level of STAT1 overexpression in transiently transfected HeLa cells used in this work.

- A.** HeLa cells were transfected with STAT1-mEGFP expression plasmid as described under Cell Culture and Transfections in the main text using 1.2 μ g DNA/well of a 24-well plate. After 20 hours, cells were fixed as described under Fluorescence Imaging in the main text, followed by blocking for 1 h in 20% (v/v) FBS/PBS prior to 15 h incubation at 4°C with primary anti-STAT1 monoclonal antibody (Becton-Dickinson 610186), washing in FBS/PBS and incubation for another hour with Cy3-coupled goat anti-mouse IgG secondary antibody (Jackson ImmunoResearch 115-165-062). Shown are four composite images of the same field of view. Top left,

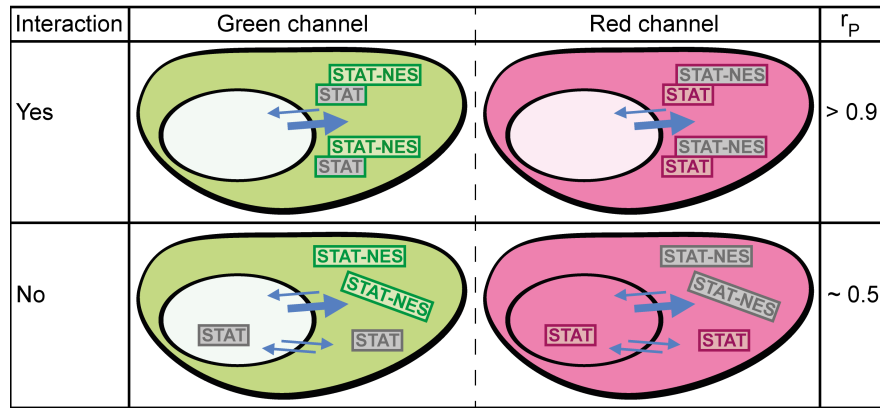
composite image of green channel (showing expression of transfected STAT1-mEGFP fusion protein) and blue channel (showing Hoechst-stained cell nuclei) with manually segmented transfected cells, based on EGFP fluorescence signals, numbered 1-9. Top right, composite of red and blue channels representing STAT1 overexpression in transfected cells. Bottom left, same as top left, but with segmentation of untransfected cells 10-18. Cell segmentations are based on green autofluorescence. Bottom right, composite of red and blue channels representing endogenous STAT1 in untransfected cells. Average red fluorescence intensities (A.U., background-subtracted) for cells 1-18 are listed on the right, together with average background fluorescence in cell-free areas recorded in the squares labelled A-C.

- B.** Violin plots summarizing the distributions of fluorescence data for untransfected and transiently transfected HeLa cells, representing endogenous and overexpressed STAT1 levels. Plots include the data collected from the micrograph shown in (A) and four more. The centre lines of the plots represent the group median, the top and bottom lines represent the 75th and 25th percentiles. Fluorescence intensities are background-subtracted. The number of cells analysed are shown in brackets.
- C.** HeLa cells expressing mEGFP fusion proteins of wild-type STAT2 (S2) or NLS-tagged C-terminally truncated STAT2 (S2- Δ C-NLS) are shown after labelling with anti-STAT1 antibody to detect the distribution of endogenous STAT1 as described above. Note co-localization of the native cellular STAT1 with the transfected bait proteins. STAT1 immunostaining was done as described in (A) above.

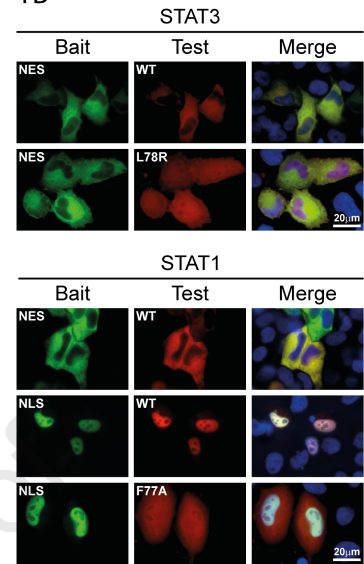
Supporting Figure 4. Calibration of image acquisition settings using a mCherry-mEGFP fusion protein.

- A.** Quantitative imaging results of HeLa cells transiently expressing a fusion protein consisting of mEGFP and mCherry connected by a rigid α -helical linker. At each of the two [red/green] exposure time ratios of 1 and 0.5 used, three different cells each were imaged repeatedly using the exposure times listed in the Figure. Given are the readings for individual cells and the resulting regression curves. Note linear relationships of red and green fluorescence emissions for different exposure time lengths and ratios. Fluorescence emission intensities given as arbitrary light units (A.U.).
- B.** Same as (A), but 32 different cells were imaged using the same exposure times of 30 ms and 68 ms for red and green channels, respectively. This [red/green] exposure time ratio of 0.44 resulted in nearly identical mEGFP and mCherry fluorescence emissions over a wide range of signal intensities. It was therefore used in all experiments to quantify the relative expression levels of co-expressed STAT proteins fused to these fluorophores. In the quantitative imaging experiments shown in Figures 2-5, exposure times between 15 ms and 300 ms were used for red fluorescence and respective times for green fluorescence. The inset shows a representative fluorescence micrograph of HeLa cells expressing the mEGFP-mCherry fusion protein.

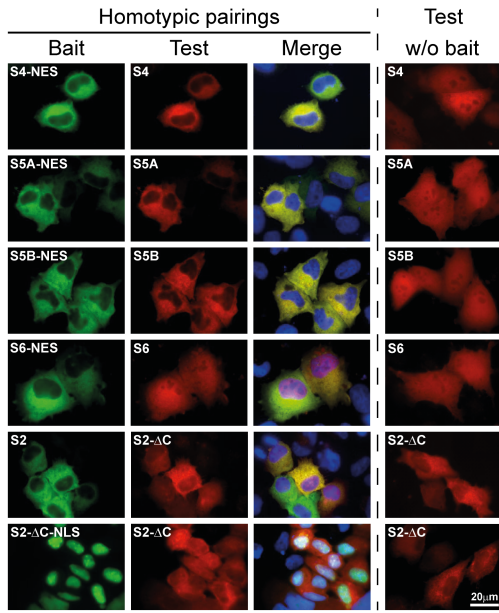
1A



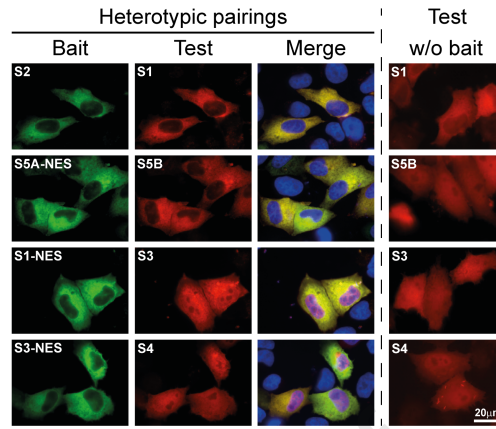
1B



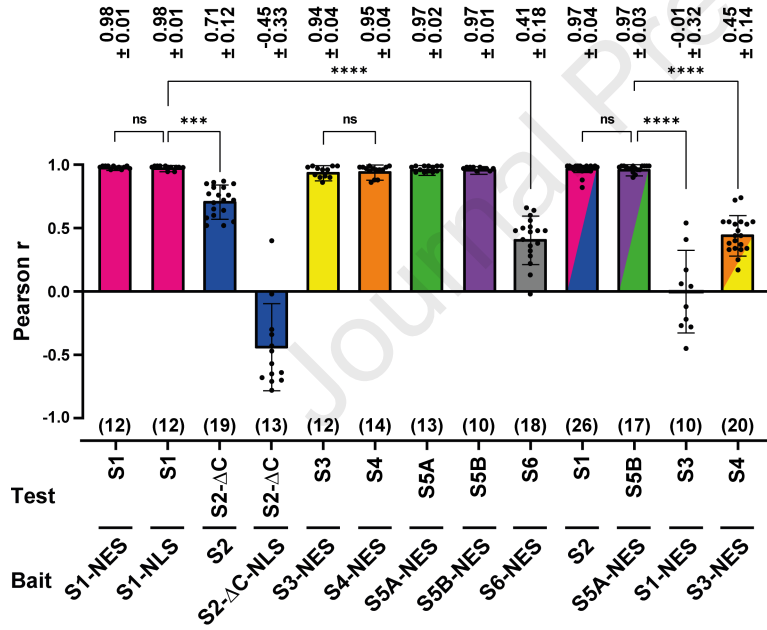
2A



2B



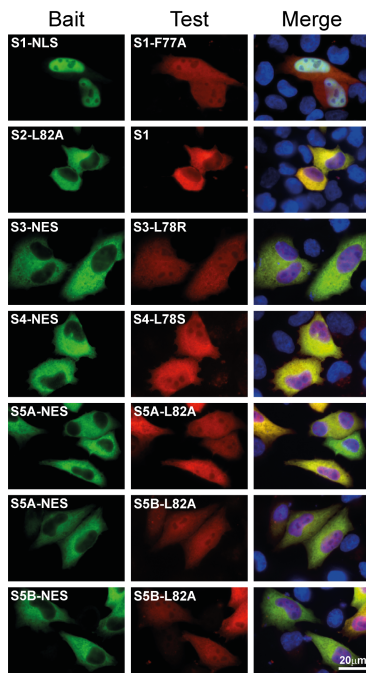
2C



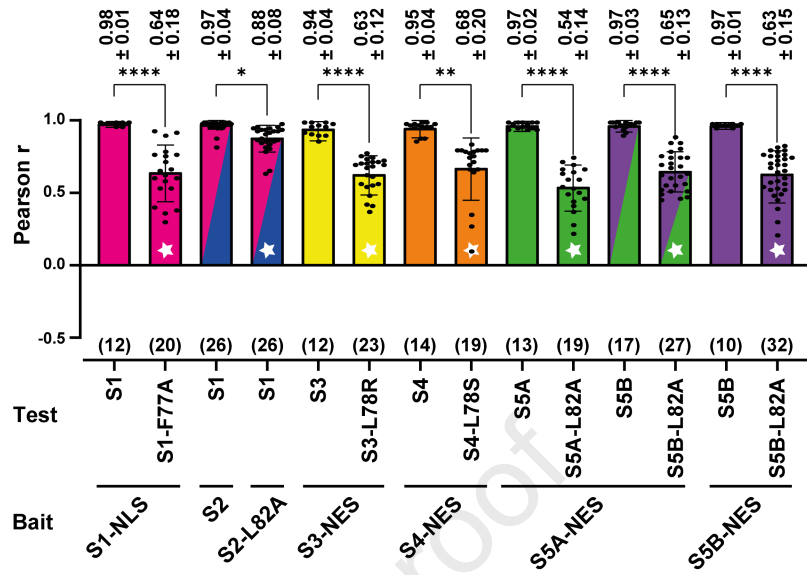
3

	STAT1	STAT2	STAT3	STAT4	STAT5A	STAT5B	STAT6
STAT1	0.98 ± 0.01 (12)	0.97 ± 0.04 (26)	-0.01 ± 0.32 (10)	0.42 ± 0.10 (12)	0.29 ± 0.19 (12)	0.24 ± 0.20 (12)	0.31 ± 0.22 (12)
STAT2		0.71 ± 0.12 [†] (19) -0.45 ± 0.3 ^{††} (13)	0.58 ± 0.14 (15)	0.26 ± 0.27 (19)	0.05 ± 0.41 (14)	0.54 ± 0.12 (12)	0.40 ± 0.19 (20)
STAT3			0.94 ± 0.04 (12)	0.45 ± 0.14 (20)	0.30 ± 0.27 (21)	0.34 ± 0.33 (15)	0.43 ± 0.17 (23)
STAT4				0.95 ± 0.04 (14)	0.37 ± 0.26 (17)	0.54 ± 0.10 (14)	0.30 ± 0.16 (16)
STAT5A					0.97 ± 0.02 (13)	0.97 ± 0.03 (17)	0.42 ± 0.18 (12)
STAT5B						0.97 ± 0.01 (10)	0.39 ± 0.23 (14)
STAT6							0.41 ± 0.18 (18)

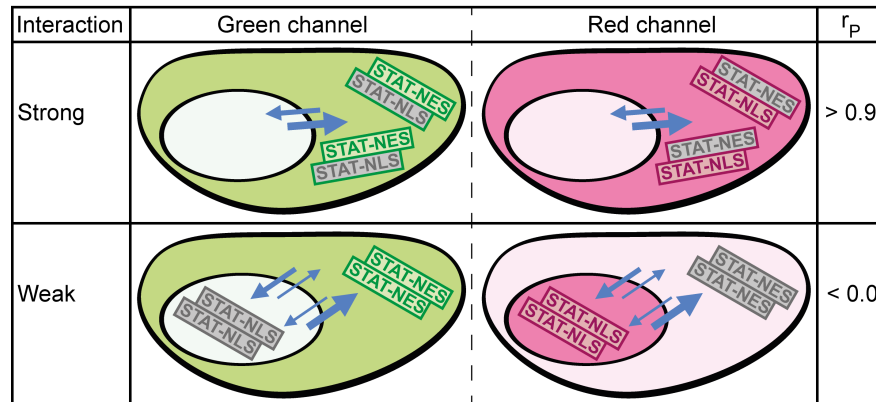
4A



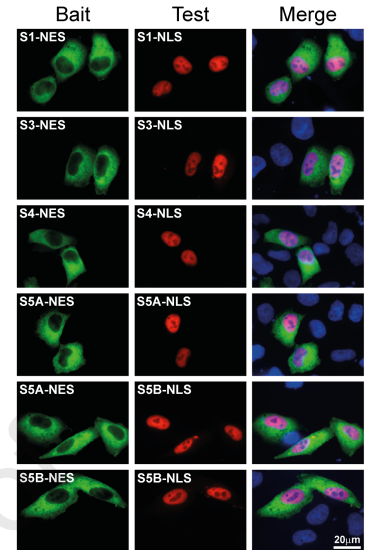
4B



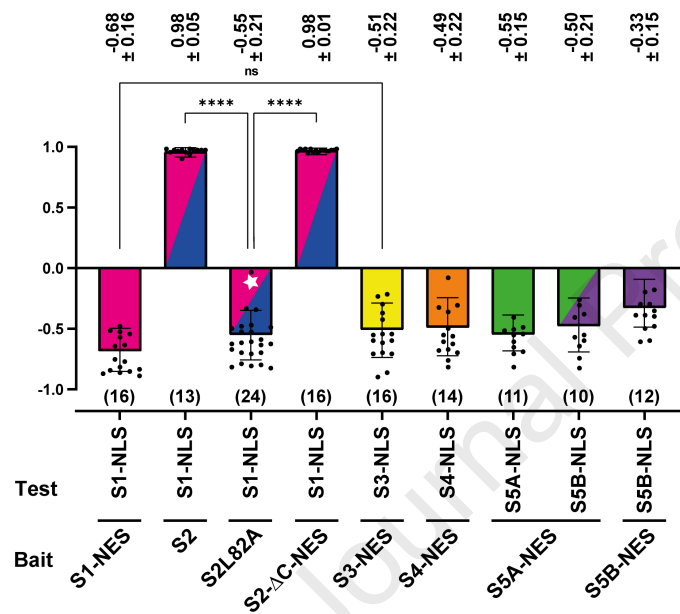
5A



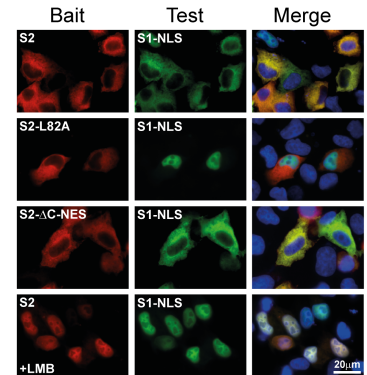
5B



5D



5C



All authors state that there are no interests to declare.

Nottingham, December 2022.

Journal Pre-proof

U. V., J. R. C., D. E. V. conceptualization; A. B., S. K. investigation; U. V. writing—original draft; U. V., A. B., J. R. C., D. E. V. writing—review & editing; A. B., U. V., S. K. visualization; U. V. supervision; U. V. funding acquisition.

Journal Pre-proof