Comparative Interactome Analysis of α-arrestin Families in Human and *Drosophila*

Kyung-Tae Lee[1,7,&], Inez K.A. Pranoto[5,&], Soon-Young Kim[6], Hee-Joo Choi[2,3,4], Ngoc Bao To[1], Hansong Chae[1], Jeong-Yeon Lee[2,3], Jung-Eun Kim[6], Young V.Kwon[5,#], and Jin-Wu Nam[1,2,7,#]

[1]Department of Life Science, College of Natural Sciences, Hanyang University, Wangsimni-ro 222, Seongdong-gu, Seoul 04763, Republic of Korea.

[2]Bio-BigData Center, Hanyang Institute for Bioscience and Biotechnology, Hanyang University, Wangsimni-ro 222, Seongdong-gu, Seoul 04763, Republic of Korea.

[3]Department of Pathology, College of Medicine, Hanyang University, Wangsimni-ro 222, Seongdong-gu, Seoul, 04763, Republic of Korea.

[4]Hanyang Biomedical Research Institute, Hanyang University, Wangsimni-ro 222, Seongdong-gu, Seoul 04763, Republic of Korea.

[5]Department of Biochemistry, University of Washington, 1959 NE Pacific St., Seattle, WA 98195, USA.

[6]Department of Molecular Medicine, Cell and Matrix Research Institute, School of Medicine, Kyungpook National University, 680 Gukchaebosang-ro, Jung-gu, Daegu 41944, Republic of Korea.

[7]Hanyang Institute of Advanced BioConvergence, Hanyang University, Wangsimni-ro 222, Seongdong-gu, Seoul 04763, Republic of Korea

#Corresponding authors: jwnam@hanyang.ac.kr, ykwon7@uw.edu

[&] These authors contributed equally to this work.

Keywords: α-arrestin, AP/MS, Human, *Drosophila*, Comparative interactomes, ATAC-seq, RNA-seq

Abstract

The α-arrestins form a large family of evolutionally conserved modulators that control diverse signaling pathways, including both G-protein-coupled receptor- (GPCR-) mediated and non-GPCR mediated pathways, across eukaryotes. However, unlike β-arrestins, only a few α-arrestin targets and functions have been characterized. Here, using affinity purification and mass spectrometry, we constructed interactomes for six human and twelve *Drosophila* α-arrestins. The resulting high-confidence interactomes comprised 307 and 467 prey proteins in human and *Drosophila*, respectively. A comparative analysis of these interactomes predicted not only conserved binding partners, such as motor proteins, proteases, ubiquitin ligases, RNA splicing factors, and GTPase-activating proteins, but also those specific to mammals, such as histone modifiers and the subunits of V-type ATPase. Given the manifestation of the interaction between the human α-arrestin, TXNIP, and the histone-modifying enzymes, including HDAC2, we undertook a global analysis of transcription signals and chromatin structures that were affected by TXNIP knockdown. We found that TXNIP activated targets by blocking HDAC2 recruitment to targets, a result that was validated by chromatin immunoprecipitation assays. Additionally, the interactome for an uncharacterized human α-arrestin ARRDC5 uncovered multiple components in the V-type ATPase, which plays a key role in bone resorption by osteoclasts. Our study presents conserved and species-specific protein-protein interaction maps for α-arrestins, which provide a valuable resource for interrogating their cellular functions for both basic and clinical research.

Introduction

The discovery of first arrestin protein in retinal rods contributed to a deeper understanding of photoreceptor signaling mediated by rhodopsin, which is one of the G-protein-coupled receptor (GPCR) class, and after its ability to arrest the GPCR signaling pathway, the protein was first named as “arrestin” (Kuhn, Hall, & Wilden, 1984; Wilden, Wust, Weyand, & Kuhn, 1986; Zuckerman & Cheasty, 1986). Shortly after this discovery of the first arrestin protein in the retina, another arrestin protein that specifically turns off β-adrenergic signaling, another type of GPCR, through “receptor desensitization” was identified and named “β-arrestin” (Benovic, DeBlasi, Stone, Caron, & Lefkowitz, 1989; Lohse, 1992; Shenoy & Lefkowitz, 2011). Further studies have revealed that β-arrestins regulate the receptor desensitization of other signaling pathways through ubiquitination and regulation of trafficking of various cargo molecules (Y. M. Kim & Benovic, 2002; Malik & Marchese, 2010; Puca & Brou, 2014).

Another class of arrestin, α-arrestin, was first studied in fungi and yeast (Andoh, Hirata, & Kikuchi, 2002) and subsequently recognized as new class of arrestins (Boase & Kelly, 2004; Herranz et al., 2005). They contain characteristic arrestin domains, arrestin_N and arrestin_C, and PPXY motifs, which are unique to the α-arrestin clan (Puca & Brou, 2014). A phylogenetic study of arrestin proteins showed that α-arrestins are the ancestral class of the arrestin family and conserved from yeast to human (Alvarez, 2008). To date, six α-arrestins, arrestin domain containing protein 1 (ARRDC1), ARRDC2, ARRDC3, ARRDC4, ARRDC5, and thioredoxin-interacting protein (TXNIP), have been found to be in the human genome (Zbieralski & Wawrzycka, 2022). These human α-arrestins were first studied in conjunction with β-arrestins in the regulation of the β2-adrenergic receptor (β2AR) in human cells. ARRDC3 and ARRDC4 works as an adaptor protein for the ubiquitination of β2AR by recruiting the NEDD4 protein, an E3 ubiquitin ligase, through its conserved PPXY motifs (S. O. Han, Kommaddi, & Shenoy, 2013; Nabhan, Pan, & Lu, 2010; Shea, Rowell, Li, Chang, & Alvarez, 2012). In addition to their

β2AR-associated roles, α-arrestins are involved in trafficking and sorting of other GPCRs and signaling molecules through post-translational modifications, including ubiquitination. For example, ARRDC1 and ARRDC3 were reported to play roles in the degradation of the Notch receptor (Puca, Chastagner, Meas-Yedid, Israel, & Brou, 2013) and in the ubiquitination of ALG-2-interacting protein X (ALIX) (Dores, Lin, N, Mendez, & Trejo, 2015). Uniquely, ARRDC1 have been reported to mediate microvesicle budding by recruiting E3 ligases, such as WW domain-containing E3 ubiquitin protein ligase2 (WWP2). This recruitment leads to its own ubiquitination. Additionally, ARRDC1 possesses a PSPA motif that binds the tumor susceptibility gene 101 (TSG101) protein, an essential component of an endosomal sorting complex that is also required for this ARRDC1 mediated microvesicle budding (Nabhan, Hu, Oh, Cohen, & Lu, 2012). Another well-known α-arrestin, TXNIP, was first named as vitamin D3-upregulated protein 1 (VDUP1) after verification that its gene is a vitamin D3 target in cancer cells (K. S. Chen & DeLuca, 1994; Qayyum, Haseeb, Kim, & Choi, 2021). Since then, TXNIP had been reported to directly interact with thioredoxin, which is an essential component of the cellular redox system, to inhibit its activity as an antioxidant (Junn et al., 2000; Nishiyama et al., 1999; Patwari, Higgins, Chutkow, Yoshioka, & Lee, 2006). TXNIP was also reported to inhibit glucose uptake by inducing the internalization of glucose transporter 1 (GLUT1) through clathrin-mediated endocytosis and by indirectly reducing GLUT1 RNA levels (Wu et al., 2013). Although the TXNIP is known to be localized in both cytoplasm and nucleus, biological functions of TXNIP have been mostly explored in cytoplasm but remained poorly characterized in nucleus.

A few α-arrestins appear to have evolutionarily conserved functions in both human and invertebrates. For instance, the Hippo signaling pathway, which impacts a variety of cellular processes such as metabolism, development, and tumor progression (Mo, Park, & Guan, 2014; Pei et al., 2015; Schutte et al., 2014; Y. Wang et al., 2010; Zhi, Zhao, Zhou, Liu, & Chen, 2012), was shown to be regulated by α-arrestin in both *Drosophila* (Y. Kwon et al., 2013) and human cells (Xiao et al., 2018). In *Drosophila*, the protein Leash was identified as

an α-arrestin and shown to down-regulate Yki by promoting its lysosomal degradation, leading to a restriction in growth (Y. Kwon et al., 2013). In human cells, ARRDC1 and ARRDC3 were shown to induce degradation of the mammalian homolog of Yki, YAP1, by recruiting the E3 ubiquitin ligase ITCH in renal cell carcinoma (Xiao et al., 2018), suggesting functional homology between human and *Drosophila*. However, because the α-arrestins interact with multiple targets, an unbiased, comparative analysis of interactome is required to determine whether other α-arrestin from human and *Drosophila* have common and specific interacting partners, which will determine their functional homology and diversification.

A comprehensive understanding of their protein-protein interactions (PPIs) and interactomes will shed light on the underlying molecular mechanisms, reveal novel regulatory axes, and enable the identification of previously unrecognized roles of α-arrestin in cellular processes. Furthermore, extensive characterization of the α-arrestin interactomes may help uncover potential therapeutic targets and provide valuable insights into the treatment of diseases associated with dysregulated signaling pathways (Diaz et al., 2005; Lu, Simin, Khan, & Mercurio, 2008; Q. Y. Wang et al., 2018; Zhou, Tardivel, Thorens, Choi, & Tschopp, 2010). However, to date, no comprehensive and comparative analysis of PPIs associated with α-arrestins has been conducted.

In this study, we conducted affinity purification/mass spectrometry (AP/MS) of six human and twelve *Drosophila* α-arrestins. A high-confidence PPI network was constructed by selecting a cut-off for receiver operating characteristic (ROC) curves of Significance Analysis of INTeractome express (SAINTexpress) scores (Teo et al., 2014). The constructed interactomes were validated using known affinities between domains of prey proteins and the short linear motifs of α-arrestins. We also investigated orthologous relationships between binding partners from human and *Drosophila* and found that many proteins with both known and novel functions could be conserved between two species. Finally, we performed experiments to provide new insights into the functions of TXNIP and ARRDC5 that were revealed in our study. Together, our results provide a valuable resource that describes the

PPI network for α-arrestins in both human and *Drosophila* and suggest novel regulatory axes of α-arrestins.

Results

High-confidence α-arrestin interactomes in human and *Drosophila*

Genome-scale sets of prey proteins interacting with α-arrestins (referred to herein as 'interactomes') were compiled by conducting AP/MS for six human and twelve *Drosophila* α-arrestin proteins (Figure 1A; Table S1). Proteins possibly interacting with α-arrestins were pulled down from total cell lysates of human embryonic kidney 293 (HEK293) and S2R+ cells stably expressing GFP-tagged α-arrestins (Figure 1B; Figure S1). All α-arrestin experiments were replicated twice, and negative control experiments were conducted multiple times. In total, 3,243 and 2,889 prey proteins involved in 9,908 and 13,073 PPIs with human and *Drosophila* α-arrestins, respectively, were initially detected through AP/MS.

To build high-confidence interactomes of α-arrestin family proteins, a probabilistic score for individual PPIs was estimated using SAINTexpress (Teo et al., 2014) and an optimal cutoff for the scores was set using positive and negative PPIs of α-arrestin from public databases and the literature (Colland et al., 2004; Dotimas et al., 2016; Draheim et al., 2010; Mellacheruvu et al., 2013; Nabhan et al., 2012; Nishinaka et al., 2004; Puca & Brou, 2014; Szklarczyk et al., 2015; Warde-Farley et al., 2010; Wu et al., 2013) (Tables S2). The resulting ROC curves showed high area under curve (AUC) values and the SAINTexpress scores at which the false discovery rate (FDR) was 0.01 were selected as cutoffs (0.85 for human and 0.88 for *Drosophila,* Figure 1C). Given the cutoffs, 1,306 and 1,732 PPIs involving 902 and 1,732 proteins were selected for human and *Drosophila*, respectively. Because proteins of low abundance (low spectral counts) are easily affected by a stochastic process (Lundgren, Hwang, Wu, & Han, 2010; Old et al., 2005), the minimum spectral count of PPIs was set at 6, allowing us to select PPIs with higher confidence. In fact, the spectral counts of the filtered PPIs were highly reproducible between replicates (Figure S2A; Pearson's correlations, 0.91 for human; 0.89 for *Drosophila*) and principal component analysis based on $\log_2$ spectral counts also confirmed a high reproducibility between replicates (Figure S2B). As a result, we successfully identified many known interaction partners of α-arrestins such as NEDD4, WWP2,

WWP1, ITCH and TSG101, previously documented in both literatures and PPI databases (Figure S2C-F) (Colland et al., 2004; Dotimas et al., 2016; Draheim et al., 2010; Mellacheruvu et al., 2013; Nabhan et al., 2012; Nishinaka et al., 2004; Puca & Brou, 2014; Szklarczyk et al., 2015; Warde-Farley et al., 2010; Wu et al., 2013). Additionally, we greatly expanded repertoire of PPIs associated with α-arrestins in human and *Drosophila*, resulting in 390 PPIs between six α-arrestins and 307 prey proteins in human, and 740 PPIs between twelve α-arrestins and 467 prey proteins in *Drosophila* (Figure S2E). These are subsequently referred to as 'high-confidence PPIs' (Table S3).

Short-linear motifs and protein domains enriched in α-arrestins and their interactomes

To validify our high-confidence PPIs, we sought to analyze known short-linear motifs in α-arrestins, which are commonly 3-15 stretches of amino acids that are known to participate in interactions with other protein domains (Dinkel et al., 2015). Utilizing the known affinities between short linear motifs in α-arrestins and protein domains in interactomes (El-Gebali et al., 2019; UniProt Consortium, 2018) from eukaryotic linear motif (ELM) database (Dinkel et al., 2015), we evaluated whether our high-confidence PPIs could be explained by the known affinities between them (Table S4). The fractions of our high-confidence PPIs (green, Figure 1D top), supported by the known affinities were significantly greater than those of all raw PPIs (red, Figure 1D top) in both species ($P < 9.37 \times 10^{-11}$ for human and $P < 0.0012$ for *Drosophila*, one-sided Fisher's exact test, Figure 1D top). One of the most well-known short-linear motifs in α-arrestin is PPxY, which is reported to bind with high affinity to the WW domain found in various proteins, including ubiquitin ligases (Ingham, Gish, & Pawson, 2004; Macias et al., 1996; Sudol, Chen, Bougeret, Einbond, & Bork, 1995). Our analysis revealed the specific enrichment of WW domain-containing proteins in the interactomes of α-arrestins with at least one PPxY motif but not in that of the human α-arrestin (ARRDC5) without a PPxY motif (Figure 1D, bottom-left). The interactomes of five out of the eight *Drosophila* α-arrestins with a PPxY motif were enriched for WW domain-containing proteins, but there was no such enrichment for any of the *Drosophila* α-arrestins without a PPxY motif (Figure 1D, bottom-right). In

conclusion, a considerable portion of the high-confidence PPIs identified in this study can be evident by known affinities between short-linear motifs and protein domains.

Next, we conducted enrichment analyses of Pfam proteins domains (El-Gebali et al., 2019; Huang da, Sherman, & Lempicki, 2009b) among interactome of each α-arrestin to investigate known and novel protein domains commonly or specifically associated (Figure S3A; Table S5). The most prominent interacting domains in both species were the Homologous to E6AP C-terminus (HECT), WW, and C2 domains (FigureS3A; Table S5). HECT and C2 domains are well known to be embedded in the E3 ubiquitin ligases such as NEDD4, HECW2, and ITCH along with WW domains (Ingham et al., 2004; Melino et al., 2008; Rotin & Kumar, 2009; Scheffner, Nuber, & Huibregtse, 1995; Weber, Polo, & Maspero, 2019) and as we observed strong preference of WW domains to PPxY containing proteins (Figure 1D), these domains were significantly enriched in binding proteins of α-arrestins with PPxY motif in human and *Drosophila* (FDR < 0.033 ~ 1.23×10^{-11} for human; FDR < 0.045 ~ 4.10×10^{-6} for *Drosophila*, Figure S3A ; Table S5). Other common protein domains involved in the protein degradation process, such as proteasome domains, were also significantly enriched in the interactomes (of ARRDC4 in human and Leash in *Drosophila*) in both species (FDR < 6.41×10^{-4} for human and FDR < 1.30×10^{-5} for *Drosophila,* Figure S3A; Table S5). Interestingly, some α-arrestins (ARRDC3 in human and Vdup1, Leash, and CG18746 in *Drosophila*) appeared to interact in common with RNA binding domains, such as DEAD box, helicase, WD40, and RNA recognition motif, but others did not. In addition, the cargo and motor protein domains IBN_N (FDR < 0.0076 for human and FDR < 2.50×10^{-4} ~ 2.11×10^{-6} for *Drosophila*) and myosin_head (FDR < 0.033 for human and FDR < 2.11×10^{-6} for *Drosophila*) also interacted with several α-arrestins in common (ARRDC4 in human and CG1105, CG18745, and CG18748 in *Drosophila,* Figure S3A; Table S5). These enriched domains explain the conserved interactomes associated with RNA splicing and protein transport in both species. In addition, human α-arrestins seem to interact with human-specific domains, such as PDZ, Rho-GEF, MCM, laminin, zinc finger, and BAG6 domains, providing an expanded

interactomes of human α-arrestins (Figure S3A; domains in black), indicating the presence of both conserved and specific protein domains interacting with α-arrestins.

Expanded functional signatures of α-arrestin interactomes

Because the functions of α-arrestins can be inferred based on their binding partners, the prey proteins were grouped based on their interactions with α-arrestins, which revealed specialized functions of the respective α-arrestins with some redundancy as well as both known and novel functions (Figure 1E). The analysis of protein class enrichment by the PANTHER classification system (Thomas et al., 2003) revealed previously reported functions, such as 'Ubiquitin ligase' (FDR < 0.0019 and 5.01 X 10^{-7} for human; Benjamini-Hochberg correction) and 'Protease' (FDR < 1.93 X 10^{-6} for human and 5.02 X 10^{-6} for *Drosophila*) (Dores et al., 2015; Y. Kwon et al., 2013; Nabhan et al., 2012; Puca et al., 2013; Rauch & Martin-Serrano, 2011; Shea et al., 2012; Xiao et al., 2018). In fact, the known binding partners, NEDD4, WWP2, WWP1, and ITCH in human and CG42797, Su(dx), Nedd4, Yki, Smurf, and HERC2 in *Drosophila*, that were detected in our data are related to ubiquitin ligases and protein degradation (C. Chen & Matesic, 2007; Ingham et al., 2004; Y. Kwon et al., 2013; Marin, 2010; Melino et al., 2008; Rotin & Kumar, 2009) (Figure 1E; Figure S2F). In addition, novel biological functions of α-arrestins were uncovered. For instance, in human, prey proteins interacting with ARRDC3 displayed enrichment of 'RNA splicing factor and helicase' functions as well as 'GTPase-activating proteins', those of ARRDC4 were enriched with 'Apolipoprotein', and those of ARRDC5 with 'ATP synthase' (Figure 1E, up). Motor protein, protease, ubiquitin ligase, RNA splicing factor, and helicase were functions that were also enriched in *Drosophila* prey proteins (Figure 1E, bottom). Among them, the motor protein and RNA splicing, and helicase functions seemed to be novel conserved functions between human and *Drosophila*. The functional compositions of the interacting proteins summarized the common or highly specialized functions of α-arrestins well (Figure 1E, right panel). For example, in human, proteins that interacted with TXNIP, ARRDC2, and ARRDC4 showed similar ubiquitination and protease-related functions, whereas ARRDC3 and ARRDC5 displayed unique

interactomes associated with other functions. For *Drosophila*, the interactomes of the [Vdup1, CG10086 and CG18744], [CG18748 and CG18747], and [CG1105 and CG14696] α-arrestin subsets each exhibited similar functional compositions, but the Leash interactome showed a distinct enrichment of ubiquitination-related and protease functions, consistent with prior reports highlighting Leash's role in the lysosomal degradation of hippo signaling pathway component, Yki (Y. Kwon et al., 2013). Taken together, these results suggest that the resulting high-confidence PPIs of α-arrestins expanded the functional interactome maps of α-arrestins in both human and *Drosophila*.

Subcellular localizations of α-arrestin interactomes

Cellular localizations of proteins often provide valuable information of their functions and activity, but only a small number of α-arrestins are known for their preferential subcellular localization. We thus examined the subcellular localizations of the interacting proteins using the cellular component feature in Gene Ontology (GO) using DAVID (Huang da, Sherman, & Lempicki, 2009a; Huang da et al., 2009b) (Figure S3B; Table S6). Prey proteins (246 for human and 245 for *Drosophila*) that were localized in at least one cellular compartment were examined. We found that prey proteins of ARRDC5 were preferentially localized in the endoplasmic reticulum and at the plasma membrane but were less often localized in the nucleus, compared to those of other human α-arrestins (Figure S3B, top). Similarly, the prey proteins of ARRDC1 and 4 were less often localized in the nucleus, instead being preferentially localized in the cytoplasm (ARRDC4) or extracellular space (ARRDC1), in agreement with previous reports (Nabhan et al., 2012; Q. Y. Wang et al., 2018). TXNIP seemed to preferentially interact with prey proteins in cytoplasm and nucleus (Figure S3B, bottom), consistent with a previous report (S. K. Kim, Choe, & Park, 2019; Saxena, Chen, & Shalev, 2010). ARRDC3, which was suggested to be localized in cytoplasm in previous study (Nabhan et al., 2010), appeared to interact with proteins preferentially localized in nucleus in addition to the ones in cytoplasm, implying novel functions of ARRDC3 in the nucleus. In *Drosophila*, the localization of interacting proteins is often uncharacterized compared to human, but a

preference for a localization for part of the interactomes can be observed (Figure S3B, bottom). Some of them are preferentially localized at the plasma membrane (CG18747), mitochondria (CG14696), peroxisome (CG14696), lysosome (CG2641), or cytoskeleton (CG18748), compared to others. However, interactomes of Leash, Vdup1, CG2641, CG18745, CG18746, and CG10086 are preferentially localized in the nucleus. Taken together, these data about the preferential localizations of interacting proteins provide evidence about the functions and activity of α-arrestins in cells.

Functional complexes in α-arrestin interactomes

The fact that protein functions are often realized in complexes (Hartwell, Hopfield, Leibler, & Murray, 1999) urged us to search for functional complexes that extensively interact with α-arrestins. For this analysis, protein complexes that are significantly connected with each α-arrestin were examined using the COMPlex Enrichment Analysis Tool (COMPLEAT) (Vinayagam et al., 2013), resulting in the detection of 99 and 18 protein complexes for human and *Drosophila*, respectively (Table S7). The complexes were iteratively combined with cellular components from GO (Huang da et al., 2009b) (Table S7) based on the overlap coefficients (Vijaymeena & Kavitha, 2016). The significance of the resulting combined complexes was then tested with the connectivity to each α-arrestin using the interquartile means (IQMs) of SAINTexpress scores compared to those from 1000 random cohorts. This approach showed that 33 clustered complexes comprising 335 protein subunits were significantly interacting with six human α-arrestins (Figure 2A; Figure S4A, FDR < 0.2) and 20 clustered complexes comprising 220 subunits were significantly interacting with *Drosophila* α-arrestins (Figure 2B; Figure S4B, FDR < 0.2).

The two largest complexes found to interact with α-arrestins were related to protein degradation (proteasome and ubiquitin-dependent proteolysis) and RNA splicing and processing in both species (Figure 2; Figure S4). ARRDC1, 2, and 4 and TXNIP in human and Leash and CG2993 in *Drosophila* were found to interact with protein degradation complexes. While the association of ARRDC3 with these ubiquitination-dependent proteolysis complexes

is statistically insignificant, ARRDC3 does interact with individual components of these complexes such as NEDD4, NEDD4L, WWP1, and ITCH (Figure S4A). This suggest their functional relevance in this context, as previously reported in both literatures and databases (Nabhan et al., 2010; Shea et al., 2012; Szklarczyk et al., 2015; Warde-Farley et al., 2010) (Puca & Brou, 2014; Xiao et al., 2018). On the other hand, ARRDC2 and 3 in human and Leash, CG18746, Vdup1, CG10086, and CG18744 in *Drosophila* were found to interact with RNA splicing and processing complexes. Although the above-mentioned α-arrestins interacted in common with the two complexes described above, they were also found to bind to distinct complexes. For instance, TXNIP specifically binds to transcriptional and histone deacetylase (HDAC) complexes, ARRDC1 to endosomal sorting and laminin complexes, ARRDC2 to the Set1C/COMPASS complex, ARRDC3 to transcription elongation factors (TEFs) and spindle assembly checkpoint (SAC) and cell polarity complexes, and ARRDC4 to clathrin-coated pit and BAT3 complexes in human. In *Drosophila*, Leash specifically binds to AP-2 adaptor and WASH complexes and CG18746 to the UTP B complex. ARRDC5 is specifically associated with V-type ATPase and vacuolar protein sorting complexes in human. CG18748 is associated with motor protein complexes including actin, myosin, and microtubule-associated complexes in *Drosophila*. Taken together, the results from this analysis provide a glimpse of underexplored roles for α-arrestins in diverse cellular processes.

Conserved interactomes of α-arrestins

Given that α-arrestins are widely conserved in metazoans (Alvarez, 2008; DeWire, Ahn, Lefkowitz, & Shenoy, 2007), we sought to exploit the evolutionally conserved interactomes of human and *Drosophila* α-arrestins. For this analysis, we searched for orthologous relationships in the α-arrestin interactomes using the DRSC Integrative Ortholog Prediction Tool (DIOPT) (Hu et al., 2011). Among high-confidence prey proteins, 68 in human and 64 in *Drosophila* were reciprocally predicted to have ortholog relationships, defining 58 orthologous prey groups (DIOPT score ≥ 2; Table S8). α-arrestins were then hierarchically clustered based on the $\log_2$-transforemd mean spectral counts of these orthologous interactome, defining

seven groups of α-arrestins. Orthologous prey proteins were grouped according to their shared biological function, defining nine functional groups and others of diverse functions (Figure 3). The resulting clusters revealed PPIs that were functionally conserved. For instance, ARRDC3 in human and CG18746 in *Drosophila* actively interact with proteins in RNA binding and splicing groups. Leash in *Drosophila* appeared to interact with proteins in similar functional groups as ARRDC3 but, like ARRDC1, it also extensively interacts with members of ubiquitin-dependent proteolysis groups. In addition, ARRDC4 interacts with proteins in the motor protein and trafficking group, similar to CG18748 in *Drosophila*, and binds to proteins in the ubiquitin-dependent proteolysis group, similar to TXNIP. Similarly, CG10086 and Vdup1, CG14696 and ARRDC5, and CG2993 and ARRDC2 appeared to have conserved interactomes between human and *Drosophila.*

The most prominent functional modules shared across both species were the ubiquitin-dependent proteolysis, endosomal trafficking, and small GTPase binding modules, which are in agreement with the well-described functions of α-arrestins in membrane receptor degradation through ubiquitination and vesicle trafficking (Dores et al., 2015; S. O. Han et al., 2013; Y. Kwon et al., 2013; Nabhan et al., 2012; Puca & Brou, 2014; Puca et al., 2013; Shea et al., 2012; Xiao et al., 2018; Zbieralski & Wawrzycka, 2022) (Figure 3). In contrast, the functional modules involving cyclin and cyclin-dependent kinase, casein kinase complex, and laminin seemed to be conserved between relatively specific sets of α-arrestins, whereas those related to motor proteins and RNA binding and splicing were more generally conserved. Taken together, the comparative analyses led us to identification of detailed, orthologous interactome maps of α-arrestins, which extend beyond the limited insights provided by sequence-based comparative analysis alone (Figure S5). Conserved roles of α-arrestins in both established and previously uncharacterized signaling pathways expand our understanding of the diverse roles of α-arrestins in cellular signaling.

Chromatin accessibility is globally decreased under TXNIP depletion

TXNIP is one of the most well-studied α-arrestins. Previous studies reported that TXNIP interacts with transcriptional repressors, such as FAZF, PLZF, and HDAC1 or HDAC3, to exert antitumor activity (S. H. Han et al., 2003) or repress NF-kB activation (H. J. Kwon et al., 2010). However, although such studies provided information about interactions with a few transcriptional repressors, they barely provided a systematic view of the roles of TXNIP in controlling the chromatin landscape and gene expression. In that sense, our PPI analysis first revealed that TXNIP extensively binds to chromatin remodeling complexes, such as the HDAC and histone H2B ubiquitination complexes, as well as to transcriptional complexes, such as the RNA polymerase II and transcription factor IIIc complexes (Figure 2A; Figure S4A). Such PPIs indicate that TXNIP could control transcriptional and epigenetic regulators. To examine how the global epigenetic landscape is remodeled by TXNIP, we knocked down its expression in HeLa cells with a small interfering RNA (siTXNIP) and confirmed a decrease at both the RNA and protein levels (Figure 4A and B). We then produced two biological replicates of ATAC- and RNA-seq experiments in HeLa cells with TXNIP depletion (Table S9) to detect differentially accessible chromatin regions (dACRs) and differentially expressed genes (DEGs) (Figure 4C; Table S10). The replicated samples of both ATAC- and RNA-seq were well grouped in principle component spaces (Figure S6A and B). The normalized ATAC-seq signal and the RNA level of expressed genes clearly showed the enrichment of open chromatin signals around the transcription start sites (TSSs) of genes that are actively transcribed (Figure S6C). We detected 70,746 high-confidence accessible chromatin regions (ACRs) across all samples, most of which were located in gene bodies (38.74%), followed by intergenic regions (32.03%) and promoter regions (29.23%, Figure S7A). TXNIP knockdown appeared to induce a global decrease in chromatin accessibility in many genomic regions including promoters (Figure 4D). Of the high-confidence ACRs, 7.38% were dACRs under TXNIP depletion; most dACRs showed reduced chromatin accessibility under this condition (dACRs(-), Figure 4D; Figure S7B). dACRs(-) were preferentially localized in gene bodies, whereas dACRs(+) were more often observed in promoter regions (Figure S7C).

The global chromatin changes induced by TXNIP knockdown could impact gene expression at corresponding loci. In fact, our gene expression analysis showed that 956 genes were downregulated, and 295 genes were upregulated by TXNIP knockdown compared to the control (Figure 4E), suggesting that the global decrease in chromatin accessibility induced by TXNIP depletion would mediate the repression of gene expression. To confirm this phenomenon, we first selected sets of differentially ("Up" and "Down" in Figure 4F) and non-differentially expressed genes ("None" in Figure 4F) with at least one detectable ACR in promoter or gene body. Next, the cumulative distribution function (CDF) of changes in chromatin accessibilities demonstrated that the genes with decreased RNA level ("Down") showed significantly reduced chromatin accessibilities at promoters compared to those with no changes in the RNA level ("None") (Figure 4F; $P < 7.24 \times 10^{-20}$ for max changes; Figure S7D; $P < 2.60 \times 10^{-24}$ for mean changes, Kolmogorov-Smirnov (KS) test). In contrast, genes with increased RNA expression ("Up") exhibited no changes in chromatin accessibilities at the promoter (Figure 4F; $P < 0.68$ for max changes; Figure S7D; $P < 0.49$ for mean changes, KS test), indicating that chromatin opening at promoters is necessary but not sufficient to induce gene expression. ACRs located in gene bodies also showed a similar trend: genes with a decreased RNA level ("Down") showing decreased chromatin accessibility upon TXNIP depletion (Figure S7E; $P < 9.3 \times 10^{-4}$ for max changes and $P < 2.58 \times 10^{-7}$ for mean changes, KS test), suggesting that TXNIP is likely to be a negative regulator of chromatin repressors that induce heterochromatin formation. We then used GO analysis(Raudvere et al., 2019) to examine the biological functions of genes that exhibited decreased chromatin accessibility at their promoter and decreased RNA expression upon TXNIP knockdown (Table S11). In general, genes associated with developmental process, signaling receptor binding, cell adhesion and migration, immune response and extracellular matrix constituents appeared to be repressed upon TXNIP depletion (Figure 4G).

TXNIP represses the recruitment of HDAC2 to target loci

Given that TXNIP knockdown led to a global reduction in chromatin accessibility with decreased transcription, we focused on identifying the potential role of the epigenetic silencer HDAC2, one of the strong binding partners of TXNIP in the AP/MS analysis, in mediating the TXNIP-dependent epigenetic and transcriptional modulation. Consistent with the AP/MS data, immunoprecipitation (IP) experiments showed that the two proteins indeed interact with each other. Furthermore, TXNIP knockdown reduced the amount of TXNIP-interacting HDAC2 protein but did not affect the HDAC2 expression level (Figure 5A). To find out how the TXNIP-HDAC2 interaction impacts the epigenetic and transcriptional reprogramming of target loci, we first checked whether the TXNIP-HDAC2 interaction causes cytosolic retention of HDAC2 to inhibit nuclear HDAC2-mediated global histone deacetylation. However, both the expression level and subcellular localization of HDAC2 were unaffected by a reduction in TXNIP, as confirmed by western blot analysis using cytoplasmic and nuclear fractions as well as by an immunofluorescence assay (Figure 5B; Figure S8A), indicating that TXNIP might modulate HDAC2 activity in a different way.

We next asked if the transcriptional suppression of TXNIP-target genes was mediated by changes in HDAC2 recruitment to and histone acetylation of chromatin. To address this question, genes that were significantly downregulated by TXNIP knockdown and that contained at least one dACR in the promoter were selected by the following additional criteria: 1) the RNA level in normal HeLa cells is ≥ 10 TPM and 2) the total ATAC-seq read count at the promoter in siTXNIP-treated HeLa cells is reduced ≥ 1.5-fold compared to that in normal cells. Among the four TXNIP-target genes selected by the above-mentioned criteria, the expression levels of CD22 and L1CAM were significantly reduced ($P < 0.05$, Student's t-test, Figure S8B). The two genes were further examined to determine whether the levels of HDAC2-binding signal and histone acetylation in their promoter regions were changed upon TXNIP knockdown (Figure 5C). We observed that RNA- and ATAC-seq coverages in exonic and promoter region of CD22 and L1CAM genes were clearly reduced upon TXNIP depletion (Figure 5C top) and an analysis of chromatin immunoprecipitation (ChIP) signals for HDAC2

and histone H3 acetylation at each dACR(-) detected in the L1CAM and CD22 promoters revealed that TXNIP knockdown increased the recruitment of HDAC2 to TXNIP-target loci, accompanied by decreased histone H3 acetylation (Figure 5C bottom). Therefore, these results suggest that the TXNIP interaction with HDAC2 inhibits the chromatin occupancy of HDAC2 and subsequently reduces histone deacetylation to facilitate global chromatin accessibility.

HDAC2 typically operates within the mammalian nucleus as part of co-repressor complexes as it lacks ability to bind to DNA directly (Hassig, Fleischer, Billin, Schreiber, & Ayer, 1997). The nucleosome remodeling and deacetylation (NuRD) complex is one of the well-recognized co-repressor complexes that contains HDAC2 (Kelly & Cowley, 2013; Seto & Yoshida, 2014) and we sought to determine if depletion of TXNIP affects interaction between HDAC2 and other components in this NuRD complex. While HDAC2 interacted with MBD3 and MTA1 under normal condition, the interaction between HDAC2 and MBD3 or MTA1 was not affected upon TXNIP depletion (Figure S8C). Next, given that HDAC2 phosphorylation is known to influence its enzymatic activity and stability (Adenuga & Rahman, 2010; Adenuga, Yao, March, Seagrave, & Rahman, 2009; Bahl & Seto, 2021; Tsai & Seto, 2002), we tested if TXNIP depletion alters phosphorylation status of HDAC2. The result indicated, however, that phosphorylation status of HDAC2 does not change upon TXNIP depletion (Figure S8D). In summary, our findings suggest a model where TXNIP plays a role in transcriptional regulation independent of these factors (Figure S8E). When TXNIP is present, it directly interacts with HDAC2, a key component of transcriptional co-repressor complex. This interaction suppresses the HDAC2 ‘s recruitment to target genomic regions, leading to the histone acetylation of target loci possibly through active complex including histone acetyltransferase (HAT). As a result, transcriptional activation of target gene occurs. In contrast, when TXNIP expression is diminished, the interaction between TXNIP and HDAC2 weakens. This restores histone deacetylating activity of HDAC2 in the co-repressor complex, leading to subsequent repression of target gene transcription.

ARRDC5 plays a role in osteoclast differentiation and function

Given that various subunits of the V-type ATPase interact with ARRDC5, we speculated that ARRDC5 might be involved in the function of this complex (Figure 6A). V-type ATPase plays an important role in the differentiation and function of osteoclasts, which are multinucleated cells responsible for bone resorption in mammals (Feng et al., 2009; Qin et al., 2012). Therefore, we hypothesized that ARRDC5 might be also important for osteoclast differentiation and function. To determine whether ARRDC5 affects osteoclast function, we prepared osteoclasts by infecting bone marrow-derived macrophages (BMMs) with lentivirus expressing either GFP-GFP or GFP-ARRDC5 and differentiating the cells into mature osteoclasts. After five days of differentiation, ectopic expression of GFP-ARRDC5 had significantly increased the total number of tartrate-resistant acid phosphatase (TRAP)-positive multinucleated cells compared to GFP-GFP overexpression (Figure 6B). In particular, the number of TRAP-positive osteoclasts with a diameter larger than 200 μm was significantly increased by GFP-ARRDC5 overexpression (Figure 6B), suggesting that ARRDC5 expression increased osteoclast differentiation. Additionally, the area of resorption pits produced by GFP-ARRDC5-expressing osteoclasts in a bone resorption pit assay was approximately 4-fold greater than that of GFP-GFP-expressing osteoclasts (Figure 6C). Depletion of ARRDC5 using short hairpin RNA (shRNA) impaired osteoclast differentiation, further affirming its crucial role in this differentiation process (Figure 6D and E)**.** These results imply that ARRDC5 promotes osteoclast differentiation and bone resorption activity.

The V-type ATPase is localized at the osteoclast plasma membrane (Toyomura et al., 2003) and its localization is important for cell fusion, maturation, and function during osteoclast differentiation (Feng et al., 2009; Qin et al., 2012). Furthermore, its localization is disrupted by bafilomycin A1, which is shown to attenuate the transport of the V-type ATPase to the membrane (Matsumoto & Nakanishi-Matsui, 2019). We analyzed changes in the expression level and localization of V-type ATPase, especially V-type ATPase V1 domain subunit (ATP6V1), in GFP-GFP and GFP-ARRDC5 overexpressing osteoclasts. The level of V-type

ATPase expression did not change in osteoclasts regardless of ARRDC5 expression levels (Figure 6F). GFP signals were detected at the cell membrane when GFP-ARRDC5 was overexpressed, indicating that ARRDC5 might also localize to the osteoclast plasma membrane (Figure 6G). In addition, we detected more V-type ATPase signals at the cell membrane in the GFP-ARRDC5 overexpressing osteoclasts, and ARRDC5 and V-type ATPase were co-localized at the osteoclast membrane (Figure 6G). Notably, bafilomycin A1 treatment reduced not only the V-type ATPase signals detected at the cell membrane but also the GFP-ARRDC5 signals (Figure 6D). These results indicate that ARRDC5 might control the plasma membrane localization of the V-type ATPase during osteoclast differentiation and function.

Discussion

In this study, we constructed high-confidence interactomes of α-arrestins from human and *Drosophila*, comprising 307 and 467 interacting proteins, respectively. The resulting interactomes greatly expanded previously known PPIs involving α-arrestins and the majority of interactomes were first reported in this study, which needs to be validated experimentally. However, some known PPIs were missed in our interactomes due to low spectral counts and SAINTexpress scores, probably resulting from different cellular contexts, experimental conditions, or other factors (Figure S2F). According to a phylogenetic analysis of arrestin family proteins, α-arrestins were shown to be ubiquitously conserved from yeast to human (Alvarez, 2008). However, compared to the more established visual/ β-arrestin proteins, α-arrestins have been discovered more recently and much of their molecular mechanisms and functions remain mostly unexplored except for budding yeast model (Zbieralski & Wawrzycka, 2022). Based on the high-confidence interactomes of α-arrestins from human and *Drosophila*, we identified conserved and specific functions of these α-arrestins. Furthermore, we uncovered molecular functions of newly discovered function of human specific α-arrestins, TXNIP and ARRDC5. We anticipate that the discovery made here will enhance current understanding of α-arrestins.

Integrative map of protein complexes that interact with α-arrestins (Figure 2; Figure S4) hint towards many aspects of α-arrestins's biology that remain uncharacterized. For example, role of α-arrestins in the regulation of β2AR in human remained controversial. One study proposed that α-arrestins might act coordinately with β-arrestins at the early step of endocytosis, promoting ubiquitination, internalization, endosomal sorting and lysosomal degradation of activated GPCRs (Shea et al., 2012). The another study, however, proposed different hypothesis suggesting that α-arrestins might act as secondary adaptor localized at endosomes to mediate endosomal sorting of cargo molecules (S. O. Han et al., 2013). Among the protein complexes that interact with α-arrestins, we identified those related with clathrin-

coated pit in human (Figure 2A; Figure S4A) and AP-2 adaptor complex in *Drosophila* (Figure 2B; Figure S4B). They are multimeric proteins to induce internalization of cargo molecules to mediate clathrin-mediated endocytosis, which suggests involvement of α-arrestins in early step of endocytosis.

The integrative map of protein complexes also highlighted both conserved and unique relationships between α-arrestins and diverse functional protein complexes. For instance, protein complexes involved in ubiquitination-dependent proteolysis, proteasome, RNA splicing, and intracellular transport (motor proteins) were prevalently linked with α-arrestins in both human and *Drosophila*. To more precisely identify conserved PPIs associated with α-arrestins, we undertook ortholog predictions within the α-arrestins' interactomes. This revealed 58 orthologous interaction groups that were observed to be conserved between human and *Drosophila* (Figure 3). Among conserved proteins, proteins known to interact with human α-arrestins, such as NEDD4, WWP2, WWP1, and ITCH, were identified along with its orthologs in *Drosophila*, which are Su(dx), Nedd4, and Smurf, implying that regulatory pathway of ubiquitination-dependent proteolysis by α-arrestins is also present in invertebrate species. Besides the known conserved functions, the novel conserved functions of α-arrestins interactomes were also identified, such as RNA splicing (Figure 3). Because our protocol did not include treatment with RNase before the AP/MS, it is possible that RNA binding proteins could co-precipitate with other proteins that directly bind to α-arrestins through RNAs, and thus could be indirect binding partners. Nevertheless, other RNA binding proteins except for RNA splicing and processing factors were not enriched in our interactomes, indicating that this possibility may be not the case. Thus, it might be of interest to explore how α-arrestins are linked to RNA processing in future. Additionally, interaction between α-arrestins and entities like motor proteins, small GTPase, ATP binding proteins, and endosomal trafficking components were identified to be conserved. Further validation of these interactions could unveil molecular mechanisms consistently associated with these cellular functions.

Some protein complexes and functional modules were found to be involved in specific cellular processes discovered in only human, suggesting that some functional roles of α-arrestins have diverged through evolution. As examples of specific cellular functions of α-arrestins, we explored the biological relevance of two interacting protein complexes: 1) the interaction between TXNIP and chromatin remodelers and 2) the interaction between ARRDC5 and the V-type ATPase complex. Given that TXNIP interacts with chromatin remodelers, such as the HDAC, we speculated that chromatin structures could be affected by the interactions. Although we showed that siTXNIP treatment directed a global decrease in chromatin accessibilities and gene expression by inhibiting the binding of HDAC2 to targets, histones themselves could be also controlled by the interaction between TXNIP and the H2B ubiquitination complex. An impact of TXNIP on histone ubiquitination could strengthen the negative regulation of target loci by siTXNIP treatment. In addition, TXNIP interacts with the proteasome, which induces the degradation of binding partners (Figure 2A; Figure S4A). However, we observed that the cellular level and localization of HDAC2 were not affected by TXNIP reduction (Figure 5A and B; Figure S8A), meaning that the proteasome seems not to be involved in TXNIP's influence on HDAC2; rather, TXNIP directly hinders HDAC2 recruitment to target loci.

Because the V-type ATPase plays a key role in osteoclast differentiation and physiology (Feng et al., 2009; Qin et al., 2012), we investigated a possible role for the ARRDC5-V-type ATPase interaction in this cell type. We demonstrated that the ectopic expression of ARRDC5 enhances both the differentiation of osteoclasts into their mature form and their bone reabsorption activity (Figure 6B and C). Conversely, depletion of ARRDC5 reduces osteoclast maturation, underscoring the pivotal role of ARRDC5 in osteoclast development and function (Figure 6D and E). Additionally, ARRDC5 co-localized with the V-type ATPase at the plasma membrane (Figure 6G). Thus, further characterization of ARRDC5 and its interactome in osteoclasts might clarify how ARRDC5 regulates the V-type ATPase to play a role in osteoclast differentiation and function. With the results, the discovery of new

binding partners and their functions of TXNIP and ARRDC5 will facilitate the further investigations to explore the novel PPIs of α-arrestins.

Given the plethora of PPIs uncovered in this study, we also anticipate that our study could provide insight into many disease models. In fact, despite a limited knowledge of their biology, α-arrestins have already been linked to a range of cellular processes and several major health disorders, such as diabetes (Batista et al., 2020; Wondafrash et al., 2020), cardiovascular diseases (Domingues, Jolibois, Marquet de Rouge, & Nivet-Antoine, 2021), neurological disorders (Tsubaki, Tooyama, & Walker, 2020), and tumor progression (Y. Chen et al., 2020; Mohankumar et al., 2015; Oka et al., 2006), making them potential therapeutic targets. We further explored association between α-arrestins' interactomes and disease pathways (Figure S9). Notably, the interactomes of α-arrestins in human showed clear links to specific diseases. For instance, ARRDC5 is closely associated with disease resulting from viral infection and cardiovascular conditions. ARRDC2, ARRDC4, and TXNIP share common association with certain neurodegenerative diseases, while ARRDC1 is implicated in cancer.

Lastly, to assist the research community, we have made comprehensive α-arrestin interactome maps on our website (big.hanyang.ac.kr/alphaArrestin_PPIN). Researchers can search and download their interactomes of interest as well as access information on potential cellular functions and protein class associated with these interactomes.

Materials and Methods

Generating *Drosophila* α-arrestin-GFP fusion DNA constructs

To create Drosophila ARRDC entry clones, we gathered cDNA sequences of twelve *Drosophila* α-arrestins : CG2993 (#2276, Drosophila Genomics Resource Center, DGRC, Bloomington, IN, USA), CG18744 (#1388606, DGRC), CG18745 (#12871, DGRC), CG18746 #9217, DGRC), CG18747 (#1635366, DGRC), CG18748 (#1387253, DGRC), CG2641 (#1649402, DGRC), CG10086 (#8816, DGRC), CG14696 (#1644977, DGRC), CG1105 (#4234, DGRC), Vdup1 (#1649326, DGRC), and Leash (Y. Kwon et al., 2013). We then subcloned each cDNA sequence of *Drosophila* α-arrestins into pCR8 entry clone vector using pCR8/GW/TOPO TA cloning kit (#K250020, Thermo Fisher Scientific, Waltham MA, USA), by following manufacturer's protocol. To generate plasmids with suitable system for protein expression in *Drosophila* cell culture, we then subcloned these α-arrestins-containing-pCR8 plasmids into pMK33-Gateway-GFP destination vector (Y. Kwon et al., 2013; Kyriakakis, Tipping, Abed, & Veraksa, 2008) using Gateway LR Clonase II enzyme mix (#11791020, Thermo Fisher Scientific), where coding sequences of α-arrestins are inserted before GFP sequence. Final constructs were validated by GENEWIZ Sanger Sequencing.

Establishing *Drosophila* α-arrestin-GFP stably expressing cell lines

S2R+ cells were maintained in Schneider's *Drosophila* Medium (#21720024, Thermo Fisher Scientific) supplemented with 10% heat inactivated FBS (#16140071, Thermo Fisher Scientific) and 1% Penicillin Streptomycin (#15070063, Thermo Fisher Scientific) at 24°C. To establish α-arrestin-GFP stably expressing *Drosophila* cell lines, 0.4×10^6 S2R+ cells were seeded in 6-well plates and were transfected with 1 µg of each pMK33-ARRDC-GFP construct using Effectene transfection reagent (#301425, Qiagen, Venlo, Netherlands). pMK33 plasmid is a copper-induced protein expression vector, which carries Hygromycin B-antibiotic-resistant gene. Therefore, we selected for α-arrestin-GFP stable cell lines by maintaining cells in Schneider's *Drosophila* Medium supplemented with 200 µM Hygromycin B (#40-005, Fisher

Scientific). The stable cells were transferred into T25 cm^2 flasks to repopulate. To induce the expression of α-arrestin-GFP fusion proteins, we exposed the stable cells to 500 μM $CuSO_4$ (#C8027, Sigma Aldrich, Burlington, MA, USA) to the media. We confirmed the GFP-tagged α-arrestin protein expressions using fluorescence microscopy.

Synthesizing human α-arrestin coding sequence

Due to the lack of commercially available stock, we utilized GENEWIZ (South Plainfield, NJ, USA) gene synthesis service to synthesize human ARRDC5 coding sequence (NM_001080523).

Generating mammalian GFP- α-arrestin fusion DNA constructs

To create human α-arrestin entry clones, we subcloned ARRDC3 (#38317, Addgene, Watertown, MA, USA) and ARRDC5 (GENEWIZ) into pCR8 entry clone vector using pCR8/GW/TOPO TA cloning kit (#K250020, (Thermo Fisher Scientific), by following manufacturer's protocol. ARRDC1 (BC032346, GeneBank), ARRDC2 (BC022516, GeneBank), ARRDC4 (BC070100, GeneBank), and TXNIP (BC093702, GeneBank) were cloned into pCR8. To generate plasmids with suitable system for protein expression in mammalian cell culture, we then subcloned these α-arrestin s-containing-pCR8 plasmids into pHAGE-GFP-Gateway destination vector (gift from Dr. Chanhee Kang at Seoul National University) using Gateway LR Clonase II enzyme mix (#11791020, Thermo Fisher Scientific), where coding sequences of α-arrestin are inserted after GFP sequence. Final constructs were validated by GENEWIZ Sanger Sequencing.

Establishing mammalian GFP- α-arrestin stably expressing cell lines

We produced GFP-α-arrestins lentiviral particles by seeding 5 $x10^6$ HEK293T cells in 10 cm^2 dish with Dulbecco's Modified Eagle Medium (#11965118, Thermo Fisher Scientific) supplemented with 25 mM HEPES, 10% heat-inactivated fetal bovine serum (#16140071, Thermo Fisher Scientific), and 1% Penicillin Streptomycin (#15070063, Thermo Fisher

Scientific) at 37°C in humidified air with 5% CO_2. Approximately after 16-24 hours (hr), at 90% cell confluency, we changed the cell media to Opti-MEM medium (#31985070, Thermo Fisher Scientific) and transfected the cells with 10 µg pHAGE-GFP-α-arrestin construct, 10 µg lentivirus packaging plasmid (pCMV-dR8.91), and 10 µg virus envelope plasmid (VSV-G) using PEIPro DNA transfection reagent (#115010, VWR, Radnor, PA, USA). GFP-α-arrestins lentiviral particles were harvested 40 hr-post transfections. To establish GFP-α-arrestins stably expressing mammalian cell lines, HEK293 cells were seeded in 10 cm^2 dish with Dulbecco's Modified Eagle Medium (#11965118, Thermo Fisher Scientific) supplemented with 25 mM HEPES, 10% heat-inactivated fetal bovine serum (#16140071, Thermo Fisher Scientific) and 1% Penicillin Streptomycin (#15070063, Thermo Fisher Scientific) at 37°C in humidified air with 5% CO_2. At 90% cell confluency, cells were infected with pHAGE-GFP-ARRDC lentivirus particle, and stable cells were selected by maintaining cells in media supplemented with1.5 µg/mL puromycin (#BP2956100, Thermo Fisher Scientific). We confirmed the GFP-tagged α-arrestin protein expressions using fluorescence microscopy.

Affinity purification of *Drosophila* and human GFP-tagged α-arrestin complexes

We seeded each of the *Drosophila* α-arrestin-GFP stable cells in six T-75 cm^2 flasks (2.1x 10^6 cells per flask) and α-arrestin-GFP expression was induced for 48 hr with 500 µM $CuSO_4$. Meanwhile, we seeded each of the human GFP-α-arrestin stable cells in eight T-75 cm^2 flasks and grown for 48 hr before collection. The cells were harvested by spinning down cells at 1,000g for 5 minutes (min) and washed once with cold PBS. We lysed the cells by resuspending cells in lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1.5 mM $MgCl_2$, 5% glycerol, 0.5% NP-40, 25 mM NaF, 1mM DTT, and 1x HALT protease and phosphatase inhibitor (#PI78442, Thermo Fisher Scientific)) and incubating them for 40 min. The lysate was separated from the insoluble fraction by centrifugation at 20,000 g for 15 min at 4℃. To capture the α-arrestins and their native interactors, each α-arrestin-containing lysate was incubated with GFP-nanobody-conjugated to Dynabeads M-270 Epoxy magnetic beads

(#14301, Thermo Fisher Scientific). The supernatant was separated from the beads using a magnetic rack, and the beads were washed five times with lysis buffer. The protein complexes were eluted from the beads by adding 200 mM glycine pH 2.5 and the pH was neutralized with Tris base pH 10.4. Purified α-arrestin proteins were confirmed by running a fraction of the eluted proteins on SDS-PAGE/Coomassie gel.

Protein sample preparation for mass spectrometry

To digest protein samples into peptides for mass spectrometry analysis, we precipitated the eluted proteins by adding trichloroacetic acid (#T0699, Sigma Aldrich) to 20% final concentration, followed by spinning down samples at maximum speed for 30 min at 4℃. The precipitates were washed with 10% trichloroacetic acid solution and three additional times with Acetone (#A929, Thermo Fisher Scientific), and left to dry in room temperature. Protein precipitations were digested with Trypsin (Promega, #V5113) diluted in Digestion buffer (100 mM Ammonium Bicarbonate and 10% Acetonitrile) in 1:40 ratio. Resulting peptides were purified using ZipTip Pipet tips (#ZTC18M096, Thermo Fisher Scientific).

LC/MS-MS analysis

We used cells stably expressing GFP and wild-type HEK293 or S2R+ cells alone as control baits. AP/MS experiments for all *Drosophila* and human α-arrestin baits were performed in two biological replicates, with the exception of human ARRDC3 baits (two technical replicates). Samples were resuspended in Mass Spectrometry buffer (5% Formic Acid and 5% Acetonitrile) and were analyzed on a Liquid Chromatography Orbitrap Fusion Lumos Tribrid Mass Spectrometer (#IQLAAEGAAPFADBMBHQ, Thermo Fisher Scientific) equipped with a nano-Acquity UPLC system and an in-house developed nano spray ionization source. Peptides were separated using a linear gradient, from 5-30% solvent B (LC-MS grade 0.1% formic acid (#A117, Thermo Fisher Scientific) and acetonitrile) in a 130 min period at a flow rate of 300 nL/min. The column temperature was maintained at a constant 5ºC during all experiments. Peptides were detected using a data dependent method. Survey scans of

peptide precursors were performed in the Orbitrap mass analyzer from 380 to 1500 m/z at 120K resolution (at 200 m/z) with a 5 x 105 ion count target and a maximum injection time of 50 milliseconds (ms). The instrument was set to run in top speed mode with 3 seconds (sec) cycles for the survey and the MS/MS scans.

Identification of high-confidence bait-prey PPIs

SAINTexpress analysis: To identify high-confidence bait-prey PPIs, spectral counts of AP/MS data from S2R+ and HEK293 cells were subjected to the SAINTexpress algorithm (version 3.6.1) (Teo et al., 2014), which calculates the probability of authenticity for each bait-prey PPI. The program outputs the SAINTexpress scores and the Bayesian false discovery rates (BFDR) based on the spectral count distribution of true and false PPI sets. Before calculating the scores, bait-to-bait self-interactions were removed manually. SAINTexpress was run with the "-R 2" parameter, which specifies the number of replicates, and the "-L 3" parameter, which specifies the number of representative negative control experiments to be considered.

PPI validation datasets: To evaluate the performance of the PPI prediction based on the SAINTexpress score, validation datasets including positive and negative PPIs were precompiled as described in previous studies (Y. Kwon et al., 2013; Vinayagam et al., 2016). Briefly, the positive PPIs were initially collected by searching for known PPIs involving α-arrestins from STRING version 10.5 (Szklarczyk et al., 2015), GeneMANIA version 3.4.1 (Warde-Farley et al., 2010), Bioplex (Huttlin et al., 2015), and DpiM (Guruharsha et al., 2011). For human, additional positive PPIs were curated from the literature (Colland et al., 2004; Dotimas et al., 2016; Nabhan et al., 2012; Nishinaka et al., 2004; Puca & Brou, 2014; Wu et al., 2013). After these steps, 30 PPIs (21 preys) for human and 46 PPIs (17 preys) for *Drosophila* were considered as positive PPIs (Tables S2A and C). Proteins manually curated from the Contaminant Repository for Affinity Purification (CRAPome) (Mellacheruvu et al., 2013) were compared to those detected in our negative controls and only those that were detected in both were considered as were negative PPIs (Tables S2B and D). As a result of

these steps, 1,372 PPIs (268 preys) for human and 1,246 PPIs (122 preys) for *Drosophila* were compiled as negative PPIs.

Construction of high-confidence PPI networks: The performance of SAINTexpress was evaluated using the positive and negative PPIs. Because there is an imbalance between positive and negative PPIs, 1000 random cohorts of negative PPIs number-matched with that of positive PPIs were generated. The average true positive and false positive rates were plotted as ROC curves over different SAINTexpress scores as a cutoff, and AUC values were calculated using the ROCR R package (version 1.0-11, https://cran.r-project.org /web/packages/ROCR). Based on these results, we chose an optimal cutoff for high-confidence PPIs with a BFDR of 0.01, where the false positive rates were less than 3% (~1.8 % for human and ~2.7% for *Drosophila*) in both species, and the true positive rates were substantially higher (~66.7 % for human and ~45.7% for *Drosophila*). The cutoffs correspond to SAINTexpress scores of 0.85 and 0.88 for human and *Drosophila*, respectively.

Identification of protein complexes associated with α-arrestins

To examine protein complexes significantly enriched in the α-arrestin interactomes, we collected known protein complexes from two databases: COMPLEAT (Vinayagam et al., 2013), which is a comprehensive resource of protein complexes built from information in the literature and predicted by orthologous relationships of proteins across species (human, *Drosophila*, and yeast), and the DAVID GO analysis of cellular components (Huang da et al., 2009a, 2009b) (Benjamini-Hochberg FDR ≤ 0.05) (Table S7B and D), from which bulk cellular compartments such as the nucleus, cytosol, and so on were excluded. From the COMPLEAT database, we evaluated the association of the resulting protein complexes with each α-arrestin by the complex association score, which is the IQM of SAINTexpress scores (Equation 1)

$$Complex\ association\ score\ (IQM) = \frac{\sum_{i=Q1}^{Q3} SAINTexpress\ score_i}{(Q3-Q1)+1}$$ [Equation 1]

, where the first quartile is *Q1* = $\frac{N}{4} + 1$, the third quartile is *Q3* = $\frac{3N}{4}$, and *N* is the total number of preys in the complex. The significance of the complex association score was estimated by comparing the score to the null distribution of the scores calculated from 1,000 random complexes of input proteins. The significance was tested through the online COMPLEAT tool, and protein complexes with $P < 0.05$ were selected for further analysis (Table S7A and C). Next, we iteratively combined (clustered) the pairs of protein complexes from any two databases (COMPLEAT and GO analysis of cellular components) that showed the highest overlap coefficients, $Overlap(X, Y)$ (Equation 2) (Vijaymeena & Kavitha, 2016), until there was no pair of complexes whose coefficients were higher than 0.5.

$$Overlap(X, Y) = \frac{|X \cap Y|}{(|X|, |Y|)} \quad \text{[Equation 2]}$$

From the clustered set of complexes, we manually removed those with fewer than three subunits or two PPIs. Subunits in the complexes that have no connection among themselves were also removed. Lastly, the significance of associations of the resulting complexes with each α-arrestin were tested in the same manner as done in COMPLEAT using complex association score. The resulting *P* values were corrected by the Benjamini-Hochberg procedure and only interactions with statistical significance (FDR < 0.2) were visualized with Cytoscape v3.5.1 (Shannon et al., 2003) (Figure 2; Figure S4).

Orthologous networks of α-arrestin interactomes

DIOPT (version 7.1) was used to search for orthologs of all prey proteins and only those with a DIOPT score ≥ 2 were selected for the identification of orthologous PPIs between *Drosophila* and human. Next, the orthologs were tested for the enrichment of GO biological process and molecular functions and Kyoto Encyclopedia of Genes and Genomes pathway using the DAVID (Huang da et al., 2009a, 2009b). In addition, manual curation of individual genes was performed through the Uniprot database (UniProt Consortium, 2018). The orthologs were manually grouped into functional modules based on the results and α-arrestins

were modularized into seven groups based on hierarchical clustering of log_2-transformed mean spectral counts using the correlation distance and the Ward linkage method. The heatmap was plotted using the pheatmap R package (version 1.0.12).

TXNIP knockdown in HeLa cells

HeLa cells (CCL-2; ATCC, Manassas, VA, USA) were cultured in complete DMEM supplemented with 10% FBS and 1% penicillin-streptomycin. Cells were cultured in an incubator at 37°C in humidified air containing 5% CO2. For siRNA-induced knockdown of TXNIP in HeLa cells, the following siRNA duplex was synthesized (Bioneer, Daejeon, South Korea): sense: 5’-GUCAGUCACUCUCAGCCAUdTdT -3', anti-sense: 5'-AUGGCUGAGAGU GACUGACdTdT-3'. Random sequence siRNAs (AccuTarget Negative control siRNA; Bioneer), which are non-targeting siRNAs that have low sequence homology with all humans, mouse, and rat genes, were used as negative controls (siCon). 100 nM of each siRNA was transfected into 105 HeLa cells using Lipofectamine RNAiMAX (#13778075, Invitrogen, Carlsbad, CA, USA; Thermo Fisher Scientific) according to the manufacturer's instructions. Transfected cells were harvested after 48 hr for RNA-seq and ATAC-seq (two biological replicates for each sequencing data).

RNA sequencing

For RNA-seq, total RNA was extracted using TRIzol (#15596018, Invitrogen; Thermo Fisher Scientific) according to the manufacturer's protocol. Total RNA concentration was calculated by Quant-IT RiboGreen (#R11490, Invitrogen; Thermo Fisher Scientific). To assess the integrity of the total RNA, samples are run on the TapeStation RNA screentape (#5067-5576, Agilent Technologies, Santa Clara, CA, USA). Only high-quality RNA preparations, with RNA integrity number greater than 7.0, were used for RNA library construction. A library was independently prepared with 1ug of total RNA for each sample by Illumina TruSeq Stranded mRNA Sample Prep Kit (#RS-122-2101, Illumina, Inc., San Diego, CA, USA). The first step in the workflow involves purifying the poly-A containing mRNA molecules using poly-T-attached

magnetic beads. Following purification, the mRNA is fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments are copied into first strand cDNA using SuperScript II reverse transcriptase (#18064014, Invitrogen, Thermo Fisher Scientific) and random primers. This is followed by second strand cDNA synthesis using DNA Polymerase I, RNase H and dUTP. These cDNA fragments then go through an end repair process, the addition of a single 'A' base, and then ligation of the adapters. The products are then purified and enriched with PCR to create the final cDNA library. The libraries were quantified using KAPA Library Quantification kits (#KK4854, KAPA BIOSYSTEMS, Wilmington, MA, USA) for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide and qualified using the TapeStation D1000 ScreenTape (#5067-5582, Agilent Technologies). Indexed libraries were then submitted to an Illumina NovaSeq 6000 (Illumina, Inc.) as the paired-end (2×100 bp) sequencing. Both library preparation and sequencing were performed by the Macrogen (Macrogen, Inc., Seoul, South Korea).

ATAC sequencing

100,000 cells were prepared using LUNA-FL™ Automated Fluorescence Cell Counter (#L20001, logos biosystems, Gyeonggi-do, South Korea). Cells were lysed using cold lysis buffer, which consist of Nuclease-free water (#10977023, Invitrogen; Thermo Fisher Scientific), IGEPAL CA-630 (#I8896, Sigma Aldrich), 1M Trizma HCl(PH7.4) (#T2194, Sigma Aldrich), 5M NaCl (#59222C, Sigma Aldrich), and 1M MgCl2 (#M1028, Sigma Aldrich). The nuclei concentration was determined using Countess II Automated Cell Counter (#AMQAX1000, Thermo Fisher Scientific) and nuclei morphology was examined using microscopy. Immediately after lysis, resuspend nuclei (50,000 cells) were put in transposition reaction mix 50 µl, which consist of TED1 2.5µl and TD 17.5 µl (#20034197, Illumina, Inc.), nuclease free water 15 µl, and the nuclei resuspension (50,000 nuclei, 15 µl). The transposition reaction was incubated for 30 min at 37°C. Immediately following transposition, the products were purified using a MinElute PCR purification Kit (#28004, Qiagen). Next, transposed DNA fragments

were amplified using Nextera DNA Flex kit (#20018704, Illumina, Inc.). To reduce GC and size bias in PCR, the appropriate number of cycles was determined as follows: qPCR side reaction was run, the additional number of cycles needed were calculated, liner Rn versus cycle was plotted and the cycle number that corresponds to 1/4 of maximum fluorescent intensity was determined. The remaining PCR reaction was run to the cycle number determined. Amplified library was purified and then quantified using KAPA library quantification kit (#07960255001, Roche, Basel, Switzerland) and Bioanalyzer (Agilent technologies). The resulting libraries were sequenced using HiSeq X Ten (Illumina, Inc.). Both library preparation and sequencing were performed by the Macrogen (Macrogen, Inc).

Processing of RNA-seq data

For quality checks and read trimming, RNA-seq data were processed by FastQC (version 0.11.8) (Andrews, 2010) and sickle (version 1.33) (Joshi NA, 2011) with default parameters. After the trimming, the reads were aligned to human transcriptomes (GENCODE version 29, GRCH38/hg38) (Frankish et al., 2019) using STAR (version 2.5.3a_modified) (Dobin et al., 2013) with default parameters and read counts were determined using RSEM (version 1.3.1) (Li & Dewey, 2011). The DEG analysis was performed using the edgeR R package (version 3.32.1) (Robinson, McCarthy, & Smyth, 2010). Batch information was added as confounding variables to adjust for batch effects. The DEGs are summarized in Supplementary Table S10.

Processing of ATAC-seq data

Each ATAC-seq dataset was processed using the ENCODE ATAC-seq pipeline implemented with Caper (https://github.com/ENCODE-DCC/atac-seq-pipeline) (Jin Lee, 2016). Briefly, reads were mapped to the human reference genome (GRCH38/hg38) using Bowtie2 (version 2.3.4.3), and unmapped reads, duplicates, and those mapped to the mitochondrial genome were removed. Peaks were called by MACS2 (Zhang et al., 2008) and optimal peaks that were reproducible across pseudo replicates were used in the downstream

analysis. The numbers of processed reads and peaks are summarized in Table S9. Plots of ATAC-seq signals around the TSSs of expressed genes were generated by the R genomation package (version 1.22.0) (Akalin, Franke, Vlahovicek, Mason, & Schubeler, 2015). The batch effects of the signals were corrected by the *removeBatchEffect* function from the limma R package (version 3.46.0) (Ritchie et al., 2015). Of the broad and narrow peaks resulting from the ENCODE ATAC-seq pipeline, the latter were used as an input to obtain consensus ACRs using the diffBind R package (version 3.0.15) (Ross-Innes et al., 2012). The dACRs were detected using the edgeR R package (version 3.32.1) (Robinson et al., 2010). In total, 70,746 ACRs and 5,219 dACRs were detected in HeLa cells and are summarized in Table S10. The genomic positions of the ACRs were annotated through the ChIPseeker R package (version 1.26.2) (Yu, Wang, & He, 2015). If the ACRs spanned more than one genomic region, their positions were assigned based on the following priority: promoters > 5' untranslated regions (UTRs) > 3'UTRs > other exons > introns > downstream > intergenic regions. The promoter of a gene was defined as the region 5 kb upstream and 500 bp downstream of the TSS.

Nuclear–cytoplasmic fractionation

Prior to transfection, HeLa cells were seeded in 100 mm cell culture dishes containing Opti-MEM medium and incubated overnight (reaching a confluency of approximately 30%-40%). The cells were then transfected with siTXNIP. Cells were harvested after 48 hr of transfection and fractionated according to the manufacturer's instructions using NE-PER Nuclear and Cytoplasmic Extraction Reagents (#78833, Thermo Fisher Scientific). Protease inhibitor cocktail (P8340; Sigma Aldrich) was added as a supplement to the lysis buffer and the protein concentration was measured using a Pierce BCA Protein Assay Kit (#23225, Thermo Fisher Scientific).

Chromatin Immunoprecipitation (ChIP) Assay

Cells were crosslinked with 1% formaldehyde at 37°C or room temperature for 15 min and the reaction was stopped by the addition of 0.125M glycine. ChIP was then performed

using a ChIP-IT High Sensitivity kit (#53040, Active Motif, Carlsbad, CA, USA) according to the manufacturer's instructions. Enrichment of the ChIP signal was detected by quantitative real-time PCR (qPCR). The data of each biological replicate were normalized with negative control IgG signals and enrichment values were calculated using the ΔΔCt method (Hellemans, Mortier, De Paepe, Speleman, & Vandesompele, 2007). The following antibodies were used: TXNIP (14715, Cell Signaling Technology, Beverly, MA), HDAC2 (57156, Cell Signaling Technology), H3ac antibody (39139; Active Motif, Carlsbad, CA), and normal rabbit IgG antibodies were used. The primers used for ChIP-qPCR are summarized in Table S12.

Osteoclast differentiation and collection of lentiviruses for ARRDC5 expression

BMMs were cultured as previously described (S. Y. Kim et al., 2019). Briefly, bone marrow was obtained from mouse femurs and tibias at 8 weeks of age, and BMMs were isolated from the bone marrow using Histopaque (1077; Sigma Aldrich). BMMs were seeded at a density of 1.2×10^5 cells/well into 24-well culture plates and incubated in α-MEM (SH30265.01; Hyclone, Rockford, IL, USA) containing 20 ng/mL macrophage colony-stimulating factor (M-CSF) (300-25; PeproTech, Cranbury, NJ, USA). To induce osteoclast differentiation, BMMs were treated for 24 hr with lentiviral-containing medium that also contained M-CSF, after which the medium was changed to α-MEM containing 20 ng/ml M-CSF and 20 ng/ml RANKL (462-TEC; R&D Systems, Minneapolis, MN, USA). The differentiation medium was changed every 24 hr during the 5-day differentiation period.

To obtain the media containing lentivirus, HEK293 cells were cultured in DMEM containing 4.5 g/L glucose (SH30243.01; Hyclone) supplemented with 10% FBS (SH30084.03; Hyclone) and 1% penicillin-streptomycin. After seeding cells at a density of 1×10^5 cells/well into 6-well culture plates, the cells were incubated with lentivirus co-transfected media for 16 hr. Lentivirus co-transfected media was prepared according to the manufacturer's instructions using the CRISPR & MISSION® Lentiviral Packaging Mix (SHP002; Sigma Aldrich) and the lentiviral transfer vector. The lentiviral vector pHAGE-GFP-ARRDC5 was used to generate GFP-ARRDC5 overexpressing osteoclasts, and pHAGE-GFP-GFP served

as the control vector. The pLKO.1-puro-CMV-tGFP vector (SHC003; Sigma Aldrich) was employed to generate ARRDC5 knockdown, which contains the target sequence 5'-CCACACCTTTGAACTTCCATTT-3', and this vector was also used to generate its non-target control. After the incubation, the medium was replaced with fresh α-MEM medium supplemented with 10% FBS and 1% penicillin-streptomycin. The medium was collected twice (after 24 and 48 hr), designated as lentiviral-containing medium, and stored in a deep freezer until used to infect BMMs.

TRAP staining and bone resorption pit assay

Osteoclast differentiation and activity were determined by TRAP staining and a bone resorption pit assay, respectively. TRAP staining was performed using a TRAP staining kit (PMC-AK04F-COS; Cosmo Bio Co., LTD., Tokyo, Japan) following the manufacturer's instructions. TRAP-positive multinucleated cells with more than three nuclei were counted under a microscope using ImageJ software (NIH, Bethesda, MD, USA). The bone resorption pit assay was performed using dentin discs (IDS AE-8050; Immunodiagnostic Systems, Tyne & Wear, UK). Cells were differentiated to osteoclasts on the discs over a 4-day period, after which the discs were stained with 1% toluidine blue solution and the resorption pit area was quantified using ImageJ software.

Immunofluorescence staining of the V-type ATPase and visualization with GFP-ARRDC5

To inhibit V-type ATPase transport to the membrane (Matsumoto & Nakanishi-Matsui, 2019), osteoclasts on the fifth day of differentiation were incubated with 100 nM bafilomycin A1 (19-148; Sigma Aldrich) for 3 hr. Then, immunofluorescence staining was performed to visualize the localization of the V-type ATPase in bafilomycin A1-treated and untreated cells. The cells were fixed using a 4% paraformaldehyde solution (PC2031-100; Biosesang, Gyeonggi-do, Korea) and permeabilized using 0.05% Triton X-100 at room temperature for 5 min. The cells were incubated with anti-V-type ATPase antibody (SAB1402125-100UG; Sigma

Aldrich) at room temperature for 1 hr, and then stained with the Alexa Fluor 594-conjugated anti-mouse antibody (A-21044; Invitrogen) at room temperature for 30 min. Finally, cells were mounted using Antifade Mountant with DAPI (P36962; Invitrogen). Fluorescence images were observed under a ZEISS confocal microscope (LSM5; Carl Zeiss, Jena, Germany).

Data availability

All AP/MS raw tables from human and *Drosophila* α-arrestins are deposited in Source Data. ATAC-seq and RNA-seq data from HeLa cells treated with siCon and siTXNIP, can be downloaded from the Korean Nucleotide Archive (KONA; PRJKA220517, https://www.kobic.re.kr/kona/).

Code availability

All source codes and in-house codes used in the study are available at GitHub (https://github.com/Kyung-TaeLee/alphaArrestins_PPI_network)

Acknowledgements

We thank all of the BIGLab and Kwon Lab members for critical reading and comments. This work was supported by the National Research Foundation (NRF) funded by the Ministry of Science & ICT (2020R1A4A1018398, 2021R1A2C3005835, and 2022M3E5F1018502 to JWN, and 2020R1A6A3A13077354 to KTL).

Author contributions

KTL performed all computational analyses; IP performed all AP/MS experiments; SYK contributed to ARRDC5-related functional study, HJC and NBT performed ChIP-related experiments, and NBT and HSC performed ATAC-seq and RNA-seq experiments; KTL contributed to writing the codes; KTL, IP, NBT, HJC, JEK, YK, and JWN contributed to writing the manuscript; YK and JWN supervised the project and conceived the idea.

References

Adenuga, D., & Rahman, I. (2010). Protein kinase CK2-mediated phosphorylation of HDAC2 regulates co-repressor formation, deacetylase activity and acetylation of HDAC2 by cigarette smoke and aldehydes. *Arch Biochem Biophys, 498*(1), 62-73. doi:10.1016/j.abb.2010.04.002

Adenuga, D., Yao, H., March, T. H., Seagrave, J., & Rahman, I. (2009). Histone Deacetylase 2 Is Phosphorylated, Ubiquitinated, and Degraded by Cigarette Smoke. *American Journal of Respiratory Cell and Molecular Biology, 40*(4), 464-473. doi:10.1165/rcmb.2008-0255OC

Akalin, A., Franke, V., Vlahovicek, K., Mason, C. E., & Schubeler, D. (2015). Genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics, 31*(7), 1127-1129. doi:10.1093/bioinformatics/btu775

Alvarez, C. E. (2008). On the origins of arrestin and rhodopsin. *BMC Evol Biol, 8*, 222. doi:10.1186/1471-2148-8-222

Andoh, T., Hirata, Y., & Kikuchi, A. (2002). PY motifs of Rod1 are required for binding to Rsp5 and for drug resistance. *FEBS Lett, 525*(1-3), 131-134. doi:10.1016/s0014-5793(02)03104-6

Andrews, S. (2010). A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bahl, S., & Seto, E. (2021). Regulation of histone deacetylase activities and functions by phosphorylation and its physiological relevance. *Cell Mol Life Sci, 78*(2), 427-445. doi:10.1007/s00018-020-03599-4

Batista, T. M., Dagdeviren, S., Carroll, S. H., Cai, W., Melnik, V. Y., Noh, H. L., . . . Lee, R. T. (2020). Arrestin domain-containing 3 (Arrdc3) modulates insulin action and glucose metabolism in liver. *Proc Natl Acad Sci U S A, 117*(12), 6733-6740. doi:10.1073/pnas.1922370117

Benovic, J. L., DeBlasi, A., Stone, W. C., Caron, M. G., & Lefkowitz, R. J. (1989). Beta-adrenergic receptor kinase: primary structure delineates a multigene family. *Science, 246*(4927), 235-240. doi:10.1126/science.2552582

Boase, N. A., & Kelly, J. M. (2004). A role for creD, a carbon catabolite repression gene from Aspergillus nidulans, in ubiquitination. *Mol Microbiol, 53*(3), 929-940. doi:10.1111/j.1365-2958.2004.04172.x

Chen, C., & Matesic, L. E. (2007). The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer Metastasis Rev, 26*(3-4), 587-604. doi:10.1007/s10555-007-9091-x

Chen, K. S., & DeLuca, H. F. (1994). Isolation and characterization of a novel cDNA from HL-60 cells treated with 1,25-dihydroxyvitamin D-3. *Biochim Biophys Acta, 1219*(1), 26-32. doi:10.1016/0167-4781(94)90242-9

Chen, Y., Ning, J., Cao, W., Wang, S., Du, T., Jiang, J., . . . Zhang, B. (2020). Research Progress of TXNIP as a Tumor Suppressor Gene Participating in the Metabolic Reprogramming and Oxidative Stress of Cancer Cells in Various Cancers. *Front Oncol, 10*, 568574. doi:10.3389/fonc.2020.568574

Colland, F., Jacq, X., Trouplin, V., Mougin, C., Groizeleau, C., Hamburger, A., . . . Gauthier, J. M. (2004). Functional proteomics mapping of a human signaling pathway. *Genome Research, 14*(7), 1324-1332. doi:10.1101/gr.2334104

DeWire, S. M., Ahn, S., Lefkowitz, R. J., & Shenoy, S. K. (2007). Beta-arrestins and cell signaling.

Annu Rev Physiol, 69, 483-510. doi:10.1146/annurev.physiol.69.022405.154749

Diaz, L. K., Cristofanilli, M., Zhou, X., Welch, K. L., Smith, T. L., Yang, Y., . . . Gilcrease, M. Z. (2005). Beta4 integrin subunit gene expression correlates with tumor size and nuclear grade in early breast cancer. *Mod Pathol, 18*(9), 1165-1175. doi:10.1038/modpathol.3800411

Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., . . . Behrendt, A. (2015). ELM 2016—data update and new functionality of the eukaryotic linear motif resource. *Nucleic acids research, 44*(D1), D294-D300. doi:10.1093/nar/gkv1291

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Domingues, A., Jolibois, J., Marquet de Rouge, P., & Nivet-Antoine, V. (2021). The Emerging Role of TXNIP in Ischemic and Cardiovascular Diseases; A Novel Marker and Therapeutic Target. *Int J Mol Sci, 22*(4). doi:10.3390/ijms22041693

Dores, M. R., Lin, H., N, J. G., Mendez, F., & Trejo, J. (2015). The alpha-arrestin ARRDC3 mediates ALIX ubiquitination and G protein-coupled receptor lysosomal sorting. *Mol Biol Cell, 26*(25), 4660-4673. doi:10.1091/mbc.E15-05-0284

Dotimas, J. R., Lee, A. W., Schmider, A. B., Carroll, S. H., Shah, A., Bilen, J., . . . Yoshioka, J. (2016). Diabetes regulates fructose absorption through thioredoxin-interacting protein. *Elife, 5*, e18313. doi:10.7554/eLife.18313

Draheim, K. M., Chen, H. B., Tao, Q., Moore, N., Roche, M., & Lyle, S. (2010). ARRDC3 suppresses breast cancer progression by negatively regulating integrin beta4. *Oncogene, 29*(36), 5032-5047. doi:10.1038/onc.2010.250

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., . . . Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res, 47*(D1), D427-D432. doi:10.1093/nar/gky995

Feng, S. M., Deng, L. F., Chen, W., Shao, J. Z., Xu, G. L., & Li, Y. P. (2009). Atp6v1c1 is an essential component of the osteoclast proton pump and in F-actin ring formation in osteoclasts. *Biochemical Journal, 417*, 195-203. doi:10.1042/Bj20081073

Frankish, A., Diekhans, M., Ferreira, A. M., Johnson, R., Jungreis, I., Loveland, J., . . . Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res, 47*(D1), D766-D773. doi:10.1093/nar/gky955

Guruharsha, K. G., Rual, J. F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., . . . Artavanis-Tsakonas, S. (2011). A protein complex network of Drosophila melanogaster. *Cell, 147*(3), 690-703. doi:10.1016/j.cell.2011.08.047

Han, S. H., Jeon, J. H., Ju, H. R., Jung, U., Kim, K. Y., Yoo, H. S., . . . Choi, I. (2003). VDUP1 upregulated by TGF-beta1 and 1,25-dihydorxyvitamin D3 inhibits tumor cell growth by blocking cell-cycle progression. *Oncogene, 22*(26), 4035-4046. doi:10.1038/sj.onc.1206610

Han, S. O., Kommaddi, R. P., & Shenoy, S. K. (2013). Distinct roles for beta-arrestin2 and arrestin-domain-containing proteins in beta2 adrenergic receptor trafficking. *EMBO Rep, 14*(2), 164-171. doi:10.1038/embor.2012.187

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature, 402*(6761 Suppl), C47-52. doi:10.1038/35011540

Hassig, C. A., Fleischer, T. C., Billin, A. N., Schreiber, S. L., & Ayer, D. E. (1997). Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell, 89*(3), 341-347. doi:10.1016/S0092-8674(00)80214-7

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol, 8*(2), R19. doi:10.1186/gb-2007-8-2-r19

Herranz, S., Rodriguez, J. M., Bussink, H. J., Sanchez-Ferrero, J. C., Arst, H. N., Jr., Penalva, M. A., & Vincent, O. (2005). Arrestin-related proteins mediate pH signaling in fungi. *Proc Natl Acad Sci U S A, 102*(34), 12141-12146. doi:10.1073/pnas.0504776102

Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., & Mohr, S. E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics, 12*(1), 357. doi:10.1186/1471-2105-12-357

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res, 37*(1), 1-13. doi:10.1093/nar/gkn923

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc, 4*(1), 44-57. doi:10.1038/nprot.2008.211

Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., . . . Gygi, S. P. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell, 162*(2), 425-440. doi:10.1016/j.cell.2015.06.043

Ingham, R. J., Gish, G., & Pawson, T. (2004). The Nedd4 family of E3 ubiquitin ligases: functional diversity within a common modular architecture. *Oncogene, 23*(11), 1972-1984. doi:10.1038/sj.onc.1207436

Jin Lee, G. C., Grey Christoforo, CS Foo, Chris Probert, Anshul Kundaje, Nathan Boley, kohpangwei, Mike Dacre, Daniel Kim. (2016). kundajelab/atac_dnase_pipelines: 0.3.3.

Joshi NA, F. J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Junn, E., Han, S. H., Im, J. Y., Yang, Y., Cho, E. W., Um, H. D., . . . Choi, I. (2000). Vitamin D3 up-regulated protein 1 mediates oxidative stress via suppressing the thioredoxin function. *J Immunol, 164*(12), 6287-6295. doi:10.4049/jimmunol.164.12.6287

Kelly, R. D., & Cowley, S. M. (2013). The physiological roles of histone deacetylase (HDAC) 1 and 2: complex co-stars with multiple leading parts. *Biochem Soc Trans, 41*(3), 741-749. doi:10.1042/BST20130010

Kim, S. K., Choe, J. Y., & Park, K. Y. (2019). TXNIP-mediated nuclear factor-kappaB signaling pathway and intracellular shifting of TXNIP in uric acid-induced NLRP3 inflammasome. *Biochem Biophys Res Commun, 511*(4), 725-731. doi:10.1016/j.bbrc.2019.02.141

Kim, S. Y., Lee, E. H., Park, S. Y., Choi, H., Koh, J. T., Park, E. K., . . . Kim, J. E. (2019). Ablation of Stabilin-1 Enhances Bone-Resorbing Activity in Osteoclasts In Vitro. *Calcif Tissue Int, 105*(2), 205-

214. doi:10.1007/s00223-019-00552-x

Kim, Y. M., & Benovic, J. L. (2002). Differential roles of arrestin-2 interaction with clathrin and adaptor protein 2 in G protein-coupled receptor trafficking. *J Biol Chem, 277*(34), 30760-30768. doi:10.1074/jbc.M204528200

Kuhn, H., Hall, S. W., & Wilden, U. (1984). Light-induced binding of 48-kDa protein to photoreceptor membranes is highly enhanced by phosphorylation of rhodopsin. *FEBS Lett, 176*(2), 473-478. doi:10.1016/0014-5793(84)81221-1

Kwon, H. J., Won, Y. S., Suh, H. W., Jeon, J. H., Shao, Y., Yoon, S. R., . . . Choi, I. (2010). Vitamin D3 upregulated protein 1 suppresses TNF-alpha-induced NF-kappaB activation in hepatocarcinogenesis. *J Immunol, 185*(7), 3980-3989. doi:10.4049/jimmunol.1000990

Kwon, Y., Vinayagam, A., Sun, X., Dephoure, N., Gygi, S. P., Hong, P., & Perrimon, N. (2013). The Hippo signaling pathway interactome. *Science, 342*(6159), 737-740. doi:10.1126/science.1243971

Kyriakakis, P., Tipping, M., Abed, L., & Veraksa, A. (2008). Tandem affinity purification in Drosophila: the advantages of the GS-TAP system. *Fly (Austin), 2*(4), 229-235. doi:10.4161/fly.6669

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Lohse, M. J. (1992). Stable overexpression of human beta 2-adrenergic receptors in mammalian cells. *Naunyn Schmiedebergs Arch Pharmacol, 345*(4), 444-451. doi:10.1007/BF00176623

Lu, S., Simin, K., Khan, A., & Mercurio, A. M. (2008). Analysis of integrin beta4 expression in human breast cancer: association with basal-like tumors and prognostic significance. *Clin Cancer Res, 14*(4), 1050-1058. doi:10.1158/1078-0432.CCR-07-4116

Lundgren, D. H., Hwang, S.-I., Wu, L., & Han, D. K. (2010). Role of spectral counting in quantitative proteomics. *Expert review of proteomics, 7*(1), 39-53.

Macias, M. J., Hyvonen, M., Baraldi, E., Schultz, J., Sudol, M., Saraste, M., & Oschkinat, H. (1996). Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature, 382*(6592), 646-649. doi:10.1038/382646a0

Malik, R., & Marchese, A. (2010). Arrestin-2 interacts with the endosomal sorting complex required for transport machinery to modulate endosomal sorting of CXCR4. *Mol Biol Cell, 21*(14), 2529-2541. doi:10.1091/mbc.E10-02-0169

Marin, I. (2010). Animal HECT ubiquitin ligases: evolution and functional implications. *Bmc Evolutionary Biology, 10*. doi:10.1186/1471-2148-10-56

Matsumoto, N., & Nakanishi-Matsui, M. (2019). Proton pumping V-ATPase inhibitor bafilomycin A1 affects Rab7 lysosomal localization and abolishes anterograde trafficking of osteoclast secretory lysosomes. *Biochemical and Biophysical Research Communications, 510*(3), 421-426. doi:10.1016/j.bbrc.2019.01.118

Melino, G., Gallagher, E., Aqeilan, R. I., Knight, R., Peschiaroli, A., Rossi, M., . . . Bernassola, F. (2008). Itch: a HECT-type E3 ligase regulating immunity, skin and cancer. *Cell Death and Differentiation, 15*(7), 1103-1112. doi:10.1038/cdd.2008.60

Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J. P., St-Denis, N. A., Li, T., . . . Nesvizhskii,

A. I. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods, 10*(8), 730-736. doi:10.1038/nmeth.2557

Mo, J. S., Park, H. W., & Guan, K. L. (2014). The Hippo signaling pathway in stem cell biology and cancer. *EMBO Rep, 15*(6), 642-656. doi:10.15252/embr.201438638

Mohankumar, K. M., Currle, D. S., White, E., Boulos, N., Dapper, J., Eden, C., . . . Gilbertson, R. J. (2015). An in vivo screen identifies ependymoma oncogenes and tumor-suppressor genes. *Nat Genet, 47*(8), 878-887. doi:10.1038/ng.3323

Nabhan, J. F., Hu, R., Oh, R. S., Cohen, S. N., & Lu, Q. (2012). Formation and release of arrestin domain-containing protein 1-mediated microvesicles (ARMMs) at plasma membrane by recruitment of TSG101 protein. *Proc Natl Acad Sci U S A, 109*(11), 4146-4151. doi:10.1073/pnas.1200448109

Nabhan, J. F., Pan, H., & Lu, Q. (2010). Arrestin domain-containing protein 3 recruits the NEDD4 E3 ligase to mediate ubiquitination of the beta2-adrenergic receptor. *EMBO Rep, 11*(8), 605-611. doi:10.1038/embor.2010.80

Nishinaka, Y., Masutani, H., Oka, S., Matsuo, Y., Yamaguchi, Y., Nishio, K., . . . Yodoi, J. (2004). Importin alpha1 (Rch1) mediates nuclear translocation of thioredoxin-binding protein-2/vitamin D(3)-up-regulated protein 1. *J Biol Chem, 279*(36), 37559-37565. doi:10.1074/jbc.M405473200

Nishiyama, A., Matsui, M., Iwata, S., Hirota, K., Masutani, H., Nakamura, H., . . . Yodoi, J. (1999). Identification of thioredoxin-binding protein-2/vitamin D-3 up-regulated protein 1 as a negative regulator of thioredoxin function and expression. *Journal of Biological Chemistry, 274*(31), 21645-21650. doi:10.1074/jbc.274.31.21645

Oka, S., Masutani, H., Liu, W., Horita, H., Wang, D., Kizaka-Kondoh, S., & Yodoi, J. (2006). Thioredoxin-binding protein-2-like inducible membrane protein is a novel vitamin D3 and peroxisome proliferator-activated receptor (PPAR)gamma ligand target protein that regulates PPARgamma signaling. *Endocrinology, 147*(2), 733-743. doi:10.1210/en.2005-0679

Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K. G., Mendoza, A., Sevinsky, J. R., . . . Ahn, N. G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics, 4*(10), 1487-1502. doi:10.1074/mcp.M500084-MCP200

Patwari, P., Higgins, L. J., Chutkow, W. A., Yoshioka, J., & Lee, R. T. (2006). The interaction of thioredoxin with Txnip. Evidence for formation of a mixed disulfide by disulfide exchange. *J Biol Chem, 281*(31), 21884-21891. doi:10.1074/jbc.M600427200

Pei, T. M., Li, Y. J., Wang, J. B., Wang, H. L., Liang, Y. J., Shi, H. W., . . . Liu, L. X. (2015). YAP is a critical oncogene in human cholangiocarcinoma. *Oncotarget, 6*(19), 17206-17220. doi:10.18632/oncotarget.4043

Puca, L., & Brou, C. (2014). Alpha-arrestins - new players in Notch and GPCR signaling pathways in mammals. *J Cell Sci, 127*(Pt 7), 1359-1367. doi:10.1242/jcs.142539

Puca, L., Chastagner, P., Meas-Yedid, V., Israel, A., & Brou, C. (2013). Alpha-arrestin 1 (ARRDC1) and beta-arrestins cooperate to mediate Notch degradation in mammals. *J Cell Sci, 126*(Pt 19), 4457-4468. doi:10.1242/jcs.130500

Qayyum, N., Haseeb, M., Kim, M. S., & Choi, S. (2021). Role of Thioredoxin-Interacting Protein in Diseases and Its Therapeutic Outlook. *Int J Mol Sci, 22*(5). doi:10.3390/ijms22052754

Qin, A., Cheng, T. S., Pavlos, N. J., Lin, Z., Dai, K. R., & Zheng, M. H. (2012). V-ATPases in osteoclasts: structure, function and potential inhibitors of bone resorption. *Int J Biochem Cell Biol, 44*(9), 1422-1435. doi:10.1016/j.biocel.2012.05.014

Rauch, S., & Martin-Serrano, J. (2011). Multiple interactions between the ESCRT machinery and arrestin-related proteins: implications for PPXY-dependent budding. *J Virol, 85*(7), 3546-3556. doi:10.1128/JVI.02045-10

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res, 47*(W1), W191-W198. doi:10.1093/nar/gkz369

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. doi:10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140. doi:10.1093/bioinformatics/btp616

Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., . . . Carroll, J. S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature, 481*(7381), 389-393. doi:10.1038/nature10730

Rotin, D., & Kumar, S. (2009). Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol, 10*(6), 398-409. doi:10.1038/nrm2690

Saxena, G., Chen, J., & Shalev, A. (2010). Intracellular shuttling and mitochondrial function of thioredoxin-interacting protein. *J Biol Chem, 285*(6), 3997-4005. doi:10.1074/jbc.M109.034421

Scheffner, M., Nuber, U., & Huibregtse, J. M. (1995). Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature, 373*(6509), 81-83. doi:10.1038/373081a0

Schutte, U., Bisht, S., Heukamp, L. C., Kebschull, M., Florin, A., Haarmann, J., . . . Feldmann, G. (2014). Hippo Signaling Mediates Proliferation, Invasiveness, and Metastatic Potential of Clear Cell Renal Cell Carcinoma. *Translational Oncology, 7*(2), 309-321. doi:10.1016/j.tranon.2014.02.005

Seto, E., & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb Perspect Biol, 6*(4), a018713. doi:10.1101/cshperspect.a018713

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498-2504. doi:10.1101/gr.1239303

Shea, F. F., Rowell, J. L., Li, Y., Chang, T. H., & Alvarez, C. E. (2012). Mammalian alpha arrestins link activated seven transmembrane receptors to Nedd4 family e3 ubiquitin ligases and interact with beta arrestins. *PLoS One, 7*(12), e50557. doi:10.1371/journal.pone.0050557

Shenoy, S. K., & Lefkowitz, R. J. (2011). beta-Arrestin-mediated receptor trafficking and signal transduction. *Trends Pharmacol Sci, 32*(9), 521-533. doi:10.1016/j.tips.2011.05.002

Sudol, M., Chen, H. I., Bougeret, C., Einbond, A., & Bork, P. (1995). Characterization of a novel protein-binding module--the WW domain. *FEBS Lett, 369*(1), 67-71. doi:10.1016/0014-5793(95)00550-s

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., . . . von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res, 43*(Database issue), D447-452. doi:10.1093/nar/gku1003

Teo, G., Liu, G., Zhang, J., Nesvizhskii, A. I., Gingras, A. C., & Choi, H. (2014). SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *J Proteomics, 100*, 37-43. doi:10.1016/j.jprot.2013.10.023

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H. Y., Karlak, B., Daverman, R., . . . Narechania, A. (2003). PANTHER: A library of protein families and subfamilies indexed by function. *Genome Research, 13*(9), 2129-2141. doi:10.1101/gr.772403

Toyomura, T., Murata, Y., Yamamoto, A., Oka, T., Sun-Wada, G. H., Wada, Y., & Futai, M. (2003). From lysosomes to the plasma membrane - Localization of vacuolar type H+-ATPase with the a3 isoform during osteoclast differentiation. *Journal of Biological Chemistry, 278*(24), 22023-22030. doi:10.1074/jbc.M302436200

Tsai, S. C., & Seto, E. (2002). Regulation of histone deacetylase 2 by protein kinase CK2. *J Biol Chem, 277*(35), 31826-31833. doi:10.1074/jbc.M204149200

Tsubaki, H., Tooyama, I., & Walker, D. G. (2020). Thioredoxin-Interacting Protein (TXNIP) with Focus on Brain and Neurodegenerative Diseases. *Int J Mol Sci, 21*(24). doi:10.3390/ijms21249357

UniProt Consortium, T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Res, 46*(5), 2699. doi:10.1093/nar/gky092

Vijaymeena, M., & Kavitha, K. (2016). A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal, 3*(2), 19-28. doi:10.5121/mlaij.2016.3103

Vinayagam, A., Hu, Y., Kulkarni, M., Roesel, C., Sopko, R., Mohr, S. E., & Perrimon, N. (2013). Protein complex-based analysis framework for high-throughput data sets. *Sci Signal, 6*(264), rs5. doi:10.1126/scisignal.2003629

Vinayagam, A., Kulkarni, M. M., Sopko, R., Sun, X., Hu, Y., Nand, A., . . . Perrimon, N. (2016). An Integrative Analysis of the InR/PI3K/Akt Network Identifies the Dynamic Response to Insulin Signaling. *Cell Rep, 16*(11), 3062-3074. doi:10.1016/j.celrep.2016.08.029

Wang, Q. Y., Yu, J. J., Kadungure, T., Beyene, J., Zhang, H., & Lu, Q. (2018). ARMMs as a versatile platform for intracellular delivery of macromolecules. *Nature Communications, 9*. doi:10.1038/s41467-018-03390-x

Wang, Y., Dong, Q. Z., Zhang, Q. F., Li, Z. X., Wang, E. H., & Qiu, X. S. (2010). Overexpression of yes-associated protein contributes to progression and poor prognosis of non-small-cell lung cancer. *Cancer Science, 101*(5), 1279-1285. doi:10.1111/j.1349-7006.2010.01511.x

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., . . . Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res, 38*(Web Server issue), W214-220. doi:10.1093/nar/gkq537

Weber, J., Polo, S., & Maspero, E. (2019). HECT E3 Ligases: A Tale With Multiple Facets. *Front Physiol, 10*, 370. doi:10.3389/fphys.2019.00370

Wilden, U., Wust, E., Weyand, I., & Kuhn, H. (1986). Rapid affinity purification of retinal arrestin (48 kDa protein) via its light-dependent binding to phosphorylated rhodopsin. *FEBS Lett, 207*(2), 292-295.

doi:10.1016/0014-5793(86)81507-1

Wondafrash, D. Z., Nire'a, A. T., Tafere, G. G., Desta, D. M., Berhe, D. A., & Zewdie, K. A. (2020). Thioredoxin-Interacting Protein as a Novel Potential Therapeutic Target in Diabetes Mellitus and Its Underlying Complications. *Diabetes Metab Syndr Obes, 13*, 43-51. doi:10.2147/DMSO.S232221

Wu, N., Zheng, B., Shaywitz, A., Dagon, Y., Tower, C., Bellinger, G., . . . McGraw, T. E. (2013). AMPK-dependent degradation of TXNIP upon energy stress leads to enhanced glucose uptake via GLUT1. *Molecular cell, 49*(6), 1167-1175. doi:10.1016/j.molcel.2013.01.035

Xiao, J., Shi, Q., Li, W., Mu, X., Peng, J., Li, M., . . . Fan, J. (2018). ARRDC1 and ARRDC3 act as tumor suppressors in renal cell carcinoma by facilitating YAP1 degradation. *Am J Cancer Res, 8*(1), 132-143.

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics, 31*(14), 2382-2383. doi:10.1093/bioinformatics/btv145

Zbieralski, K., & Wawrzycka, D. (2022). alpha-Arrestins and Their Functions: From Yeast to Human Health. *Int J Mol Sci, 23*(9). doi:10.3390/ijms23094988

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. doi:10.1186/gb-2008-9-9-r137

Zhi, X., Zhao, D., Zhou, Z. M., Liu, R., & Chen, C. S. (2012). YAP Promotes Breast Cell Proliferation and Survival Partially through Stabilizing the KLF5 Transcription Factor. *American Journal of Pathology, 180*(6), 2452-2461. doi:10.1016/j.ajpath.2012.02.025

Zhou, R., Tardivel, A., Thorens, B., Choi, I., & Tschopp, J. (2010). Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nat Immunol, 11*(2), 136-140. doi:10.1038/ni.1831

Zuckerman, R., & Cheasty, J. E. (1986). A 48 kDa protein arrests cGMP phosphodiesterase activation in retinal rod disk membranes. *FEBS Lett, 207*(1), 35-41. doi:10.1016/0014-5793(86)80008-4

Figure legends

Figure 1. Identification of high-confidence α-arrestin PPIs

(A) Phylogenetic tree of α-arrestins from human (6, top) and *Drosophila* (12, bottom) based on protein sequences. The numbers in parentheses indicate the length of each protein. aa, amino acids; Arr_N: Arrestin N domain; Arr_C: Arrestin C domain; PPxY: PPxY motif. **(B)** Shown is a schematic flow of AP/MS experiments and computational analysis. **(C)** ROC curves of SAINTexpress scores along with mean AUC values. The arrows point to the cutoff scores used in subsequent studies in human (left) and *Drosophila* (right). **(D)** (Top) The fraction of "high-confid" (high-confidence) and raw (unfiltered) PPIs that are supported by known affinities between short linear motifs and protein domains in human (left) and *Drosophila* (right). One-sided, Fisher's exact test was performed to test the significance. (Bottom) The sum of $\log_2$ spectral counts ("$\log_2$ spec") of proteins with WW domains that were reported to interact with each α-arrestin in the high-confidence or raw PPI sets are depicted as heatmap. **(E)** The α-arrestins and their interactomes were hierarchically clustered based on the $\log_2$ mean spectral counts and summarized for human (top) and *Drosophila* (bottom) in the heatmaps. The functionally enriched protein class in the clustered interactomes are indicated on the top. Proteins that were reported to interact with α-arrestins in literatures and databases are selectively labeled on the bottom. On the right, the functional composition of the clustered α-arrestin interactomes are summarized as the sum of $\log_2$ mean spectral counts, which are colored to correspond with the labels on the left.

Figure 2. Network of α-arrestins and their associated protein complexes

Network of α-arrestins and the functional protein complexes that significantly interact with them in human **(A)** and *Drosophila* **(B)**. α-arrestins are colored yellow and prey proteins in protein complexes are colored according to the SAINTexpress scores (averaged if the protein interacts with multiple α-arrestins). The thickness of the green arrows indicates the strength of the interaction between α-arrestins and the indicated protein complexes, which was

estimated with FDR of complex association scores (see “Materials and Methods”). UB, ubiquitination; HDAC, histone deacetylase; COMPASS, complex proteins associated with Set1; SMN, survivor of motor neurons; TFIIIC, transcription factor III C; RNA polII, RNA polymerase II; MCM, minichromosome maintenance protein complex; SAC, spindle assembly checkpoint; NSL, non-specific lethal; WASH, Wiskott-Aldrich syndrome protein and scar homolog; Arp2/3, actin related protein 2/3; TEF, transcription elongation factor.

Figure 3. A substantial fraction of α-arrestin-PPIs are conserved across species

Human and *Drosophila* α-arrestins and their orthologous interactomes are hierarchically clustered based on $\log_2$-transformed mean spectral counts. They are then manually grouped based on their shared biological functions and assigned distinct colors. The names of orthologous proteins that interact with α-arrestins are displayed on the right side of the heatmap.

Figure 4. TXNIP knockdown induces a global decrease in chromatin accessibility and gene expression

(A-B) HeLa cells were treated with either siRNA against TXNIP (siTXNIP) or negative control (siCon) for 48 hours (hr) and analyzed of changes in the mRNA **(A)** and protein levels **(B)** of TXNIP. Gray dots depict actual values of each experiment and bar plots indicate mean ± standard deviation (sd). ***FDR < 0.001 (see “Materials and Methods”) for RNA-seq. *P < 0.05, *** P < 0.001 (two-sided paired Student‘s t-test) for RT-qPCR and western blots. **(A)** Expression levels of RNAs were quantified by RNA-seq (left, log2 counts per million mapped reads (CPM), see “Materials and Methods”) and RT-qPCR. **(B)** Protein levels were first visualized by western blot analysis of lysates from HeLa cells and band intensities of three independent experiments were quantified using image J software (right). **(C)** A schematic workflow for detecting dACRs and DEGs using ATAC- and RNA-seq analyses, respectively. **(D)** Volcano plots of differential chromatin accessibility for all ACRs (left) and those associated with promoters (right). **(E)** Volcano plots of differential gene expression. **(D-E)** Blue dots

denote “dACRs(-)” of significantly decreased chromatin accessibility **(D)** and “Down” genes of significantly down-regulated genes **(E)** in siTXNIP-treated cells compared to control (FDR $\leq$ 0.05, $\log_2$(siTXNIP / siCon) $\leq$ -1); red dots denote “dACRs(+)” of significantly increased chromatin accessibility **(D)** and “Up” genes of significantly up-regulated genes **(E)** in siTXNIP-treated cells compared to control (FDR $\leq$ 0.05, $\log_2$(siTXNIP / siCon) $\geq$ 1). Black dots denote data points with no significant changes. **(F)** Changes in chromatin accessibility of ACRs located in the promoter region of genes were plotted as CDFs. Genes were categorized into three groups based on changes in RNA levels (“Up”, “Down” as in **(E)** and “None” indicating genes with -0.5 $\leq$ $\log_2$(siTXNIP / siCon) $\leq$ 0.5). The number of genes in each group are shown in parentheses and *P* values in the left upper corner were calculated by one-sided KS test. **(G)** Top 10 GO terms (biological process and molecular function) enriched in genes that exhibited decreased chromatin accessibility at their promoter and decreased RNA expression upon TXNIP knockdown (Table S11).

Figure 5. TXNIP directly represses the recruitment of HDAC2 to target loci

(A) Co-IP assay showing the interaction between TXNIP and HDAC2 proteins. Lysates from HeLa cells that had been treated with either siCon or siTXNIP for 48 hr were subjected to IP and immunoblotting with antibodies recognizing TXNIP and HDAC2. IgG was used as the negative control. **(B)** Nuclear and cytoplasmic fractions of HeLa cells were analyzed with western blots following transfection with siCon or siTXNIP for 48 hr (left). Lamin B1 and GAPDH were used as nuclear and cytoplasmic markers, respectively. Western blot results from three independent experiments for TXNIP and HDAC2 were quantified as in Figure 4B. C, cytoplasm; N, nucleus. **(C)** Genomic regions showing RNA expression and chromatin accessibility at CD22 and L1CAM gene loci (top). Through the ChIP-qPCR analysis, the fold enrichment of HDAC2 and histone H3 acetylation (H3ac) at the CD22 and L1CAM promoter regions in HeLa cells treated with either siCon or siTXNIP for 48 hr were quantified (bottom). Data are presented as the mean $\pm$ sd (*n*=3, biological replicates). Gray dots depict actual

values of each experiment. *P < 0.05, **P < 0.01, ns: not significant (two-sided paired Student's t-test).

Figure 6. Interaction of ARRDC5 with the V-type ATPases in osteoclasts

(A) The human ARRDC5-centric PPI network. V-type and P-type ATPases, their related components, and extracellular exosomes are labeled and colored. Other interacting proteins are indicated with gray circles. **(B)** TRAP staining of osteoclasts. Cell differentiation was visualized with TRAP staining of GFP-GFP or GFP-ARRDC5 overexpressing osteoclasts (scale bar = 500 μm). TRAP-positive multinucleated cells (TRAP+ MNC) were quantified as the total number of cells and the number of cells whose diameters were greater than 200 μm. * P < 0.05. **(C)** Resorption pit formation on dentin slices. Cell activity was determined by measuring the level of resorption pit formation in GFP-GFP or GFP-ARRDC5 overexpressing osteoclasts (scale bar = 200 μm). Resorption pits were quantified as the percentage of resorbed bone area per the total dentin disc area using ImageJ software. The resorption area is relative to that in dentin discs seeded with GFP-GFP overexpressing osteoclasts, which was set to 100%. The colors of the bar plots are same as in **(B)**. ** P < 0.01. **(D)** Relative mRNA levels of ARRDC5 in non-target control (Control) or shARRDC5-expressing osteoclasts (shARRDC5) measured by qPCR. * P < 0.05 (Student's t-test, one-sided). **(E)** TRAP staining of osteoclasts. Cell differentiation was visualized with TRAP staining of "Control" or "shARRDC5" expressing osteoclasts (scale bar = 500 μm, left). TRAP-positive multinucleated cells (TRAP+ MNC) were quantified as the total number of cells (right). Colors of the bar plots are same as in **(D)**. *** P < 0.001 (Student's t-test). **(F)** The protein level of ATP6V1 in GFP-GFP or GFP-ARRDC5 overexpressing osteoclasts. The numbers represent independent samples for western blot analysis (left) and band intensities of three independent experiments were quantified (right). Colors of the bar plots are same as in **(B)** and **(C)**. ns, not significant. **(G)** Localization of ARRDC5 and V-type ATPase V1 domain subunit (ATP6V1) in osteoclasts. ATP6V1 was visualized with immunofluorescence (red), GFP-GFP and GFP-ARRDC5 were visualized with GFP fluorescence (green), and nuclei were visualized with DAPI (blue).

Representative fluorescence images are shown (scale bar = 100 μm). The region of interest, marked by the red boxes, was high-magnified and presented below. The integrated density of fluorescence was quantified using Image J software and expressed as relative fluorescence (right). The integrated density of fluorescence in GFP-GFP osteoclasts was established as the reference value, which was set to 1. * $P < 0.05$, ** $P < 0.01$ (Student's t-test).

Figure 1. Identification of high-confidence α-arrestin PPIs

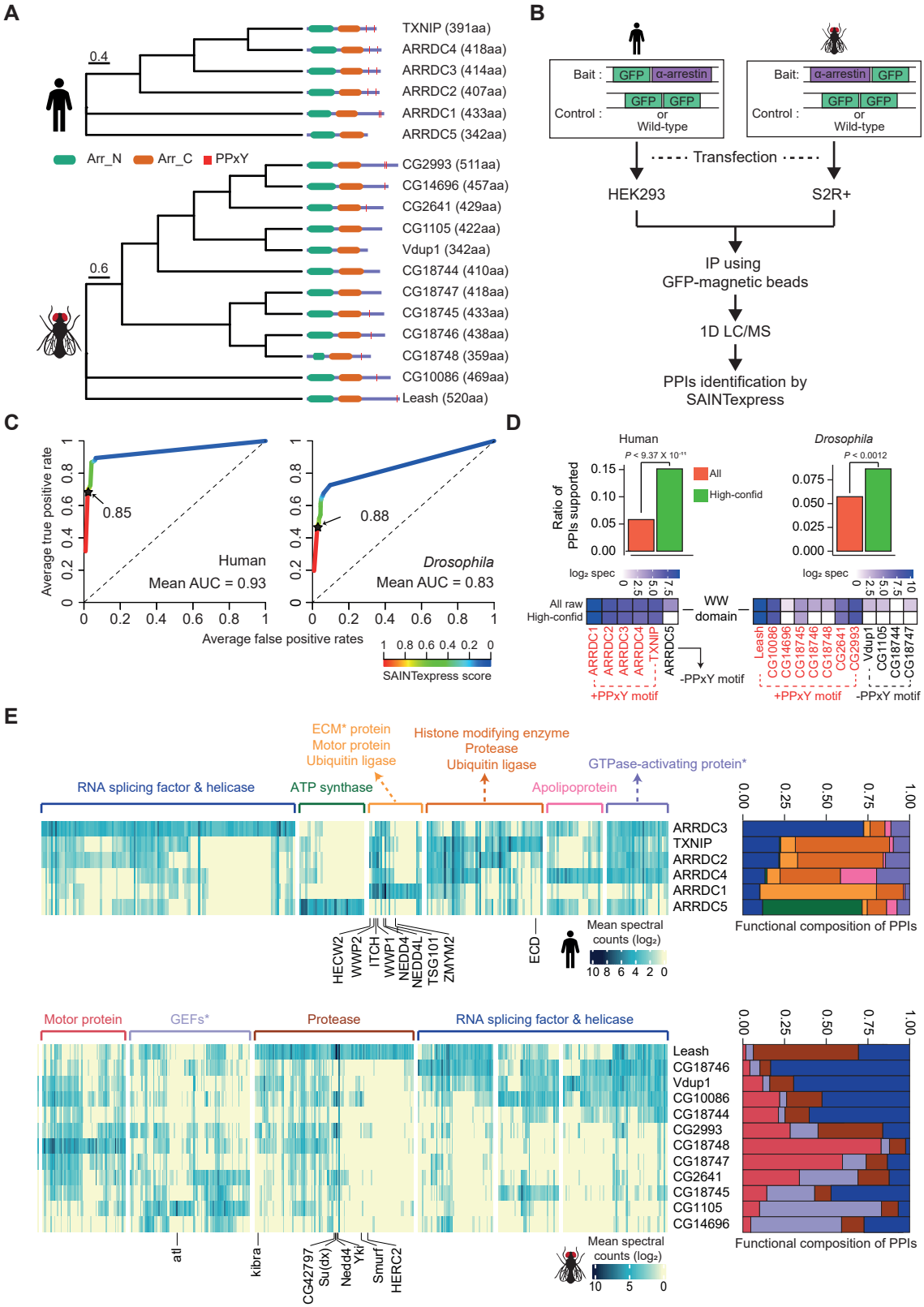
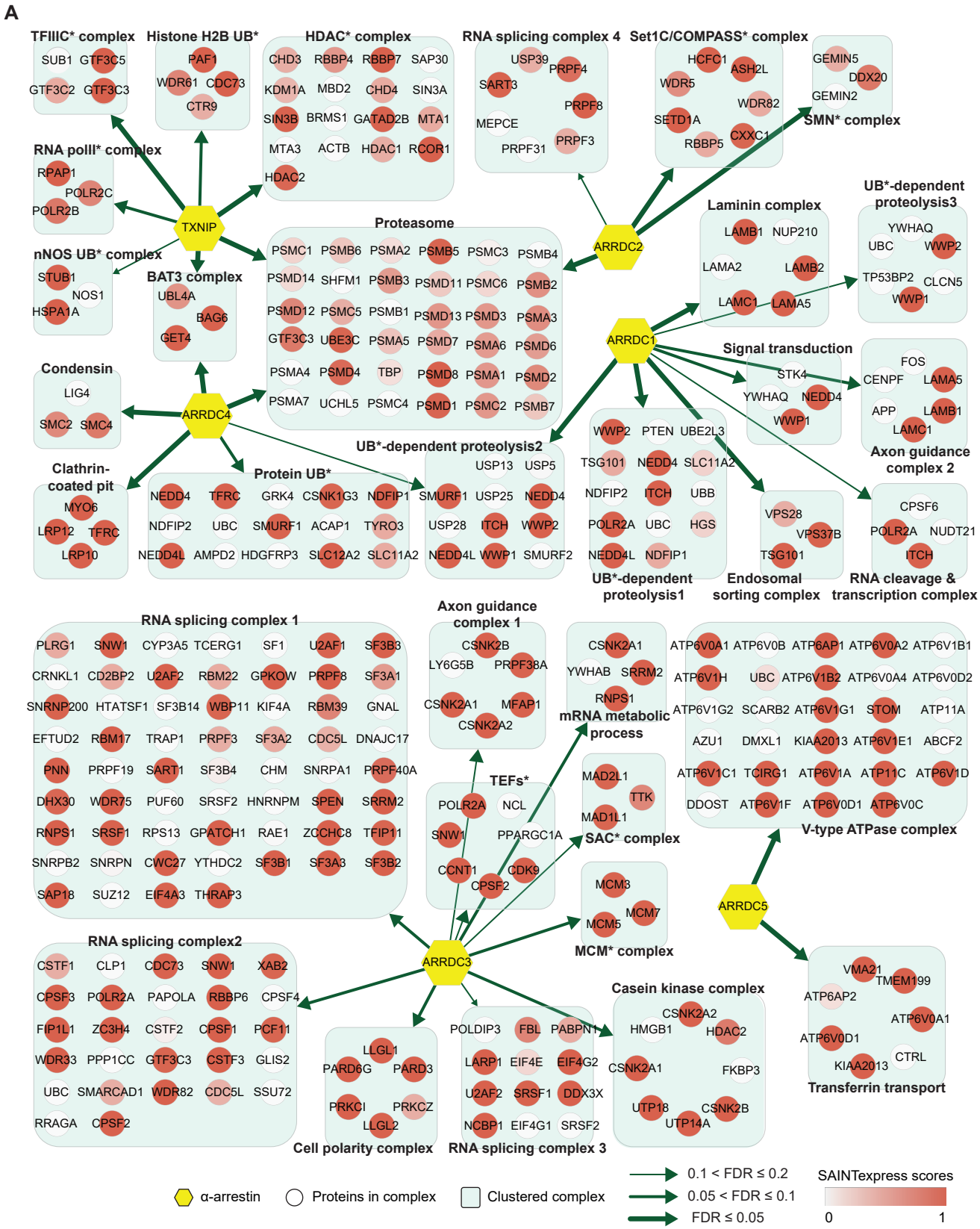


Figure 2. Network of α-arrestins and their interacting protein complexes



B

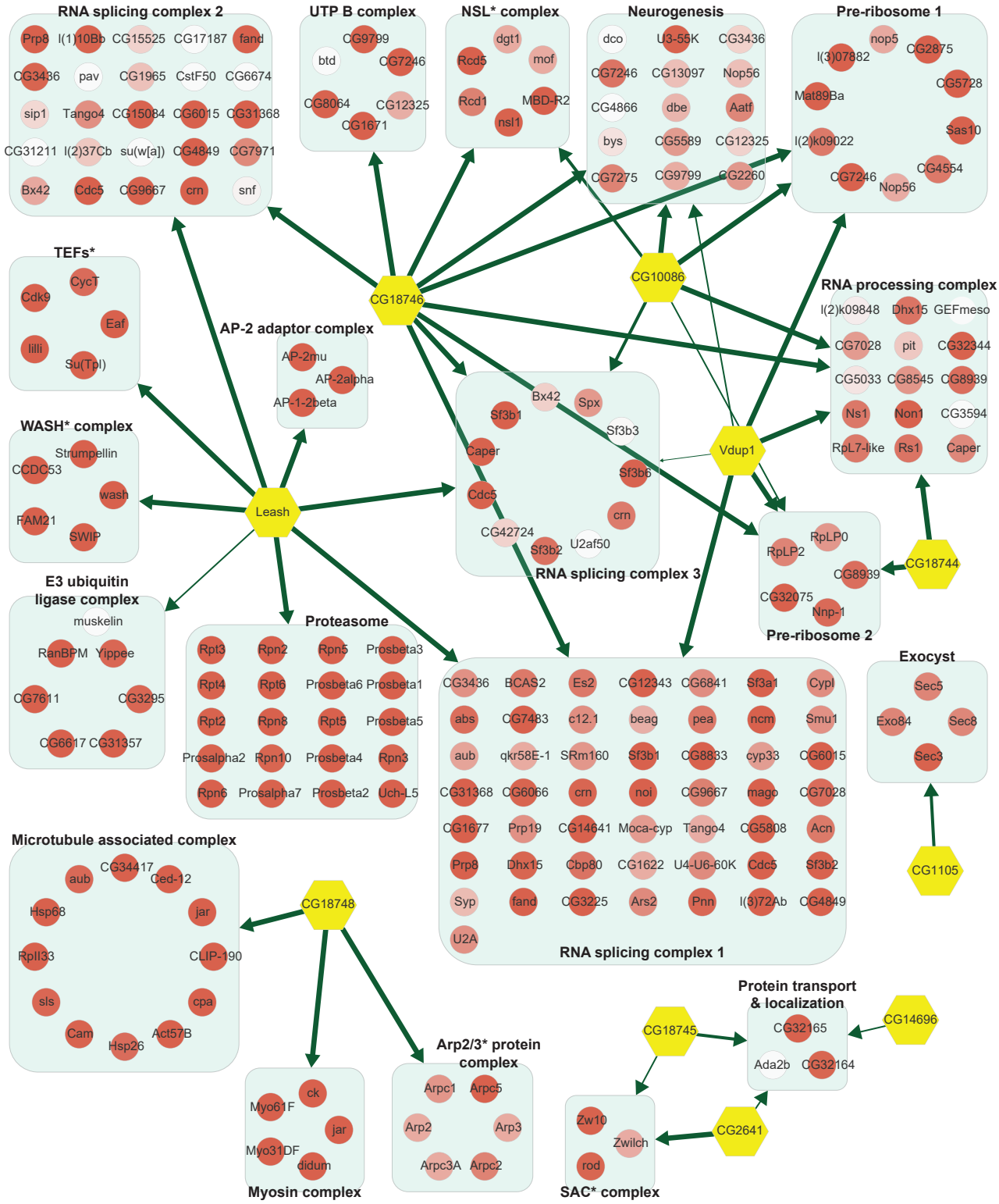

RNA splicing complex 2
Prp8
l(1)10Bb
CG15525
CG17187
fand
CG3436
pav
CG1965
CstF50
CG6674
sip1
Tango4
CG15084
CG6015
CG31368
CG31211
l(2)37Cb
su(w[a])
CG4849
CG7971
Bx42
Cdc5
CG9667
crn
snf
UTP B complex
CG9799
btd
CG7246
CG8064
CG12325
CG1671
NSL* complex
dgt1
Rcd5
mof
Rcd1
MBD-R2
nsl1
Neurogenesis
dco
U3-55K
CG3436
CG7246
CG13097
Nop56
CG4866
dbe
Aatf
bys
CG5589
CG12325
CG7275
CG9799
CG2260
Pre-ribosome 1
nop5
l(3)07882
CG2875
CG5728
Mat89Ba
Sas10
l(2)k09022
CG4554
CG7246
Nop56
TEFs*
CycT
Cdk9
Eaf
lilli
Su(Tpl)
CG18746
CG10086
RNA processing complex
l(2)k09848
Dhx15
GEFmeso
CG7028
pit
CG32344
CG5033
CG8545
CG8939
Ns1
Non1
CG3594
RpL7-like
Rs1
Caper
AP-2 adaptor complex
AP-2mu
AP-2alpha
AP-1-2beta
Bx42
Spx
Sf3b1
Sf3b3
Caper
Sf3b6
Cdc5
crn
CG42724
Sf3b2
U2af50
RNA splicing complex 3
Vdup1
WASH* complex
Strumpellin
CCDC53
wash
FAM21
SWIP
Leash
RpLP0
RpLP2
CG8939
CG32075
Nnp-1
Pre-ribosome 2
CG18744
E3 ubiquitin
ligase complex
muskelin
RanBPM
Yippee
CG7611
CG3295
CG6617
CG31357
Proteasome
Rpt3
Rpn2
Rpn5
Prosbeta3
Rpt4
Rpt6
Prosbeta6
Prosbeta1
Rpt2
Rpn8
Rpt5
Prosbeta5
Prosalpha2
Rpn10
Prosbeta4
Rpn3
Rpn6
Prosalpha7
Prosbeta2
Uch-L5
CG3436
BCAS2
Es2
CG12343
CG6841
Sf3a1
Cypl
abs
CG7483
c12.1
beag
pea
ncm
Smu1
aub
qkr58E-1
SRm160
Sf3b1
CG8883
cyp33
CG6015
CG31368
CG6066
crn
noi
CG9667
mago
CG7028
CG1677
Prp19
CG14641
Moca-cyp
Tango4
CG5808
Acn
Prp8
Dhx15
Cbp80
CG1622
U4-U6-60K
Cdc5
Sf3b2
Syp
fand
CG3225
Ars2
Pnn
l(3)72Ab
CG4849
U2A
RNA splicing complex 1
Exocyst
Sec5
Exo84
Sec8
Sec3
CG1105
Microtubule associated complex
CG34417
aub
Ced-12
Hsp68
jar
RpII33
CLIP-190
sls
cpa
Cam
Act57B
Hsp26
CG18748
Protein transport
& localization
CG18745
CG32165
CG14696
Ada2b
CG32164
Arp2/3* protein
complex
Arpc1
Arpc5
Arp2
Arp3
Arpc3A
Arpc2
ck
Myo61F
jar
Myo31DF
didum
Myosin complex
Zw10
Zwilch
rod
SAC* complex
CG2641

Figure 3. A substantial fraction of α-arrestin-PPIs are conserved across species

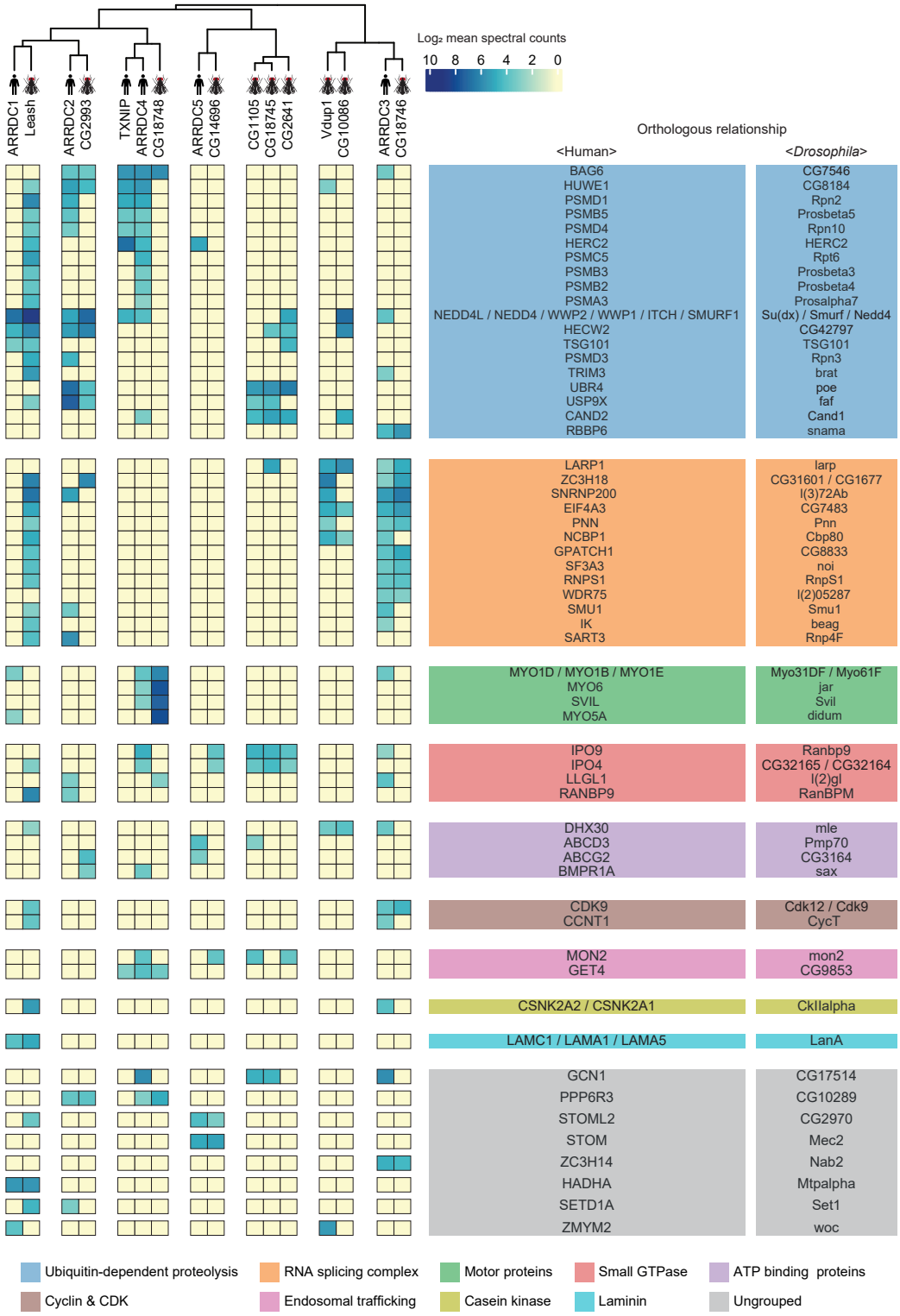


Figure 4. TXNIP knockdown induces a global decrease in chromatin accessibility and gene expression

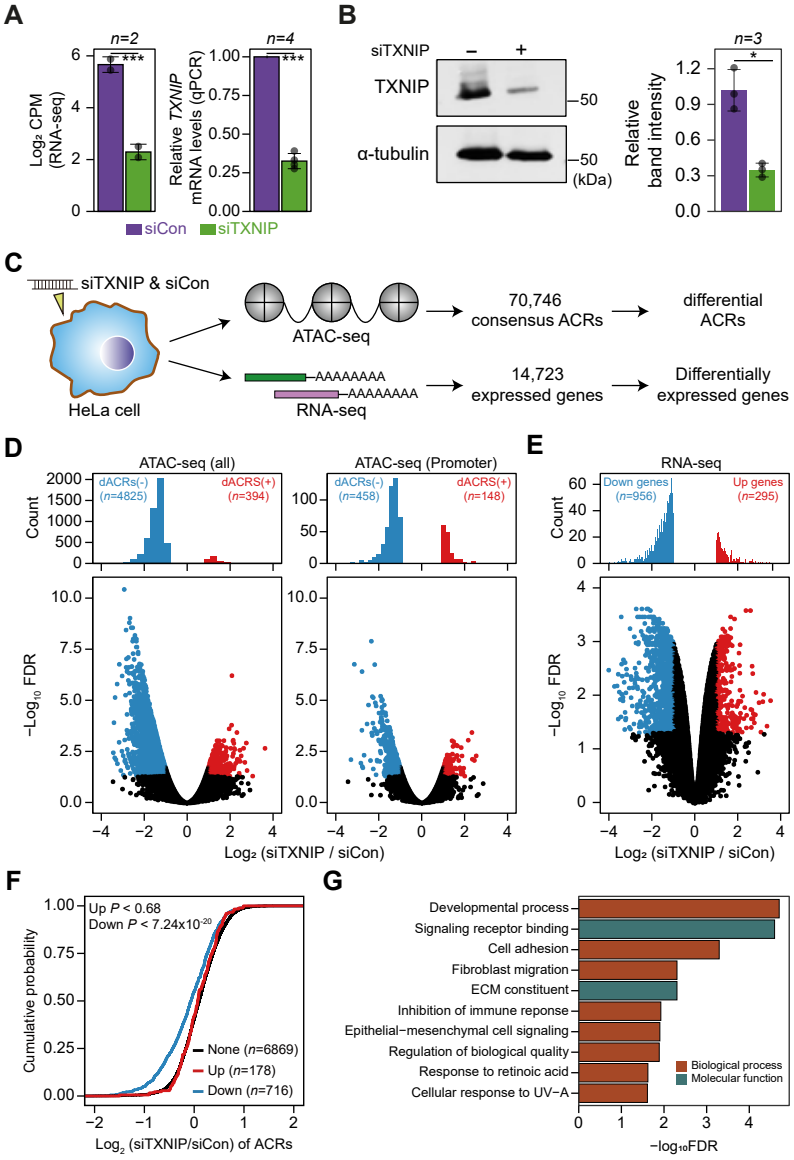


Figure 5. TXNIP directly represses the recruitment of HDAC2 to target loci

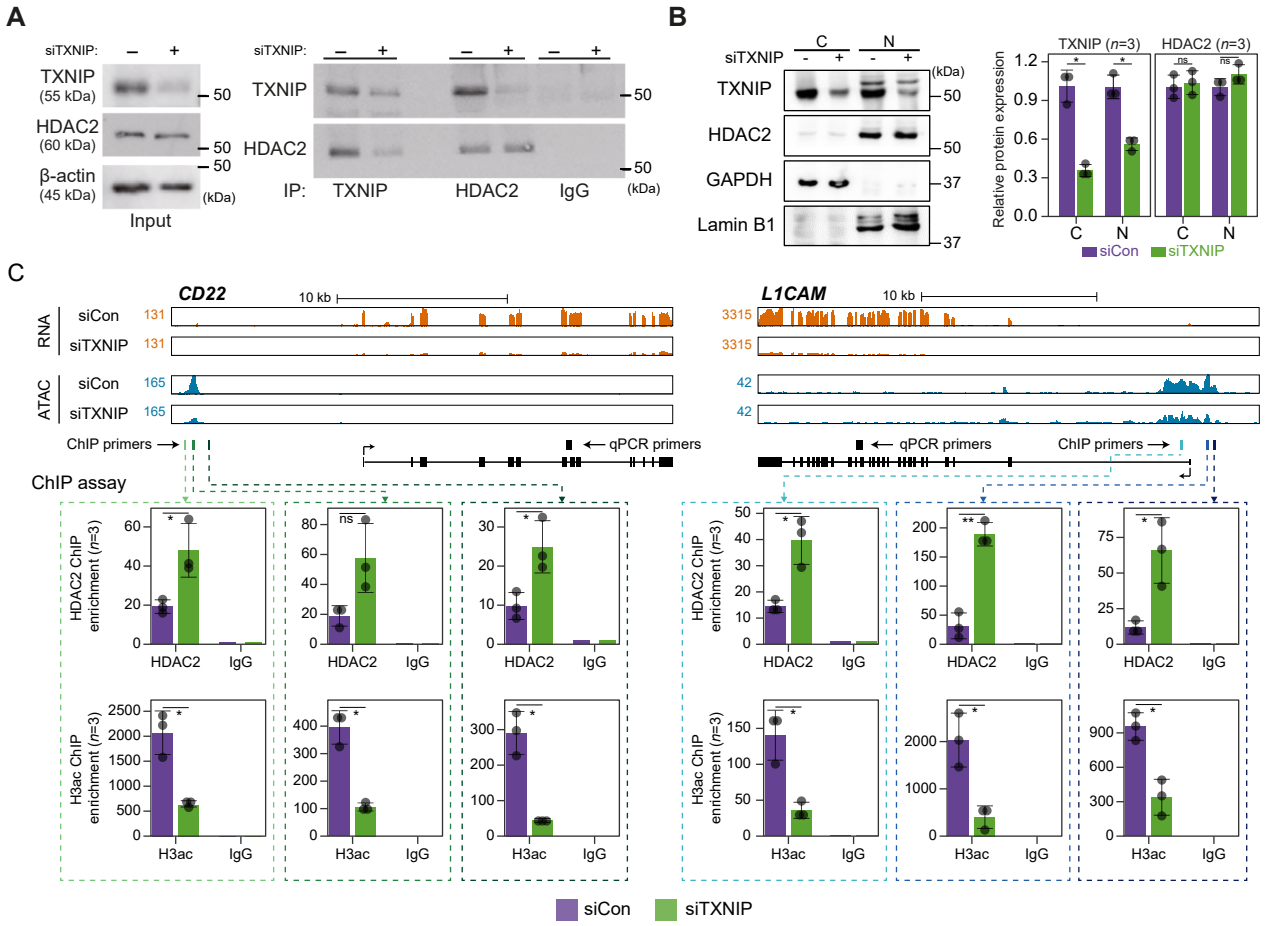
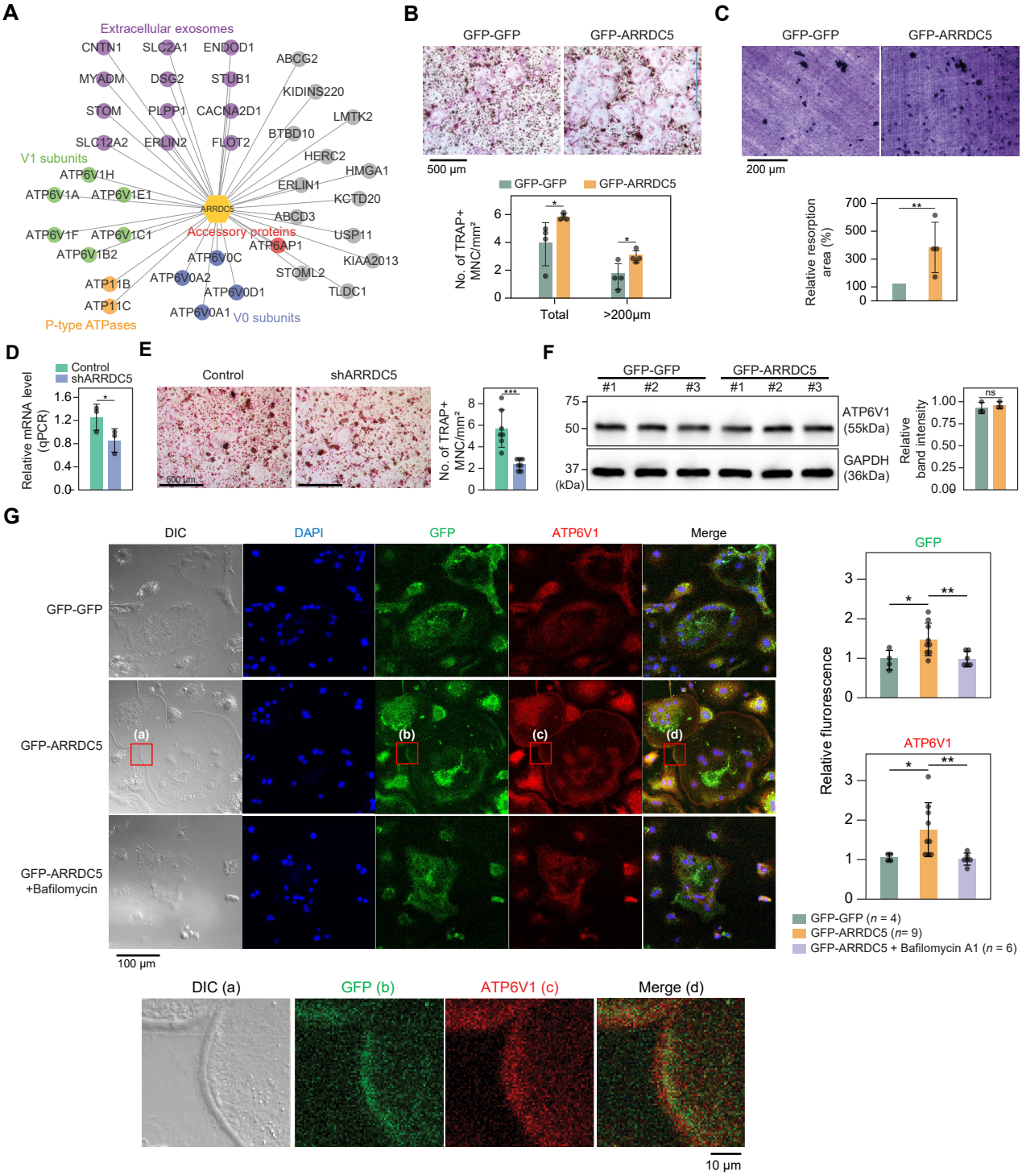



Figure 6. Interaction of ARRDC5 with the V-type ATPases in osteoclasts