# Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly

Bertolini, Matilde; Fenzl, Kai; Kats, Ilia; Wruck, Florian; Tippmann, Frank; Schmitt, Jaro; Auburger, Josef Johannes; Tans, Sander; Bukau, Bernd; Kramer, Günter

**DOI**
10.1126/science.abc7151

**Publication date**
2021

**Document Version**
Accepted author manuscript

**Published in**
Science

**Citation (APA)**
Bertolini, M., Fenzl, K., Kats, I., Wruck, F., Tippmann, F., Schmitt, J., Auburger, J. J., Tans, S., Bukau, B., & Kramer, G. (2021). Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. *Science*, *371*(6524), 57-64. Article eabc7151. https://doi.org/10.1126/science.abc7151

# Title: Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly

**Authors:** Matilde Bertolini[1,†], Kai Fenzl[1,†], Ilia Kats[1,2], Florian Wruck[3], Frank Tippmann[1], Jaro Schmitt[1], Josef Johannes Auburger[1], Sander Tans[3,4], Bernd Bukau[1,*] and Günter Kramer[1,*]

**Affiliations:**

[1]Center for Molecular Biology of Heidelberg University (ZMBH) and German Cancer Research Center (DKFZ), DKFZ-ZMBH Alliance, Im Neuenheimer Feld 282, Heidelberg D-69120, Germany

[2]Current address: Computational Genomics and System Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

[3]AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

[4]Bionanoscience Department of Delft University of Technology and Kavli Institute of Nanoscience Delft, 2629HZ Delft, The Netherlands

[†]These authors contributed equally to this work

[*]Corresponding authors

g.kramer@zmbh.uni-heidelberg.de

bukau@zmbh.uni-heidelberg.de

**Abstract:**

Faithful assembly of newly-synthesized proteins into functional oligomers is crucial for cell activity. Here, we asked whether direct interactions of two nascent proteins, emerging from nearby ribosomes (co-co assembly), is a general mechanism for oligomer formation. We used a proteome-wide screen to detect nascent chain-connected ribosome pairs and identified hundreds of homomer subunits that co-co assemble in human cells. Interactions were mediated by five major domain classes, among which N-terminal coiled coils were the most frequent. We were able to reconstitute co-co assembly of nuclear lamin in *E. coli*, demonstrating that dimer formation was independent of dedicated assembly machineries. Co-co assembly may thus constitute an efficient way to limit protein aggregation risks posed by diffusion-driven assembly routes and ensure isoform-specific homomer formation.

**One Sentence Summary:** Co-translational homomer assembly occurs by N-terminal dimerization of two nascent proteins and supports isoform-specificity.

Sophisticated mechanisms have evolved to ensure efficient and accurate protein complex biogenesis, including the fine-tuning of subunit expression to match complex stoichiometries (*1*), the employment of general or dedicated chaperones to guide oligomerization (*2–4*), the co-localization of subunit synthesis (*5–7*), and the timely oligomerization by coupling translation and subunit interactions (co-translational assembly) (*3*, *8*, *9*). Selective Ribosome Profiling (SeRP) has provided mechanistic details of co-translational assembly for *Vibrio harveyi* luciferase expressed in *E. coli* (*3*) and several heteromeric complexes in yeast (*8*). In all cases studied, a freely diffusing, presumably folded protein engages its nascent partner subunit (co-post assembly).

Here, we tested whether co-translational assembly of protein complexes may also occur via association of two nascent subunits concurrently translated by two ribosomes (co-co assembly). A priori, co-co assembly may involve nascent chains synthesized on two different mRNAs (in trans) or, for homo-oligomer assembly, on the same mRNA (in cis). Importantly, cis-assembly does not require that distinct mRNA molecules co-localize in the cytosol and enables transcript-specific homomeric complex generation, avoiding undesired interactions between closely related proteins or wildtype and mutant alleles (*10*). Although co-co assembly has already been proposed for individual protein complexes in different organisms (*10–14*), direct experimental evidence that two ribosome-nascent chain complexes interact is still missing, and we lack any information on the prevalence, molecular mechanisms and relevance of this proposed assembly process. Here, we developed an unbiased, proteome-wide screen based on ribosome profiling (*15*), termed Disome Selective Profiling (DiSP), to reveal the co-co assembly proteome in human cells.

## DiSP reveals widespread disome formation mediated by nascent chain interactions

To identify co-co assembling complexes across the proteome, we reasoned that ribosome pairs (disomes) connected by their exposed nascent chains will remain connected even upon mRNA digestion. Thus, it should be possible to detect co-co assembly candidates by RNase treatment of cell lysates, followed by separation of monosomes and disomes in sucrose gradients and deep-sequencing of 30 nt ribosomal footprints from both fractions (DiSP, Fig. 1A and S1A). The disome fraction will also contain RNase-resistant disomes that form upon collision of ribosomes translating the same mRNA; however, these disomes will protect double length (60 nt) mRNA fragments (*16*) and are not analyzed by DiSP. Translating ribosomes engaged in co-co assembly will shift from the monosome to the disome fraction upon nascent chain dimerization, which could be detected by analyzing the relative footprint density of both samples (separately or as enrichment of disome over monosome) along a gene's coding sequence (Fig. 1A). In contrast to SeRP, which has been used to explore co-post assembly of selected protein complexes (*3*, *8*), DiSP can provide proteome-wide interaction profiles of all translating ribosomes.

We initially performed DiSP of HEK293-T cells. To identify co-co assembly candidates, we first compared gene-specific footprint densities in the disome and monosome fractions, revealing over 1300 genes with a disome over monosome enrichment higher than two (Fig. 1B, top). A metagene profile of the averaged monosome and disome density along all coding sequences showed that early during translation, when nascent chains are short, ribosomes mostly migrated as monosomes, followed by a steady disome enrichment that leveled out at about 200 codons (Fig. 1B, bottom). The monosome to disome shift of translating ribosomes occurred only on a subset of genes, supporting the assumption that it depended on interaction properties of nascent chains (Fig. 1B, top and S1B). One example among the two-fold disome enriched genes is *DCTN1*, encoding $p150^{glued}$, a subunit of the dynactin motor complex. Ribosomes translating *DCTN1* convert from

monosomes to disomes near codon 430, when about 400 amino acids of nascent $p150^{glued}$ are exposed on the ribosomal surface. This N-terminal segment includes major parts of the coiled coil dimerization domain, suggesting the disome shift was caused by co-translational homodimerization (Fig. 1C, top). Repeating DiSP in U2OS cells, we found a large overlap of disome enriched genes and robustly correlated enrichment profiles (Fig. 1C and S1B, C), demonstrating that disome formation is a general feature of a specific subset of nascent proteins across different cell types.

To challenge our model that disome formation is mediated by nascent proteins, we explored whether disome shifts were sensitive to release or degradation of nascent chains. Treating lysates with Puromycin (Puro) or increasing concentrations of Proteinase K (PK) efficiently suppressed the shift of footprints from monosome to disome. This was apparent from a general reduction of the disome enrichment (Fig. 2A) and a flattening of enrichment profiles at both, the metagene level (Fig. 2B) and for individual genes (Fig. 2C and S1D-G). Thus, the stability of DiSP-detected disomes critically depends on the integrity of nascent chains, in agreement with the model of co-co assembly.

## A high confidence list of co-co assembly candidates enriched for homomers

We developed an unbiased bioinformatics selection regime to classify proteins based on their proficiency to co-co assemble. Accordingly, a protein qualified as high confidence candidate if all of the following criteria were fulfilled: (i) The gene's enrichment profile had a sigmoidal shape, indicating that with progressing translation, ribosomes shifted from the monosome to the disome fraction. If one of the interacting ribosomes terminates earlier, the other ribosome in the pair will shift back to the monosome fraction before it reaches the end of the coding sequence, resulting in a double sigmoidal shift (Fig. 3A). (ii) The enrichment profile becomes less sigmoidal upon treatment of the lysate with Puromycin and (iii) similarly with PK. (iv) The mature protein localizes to either the cytoplasm or the nucleus. We decided to categorize translocated protein candidates as low confidence because we cannot formally exclude the possibility that these ribosomes interact with membrane components of the translocation machinery and therefore migrate in the disome fraction. In addition, our validation experiments focused on cytosolic and nuclear candidates (Fig. 5 and S4) and poor structural annotation of membrane proteins complicates the downstream bioinformatics analysis. Out of a total of 15898 detected genes, 829 fulfilled all criteria and were classified as high confidence co-co assembly candidates (Table S1). A large number of genes (3301) fulfilled the important criterion (i) but not all of criteria (ii) to (iv) and were therefore categorized as low confidence candidates (Table S1). The low confidence list included 1404 proteins that are translocated across or inserted into organelle membranes, mainly the Endoplasmic Reticulum (ER), of which 443 fulfilled all other criteria. The latter fraction reflects the general frequency of ER-translocated proteins in the human proteome, indicating that co-co assembly may be an equally important mechanism to assemble cytosolic/nuclear and ER complexes, in agreement with previous experimental indications (*17–19*). The disome shift of ribosomes synthesizing membrane proteins frequently occurs after exposure of the first transmembrane domain (TMD) (Fig. S2A), which may suggest that co-co assembly involves interactions of two TMDs in the ER membrane.

Our next aim was to quantitatively assess what fraction of each high confidence candidate assembles co-translationally (from hereon named "efficiency" of co-co assembly). The efficiency was estimated by determining the reduction of footprints in the monosome fraction after initiation

of co-co assembly compared to the total translatome (including all translating ribosomes, determined by classical ribosome profiling (*15*, *20*)). Metagene analyses of footprint densities of all high-confidence genes aligned to the onset of assembly revealed a reduction of footprints in the monosome fraction from a DiSP experiment but not in the total translatome (Fig. 3B, top). This confirmed that the monosome depletion was caused by a shift of ribosomes to the disome fraction. The median monosome footprint reduction after the detected co-co assembly onset of high confidence genes was around 40%, and for some genes even exceeded 90%, indicating that in many cases the majority of nascent chains assembled co-translationally (Fig. 3B, bottom). Although to a smaller extent, monosome depletion was also observed for many low confidence candidates, suggesting that this list includes additional proteins that employ co-co assembly as a main route for complex formation (Fig. S2B, C). Importantly, the calculated depletion value most likely under-estimates the in vivo co-co assembly efficiency due to (i) the inevitable slight cross-contamination between the monosome and disome fractions and (ii) the possibility of a partial loss of disomes during sucrose gradient centrifugation that are connected by comparably weak nascent chain interactions. Supporting this notion, the three proteins featuring the highest efficiency (≥ 90% depletion, namely TPR, EEA1 and CLIP1) contained extremely long coiled coil homodimerization domains (between 1000 and 1500 amino acids, compared to a median coiled coil length of 66 amino acids in the cellular proteome) suggesting high stability.

We went on to analyze the features of proteins included in the high and low confidence lists. Consistently, annotated monomeric proteins were depleted in both lists of co-co assembly candidates, most strongly among the high confidence proteins (Fig. 3C, Table S2). Both classes showed a significant enrichment of homomers while heteromers were not significantly enriched. Furthermore, we often found only one subunit of a heterodimer in our candidate list, suggesting that this subunit rather formed a homo-oligomer or co-co assembled with a so far unknown partner subunit.

We used available crystal structures of protein complexes to determine the position of residues involved in subunit interaction at the onset of the disome shift. This analysis showed that the onset of assembly often coincided with the emergence of nascent chain segments that form the interfaces for the homo-oligomers (Fig. 3D, left). This correlation was not detected for heteromeric high confidence candidates (Fig. 3D, right). While these findings do not exclude the possibility that individual heteromers co-co assemble, as previously reported (*13*, *14*, *19*), they rather suggest that co-co assembly is predominantly employed for the formation of homomeric protein complexes.

## Co-co assembly is driven by exposure of conserved N-terminal homodimerization domains

Most detected co-co assembly interactions were established at early translation stages (Fig. S3A). Consistently, homodimerization interfaces are enriched in the N-terminal halves of high confidence candidates (Fig. S3B, left). This is different in the human proteome, where homodimerization interfaces are more often located in the C-terminal half of the protein, as previously reported (*21*) (Fig. S3B, right).

We next aimed to identify protein motifs or folds that mediate co-co assembly, by studying the enrichment of exposed domains at the onset of assembly. This analysis identified seven domain clusters mediating co-co assembly (color-coded in Fig. 4A), of which five are established homodimerization units.

Among our high confidence candidates, coiled coils were the most prevalent annotated domain class that is exposed on the ribosome surface at assembly onset (193 of 829 proteins according to

UniprotKB, Fig. 4B, left). Furthermore, the DeepCoil prediction tool (*22*) identified coiled coil segments on the exposed nascent chains in 408 genes (Fig. S3C), suggesting that up to 50% of high confidence candidates employ this fold for co-co assembly. In many cases, the coiled coil is only partially exposed at assembly onset (Fig. 4B, left). The number of exposed residues involved in coiled coil formation varied (median of 111 residues in the high confidence class, Fig. S3D), which may suggest that different lengths of the coiled coil are needed to form a stable dimer.

We found seven additional domains that are generally positioned N-terminally to coiled coil domains in myosins, kinesins and AGC-kinases (orange in Fig. 4A), and were therefore exposed at the onset of co-co assembly. However, disome enrichment generally required the partial or complete exposure of the coiled coil segment, suggesting that these domains do not contribute to oligomerization.

A second domain class that was often only partially exposed at the onset of assembly are BAR domains (named after Bin, Amphiphysin and Rvs, Fig. 4B, right); conserved dimerization domains found in many proteins mediating membrane curvature. They consist of three (classical BAR) to five (F-BAR) bent antiparallel alpha-helices. According to our dataset, co-co assembly generally required the exposure of the most N-terminal alpha-helix (helix1, Fig. 4B, right), that interacts with its partner helix1' in an antiparallel fashion.

All other enriched domain classes were globular and fully exposed at assembly onset, implying that their co-translational folding was required for assembly, including BTB (Broad-Complex, Tramtrack and Bric a brac), RHD (Rel Homology Domain) and SCAN (SRE-ZBP, CTfin51, AW-1 and Number 18 cDNA) domains (Fig. 4C). BTBs are highly conserved globular dimerization domains located at the N-termini of many transcription factors, ion channels and E3 ligase subunits, and found in 36 of our high confidence candidates (Fig. 4C, left). The less abundant RHDs are found at the N-terminus of proteins involved in nuclear factor kappa-B (NF-kB) complex formation and create the interface of homo- and heteromeric interactions. According to our DiSP, all NF-kB homologs co-co assemble, confirming earlier indications that proteins encoded by *NFKB1* may co-translationally assemble in cis and that early assembly is required for native biogenesis of the p50 transcription factor (*12*, *23*) (Fig. 4C, middle, Fig. S1B, right). This very likely also holds true for the *RELB* encoded homolog, however, because *RELB* is low expressed in HEK293-T cells, we cannot make a definite statement.

The high confidence list also included 12 transcription factors that employ SCAN domains for co-co assembly (Fig. 4C, right). SCAN domains are leucine-rich, N-terminal motifs composed of five packed alpha-helices that mediate homo- and hetero-oligomerization of a large family of C2H2 zinc finger proteins by intercalating helix 2 of one monomer between helices 3 and 5 of the opposing monomer.

Comparing the co-co assembly efficiency of these five major dimerization domains, we found that coiled coils conferred the highest, yet very variable stability to the nascent chain interactions, followed by BTB, BAR, RHD and SCAN domains (Fig. S3E).

Finally, our dataset included two less characterized domains that were significantly enriched (Fig. 4A). The first were STI1 repeats of ubiquilin proteins. This domain mediates homo- and hetero-dimerization of ubiquilin 1 and 2 (*24*), which both were high confidence candidates that fully exposed the second STI1 repeat (STI1 2) at the assembly onset (Fig. S3F).

The second are GBD/FH3; conserved N-terminal regulatory elements in Diaphanous-related formins, a protein class involved in nucleation and remodeling of the actin cytoskeleton. The FH3 domain has been implicated in dimerization of the mouse homologue of human *DIAPH1* (*25*). We found six human formins among our high confidence proteins and in all cases the FH3 domain

was exposed at assembly onset, suggesting that formins may co-translationally assemble via the FH3 domain (Fig. S3G).

## Co-co assembly is independent of eukaryotic assembly factors

We next examined whether ribosome exposure of co-co assembly-competent nascent chains suffices for disome formation, and if it could occur outside the eukaryotic folding environment. To investigate this question, we performed DiSP of *E. coli* synthesizing human lamin C (*LMNA*), one of the mammalian intermediate filaments that were all high confidence candidates of our DiSP screen. Lamins form homodimers in the cytosol and assemble into higher-order polymers in the nucleus. Dimerization involves the N-terminal rod domain, a long discontinuous coiled coil that includes three segments named coil 1A, 1B, and 2AB. *LMNA* overexpression generated a disome peak in the RNase-digested lysate (Fig. 5A). DiSP revealed that these disomes were enriched with ribosomes translating *LMNA* (Fig. 5B), indicating nascent lamin C can co-translationally dimerize in bacteria. The minimal length of nascent lamin C mediating the disome shift in *E. coli* was close to that of the endogenously expressed lamin C in mammalian cells (Fig. 5B). Likewise, overexpression of *DCTN1* generated a disome peak that was enriched with ribosomes exposing the coiled coil of $p150^{glued}$, and the assembly onset was similar to human cells (Fig. S4A). This indicates that co-co assembly of coiled coils is independent of eukaryote-specific assembly factors or mRNA subcellular localization.

To test our hypothesis that the formation of a coiled coil between two nascent chains is minimally required and sufficient to induce disome shifts in bacteria, we used coil 1B of lamin C as a paradigm. First, we employed an established in vivo dimerization assay based on a lambda repressor fusion system (*26*) to show that the isolated 1B efficiently dimerized in *E. coli* (Fig. S4B). Second, we performed DiSP to verify that nascent 1B, N-terminally fused to mCherry, efficiently mediated co-co assembly (Fig. 5C, left). Third, we perturbed the periodicity of nonpolar and charged amino acids required for coiled coil formation of 1B by swapping the position ‘a’ and ‘e’ of the coiled coil heptameric repeats (1B*; Fig. 5C, middle). These swaps do not change the overall amino acid composition, nor the hydrophobicity, or the predicted propensity to form alpha helices, but eliminated the proficiency of 1B to form a coiled coil (Fig. 5C, insets). In contrast to 1B, the mutated 1B* did not confer co-translational disome formation in *E. coli* (Fig. 5C, right), further supporting that DiSP detects productive, in vivo interactions between nascent chains that drive protein oligomer formation.

## Co-co assembly in cis may ensure isoform-specific coiled coil formation

Lamin A and C are isoforms encoded by the same gene but translated on two alternatively spliced transcripts. Although they share the same N-terminal rod dimerization domain, lamin A and C exclusively form homodimers in vivo (*27*). How this isoform-specificity is achieved in the cellular environment is not known. Co-co assembly may provide a simple answer to this conundrum: isoform-specific assembly may be achieved by co-co assembly in trans on co-localized mRNAs of the same kind (which might segregate in the cytosol due to their unique 3’UTRs), or - even simpler - in cis, facilitated by interaction of nascent proteins synthesized by neighboring ribosomes organized in a polysome (Fig. 5D, left).

To discriminate between these possibilities, we generated a heterozygous HEK293-T cell line, in which one *LMNA* allele encodes a C-terminally TwinStrep-tagged lamin C. We performed a series

of affinity purification experiments which revealed that tagged lamin C never co-purified the untagged counterpart, even though both proteins derive from identically spliced mRNAs with identical UTRs (Fig. 5D, right). This result supports the model that co-co assembly in cis facilitates isoform-specific lamin dimerization in human cells.

## Discussion

We provide a comprehensive analysis of co-translational protein complex assembly mediated by two nascent subunits. The ribosome profiling-based approach developed here (DiSP) allowed us to identify hundreds of high confidence and thousands of low confidence candidates in human cells, revealing co-co assembly as a major route to complex formation.

We decided to include all translocated proteins into the low confidence list. Many of them are membrane proteins that are often partially or fully resistant to PK but sensitive to Puromycin, in particular small proteins (up to 35 kDa) with multiple annotated TMDs. PK resistance may be conferred by ribosome docking to the translocon that limits the access of PK to the nascent protein. We speculate that docking of ribosomes that closely follow each other in a polysome may spatially organize translocons in the membrane and facilitate homomer assembly.

Our data show that predominantly homodimers co-co assemble. We did not find clear evidence that heteromers co-co assemble in trans, because our high confidence list in most cases only contained one of the subunits of an annotated heteromer. The absence of a known partner subunit may be caused by the less complete structural characterization of heteromeric complexes.

We also did not find clear evidence that the recently described assembly of TAF6-TAF9 nuclear complex includes nascent chain interactions (*14*). Both subunits are included in the low confidence list, but the length of the disome shift and the enrichment efficiency is very different between the two proteins, which does not agree with a model of co-co assembly in trans.

Co-co assembly of homomers in cis may be facilitated by a generally high ribosome occupancy to ensure close proximity of the interacting nascent chains. In addition, both heteromer assembly (in trans) and homomer assembly (in cis or in trans) may benefit from the slowdown of ribosomes at the onset of assembly, to allow the trailing ribosome translating the same mRNA to catch up or to provide an extended timeframe to establish the interaction with another nascent chain translated on a distinct mRNA (*13*).

We discovered two different types of nascent chain dimerization. First, a zipper-like formation of coiled coils and BAR domains. In these cases, the interaction strength may gradually increase as both nascent chains grow, until enough residues involved in dimerization are ribosome-exposed to drive the co-co assembly of stable dimers.

The second type of nascent chain dimerization may require the prior folding of a fully emerged, globular interaction domain, i.e. BTB, RHD and SCAN domains, a feature already reported for co-post assembly (*3*, *8*).

Homodimerization contact regions are evolutionarily selected to be enriched in C-terminal halves of proteins, supposedly to ensure that folding occurs undisturbed by the vicinity of another identical, incompletely folded subunit (*28*). Our analysis supports this C-terminal enrichment for the majority of the human proteome, except for the proteins enclosed in our high confidence list. For the latter proteins, the selective pressure to assemble early apparently overrules the penalty that is inferred by enhanced folding problems of yet to be synthesized C-terminal domains. We speculate that productive folding of the native dimer, beyond co-co assembly, is likely supported by extensive, finely tuned intervention of molecular chaperones.

There are multiple reasons that may create selective pressure against diffusion-driven assembly and favoring co-co assembly. (i) Co-co assembly may increase the efficiency and rate of complex formation. This advantage is most evident for the cis assembly mode where dimerizing nascent chains are already adjacent within polysomes. (ii) Synthesis-coupled assembly may suppress unproductive interactions and facilitate native folding, by limiting the exposure of aggregation-prone dimerization interfaces to the crowded cellular environment. (iii) Cis assembly creates mRNA-specific homomers. Coiled coils and BTB domains are recurrent dimerization modules in the human proteome, bearing high potential for promiscuous, potentially deleterious heteromeric interactions (*29*, *30*). Such interactions would be efficiently prevented by in cis assembly, including the mix of splicing-derived isoforms that share identical dimerization domains, as in the case of human lamin A and C (*27*, *29*). Misassembled subunits that failed to co-co assemble may be recognized by a recently described pathway, that specifically detects and eliminates complexes of aberrant composition (DQC – Dimerization Quality Control (*31*)). Interestingly, DQC has been reported as a surveillance mechanism for BTB complexes, but a similar molecular machinery may exist that monitors the composition of other complexes, including coiled coils. Here, our proteome-wide study reveals that co-translational interactions between nascent subunits are a general and efficient strategy to guide the isoform-specific formation of protein complexes.

**Materials and Methods:**

Detailed materials and methods can be found in the supplementary materials.

Human osteosarcoma U2OS (ATCC Cat# HTB-96), human embryonal kidney HEK293-T (DSMZ Cat# ACC 635) and E. coli Rosetta cells (Novagene) were employed for DiSP experiments.

All ribosome profiling libraries were prepared as described (*20*) and sequenced on a NextSeq550 (Illumina) according to the manufacturer's protocol, except for libraries of U2OS samples, which were prepared as described (*8*) and sequenced on a HiSeq 2000 (Illumina).

DiSP with PK treatment included incubation of the cell lysates for 30 minutes at 4°C with following PK to total protein amounts: (i) Low PK = 1:20000; (ii) Mid PK = 1:6000; (iii) High PK = 1:2000; (iv) Very High PK = 1:200.

DiSP with Puromycin omitted cycloheximide from all buffers; cell lysates were incubated for 25 minutes with 2 mM Puromycin and crosslinked with 0.5% formaldehyde.

## References and Notes:


1. G.-W. Li, D. Burkhardt, C. Gross, J. S. Weissman, Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell*. **157**, 624–635 (2014).

2. G. Tian, S. A. Lewis, B. Feierbach, T. Stearns, H. Rommelaere, C. Ampe, N. J. Cowan, Tubulin Subunits Exist in an Activated Conformational State Generated and Maintained by Protein Cofactors. *J. Cell Biol.* **138**, 821–832 (1997).

3. Y.-W. Shieh, P. Minguez, P. Bork, J. J. Auburger, D. L. Guilbride, G. Kramer, B. Bukau, Operon structure and cotranslational subunit association direct protein assembly in bacteria | Science. *Science*. **350**, 678–680 (2015).

4. A. Rousseau, A. Bertolotti, Regulation of proteasome assembly and activity in health and disease. *Nat. Rev. Mol. Cell Biol.* **19**, 697–712 (2018).

5. L. A. Mingle, Localization of all seven messenger RNAs for the actin-polymerization nucleator Arp2/3 complex in the protrusions of fibroblasts. *J. Cell Sci.* **118**, 2425–2433 (2005).

6. M. Pizzinga, C. Bates, J. Lui, G. Forte, F. Morales-Polanco, E. Linney, B. Knotkova, B. Wilson, C. A. Solari, L. E. Berchowitz, P. Portela, M. P. Ashe, Translation factor mRNA granules direct protein synthetic capacity to regions of polarized growth. *J. Cell Biol.* **218**, 1564–1581 (2019).

7. B. Hampoelz, A. Schwarz, P. Ronchi, H. Bragulat-Teixidor, C. Tischer, I. Gaspar, A. Ephrussi, Y. Schwab, M. Beck, Nuclear Pores Assemble from Nucleoporin Condensates During Oogenesis. *Cell*. **179**, 671-686.e17 (2019).

8. A. Shiber, K. Döring, U. Friedrich, K. Klann, D. Merker, M. Zedan, F. Tippmann, G. Kramer, B. Bukau, Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling. *Nature*. **561**, 268–272 (2018).

9. G. Kramer, A. Shiber, B. Bukau, Mechanisms of Cotranslational Maturation of Newly Synthesized Proteins. *Annu. Rev. Biochem.* **88**, 337–364 (2019).

10. C. D. Nicholls, K. G. McLure, M. A. Shields, P. W. K. Lee, Biogenesis of p53 Involves Cotranslational Dimerization of Monomers and Posttranslational Dimerization of Dimers: IMPLICATIONS ON THE DOMINANT NEGATIVE EFFECT. *J. Biol. Chem.* **277**, 12937–12945 (2002).

11. R. Gilmore, M. C. Coffey, G. Leone, K. McLure, P. W. Lee, Co-translational trimerization of the reovirus cell attachment protein. *EMBO J.* **15**, 2651–2658 (1996).

12. L. Lin, G. N. DeMartino, W. C. Greene, Cotranslational dimerization of the Rel homology domain of NF-κB1 generates p50–p105 heterodimers and is required for effective p50 production. *EMBO J.* **19**, 4712–4722 (2000).

13. O. O. Panasenko, S. P. Somasekharan, Z. Villanyi, M. Zagatti, F. Bezrukov, R. Rashpa, J. Cornut, J. Iqbal, M. Longis, S. H. Carl, C. Peña, V. G. Panse, M. A. Collart, Co-translational assembly of proteasome subunits in NOT1-containing assemblysomes. *Nat. Struct. Mol. Biol.* **26**, 110–120 (2019).

14. I. Kamenova, P. Mukherjee, S. Conic, F. Mueller, F. El-Saafin, P. Bardot, J.-M. Garnier, D. Dembele, S. Capponi, H. T. M. Timmers, S. D. Vincent, L. Tora, Co-translational assembly of mammalian nuclear multisubunit complexes. *Nat. Commun.* **10**, 1–15 (2019).

15. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (80-. ).* **324**, 218–223 (2009).

16. P. Han, Y. Shichino, T. Schneider-Poetsch, M. Mito, S. Hashimoto, T. Udagawa, K. Kohno, M. Yoshida, Y. Mishima, T. Inada, S. Iwasaki, Genome-wide Survey of Ribosome Collision. *Cell Rep.* **31**, 107610 (2020).

17. S. D. Redick, J. E. Schwarzbauer, Rapid intracellular assembly of tenascin hexabrachions suggests a novel cotranslational process. *J. Cell Sci.* **108**, 1761–1769 (1995).

18. J. Lu, J. M. Robinson, D. Edwards, C. Deutsch, T1−T1 Interactions Occur in ER Membranes while Nascent Kv Peptides Are Still Attached to Ribosomes †. *Biochemistry*. **40**, 10934–10946 (2001).

19. F. Liu, D. K. Jones, W. J. de Lange, G. A. Robertson, Cotranslational association of mRNA encoding subunits of heteromeric ion channels. *Proc. Natl. Acad. Sci.* **113**, 4859–4864 (2016).

20. N. J. MGlincy, N. T. Ingolia, Transcriptome-wide measurement of translation by ribosome profiling. *Methods*. **126**, 112–129 (2017).

21. E. Natan, T. Endoh, L. Haim-Vilmovsky, T. Flock, G. Chalancon, J. T. S. Hopper, B. Kintses, P. Horvath, L. Daruka, G. Fekete, C. Pál, B. Papp, E. Oszi, Z. Magyar, J. A. Marsh, A. H. Elcock, M. M. Babu, C. V Robinson, N. Sugimoto, S. A. Teichmann, Cotranslational protein assembly imposes evolutionary constraints on homomeric proteins. *Nat. Struct. Mol. Biol.* **25**, 279–288 (2018).

22. J. Ludwiczak, A. Winski, K. Szczepaniak, V. Alva, S. Dunin-Horkawicz, DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. *Bioinformatics*. **35**, 2790–2795 (2019).

23. L. Lin, G. N. DeMartino, W. C. Greene, Cotranslational biogenesis of NF-κB p50 by the 26S proteasome. *Cell*. **92**, 819–828 (1998).

24. D. L. Ford, M. J. Monteiro, Dimerization of ubiquilin is dependent upon the central region of the protein: evidence that the monomer, but not the dimer, is involved in binding presenilins. *Biochem. J.* **399**, 397–404 (2006).

25. R. Rose, M. Weyand, M. Lammers, T. Ishizaki, M. R. Ahmadian, A. Wittinghofer, Structural and mechanistic insights into the interaction between Rho and mammalian Dia. *Nature*. **435**, 513–518 (2005).

26. J. C. Hu, E. K. O&amp, apos, Shea, P. S. Kim, R. T. Sauer, E. K. O'Shea, P. S. Kim, R. T. Sauer, Sequence requirements for coiled-coils: analysis with lambda repressor-GCN4 leucine zipper fusions. *Science*. **250**, 1400–1404 (1990).

27. T. Kolb, K. Maass, M. Hergt, U. Aebi, H. Herrmann, Lamin A and lamin C form homodimers and coexist in higher complex forms both in the nucleoplasmic fraction and in the lamina of cultured human cells. *Nucleus*. **2**, 425–433 (2011).

28. E. Natan, T. Endoh, L. Haim-Vilmovsky, T. Flock, G. Chalancon, J. T. S. Hopper, B. Kintses, P. Horvath, L. Daruka, G. Fekete, C. Pál, B. Papp, E. Őszi, Z. Magyar, J. A. Marsh, A. H. Elcock, M. M. Babu, C. V Robinson, N. Sugimoto, S. A. Teichmann, E. Oszi, Z. Magyar, J. A. Marsh, A. H. Elcock, M. M. Babu, C. V Robinson, N. Sugimoto, S. A. Teichmann, Cotranslational protein assembly imposes evolutionary constraints on homomeric proteins. *Nat. Struct. Mol. Biol.* **25**, 279–288 (2018).

29. Q. Ye, H. J. Worman, Protein-protein interactions between human nuclear lamins expressed in yeast. *Exp. Cell Res.* **219**, 292–298 (1995).

30. G. Schreiber, A. E. Keating, Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).

31. E. L. Mena, R. A. S. Kjolby, R. A. Saxton, A. Werner, B. G. Lew, J. M. Boyle, R. Harland, M. Rape, Dimerization quality control ensures neuronal development and survival. *Science (80-. ).* **362**, eaap8236 (2018).

32. ilia-kats, ilia-kats/RiboSeqTools: v0.1 (2020), doi:10.5281/ZENODO.4016066.

33. Materials and methods are available as supplementary materials at the Science website.

34. M. D. Young, M. J. Wakefield, G. K. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).

35. A. Yeliseev, L. Zoubak, K. Gawrisch, Use of dual affinity tags for expression and purification of functional peripheral cannabinoid receptor. *Protein Expr. Purif.* **53**, 153–163 (2007).

36. A. Paix, A. Folkmann, D. H. Goldman, H. Kulaga, M. J. Grzelak, D. Rasoloson, S. Paidemarry, R. Green, R. R. Reed, G. Seydoux, Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci.* **114**, E10745–E10754 (2017).

37. G. V Shivashankar, in *Nuclear Mechanics and Genome Regulation* (Academic Press, 2010), vol. 98, pp. 111–112.

38. G. Blobel, D. Sabatini, Dissociation of Mammalian Polyribosomes into Subunits by Puromycin. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 390–394 (1971).

39. C. V Galmozzi, D. Merker, U. A. Friedrich, K. Döring, G. Kramer, Selective ribosome profiling to study interactions of translating ribosomes in yeast. *Nat. Protoc.* **14**, 2279–2317 (2019).

40. K. Döring, N. Ahmed, T. Riemer, H. G. Suresh, Y. Vainshtein, M. Habich, J. Riemer, M. P. Mayer, E. P. O'Brien, G. Kramer, B. Bukau, Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. *Cell*. **170**, 298-311.e20 (2017).

41. D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*. **6**, 343–345 (2009).

42. A. H. Becker, E. Oh, J. S. Weissman, G. Kramer, B. Bukau, Selective ribosome profiling as a tool for studying the interaction of chaperones and targeting factors with nascent polypeptide chains and ribosomes. *Nat. Protoc.* **8**, 2212–2239 (2013).

43. J. Schaefer, G. Jovanovic, I. Kotta-Loizou, M. Buck, Single-step method for β-galactosidase assays in Escherichia coli using a 96-well microplate reader. *Anal. Biochem.* **503**, 56–57 (2016).

44. U. Fiedler, V. Weiss, A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. *EMBO J.* **14**, 3696–3705 (1995).

45. J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, Julia: A Fresh Approach to Numerical Computing. *SIAM Rev.* **59**, 65–98 (2017).

46. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

47. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

48. A. Agresti, B. A. Coull, Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *Am. Stat.* **52**, 126 (1998).

49. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B.* **57**, 289–300 (1995).

50. P. J. Thul, L. Akesson, M. Wiking, D. Mahdessian, A. Geladaki, H. Ait Blal, T. Alm, A. Asplund, L. Björk, L. M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, A. Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. Von Feilitzen, K. S. Lilley, M. Uhlén, E. Lundberg, A subcellular map of the human proteome. *Science (80-. ).* **356**, eaal3321 (2017).

51. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

52. J. Sprenger, J. L. Fink, S. Karunaratne, K. Hanson, N. A. Hamilton, R. D. Teasdale, LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res.* **36**, D230–D233 (2008).

53. K.-C. Chou, Z.-C. Wu, X. Xiao, iLoc-Euk: A Multi-Label Classifier for Predicting the Subcellular Localization of Singleplex and Multiplex Eukaryotic Proteins. *PLoS One*. **6**, e18258 (2011).

54. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

55. M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).

56. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

57. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).

**Acknowledgments:**

We thank all members of B.B.'s laboratory, for discussions and advice; David Coombs for help with optimization of ribosome separation on sucrose gradients; Ulrike Friedrich for help on establishing pipelines for processing ribosome profiling data; Simon Anders for valuable advice concerning development of DiSP data analysis tools; the ZMBH Flow Cytometry & FACS Core Facility, the DKFZ Sequencing Core Facility and the DKFZ Vector and Clone Repository for support of experimental work. M.B., K.F. and J.S. are members of the Heidelberg Biosciences International Graduate School (HBIGS).

**Funding:** M.B. and K.F. were supported by a HBIGS PhD fellowship. M.B. was additionally supported by a Boehringer Ingelheim Fonds (BIF) PhD fellowship. F.W. received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 745798. This work was supported by the Helmholtz-Gemeinschaft (DKFZ NCT3.0 Integrative Project in Cancer Research (DysregPT_Bukau 1030000008 G783)), the Deutsche Forschungsgemeinschaft (SFB 1036), the European Research Council (ERC Advanced grant (743118)) and the Klaus Tschira Foundation. Work in the Tans laboratory was supported by the Netherlands Organization for Scientific Research (NWO).

**Authors contributions:**
Conceptualization: M.B., K.F., F.W., J.S., S.T., B.B., and G.K.
Methodology: M.B., K.F., J.A., B.B. and G.K.
Investigation: M.B., K.F.
Software: I.K., F.T., M.B. and K.F.
Formal Analysis, Data curation and Visualization: M.B., K.F., I.K., F.T., B.B. and G.K.
Writing – Original Draft: M.B., K.F., I.K. and G.K.
Writing – Review & Editing: all authors
Supervision: S.T., B.B. and G.K.

**Competing interests:** All authors declare no competing interests.

**Data and materials availability:** All sequencing data reported in this study are available at GEO under accession number GSE151959. Explicit Julia code is available as supplementary material; explicit R code will be made available upon request. Data analysis of ribosome profiling datasets were performed with RiboSeqTools (available at https://github.com/ilia-kats/RiboSeqTools and (*32*)).

**Supplementary Materials:**
Materials and Methods
Figures S1-S4
Tables S2 and S4
Captions for Tables S1, S3 and S5
Captions for Custom Julia Scripts 1-3
References (*35-57*)

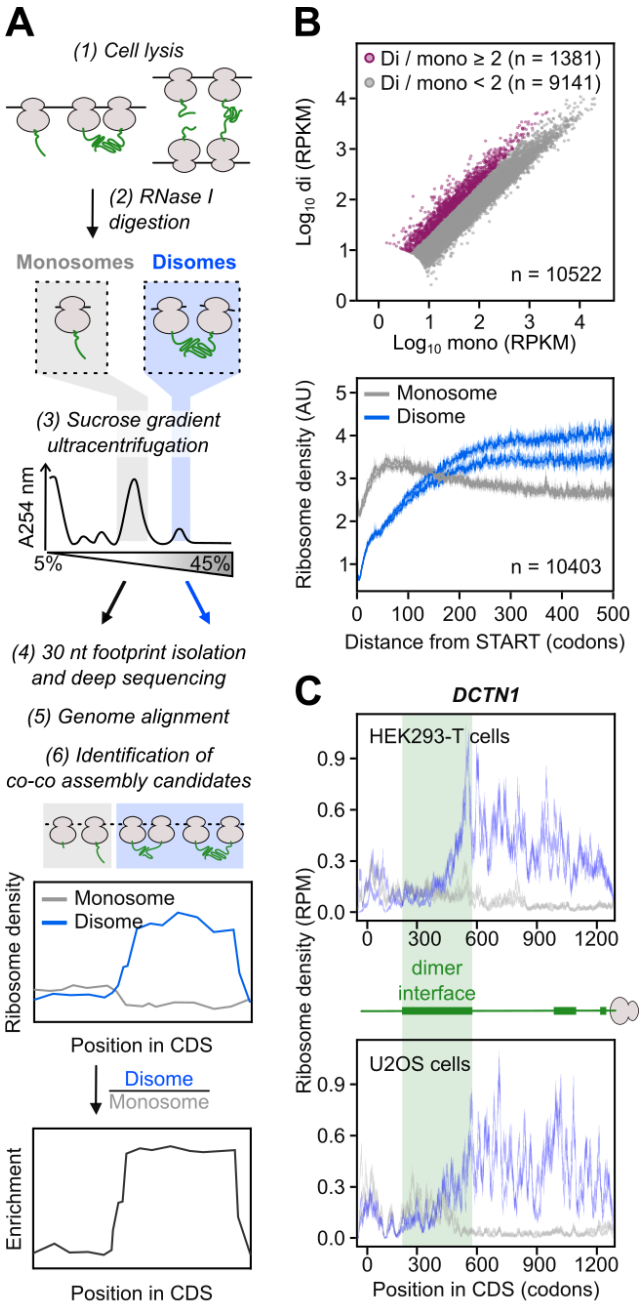


**Fig. 1. Disome Selective Profiling (DiSP) reveals widespread disome formation**

A) Experimental procedure of DiSP: cell lysates are RNase-treated (1, 2), monosomes and disomes are separated by sucrose gradient ultracentrifugation (3) and ribosome footprints with a length of about 30 nucleotides are extracted, converted into a DNA library and sequenced (4). Co-co assembly candidates are identified by a shift of the footprint density from monosome to disome fraction, or by a disome over monosome enrichment profile (5, 6).

B) Comparison of disome (di) and monosome (mono) footprint density (RPKM = Reads Per Kilobase per Million mapped reads) of all detected genes in HEK293-T cells (top, one replicate shown). Average footprint density along the coding sequence of all detected genes (metagene) aligned to translation start (bottom, n = 2).

C) Monosome (grey) and disome (blue) footprint density along the coding sequence (CDS) of *DCTN1* (RPM = Reads Per Million). Cartoon shows exposed nascent chain segments during translation, green bars indicate dimerization interfaces. DiSP data of HEK293-T (n = 2) and U2OS cells (n = 2) are compared.

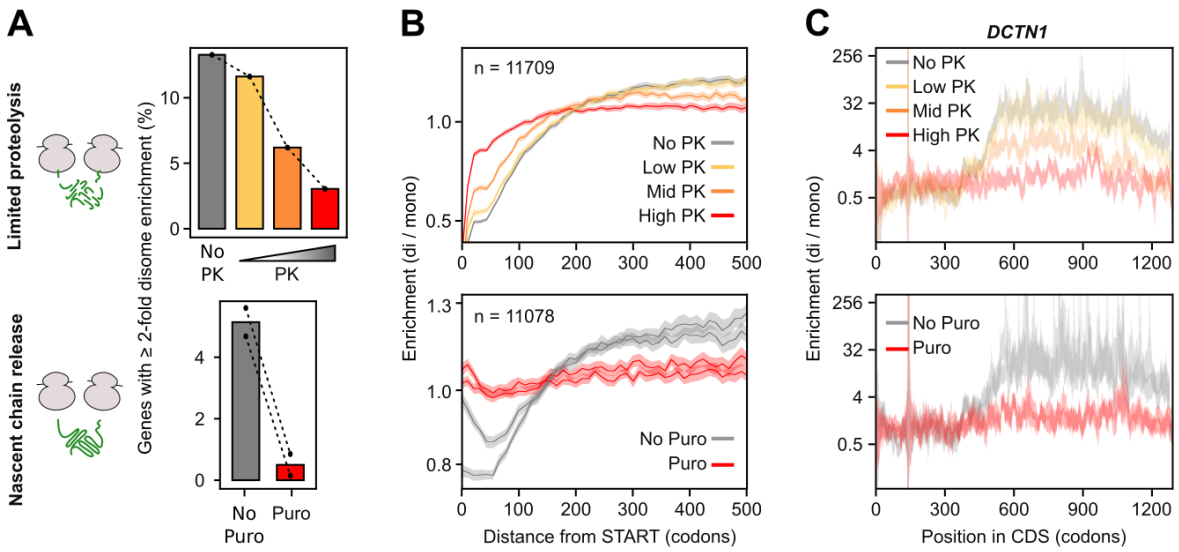


**Fig. 2. Disome formation is nascent chain dependent**

A) DiSP was performed on lysates treated with increasing Proteinase K (PK, n = 1) concentrations or with Puromycin (Puro, n = 2) to degrade or release nascent chains. Both treatments resulted in a large depletion of genes with ≥ 2-fold higher footprint density in the disome compared to the monosome fraction.

B) Metagene enrichment profiles (disome / monosome) aligned to translation start of all detected genes in PK (top) and Puro (bottom) DiSP experiments.

C) Enrichment profiles (disome / monosome) of *DCTN1* of untreated DiSP samples and samples treated with increasing concentrations of Proteinase K (PK, top) or with Puromycin (Puro, bottom).

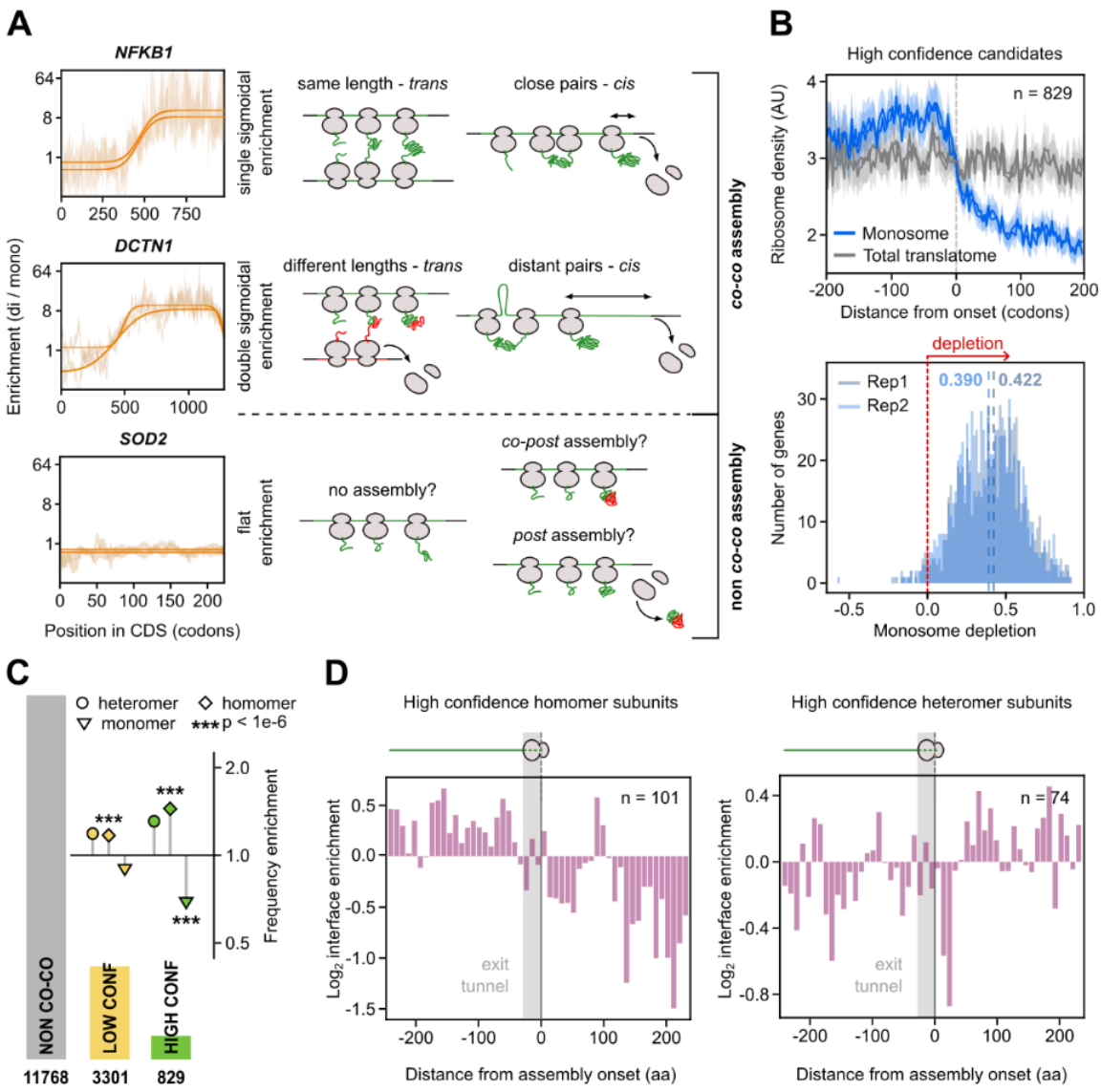


**Fig. 3. High confidence co-co assembly proteins are enriched in homo-oligomers**

A) Examples of gene-specific disome over monosome enrichment profiles (DiSP data, in the background, n = 2) and the corresponding fitting (solid lines) for each of the three possible shapes of DiSP enrichments. The single sigmoid agrees with nascent chain-connected ribosomes that terminate translation simultaneously, either by co-co assembly in trans if the mRNA segments translated by both ribosomes after co-co assembly have similar lengths, or in cis, with ribosomes that closely follow each other on the same mRNA (top). The double sigmoid agrees with co-co assembly involving two ribosomes that do not terminate at the same time; this may occur in trans if the mRNA segments translated by both ribosomes after co-co assembly have different lengths, or in cis, if the leading ribosome is distant from the trailing one (middle). Flat enrichment profiles indicate that nascent proteins do not co-co assemble.

B) Metagene profiles of all high confidence candidates aligned to assembly onset (top). Footprint density in the monosome fraction and the total translatome are shown (n = 2). Gene-specific

quantification of the efficiency of co-co assembly, calculated as the relative depletion of footprint density in monosome compared to total translatome after assembly onset (bottom). The median monosome depletion for each replicate is indicated by blue dashed lines.

C) Frequency enrichment of annotated subunits of protein complexes in high and low confidence lists compared to the whole proteome (absolute and relative numbers are provided in Table S2) (*33*). The number of genes included in each assembly class is indicated in the bar plot. The p-values were calculated using an enrichment test adjusted for expression bias (*33*, *34*).

D) Distribution of residues forming the inter-subunit interface of protein complexes determined from available crystal structures. The position of interface residues on the proteins' primary sequence is aligned to assembly onset of high confidence homomers (left) or heteromers (right).

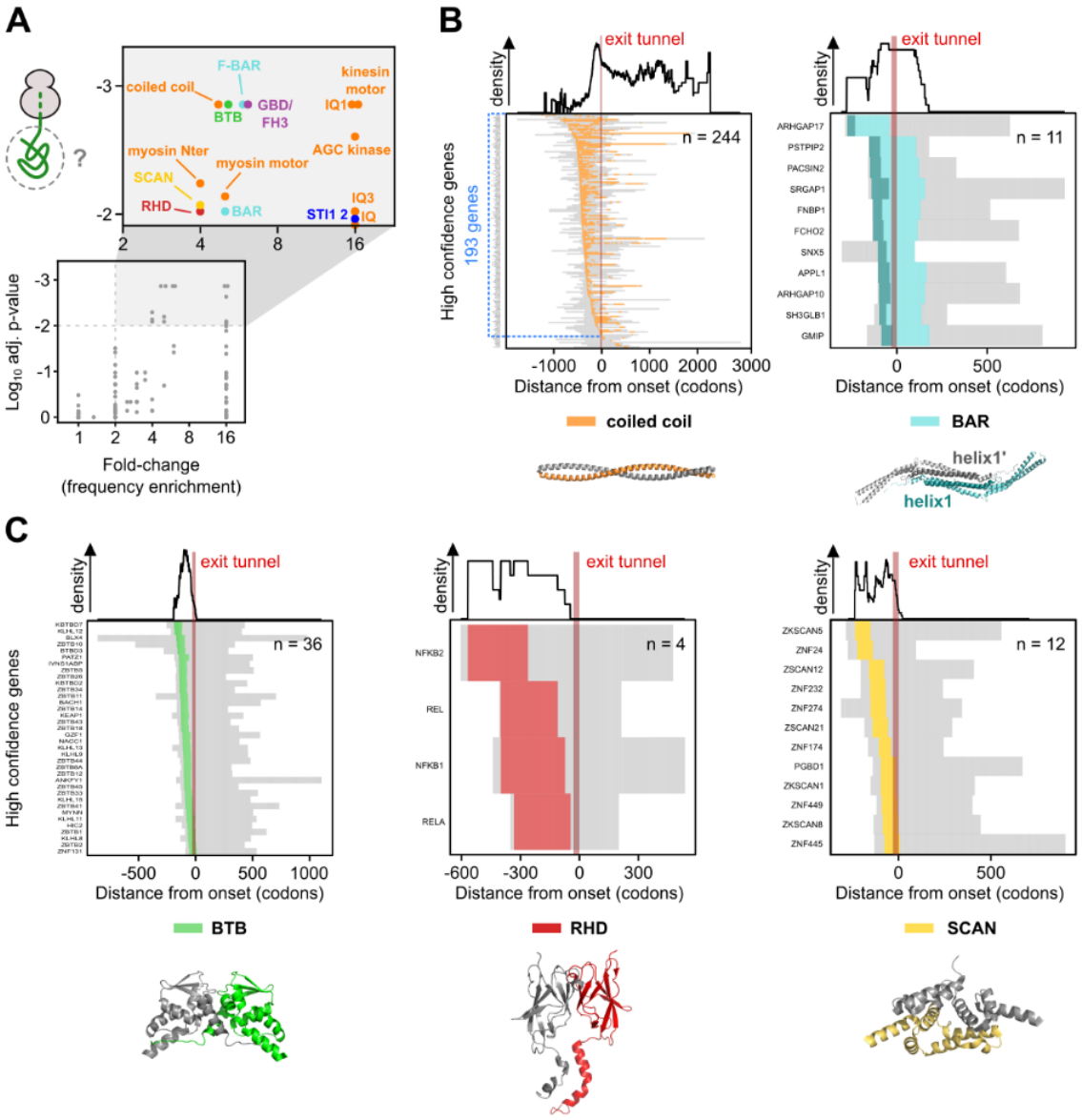


**Fig. 4. Co-co assembly is coordinated with exposure of five major dimerization domain classes**

A) Analysis of protein domains on nascent chain segments exposed at assembly onset. The frequency of each domain in the high confidence class is compared to their general frequency in the proteome (*33*). We used a Monte-Carlo simulation of the null hypothesis to calculate the p-value (*33*) and the Benjamini-Yekutieli procedure to correct for multiple testing. The adjusted p-value is plotted against the respective fold-change (frequency enrichment). Domains passing a significance (p-adj. ≥ 0.01) and fold-change (≥ 2) threshold are shown in the magnified rectangle and further analyzed.

B) Heatmaps of partially exposed domains: coiled coil (left) and BAR (right). In the heatmaps, nascent chain segments left from the ribosome exit tunnel (approximated to 30 codons, shown by a red bar) are exposed when assembly starts. The subset of genes exposing a coiled coil segment on the nascent chain at the onset of assembly is highlighted in blue (n = 193). Residues forming

helix1 of BAR domains are colored dark green in the heatmap and in the exemplary structure. Corresponding domain density profiles shown on top. Representative structures are PDB: 1D7M, 3Q0K.

C) Heatmaps of completely exposed domains: BTB (left), RHD (middle) and SCAN (right). Corresponding domain density profiles shown on top. Representative structures are PDB: 1BUO, 1K3Z, 3LHR.

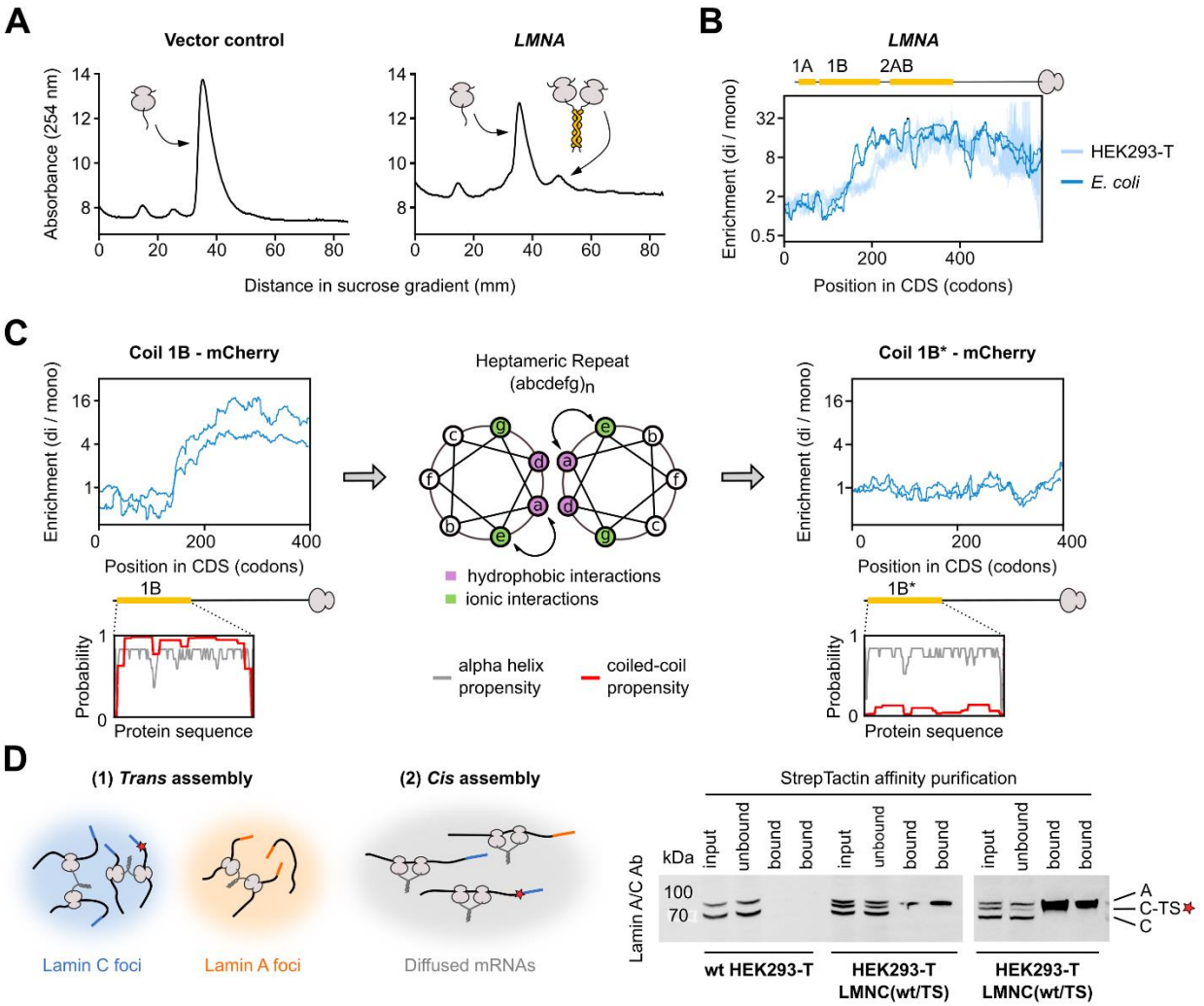


**Fig. 5. Co-co assembly does not rely on eukaryote-specific factors and facilitates native biogenesis of lamin C homodimers**

A) Sucrose gradient sedimentation analysis of *E. coli* ribosomes from cells transformed with a control plasmid (left) or a plasmid enclosing human *LMNA* encoding lamin C (right), lacking the unstructured N-terminal head domain (*33*).

B) Disome over monosome enrichment profile of plasmid-encoded *LMNA* expressed in *E. coli* (dark blue, n = 2), and endogenously expressed *LMNA* in HEK293-T cells (light blue, n = 2). The ribosome-exposed coiled coil interfaces are indicated by yellow bars.

C) Disome over monosome enrichment profiles of *LMNA* encoding lamin coil 1B (left) or the a↔e swapped version of 1B (1B*, right) fused N-terminally to mCherry and expressed in *E. coli* (n = 2). The ribosome-exposed coiled coil interfaces are indicated by yellow bars. Helical wheel projection shows residue arrangements (a-g) of the heptad repeat (middle). Coiled coil (red) and alpha-helical (grey) probability predictions are shown for both wild type and mutant 1B (insets).

D) Hypothetical models of co-co assembly supporting isoform-specific homodimerization (left). A red star represents the TwinStrep tag (TS). Affinity purification of tagged lamin C (C-TS) from

wild type or heterozygous *LMNC*(wt/TS) HEK293-T cells (bottom, technical replicates shown). Bands are labeled: A (lamin A), C (lamin C), C-TS (lamin C – TwinStrep).

# Supplementary Materials for

## Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly

Matilde Bertolini, Kai Fenzl, Ilia Kats, Florian Wruck, Frank Tippmann, Jaro Schmitt, Josef Johannes Auburger, Sander Tans, Bernd Bukau and Günter Kramer

correspondence to: g.kramer@zmbh.uni-heidelberg.de
bukau@zmbh.uni-heidelberg.de

**This PDF file includes:**



**Other Suplementary Materials for this manuscript includes the following:**

## Materials and Methods

Cell culture:

U2OS cells (Homo sapiens osteosarcoma, ATCC Cat# HTB-96, RRID: CVCL_0042) and HEK293-T cells (Homo sapiens embryonal kidney, DSMZ Cat# ACC 635) were cultivated in high glucose DMEM media containing GlutaMAX™ and pyruvate (Gibco) supplemented with 10% heat-inactivated FCS (Gibco), 100 units/mL penicillin and 100 µg/mL streptomycin (Gibco). Cells were passaged regularly through trypsinization (Gibco) and grown in a humidified incubator with 5% $CO_2$ at 37°C (HERAcell 150i). For all experiments, cells were seeded 18-24 hours before lysis in 15 $cm^2$ dishes (3.5 million U2OS or 6 million HEK293-T cells) to reach 70-90% confluency at the time of harvesting. A single dish of cells seeded in this way is enough for performing one DiSP experiment.

Cell line generation:

The sequence encoding for GFP11-TwinStrep was inserted upstream of lamin C stop codon at the *LMNA* endogenous locus via CRISPR homology-directed repair. GFP11 was included to allow FACS selection of positive edits by complementation with plasmid-expressed GFP1-10. The sequence encoding for the TwinStrep tag includes a shortened linker as described in (*35*) to reduce the template size. The single-stranded donor oligonucleotide (ssODN) was designed according to (*36*) with 35 nt homology arms at each side of insertion and purchased from IDT (ultramer oligo, desalted) (T5, Table S3.2). crRNA was designed according to the Dharmacon online design tool (http://dharmacon.horizondiscovery.com/gene-editing/crispr-cas9/crispr-design-tool/), (crRNA5, Table S3.2).

Genome editing was performed by transfection of Ribo-Nucleo-Proteins (RNPs) using Invitrogen™ TrueGuide™ Synthetic gRNA reagents and user guide. Briefly, 60,000 HEK293-T cells / well were seeded on poly-L-lysine coated 24-well plates (Greiner) the day before transfection. On the next day, Cas9/gRNA/Cas9 Plus solution mix was prepared in RNase-free tubes (7.5 pmol TrueCut Cas9 protein v2, 7.5 pmol crRNA:tracrRNA duplex and 1:10 v/v Lipofectamine™ Cas9 Plus™ Reagent in Opti-MEM™ medium, 20 µl per well). After incubation for 5 min at room temperature, 5.5 pmol of ssODN template were added. Diluted Lipofectamine™ CRISPRMAX™ reagent (1.5 µl in 25 µl opti-MEM™ / well) was added to the transfection RNP mix and 55 µl final transfection complex was distributed on each well. After 24 hours, cells were trypsinized and passed to poly-L-lysine coated 6-well (Greiner) with fresh DMEM supplemented with 10% FBS. On the following day, cells were transfected with 1.5 µg of pcDNA3.1-GFP1-10 plasmid, 4.5 µl Invitrogen™ Lipofectamine™ 2000 Reagent in opti-MEM™ (180 µl transfection mix per well). After 24 hours, cells were FACS-sorted at the ZMBH Flow Cytometry and FACS facility to enrich for positive edits. Single clones were grown and the edit was validated by genome extraction, PCR (primers MB132 + MB133, Table S3.1) and sequencing.

Affinity purification of lamin C – TwinStrep:

Wild type and heterozygous *LMNC*(wt/TS) HEK293-T cells were grown until confluence in one T75 flask each, harvested by trypsinization and washed in 1x PBS. Each cell pellet was resuspended in 0.5 ml hypotonic buffer (10 mM HEPES pH 7, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM EDTA, 0.05% NP-40), nuclei were pelleted at 3,300 ×g for 10 min and washed once more in 0.5 ml hypotonic buffer. Nuclei were lysed in 200 µl lamin extraction buffer (25 mM Tris pH 8.6, 1% NP-40, 0.5% DOC, 0.1% SDS, 500 mM NaCl, 1 µl Benzonase (E1014 Millipore), EDTA Free protease inhibitor tablet Roche), which is a modified version of standard RIPA buffer, optimized

according to (*37*) to allow solubilization of lamin dimers from the nuclear lamina. Nuclear lysates were incubated for 10 min in ice with occasional shaking and cleared by centrifugation for 10 min at 20,000 ×g. Each cleared lysate was subjected to affinity purification with 40 µl MagStrep "type3" XT beads (5% suspension, iba) according to provider's instructions. Elution was performed by incubating beads with 20 µl lamin extraction buffer supplemented with 1x Buffer BXT (iba) for at least 10 min at RT. Input, flow-through and elution samples were analyzed by Western blotting using anti-Lamin A/C antibody (Santa Cruz Biotechnology Cat# sc-376248, RRID: AB_10991536).

Disome Selective Profiling (DiSP):

Lysis protocols varied slightly for different experiments. Standard lysis buffer contained 50 mM HEPES pH 7.0, 10 mM $MgCl_2$, 150 mM KCl, 1% NP40, 10 mM DTT, 100 µg/ml CHX, 25 U/ml recombinant Dnase1 (Roche) and protease inhibitor (complete EDTA free, Roche). Given the requirement for high salt concentrations in the Puromycin DiSP experiment (*38*), we employed a high-salt lysis buffer containing 500 mM KCl for all DiSP experiments of HEK293-T cells to allow comparison of the main and control datasets. Standard lysis buffer (containing 150 mM KCl) was employed for DiSP of U2OS cells (data shown in Fig. 1 and S1) and for an additional dataset of HEK293-T cells (not shown in this study), which revealed highly similar results to the DiSP results obtained under high salt conditions of HEK293-T cells (main dataset of this study).

Cells were taken from the incubator immediately before harvesting (maximum three dishes per time). After removing the growth media by inversion, all subsequent steps were performed on ice, using ice-cold and RNase-free solutions and tools.

HEK293-T cells were detached by pipetting 10 ml of 1x PBS supplemented with 100 µg/ml CHX and 10 mM $MgCl_2$ on dish, they were collected in falcon tubes and pelleted for 3 min at 2000 ×g, 4°C. The cell pellet derived from one dish was resuspended in 200 µl 1x high-salt lysis buffer and incubated for 15 min on ice.

U2OS cells are less easily detached by pipetting, therefore lysis was performed on dish: cells were first washed by gently pouring 10 ml of 1x PBS supplemented with 100 µg/ml CHX and 10 mM $MgCl_2$ to cover the whole dish surface; next, the PBS solution was removed completely and 100 µl 5x concentrated standard lysis buffer was added and cells were scraped from the plate. For all U2OS samples, RNase 1 was directly supplemented in the 5x lysis buffer (6.6 units/µl). The cell lysate of one plate (around 500 µl after scraping) was transferred to a 1.5 ml non-stick RNase-free tube (Ambion) and incubated for 15 min on ice.

Both HEK293-T and U2OS cell lysates were triturated five times through a 26-G needle and cleared by centrifugation for 5 min at 20,000 ×g at 4°C. For HEK293-T samples, RNA concentration in the cleared lysate was determined by Qubit HS RNA assay with 1:100 dilutions in water. Lysates were digested with 150U RNase1 (Ambion) / 40 µg RNA for 30 min at 4°C and 500 rpm on a thermomixer.

5-45% and 10-25% sucrose gradients were used for separation of monosome and disome fractions with similar results. Briefly, gradients were prepared with the Gradient Station (BioComp) using SW40 centrifugation tubes (SETON). Sucrose was dissolved in sucrose buffer (50 mM HEPES pH 7.0, 5 mM $MgCl_2$, 150 mM KCl, 100 µg/ml Cycloheximide, EDTA Free protease inhibitor tablet Roche) and solutions were filtered. Short caps were used to seal the tubes and 5 - 45% gradients were formed with the following custom mixing program: M#1: 09 sec/83.0°/30 rpm M#2: 09 sec/83.0°/0 rpm M#3: 01 sec/86.0°/40 rpm M#4: 7 min/90.0°/0 rpm, sequence 12121212121234. Alternatively, 10-25% sucrose gradients were mixed with a one-step mixing

program (2:19 min/81.5°/14 rpm). Gradients were stored at 4°C for at least 1 hour before use. Up to 300 µg total RNA was loaded per gradient, 5-45% gradients were centrifuged for 3.5 hours and 10-25% gradients for 3 hours at 35,000 rpm, 4°C (SW40-rotor, Sorvall Discovery 100SE Ultracentrifuge) to allow maximum separation of monosome and disome peaks. After centrifugation, absorbance profiles at 254 nm were recorded using the Piston Gradient Fractionator™ (Biocomp) and gradients were fractionated in 60 fractions of 200 µl that were immediately frozen in liquid nitrogen. Fractions corresponding to monosome and disome peaks were pooled separately and subjected to acid phenol RNA extraction (*39*). Note that 5 to 8 fractions between the monosome and disome peaks were usually excluded to minimize contamination between the two samples. Ribosome profiling libraries of U2OS samples were prepared as described in (*40*) and sequenced on a HiSeq 2000 (Illumina) at the DKFZ Core Facility for Sequencing. All other libraries were prepared as described in (*20*, *39*), in combination with a custom rRNA depletion (see below) and sequenced on a NextSeq550 (Illumina) according to the manufacturer's protocol.

Ribosome Profiling:

Total translatomes were generated by classical ribosome profiling as described in (*20*), in combination with rRNA depletion (see below) and sequenced on a NextSeq550 (Illumina) according to the manufacturer's protocol.

Custom rRNA depletion:

We removed the most prevalent rRNA fragments from our libraries by hybridization of custom biotinylated reverse complement DNA oligonucleotides (developed in collaboration with siTOOLs Biotech, Table S4), followed by a pull-down via magnetic Streptavidin beads (NEB). We generally performed rRNA depletion on the adaptor-ligated RNA footprints. To maximize efficiency, an additional depletion step was optionally performed on the circularized DNA using a reverse-complement pool of biotinylated oligos. Briefly, 5 µl ligated RNA or circularized cDNA was mixed with a 4-fold molar excess of the respective rRNA depletion oligo pool and DEPC water to a final volume of 25 µl. 2x wash/binding buffer (40 mM Tris pH7, 1 M NaCl, 2 mM EDTA, 0.1% Tween 20 supplemented with 2 µl murine RNase inhibitor) was added to a final volume of 50 µl. Nucleic acids were denatured in a thermocycler for 90 s at 99°C and hybridization was performed by decreasing the temperature by 0.1°C per second to 37°C, followed by a 15 min incubation at 37°C. For each reaction, a 2-fold excess Streptavidin Magnetic Beads (NEB) was calculated based on the beads binding capacity and the amounts of biotinylated oligos in reaction. Beads were washed three times with 750 µl 1x wash/binding buffer and resuspend in 10 µl 1x wash/binding buffer. Beads were added to the hybridized RNA/DNA-oligo mix and incubated for 15 min at room temperature (with occasional mixing). Biotinylated oligos hybridized to target rRNA were then magnetized and removed from the sample. The remaining nucleic acids were precipitated according to (*40*).

DiSP with Proteinase K treatment:

10 mg lyophilized Proteinase K from Tritirachium album (Sigma) were mixed with 1 ml ice-cold PK storage buffer (50 mM Tris pH 7.5, 5 mM $CaCl_2$, 40% glycerol). The stock was aliquoted and stored at -80°C. For PK treatments one aliquot was thawed and immediately used. All steps were carried out on ice, using pre-cooled ice-cold solutions and tools. DiSP with PK treatment was performed as described above using HEK293-T cells with some modifications. Briefly, cells were

harvested and resuspended in 1x high salt lysis buffer without protease inhibitors. Protein concentration in the cleared lysate was determined by Bradford assay (BioRad Protein Assay) and RNA digestion was performed as for standard DiSP.

Next, lysates were supplemented with different PK concentrations and incubated for additional 30 min at 10 rpm on a rotation wheel at 4°C. According to the protein content in the lysate, PK was titrated as follows:

- No PK = PK storage buffer was added in place of PK
- Low PK = 1:20,000 (PK to total protein amount)
- Mid PK = 1:6,000
- High PK = 1:2,000
- Very High PK = 1:200

Note that data derived from all five PK experiments were employed for bioinformatics determination of PK sensitivity of single gene candidates (see "Defining high confidence candidates" below), however, the "Very High PK" condition was omitted in graphs of Fig. 2 and S1 for simplicity.

Samples were loaded on 10-25% linear sucrose gradients containing protease inhibitors (complete EDTA free, Roche). RNaseI digestion was omitted in control samples to verify polysome integrity after PK digestion by polysome profiling (Fig. S1E, left). Total lysates were also analyzed on SDS PAGE to visualize the degree of protein degradation upon different PK treatments (Fig. S1E, right).

DiSP with Puromycin treatment:

Conditions suited to release nascent chains with Puromycin without dissociating ribosomes from mRNAs were adapted from (*38*). Cycloheximide had to be omitted from all solutions because incompatible with Puromycin activity. All steps were carried out working on ice with ice-cold solutions and tools. HEK293-T cells were seeded on poly-L-lysine coated 15 $cm^2$ dishes and lysed on dish as follows: cells were rinsed with ice-cold PBS supplemented with 10 mM $MgCl_2$ and lysed by scraping in 100 µl 5x concentrated standard lysis buffer lacking cycloheximide. Next, cleared lysates (roughly 500 µl / dish after scraping) were supplemented with KCl to obtain a final concentration of 500 mM. Puromycin samples were supplemented with 2 mM Puromycin (Gibco™ Puromycin Dihydrochloride) and control samples with the same volume of 1x lysis buffer. We found RNaseI to be considerably less active at 0°C compared to 4°C, therefore, RNA digestion was performed with 750U RNase1 (Ambion) / 40 µg RNA in an ice-bath for 25 min with occasional shaking. After incubation, lysates were cross-linked using 0.5% formaldehyde (Pierce™ 16% Formaldehyde (w/v), Methanol-free) and incubated for 30 additional minutes in an ice-bath. Samples were loaded on linear 5-45% sucrose gradients and all downstream steps were carried out as described for standard DiSP.

RNaseI digestion was omitted in control samples to verify polysome integrity after Puromycin treatment by polysome profiling (Fig. S1F, left). In these cases, sucrose fractions corresponding to the supernatant (containing released nascent proteins) and to polysomes (containing ribosome-bound nascent proteins) were collected. Proteins were precipitated with Trichloroacetic acid (TCA) and separated by SDS PAGE. Puromycilated nascent proteins were detected by Western blot using anti-Puromycin antibody (Millipore Cat# MABE343, RRID: AB_2566826) (Fig. S1F, right).

Cloning:

All primer sequences used for cloning are available in Table S3.1.

For DiSP experiments, *LMNA* residues 31-542, corresponding to lamin C lacking the unstructured head domain, was PCR-amplified from a self-made U2OS cDNA library (SuperScript™ III first-strand synthesis kit, ThermoFisher). The employed PCR primers (MB143 + MB144) added a NdeI restriction site followed by a splitFlAsH tag (SF: MAGSCCGG) at the 5' end and a TwinStrep tag (TS: GGSGSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKGA) with a BamHI overhang at the 3' end of the construct (final sequence named SFLMNCTS available in Table S5). T4 DNA ligase was used to ligate the gel-purified PCR fragment into a BamHI/NdeI restricted pET3a vector. The resulting plasmid was sequenced with standard Eurofins primers (T7 forward and pET reverse primers) and custom primers (MB75 + MB76).

This plasmid was further used as template for amplification of coil 1B (MB212 + MB213). The pET3a-SF-coil1B*-mcherry-TS plasmid was ordered (via BioCat), with a SpeI and XhoI restriction site flanking the mutated coil 1B* sequence (SF_Coil1B_Mut_mCherry_TS, Table S5). This plasmid was used to substitute the mutated coil 1B* sequence by the PCR amplified wild type coil 1B sequence via restriction and ligation (SF_Coil1B_WT_mCherry_TS, Table S5).

*DCTN1* was PCR amplified (MB209 + MB210) from a pENTR221-*DCTN1* (p150glued) plasmid ordered from the DKFZ vector and clone repository. Gibson assembly (*41*) was used to transfer the PCR amplified *DCTN1* sequence from the ordered plasmid into a pET3d-vector, flanked by an N-terminal splitFlAsH tag and a C-terminal TwinStrep tag (MB205 + MB206). The resulting plasmid (SF_DCTN1_TS, Table S5) was sequenced with standard Eurofins primers (M13 forward and reverse primers) and custom primers (MB197 + MB198).

Plasmids used for the dimerization assay were generated by PCR amplification of coil1A (MB159 + MB160), coil1B (MB161 + MB162) and coil2AB (MB163 + MB164), each flanked by homologous regions to the target vector, from a synthetic full length lamin sequence (Invitrogen). Gibson assembly was used to clone each fragment into a SalI/BamHI digest pJH391 plasmid containing a C-terminal TwinStrep tag. The resulting plasmids were sequenced with custom primers (#1229 + #1230).

DiSP in *E. coli*:

All generated plasmids were freshly transformed into competent *E. coli* cells (Rosetta F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR), Novagene), and selected on LB agar plates with the required antibiotics. Colonies were picked for overnight cultures in EZ Rich Defined Medium, which were used on the next day to inoculate 200 ml EZ-RDM to an initial $OD_{600}$ of 0.05. Cells were grown at 37°C in 1L baffled Erlenmeyer flasks with shaking at 120 rpm. Following procedures were performed as described in (*3*, *42*) with minor modifications. Briefly, cells were harvested during log phase ($OD_{600}$ = 0.5-0.6); if not otherwise stated, cells were induced for 16 min with 1 mM IPTG, isolated by fast-filtration and flash-frozen in liquid nitrogen. Frozen cell pellets were lysed by mixer milling (2 min, 30 Hz, Retsch) in the presence of 500 µl frozen lysis buffer (50 mM HEPES pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 5 mM $CaCl_2$, 0.4% Triton X-100, 0.1% NP-40, 1 mM chloramphenicol, protease inhibitor tablets (Roche), DNase I (Roche), and 1 mM TCEP or 1 mM DTT). Lysates were digested with MNase (produced in house, 150U MNase / 40 µg RNA) at 25 °C and 650 rpm on a thermomixer. Digestion was stopped by placing samples in ice and supplementing 6 mM EGTA. Lysates were loaded on pre-cooled 5-45% sucrose gradients (sucrose dissolved in 50 mM HEPES pH 7.0, 100 mM KCl, 10 mM $MgCl_2$, 1 mM chloramphenicol, protease inhibitor tablets (Roche), and 1 mM TCEP or 1 mM DTT), and centrifuged for 3.5 h at 35,000 rpm, 4°C. Fractions corresponding to monosomes and disomes

were isolated and ribosome-protected RNA footprints were processed as described above, in combination with an rRNA depletion step as described in (*42*).

Dimerization Assay:

This assay is based on (*43*) and aims to combine (i) $OD_{600}$ measurement, (ii) cell permeabilization, (iii) ONPG breakdown, and (iv) kinetic $OD_{420}$ quantification into a single step. The required FI8202 *E. coli* strain [ΔntrBCfadAB101::Tn10 laqIq lacL8/λ202] (*44*) has a lac repressor (lac1q) deletion, therefore it is galactosidase positive. Strains transformed with a plasmid expressing an active dimerization domain fused to the N-terminal part of the lambda repressor (residues 1 to 102 of λ repressor) will have reduced galactosidase activity. The pKH101 plasmid (expressing only N-terminal part of the lambda repressor) (*26*) was used as negative control, and pFG157 (expressing the full-length lambda repressor) (*26*) as positive control. Freshly transformed FI8200 colonies were picked from LB plates for overnight cultures in LB media. 80 µL of each overnight culture were transferred into a 96-well Greiner® flat bottom microplate (transparent), 120 µL freshly prepared master-mix (60 mM $Na_2HPO_4$, 40 mM $NaH_2PO_4$, 10 mM KCl, 1 mM $MgSO_4$, 36 mM β-mercaptoethanol, 6.70% (v/v) PopCulture® Reagent, 1.1 mg/ml ONPG, Lysozyme) were quickly added and the measurement started using SPECTROstar Nano Microplate Reader (program: $OD_{600}$ and $OD_{420}$ readings taken every 60 sec for 1 h, at room temperature, shook at 500 rpm (double orbital shaking) for 30 seconds before each cycle). The linear slope of $OD_{420}$ over time ($OD_{420}$/min) was multiplied by 5000, and adjusted for the $OD_{600}$ reading at the first time point (defined as Miller units). $OD_{600}$ was assumed to be constant since lysis of cells had only minor effect on the $OD_{600}$ values over time. Repression efficiencies were calculated as in (*26*).

Processing of DiSP raw sequencing data:

Samples obtained by DiSP of U2OS cells were sequenced on a HiSeq 2000 (Illumina) and data were processed as described in (*40*).

All other samples were sequenced on a NextSeq 550 (Illumina) and data were processed as follows:

3' adaptor sequences were trimmed with Cutadapt v1.13 using following command:

```
cutadapt -q20 -m23 --discard-untrimmed -O6 -a ATCGTAGATCGGAAGAG-
CACACGTCTGAACTCCAGTCAC -o <path_to_output>/outfile.fastq.gz
<path_to_input>/infile.fastq.gz 1>
<path_to_output>/Cutadapt_report.txt
```

Unique molecular identifiers (UMIs) were extracted from each read using a custom Julia script (Script1) (*45*) with the following command:

```
julia <path_to_script>/Script1.jl
<path_to_input>/infile.fastq.gz
<path_to_output>/outfile.fastq.gz --umi3 5 --umi5 2
```

The resulting fastq file contains the 7 nt long UMI in the read name, consisting of five random 3' and two random 5' nucleotides implemented in the library preparation to prevent ligation biases (*20*).

The trimmed reads containing the UMI information in the read name (outfile of Script1) were aligned to human or *E. coli* rRNA sequences with bowtie2 v.2.3.5.1 (*46*), using following command:

```
bowtie2 -t -x <path_to_index>/index_base -q
<path_to_input>/infile.fastq.gz --un
<path_to_output>/outfile.fastq -L 13 -S /dev/null 2>
<path_to_output>/Bowtie2_report.txt
```

Reads that did not align to rRNA were aligned to the human genome (GRCh38p10) or *E. coli* BL21 (DE3) genome (GCA_000022665.2 modified to include additional chromosomes consisting of plasmid-encoded gene sequences, see Table S5 for sequences and respective gene names) using STAR v2.7.1a (*47*) and following command:

```
STAR --runThreadN 24 --genomeDir <path_to_indexed_genome> --
readFilesIn <path_to_input>infile.fastq --outFilterMultimapNmax
1 --outFilterType BySJout --alignIntronMin 5 --outFileNamePrefix
<path_to_output> --outReadsUnmapped Fastx --outSAMtype BAM
SortedByCoordinate --outSAMattributes All XS --quantMode
GeneCounts --twopassMode Basic
```

For each gene, the transcript with the longest coding sequence was selected and reads were assigned (a-, p-, e-site) via a custom Julia script (Script2) using following command:

```
julia <path_to_script>/Script2.jl -c 1 -g
<path_to_genome_annotation>annotation.gff' -u -o
<path_to_output> <path_to_input>infile.bam
```

Each output HDF5 file contains one data set per gene. Each data set consists of a 2-row matrix, with the first row containing the 1-based position within the CDS, and the second row the number of detected p-site reads at this position. Additional information is stored in the data set attributes, including: gene and protein names, transcript isoform used for position assignment, length of the coding sequence, chromosome and strand location of the gene.
All analyses in this study were performed on p-site assigned reads aligned to the coding sequence (CDS) only, which were further analyzed with RiboSeqTools (available at: https://github.com/ilia-kats/RiboSeqTools and (*32*)) and custom scripts (see below).

Single gene enrichment profiles:
Ribosome profiling data are typically sparse and noisy. Simply plotting position-wise enrichment, as is often done, can convey a false sense of precision, even though the value may have been calculated from only a few reads and therefore carries considerable uncertainty. We therefore calculate position-wise enrichment confidence intervals (https://github.com/ilia-kats/RiboSeqTools and (*32*)).
In particular, let $D_i$ denote the number of disome reads with an assigned P-site at position $i$ for gene $g$ and $M_i$ the corresponding number of monosome reads (the subscript $g$ is omitted for the sake of notational simplicity). As usual, we assume that read counts follow a Poisson distribution: $D_i \sim \mathrm{Pois}(\lambda_{d,i})$ and $M_i \sim \mathrm{Pois}(\lambda_{m,i})$. We furthermore assume that $D_i$ is stochastically independent of $M_i$, in which case it can be shown that $D_i \mid D_i + M_i \sim \mathrm{Bin}(D_i + M_i, \frac{\lambda_{d,i}}{\lambda_{d,i}+\lambda_{m,i}})$. Writing $p_i := \frac{\lambda_{d,i}}{\lambda_{d,i}+\lambda_{m,i}}$, we calculate a 95% confidence interval for $p_i$ using the Agresti-Coull method (*48*). The enrichment confidence interval is then given by $b_{e_i} = \frac{b_{p_i}}{1-b_{p_i}}$, where $b_{e_i}$ and $b_{p_i}$ are confidence

bounds of $e_i$, the enrichment at position i, and $p_i$, respectively. We adjust for library size differences by decomposing the Poisson means $\lambda_{d,i} := \mu_{d,i}D$ and $\lambda_{m,i} := \mu_{m,i}M$, where D and M are total read counts for the mono- and disome libraries, respectively, and $\mu_{d,i}$ and $\mu_{m,i}$ are the parameters of interest. The library-size adjusted enrichment confidence interval is given by $\tilde{b}_{e_i} = b_{e_i}\frac{M}{D}$ and is shown in the single-gene plots. To minimize the impact of spurious peaks, which can arise due to amplification and/or sequencing biases, we set $\widetilde{D}_i = \sum_{k=i-7}^{i+7} D_k$ and $\widetilde{M}_i = \sum_{k=i-7}^{i+7} M_k$ and use $\widetilde{D}_i$ and $\widetilde{M}_i$ to calculate the confidence interval, that is we smooth the read counts with a 15 codon wide sliding window.

Single gene density profiles:

For monosome and disome density profiles, we show the position-wise 95% Poisson confidence interval corrected for library size. Read counts are again smoothed with a 15-codon wide sliding window.

Metagene profiles:

Only genes for which the summed coverage (monosome + disome raw counts in two replicates) is higher than 0.5 read/codon (corresponding to 0.25 reads / codon in average in each replicate) are included in the analysis. The contribution of each gene is normalized to its expression level by dividing the read density at each codon position by the normalized read density of the gene in the total translatome (expressed in RPKM).

Finally, average or enrichment metagene profiles are calculated as the position-wise arithmetic mean or the position-wise enrichment of disome over monosome, respectively. Profiles are computed separately for each experiment and replicate from the full data set (all genes) as well as bootstrapping samples (sampling genes). Metagene profiles including all genes are plotted as solid lines, with the shading indicating the 95% bootstrapping confidence interval (https://github.com/ilia-kats/RiboSeqTools and (*32*).

Sigmoid fitting for the identification of co-co assembly candidates:

Proteins undergoing co-co assembly should show a sigmoidal disome/monosome enrichment profile, with low enrichment at the N-terminus and high enrichment at the C-terminus. If the distance between two ribosomes bridged by interacting nascent chains is large or if the protein is subject to trans co-co assembly, the leading ribosome may terminate with a sufficient lead time to the lagging ribosome that an enrichment drop-off at the C-terminus is evident. In this case, the enrichment profile would approximately follow a double sigmoidal model (Fig. 3A).

Uncertainty in the shape of the enrichment profile due to sequencing noise must be taken into account for candidate identification. Let $D_i$ denote the number of disome reads with an assigned P-site at position $i$ for gene $g$ and $M_i$ the corresponding number of monosome reads (the subscript $g$ is omitted for the sake of notational simplicity). As usual, we assume that read counts follow a Poisson distribution: $D_i \sim \mathrm{Pois}(\lambda_{d,i})$ and $M_i \sim \mathrm{Pois}(\lambda_{m,i})$. We furthermore assume that $D_i$ is

stochastically independent of $M_i$, in which case it can be shown that $D_i \mid D_i + M_i \sim \mathrm{Bin}(D_i + M_i, \frac{\lambda_{d,i}}{\lambda_{d,i}+\lambda_{m,i}})$. Writing $p(i) \coloneqq \frac{\lambda_{d,i}}{\lambda_{d,i}+\lambda_{m,i}}$, we consider three parametrizations for $p(i)$:

1. $p(i) \equiv p$, the null model with constant enrichment along the gene
2. $p(i) = \frac{I_{\max}-I_{\mathrm{init}}}{1+\exp(-a(i-i_{\mathrm{mid}}))} + I_{\mathrm{init}}$, the single sigmoidal enrichment profile. The free parameters are $I_{\mathrm{init}} \in (0,1)$, $I_{\max} \in (0,1)$, $a \in [0,0.5]$, and $i_{\mathrm{mid}} \in [1, l]$, where $l$ is the gene length.
3. $p(i) = (\frac{I_{\max}-I_{\mathrm{init}}}{1+\exp(-a_1(i-i_{\mathrm{mid}}))} + I_{\mathrm{init}})(\frac{1-I_{\mathrm{final}}}{1+\exp(-a_2(i-(i_{\mathrm{mid}}+i_{\mathrm{dist}})))} + I_{\mathrm{final}})$, the double sigmoidal model. The free parameters are $I_{\mathrm{init}} \in (0,1)$, $I_{\max} \in (0,1)$, $I_{\mathrm{final}} \in (0,1)$, $a_1 \in [0,0.5]$, $a_2 \in [-0.5,0]$, $i_{\mathrm{mid}} \in [1, l]$, and $i_{\mathrm{dist}} \in [1, l]$, where $l$ is the gene length.

For each model, parameters were estimated by maximum likelihood, and we select the best model using the Bayesian Information Criterion (BIC). Genes for which models 2 or 3 are selected are considered to be candidates for co-co assembly, unless the determined onset (the inflection point of the sigmoid) falls into the ribosome exit tunnel (codons 1-30) or the last codon.

These calculations are included in a sigmoid fitting script (Script3), which can be invoked by the following command:

```
julia <path_to_script>/Script3.jl <path_to_input>.hdf5
```

Defining high confidence candidates:

Treatment with Puromycin, which releases nascent chains from the ribosome, or Proteinase K (PK), which digests nascent chains, should disrupt disomes of proteins undergoing co-co assembly. The corresponding footprints would be detected in the monosome fraction. We therefore expect the enrichment profile of co-co assembling proteins to have a considerably less sigmoidal shape in our control experiments with Puromycin or PK treatment.

The Puromycin control experiment consists of two samples, one treated and one untreated. We used co-co assembly candidates and assembly onsets determined using the main experiment. Read counts before and after the assembly onset were summed up separately for the treated and untreated datasets. Note that for genes classified as double sigmoid in at least one replicate, "after onset" refers to after onset and before the end of co-co assembly. We then fitted a beta-binomial GLM of the form $\mathrm{logit}\left(\frac{d}{d+m}\right) = \beta_1 + \beta_2 a + \beta_3 p + \beta_4 ap - \log(s)$ to each gene, where $\beta$ is the weight vector to be estimated, $d$ is the number of reads in the disome sample, $m$ the number of reads in the monosome sample, $a \in \{0,1\}$ signifies whether the response variable is measured after onset of co-co assembly, and $p \in \{0,1\}$ signifies whether Puromycin was added. $s = \frac{\sum_g \sum_i m_{gi}}{\sum_g \sum_i d_{gi}}$ is a scaling factor accounting for differences in library size, where $m_{gi}$ and $d_{gi}$ are monosome and disome counts for gene $g$ at position $i$, respectively. The beta-binomial error model was chosen to account for overdispersion caused by biological or pre-sequencing technical variability. This model was compared to a simpler GLM lacking the interaction term using the likelihood-ratio test. False discovery rate was controlled using the Benjamini-Hochberg procedure (*49*).

The PK control experiment consists of an untreated sample and multiple samples treated with different PK concentrations. A sigmoidal dose-response model would be appropriate for this experimental setup. However, in this case it is not clear what the response and the appropriate error model would be and how to include additional covariates such as sequencing library size. We therefore used a GLM approximation. We first determined a predictor value for each PK concentration such that the predictors had a linear relationship with the response. More precisely, we used the 100 genes with the highest disome/monosome ratio after onset in the untreated sample

with at least 200 reads in the monosome sample, and we optimized the predictor values using maximum likelihood with a binomial error model, such that $x_0 = 0$ and $\text{logit}\left(\frac{d_g}{d_g+m_g}\right) = a_g x - \log(s) + \log\left(s_0 \frac{d_{g,0}}{m_{g,0}}\right)$, where $d$ is the number of reads in the disome sample, $m$ the number of reads in the monosome sample, $a_g$ and $x$ are free parameters and $g$ indexes over genes. The 0 subscript indicates the untreated sample and $s = \frac{\sum_g \sum_i m_{gi}}{\sum_g \sum_i d_{gi}}$ is a scaling factor accounting for differences in library size, where $m_{gi}$ and $d_{gi}$ are monosome and disome counts for gene $g$ at position $i$, respectively. We then used the determined $x$ values as surrogates for PK concentration. Similar to the analysis of the Puromycin experiment, we used co-co assembly candidates and assembly onsets determined using the main experiment. Read counts before and after the assembly onset were summed up separately for each dataset. Note that for double sigmoid fits, "after onset" refers to after onset and before the end of co-co assembly. We then fitted a beta-binomial GLM of the form $\text{logit}\left(\frac{d}{d+m}\right) = \beta_1 + \beta_2 a + \beta_3 x + \beta_4 a x - \log(s)$ to each gene, where $\beta$ is the weight vector to be estimated, $d$ is the number of reads in the disome sample, $m$ the number of reads in the monosome sample, $a \in \{0,1\}$ signifies whether the response variable is measured after onset of co-co assembly, $x$ is the surrogate PK concentration, and $s$ is the scaling factor. This model was compared to a simpler GLM lacking the interaction term using the likelihood-ratio test. False discovery rate was controlled using the Benjamini-Hochberg procedure (*49*).

We defined high confidence co-co assembly candidates as proteins which showed significant responses to both Puromycin and PK treatment at FDR ≤ 0.01 and for which both PK and Puromycin effects (the coefficients of the interaction term from the respective model) were negative. We further restricted high confidence candidates to cytosolic or nuclear proteins, using a custom annotation combining information from several sources, as explained in the next paragraph.

Calculation of monosome depletion:

We observed a distinct downward trend in total translatome ribosome density towards the C-terminus of some genes. We therefore normalize monosome reads to total translatome read counts to quantify the depletion of monosomes after onset of co-co assembly. Specifically, we calculate a gene-wise density ratio as $r_g = \frac{\frac{\sum_{i=o_g+1}^{l_g} M_{g,i}}{T_{g,a}}}{\frac{\sum_{i=1}^{o_g} M_{g,i}}{T_{g,b}}}$, where $T_{g,b}$ and $T_{g,a}$ are the number of reads in the total translatome for gene g before and after onset, respectively, $l_g$ is either the length of gene g or the end of co-co assembly for genes classified as double sigmoids, and $M_{g,i}$ is the number of monosome reads for gene g at position i. $T_{g,b}$ and $T_{g,a}$ are averages of RPM over replicates. We define monosome depletion as $1 - r_g$.

As a control, we repeated the analysis with randomized assembly onsets. Since we observed a log-log-linear relationship between CDS length and both assembly onset and end of assembly for double sigmoid genes, randomized onsets and assembly endpoints were generated conditional on the CDS length. Specifically, we fitted a linear regression using log CDS length as predictor and log onset (or log endpoint) as the response variable. For each gene, a new onset (and endpoint for double sigmoid genes) was drawn from a truncated Normal distribution with mean and standard

deviation given by the linear regression prediction and the regression's residual standard deviation, respectively, truncated to 1 and the CDS length. We then calculated monosome depletions as described above. The entire process was repeated 10000 times. In each iteration, we took the median monosome depletion. The distribution of median depletions from the randomized control is compared to the value obtained using real data in Fig. S2C.

Comprehensive annotation of the human proteome:

To obtain a complete annotation of the subcellular localization of human proteins, we retrieved and merged information from different databases: Human Proteome Atlas (*50*), UniProtKB (*51*), LOCATE (*52*), and the benchmark dataset of iLoc-Euk (*53*). Annotations from mouse/rat homologs were employed in case no annotation was available for the human protein. To classify a protein as 'cyto-nuclear' it required the occurrence of at least one of the following keywords ('cytosol', 'nucleoplasm', 'nucleus', 'cytoplasm', 'nucleoli', 'nucleolus', 'perinuclear region of cytoplasm') in the merged annotation file and the absence of any TMD annotated in UniprotKB.

Annotation of the proteins' oligomeric state was retrieved from another set of databases: UniprotKB (*51*), PDB (*54*), Corum (*55*), Swissmodel (*56*).

We implemented a hierarchical annotation scheme in order to avoid multiple annotations for the same proteins:

(i) in case of multiple annotations from different organisms, we ranked human > mouse > rat;

(ii) in case of annotations from different databases within an organism, we ranked UniprotKB > PDB > Corum > Swissmodel;

(iii) in case of multiple oligomeric states within a database, we ranked "homomer" > "heteromer" > "monomer". Proteins annotated as "heteromer of homomers" were excluded to avoid noise from subunits whose assembly partner is uncertain.

Enrichment of protein domains:

Annotation of protein domains and the respective positions in protein sequences was retrieved from UniprotKB (*51*) ("Domain[FT]" and "Coiled coil" fields). Each domain was considered "exposed" in high confidence candidates if its N-terminal boundary was included in the ribosome-exposed nascent chain at assembly onset (calculated as DiSP onset – 30 residues to account for the ribosomal exit tunnel).

A simple comparison of the frequency of "exposed" domains at assembly onset in the high confidence class to their general frequency in the human proteome (including full-length proteins) would be biased towards detection of domains that are generally found at the N-terminus of proteins. To reveal genuine co-co assembly-driving domains we compared N-terminal portions of high confidence candidates to similar N-terminal portions of proteins in the human proteome. Therefore, we defined the background (denominator of the enrichment analysis) as the protein

segments upstream of randomized assembly onsets of all "cyto-nuclear" proteins (belonging to any assembly class), and computed the significance of enrichment by a resampling approach:

(i) A sample of the same size as the high confidence class (829 genes) is first drawn from all "cyto-nuclear" proteins.

(ii) A randomized onset is assigned to each protein in the sample as explained above (see "Calculation of monosome depletion")

(iii) Steps (i) and (ii) are repeated $10^5$ times.

We calculated the proportion of high confidence proteins exposing each domain at DiSP assembly onset (prop_highconf) and compared it with the proportion of proteins exposing the same domain at randomized assembly onsets in each of the control random samples (prop_control). A median enrichment ("Fold-change (frequency enrichment)" in Fig. 4A) was defined for each domain as prop_highconf divided by the median of prop_control.

For significance analysis, we defined N as the number of samples for each domain where prop_control is equal or larger than prop_high and calculated p-values as $(N+1)/(10^5+1)$. Finally, p-values were adjusted for multiple comparisons using the Benjamini & Yekutieli method ("adj. p-value" in Fig. 4A) (*57*).

Enrichment of complex subunits:

Enrichment of complex subunits ("Frequency enrichment" in Fig. 3C) was calculated as the frequency of proteins annotated as monomers or part of oligomeric complexes in the low- or high confidence class divided by their frequency in the human proteome (background). Note that since the high confidence class only includes "cyto-nuclear" proteins, we employed "cyto-nuclear" proteins as background for the high confidence class. Abundance in the low confidence class was instead compared to all proteins.

The subset of proteins detected by DiSP and included in high- and low confidence classes are biased towards highly expressed genes. We used the goseq package (*34*) to perform bias-corrected analysis of enrichments and significance calculation.

Prediction of coiled coils based solely on the proteins' primary sequence by DeepCoil (*22*) was performed following the instructions at https://github.com/labstructbioinf/DeepCoil.

Since analysis is restricted to a maximum of 500 residues, a FASTA file including the sequence spanning 250 residues upstream and downstream of DiSP assembly onsets of all high confidence proteins was first generated (onset_aligned.fasta). As control, a similar FASTA file was generated including "cyto-nuclear" non co-co assembly proteins aligned to simulated assembly onsets (defined as described in "Calculation of monosome depletion" section).

Finally, the following command was employed:

```
python <path_to_script>/deepcoil.py -i
<path_to_infile>/onset_aligned.fasta -out_path
<path_to_outfolder>/predictions_out/
```

Structure interface analysis:

All X-ray structures with annotated human proteins were retrieved from PDB (*54*). For every gene, the structure with the highest sequence coverage and highest resolution was chosen, structure components not based on the 20 proteinogenic amino acids or protein chains with a length below 10 amino acids were ignored. The residue-specific solvent accessible surface area was calculated with FreeSASA (https://freesasa.github.io). Protein-wide structure interface analysis was performed as described in (*28*) and included only exclusively homomeric structures (Fig. S3B).

The same analysis was repeated to calculate onset-aligned interface enrichment in high confidence proteins (Fig. 3D), with following changes: onsets of high confidence candidates were set to position zero and only interfaces located in a window of 500 amino acids around the onset were considered. Analysis of homomer subunits was limited to exclusively homomeric structures and included interfaces between human proteins with identical UniProt ID within the same structure. Analysis of heteromeric subunits was limited to exclusively heteromeric structures (where no subunit was repeated more than once) and considered interfaces between proteins with different UniProt ID, where at least one subunit was enclosed in the high confidence list. For plotting, each data point was normalized by the arithmetic mean of all data points ("Interface enrichment").

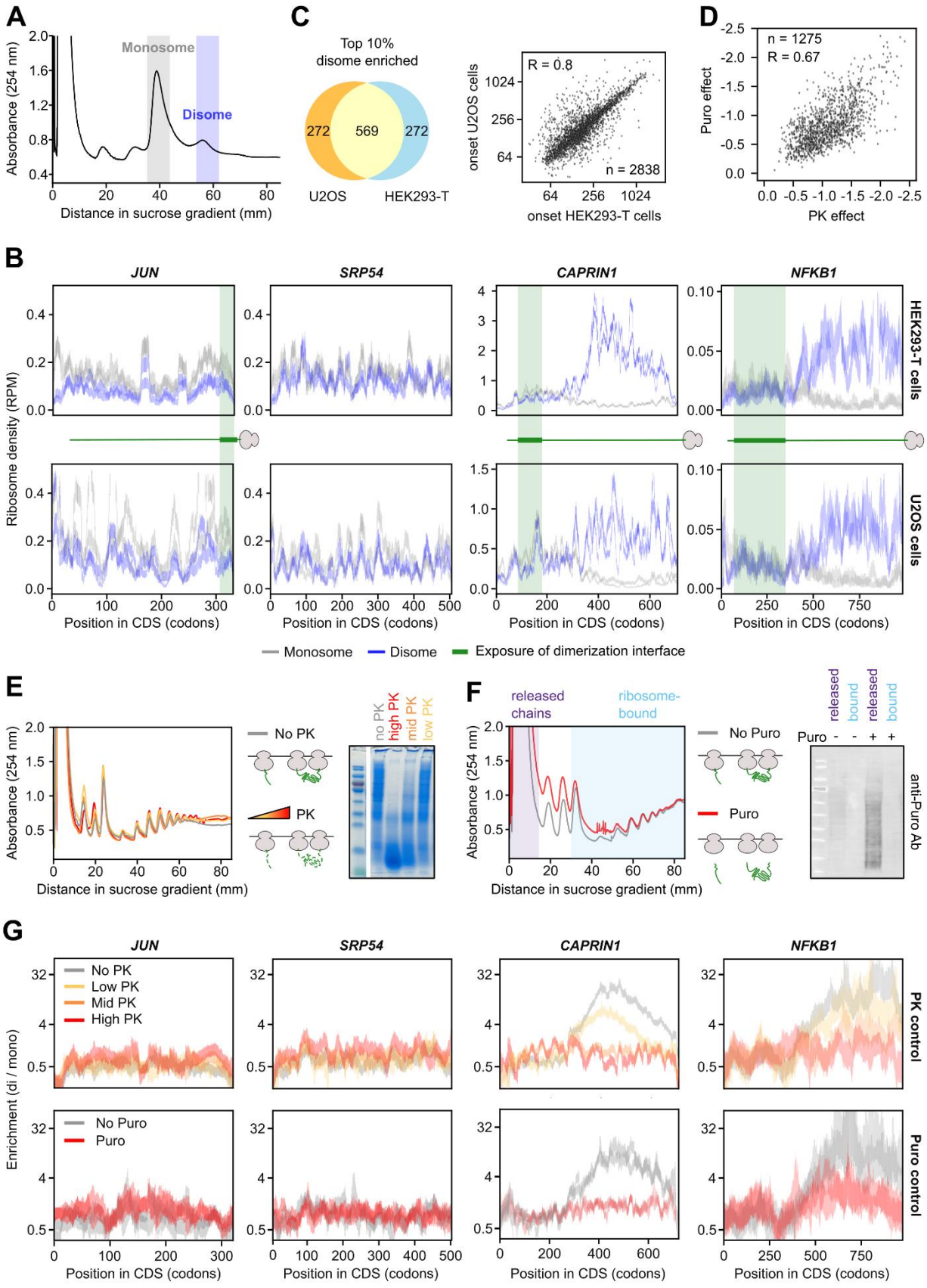
A
Absorbance (254 nm)
Monosome
Disome
Distance in sucrose gradient (mm)
C
Top 10%
disome enriched
272
569
272
U2OS
HEK293-T
R = 0.8
onset U2OS cells
onset HEK293-T cells
n = 2838
D
n = 1275
R = 0.67
Puro effect
PK effect
B
JUN
SRP54
CAPRIN1
NFKB1
Ribosome density (RPM)
HEK293-T cells
U2OS cells
Position in CDS (codons)
Monosome
Disome
Exposure of dimerization interface
E
Absorbance (254 nm)
Distance in sucrose gradient (mm)
No PK
PK
no PK
high PK
mid PK
low PK
F
released
chains
ribosome-
bound
Absorbance (254 nm)
Distance in sucrose gradient (mm)
No Puro
Puro
released
bound
released
bound
Puro - - + +
anti-Puro Ab
G
JUN
SRP54
CAPRIN1
NFKB1
Enrichment (di / mono)
No PK
Low PK
Mid PK
High PK
PK control
No Puro
Puro
Puro control
Position in CDS (codons)

**Fig. S1. Disome Selective Profiling (DiSP) reveals widespread nascent chain dependent disome formation**

A) Absorbance at 254 nm along a 10-25% sucrose gradient loaded with RNase1 digested lysate of HEK293-T cells (SW40 rotor, centrifugation: 3h, 35 000 rpm, 4°C). Isolated monosome and disome fractions for DiSP are indicated with a grey and a blue box, respectively.

B) Normalized monosome and disome footprint density distributions along the coding sequence (CDS) of two disome-enriched candidates (*CAPRIN1*, *NFKB1*) and two non-enriched candidates (*JUN*, *SRP54*). DiSP of HEK293-T (upper row, n = 2) and U2OS cells (lower row, n = 2) are shown. Cartoons indicate the exposed nascent chain segments during translation (assuming that the ribosomal tunnel covers the C-terminal 30 residues), green bars indicate dimer interfaces. RPM = Reads Per Million.

C) Top 10% disome enriched genes overlap between HEK293-T and U2OS cells (including only genes expressed in both cell lines, left). Onsets of disome shifts (turning point of a sigmoidal curve fitted to the enrichment profiles) highly correlate between HEK293-T and U2OS cells (right).

D) The loss of sigmoidal shape of disome enrichment profiles after Proteinase K (PK) and Puromycin (Puro) treatment (effect, see (*33*)) is correlated (only candidates significantly affected by both controls included).

E) Polysome profiles of control and PK-treated lysates indicate that ribosome integrity is not visibly affected under the employed protease concentrations (left), while the effect on the whole proteome is visible by SDS-PAGE (right).

F) Polysome profiles of control and Puro-treated lysates show that Puro does not lead to ribosome disassembly under the employed experimental conditions (left). Immunostaining of puromycylated nascent chains indicates efficient release from ribosomes to the supernatant fraction.

G) Enrichment profiles (disome / monosome) along the coding sequence (CDS) of the candidates shown in (A) upon treatment of lysates with different Proteinase K (PK) concentrations or Puromycin (Puro).

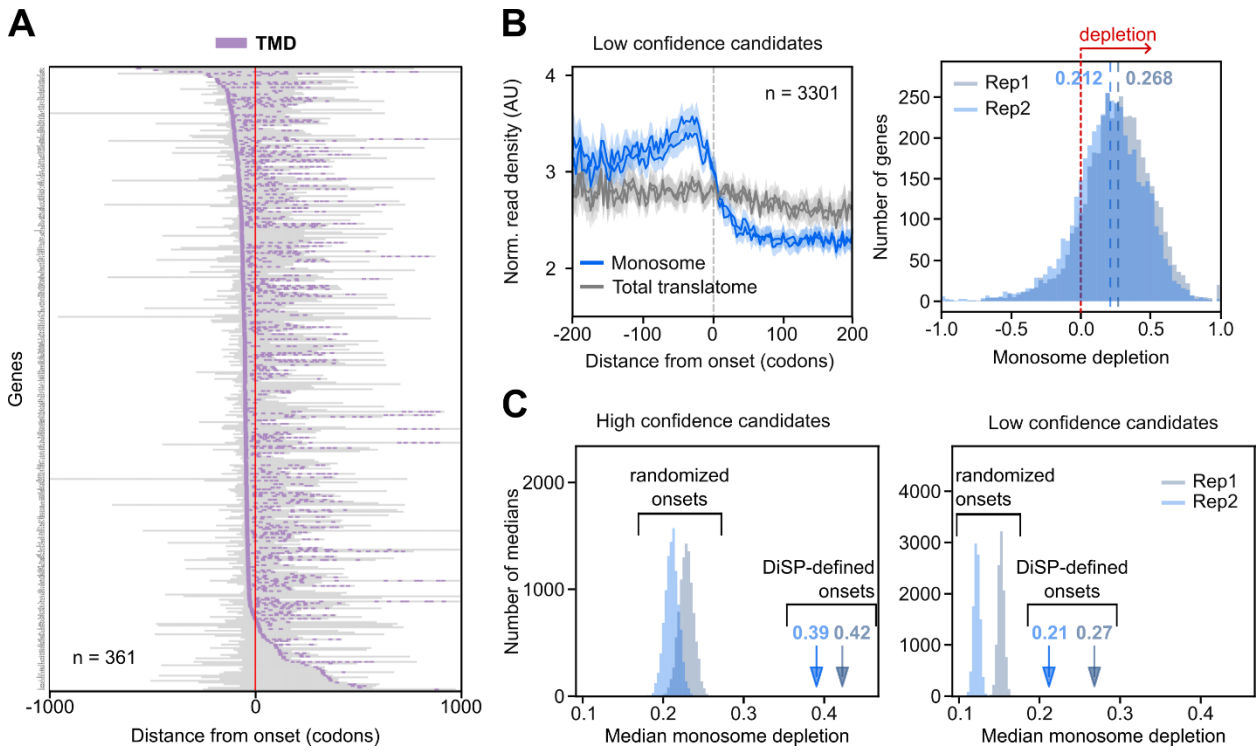


**Fig. S2. Features of high and low confidence co-co assembly candidates**

A) Heatmap of transmembrane domain positions (TMD, violet) aligned to the onset of co-co assembly. Low confidence candidates that contain an annotated TMD and fulfill criteria (i) to (iii) are analyzed.

B) Metagene profiles of low confidence candidates aligned to assembly onset (left); footprint density in the monosome fraction and the total translatome are shown (n=2). Monosome depletion is quantified for each gene separately by analyzing the fraction of remaining footprints downstream compared to upstream assembly onset, normalized by the total translatome (right). Median monosome depletion in two replicates are shown by blue dashed lines.

C) To verify that monosome depletion of high confidence (Fig. 3B) and low confidence candidates (panel B of this figure) is not observed by chance but depends on the specific onset positions determined by DiSP, monosome depletion is calculated with randomized onsets (and offsets in case of double sigmoidal profiles (*33*)) in $10^5$ iterations. The median monosome depletion of each randomized sample is calculated and plotted separately for two replicates of high confidence (left) and low confidence candidates (right). No median depletion from random sampling is equal or higher than the median depletion calculated from the real DiSP data (shown by blue arrows), demonstrating that monosome depletion after onset of co-co assembly is not observed by chance.

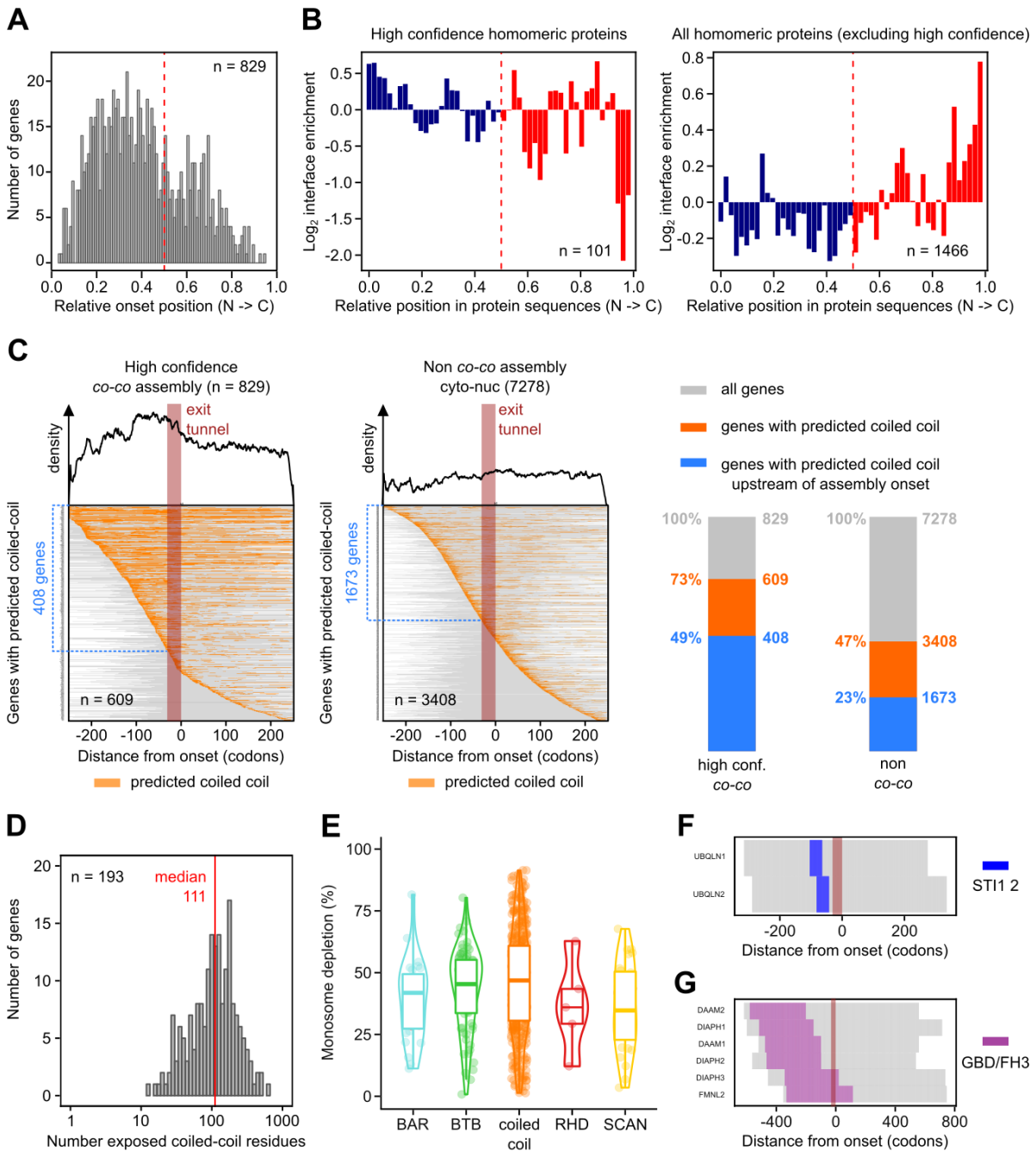


**Fig. S3. Co-co assembly is coordinated with the exposure of N-terminal dimerization domains**

A) Relative onset positions of high confidence co-co assembly candidates. All genes are normalized to the same length. The red dashed line separates the N-terminal and C-terminal halves of proteins.

B) Relative enrichment of segments forming the complex subunit interface for high confidence homomeric complexes (left) or including all homomeric complexes in the human proteome excluding high confidence candidates (right). Interface positions were determined from crystal

structures. All genes are normalized to the same length. Therefore, blue and red bars left and right of the vertical dashed line indicate interface enrichment in the N-terminal and C-terminal halves of proteins, respectively.

C) Left and Center: Heatmaps showing the positions of predicted coiled coils using the DeepCoil algorithm for all proteins in the high confidence proteome (left) and the non co-co assembly proteome (center). Proteins are aligned to assembly onsets determined by DiSP (high confidence proteome, left) or by bioinformatics simulations (non co-co assembly proteome, right (*33*)). 500 residues surrounding the assembly onset are analyzed. Residues left from the highlighted ribosome exit tunnel area (red bar) are exposed at the time-point of co-co assembly.

Right: 73% of high confidence candidates (609 out of 829) contained a predicted coiled coil, compared to 47% of the general proteome (3408 out of 7278). About 49% of all high confidence candidates exposed a predicted coiled coil at assembly onset, compared to about 23% of the general proteome. A higher frequency of exposed coiled coil residues was observed in the high confidence group (39 in median), compared to the non co-co assembly proteome (14 in median). Together, this data indicates that coiled coil exposure is a specific feature of co-co assembly.

D) Distribution of the number of residues involved in coiled coil formation on the ribosome-exposed nascent chains at the time point of assembly. High confidence proteins with annotated coiled coils (according to UniprotKB, see Fig. 4B, left) upstream of assembly onset are included in the analysis.

E) Monosome depletion (%) after onset of co-co assembly reveals variable assembly efficiencies conferred by the five major dimerization domains.

F-G) Heatmaps indicating the STI1 2 (F) and GBD/FH3 (G) domain positions at the assembly onset of high confidence candidates. Residues left from the exit tunnel area (red) are ribosome-exposed.

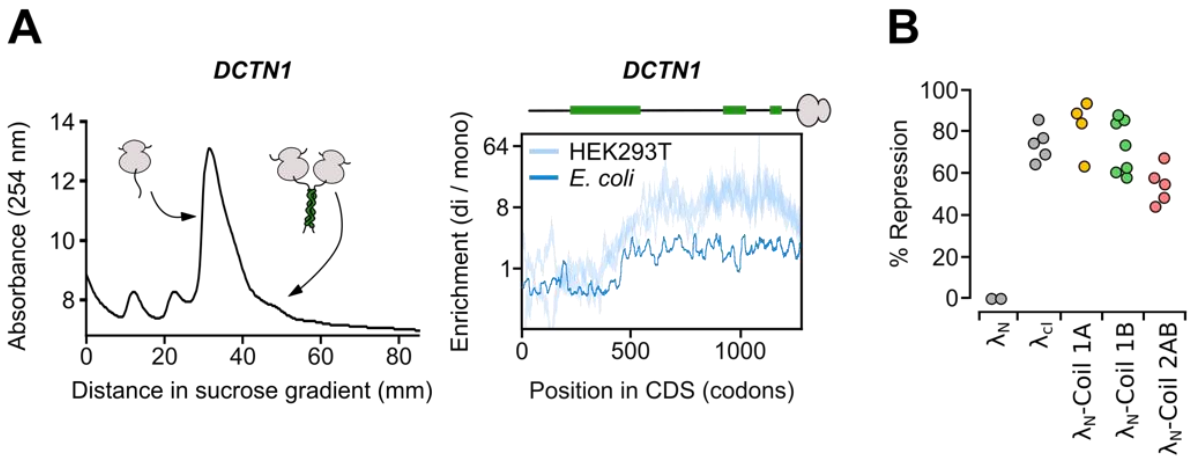


**Fig. S4. Dimerization of human co-co candidates in *E. coli***

A) Left: sucrose gradient centrifugation analysis of *E. coli* expressing plasmid-encoded *DCTN1* (encoding dynactin p150$^{glued}$ subunit). Right: DiSP enrichment profiles (disome / monosome) of *E. coli* expressing *DCTN1* (dark blue) and of human HEK293-T cells expressing endogenously encoded p150$^{glued}$ (light blue). The green boxes in the cartoon indicate the position of coiled coil interfaces on nascent p150$^{glued}$ subunit.

B) The dimerization propensity of individual lamin C rod sub-domains determined in vivo. The dimerization assay employs the monomeric N-terminal DNA binding domain ($\lambda_N$) of the phage lambda repressor protein ($\lambda_{cI}$), which efficiently binds its DNA operator sequence only upon dimerization (*26*). By expressing hybrid proteins consisting of $\lambda_N$ and a C-terminally fused protein or domain in *E. coli* encoding *lacZ* under control of the $\lambda$ promoter, the dimerization propensity of hybrid proteins can be measured. Only dimeric $\lambda_N$ fusion constructs repress *lacZ* expression. Monomeric $\lambda_N$ and dimeric wild-type lambda repressor ($\lambda_{cI}$) serve as control. All $\lambda_N$ fusion proteins enclosing lamin coiled coil segments repress *lacZ* expression, indicating they form dimers in *E. coli*.

**Table S2:** Absolute and relative amounts of proteins annotated as homo-, hetero- or monomeric in each protein class. NA = not assigned, includes proteins annotated as "homo-heteromers" for which the assembly partner is uncertain.

Enrichment of complex subunits (Frequency enrichment, plotted in Fig. 3C) was calculated by dividing the frequency in each assembly class by the frequency in the respective background proteome (cyto/nuclear proteome for the high confidence and total proteome for the low confidence class).

| PROTEIN CLASS | OLIGOMER STATE | ABSOLUTE NUMBER | FRACTION OF PROTEIN CLASS | FREQUENCY ENRICHMENT (plotted in Fig. 3C) |
|---|---|---|---|---|
| High confidence (cyto / nuc) | Homomer | 245 | 0,296 | 1,453 |
| | Heteromer | 267 | 0,322 | 1,315 |
| | Monomer | 246 | 0,297 | 0,678 |
| | NA | 71 | 0,086 | 0,751 |
| Human proteome (cyto / nuc) for normalization of high confidence class | Homomer | 2060 | 0,203 | |
| | Heteromer | 2480 | 0,245 | |
| | Monomer | 4431 | 0,438 | |
| | NA | 1155 | 0,114 | |
| Low confidence | Homomer | 796 | 0,241 | 1,172 |
| | Heteromer | 819 | 0,248 | 1,186 |
| | Monomer | 1291 | 0,391 | 0,884 |
| | NA | 395 | 0,120 | 0,838 |
| Human proteome for normalization of low confidence class | Homomer | 3270 | 0,206 | |
| | Heteromer | 3326 | 0,209 | |
| | Monomer | 7033 | 0,442 | |
| | NA | 2269 | 0,143 | |

**Table S4:** Biotinylated oligos employed for depletion of human rRNA fragments from ribosome profiling libraries.

| ID | Oligos for ligated RNA | Oligos for circ. DNA |
|---|---|---|
| 1 | ACCGGCTATCCGAGGCCAAC | GTTGGCCTCGGATAGCCGGT |
| 2 | GACCGGCTATCCGAGGCCAA | TTGGCCTCGGATAGCCGGTC |
| 3 | CGGCTATCCGAGGCCAACCG | CGGTTGGCCTCGGATAGCCG |
| 4 | CCGGCTATCCGAGGCCAACC | GGTTGGCCTCGGATAGCCGG |
| 5 | CGGGCGCTTGGCGCCAGAAG | CTTCTGGCGCCAAGCGCCCG |
| 6 | CCGGGCGCTTGGCGCCAGAA | TTCTGGCGCCAAGCGCCCGG |
| 7 | CAGACAGGCGTAGCCCCGGG | CCCGGGGCTACGCCTGTCTG |
| 8 | GACGCTCAGACAGGCGTAGC | GCTACGCCTGTCTGAGCGTC |
| 9 | CGACGCTCAGACAGGCGTAG | CTACGCCTGTCTGAGCGTCG |
| 10 | GCGACGCTCAGACAGGCGTA | TACGCCTGTCTGAGCGTCGC |
| 11 | AGCGACGCTCAGACAGGCGT | ACGCCTGTCTGAGCGTCGCT |
| 12 | GACAGGCGTAGCCCCGGGAG | CTCCCGGGGCTACGCCTGTC |
| 13 | GCCGGGCGCTTGGCGCCAGA | TCTGGCGCCAAGCGCCCGGC |
| 14 | CCTCGATCAGAAGGACTTGG | CCAAGTCCTTCTGATCGAGG |
| 15 | GCCTCGATCAGAAGGACTTG | CAAGTCCTTCTGATCGAGGC |
| 16 | TGCGATCGGCCCGAGGTTAT | ATAACCTCGGGCCGATCGCA |
| 17 | CGATCGGCCCGAGGTTATCT | AGATAACCTCGGGCCGATCG |

| | | |
|---|---|---|
| 18 | GCGATCGGCCCGAGGTTATC | GATAACCTCGGGCCGATCGC |
| 19 | GGGCCGGTGGTGCGCCCTCG | CGAGGGCGCACCACCGGCCC |
| 20 | CGGGCCGGTGGTGCGCCCTC | GAGGGCGCACCACCGGCCCG |
| 21 | GACGGCGCGACCCGCCCGGG | CCCGGGCGGGTCGCGCCGTC |
| 22 | ACCGGGTCAGTGAAAAAACG | CGTTTTTTCACTGACCCGGT |
| 23 | ACTCCGCACCGGACCCCGGT | ACCGGGGTCCGGTGCGGAGT |
| 24 | ACAGGCGTAGCCCCGGGAGG | CCTCCCGGGGCTACGCCTGT |
| 25 | ACAGGCGTAGCCCCGGGAGA | TCTCCCGGGGCTACGCCTGT |
| 26 | CGACGGCGCGACCCGCCCGG | CCGGGCGGGTCGCGCCGTCG |
| 27 | AGGACTTGGGCCCCCCACGA | TCGTGGGGGGCCCAAGTCCT |
| 28 | CCGGGTCAGTGAAAAAACGA | TCGTTTTTTCACTGACCCGG |
| 29 | CGGGTCGACTCCGTGTACAT | ATGTACACGGAGTCGACCCG |
| 30 | AGGCCTCGGGATCCCACCTC | GAGGTGGGATCCCGAGGCCT |

**Additional data Tables and supplementary materials (separate files):**

**Table S1:** High and low confidence candidates from HEK293-T cells

**Table S3.1:** Primer sequences used in this study

**Table S3.2:** Sequences used for genome editing

**Table S3.3:** Plasmids generated for this study

**Table S5:** Sequences (5' - 3') of genes that were over-expressed in *E. coli* for DiSP experiments. Each gene sequence is flanked by a short region corresponding to the plasmid backbone. Open reading frames are highlighted (bold). Data analysis included alignment to the *E. coli* genome bearing the relevant gene sequence as an additional chromosome. The indicated gene names are the same as included in the processed HDF5 files.

**Custom Julia Script 1:** Generates a unique molecular identifier (UMI) for each sequenced read by combing the random nucleotides at the 5' and 3' end of the footprint, which are implemented in the library preparation.

**Custom Julia Script 2:** Performs the a-, p- or e-site assignment of reads.

**Custom Julia Script 3:** Sigmoidal fitting algorithm, that estimates the sigmoidal parameters and selects the best model using the Bayesian Information Criterion (BIC).