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 β 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 β 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").

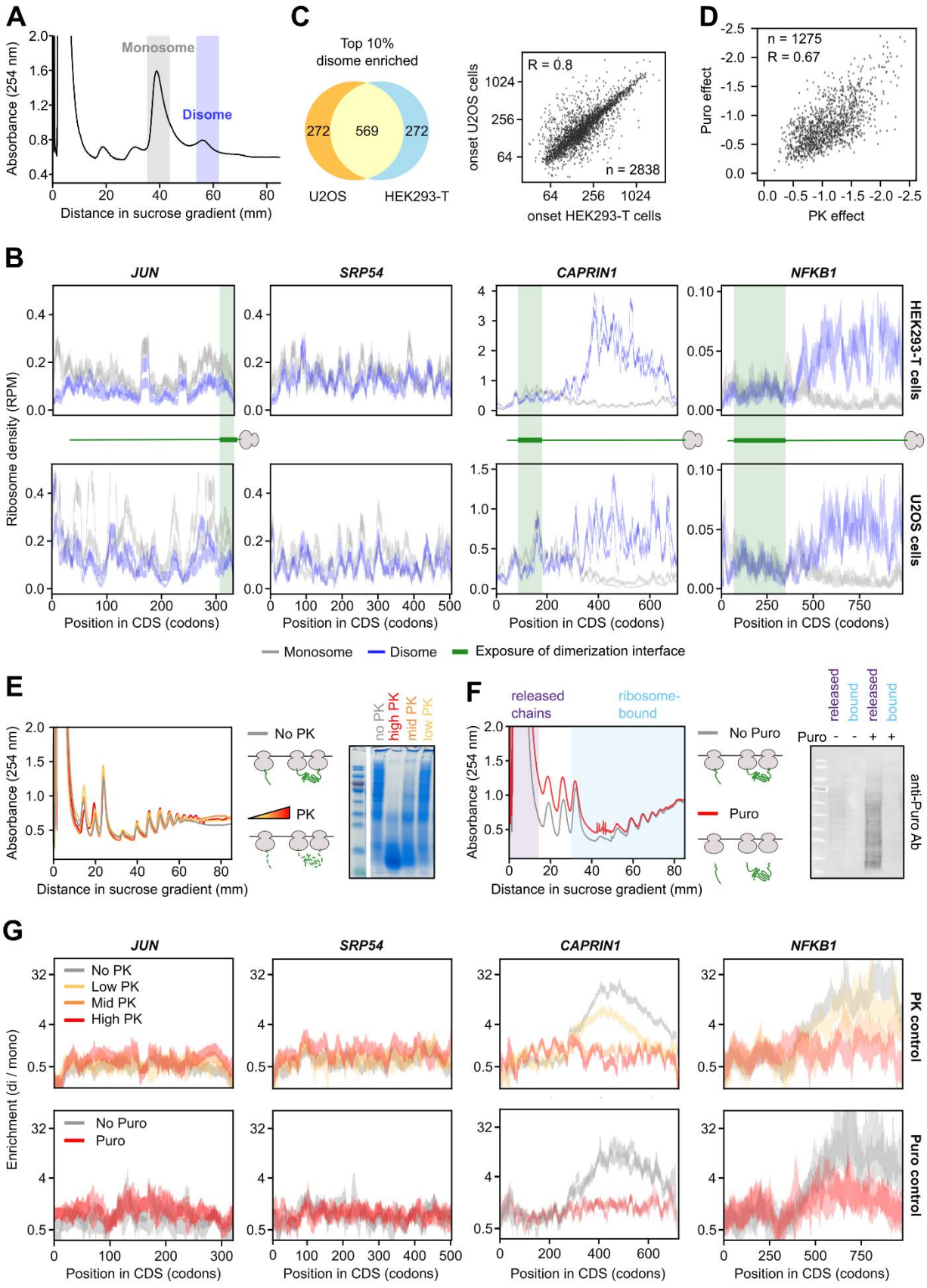
A
Absorbance (254 nm)
Monosome
Disome
Distance in sucrose gradient (mm)
C
Top 10%
disome enriched
272
569
272
U2OS
HEK293-T
R = 0.8
onset U2OS cells
onset HEK293-T cells
n = 2838
D
n = 1275
R = 0.67
Puro effect
PK effect
B
JUN
SRP54
CAPRIN1
NFKB1
Ribosome density (RPM)
HEK293-T cells
U2OS cells
Position 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).

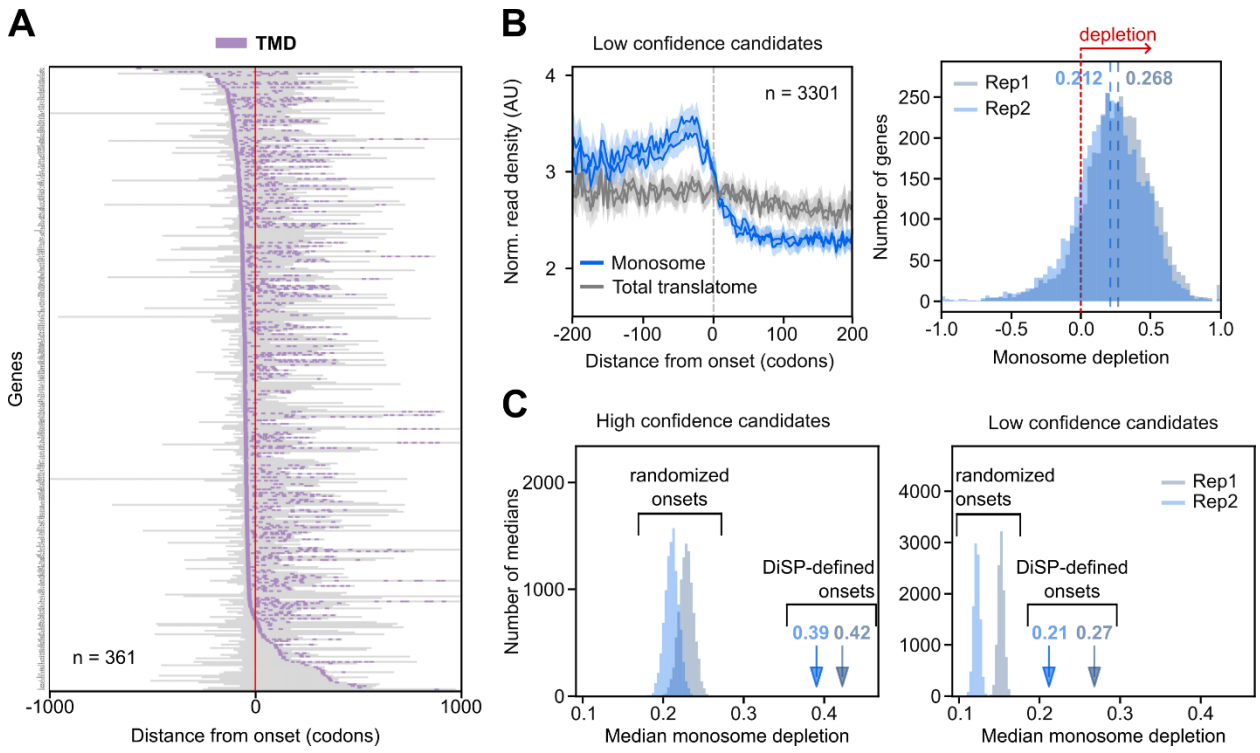

Fig. S2. Features of high and low confidence co-co assembly candidates

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.

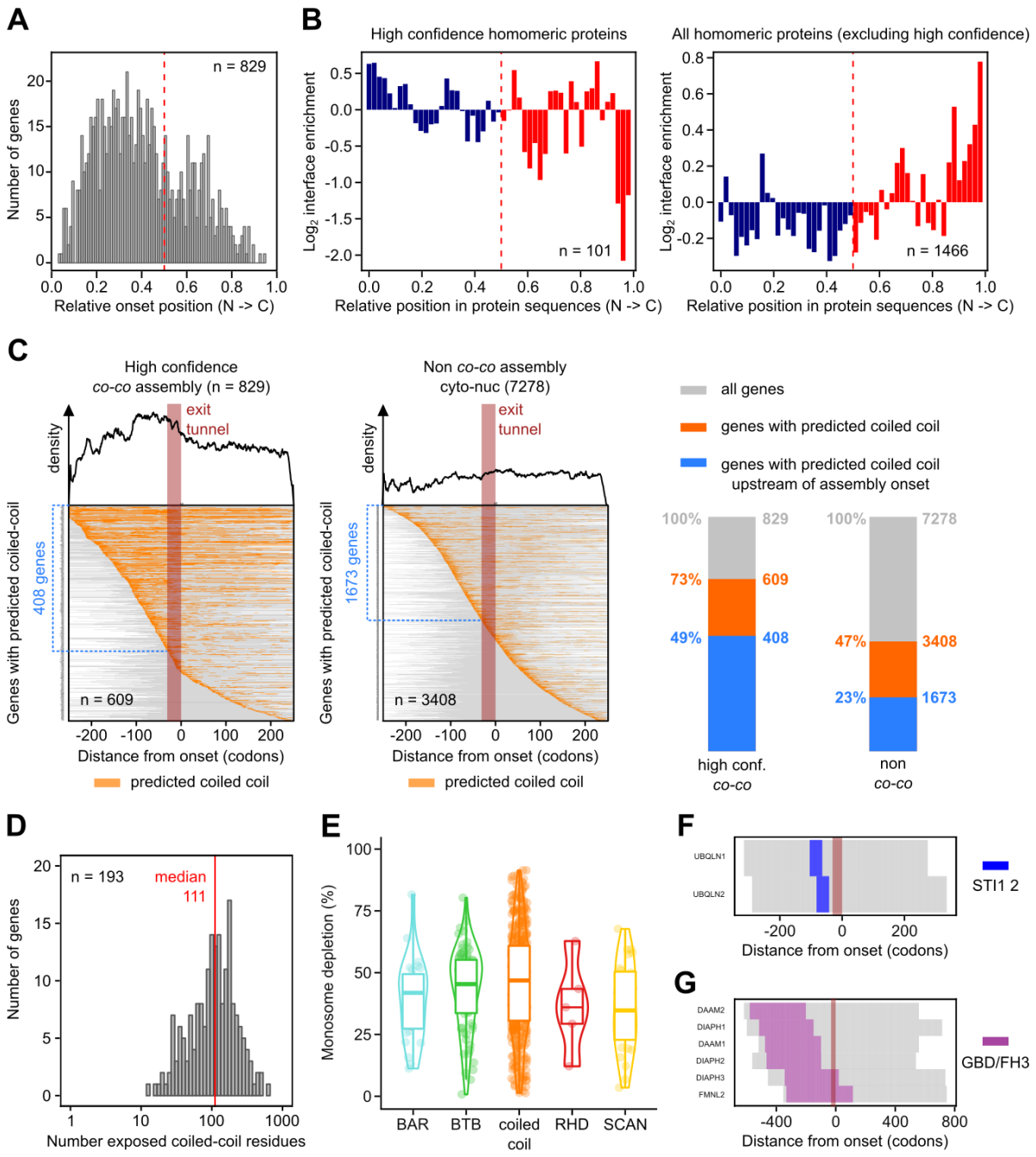

Fig. S3. Co-co assembly is coordinated with the exposure of N-terminal dimerization domains

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.

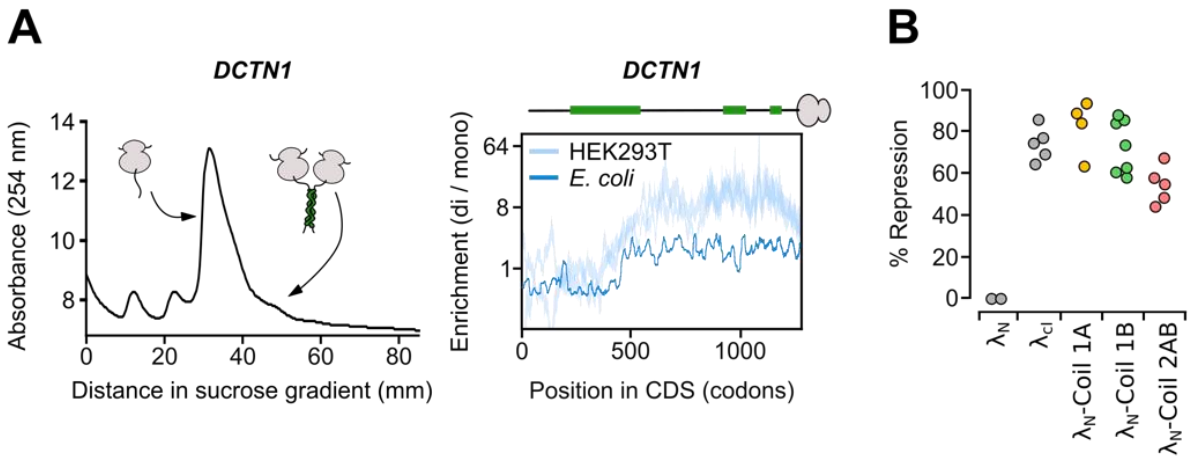

Fig. S4. Dimerization of human co-co candidates in *E. coli*

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 (λ_N) of the phage lambda repressor protein (λ_{cI}), which efficiently binds its DNA operator sequence only upon dimerization (*26*). By expressing hybrid proteins consisting of λ_N and a C-terminally fused protein or domain in *E. coli* encoding *lacZ* under control of the λ promoter, the dimerization propensity of hybrid proteins can be measured. Only dimeric λ_N fusion constructs repress *lacZ* expression. Monomeric λ_N and dimeric wild-type lambda repressor (λ_{cI}) serve as control. All λ_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).