Supporting Information for "Selective Assembly of DNA Conjugated Single-Walled Carbon Nanotubes from the Vascular Secretome"

AUTHOR NAMES. *Xun Gong[1], Anil K. Sharma[2], Michael S. Strano[3], Debabrata Mukhopadhyay[1,2].*

AUTHOR ADDRESS.

[1]Department of Physiology and Biomedical Engineering, Mayo Clinic College of Medicine, Rochester, MN 55904, USA

[2]Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine, Rochester, MN 55904, USA

[3]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*E-mail: Mukhopadhyay.Debabrata@mayo.edu Office Phone: 507-538-5381. Fax: 507-293-1058

SUPPORTING INFORMATION

A. SUPPLEMENTAL FIGURES

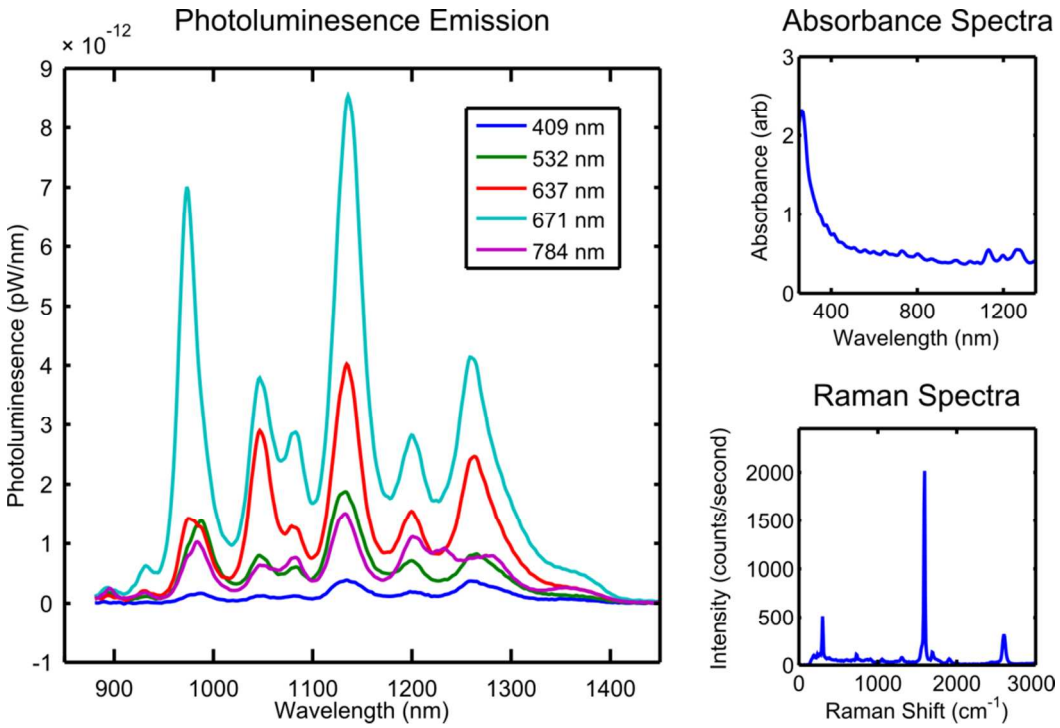

Figure S1. Nanotube Sample Characterization. (A) nIR spectra of AT_{15}-SWCNT from 5 laser excitations show E_{11} emissions of singly-dispersed SWCNT with high optical efficiency. (B) Absorbance spectra of AT_{15}-SWCNT show peaks corresponding to E_{22} and some of the E_{11} absorbance of different SWCNT chiralities. (C) Raman spectra of AT_{15}-SWCNT.

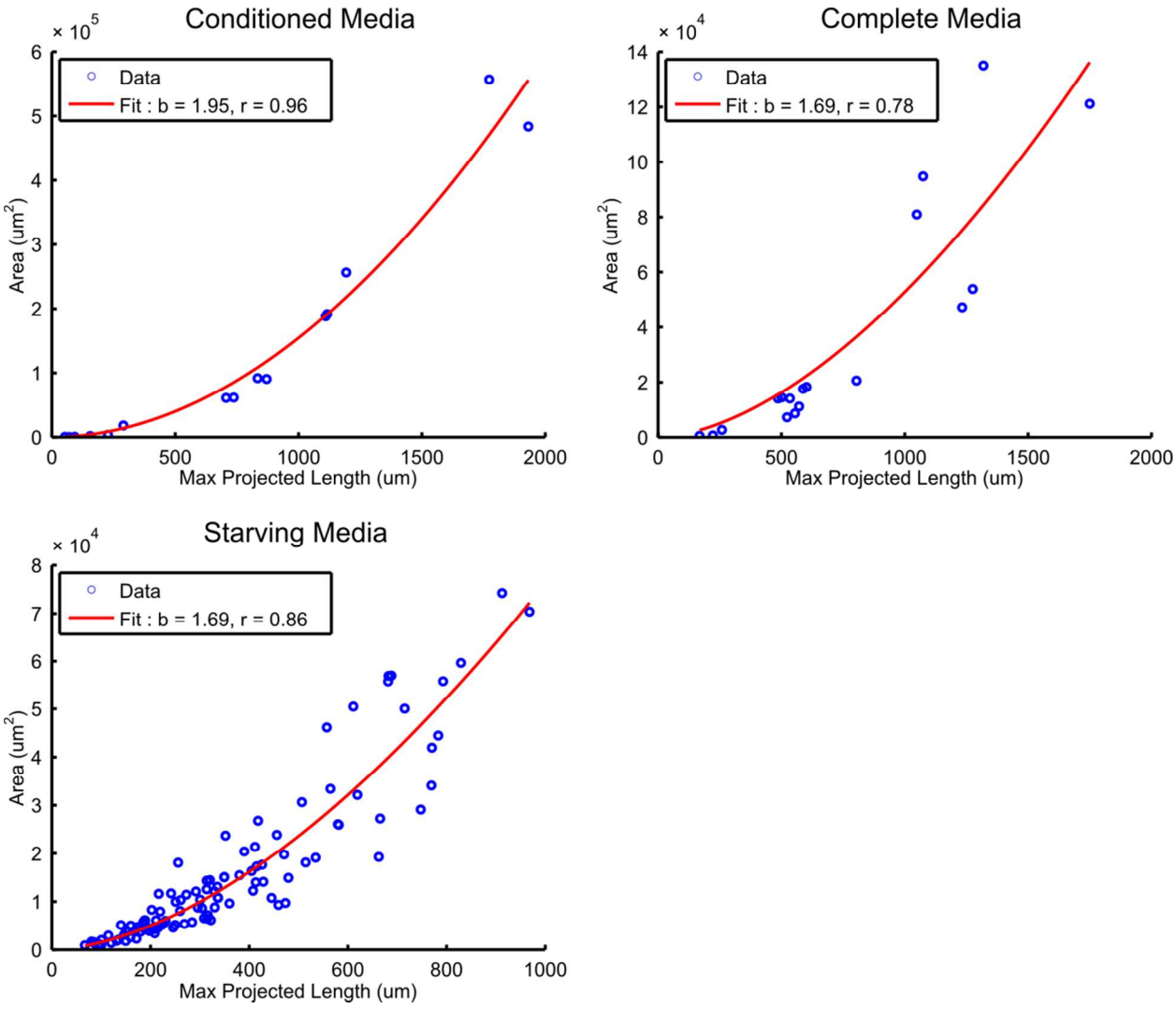

Figure S2..Fractal Dimension Fitting. Exponential fitting of fractional occupation area versus maximum projected length collected from processed bright-field images of aggregates to determine D_2 fractal coefficients. The r-values of fit are comparable with literature.[1]

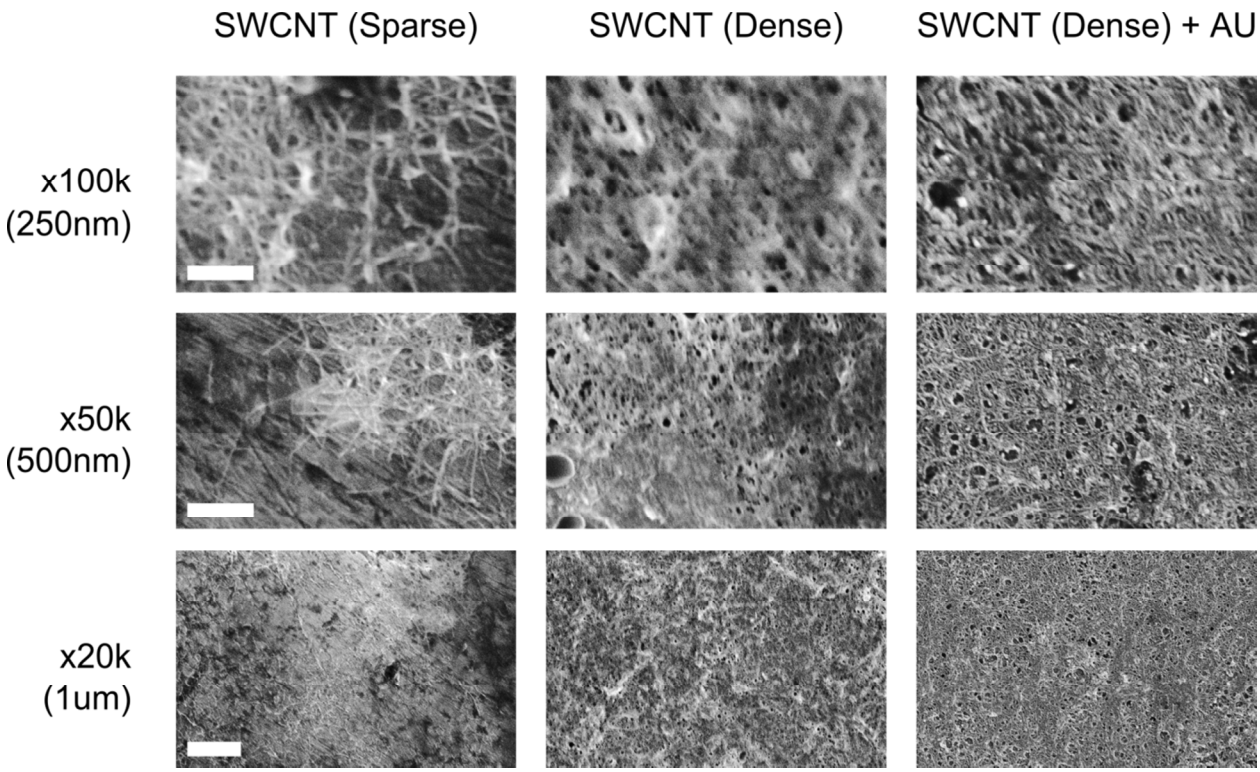

Figure S3. **SEM Control.** Control SWCNT deposited on aluminum stub showed no presence of proteins. Gold coating revealed more densely packed SWCNT with globular DNA wrapping. Sparse SWCNT Deposition is used to enhance visualization of individual tubes. nSWCNT unassociated with proteins pack more tightly in the dried state, which explains the differences observed in Raman RBM.

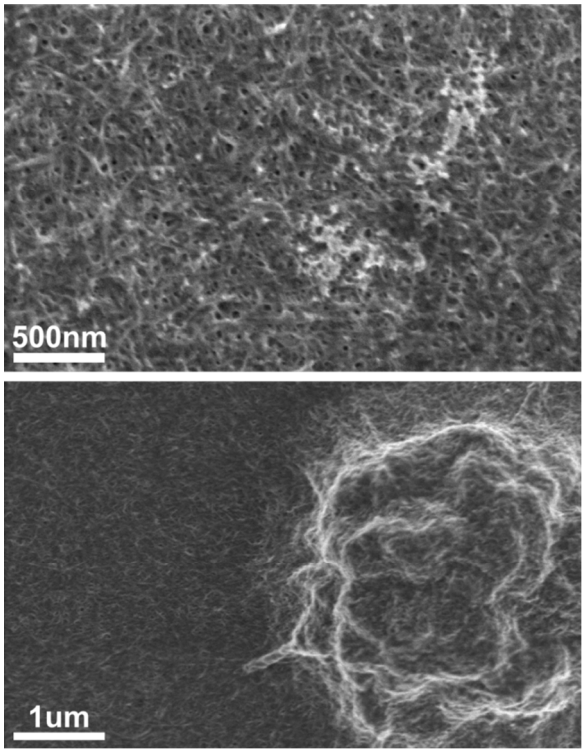

Figure S4. **SEM Protein-SWCNT Signal.** SEM without gold coating show unique contrast in protein scattering patterns. Interweaved structures between proteins and SWCNT were observed. This is another observation that protein association played a role in aggregate macrostructure organization.

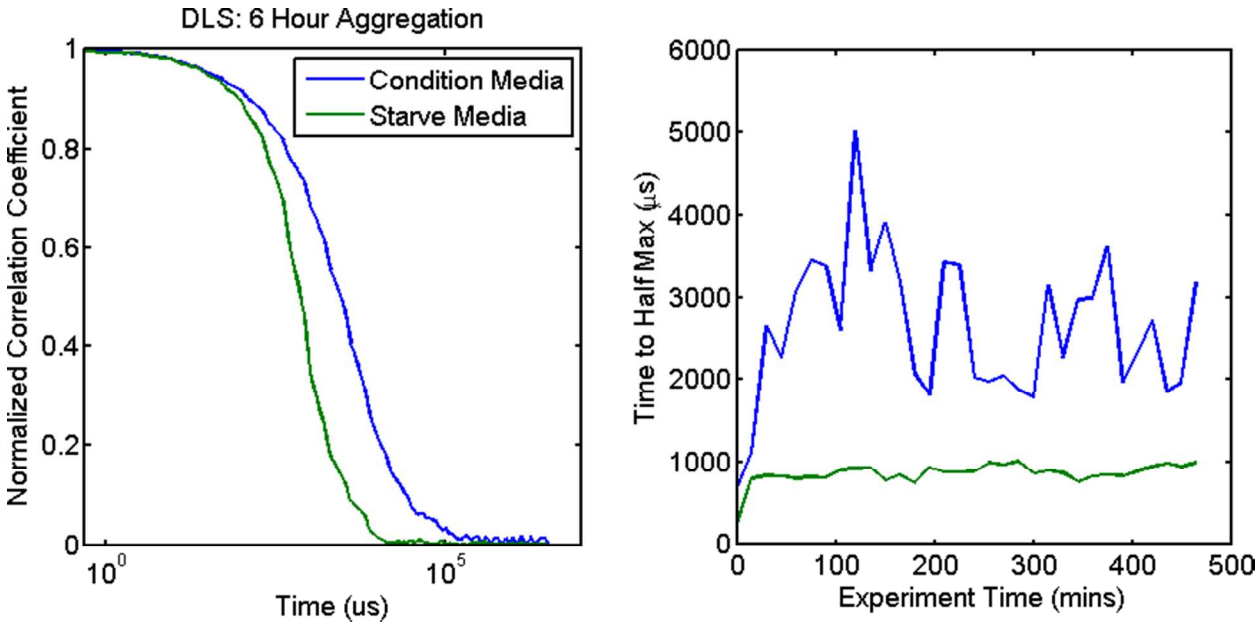

Figure S5. DLS Aggregation. (A) After 6 hours of aggregation, conditioned media showed significant right shift of normalized correlation coefficient versus time curve, indicating an increase in particle size. (B) Time to half max of the correlation curves were plotted for the duration of the 6 hours of aggregation, showing distinct particle size differences between control and conditioned media. Noise in the conditioned media experiment can be attributed to high sample polydispersity.

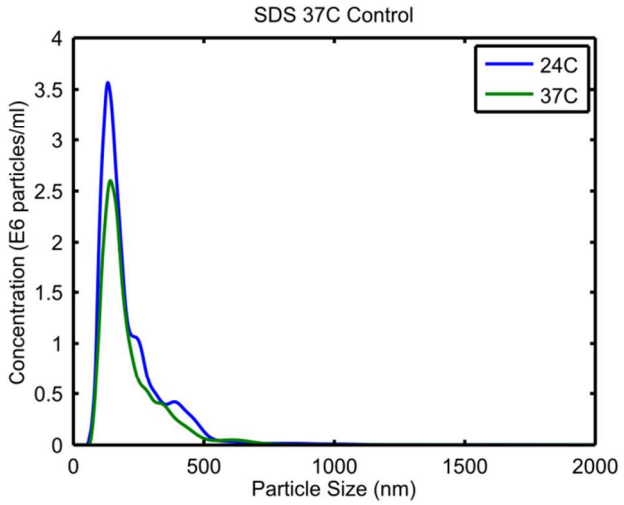

Figure S6. NTA SDS Aggregation Control. SDS-SWCNT following temperature change did not exhibit significant size distribution right shift. The fact that SDS-SWCNT did not undergo oligomerization reinforces the hypothesis that DNA is responsible for the observed temperature dependent phenomena.

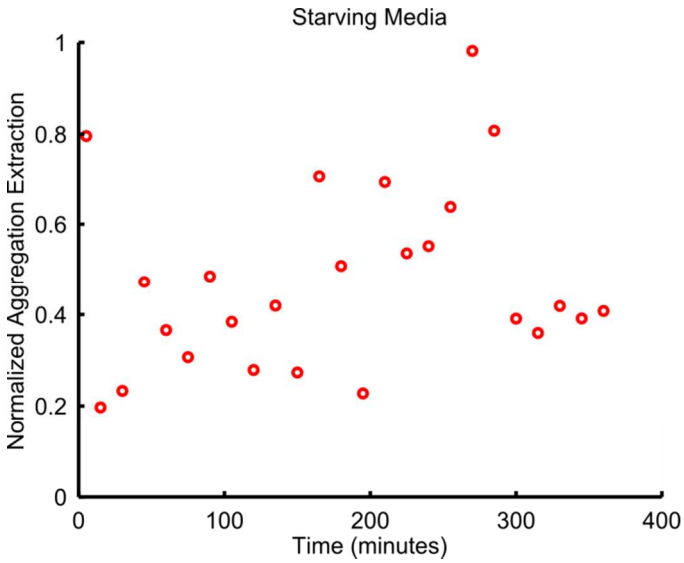

Figure S7. NTA Aggregation Kinetics Control. Starving media control of aggregation kinetics does not show strong decreasing trend in total SWCNT volume. As temperature changes, different intermediate sized aggregates are formed and were sampled within the detection area.

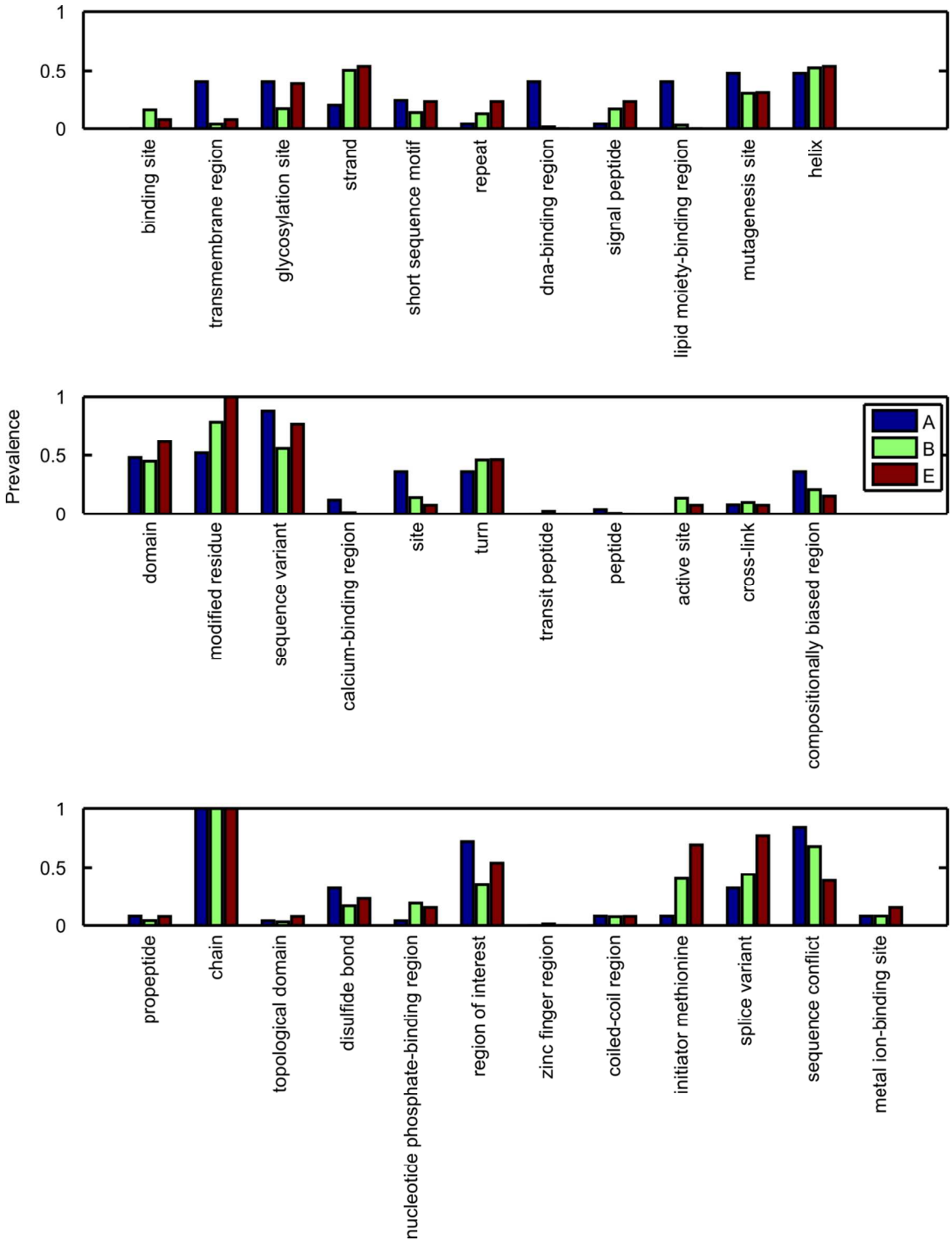

Figure S8. MS Functional Group Comparison. Functional group comparison showed that species within the high molecular weight aggregated protein bands contain preference for

transmembrane regions, lipid moiety binding regions and DNA binding regions. Similar to the major extracted species, these preferences are potentially due to hydrophobicity of the SWCNT and presence of the DNA wrapping.

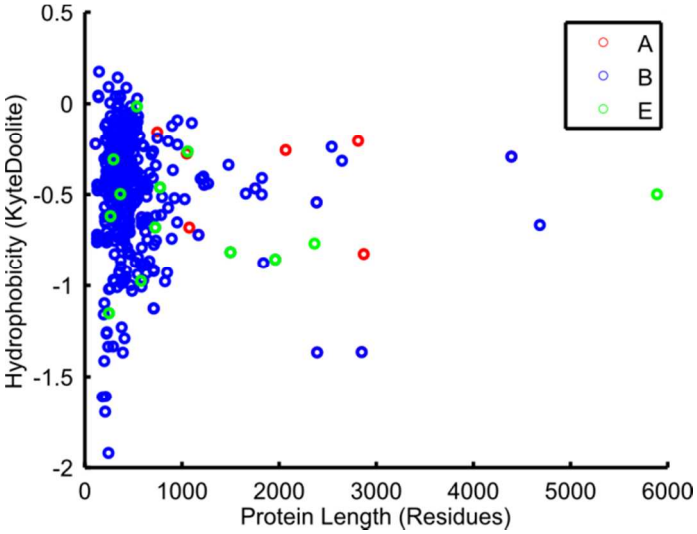

Figure S9. Hydrophobicity versus Protein Length. Hydrophobicity versus length for detected protein species in the three different groups were scatter plotted for visualization.

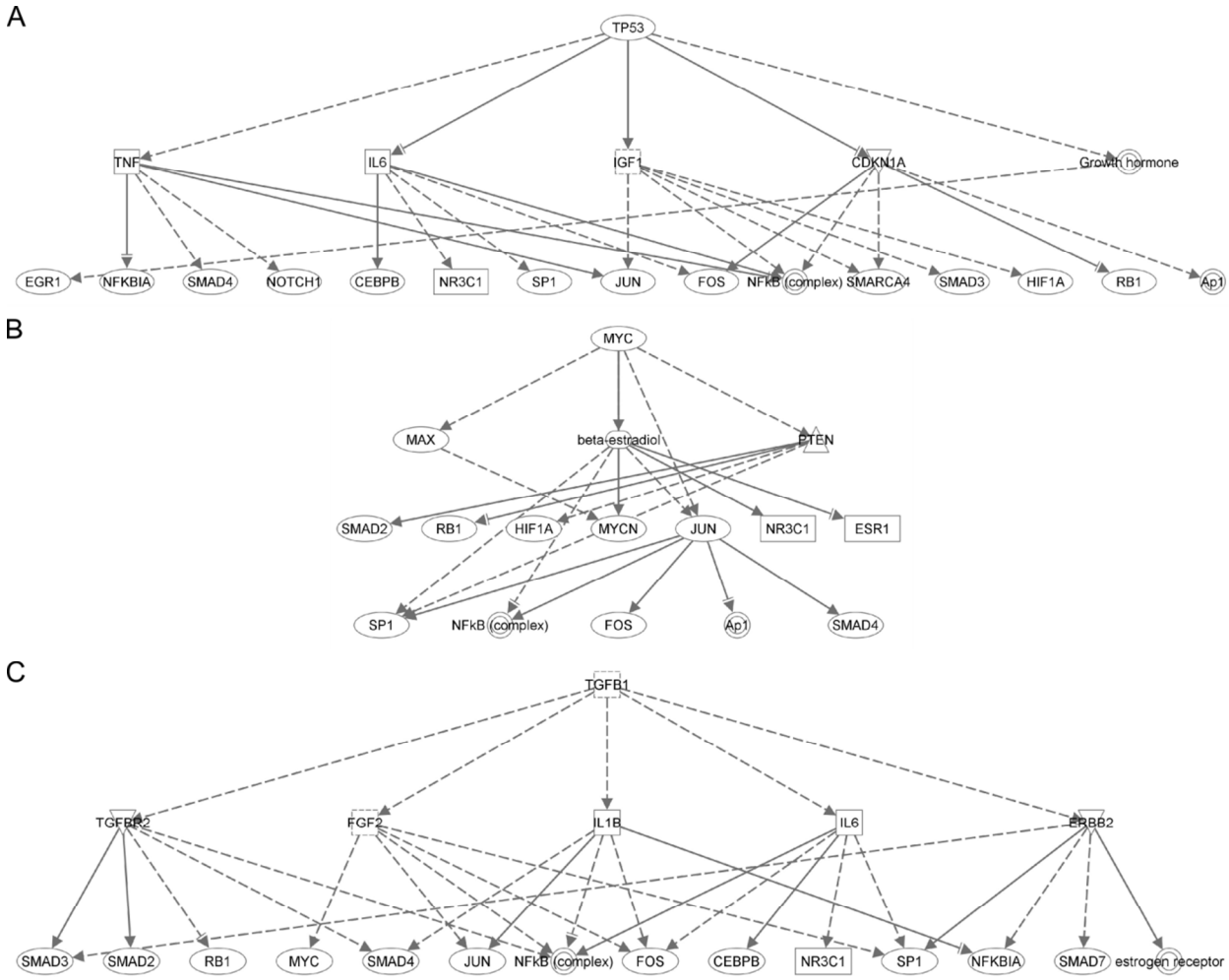

Figure S10. Network Analysis. Three major networks and upstream regulators of proteins detected within the aggregates are diagramed.

B. SUPPLEMENTAL Tables

Table S1: Top Canonical Pathways

Name	p-value
Protein Ubiquitination Pathway	2.40E-12
Remodeling of Epithelial Adherens Junctions	2.46E-11
Glycolysis	2.58E-10
IGF-1 Signaling	8.22E-10
Caveolar-mediated Endocytosis Signaling	9.85E-10

Table S2: Associated Network Functions

Name	Score
RNA Post-Transcriptional Modification, DNA Replication, Recombination, and Repair, Protein Synthesis	57
Infectious Disease, RNA Post-Transcriptional Modification, Connective Tissue Disorders	51
Cardiovascular System Development and Function, Cellular Development, Cellular Growth and Proliferation	50
Dermatological Diseases and Conditions, Hereditary Disorder, Organismal Injury and Abnormalities	46
Carbohydrate Metabolism, Nucleic Acid Metabolism, Small Molecule Biochemistry	44

Table S3: Molecular and Cellular Functions

Name	p-value	#molecules
Cellular Growth and Proliferation	1.66E-33	250
Cell Death and Survival	1.99E-31	247
Cellular Assembly and Organization	1.37E-23	158
Cellular Function and Maintenance	1.37E-23	155
Cellular Movement	7.03E-21	160

Table S4: Physiological System Development and Function

Name	p-value	#molecules
Organismal Survival	2.15E-14	157
Cardiovascular System Development and Function	3.57E-12	97
Organismal Development	1.24E-10	146
Tissue Development	1.26E-10	144
Skeletal and Muscular System Development and Function	5.85E-10	33

C. SUPPLEMENTAL METHODS

C.1 Bright-field Image Collection, Segmentation and D_3 Calculation. Bright-field images were taken in a manner such that a semi-automated processing code can be applied to extract statistics of individual aggregates. Code performing the following computations were written in Matlab.

1. *Image Acquisition.* An area of solution containing aggregates was found with the 5x objective. A low 1 ms camera acquisition time was chosen so that solution movements do not affect image quality. Illumination (2.7V in the setup) was calibrated *via* histogram to give the maximum amount of contrast without saturating pixels. An optimal number of z-slices were acquired such that for each aggregate there exists an image that contains the largest cross section of that aggregate in focus.
2. *Aggregate Segmentation.*
    a. Individual optimal detection. The above method can detect tens of aggregates with a single image z-stack, significantly improving data collection. Since the aggregates remain in suspension, finding the image with the best contrast for each aggregate is necessary. To determine the approximate location and boundary of each aggregate, image segmentation was performed on the maximum intensity projection of the z-stack. Any background trend was first subtracted *via* an image opening operation, and then a smoothing filter was applied to removed high frequency noise before a threshold was applied to construct a segmentation mask. The area in which individual aggregates exist were identified, and the optimal z-slice for each particular aggregate was found *via* maximizing a steerable filters-based contrast measure.[2]

b. Fine segmentation. Using the optimal image found for each aggregate, the same segmentation method is applied again. In this case the threshold value was chosen to accurately detect any internal and external aggregate boundaries without bias.

3. *Extraction of Parameters*. Two parameters are needed to calculate D_2, the maximum projected length (L_P) and the area of occupation (A_O). L_P was calculated by identifying the external border pixels of the fine segmentation mask, and then calculating pair-wise distances to find the maximum length. To calculate A_O, we devised a normalized approach. As all images were taken together with the same illumination, pixel intensities were comparable. Each pixel represents the amount of SWCNT within an area and light transmittance decreases with increasing SWCNT due to the material absorbance. The images were inverted, and then normalized to the pixel with the maximum amount of SWCNT, which we define as 100% SWCNT occupancy of the pixel area. The A_O of each aggregate was then found by division of the sum of total SWCNT occupancy by the total segmented pixel area. The entire process was automated to avoid operator bias.

C.2 SWCNT Parameter Calculations

Approximate Length Estimate from HD. Through linear fitting, we found that each SWCNT contributes on average 183 nm HD aggregate size. To convert this to an approximate SWCNT length, we combine Eqn. 3 and Eqn. 4 to solve for the relationship between HD and rod length.

$$L = 2r\ln(p) + C \tag{S1}$$

L is rod length, r is half of HD and p is aspect ratio which is L/nSWCNT radius. Using 5 ± 2 nm as the DNA coated SWCNT radius (approximated by AFM and take from previous works),[3] - 0.614 as C,[4] we then solve the equation for L, which we determine to be 815 ± 73 nm. Error propagation is calculated with equation S2, where n is the DNA coated SWCNT radius.

$$\delta L = \left[\frac{dL}{dn}\right] \delta n \qquad \text{(S2)}$$

This is an over-estimate because SWCNT not a rigid rod, as observed in SEM, which would significantly decrease the aspect ratio subsequently L. However, knowing that the SWCNT material ranges in length from 100-1000nm, the calculated result is reasonable.

C.3 Mass Spectroscopy Analysis (Full MS Protocol).

Mass spectrometry of the samples was performed by nano-flow liquid chromatography electrospray tandem mass spectrometry (nanoLC-ESI-MS/MS) using a Thermo Scientific Orbitrap Elite Hybrid Mass Spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to a Thermo Ultimate 3000 RSLCnano HPLC system. The resulting Tandem mass spectra were extracted by msconvert (version 3.0.4019; ProteoWizard) and all MS/MS samples were analyzed using Mascot (Matrix Science, London, UK; version 2.4.0), Sequest (Thermo Fisher Scientific, San Jose, CA, USA; version 27, rev. 12) and X! Tandem (The GPM, thegpm.org; version CYCLONE (2010.12.01.1)). Mascot, Sequest, and X! Tandem were set up to search the UniProt June 2013 human reference protein database (175236 entries), including a decoy reverse database and assuming the digestion enzyme trypsin. Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10.0 PPM. Sequest was searched with a fragment ion mass tolerance of 0.60 and a parent ion tolerance of 0.01 Da.

Oxidation of methionine and iodoacetamide derivative of cysteine were specified in Mascot, Sequest and X! Tandem as variable modifications. Scaffold (version Scaffold_4.2.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they can be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm.[5] Protein probabilities are assigned by the Protein Prophet algorithm.[6] Proteins that contained similar peptides and can't be differentiated based on MS/MS analysis alone are grouped to satisfy the principles of parsimony.

D. SUPPLEMENTAL REFERENCES

1. Lee, C.; Kramer, T. A. Prediction of Three-Dimensional Fractal Dimensions Using the Two-Dimensional Properties of Fractal Aggregates. *Adv. Colloid Interface Sci.* 2004, 112, 49-57.
2. Minhas, R.; Mohammed, A.; Wu, Q. M. J.; Sid-Ahmed, M. 3d Shape from Focus and Depth Map Computation Using Steerable Filters. In *Image Analysis and Recognition*, Kamel, M.; Campilho, A., Eds. Springer Berlin Heidelberg: 2009; Vol. 5627, pp 573-583.
3. Cherukuri, T. K.; Tsyboulski, D. A.; Weisman, R. B. Length- and Defect-Dependent Fluorescence Efficiencies of Individual Single-Walled Carbon Nanotubes. *ACS Nano* 2012, 6, 843-850.
4. Broersma, S. Viscous Force and Torque Constants for a Cylinder. *J. Chem. Phys.* 1981, 74, 6989-6990.
5. Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by Ms/Ms and Database Search. *Anal. Chem.* 2002, 74, 5383-5392.
6. Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* 2003, 75, 4646-4658.