Title: Accelerating biological understanding for genes of unknown function

Writing Team: Kimberly A. Reynolds (kimberly.reynolds@utsouthwestern.edu), A. Mark Settles (settles@ufl.edu), Rob Ward (robward@ku.edu), Hongbin Zhang (hbz7049@tamu.edu), Bree Urbanowicz (breeanna@uga.edu), Eduardo Rosa-Molinar (erm@ku.edu)

Summary: The rapid expansion of genome sequence data is increasing the discovery of functionally unannotated protein-coding genes across all domains of life. Annotating these genes with reliable functional information is necessary to define the full biochemical space accessed by nature and to understand how taxonomically diverse organisms function and evolve. Recent advances in comparative statistical genomics, machine learning, cell biology, microscopy, genetics, and high-throughput biochemistry suggest new potential for the characterization of proteins and domains of unknown function (so-called PUFs and DUFs). By coupling interdisciplinary teams of informaticians, cell biologists, geneticists, and biochemists with well-maintained, open access databases of function and phenotype, it should be possible to dramatically accelerate annotation of the so-called “dark proteome”.

Introduction and Statement of Problem:

In the last 20 years, our capacity to sequence genomes has grown exponentially. At present, NCBI catalogs well over 200,000 genomes, assembled at varying levels of completeness (http://www.ncbi.nlm.nih.gov/genome/browse/). Despite this rich catalog of genetic information, the function of the vast majority of the genes contained in a species remain unknown. Recent estimates suggest that depending on the organism, anywhere from 20-60% of genes are unannotated (1,2). These numbers do not fully account for the potential for incomplete or mis-annotated genes. Further, this challenge is dramatically higher for Intrinsically Disordered Proteins -- estimated to comprise 15-40% of human proteome (3) -- as variations in sequence length and composition often preclude alignment, and the lack of a folded domains prohibits the use of structural information.

Providing accurate functional and phenotypic information for individual proteins is nevertheless critical to discover new chemistry and cellular processes. It is also fundamental o understand the mapping between genotype and phenotype. Traditionally, protein function is defined through a one-gene-at-a-time approach requiring years of intensive research to associate a biochemical, cellular, and organismal role for a protein. Even with intense investigation, consensus may not be reached on an individual protein’s function. Computational approaches aim to expand annotation capacity, but simple homology-based predictions often provide limited or misleading insights into function. Therefore, it is imperative to develop technologies or approaches to rapidly determine the biochemical and biological functions of functionally unknown genes.

We describe recent advances and challenges in function annotation and lay out a potential pipeline for prioritizing and characterizing PUFs and DUFs. This process demands collaborations between computational biologists and experimentalists from diverse biological disciplines. Such interdisciplinary collaboration is necessary to: 1) maximally leverage the experimental data to develop and refine computational tools, 2) create high-dimensional annotations that describe multiple aspects of protein function and 2) computationally extend annotations to experimentally inaccessible systems. Importantly, this approach is necessarily

species agnostic, but will provide cascading benefits as more genomes are sequenced and annotated.

Key Advances and Challenges:
Recent years have brought a host of advances in both computational and experimental approaches to the annotation of function. Here we highlight several that seem exceptionally promising and discuss present barriers and challenges.

Informatics: The rapid increase in genomic "big data" provides rich opportunities for prediction of function by statistically-driven evolutionary analyses and machine learning. Signals of sequence homology, conserved chromosomal proximity (e.g. gene neighborhoods, synteny), the presence/absence of genes (e.g. phylogenetic profiling), and protein sequence co-evolution (e.g. evolutionary rate correlations, mirrortree) can all provide predictions of function and functional interactions (4–12). As more protein structural data becomes available, homology-based structural approaches (e.g. HHPRED) will become increasingly powerful (13). Advances in machine learning are also dramatically increasing the potential of computational annotation (14). It is further interesting to consider the possibility of functional annotations for units larger than a single gene: co-evolutionary analyses suggest the possibility that cellular systems might be decomposed into small multi-gene functional units (6,15). In this sense, we might recognize or annotate a small group of genes as a physical complex or the unit underlying a particular phenotypic trait.

However, these approaches suffer from a few limitations. First, gold-standard training and test set data are often limited. Expanding the availability of high-quality experimental functional annotations is important to refining and validating the performance of many computational algorithms. For example, the community-driven CAFA challenge (Critical Assessment of Functional Annotation), has been valuable in benchmarking progress and identifying directions requiring further work (16). Increasing the annotations available to CAFA would further accelerate progress. Secondly, computational annotation methods are typically used by a limited set of expert practitioners, and analysis results (and code) are sometimes inaccessible or restricted in scope. It is often difficult for experimental biologists to know which tools are the most cutting-edge and reliable, or assess which would be the best for their application. In this direction, a combination of educational workshops, open-access initiatives, web-based annotation tools, and collaborative, review-like publications are necessary to broaden and democratize usage.

Genetics: One standard approach to assessing function is creating genetic knockouts or knockdowns, and observing the resulting functional phenotype. For example, recent work in bacteria used transposon-based mutagenesis and high-throughput measurements of growth rate to measure fitness effects for more than 10,000 genes in 32 diverse bacterial species (17). In these types of studies, characterizing phenotypes across many environments or growth conditions is important to expose as many functional roles as possible (17–20). To this end, recent advances in multiplexed continuous culture devices provide one avenue to broadly sample well-controlled environments for bacteria and yeast (21,22).

Increasing the throughput of “knock-in” or gene complementation studies is another interesting possibility: assessing the capacity of DUFs and PUFs to rescue loss-of-function knock-out phenotypes in model organisms presents an additional avenue to rich functional information (23,24). Existing gene knockout collections in several model organisms (E. coli, yeast, Drosophila, Arabidopsis, and mammalian cell culture lines) provide a solid foundational toolkit for these sort of genetic studies, while new CRISPR-based tools (allowing both loss-of-function and gain-of-function gene expression studies) enhance model system genetic analysis and provide powerful tools for generating genetic perturbations in non-model organisms (25).

Nonetheless, genetic manipulation in most cell or organism-level systems is arduous. Model species contain only a small subset of the diverse protein coding sequences that have been discovered through genome sequencing. Unknown function genes may be identified in organisms that cannot be grown in a laboratory, making these species inaccessible for genetic manipulation. Moreover, genetic redundancy within individual species can require whole gene families to be mutated before any discernible phenotype can be observed. Technologies that enable large scale, site-directed mutagenesis in a diversity of species are necessary to characterize unknown function genes at scale. Importantly, these genetic tools must be coupled with high-throughput, quantitative measurements of phenotype.

Biochemistry: Annotation of biochemical function -- including the reactions catalyzed, associated catalytic parameters, and knowledge of substrate specificity or ligand binding affinities -- is fundamental to identifying new chemistry, defining metabolic pathways, and discovering drug targets. Nonetheless, expressing, purifying and biochemically characterizing individual proteins is a labor-intensive and typically low-throughput process. Combining comparative genomics, structural modeling and high-throughput biochemical assays suggests a path towards broadly defining enzyme function (26,27), but requires additional scaling to keep pace with the expansion in genomic datasets.

Recent advances in microfluidics based assays, which can enable measurements of catalytic parameters for tens of thousands of enzyme variants, provide one strategy for rapidly characterizing homologs or mutants. In these systems, proteins are often transcribed and translated *in vitro*, and measurements of catalytic activity, binding affinity and substrate specificity are collected within isolated microfluidic chambers or droplets (28,29). However, these systems are presently limited to relatively well-behaved model enzymes, and typically require known fluorescent substrates. Extending the utility of these approaches would broadly enable the fields of function annotation and protein engineering.

Moreover, we emphasize that the resulting data should be collected in an accessible public database in machine-readable format with appropriate meta-data. The BRENDA database (https://www.brenda-enzymes.org/) provides one existing platform for accomplishing this, but the data are not easily searched or downloaded in a readily machine-readable format. KEGG provides an organized view of metabolic enzymes and pathway structure

(https://www.genome.jp/kegg/), but does not make biochemical parameters for enzyme orthologs readily available.

Imaging: Imaging and analyzing the spatial~~,~~ intra- and inter-cellular PPI of the "dark matter proteome" will be incredibly difficult because of its **1)** low numbers, **2)** transient nature, **3)** poor to no evolutionary conservation, **4)** lack of whole or regional protein sequence information, **5)** weak to no PPIs, and **6)** detectability (30,31,32) and because all reagents and technologies have been developed for highly abundant, well characterized, ordered, strongly interacting, and highly detectable proteins (33-36).

For example, the issues involved in using commercially available antibodies generated for PPI studies included, high quality antibodies required for antibody-dependent methodologies and workflows may be poorly characterized, may have been developed and used for single event detection and analysis, or they are simply unavailable; similarly, methodologies and workflows required to study PPIs are hampered by a lack of controls required to assure optimization, validation, and reproducibility (37-43).

Although advances in new tools such as CRISPR-mediated gene editing coupled with generic fluorescent tag-donor plasmids are very promising for PPI studies, new high throughput tools for effective tagging endogenous proteins are needed (44).

Similarly, academic entrepreneurial microscopists' development of new optical and electron microscopic imaging reagents and technologies and using them in combination with common and uncommon optical and electron microscopic-based methods and technologies have permitted the discovery, study, and quantitative analysis of dark matter proteome PPIs and will promote and further efforts to connect genotypes-to-phenotypes and disruptions that often lead to cellular dysregulation (45,46).

However, all combinations of new reagents and technologies with common (*e.g.,* fluorescence cross-correlation spectroscopy and Förster resonance energy transfer [FRET]) and uncommon (single-molecule FRET; FRET-fluorescence lifetime imaging microscopy; cytoskeleton-based assay for protein-protein interaction; single-molecule protein proximity index; concentric FRET; homogeneous time resolved fluorescence; acceptor photo-bleaching FRET, and correlative acceptor photo-bleaching FRET) optical and electron microscopic-based methods and technologies have pitfalls when used in the search for dark matter proteome PPIs (47-53).

These methods and technologies 1) are labor intensive; 2) make it difficult to differentiate and determine proximity among direct PPIs; 3) are poorly scalable; 4) require many negative controls, and 5) intensive optimization and trouble-shooting; and all require better probe design for longer acceptor half-live and lower time-resolution, as well as smaller size probes (47-53).

Thus, to advance PPI studies of the dark matter proteome, imaging obstacles must be overcome. A promising direction is the development of correlating FRET-based multiplexed techniques that use multiple donors/acceptor pairs and multiple excitation sources to monitor multiple events and to permit correlation of multiple *in vivo* spatiotemporal characteristics of PPI between dark matter proteome proteins within intact multicellular organism at multiscale resolution (54-58).

This section is not intended to provide an exhaustive review of peer reviewed literature, relevant books, and reports. Rather, it seeks to provide some understanding of and insight into PPI

studies' available imaging tools and technologies and their pitfalls and to point to what new technologies and tools will advance PPI studies of the dark matter proteome.

A multidisciplinary pipeline.
Annotating protein function is a long-standing problem, and is well assed by establishing a multidisciplinary, community-driven, collaborative pipeline (2,59,60). Given the large scope of the problem, here we sketch out a proof-of principle pipeline for functional annotation that could be accomplished today with current technologies. The first step is to generate a prioritized list of DUFs and PUFs; we suggest that proteins with homologs in numerous taxa from all three kingdoms are an excellent starting point. Next, informaticians can generate rough functional annotations or hypotheses using co-evolutionary studies, correlative gene expression patterns with known genes in multiple taxa/conditions, correlative protein interactions studies (i.e. yeast two-hybrid) with known proteins in multiple taxa/conditions, etc. This subset of genes will then be assayed using a variety of approaches to assign function. For example, growth rate complementation in bacteria and yeast provides one strategy to identify genes encoding proteins that likely have essential biochemical functions. These studies could be further coupled with metabolomic profiling. Fluorescently-tagged proteins will be generated and coupled with advanced imaging technologies to address questions of subcellular localization, tissue distribution (in multicellular organisms) and colocalization with landmark proteins. Loss of function studies in model multicellular organisms (e.g. Arabidopsis, Drosophila, C. elegans) using existing reagents or with CRISPR interference (CRISPRi) will be used to reveal developmental and signaling functions. This collected experimental information should be compiled in a publicly accessible, readily searchable online database. Making these data available in machine-readable formats (e.g. tab delimited text) with appropriate metadata (date and location of collection, assay conditions,etc.) is important for the training and validation of computational algorithms. Existing pipelines - like the Enzyme Function Initiative (EFI, (61,62,63) - provide an exemplar for extending and refining annotation efforts.

Broader Impacts and Educational Outcomes. Many of these pipeline tasks are amenable to science crowd-sourcing and undergraduate laboratory curricula. We envision the possibility of nation-wide laboratory courses that assign small groups of students to gene products or protein domains of interest. These proteins can then be characterized using a small set of pre-defined and relatively straightforward assays, including growth rate complementation studies, microscopy-based characterization of localization, assays of cellular phenotypes for knockouts or knockdowns, as well as expression and purification trials. Importantly, these experimental efforts can be combined with informatics and comparative genomics analyses, providing a rich opportunity for interdisciplinary cross-training.

Summary and Potential Impacts
Increasing the throughput and information content of protein annotations is essential to enabling synthetic biology and metabolic engineering, discovering new drug targets, and most fundamentally, understanding the molecular basis of phenotype. Recent advances in high throughput phenotyping, imaging, and biochemistry indicate the potential for high-dimensional annotations of function. By going beyond simple classifications or categories, we might provide

rich annotations of protein localization, physiological roles in varied conditions, and quantitative biochemical parameters. Collecting these data into a single, well-maintained and well-organized database would richly enable experimental biology and computational efforts to predict and design protein function.

Bibliography

1. Osterman A, Overbeek R. Missing genes in metabolic pathways: a comparative genomics approach. Curr Opin Chem Biol. 2003 Apr 1;7(2):238–51.
2. Earnshaw WC. Deducing Protein Function by Forensic Integrative Cell Biology. PLOS Biol. 2013 Dec 17;11(12):e1001742.
3. Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci. 2012 Dec 1;37(12):509–16.
4. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles. Proc Natl Acad Sci. 1999 Apr 13;96(8):4285–8.
5. Conserved Units of Co-Expression in Bacterial Genomes: An Evolutionary Insight into Transcriptional Regulation [Internet]. [cited 2019 Dec 21]. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155740
6. Rivoire O. Elements of Coevolution in Biological Sequences. Phys Rev Lett. 2013 Apr 23;110(17):178102.
7. Kim P-J, Price ND. Genetic Co-Occurrence Network across Sequenced Microbes. PLOS Comput Biol. 2011 Dec 29;7(12):e1002340.
8. Engelhardt BE, Jordan MI, Srouji JR, Brenner SE. Genome-scale phylogenetic function annotation of large and diverse protein families. Genome Res. 2011 Nov 1;21(11):1969–80.
9. Janga SC, Collado-Vides J, Moreno-Hagelsieb G. Nebulon: a system for the inference of functional relationships of gene products from the rearrangement of predicted operons. Nucleic Acids Res. 2005 Apr 1;33(8):2521–30.
10. Huynen M, Snel B, Lathe W, Bork P. Predicting Protein Function by Genomic Context: Quantitative Evaluation and Qualitative Inferences. Genome Res. 2000 Aug 1;10(8):1204–10.
11. Snel B, Bork P, Huynen MA. The identification of functional modules from the genomic association of genes. Proc Natl Acad Sci. 2002 Apr 30;99(9):5890–5.
12. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan 8;47(D1):D607–13.
13. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018 Jul 20;430(15):2237–43.
14. Bileschi ML, Belanger D, Bryant D, Sanderson T, Carter B, Sculley D, et al. Using Deep Learning to Annotate the Protein Universe. bioRxiv. 2019 Jul 15;626507.
15. Schober AF, Mathis AD, Ingle C, Park JO, Chen L, Rabinowitz JD, et al. A Two-Enzyme Adaptive Unit within Bacterial Folate Metabolism. Cell Rep. 2019 Jun 11;27(11):3359-3370.e7.

16. Zhou N, Jiang Y, Bergquist TR, Lee AJ, Kacsoh BZ, Crocker AW, et al. The CAFA challenge reports improved protein function prediction and new functional annotations for hundreds of genes through experimental screens. Genome Biol. 2019 19;20(1):244.
17. Deutschbauer A, Price MN, Wetmore KM, Tarjan DR, Xu Z, Shao W, et al. Towards an informative mutant phenotype for every bacterial gene. J Bacteriol. 2014 Oct;196(20):3643–55.
18. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science. 2008 Apr 18;320(5874):362–5.
19. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. Nature. 2018;557(7706):503–9.
20. Jaffe M, Dziulko A, Smith JD, St.Onge RP, Levy SF, Sherlock G. Improved discovery of genetic interactions using CRISPRiSeq across multiple environments. Genome Res. 2019 Apr 1;29(4):668–81.
21. Wong BG, Mancuso CP, Kiriakov S, Bashor CJ, Khalil AS. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat Biotechnol. 2018 Aug;36(7):614–23.
22. Toprak E, Veres A, Yildiz S, Pedraza JM, Chait R, Paulsson J, et al. Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. Nat Protoc. 2013 Mar;8(3):555–67.
23. Thiaville JJ, Flood J, Yurgel S, Prunetti L, Elbadawi-Sidhu M, Hutinet G, et al. Members of a Novel Kinase Family (DUF1537) Can Recycle Toxic Intermediates into an Essential Metabolite. ACS Chem Biol. 2016 Aug 19;11(8):2304–11.
24. Kim J, Kershner JP, Novikov Y, Shoemaker RK, Copley SD. Three serendipitous pathways in E. coli can bypass a block in pyridoxal-5′-phosphate synthesis. Mol Syst Biol. 2010 Jan 1;6(1):436.
25. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014 Nov 28;346(6213):1258096.
26. Bastard K, Smith AAT, Vergne-Vaxelaire C, Perret A, Zaparucha A, Melo-Minardi RD, et al. Revealing the hidden functional diversity of an enzyme family. Nat Chem Biol. 2014 Jan;10(1):42–9.
27. Zhao S, Kumar R, Sakai A, Vetting MW, Wood BM, Brown S, et al. Discovery of new enzymes and metabolic pathways by using structure and genome context. Nature. 2013 Oct;502(7473):698–702.
28. Mazutis L, Gilbert J, Ung WL, Weitz DA, Griffiths AD, Heyman JA. Single-cell analysis and sorting using droplet-based microfluidics. Nat Protoc. 2013 May;8(5):870–91.
29. Fordyce PM, Gerber D, Tran D, Zheng J, Li H, DeRisi JL, et al. De Novo Identification and Biophysical Characterization of Transcription Factor Binding Sites with Microfluidic Affinity Analysis. Nat Biotechnol. 2010 Sep;28(9):970–5.
30. Perdigão, N., Heinrich, J., Stolte, C., Sabir, K.S., Buckley, M.J., Tabor, B., Signal, B., Gloss, B.S., Hammang, C.J., Rost, B., Schafferhans, A., O'Donoghue, S.I. (2015). Unexpected features of the dark proteome. Proc. Natl. Acad. Sci. USA 2015, 112, 15898–15903.
31. Lieutaud, P.; Ferron, F.; Uversky, A.V.; Kurgan, L.; Uversky, V.N.; Longhi, S. (2016). How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe. Intrinsically Disord. Proteins 2016, 4, e1259708.
32. Perdigão, N. and Rosa, A. (2019). Dark Proteome Database: Studies on Dark Proteins. High-Throughput 8, 8; doi:10.3390/ht8020008.
33. Daniels, G.J., Jenkins, R.D., Bradshaw, D.S., and Andrews, D.L. (2003). Resonance energy transfer: The unified theory revisited. J Chem Phys 119, 2264–2274.

34. Costes, S.V., Daelemans, D., Cho, E.H., Dobbin, Z., Pavlakis, G., and Lockett, S. (2004). Automatic and quantitative measurement of protein-protein co-localization in live cells. Biophys J 86, 3993–4003.
35. Bolte, S., and Cordelières, F.P. (2006). A guided tour into subcellular co-localization analysis in light microscopy. J Microsc 224, 213–232.
36. Comeau, J.W.D., Kolin, D.L., and Wiseman, P.W. (2008). Accurate measurements of protein interactions in cells via improved spatial image cross-correlation spectroscopy. Mol Biosyst 4, 672–685.
37. Snapp, E.L. and Hegde, R.S. (2006). Rational design and evaluation of FRET experiments to measure protein proximities in cells. Curr Protoc Cell Biol. 17:Unit 17.9. doi: 10.1002/0471143030.cb1709s32.
38. Uhlén, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Björling, L., and Ponten, F. (2010). Towards a knowledge-based human protein atlas. Nat Biotechnol. 28: 1248–1250.
39. Burry, R.W. (2011). Controls for immunocytochemistry: an update. J Histochem Cytochem. 1:6-12.
40. Hewitt, S.M., Baskin, D.G., Frevert, C.W., Stahl, W.L., and Rosa-Molinar, E. (2014). Controls for Immunohistochemistry: The Histochemical Society's standards of practice for validation of immunohistochemical assays. J. Histochem Cytochem. 62: 693–697.
41. Uhlén, M. and Ponten, F. (2015). Antibody-based proteomics for human tissue profiling. Mol Cell Proteomics. 4:384-93.
42. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., and Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. Science. 347:1260419.
43. Gautron, L. (2019). On the necessity of validating antibodies in the immunohistochemistry literature. Front Neuroanat. 2019; 13: 46.
44. Thöne, F.M.B., Kurrle, N.S., von Melchner, H., and Schnütgen, F. (2019). CRISPR/Cas9-mediated generic protein tagging in mammalian cells. Methods. 2019 Jul 15;164-165:59-66.
45. Colón-Ramos, D.A., La Riviere, P., Shroff, H., and Oldenbourg, R. (2019). Transforming the development and dissemination of cutting-edge microscopy and computation. Nat Methods. Nat Methods 16, 667–669.
46. Power, R.M. and Huisken (2019). Putting advanced microscopy in the hands of biologists. Nat Methods. 16:1069-1073.
47. Periasamy, A., and Day, R.N. (1999). Visualizing protein interactions in living cells using digitized GFP imaging and FRET microscopy. Method Cell Biol 58, 293–314.
48. Zaccolo, M. (2004). Use of chimeric fluorescent proteins and fluorescence resonance energy transfer to monitor cellular responses. Circ Res. 2004;94:866-873.
49. Ciruela, F. (2008). Fluorescence-based methods in the study of protein-protein interactions in living cells. Curr Opin Biotech 19, 338–343.

50. Bucherl, C., Aker, J., de Vries, S., and Borst, J.W. (2010). Probing protein-protein interactions with FRET-FLIM. Methods Mol Biol 655, 389–399.
51. De Los Santos, C., Chang, C.W., Mycek, M.A., and Cardullo, R.A. (2015). FRAP, FLIM, and FRET: Detection and analysis of cellular dynamics on a molecular scale using fluorescence microscopy. Mol Reprod Dev 82, 587–604.
52. Shrestha, D., Jenei, A., Nagy, P., Vereb, G., and Szöllosi, J. (2015). Understanding FRET as a research tool for cellular studies. IJMS 16, 6718–6756.
53. Shinogle-Decker, H., Martinez-Rivera, N., O'Brien, J., Powell, R. D., Joshi, V. N., Connell, S., and Rosa-Molinar, E. (2018). Correlative FRET: new method improves rigor and reproducibility in determining distances within synaptic nanoscale architecture. Proc. SPIE 10499, Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXV, 1049920; 23 February 2018.
54. Bunt, G. and Wouters F.S. (2017). FRET from single to multiplexed signaling events. Biophys Rev. 9(2):119-129.
55. Geißler, D. and Hildebrandt, N. (2016). Recent developments in Förster resonance energy transfer (FRET) diagnostics using quantum dots. Anal Bioanal Chem. 408(17):4475-83.
56. Chenab, K.K., Eivazzadeh-Keihan R., Maleki, A., Pashazadeh-Panahi, P., Hamblin, M.R., and Mokhtarzadeh, A. (2019). Biomedical applications of nanoflares: Targeted intracellular fluorescence probes. Nanomedicine. 17:342-358.
57. Tsai, H.Y., Kim, H., Massey, M., Krause, K.D., and Algar, W.R. (2019). Concentric FRET: a review of the emerging concept, theory, and applications. Methods Appl Fluoresc. 7(4):042001.
58. Ellen's KW, Christian N, Singh C, Satagopam VP, May P, Linster CL. Confronting the catalytic dark matter encoded by sequenced genomes. Nucleic Acids Res. 2017 Nov 16;45(20):11495–514.
59. Roberts RJ. Identifying Protein Function—A Call for Community Action. PLOS Biol. 2004 Mar 16;2(3):e42.
60. Gerlt JA, Allen KN, Almo SC, Armstrong RN, Babbitt PC, Cronan JE, et al. The Enzyme Function Initiative. Biochemistry. 2011 Nov 22;50(46):9950–62.
61. Zallot R, Oberg N, Gerlt JA. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. Biochemistry. 2019 Oct 15;58(41):4169–82.
62. Zallot R, Oberg NO, Gerlt JA. 'Democratized' genomic enzymology web tools for functional assignment. Curr Opin Chem Biol. 2018 Dec 1;47:77–85.