ChIP-Seq is the mainstream experimental method for genome-wide identification of transcription factor (TF) binding and chromatin modification sites. The Gerstein lab developed two ChIP-Seq peak calling tools: PeakSeq (84) and MUSIC (85). PeakSeq is a versatile tool for identification of TF binding sites and is one of the standard peak calling programs used by the ENCODE and modENCODE consortia for numerous ChIP-Seq datasets (86). PeakSeq constructs local thresholds using input signals from input genomic DNA to simulate the null process for background. PeakSeq then identifies significantly enriched peaks relative to this background. MUSIC is a recently introduced signal processing approach for identification of enriched regions in ChIP-Seq when strict peaks are not observed. Utilizing multiscale decomposition of ChIP-Seq signal profiles in conjunction with mappability correction, MUSIC allows identification of broad enrichment domains (85). MUSIC performs well in both narrow and broad peak calling. The Gerstein lab has also developed tools that use ChIP-Seq signals to identify regulatory regions and predict of gene expression (37, 87-90). For example, we integrated >100 ChIP-Seq signals from the ENCODE project to predict enhancers for distal gene regulatory modules (90).

For RNA-Seq analysis, we have developed RSEQtools, a computational package that enables expression quantification of annotated RNAs and identification of splice sites and gene models (91). In addition, we have developed IQseq, a computationally efficient method to quantify isoforms for alternatively spliced transcripts (92). Comparisons between RNA-Seq samples and other genome-wide data can be facilitated in part by our
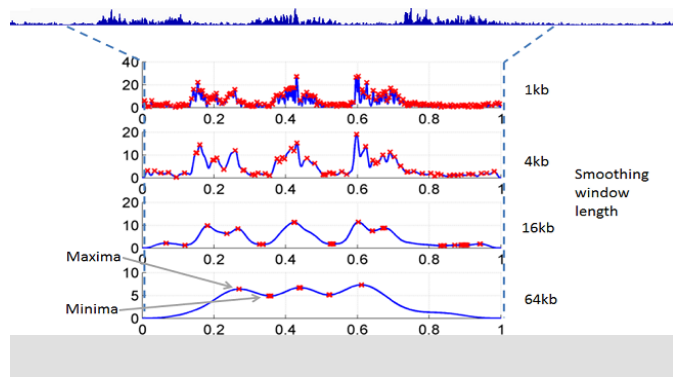

Fig 7 ChIP-Seq analysis through MUSIC

Aggregation and Correlation Toolbox (ACT), a general purpose tool for comparing genome signal tracks (93). An important challenge in RNA-Seq analysis is detecting unannotated transcription hard to distinguish from noise. This topic has been central to many of our expression analysis tools (35, 94-98). Our Database of Annotated Regions with Tools (DART) package contains tools for identifying unannotated genomic regions enriched for transcription, as well as a framework for storing and querying this information (99). We developed incRNA, a method that uses known ncRNAs of various classes as a gold standard training set to predict and analyze novel ncRNAs (100). We have also developed specific tools to identify types of transcripts that are difficult to detect using standard analysis pipelines, including fusion transcripts and pseudogenes (101, 102)(103).

Network representations can be applied consistently to many different types of biological data and therefore can be used as a powerful framework for integration of heterogeneous datasets (109-111). We have developed novel approaches for studying

biological networks. We developed methods to construct and analyze the regulatory networks of human and model organisms (36, 88, 112, 113) based on ENCODE and modENCODE datasets (34). Regulatory networks were integrated with gene expression to uncover different types of functional modules (114-117). We constructed and analyzed hierarchical regulatory networks in model organisms and humans (88, 118-121), discovering that the hierarchy rather than centrality ("hubiness") better reflects the importance of regulators. We built a multi-layered network that incorporated information from heterogeneous data sources such as protein-protein interactions and metabolic, phosphorylation, signaling, genetic, and regulatory networks (122) and used networks to improve understanding of genomic variants (123) We also introduced several software tools for network analysis, including Topnet, (124) tYNA (125) and PubNet (126).

References

1. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature **489:**57-74. **Agarwal, A., D. Koppstein, J. Rozowsky, A. Sboner, L. Habegger, L. W. Hillier, R. Sasidharan, V. Reinke, R. H. Waterston, and M. Gerstein.** 2010. Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. BMC Genomics **11:**383.

2. **Anko, M. L., L. Morales, I. Henry, A. Beyer, and K. M. Neugebauer.** 2010. Global analysis reveals SRp20- and SRp75-specific mRNPs in cycling and neural cells. Nature structural & molecular biology **17:**962-970.

3. **Anko, M. L., M. Muller-McNicoll, H. Brandl, T. Curk, C. Gorup, I. Henry, J. Ule, and K. M. Neugebauer.** 2012. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. Genome Biol **13:**R17.

4. **Anko, M. L., and K. M. Neugebauer.** 2012. RNA-protein interactions in vivo: global gets specific. Trends Biochem Sci **37:**255-262.

5. **Baltz, A. G., M. Munschauer, B. Schwanhausser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, and M. Landthaler.** 2012. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Molecular cell **46:**674-690.

6. **Barcaroli, D., L. Bongiorno-Borbone, A. Terrinoni, T. G. Hofmann, M. Rossi, R. A. Knight, A. G. Matera, G. Melino, and V. De Laurenzi.** 2006. FLASH is required for histone transcription and S-phase progression. Proceedings of the National Academy of Sciences of the United States of America **103:**14808-14812.

7. **Barcaroli, D., D. Dinsdale, M. H. Neale, L. Bongiorno-Borbone, M. Ranalli, E. Munarriz, A. E. Sayan, J. M. McWilliam, T. M. Smith, E. Fava, R. A. Knight, G. Melino, and V. De Laurenzi.** 2006. FLASH is an essential component of Cajal bodies. Proceedings of the National Academy of Sciences of the United States of America **103:**14802-14807.

8. **Battle, D. J., C. K. Lau, L. Wan, H. Deng, F. Lotti, and G. Dreyfuss.** 2006. The Gemin5 protein of the SMN complex identifies snRNAs. Molecular cell **23:**273-279.

9. **Bernstein, B. E., J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham, M. Hirst, E. S. Lander, T. S. Mikkelsen, and J. A. Thomson.** 2010. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol **28:**1045-1048.

10.   **Bertone, P., V. Stolc, T. E. Royce, J. S. Rozowsky, A. E. Urban, X. Zhu, J. L. Rinn, W. Tongprasit, M. Samanta, S. Weissman, M. Gerstein, and M. Snyder.** 2004. Global identification of human transcribed sequences with genome tiling arrays. Science **306:**2242-2246.

11.   **Bhardwaj, N., M. B. Carson, A. Abyzov, K. K. Yan, H. Lu, and M. B. Gerstein.** 2010. Analysis of combinatorial regulation: scaling of partnerships between regulators with the number of governed targets. PLoS Comput Biol **6:**e1000755.

12.   **Bhardwaj, N., P. M. Kim, and M. B. Gerstein.** 2010. Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators. Sci Signal **3:**ra79.

13.   **Bhardwaj, N., K. K. Yan, and M. B. Gerstein.** 2010. Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. Proceedings of the National Academy of Sciences of the United States of America **107:**6841-6846.

14.   **Bieberstein, N. I., K. Straube, and K. M. Neugebauer.** 2014. Chromatin immunoprecipitation approaches to determine co-transcriptional nature of splicing. Methods Mol Biol **1126:**315-323.

15.   **Boisvert, F. M., J. Cote, M. C. Boulanger, P. Cleroux, F. Bachand, C. Autexier, and S. Richard.** 2002. Symmetrical dimethylarginine methylation is required for the localization of SMN in Cajal bodies and pre-mRNA splicing. The Journal of cell biology **159:**957-969.

16.   **Bond, C. S., and A. H. Fox.** 2009. Paraspeckles: nuclear bodies built on long noncoding RNA. J Cell Biol **186:**637-644.

17.   **Bongiorno-Borbone, L., A. De Cola, P. Vernole, L. Finos, D. Barcaroli, R. A. Knight, G. Melino, and V. De Laurenzi.** 2008. FLASH and NPAT positive but not Coilin positive Cajal Bodies correlate with cell ploidy. Cell Cycle **7:**2357-2367.

18.   **Boulon, S., C. Verheggen, B. E. Jady, C. Girard, C. Pescia, C. Paul, J. K. Ospina, T. Kiss, A. G. Matera, R. Bordonne, and E. Bertrand.** 2004. PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. Molecular cell **16:**777-787.

20.   **Boulon, S., B. J. Westman, S. Hutten, F. M. Boisvert, and A. I. Lamond.** 2010. The nucleolus under stress. Molecular cell **40:**216-227.

21.   **Brannan, K., H. Kim, B. Erickson, K. Glover-Cutter, S. Kim, N. Fong, L. Kiemele, K. Hansen, R. Davis, J. Lykke-Andersen, and D. L. Bentley.** 2012. mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription. Molecular cell **46:**311-324.

22.   **Brouwer, A. K., J. Schimmel, J. C. Wiegant, A. C. Vertegaal, H. J. Tanke, and R. W. Dirks.** 2009. Telomeric DNA mediates de novo PML body formation. Molecular biology of the cell **20:**4804-4815.

23.   **Carmo-Fonseca, M., R. Pepperkok, M. T. Carvalho, and A. I. Lamond.** 1992. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. J Cell Biol **117:**1-14.

24.   **Carvalho, T., F. Almeida, A. Calapez, M. Lafarga, M. T. Berciano, and M. Carmo-Fonseca.** 1999. The spinal muscular atrophy disease gene product, SMN: A link between snRNP biogenesis and the Cajal (coiled) body. J Cell Biol **147:**715-728.

25.   **Castello, A., B. Fischer, K. Eichelbaum, R. Horos, B. M. Beckmann, C. Strein, N. E. Davey, D.**

T. Humphreys, T. Preiss, L. M. Steinmetz, J. Krijgsveld, and M. W. Hentze. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell **149:**1393-1406.

26. **Cheng, C., R. Alexander, R. Min, J. Leng, K. Y. Yip, J. Rozowsky, K. K. Yan, X. Dong, S. Djebali, Y. Ruan, C. A. Davis, P. Carninci, T. Lassman, T. R. Gingeras, R. Guigo, E. Birney, Z. Weng, M. Snyder, and M. Gerstein.** 2012. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome research **22:**1658-1667.

27. **Cheng, C., X. Fu, P. Alves, and M. Gerstein.** 2009. mRNA expression profiles show differential regulatory effects of microRNAs between estrogen receptor-positive and estrogen receptor-negative breast cancer. Genome Biol **10:**R90.

28. **Cheng, C., R. Min, and M. Gerstein.** 2011. TIP: a probabilistic method for identifying transcription factor target genes from ChIP-seq binding profiles. Bioinformatics **27:**3221-3227.

29. **Cheng, C., K. K. Yan, W. Hwang, J. Qian, N. Bhardwaj, J. Rozowsky, Z. J. Lu, W. Niu, P. Alves, M. Kato, M. Snyder, and M. Gerstein.** 2011. Construction and analysis of an integrated regulatory network derived from high-throughput sequencing data. PLoS Comput Biol **7:**e1002190.

30. **Clark, M. B., P. P. Amaral, F. J. Schlesinger, M. E. Dinger, R. J. Taft, J. L. Rinn, C. P. Ponting, P. F. Stadler, K. V. Morris, A. Morillon, J. S. Rozowsky, M. B. Gerstein, C. Wahlestedt, Y. Hayashizaki, P. Carninci, T. R. Gingeras, and J. S. Mattick.** 2011. The reality of pervasive transcription. PLoS Biol **9:**e1000625; discussion e1001102.

31. **Clemson, C. M., J. N. Hutchinson, S. A. Sara, A. W. Ensminger, A. H. Fox, A. Chess, and J. B. Lawrence.** 2009. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. Molecular cell **33:**717-726.

32. **Cong, R., S. Das, I. Ugrinova, S. Kumar, F. Mongelard, J. Wong, and P. Bouvet.** 2012. Interaction of nucleolin with ribosomal RNA genes and its role in RNA polymerase I transcription. Nucleic acids research **40:**9441-9454.

33. **Cristofari, G., E. Adolf, P. Reichenbach, K. Sikora, R. M. Terns, M. P. Terns, and J. Lingner.** 2007. Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation. Molecular cell **27:**882-889.

34. **Dekker, J., M. A. Marti-Renom, and L. A. Mirny.** 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet **14:**390-403.

35. **Deryusheva, S., and J. G. Gall.** 2013. Novel small Cajal-body-specific RNAs identified in Drosophila: probing guide RNA function. RNA **19:**1802-1814.

36. **Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Roder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B.**

Wold, P. Carninci, R. Guigo, and T. R. Gingeras. 2012. Landscape of transcription in human cells.

Nature **489:**101-108.

37.   **Douglas, S. M., G. T. Montelione, and M. Gerstein.** 2005. PubNet: a flexible system for visualizing literature derived networks. Genome Biol **6:**R80.

38.   **Du, J., J. Leng, L. Habegger, A. Sboner, D. McDermott, and M. Gerstein.** 2012. IQSeq: integrated isoform quantification analysis based on next-generation sequencing. PLoS One **7:**e29175.

39.   **Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau, and W. S. Noble.** 2010. A three-dimensional model of the yeast genome. Nature **465:**363-367.

40.   **Dundr, M., M. D. Hebert, T. S. Karpova, D. Stanek, H. Xu, K. B. Shpargel, U. T. Meier, K. M. Neugebauer, A. G. Matera, and T. Misteli.** 2004. In vivo kinetics of Cajal body components. J Cell Biol **164:**831-842.

41.   **Edwards, A. M., B. Kus, R. Jansen, D. Greenbaum, J. Greenblatt, and M. Gerstein.** 2002. Bridging structural biology and genomics: assessing protein interaction data with known complexes. Trends Genet **18:**529-536.

42.   **Engreitz, J. M., K. Sirokman, P. McDonel, A. A. Shishkin, C. Surka, P. Russell, S. R. Grossman, A. Y. Chow, M. Guttman, and E. S. Lander.** 2014. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell **159:**188-199.

43.   **Frey, M. R., A. D. Bailey, A. M. Weiner, and A. G. Matera.** 1999. Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. Curr Biol **9:**126-135.

44.   **Frey, M. R., and A. G. Matera.** 1995. Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells. Proc Natl Acad Sci U S A **92:**5915-5919.

45.   **Frey, M. R., and A. G. Matera.** 2001. RNA-mediated interaction of Cajal bodies and U2 snRNA genes. J Cell Biol **154:**499-509.

46.   **Gall, J. G.** 2000. Cajal bodies: the first 100 years. Annu Rev Cell Dev Biol **16:**273-300.

47.   **Gerstein, M. B., A. Kundaje, M. Hariharan, S. G. Landt, K. K. Yan, C. Cheng, X. J. Mu, E. Khurana, J. Rozowsky, R. Alexander, R. Min, P. Alves, A. Abyzov, N. Addleman, N. Bhardwaj, A. P. Boyle, P. Cayting, A. Charos, D. Z. Chen, Y. Cheng, D. Clarke, C. Eastman, G. Euskirchen, S. Frietze, Y. Fu, J. Gertz, F. Grubert, A. Harmanci, P. Jain, M. Kasowski, P. Lacroute, J. Leng, J. Lian, H. Monahan, H. O'Geen, Z. Ouyang, E. C. Partridge, D. Patacsil, F. Pauli, D. Raha, L. Ramirez, T. E. Reddy, B. Reed, M. Shi, T. Slifer, J. Wang, L. Wu, X. Yang, K. Y. Yip, G. Zilberman-Schapira, S. Batzoglou, A. Sidow, P. J. Farnham, R. M. Myers, S. M. Weissman, and M. Snyder.** 2012. Architecture of the human regulatory network derived from ENCODE data. Nature **489:**91-100.

48.   **Gerstein, M. B., Z. J. Lu, E. L. Van Nostrand, C. Cheng, B. I. Arshinoff, T. Liu, K. Y. Yip, R. Robilotto, A. Rechtsteiner, K. Ikegami, P. Alves, A. Chateigner, M. Perry, M. Morris, R. K. Auerbach, X. Feng, J. Leng, A. Vielle, W. Niu, K. Rhrissorrakrai, A. Agarwal, R. P. Alexander, G. Barber, C. M. Brdlik, J. Brennan, J. J. Brouillet, A. Carr, M. S. Cheung, H. Clawson, S. Contrino, L. O. Dannenberg, A. F. Dernburg, A. Desai, L. Dick, A. C. Dose, J. Du, T. Egelhofer, S. Ercan, G. Euskirchen, B. Ewing, E. A. Feingold, R. Gassmann, P. J. Good, P. Green, F. Gullier, M. Gutwein, M. S. Guyer, L. Habegger, T. Han, J. G. Henikoff, S. R. Henz, A. Hinrichs, H. Holster, T. Hyman, A. L. Iniguez, J. Janette, M. Jensen, M. Kato, W. J. Kent, E. Kephart, V. Khivansara, E. Khurana, J. K. Kim, P. Kolasinska-Zwierz, E. C. Lai, I. Latorre, A. Leahey, S. Lewis, P. Lloyd, L. Lochovsky, R. F. Lowdon, Y. Lubling, R. Lyne, M. MacCoss, S.**

D. Mackowiak, M. Mangone, S. McKay, D. Mecenas, G. Merrihew, D. M. Miller, 3rd, A. Muroyama, J. I. Murray, S. L. Ooi, H. Pham, T. Phippen, E. A. Preston, N. Rajewsky, G. Ratsch, H. Rosenbaum, J. Rozowsky, K. Rutherford, P. Ruzanov, M. Sarov, R. Sasidharan, A. Sboner, P. Scheid, E. Segal, H. Shin, C. Shou, F. J. Slack, et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science **330:**1775-1787.

49.   Gerstein, M. B., J. Rozowsky, K. K. Yan, D. Wang, C. Cheng, J. B. Brown, C. A. Davis, L. Hillier, C. Sisu, J. J. Li, B. Pei, A. O. Harmanci, M. O. Duff, S. Djebali, R. P. Alexander, B. H. Alver, R. Auerbach, K. Bell, P. J. Bickel, M. E. Boeck, N. P. Boley, B. W. Booth, L. Cherbas, P. Cherbas, C. Di, A. Dobin, J. Drenkow, B. Ewing, G. Fang, M. Fastuca, E. A. Feingold, A. Frankish, G. Gao, P. J. Good, R. Guigo, A. Hammonds, J. Harrow, R. A. Hoskins, C. Howald, L. Hu, H. Huang, T. J. Hubbard, C. Huynh, S. Jha, D. Kasper, M. Kato, T. C. Kaufman, R. R. Kitchen, E. Ladewig, J. Lagarde, E. Lai, J. Leng, Z. Lu, M. MacCoss, G. May, R. McWhirter, G. Merrihew, D. M. Miller, A. Mortazavi, R. Murad, B. Oliver, S. Olson, P. J. Park, M. J. Pazin, N. Perrimon, D. Pervouchine, V. Reinke, A. Reymond, G. Robinson, A. Samsonova, G. I. Saunders, F. Schlesinger, A. Sethi, F. J.

Slack, W. C. Spencer, M. H. Stoiber, P. Strasbourger, A. Tanzer, O. A. Thompson, K. H. Wan, G. Wang, H. Wang, K. L. Watkins, J. Wen, K. Wen, C. Xue, L. Yang, K. Yip, C. Zaleski, Y. Zhang, H. Zheng, S. E. Brenner, B. R. Graveley, S. E. Celniker, T. R. Gingeras, and R. Waterston. 2014. Comparative analysis of the transcriptome across distant species. Nature **512:**445-448.

50.   Ghule, P. N., Z. Dominski, J. B. Lian, J. L. Stein, A. J. van Wijnen, and G. S. Stein. 2009. The subnuclear organization of histone gene regulatory proteins and 3' end processing factors of normal somatic and embryonic stem cells is compromised in selected human cancer cell types. J Cell Physiol **220:**129-135.

51.   Gornemann, J., K. M. Kotovic, K. Hujer, and K. M. Neugebauer. 2005. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell **19:**53-63.

52.   Granneman, S., G. Kudla, E. Petfalski, and D. Tollervey. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proceedings of the National Academy of Sciences of the United States of America **106:**9613-9618.

53.   Grob, A., C. Colleran, and B. McStay. 2014. Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. Genes & development **28:**220-230.

54.   Habegger, L., A. Sboner, T. A. Gianoulis, J. Rozowsky, A. Agarwal, M. Snyder, and M. Gerstein. 2011. RSEQtools: a modular framework to analyze RNA-Seq data using compact, anonymized data summaries. Bioinformatics **27:**281-283.

55.   Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, Jr., A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell **141:**129-141.

56.   Han, T. W., M. Kato, S. Xie, L. C. Wu, H. Mirzaei, J. Pei, M. Chen, Y. Xie, J. Allen, G. Xiao, and S. L. McKnight. 2012. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell **149:**768-779.

57.   Harmanci, A., J. Rozowsky, and M. Gerstein. 2014. MUSIC: Identification of Enriched Regions in ChIP-Seq Experiments using a Mappability-Corrected Multiscale Signal Processing Framework. Genome Biol **15:**474.

58.   Hebert, M. D., K. B. Shpargel, J. K. Ospina, K. E. Tucker, and A. G. Matera. 2002. Coilin

methylation regulates nuclear body formation. Dev Cell **3:**329-337.

59. **Heyn, P., M. Kircher, A. Dahl, J. Kelso, P. Tomancak, A. T. Kalinka, and K. M. Neugebauer.** 2014. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep **6:**285-292.

60. **Holme, P., and J. Saramaki.** 2012. Temporal Networks. Physics Reports **519:**97-125.

61. **Huppertz, I., J. Attig, A. D'Ambrogio, L. E. Easton, C. R. Sibley, Y. Sugimoto, M. Tajnik, J. Konig, and J. Ule.** 2014. iCLIP: protein-RNA interactions at nucleotide resolution. Methods **65:**274-287.

62. **Hutchinson, J. N., A. W. Ensminger, C. M. Clemson, C. R. Lynch, J. B. Lawrence, and A. Chess.** 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics **8:**39.

63. **Jacobs, E. Y., M. R. Frey, W. Wu, T. C. Ingledue, T. C. Gebuhr, L. Gao, W. F. Marzluff, and A. G. Matera.** 1999. Coiled bodies preferentially associate with U4, U11, and U12 small nuclear RNA genes in interphase HeLa cells but not with U6 and U7 genes. Mol Biol Cell **10:**1653-1663.

64. **Jansen, R., N. Lan, J. Qian, and M. Gerstein.** 2002. Integration of genomic datasets to predict protein complexes in yeast. J Struct Funct Genomics **2:**71-81.

65. **Jee, J., J. Rozowsky, K. Y. Yip, L. Lochovsky, R. Bjornson, G. Zhong, Z. Zhang, Y. Fu, J. Wang, Z. Weng, and M. Gerstein.** 2011. ACT: aggregation and correlation toolbox for analyses of genome tracks. Bioinformatics **27:**1152-1154.

66. **Ji, X., Y. Zhou, S. Pandit, J. Huang, H. Li, C. Y. Lin, R. Xiao, C. B. Burge, and X. D. Fu.** 2013. SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell **153:**855-868.

67. **Kato, M., T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko, and S. L. McKnight.** 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell **149:**753-767.

68. **Khurana, E., Y. Fu, J. Chen, and M. Gerstein.** 2013. Interpretation of genomic variants using a unified biological network approach. PLoS Comput Biol **9:**e1002886.

69. **Khurana, E., Y. Fu, V. Colonna, X. J. Mu, H. M. Kang, T. Lappalainen, A. Sboner, L. Lochovsky, J. Chen, A. Harmanci, J. Das, A. Abyzov, S. Balasubramanian, K. Beal, D. Chakravarty, D. Challis, Y. Chen, D. Clarke, L. Clarke, F. Cunningham, U. S. Evani, P. Flicek, R. Fragoza, E. Garrison, R. Gibbs, Z. H. Gumus, J. Herrero, N. Kitabayashi, Y. Kong, K. Lage, V. Liluashvili, S. M. Lipkin, D. G. MacArthur, G. Marth, D. Muzny, T. H. Pers, G. R. Ritchie, J. A. Rosenfeld, C. Sisu, X. Wei, M. Wilson, Y. Xue, F. Yu, E. T. Dermitzakis, H. Yu, M. A. Rubin, C. Tyler-Smith, and M. Gerstein.** 2013. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science **342:**1235587.

70. **Klingauf, M., D. Stanek, and K. M. Neugebauer.** 2006. Enhancement of U4/U6 small nuclear ribonucleoprotein particle association in Cajal bodies predicted by mathematical modeling. Mol Biol Cell **17:**4972-4981.

71. **Konig, J., K. Zarnack, N. M. Luscombe, and J. Ule.** 2011. Protein-RNA interactions: new genomic technologies and perspectives. Nat Rev Genet **13:**77-83.

72. **Konig, J., K. Zarnack, N. M. Luscombe, and J. Ule.** 2012. Protein-RNA interactions: new

genomic technologies and perspectives. Nat Rev Genet **13:**77-83.

73. **Konig, J., K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe, and J. Ule.** 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nature structural & molecular biology **17:**909-915.

74. **Kotovic, K. M., D. Lockshon, L. Boric, and K. M. Neugebauer.** 2003. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol Cell Biol **23:**5768-5779.

75. **Kwon, I., S. Xiang, M. Kato, L. Wu, P. Theodoropoulos, T. Wang, J. Kim, J. Yun, Y. Xie, and S. L. McKnight.** 2014. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science **345:**1139-1145.

76. **Lallemand-Breitenbach, V., and H. de The.** 2010. PML nuclear bodies. Cold Spring Harb Perspect Biol **2:**a000661.

77. **Landt, S. G., G. K. Marinov, A. Kundaje, P. Kheradpour, F. Pauli, S. Batzoglou, B. E. Bernstein, P. Bickel, J. B. Brown, P. Cayting, Y. Chen, G. DeSalvo, C. Epstein, K. I. Fisher-Aylor, G. Euskirchen, M. Gerstein, J. Gertz, A. J. Hartemink, M. M. Hoffman, V. R. Iyer, Y. L. Jung, S. Karmakar, M. Kellis, P. V. Kharchenko, Q. Li, T. Liu, X. S. Liu, L. Ma, A. Milosavljevic, R. M. Myers, P. J. Park, M. J. Pazin, M. D. Perry, D. Raha, T. E. Reddy, J. Rozowsky, N. Shoresh, A. Sidow, M. Slattery, J. A. Stamatoyannopoulos, M. Y. Tolstorukov, K. P. White, S. Xi, P. J. Farnham, J. D. Lieb, B. J. Wold, and M. Snyder.** 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome research **22:**1813-1831.

78. **Lau, C. K., J. L. Bachorik, and G. Dreyfuss.** 2009. Gemin5-snRNA interaction reveals an RNA binding function for WD repeat domains. Nature structural & molecular biology **16:**486-491.

79. **Listerman, I., A. K. Sapra, and K. M. Neugebauer.** 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nature structural & molecular biology **13:**815-822.

80. **Liu, J. L., C. Murphy, M. Buszczak, S. Clatterbuck, R. Goodman, and J. G. Gall.** 2006. The Drosophila melanogaster Cajal body. J Cell Biol **172:**875-884.

81. **Liu, J. L., Z. Wu, Z. Nizami, S. Deryusheva, T. K. Rajendra, K. J. Beumer, H. Gao, A. G. Matera, D. Carroll, and J. G. Gall.** 2009. Coilin is essential for Cajal body organization in Drosophila melanogaster. Mol Biol Cell **20:**1661-1670.

82. **Lu, Z. J., K. Y. Yip, G. Wang, C. Shou, L. W. Hillier, E. Khurana, A. Agarwal, R. Auerbach, J. Rozowsky, C. Cheng, M. Kato, D. M. Miller, F. Slack, M. Snyder, R. H. Waterston, V. Reinke, and M. B. Gerstein.** 2011. Prediction and characterization of noncoding RNAs in C. elegans by integrating conservation, secondary structure, and high-throughput sequencing and array data. Genome research **21:**276-285.

83. **Luscombe, N. M., M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann, and M. Gerstein.** 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature **431:**308- 312.

84. **Ma, T., B. A. Van Tine, Y. Wei, M. D. Garrett, D. Nelson, P. D. Adams, J. Wang, J. Qin, L. T. Chow, and J. W. Harper.** 2000. Cell cycle-regulated phosphorylation of p220(NPAT) by cyclin E/Cdk2 in Cajal bodies promotes histone gene transcription. Genes & development **14:**2298-2313.

85. **Machyna, M., P. Heyn, and K. M. Neugebauer.** 2013. Cajal bodies: where form meets function. Wiley Interdiscip Rev RNA **4:**17-34.

86.   **Machyna, M., S. Kehr, K. Straube, D. Kappei, F. Buchholz, F. Butter, J. Ule, J. Hertel, P. F. Stadler, and K. M. Neugebauer.** 2014. The Coilin Interactome Identifies Hundreds of Small Noncoding RNAs that Traffic through Cajal Bodies. Molecular cell **56:**389-399.

87.   **Mahmoudi, S., S. Henriksson, I. Weibrecht, S. Smith, O. Soderberg, S. Stromblad, K. G. Wiman, and M. Farnebo.** 2010. WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies. PLoS Biol **8:**e1000521.

88.   **Mao, Y. S., H. Sunwoo, B. Zhang, and D. L. Spector.** 2011. Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. Nat Cell Biol **13:**95-101.

89.   **Mao, Y. S., B. Zhang, and D. L. Spector.** 2011. Biogenesis and function of nuclear bodies. Trends Genet **27:**295-306.

90.   **Marzluff, W. F., E. J. Wagner, and R. J. Duronio.** 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet **9:**843-854.

91.   **Matera, A. G., and Z. Wang.** 2014. A day in the life of the spliceosome. Nature reviews. Molecular cell biology **15:**108-121.

92.   **McCann, K. L., and S. J. Baserga.** 2013. Genetics. Mysterious ribosomopathies. Science **341:**849- 850.

93.   **Mercer, T. R., M. B. Clark, J. Crawford, M. E. Brunck, D. J. Gerhardt, R. J. Taft, L. K. Nielsen, M. E. Dinger, and J. S. Mattick.** 2014. Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc **9:**989-1009.

94.   **Morimoto, M., and C. F. Boerkoel.** 2013. The role of nuclear bodies in gene expression and disease. Biology (Basel) **2:**976-1033.

95.   **Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder.** 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320:**1344-1349.

96.   **Naumova, N., M. Imakaev, G. Fudenberg, Y. Zhan, B. R. Lajoie, L. A. Mirny, and J. Dekker.** 2013. Organization of the mitotic chromosome. Science **342:**948-953.

97.   **O'Keefe, R. T., A. Mayeda, C. L. Sadowski, A. R. Krainer, and D. L. Spector.** 1994. Disruption of pre-mRNA splicing in vivo results in reorganization of splicing factors. The Journal of cell biology **124:**249-260.

98.   **Pabis, M., N. Neufeld, M. C. Steiner, T. Bojic, Y. Shav-Tal, and K. M. Neugebauer.** 2013. The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. Rna **19:**1054-1063.

99.   **Passon, D. M., M. Lee, O. Rackham, W. A. Stanley, A. Sadowska, A. Filipovska, A. H. Fox, and C. S. Bond.** 2012. Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. Proceedings of the National Academy of Sciences of the United States of America **109:**4846-4850.

100.   **Pederson, T.** 2011. The nucleolus. Cold Spring Harb Perspect Biol **3**.

101.   **Pei, B., C. Sisu, A. Frankish, C. Howald, L. Habegger, X. J. Mu, R. Harte, S. Balasubramanian, A. Tanzer, M. Diekhans, A. Reymond, T. J. Hubbard, J. Harrow, and M. B. Gerstein.** 2012. The GENCODE pseudogene resource. Genome Biol **13:**R51.

102.   **Pflueger, D., S. Terry, A. Sboner, L. Habegger, R. Esgueva, P. C. Lin, M. A. Svensson, N. Kitabayashi, B. J. Moss, T. Y. MacDonald, X. Cao, T. Barrette, A. K. Tewari, M. S. Chee, A. M. Chinnaiyan, D. S. Rickman, F. Demichelis, M. B. Gerstein, and M. A. Rubin.** 2011. Discovery of non-ETS gene fusions in human prostate cancer using next-generation RNA sequencing. Genome research **21:**56-67.

103.   **Poser, I., M. Sarov, J. R. Hutchins, J. K. Heriche, Y. Toyoda, A. Pozniakovsky, D. Weigl, A. Nitzsche, B. Hegemann, A. W. Bird, L. Pelletier, R. Kittler, S. Hua, R. Naumann, M. Augsburg, M. M. Sykora, H. Hofemeister, Y. Zhang, K. Nasmyth, K. P. White, S. Dietzel, K. Mechtler, R. Durbin, A. F. Stewart, J. M. Peters, F. Buchholz, and A. A. Hyman.** 2008. BAC TransgeneOmics: a high- throughput method for exploration of protein function in mammals. Nat Methods **5:**409-415.

104.   **Qian, J., J. Lin, N. M. Luscombe, H. Yu, and M. Gerstein.** 2003. Prediction of regulatory networks: genome-wide identification of transcription factor targets from gene expression data. Bioinformatics **19:**1917-1926.

105.   **Rao, S. S., M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden.** 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell **159:**1665-1680.

106.   **Reid, B. D., and P. Parsons.** 1971. Partial purification of mitochondrial RNA polymerase from rat liver. Proceedings of the National Academy of Sciences of the United States of America **68:**2830-2834.

107.   **Richard, P., X. Darzacq, E. Bertrand, B. E. Jady, C. Verheggen, and T. Kiss.** 2003. A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. Embo J **22:**4283-4293.

108.   **Rivera, C. G., R. Vakil, and J. S. Bader.** 2010. NeMo: Network Module identification in Cytoscape. BMC Bioinformatics **11 Suppl 1:**S61.

109.   **Rodriguez, A., and A. Laio.** 2014. Machine learning. Clustering by fast search and find of density peaks. Science **344:**1492-1496.

110.   **Rozowsky, J., G. Euskirchen, R. K. Auerbach, Z. D. Zhang, T. Gibson, R. Bjornson, N. Carriero, M. Snyder, and M. B. Gerstein.** 2009. PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. Nat Biotechnol **27:**66-75.

111.   **Rozowsky, J. S., D. Newburger, F. Sayward, J. Wu, G. Jordan, J. O. Korbel, U. Nagalakshmi, J. Yang, D. Zheng, R. Guigo, T. R. Gingeras, S. Weissman, P. Miller, M. Snyder, and M. B. Gerstein.** 2007. The DART classification of unannotated transcription within the ENCODE regions: associating transcription with known and novel loci. Genome research **17:**732-745.

112.   **Sabath, I., A. Skrajna, X. C. Yang, M. Dadlez, W. F. Marzluff, and Z. Dominski.** 2013. 3'-End processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. RNA **19:**1726-1744.

113.   **Salzler, H. R., D. C. Tatomer, P. Y. Malek, S. L. McDaniel, A. N. Orlando, W. F. Marzluff, and R. J. Duronio.** 2013. A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs. Dev Cell **24:**623-634.

114.   **Sapra, A. K., M. L. Anko, I. Grishina, M. Lorenz, M. Pabis, I. Poser, J. Rollins, E. M. Weiland, and K. M. Neugebauer.** 2009. SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. Mol Cell **34:**179-190.

115.   **Sboner, A., L. Habegger, D. Pflueger, S. Terry, D. Z. Chen, J. S. Rozowsky, A. K. Tewari, N.**

Kitabayashi, B. J. Moss, M. S. Chee, F. Demichelis, M. A. Rubin, and M. B. Gerstein. 2010. FusionSeq: a modular framework for finding gene fusions by analyzing paired-end RNA-sequencing data. Genome Biol **11:**R104.

116. **Shepard, P. J., and K. J. Hertel.** 2009. The SR protein family. Genome Biol **10:**242.

117. **Shevtsov, S. P., and M. Dundr.** 2011. Nucleation of nuclear bodies by RNA. Nat Cell Biol **13:**167-173.

118. **Shiels, C., S. A. Islam, R. Vatcheva, P. Sasieni, M. J. Sternberg, P. S. Freemont, and D. Sheer.** 2001. PML bodies associate specifically with the MHC gene cluster in interphase nuclei. J Cell Sci **114:**3705-3716.

119. **Stanek, D., and K. M. Neugebauer.** 2004. Detection of snRNP assembly intermediates in Cajal bodies by fluorescence resonance energy transfer. J Cell Biol **166:**1015-1025.

120. **Stanek, D., S. D. Rader, M. Klingauf, and K. M. Neugebauer.** 2003. Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. J Cell Biol **160:**505-516.

121. **Stejskalova, E., and D. Stanek.** 2014. The splicing factor U1-70K interacts with the SMN complex and is required for nuclear gem integrity. J Cell Sci **127:**3909-3915.

122. **Strzelecka, M., A. C. Oates, and K. M. Neugebauer.** 2010. Dynamic control of Cajal body number during zebrafish embryogenesis. Nucleus **1:**96-108.

123. **Strzelecka, M., S. Trowitzsch, G. Weber, R. Luhrmann, A. C. Oates, and K. M. Neugebauer.** 2010. Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. Nature structural & molecular biology **17:**403-409.

124. **Sugimoto, Y., J. Konig, S. Hussain, B. Zupan, T. Curk, M. Frye, and J. Ule.** 2012. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. Genome Biol **13:**R67.

125. **Tripathi, V., J. D. Ellis, Z. Shen, D. Y. Song, Q. Pan, A. T. Watt, S. M. Freier, C. F. Bennett, A. Sharma, P. A. Bubulya, B. J. Blencowe, S. G. Prasanth, and K. V. Prasanth.** 2010. The nuclear- retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Molecular cell **39:**925-938.

126. **Tripathi, V., D. Y. Song, X. Zong, S. P. Shevtsov, S. Hearn, X. D. Fu, M. Dundr, and K. V. Prasanth.** 2012. SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. Molecular biology of the cell **23:**3694-3706.

127. **Tripsianes, K., T. Madl, M. Machyna, D. Fessas, C. Englbrecht, U. Fischer, K. M. Neugebauer, and M. Sattler.** 2011. Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins. Nature structural & molecular biology **18:**1414-1420.

128. **Tycowski, K. T., M. D. Shu, A. Kukoyi, and J. A. Steitz.** 2009. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. Molecular cell **34:**47-57.

129. **Wang, Z., M. Gerstein, and M. Snyder.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10:**57-63.

130. **Wei, Y., J. Jin, and J. W. Harper.** 2003. The cyclin E/Cdk2 substrate and Cajal body component p220(NPAT) activates histone transcription through a novel LisH-like domain. Molecular and cellular biology **23:**3669-3680.

131. **West, J. A., C. P. Davis, H. Sunwoo, M. D. Simon, R. I. Sadreyev, P. I. Wang, M. Y. Tolstorukov, and R. E. Kingston.** 2014. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Molecular cell **55:**791-802.

132. **Windhager, L., T. Bonfert, K. Burger, Z. Ruzsics, S. Krebs, S. Kaufmann, G. Malterer, A. L'Hernault, M. Schilhabel, S. Schreiber, P. Rosenstiel, R. Zimmer, D. Eick, C. C. Friedel, and L. Dolken.** 2012. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. Genome research **22:**2031-2042.

133. **Xia, Y., L. J. Lu, and M. Gerstein.** 2006. Integrated prediction of the helical membrane protein interactome in yeast. J Mol Biol **357:**339-349.

134. **Yan, K. K., G. Fang, N. Bhardwaj, R. P. Alexander, and M. Gerstein.** 2010. Comparing genomes to computer operating systems in terms of the topology and evolution of their regulatory control networks. Proceedings of the National Academy of Sciences of the United States of America **107:**9186-9191.

135. **Yip, K. Y., C. Cheng, N. Bhardwaj, J. B. Brown, J. Leng, A. Kundaje, J. Rozowsky, E. Birney, P. Bickel, M. Snyder, and M. Gerstein.** 2012. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. Genome Biol **13:**R48.

136. **Yip, K. Y., H. Yu, P. M. Kim, M. Schultz, and M. Gerstein.** 2006. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. Bioinformatics **22:**2968-2970.

137. **Yu, H., and M. Gerstein.** 2006. Genomic analysis of the hierarchical structure of regulatory networks. Proceedings of the National Academy of Sciences of the United States of America **103:**14724-14731.

138. **Yu, H., N. M. Luscombe, J. Qian, and M. Gerstein.** 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends Genet **19:**422-427.

139. **Yu, H., A. Paccanaro, V. Trifonov, and M. Gerstein.** 2006. Predicting interactions in protein networks by completing defective cliques. Bioinformatics **22:**823-829.

140. **Yu, H., X. Zhu, D. Greenbaum, J. Karro, and M. Gerstein.** 2004. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. Nucleic acids research **32:**328-337.