Big (sequencing) future of non-coding RNA research for the understanding of cocaine

Da-Yu Wu*

Division of Basic Neuroscience and Behavior Research, National Institute on Drug Abuse, National Institutes of Health, Bethesda, MD, USA
*Correspondence: wudy@nida.nih.gov

Edited by:
Andre Pietrzykowski, Rutgers University, USA

Reviewed by:
Andre Pietrzykowski, Rutgers University, USA

Recent advancement in genomic and genetic sequencing technology has ushered in a new era in which unprecedented amount of genomic sequence and transcript sequence data with extraordinary detail can be generated at incredibly short time. As part of this advancement, the approaches that utilize the powerful next generation sequencing technology, such as RNA-Seq, may have dramatically changed the field of non-coding RNA (ncRNA) research. Since these new approaches do not require prior knowledge of annotated transcripts for probes, theoretically the entire transcriptome of a given sample can be sequenced. This enables the detection of novel transcripts, including both protein coding and ncRNA, as well as RNA with somatic mutations and alternative splicing forms (Trapnell et al., 2009, 2010; Au et al., 2010; Griffith et al., 2010). Before this recent technological advancement, in the more traditional approaches for ncRNA studies, such as using RNA microarray technology, prior sequence knowledge of ncRNA in a cell, tissue, or an organism is required to first generate the microarray. Known RNA and RNA variants of a sample are then hybridized to the array, detected, quantified, and analyzed, while unknown RNA or RNA variants remain undetected and unanalyzed. This is especially limiting for the studying of many classes of ncRNA, such as long ncRNA (lncRNA, which are ncRNA that are longer than 200 nucleotides), due to somatic mutations, epigenetic consequences, and the possibility of secondary structures that longer RNA often take (Wang and Chang, 2011; Mercer et al., 2012; Moran et al., 2012). Some additional advantages of RNA-Seq that have been reported in studying protein coding transcripts, including the capability of detecting high dynamic range of gene expression level, and its high precision and high reproducibility (Marioni et al., 2008; Mane et al., 2009; Bradford et al., 2010; Chen et al., 2010; Twine et al., 2011; Zhang et al., 2012), will likely further ratify this new trend of ncRNA research. Nevertheless, currently microarray is still a valid technology for RNA studies in many other aspects, for its speed, specificities for targeted probing, and its low cost. For example, since statistically 5% of genes give rise to 75% of housekeeping transcripts, much effort is required in RNA-Seq to filter out high abundance transcripts before one can piece together and quantify the interested lower abundance transcripts in a typical RNA sample. Cost is another limiting factor for the migration to RNA-Seq. A typical RNA-Seq run can cost a few thousands, while a typical run for a targeted high throughput microarray costs around hundreds in dollar amount. In the foreseeable future, microarray and RNA-Seq may co-exist and each provides its own aspects of advantage.

If the size of human genome truly reflects the evolution pressure and practical genomic requirements for biological function and survival, one may conclude that the number of genes we have studied and all the facts we have discovered on genes and gene regulation is just the tip of an iceberg, because merely 1.5–2% of nucleotide (nt) bases in the human genome are transcribed into genes that code for proteins (Wolfsberg et al., 2001; The ENCODE Project Consortium, 2007), which has been the main research emphasis for genetic research for the last several decades. With the advancement of modern technologies, including RNA-Seq and microarray technology, it is clear that new information on genomes, transcripts, and their regulation will increase exponentially in the coming decades (Schulz et al., 2008; Zerbino et al., 2012). How to manage, mine, and comprehend these large data will be a real challenge. Meantime the reward can be high and the effort may be well justifiable, considering the likely impact the new knowledge will have on disease research and treatment.

A prominent member of ncRNA in brain function and psychiatric disorder is microRNA (miRNA), which is believed to be the transcripts of 1–3% of the human genome. miRNA is highly involved in brain development and plasticity at the neuronal level (Kapsimali et al., 2007; Choi et al., 2008; Cheng et al., 2009; Liu et al., 2010; Shi et al., 2010; also see reviews by Saba and Schratt, 2010; Konopka et al., 2011; Im and Kenny, 2012). This ncRNA regulates gene expression through RNA interference, RNA degradation, DNA methylation, and chromatin remodeling. As reviewed by two articles in this special issue from Kenny and Mains groups, recent discoveries of miRNA function and mechanisms in cocaine addiction have demonstrated important roles of ncRNA in this psychiatric disorder. For example, in the dorsal striatum, miR-212 was upregulated in rats addicted to cocaine and this upregulation enhances miR-212 regulated CREB signaling and counters the motivational property of cocaine (Hollander et al., 2010). Further studies found this specific function of miR-212 involves its negative homeostatic interaction with transition of acquizition period of addiction to compulsive-like increase of craving of cocaine. The downstream signaling pathway is further revealed in the same study to be through the controlling of BDNF levels in dorsal striatum (Im et al., 2010). Corroborative evidence of cocaine induced various miRNA expression and their functional role in cocaine addiction have also emerged in more and more other laboratories, complementing many aspects of our understanding (Nudelman et al.,

2010; Schaefer et al., 2010; Chandrasekar and Dreyer, 2011; Eipper-Mains et al., 2011). Particularly remarkable in one of these studies is that, using RNA-Seq approaches, cocaine is found to both upregulate and downregulate specific miRNA or family of miRNA (Eipper-Mains et al., 2011). While similar findings have been reported earlier in the nervous system (Vo et al., 2005; Krol et al., 2010; Nudelman et al., 2010), this RNA-Seq study is the first comprehensive analysis of cocaine induced changes of miRNA, and has demonstrated the advantage of next generation sequencing technology over microarray technology, which is often weakened by cross hybridization and high background signal, and critically limited by its prerequisite of prior knowledge of ncRNA identity it analyzes.

A much less tapped area in ncRNA research for the understanding of addiction of cocaine, or other substances of abuse, is the lncRNA; Wapinski and Chang, 2011; Huang et al., 2012; Jeggari et al., 2012; Moran et al., 2012). lncRNA are those ncRNA that are longer than 200 nucleotides and are oftentimes transcripts of intergenic regions between transcription clusters, or "foci," in the genome. Some earlier projects identified around 35,000 long non-coding transcripts from estimated 10,000 distinct loci in mammalian genome, often bearing signatures of protein coding mRNA, such as 5′ capping, splicing, and poly adenylation, but have no apparent open reading frame (ORF) for protein translation (Carninci, 2005, 2006; Kapranov et al., 2010). Remarkably, many more lncRNA may exist but not recognized due to the fact that the majority of lncRNA transcripts may not be poly adenylated (Cheng et al., 2005; Carninci, 2006; Kapranov et al., 2010), while many of the transcripts detection approaches rely on first hybridizing the poly adenylated region. Functionally, lncRNA have been found to have roles in epigenetic regulation, imprinting, and X-chromosome inactivation, in addition to regulation of gene transcription and translation (Mercer et al., 2009; Wang and Chang, 2011; Wapinski and Chang, 2011; Harries, 2012; Huang et al., 2012; Moran et al., 2012). Consequently, lncRNA may play significant roles in cocaine addiction. Indeed, it is likely that future work will identified high numbers of lncRNA in brain areas of the putative learning-reward-addiction neural

circuits such as nucleus accumbens, and will likely shed light on important involvement of lncRNA in these brain regions for reward and addiction. How genetic and epigenetic factors interact with these lncRNA, and how changes were brought upon these lncRNA by brain activities, in terms of lncRNA mutation, expression level, dynamics, and function, may provide significant insights on the mechanisms of cocaine addiction. Because substance abuse and dependence involves substantial gene-environment interactions and epigenetic factors, the capability of RNA-Seq in detecting novel gene variations and *de novo* gene mutations also provides valuable and necessary tools for the addiction research field.

The complexity and abundance of ncRNA, especially lncRNA, mandate advanced non-traditional, large scale and high throughput transcript sequencing, and analysis. While microarray technology enjoys established base and proven utility, the next generation sequencing technology represented by RNA-Seq emerges as the future choice approach for the task because of its unbiased coverage of the entire transcriptome of the genome. In addition, RNA-Seq allows versatile experimental design for sequencing depth and adjustable detection sensitivity, using machines of different throughputs and multiplexing different numbers of samples on a sequencing lane. Furthermore, the inherited benefit of RNA-Seq (Mane et al., 2009; Bradford et al., 2010; Chen et al., 2011; Eipper-Mains et al., 2011; also see reviews in this issue), including low background noise, low error rate and low technical variance, the high accuracy of the digital nature, and the capability of detecting novel transcripts and alternative splicing forms, will definitely benefit not only ncRNA analysis, but all other forms of RNA research as well. All these will greatly enable researchers in the field of substance abuse and dependence for future discoveries and breakthroughs.

REFERENCES

Au, K. F., Jiang, H., Lin, L., Xing, Y., and Wong, W. H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res.* 38, 4570–4578.

Bradford, J. R., Hey, Y., Yates, T., Li, Y., Pepper, S. D., Crispin, J., and Miller, C. J. (2010). A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genet.* 11, 282. doi: 10.1186/1471-2164-11-282

Carninci, P. (2006). Tagging mammalian transcription complexity. *Trends Genet.* 22, 501–510.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis V., Allen, J. E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K. C., Pavan, W. J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S. L., Sandelin, A., Schneider, C., Schönbach, C., Sekiguchi, K., Semple, C. A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegner, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R., Wei, C. L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S. M., Teasdale, R. D., Liu, E. T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J. S., Hume, D. A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., and Hayashizaki, Y. (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.

Chandrasekar, V., and Dreyer, J. L. (2011). Regulation of MiR-124, Let-7d, and MiR-181a in the accumbens affects the expression, extinction, and reinstatement of cocaine-induced conditioned place preference. *Neuropsychopharmacology* 36, 1149–1164.

Chen, H., Liu, Z., Gong, S., Wu, X., Taylor, W. L., Williams, R. W., Matta, S. G., and Sharp, B. M. (2011). Genome-wide gene expression profiling of nucleus accumbens neurons projecting to ventral pallidum using both microarray and transcriptome sequencing. *Front. Neurosci.* 5:98. doi: 10.3389/fnins.2011.00098

Chen, S., Yang, P., Jiang, F., Wei, Y., Ma, Z., and Kang, L. (2010). De novo analysis of transcriptome dynamics in the migratory locust during the development of phase traits. *PLoS One* 5, e15633. doi: 10.1371/journal.pone.0015633

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D. K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D. S., and Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308, 1149–1154.

Cheng, L. C., Pastrana, E., Tavazoie, M., and Doetsch, F. (2009). miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. *Nat. Neurosci.* 12, 399–408.

Choi, P. S., Zakhary, L., Choi, W. Y., Caron, S., Alvarez-Saavedra, E., Miska, E. A., McManus, M., Harfe, B., Giraldez, A. J., Horvitz, H. R., Schier, A. F., and Dulac, C. (2008). Members of the miRNA-200 family regulate olfactory neurogenesis. *Neuron* 57, 41–55.

Eipper-Mains, J. E., Kiraly, D. D., Palakodeti, D., Mains, R. E., Eipper, B. A., and Graveley, B. R. (2011). microRNA-Seq reveals cocaine-regulated expression of striatal microRNAs. *RNA* 17, 1529–1543.

Griffith, M., Griffith, O. L., Mwenifumbo, J., Goya, R., Morrissy, A. S., Morin, R. D., Corbett, R., Tang, M. J., Hou, Y. C., Pugh, T. J., Robertson, G., Chittaranjan, S., Ally, A., Asano, J. K., Chan, S. Y., Li, H. I., McDonald, H., Teague, K., Zhao, Y., Zeng, T., Delaney, A., Hirst, M., Morin, G. B., Jones, S. J., Tai, I. T., and Marra, M. A. (2010). Alternative expression analysis by RNA sequencing. *Nat. Methods* 7, 843–847.

Harries, L. (2012). Long non-coding RNAs and human disease. *Biochem. Soc. Trans.* 40, 902–906.

Hollander, J. A., Im, H. I., Amelio, A. L., Kocerha, J., Bali, P., Lu, Q., Willoughby, D., Wahlestedt, C., Conkright, M. D., and Kenny, P. J. (2010). Striatal microRNA controls cocaine intake through CREB signaling. *Nature* 466, 197–202.

Huang, Y., Liu, N., Wang, J. P., Wang, Y. Q., Yu, X. L., Wang, Z. B., Cheng, X. C., and Zou, Q. (2012). Regulatory long non-coding RNA and its functions. *J. Physiol. Biochem.* PMID: 22535282. [Epub ahead of print].

Im, H. I., Hollander, J. A., Bali, P., and Kenny, P. J. (2010). MeCP2 controls BDNF expression and cocaine intake through homeostatic interactions with microRNA-212. *Nat. Neurosci.* 13, 1120–1127.

Im, H. I., and Kenny, P. J. (2012). MicroRNAs in neuronal function and dysfunction. *TINS* 35, 325–334.

Jeggari, A., Marks, D. S., and Larsson, E. (2012). miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics* 28, 2062–2063.

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C. P., Sorensen, P. H., Reaman, G., Milos, P., Arceci, R. J., Thompson, J. F., and Triche, T. J. (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is "dark matter" un-annotated RNA. *BMC Biol.* 8, 149. doi: 10.1186/1741-7007-8-149

Kapsimali, M., Kloosterman, W. P., de Bruijn, E., Rosa, F., Plasterk, R. H. A., and Wilson, S. W. (2007). MicroRNAs show a wide diversity of expression profiles in the developing and mature central nervous system. *Genome Biol.* 8, R173.

Konopka, W., Schtz, G., and Kaczmarek, L. (2011). The microRNA contribution to leaning and memory. *Neuroscientist* 17, 468–474.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M. B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H. J., Schübeler, D., Oertner, T. G., Schratt, G., Bibel, M., Roska, B., and Filipowicz, W. (2010). Characterizing light-regulated retinal microRNAs reveals rapid turn-over as a common property of neuronal microRNAs. *Cell* 141, 618–631.

Liu, C., Teng, Z. Q., Santistevan, N. J., Szulwach, K. E., Guo, W., Jin, P., and Zhao, X. (2010). Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. *Cell Stem Cell* 6, 433–444.

Mane, S. P., Evans, C., Cooper, K. L., Crasta, O. R., Folkerts, O., Hutchison, S. K., Harkins, T. T., Thierry-Mieg, D., Thierry-Mieg, J., and Jensen, R. V. (2009). Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing. *BMC Genomics* 10, 264. doi: 10.1186/1471-2164-10-264

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18, 1509–1517.

Mercer, T. R., Dinger, M. E., and Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10, 155–159.

Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., Mattick, J. S., and Rinn, J. L. (2012). Targeted RNA sequenciung reveals the deep complexity of the human transcriptome. *Nat. Biotech.* 30, 99–104.

Moran, V. A., Perera, R. J., and Khalil, A. M. (2012). Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res.* 2012, 1–12.

Nudelman, A. S., DiRocco, D. P., Lambert, T. J., Garelick, M. G., Le, J., Nathanson, N. M., and Storm, D. R. (2010). Neuronal activity rapidly induces transcription of the CREB-regulated microRNA-132, in vivo. *Hippocampus* 20, 492–498.

Saba, R., and Schratt, G. M. (2010). MicroRNAs in neuronal development, function and dysfunction. *Brain Res.* 1338, 3–13.

Schaefer, A., Im, H. I., Veno, M. T., Fowler, C. D., Min, A., Intrator, A., Kjems, J., Kenny, P. J., O'Carroll, D., and Greengard, P. (2010). Argonaute 2 in dopamine 2 receptor-expressing neurons regulates cocaine addiction. *J. Exp. Med.* 207, 1843–1851.

Schulz, M. H., Zerbino, D. R., Vingron, M., and Birney, E. (2008). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092.

Shi, Y., Zhao, X., Hsieh, J., Wichterle, H., Impey, S., Banerjee, S., Neveu, P., and Kosik, K. S. (2010). MicroRNA regulation of neural stem cells and neurogenesis. *J. Neurosci.* 30, 14931–14936.

The ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.

Twine, N. A., Janitz, K., Wilkins, M. R., and Janitz, M. (2011). Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS ONE* 6, e16266. doi: 10.1371/journal.pone.0016266

Vo, N., Klein, M. E., Varlamova, O., Keller, D. M., Yamamoto, T., Goodman, R. H., and Impey, S. (2005). A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 102, 16426–16431.

Wang, K. C., and Chang, H. Y. (2011). Molecular mechanisms of long non-coding RNAs. *Mol. Cell* 43, 904–914.

Wapinski, O., and Chang, H. Y. (2011). Long non-coding RNAs and human disease. *Trends Cell Biol.* 21, 354–361.

Wolfsberg, T. G., McEntyre, J., and Schuler, G. D. (2001). Guide to the draft human genome. *Nature* 409, 824–826.

Zerbino, D. R., Paten, B., and Haussler, D. (2012). Integrating genomes. *Science* 336, 179–182.

Zhang, L. Q., Cheranova, D., Gibson, M., Ding, S., Daniel, P., Heruth, D. P., Fang, D., and Ye, S. Q. (2012). RNA-seq reveals novel transcriptome of genes and their isoforms in human pulmonary microvascular endothelial cells treated with thrombin. *PLoS ONE* 7, e31229. doi: 10.1371/journal.pone.0031229