RNA-Seq based characterization of long non-coding RNA involved in respiratory viruses pathogenesis

Laurence Josset[1,2], Martin T. Ferris[2,3,4], Lisa E. Gralinski[2,5], Nicolas Tchitchek[1,2], Richard Green[1,2], Matthew J. Thomas[1,2], Fernando Pardo-Manuel de Villena[4,5], Ralph S. Baric[2,5], Mark T. Heise[2,4], Xinxia Peng[1,2], and Michael G. Katze[1,2]

[1]Department of Microbiology, School of Medicine, University of Washington, Seattle, Washington, United States of America
[2]Pacific Northwest Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Research, Portland, Oregon, United States of America
[3]Carolina Vaccine Institute, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, United States of America
[4]Department of Genetics, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, United States of America
[5]Department of Epidemiology, University of North Carolina-Chapel Hill, Chapel Hill, North Carolina, United States of America

Outcome of respiratory virus infection is determined by a complex interplay of viral and host factors. Some potentially important host factors for the antiviral response, whose functions remain largely unexplored, are non-protein-coding RNAs (ncRNAs). Long non-coding RNAs (lncRNAs) are endogenous cellular RNA that are mRNA-like in length (> 200 nt) but are lacking any positive-strand open reading frames greater longer than 30 amino acids. Recent studies suggest that lncRNAs play regulatory roles in host response to pathogens. Here we aimed at systematically inferring the regulatory functions of host lncRNAs in response to influenza A virus and severe acute respiratory syndrome coronavirus (SARS-CoV) in the mouse model, using a 'guilt-by-association' approach which relies on finding which lncRNAs have similar expression profiles to protein-coding genes of known function. To build a large panel of diverse host responses to viral infection, we took advantage of the genetic diversity present in the 8 founder strains of the Collaborative Cross (CC) mouse resource. Mouse strains were grouped into susceptible or resistant groups based on weight loss and viral titers. Extensive pulmonary host-response profiling was performed on mock and viral-infected lungs at 2 and 4 days post-infection using total RNA-Seq. Overall lncRNAs accounted for about one-fourth of total genes differentially expressed upon infections. To predict the functions of these lncRNAs, we constructed a co-expression network using the weighted correlation network analysis (WGCNA) and identified modules of co-expressed genes. Several lncRNAs were identified as belonging to gene modules associated with viral replication or weight loss, and enriched in various infection-related biological processes such as immune response. Additional validation of lncRNA roles during viral infection was performed by examining their expression changes across additional RNA-Seq datasets, including interferon-treated mice and mice infected with highly pathogenic H5N1 virus. Altogether, these results provide a broad categorization of lncRNA functions and identify subsets of lncRNAs with potential key roles for respiratory virus pathogenesis.