Imaging Transcriptomics Linking To Brain-Wide Gene Expression and Neuroimaging Data

Paulette Jennifer[*]

Department of Biotechnology, Illinois University, Chicago, USA

ABOUT THE STUDY

Human imaging genetics has emerged as a powerful strategy for understanding the molecular basis of macroscopic neural phenotypes measured across the entire brain over the last two decades. This work has traditionally involved correlating allelic variation at one or more genetic loci with variation in one or more Imaging-Derived Phenotypes (IDPs), initially through candidate gene studies and more recently at the genome-wide level. The formation of large consortia, such as Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA), has aided the latter. A common assumption in this work is that variants associated with an IDP (or nearby variants tagged by the associated variant) influence gene expression or protein abundance, which affects cellular function and, ultimately, the studied IDP. However, a variety of environmental and other factors can influence gene activity, and the functional roles of many IDP-linked variants, which are typically discovered through large-scale statistical analyses, are frequently unknown. As a result, the mechanisms by which a particular variant may influence phenotypic variation are not always clear. Furthermore, the expression levels of many genes vary significantly across brain regions, and these spatial differences cannot be inferred solely from DNA sequence. Gene expression assays provide a more direct measure of gene function. Because expression assays are invasive and require direct access to neural tissue, studies of gene expression in the brain have historically been limited to small groups of areas studied in isolation. These constraints have recently been overcome by developments in high-throughput tissue processing and bioinformatics pipelines, leading to datasets of gene expression over a significant portion of the genome in numerous brain regions and at different phases of development. Only the Allen Human Brain Atlas (AHBA), which includes expression measurements for more than 20,000 genes obtained from 3702 geographically different tissue samples, offers high resolution coverage of virtually the entire brain, whereas several human atlases cover multiple brain regions. Importantly, the samples' stereotaxic mapping enables researchers to directly link spatial changes in gene expression to spatial variations in IDP. The emerging field of imaging transcriptomics has emerged as a result of this unprecedented ability to connect molecular function to large-scale brain organization. This field has begun to reveal new information about how regional differences in gene expression relate to functional connectivity within canonical resting-state networks, fiber tract connectivity between brain regions, temporal and topological properties of large-scale brain functional networks, and the specialization of cortical and subcortical networks. Additionally, software toolboxes for integrating data from transcriptome and imaging studies across the entire brain have been created. Imaging transcriptomics analyses are sometimes very multivariate, comprising measurements of the expression of about 20,000 genes in each of the 102–103 brain areas and being associated with one or more different IDPs assessed in each region. This requires quite a bit of data processing. Strategies for the correction of motion-related and global signal fluctuations in functional MRI serve as a perfect illustration of how data processing decisions affect the outcomes of neuroimaging research. The numerous processing options that can influence how transcriptome atlases are analyzed and how they relate to IDPs have not yet been subjected to comparable examination. As of this writing, more than 30 researches have connected AHBA gene expression measurements to information from human neuroimaging.

Because there is no standard processing pipeline for gene expression data, it is unclear how the results of this work are to different methodological choices. As the field evolves, it is critical to develop methodological guidelines to ensure consistent and reproducible results, as well as to support valid interpretation. In this paper, we provide a practical guide to some of the key steps in processing the AHBA gene expression data, as well as an examination of the potential impact of methodological options available at each step. We concentrate on the AHBA because it is the most comprehensive and widely used gene expression atlas in the field. We begin by summarizing some fundamental aspects of gene expression quantification and

Correspondence to: Paulette Jennifer, Department of Biotechnology, Illinois University, Chicago, USA, E-mail:paulette@iit.edu

Citation: Jennifer P (2022) Imaging Transcriptomics Linking To Brain-Wide Gene Expression and Neuroimaging Data. J Data Mining Genomics Proteomics. 13:258.

AHBA characteristics. Then, we outline several key steps in a basic workflow for relating gene expression measures to imaging data, and we investigate the impact of methodological choices at each stage. In the final section, we make some best practice recommendations and point you in the direction of additional research.

The recent availability of comprehensive, brain-wide gene expression atlases such as the Allen Human Brain Atlas (AHBA) has opened up new avenues for understanding how spatial variations on the molecular scale relate to macroscopic neuroimaging phenotypes. Although there is a rapidly growing body of literature demonstrating relationships between gene expression and various properties of brain structure and function, approaches for combining expression atlas data with neuroimaging are highly inconsistent, with significant variations in how the expression data are processed.

The extent to which these methodological variations influence findings is unknown. We present a seven-step analysis pipeline for relating brain-wide transcriptomic and neuroimaging data and compare how different processing options influence the resulting data. We propose that AHBA studies strive for a unified data processing pipeline to ensure consistent and reproducible results in this burgeoning field.