A Quantitative Comparison of Region-Specific Microarray Gene Expression Profiles Across Diseases in the Human Brain

Yasutaka Tanaka, Chenchao Zang, Ian Johnson

We propose a project to compare expression microarray data from disparate datasets in order to find relationships among various disease states in the human brain. We will gather data from the NCBI Gene Expression Omnibus and EBI ArrayExpress, among other sources. We will then normalize the datasets relative to control samples from the Allen Brain Atlas, which provides comprehensive, region-specific expression values, from a non-disease state. After normalization, we will calculate several statistical measurements to quantify the relationships among the differentially-expressed genes, including euclidian distance, Pearson correlation, and cosine similarity.

Initial GEO Data Set candidates (to be filtered down)

Human:

- GDS5204 - Age effect on normal adult brain: frontal cortical region
- GDS4879 - Chronic high-level alcohol consumption effect on brain: post-mortem hippocampus
- GDS4859 - Analysis of prolactinomas and normal pituitary tissues.
- GDS4838 - Analysis of central nervous system primitive neuroectodermal tumors (CNS PNET).
- GDS4758 - Analysis of postmortem brain tissues from Alzheimer's disease (AD) or an AD-like disorder.
- GDS4532 - Analysis of prefrontal/orbito-frontal and parieto-temporal cortex at 17 and 19 gestational weeks.
- GDS4523 - Analysis of post-mortem tissue from the anterior prefrontal cortex (Brodmann Area 10, BA10) of schizophrenic patients.
- GDS4522 - Analysis of post-mortem tissue from the superior temporal cortex (Brodmann Area 22, BA22) of schizophrenic patients.
- GDS4477 - Analysis of pediatric glioblastoma (GBM) tumors with mutations in H3F3A gene.
- GDS4414 - Amyloid Precursor Protein APP family members: adult prefrontal cortex
- GDS4358 - Two types of HIV-associated neurocognitive impairment: brain regions
- GDS4231 - Antiretroviral therapy effect on brain of patients with HIV-associated neurocognitive disorders
- GDS4218 - Multiple sclerosis: brain lesions

- GDS4154 - Parkinson's disease: post-mortem medullary regions
- GDS4136 - Various stages of Alzheimer's disease: laser-captured hippocampal CA1 gray matter
- GDS4135 - Postmortem temporal cortex from Medical Research Council Cognitive Function and Ageing Study (MRC-CFAS) cohort: astrocytes
- GDS3834 - Multiple normal tissues
- GDS3502 - Schizophrenia: endothelial and neuronal cells from postmortem dorsolateral prefrontal cortex tissue
- GDS3459 - Frontotemporal lobar degeneration with ubiquitinated inclusions and progranulin mutations: various brain regions
- GDS3345 - Various mental disorders: postmortem brains
- GDS3129 - Parkinson's disease: substantia nigra (HG-U133B)
- GDS3128 - Parkinson's disease: substantia nigra (HG-U133A)
- GDS3113 - Various normal tissues
- GDS3110 - Hypothalamic hamartoma and central precocious puberty
- GDS3069 - Various brain tumors
- GDS2978 - Multiple sclerosis: brain
- GDS2941 - Down syndrome: brain
- GDS2821 - Parkinson's disease: substantia nigra
- GDS2795 - Alzheimer's disease: neurofibrillary tangles
- GDS2613 - Rett syndrome: brain frontal cortex
- GDS2191 - Bipolar disorder: orbitofrontal cortex
- GDS2190 - Bipolar disorder: dorsolateral prefrontal cortex
- GDS2154 - Inflammatory dilated cardiomyopathy
- GDS1962 - Glioma-derived stem cell factor effect on angiogenesis in the brain
- GDS1917 - Cerebellar cortex in schizophrenia
- GDS1912 - X-linked recessive dystonia-parkinsonism
- GDS1835 - Various cell lines and Universal Reference RNA (II)
- GDS1816 - High-grade gliomas (HG-U133B)
- GDS1815 - High-grade gliomas (HG-U133A)
- GDS1813 - Glial brain tumors
- GDS1726 - HIV encephalitis: brain frontal cortex
- GDS1253 - Pituitary adenoma subtypes
- GDS1096 - Normal tissues of various types
- GDS1085 - Normal tissues of diverse types (SHBW)
- GDS910 - Alternative splicing in five tissues (dye-swap)
- GDS909 - Alternative splicing in five tissues
- GDS833, GDS832, GDS831, GDS830, GDS829 - lternative pre-mRNA splicing in various tissues and cell lines (Rosetta/Merck Splicing Chip 1 - 5)
- GDS811 - Amyloid precursor protein mutant overexpression in strain resistant to neuronal cell loss

- GDS810 - Alzheimer's disease at various stages of severity
- GDS707 - Aging brain: frontal cortex expression profiles at various ages
- GDS596 - Large-scale analysis of the human transcriptome (HG-U133A)
- GDS564 - Sex specific transcription in hypothalamus
- GDS426, GDS425, GDS424, GDS423, GDS422 - Normal human tissue expression profiling (HG-U95A)
- GDS232 - Medulloblastoma metastasis
- GDS181 - Large-scale analysis of the human transcriptome (HG-U95A)

Mouse:

- GDS682 - Down syndrome and brain (MG-U74B)
- GDS681 - Down syndrome and brain (MG-U74A)
- GDS6016 - Transcription factor engrailed-2 loss-of-function model of autism spectrum disorder: hippocampus
- + Optionally many more candidates

Issues:

- Ensuring that datasets are comparable
    - Used in same brain regions?
    - Targeted same cell types?
    - Enough overlap of probes?
    - Different probes —> Different binding affinities —> Incomparable data?

How to confirm results:

- Validate findings with known genetic associations between diseases

Possible additional and/or alternative steps:

- Test various machine learning models to predict disease given microarray expression values (Enough features / data?)
- Explore RNA-seq datasets for mutation differences across brain diseases