



**Karolinska  
Institutet**

**Institutionen för medicinsk epidemiologi och biostatistik**

# Genomics and bioinformatics strategies in the study of aging and Alzheimer disease

**AKADEMISK AVHANDLING**

som för avläggande av medicine doktorsexamen vid Karolinska  
Institutet offentligen försvaras i Atrium, Nobels väg 12b.

**Fredagen den 11 mars, 2011, kl 09.00**

av

**Mun-Gwan Hong**

*Huvudhandledare:*

Docent Jonathan A. Prince  
Karolinska Institutet  
Institutionen för medicinsk epidemiologi och  
biostatistik

*Bihandledare:*

Professor Nancy L. Pedersen  
Karolinska Institutet  
Institutionen för medicinsk epidemiologi och  
biostatistik

*Fakultetsopponent:*

Professor Kevin Morgan  
University of Nottingham  
Centre for Genetics and Genomics  
Nottingham, United Kingdom

*Betygsnämnd:*

Docent Johan Björkegren  
Karolinska Institutet  
Institutionen för Medicinsk Biokemi och  
Biofysik

Docent Maria Ankarcrona  
Karolinska Institutet  
Institutionen för Neurobiologi, Vårdvetenskap  
och Samhälle

Professor Mats Nilsson  
Uppsala universitet  
Institutionen för immunologi, genetik och  
patologi

**Stockholm 2011**

## ABSTRACT

To understand complex phenotypes, medical research has evolved from the study of single genes and proteins to approaches that encompass more comprehensive catalogues of molecules. Among the more widely used are genome-wide expression and high-throughput genotyping, the latter primarily making use of single nucleotide polymorphisms (SNPs) in what has been termed genome-wide association studies (GWAS). Because of the scale of the data sets that are being produced, unique problems have emerged that necessitate the extensive use of bioinformatics tools. This thesis has entailed the analysis of several such large data sets in the context of biological pathways and introduces several bioinformatics solutions. Paper III, IV, and V deal with this topic. This thesis is primarily oriented around the study of Alzheimer disease (AD) and aging. The questions about the etiology of AD are often concurrent with questions about the biology of aging. This thesis pursues insight on genomic factors pertaining to both inquiries, acknowledging that both the AD state and aging itself are complex and multi-factorial. Two constituent papers (I and III) address aging and two papers (II and V) deal with genetic models in the study of AD.

In paper I, we examined the association of age with several genetic markers in the insulin degrading enzyme (*IDE*) and explored possible molecular mechanisms. In contrast to women, both age-at-sampling and age-at-death of the males were significantly lower in individuals that were heterozygous at genetic loci spanning the *IDE* locus. Plasma insulin levels and the expression levels of the gene were found to be higher in those same heterozygous males.

In paper II, SNPs in 25 genes involved in cholesterol metabolism were tested for association with AD and dementia. Genetic markers in a large linkage disequilibrium block spanning *SREBF1*, *TOMIL2*, and *ATPAF2* were significantly associated with disease. Gene expression and gene network analyses supported the findings.

In paper III, we investigated the biological pathway basis of age in human brain and lymphocytes. Mitochondrial genes were negatively regulated in both tissue samples, while the protein translation genes appeared to decrease in lymphocytes but increase in brain. Those observations indicated that there are common themes across tissues, but also tissue specific changes in gene regulation. We also examined the genomic architecture of the age-regulated genes, and found that the expression of non-compact genes tend to decrease with advancing age.

A large number of genome-wide association studies (GWAS) have now been performed over the past few years. In paper IV, we developed a program that automates the conversion of SNPs to representative gene lists in order to facilitate the exploration of biological pathway in the context of GWAS.

In paper V, we employed the software developed in study IV to identify biological pathways enriched among the genes that were significantly associated from a GWAS of AD. Genes involved in intracellular protein transmembrane transport were found to be significantly overrepresented. These results highlighted the possibility that *TOMM40* contributes to AD pathology together with other translocases.

Through this thesis, several biological relationships have been identified linking AD and aging. Genetic markers in *IDE*, a gene previously claimed to be associated with AD, also associate with age. With advancing age, mitochondrial gene expression deteriorates significantly. *TOMM40* may contribute the AD pathology, together with other genes that encode proteins of the intracellular transmembrane protein transport pathway. Methodologically, pathway analyses were conducted successfully with the program, ProxyGeneLD. This enabled discoveries and discussion of the challenges that face the exploration of GWAS data sets in a pathway context. In the future, more sophisticated bioinformatics tools and enhanced gene annotation may lead to the discovery of the molecular mechanisms that dominate complex diseases and traits.