



Functional genomic correlation among complex diseases

Ya-Hsin Tien¹, Chueh-Lin Hsu¹, Feng-Hsiang Chung¹, Hoong-Chien Lee¹⁻³

¹Department of Biomedical Sciences and Engineering, National Central University, Chungli, Taoyuan City, Taiwan 32001

²Center for Dynamical Biomarkers and Translational Medicine, National Central University, Chungli, Taoyuan City, Taiwan 32001

³Department of Physics, Chung Yuan Christian University, Chungli, Taoyuan City, Taiwan 320238

Background

Complex diseases such as cancer, neurodegenerative disorders, diabetes mellitus, and cardiovascular disease are associated with altered interactions between thousands of genes. Although variants in many genetic and environmental factors have been associated with complex diseases, the inter-relationship among the genetic variations of these diseases remains largely unknown. Recently a matrix of significant comorbidity relationships between complex disorder and Mendelian disease pairs that gave insights to disease risk was constructed (Blair et al., 2013). Here, we construct an association matrix based on relative genetic variations among a set of five common diseases – type 2 diabetes, kidney disease, asthma, colon cancer, thyroid cancer – and six rare complex diseases – multiple sclerosis, Huntington disease, Gaucher disease, adrenoleukodystrophy, Williams syndrome, and Hutchinson-Gilford progeria syndrome.

Methods

We collected gene expression microarray data from the NCBI Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) database and other sources where available. For quality control we used Principle Component Analysis (PCA) (Abdi & Williams, 2010) to determine sample quality, and selected samples with PCA scores not less than 0.75. Twenty two datasets involving 11 complex diseases were selected: type 2 diabetes (T2D, 2 datasets selected), kidney disease (KD, 4), asthma (Asthma, 3), colon cancer (CA, 1), thyroid cancer (Thyroid, 5), Gaucher disease (GD, 2), adrenoleukodystrophy (ALD, 1), multiple sclerosis (MS, 1), Huntington disease (HD, 1), Williams syndrome (WS, 1), and Hutchinson-Gilford progeria syndrome (HGPS, 1) (Table 1). We used Gene Set Enrichment Analysis (GSEA, <http://www.broad.mit.edu/gsea/>) (Subramanian et al., 2005) to compute the enrichment scores (ESs) and corresponding p-values of the 8426 molecular signatures (MS) in the Molecular Signatures Database (MSigDB v5.0) that have 10 to 200 genes against the 22 disease datasets to generate a 8426 by 22 matrix of ES (and p-) values. Subsets of this matrix were used to cluster the 22 datasets to obtain two-way MS versus disease clusters. Geneset representing an MS cluster was obtained by collecting the leading-edge genes (LEGs) of the MSs within a cluster against all the disease datasets and selecting those genes among the LEGs having the highest 5% frequency of occurrence. The MS cluster geneset (MSCG) is then given a Gene Ontology (GO) analysis to identify significantly (over or under) represented GO terms.

Results

In our preliminary results we found 590 of the 8426 molecular signatures were significant, where a significant molecular signature (SMS) is one having a nominal p-value <0.05 (for normalized ES) for at least 6 of 22 disease datasets. These 590 SMSs were used to two-way hierarchically cluster the 22 disease datasets. The disease datasets formed two large clusters, one including T2D, GD, MS, WS, KD, and HGPS, and a second including Thyroid, CA, Asthma, ALD, and HD (Figure 1). The 590 SMSs formed three large clusters: C1 (cluster 1) with 83 SMSs; C2, 364; and C3, 143. The MSCG for each cluster was computed. Here we report only the GO analysis of the SMS clusters. The MSCG of C1 contains 160 genes. Significantly (up or down) expressed GO terms include metabolic process, catabolic process, and mitochondrion. Mitochondrial dysfunction has been reported as a cause for T2D (Lowell & Shulman, 2005), and mitochondrial DNA content in T2D patients has been reported to be down-regulated (Hsieh et al., 2011). The MSCG of C2 has 441 genes, and significantly expressed GO terms include cell cycle and protein complex. It has been reported that β -cell mass is decreased in type 2 diabetes and that the mechanism underlying this is increased β -cell apoptosis (Butler et al., 2003). The MSCG of C3 has 305 genes, and significantly expressed GO terms include expression of developmental process, extracellular matrix, and response to stimulus. Abnormal extracellular matrix affects cancer progression by directly promoting cellular transformation and metastasis (Lu, Weaver, & Werb, 2012) (Figure 2).

Conclusion

New functional genomic inter-disease relations revealed by this study may lead to new understanding of inter-relationship among common and complex diseases, disease risks, diagnosis, and prevention.

Table and Figures

Table 1. Twenty two microarray datasets for complex diseases.

Name	Type of cell/tissue	Type of source	PCA score	QC report
Common diseases				
Type 2 Diabetes, T2D_1	blood	patient	0.75	Success
Type 2 Diabetes, T2D_2	pancreatic islets	patient	0.85	Success
Kidney Disease, KD_1	renal cortical tissue	patient	1	Success
Kidney Disease, KD_2	renal cortical tissue	patient	1	Success
Kidney Disease, KD_3	renal cortical tissue	patient	1	Success
Kidney Disease, KD_4	kidney glomerulus	patient	0.94	Success
Asthma, Asthma_1	CD4+ T cells	patient	1	Success
Asthma, Asthma_2	CD4+ T cells	patient	1	Success
Asthma, Asthma_3	bronchial biopsies	patient	0.75	Success
Colon Cancer, CA	descending colon	patient	1	Success
Thyroid Cancer, Thyroid_1	thyroid	patient	0.88	Success
Thyroid Cancer, Thyroid_2	thyroid	patient	0.95	Success
Thyroid Cancer, Thyroid_3	thyroid	patient	1	Success
Thyroid Cancer, Thyroid_4	thyroid	patient	1	Success
Thyroid Cancer, Thyroid_5	thyroid	patient	1	Success
Rare diseases				
Gaucher Disease, GD_1	skin fibroblasts	cell line	1	Success
Gaucher Disease, GD_2	skin fibroblasts	cell line	1	Success
Adrenoleukodystrophy, ALD	dermal fibroblast	patient	0.9	Success
Multiple Sclerosis, MS	peripheral blood mononuclear cells	patient	0.81	Success
Huntington Disease, HD	whole blood	patient	0.93	Success
Williams Syndrome, WS	skin fibroblasts	patient	0.83	Success
Hutchinson-Gilford Progeria Syndrome, HGPS	primary fibroblast	cell line	0.94	Success

Figure 1. Gene-set correlation matrix (GSCM) was produced using 590 significantly enriched/depleted gene-sets (nominal p < 0.05 by GSEA algorithm) with normalized enrichment scores (NES).

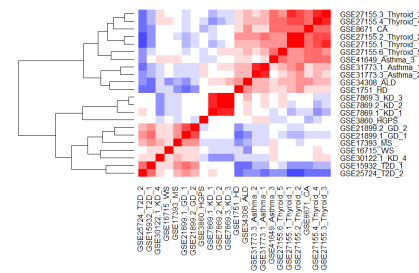
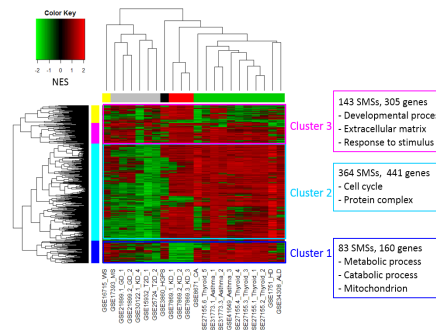


Figure 2. The figure shows the hierarchical clustering of 22 complex datasets, and report only the GO analysis of the SMS clusters in our studying.



References

- [1] Blair, David R, Lytle, Christopher S, Mortensen, Jonathan M, Bearden, Charles F, Jensen, Anders Boeck, Khabibani, Hossein, . . . Brunak, Søren. (2013). A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk. *Cell*, 155(1), 70-80.
- [2] Abdi, Hervé, & Williams, Lynne J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433-459.
- [3] Subramanian, Aravind, Tamayo, Pablo, Mootha, Vamsi K, Mukherjee, Sayan, Ebert, Benjamin L, Gillette, Michael A, . . . Lander, Eric S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43), 15545-15550.
- [4] Lowell, Bradford B, & Shulman, Gerald I. (2005). Mitochondrial dysfunction and type 2 diabetes. *Science*, 307(5708), 384-387.
- [5] Hsieh, Ching-Jung, Weng, Shao-Wen, Liou, Chia-Wei, Lin, Tsu-Kung, Chen, Jin-Bor, Tiao, Mao-Meng, . . . Wang, Pei-Wen. (2011). Tissue-specific differences in mitochondrial DNA content in type 2 diabetes. *Diabetes research and clinical practice*, 92(1), 106-110.
- [6] Butler, Alexandra E, Janson, Juliette, Bonner-Weir, Susan, Ritzel, Robert, Rizza, Robert A, & Butler, Peter C. (2003). β -cell deficit and increased β -cell apoptosis in humans with type 2 diabetes. *Diabetes*, 52(1), 102-110.
- [7] Lu, Pengfei, Weaver, Valerie M, & Werb, Zena. (2012). The extracellular matrix: a dynamic niche in cancer progression. *The Journal of cell biology*, 196(4), 395-406.