IMPROVER
systemsbiologyverification

MS: Genomic Biomarker Status and Challenge Scoring

Tamir Tuller,

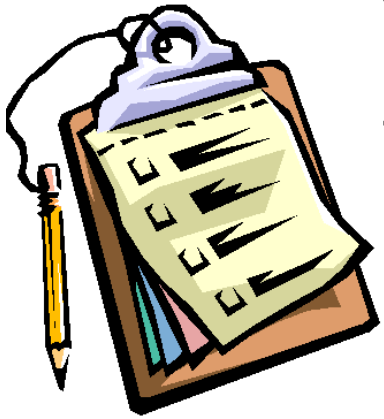
Laboratory of Computational Systems Biology,

Tel Aviv University

2nd October 2012

Agenda

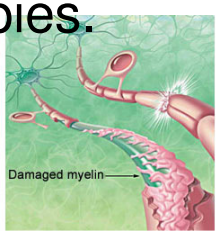
- **Introduction:** Multiple Sclerosis, autoimmune diseases, PBMC, previous studies.
- **The data**…
- **Challenges in MS predictions based on gene expression.**
- **The features** -- a short biological analysis.
- **Conclusions**



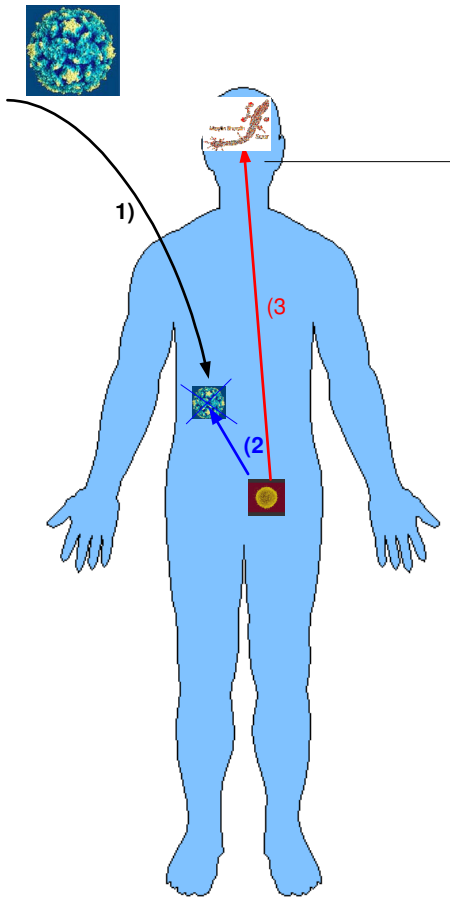
Introduction: Few Facts about Multiple Sclerosis

A chronic, inflammatory, demyelinating disease that affects the central nervous system.

- Up to 150 cases per 100,000 people.
- Up to 35% genetics (concordance in genetically identical twins), 65% environmental.
- Results from attacks to the nervous system by the body's own immune system. **An autoimmune disease.**
- The cause of MS remains elusive: A metabolically dependent disease?.. Might be caused by a virus (*e.g.* Epstein-Barr)?… Deficiency of vitamin D during childhood?..
- The disease does not have a cure, but there are several helpful therapies.



Multiple Sclerosis - Symptoms



- Attack the myelin (brain spinal cord).

- Problem in electrical signal conduction.

Muscle weakness

Abnormal muscle spasms

Difficulty in moving

Cognitive decline

Sensory problems

Difficulties with coordination and balance

Problems in speech or swallowing

Visual problems

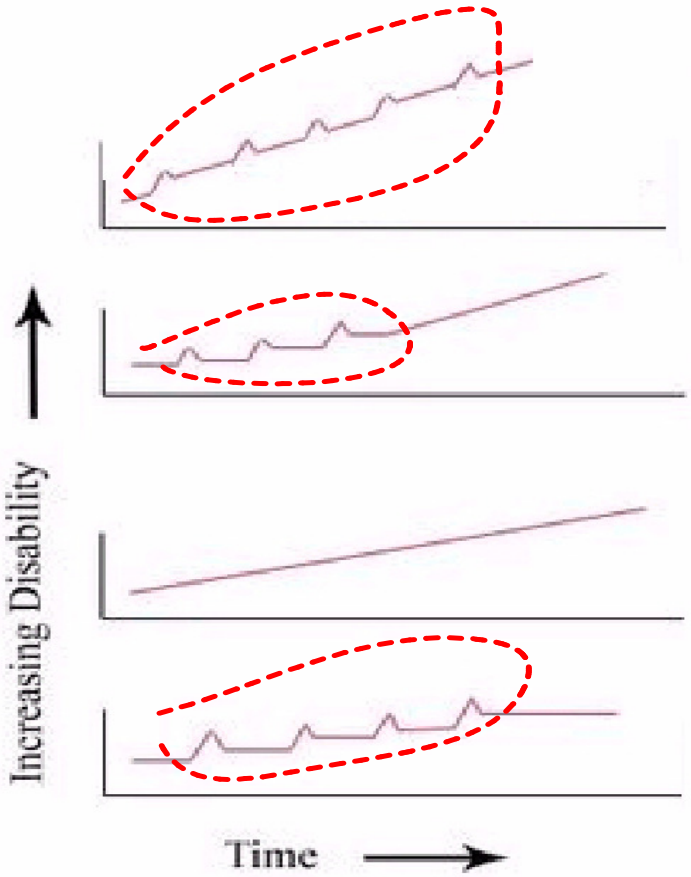
Fatigue and acute or chronic pain syndromes

Bladder and bowel difficulties

Cognitive impairment of varying degrees

Emotional symptomatology in the form of depression or pseudobulbar affect

Relapse vs. Remission in Multiple Sclerosis.



Progressive-relapsing multiple sclerosis: steady decline since onset with superimposed attacks

Secondary progressive multiple sclerosis: initial relapsing-remitting multiple sclerosis that suddenly begins to have decline without periods of remission

30% of the cases.

Primary progressive multiple sclerosis: steady increase in disability without attacks.

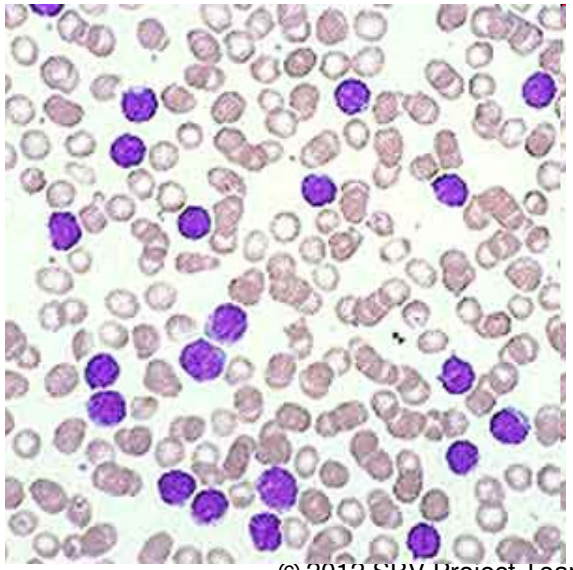
Relapsing-remitting multiple sclerosis: unpredictable attacks which may or may not leave permanent deficits followed by periods of remission

55% of the cases.

Peripheral Blood Mononuclear Cells (PBMC)

- T cells (CD4 and CD8 positive ~75%), B cells and NK cells (~25% combined).

- Critical components in the immune system.



A Few Previous Studies MS *vs.* Healthy

- Bomprezzi *et al.* 2003: 112 genes can differentiate MS from healthy.
- Achiron *et al.* 2004: Transcriptional signature of 1,109 genes in PBMCs from 26 MS patients, irrespective of disease activation state or immunomodulatory treatment: T-cell activation and expansion, inflammation, and apoptosis

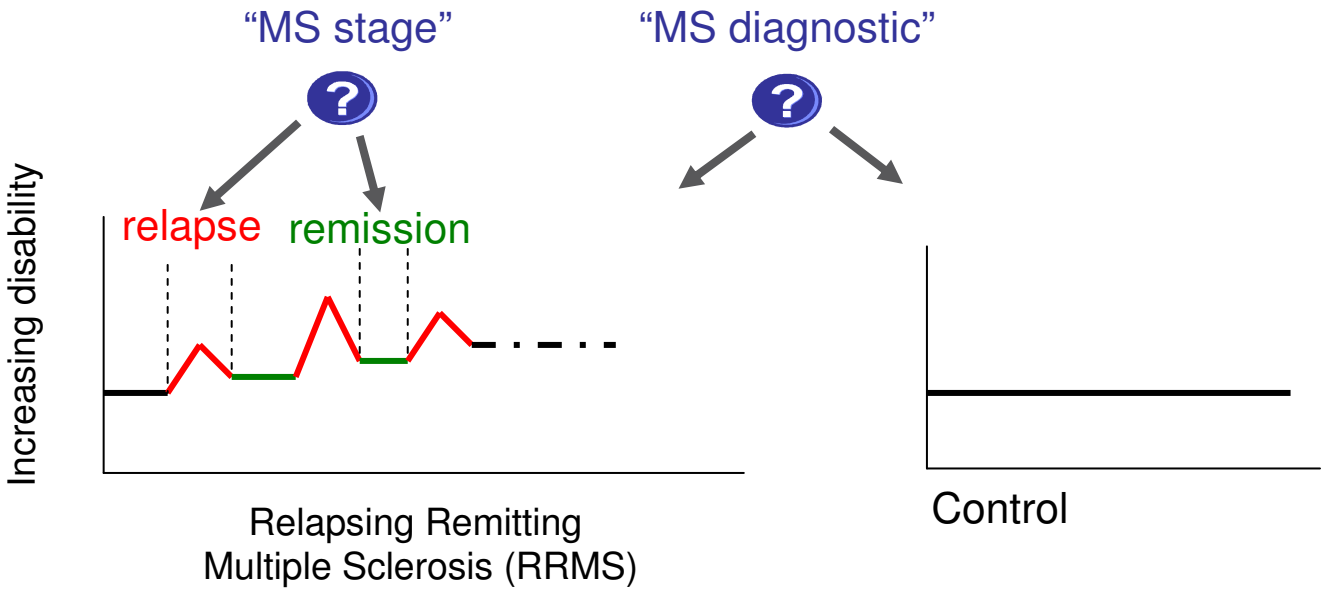
A Few Previous Studies Relapse *vs.* Remission

- Achiron *et al.* 2007: 1578 transcripts differ between acute relapse and remission in RRMS (among which many apoptosis-related).
- Arthur *et al.* 2007: Looks at relapse *vs.* remission in RRMS: genes involved in inflammation and apoptosis; paper focuses on few genes and point mutations
- Lindsey *et al.* 2011: 14 patients in relapse vs. remission; 71 transcripts change expression with $p < 0.001$.
- Gurevich *et al.* 2009: Prediction of acute multiple sclerosis relapses by transcription levels of peripheral blood cells.
- Tuller *et al.* 2011: Global signal of MS and rel/rem when considering protein-protein interactions.

Multiple Sclerosis (MS) Challenges

- Two challenges addressed MS. Their goals were to distinguish between:
  - Healthy *vs.* MS
  - Remitting *vs.* Relapsing RRMS patients

  Using the transcriptome of Peripheral Blood Mononuclear Cells (PBMC).



Multiple Sclerosis Diagnostic (MSD) Challenge – Test Dataset

Class	No Samples
Control	32
MS	28

All samples were acquired from living patients

Transcriptomics profiles were obtained from PBMCs

Additional information included age, gender, treatment, and comorbidities

Test dataset was licensed from GeneLogic (http://www.genelogic.com/)

Multiple Sclerosis Stage (MSS) Challenge – Test Dataset

Class	No Samples
Relapse	34
Remission	35

All samples were acquired from living patients

Transcriptomics profiles were obtained from PBMCs

Additional information included age and gender

Test dataset was provided by Tamir Tuller

This challenge was excluded from overall scoring because none of the teams could predict at levels better than random.

Challenges in MS Diagnosis Based on Gene Expression

- **Indirect measurements**: the tissue under attack is the brain myelin while we study PBMC (instead if CSF/Cerebrospinal Fluid).

- **The control group includes different set of subjects.** In the case of cancer we can compare healthy and cancerous tissue of the same patient.

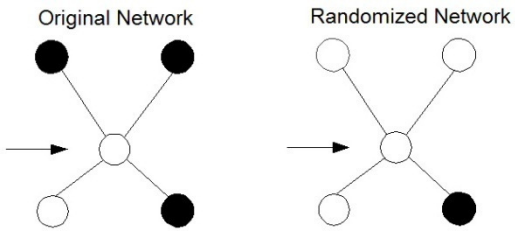
.

Challenges in MS Diseases Stage Predictions.

- The signal of relapse *vs.* remission is ***weaker*** than the signal of MS *vs.* control (*e.g.* Tuller *et al. Human Molecular Genetics* 2011).
- **Less training data** in the case of relapse *vs.* remission in comparison to MS *vs.* control.
- Probably there are **diagnostic errors** in the case of classifying patients to relapse/remission.
- Previous studies by us and others included **less trivial/conventional normalizations** and p-values to remove batch effects and demographical variance..
- Previous studies by us the **train and the test was from the same lab** (and generated by the same technician)… control for batch effects but also demographical/genetic variables.
- **The relapse group and the remission group includes different set of subjects.** In practice it is hard to follow the same patient over long period.

Do the Selected Features (Genes) Make Sense ?

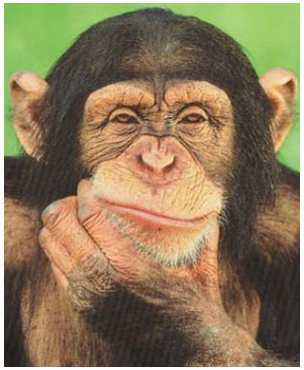
- GO enrichment of the consensus genes/features (used by at least one group) do not seems specific.
- PPI p-value: based on the number of neighbors (*vs.* the expected *no.* of neighbors) in the networks that are **consensus features**. (See Tuller *et al. Hum Mol Genet.* 2011).



- GO enrichment (+ FDR) of the consensus genes/features and corresponding PPI significant genes: apoptosis/anti-apoptosis/cell death, cell migration, T cell activation, phosphorylation..

Conclusions

- MS Diagnosis based on gene expression is not a trivial task .. Prediction of MS Stage is even more challenging ..
- The results of the MS Diagnosis challenge are significant and the selected features are biologically relevant ..
- The results of the MS Stage challenge are not significant .. However, I believe that there is a weak but significant signal .. Among others, more/'better' data, biological knowledge, experience with the biases of the data are needed to detect it ..



Thank You!