

Reproducible Research

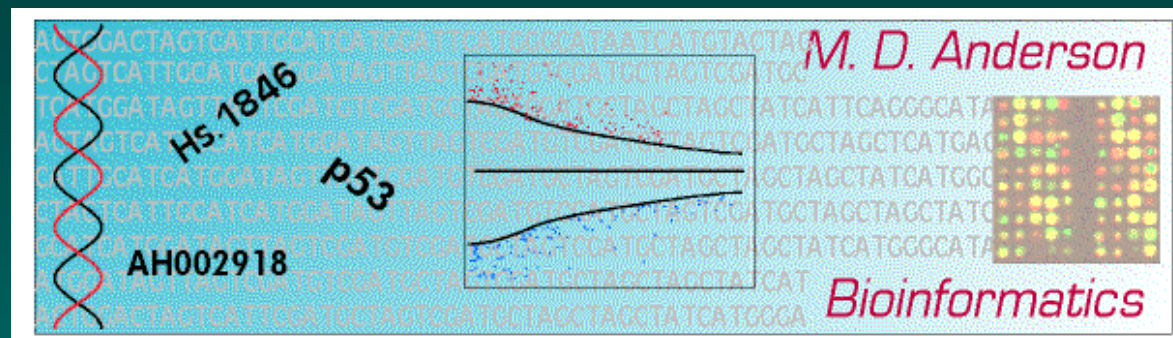
Kevin R. Coombes

Bioinformatics and Computational Biology

UT M. D. Anderson Cancer Center

kcoombes@mdanderson.org

BIOTECHNO; Venice, May 2011



Background

Duke University researchers have recently terminated three clinical trials and retracted four major publications.

These actions were in response to our unsuccessful attempts to reproduce the Duke analyses. We spent between 1500 and 2000 person-hours trying to figure out what computer code was applied to which data sets in order to assess the results.

- Nature Medicine. 2007; 13(11): 1276-7.
- J Clin Oncol. 2008; 26(7): 1186-7.
- Ann Appl Stat. 2010; 3(4): 1309-1334.
- Nature. 2010; 467(7314): 401.
- Clin Chem. 2011 Mar 1.
- <http://groups.google.com/group/reproducible-research>

Topics for Discussion

We published a letter in Nature recommending that the following material should be required:

- Primary data should be provided.
- Provenance of all data should be documented.
- All software codes/scripts should be provided.
- All non-scriptable analytical protocols should be described.
- Pre-specified research plans should be provided.

Questions for Discussion:

- Are the recommendations appropriate?
- How do we get journals, funding agencies, and academic institutions to implement these steps?
- What tools should be developed to make reproducibility easier?



INTERNAL FORCE FIELD IN PROTEINS

Damian Marchewka, Mateusz Banach, Irena Roterman

Department of Bioinformatics and Telemedicine

Jagiellonian University – Collegium Medicum

Faculty of Physics, Astronomy and Applied Computer Science

Jagiellonian University

Cracow – POLAND

FUZZY OIL DROP MODEL

THEORETICAL DISTRIBUTION - 3-D GAUSS FUNCTION

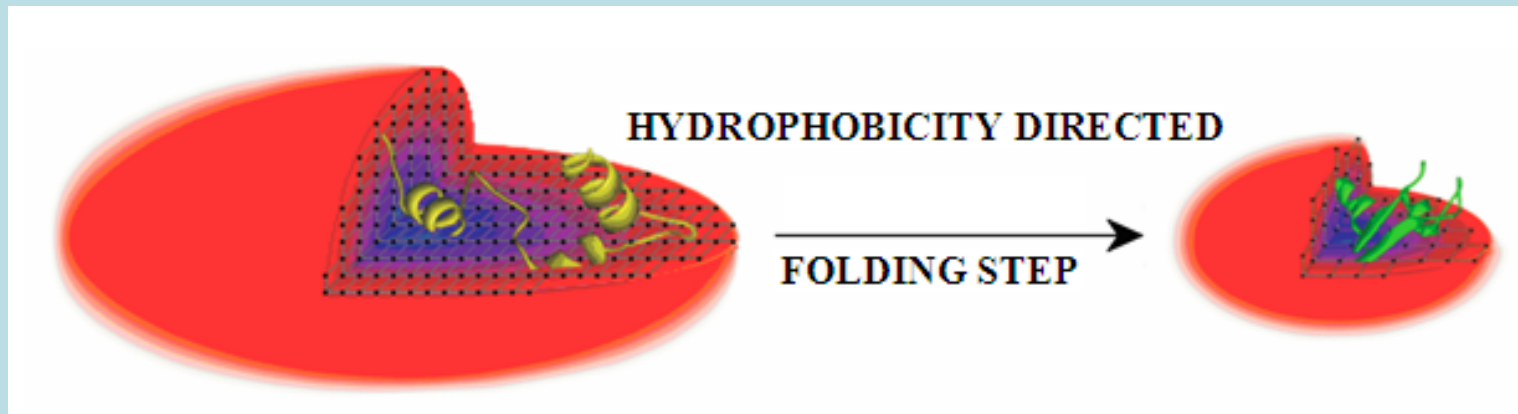
$$\tilde{H}t_j = \frac{1}{\tilde{H}t_{sum}} \exp\left(\frac{-(x_j - \bar{x})^2}{2\sigma_x^2}\right) \exp\left(\frac{-(y_j - \bar{y})^2}{2\sigma_y^2}\right) \exp\left(\frac{-(z_j - \bar{z})^2}{2\sigma_z^2}\right)$$

EMPIRICAL DISTRIBUTION – LEVITT FUNCTION

$$\tilde{H}o_j = \frac{1}{\tilde{H}o_{sum}} \sum_{i=1}^N H_i^r \begin{cases} \left[1 - \frac{1}{2} \left(7 \left(\frac{r_{ij}}{c} \right)^2 - 9 \left(\frac{r_{ij}}{c} \right)^4 + 5 \left(\frac{r_{ij}}{c} \right)^6 - \left(\frac{r_{ij}}{c} \right)^8 \right) \right] & \text{FOR } r_{ij} \leq c \\ 0 & \text{FOR } r_{ij} > c \end{cases}$$

FUZZY OIL DROP MODEL

HYDROPHOBIC
INTERACTION



Advances in BioSystems: Computational Modeling of Brain Functions

Marc Ebner

Eberhard Karls Universität Tübingen
Wilhelm Schickard Institut für
Informatik

Cognitive Systems

Sand 1, 72076 Tübingen, Germany

marc.ebner@wsii.uni-tuebingen.de

Brain Functions (Current Status)

Computational model for color perception/color constancy (Ebner 2007)

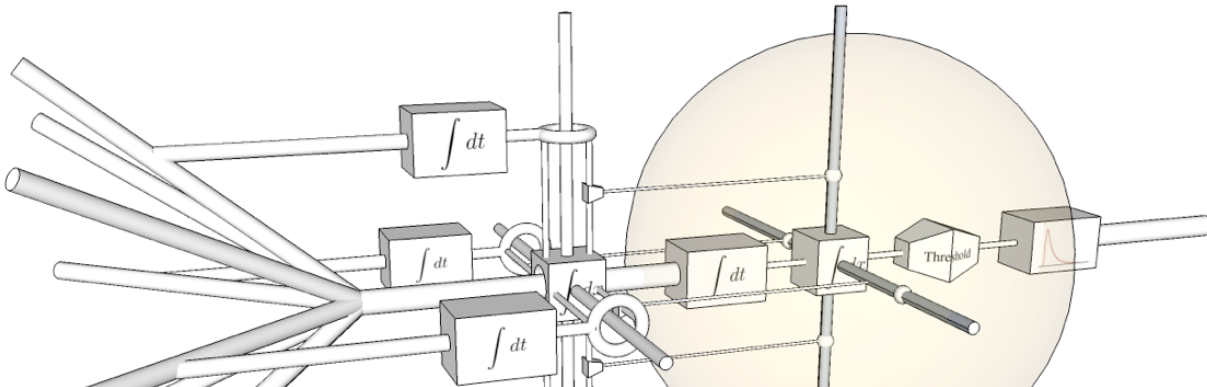
Computational model for figure/ground Separation (Ebner and Hameroff 2011)



Marc Ebner

COLOR CONSTANCY

WILEY



Computational Modeling of Brain Functions (Vision)

- Computational modeling required for understanding how the brain works.
- Experimental results have to be mapped to computational models.
- Eventually, modeling of higher brain functions will be possible.



Advances in Bio-Systems

Stephen Anthony, Enrico Coiera

The University of New South Wales, Centre for Health Informatics

Vitali Sintchenko

The University of Sydney, Centre for Infectious Diseases and Microbiology

Seamless Biological Visualisation

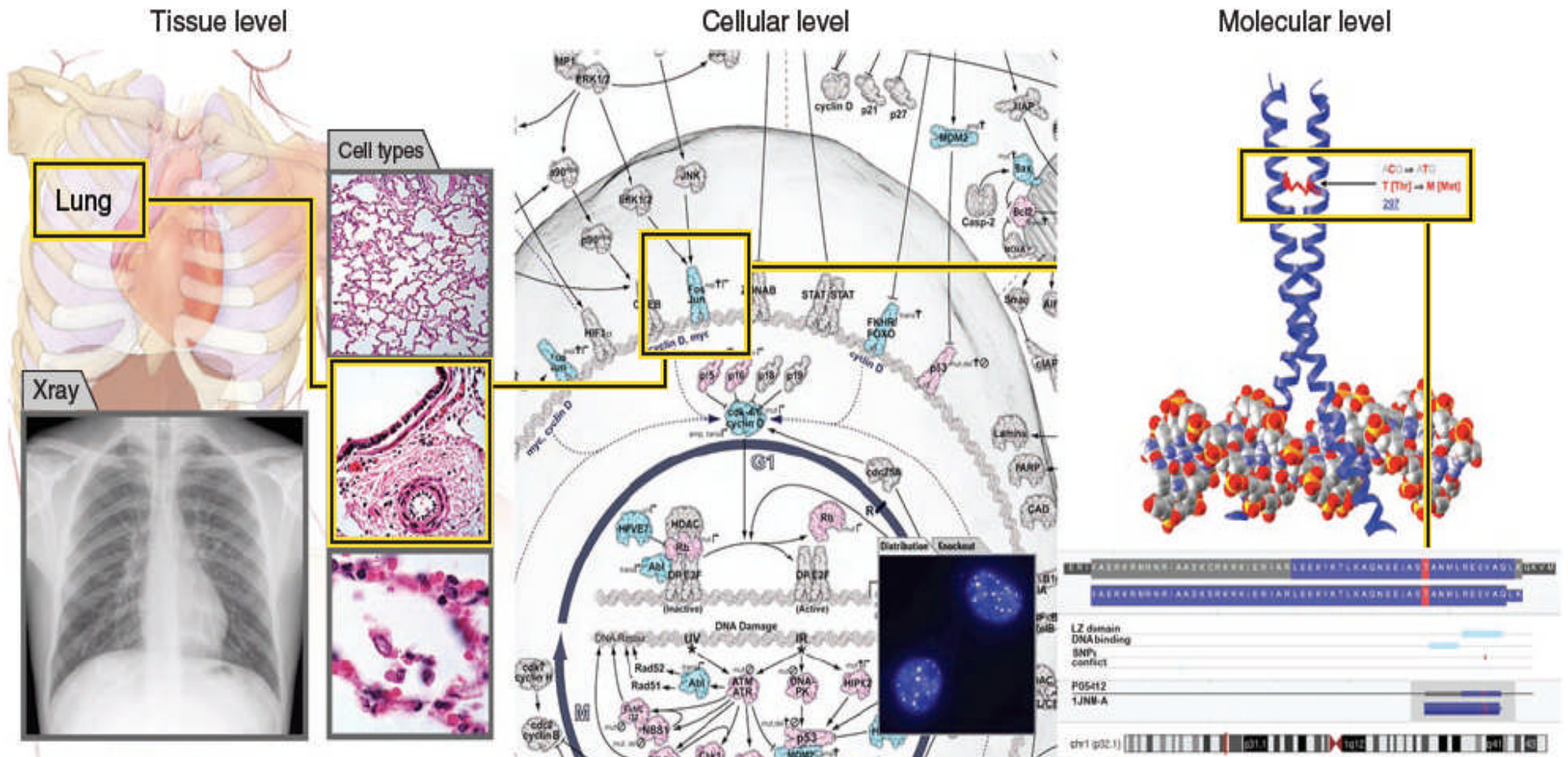


Image source: O' Donoghue, S. *et al.*, *Nature* 7(3), 2010.

Data Sources

Systemic diseases and sepsis

- Angiomatosis
- Bacteremia
- Endocarditis
- Endotoxic shock
- Influenza
- Lymphadenitis
- Sepsis
- Septic shock
- Toxic shock
- Reiter's
- Rheumatic fever
- Vasculitis
- Pleurodynia
- Prosthetic valve
- Purpuric fever

Respiratory tract infection

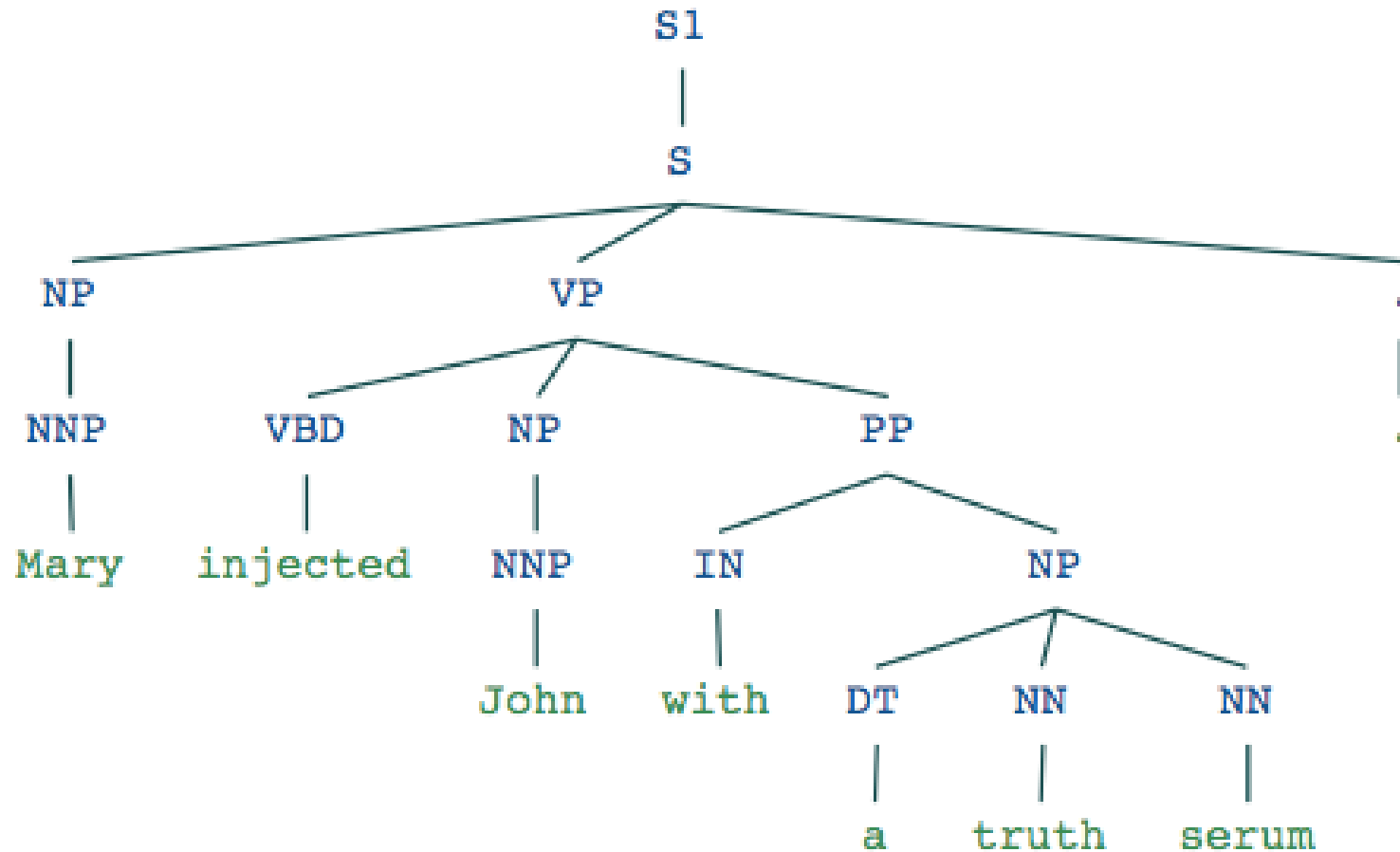
- Bronchiolitis
- Bronchitis
- Bronchopulmonary disease
- Coryza
- Croup
- Cystic fibrosis
- Empyema
- Epiglottitis
- Laryngitis
- Pneumonia
- Pneumonitis
- Tracheitis
- Tracheobronchitis

CNS infection

- Brain abscess
- Choriomeningitis
- Encephalitis
- Encephalopathy
- Mastoiditis
- Meningitis
- Meningoencephalitis
- Otitis
- Paralysis
- Sinusitis



Language Processing: Syntax



Language Processing: Semantics



Mary	injector [A0]
injected	V: inject
John	Injected into what? [A2]
with	
a	stuff being put in [A1]
truth	
serum	
.	



Charniak's Parse Tree

```
(S1 (S (NP (NNP Mary))
        (VP (VBD injected)
            (NP (NNP John))
            (PP (IN with)
                (NP (DT a)
                    (NN truth)
                    (NN serum))))))
      (. .)))
```



Fine-grained Information Extraction

Process
Virulence Factors

Protein Streptococci elaborate several factors implicated in **Process** infection, including surface-exposed adhesins and secreted toxigenic proteins (reviewed in [7,14,24]).

The initial statistical analysis identified four differentially expressed **Process** virulence genes (Tables 1 and 2).

Genes encoding streptolysin O (**Protein** slo or **Pro** spy0167) and the SpeB protease (**Protein** spy2039) were **-Reg -Reg** downregulated, while genes encoding pyrogenic exotoxin H (**Protein** speH or **Pro** spy1008) and a putative fibronectin-binding protein (**Protein** spy0130) were **+Reg +Reg** upregulated.

We verified the differential **Expression Expression** of spy2039 and spy0130 by qRT-PCR.

The downregulation of **Process** virulence loci during presumably inappropriate stages of **Process** infection was not surprising.

Protein Streptolysin O is a cytotoxin that damages **Org** human tissue and increases host cell cytotoxicity [7,25]. The resulting cellular damage, particularly to polymorphonuclear leukocytes [26], decreases internalization and subsequent **Organism** intracellular killing of streptococci [27].

Image source: BioNLP 2011 Bio-Event Extraction