

### **Supporting Materials for Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates**

This package includes the supporting materials to reproduce the results shown in the paper "Characterising the Alteration in the AP Distribution with the RSS Distance and the Position Estimates" accepted in the Ninth International Conference on Indoor Positioning and Indoor Navigation (IPIN 2018).

Organisations: Electrical Engineering

People: Torres-Sospedra, J. (Creator), Richter, P. (Creator), Mendoza-Silva, G. M. (Creator), Lohan, E. (Creator), Huerta, J. (Creator)

Publisher: Zenodo

Date made available: 4 Jul 2018

DOI: 10.5281/zenodo.1305414

Dataset

### **Data used in the paper: A compact constitutive model to describe the viscoelastic-plastic behaviour of glassy polymers: Comparison with monotonic and cyclic experiments and state-of-the-art models. Int. J Plasticity (2019).**

Tests were conducted at Univ. Bourgogne Franche-Comte, FEMTO-ST Institute, Department of Applied Mechanics, Besancon, France. The tests were conducted under isothermal conditions at room temperature when the frequency of 5 Hz and stress ratio of 0.5 were applied (for a PC -polymer). The elongation of dumbbell -shaped PC-specimens versus the applied force  $F$  was measured. The elongation was measured by an extensometer. The stress was defined by  $F/A$ , where the area of the initial cross section (gauge section) is given by  $A$ . Hysteresis loops given in the paper are based on this stress vs strain -data.

Organisations: Civil Engineering

People: Holopainen, S. (Creator), Barriere, T. (Creator)

Publisher: Tampere University of Technology

Date made available: 18 Aug 2019

DOI: 10.17632/7dpsyn8y62.1

Dataset

### **GATEMAN project – Wide-bandwidth, high-precision GNSS and jammer raw data**

This dataset contains high quality GNSS and jammer raw data generated during the in-lab validation activities of jamming detection and localization performed in the frame of the GATEMAN project.

Organisations: Research group: Wireless Communications and Positioning, Electronics and Communications Engineering

People: Richter, P. (Creator), Morales Ferre, R. (Creator), Lohan, E. (Creator)

Publisher: Tampere University of Technology

Date made available: 5 Mar 2019

DOI: 10.5281/zenodo.1161297

Dataset

### **PortSmash Proof-of-Concept exploit**

This is a proof-of-concept exploit of the PortSmash microarchitecture attack, tracked by CVE-2018-5407.

Organisations: Pervasive Computing, Research area: Information security, Research area: Information security, Department of Pervasive Computing

People: Aldaya, A. C. (Creator), Brumley, B. (Creator), Sohaib ul Hassan, N. (Creator), Pereida Garcia, C. (Creator), Tuveri, N. (Creator)

Publisher: Zenodo

Date made available: 29 Jan 2019

DOI: 10.5281/zenodo.2552315

Dataset

### **Electromagnetic (EM) side-channel traces of elliptic curve point multiplication during SM2 decryption in OpenSSL**

The dataset was used for Test Vector Leakage Assessment (TVLA) of SM2 decryption. The results are presented at ACSAC 2018 (<https://www.acsac.org/2018/>) in "Side-Channel Analysis of SM2: A Late-Stage Featurization Case Study" (DOI: 10.1145/3274694.3274725). The dataset contains Electromagnetic (EM) side-channel traces of elliptic curve point multiplication during SM2 decryption in OpenSSL. Folders Description EM\_traces\_before\_countermeasures : contains EM traces that were taken on unprotected version of SM2 decryption; EM\_traces\_after\_countermeasures : contains EM traces that were taken after all the countermeasures described in the paper were applied; Data Description These are Matlab generated files (.mat). Each file contains a single Matrix where each column represent an EM trace. The data type is double-precision floating-point (8-bytes). File Name Description TRACE\_fixed\* : contains traces with fixed key and fixed point (cipher text) TRACE\_rand\_key\* : contains traces with random key and fixed point (cipher text) TRACE\_rand\_point\* : contains traces with fixed key and random point (cipher text)

Organisations: Pervasive Computing, Research area: Information security

People: Tuveri, N. (Creator), Sohaib ul Hassan, N. (Creator), Pereida Garcia, C. (Creator), Brumley, B. (Creator)

Publisher: Tampere University of Technology

Date made available: 3 Dec 2018  
DOI: 10.5281/zenodo.1436828  
Dataset

### **ESR dataset**

Dataset for 'Comprehensive survey of Distance Function Effect on Classification of Electron Spin Resonance Spectra' article

Organisations: Signal Processing, Faculty of Computing and Electrical Engineering, Research group: MMDM  
People: Minaev, G. (Creator), Visa, A. (Supervisor)  
Publisher: Tampere University of Technology  
Date made available: 6 Mar 2018  
DOI: 10.5281/zenodo.1189252  
Dataset

### **WLAN (WiFi) RSS database for fingerprinting positioning**

This data set contains two WLAN Received Signal Strengths (RSS) databases suitable for fingerprinting positioning. One database contains training data (Training\_rss.csv, Training\_coordinates.csv), the radio map, the second database contains test data (Test\_rss.csv, Test\_coordinates.csv), RSS measurements on a path and the coordinates of that path. The data was collected in a three-floor building at Tampere University of Technology. The files Training\_rss.csv and Test\_rss.csv represent a matrix, with a reference point per row and an access point per column. The radio map consists of 446 reference points and 489 access points. Empty RSS values are set to 100. The files Training\_coordinates.csv and Test\_coordinates.csv contain the reference positions, 3D coordinates in a metric local reference frame. The data format allows to use previously published software (<https://doi.org/10.5281/zenodo.889797>) to analyze the data. The data is postprocessed: The reference positions of each floor are mapped onto a regular grid with 5 meter grid point spacing and the RSS values at each reference position are spatial averages of the RSS values in the resulting cells. The test data is mapped as well, but to a grid of 1 meter grid point spacing, from which only every third value was selected.

Organisations: Electronics and Communications Engineering, Research group: Wireless Communications and Positioning  
People: Richter, P. (Creator), Lohan, E. (Creator), Talvitie, J. (Creator)  
Publisher: Zenodo  
Date made available: 27 Jan 2018  
DOI: 10.5281/zenodo.1161524  
Links: <https://doi.org/10.5281/zenodo.1161524>  
Dataset

### **Geodatabase on Human activity patterns of Tampere**

A geographic database of human activity patterns extracted from manually categorised social media data platforms - Instagram.

Organisations: Architecture, Research group: Urban Planning and Design  
People: Cerrone, D. (Creator), Lehtovuori, P. (Supervisor)  
Publisher: Tampere University of Technology  
Date made available: 2017  
Dataset

### **Geodatabase on Human activity patterns of Turku**

A geographic database of human activity patterns extracted from manually categorised social media data platforms - Instagram.

Organisations: Architecture, Research group: Urban Planning and Design  
People: Cerrone, D. (Creator), Lehtovuori, P. (Supervisor), Ruoppila, S. (Supervisor)  
Publisher: Tampere University of Technology  
Date made available: 2017  
Dataset

### **Geodatabase on Human activity patterns of Helsinki**

A geographic database of human activity patterns extracted from manually categorised social media data platforms - Instagram.

Organisations: Architecture, Research group: Urban Planning and Design  
People: Cerrone, D. (Creator), Lehtovuori, P. (Supervisor)  
Publisher: Tampere University of Technology  
Date made available: 2017  
Dataset

### **Foursquare geodatabase for Turku**

I produced a dataset containing all available Foursquare amenities reclassified according to human activity patterns.

Organisations: Architecture, Research group: Urban Planning and Design

People: Cerrone, D. (Creator), Lehtovuori, P. (Supervisor), Ruoppila, S. (Supervisor)

Publisher: Tampere University of Technology

Date made available: 2017

Dataset

### **Foursquare geodatabase for Helsinki**

I produced a dataset containing all available Foursquare amenities reclassified according to human activity patterns.

Organisations: Architecture, Research group: Urban Planning and Design

People: Cerrone, D. (Creator), Lehtovuori, P. (Supervisor)

Publisher: Tampere University of Technology

Date made available: 2017

Dataset

### **Foursquare geodatabase for Tampere**

I produced a dataset containing all available Foursquare amenities reclassified according to human activity patterns.

Organisations: Architecture, Research group: Urban Planning and Design

People: Cerrone, D. (Creator)

Publisher: Tampere University of Technology

Date made available: 2016

Dataset

### **Long-Term Wi-Fi fingerprinting dataset and supporting material**

WiFi measurements database for UJI's library and supporting material. The measurements were collected by one person using one Android smartphone during 15 months at two floor of the library building from Universitat Jaume I, in Spain. It contains 63,504 WiFi fingerprints, which are organized into datasets. Each dataset is the result of a collection campaign. The supporting material includes Matlab® scripts to load and filter the desired data, and provides examples on possible studies that the database may enable. The supporting material also includes the bookshelve local coordinates.

Organisations: Electronics and Communications Engineering, Research group: Wireless Communications and Positioning

People: Mendoza-Silva, G. M. (Creator), Richter, P. (Contributor), Torres-Sospedra, J. (Contributor), Lohan, E. (Contributor), Huerta, J. (Contributor)

Publisher: Zenodo

Date made available: 24 Nov 2017

DOI: 10.5281/zenodo.1066040

Dataset

### **Crowdsourced WiFi database and benchmark software for indoor positioning**

This dataset contains two Wi-Fi databases (one for training and one for test/estimation purposes in indoor positioning applications), collected in a crowdsourced mode (i.e., via 21 different devices and different users), together with a benchmarking utility software (in Matlab and Python) to illustrate various algorithms of indoor positioning based solely on WiFi information (MAC addresses and RSS values). The data was collected in a 4-floor university building in Tampere, Finland, during Jan-Aug 2017 and it comprises 687 training fingerprints and 3951 test or estimation fingerprints.

Organisations: Electronics and Communications Engineering, Research group: Wireless Communications and Positioning, Research group: Wireless Communications and Positioning, Automation and Hydraulic Engineering, Research area: Dynamic Systems, Research group: Positioning, Research area: Dynamic Systems

People: Lohan, E. (Creator), Torres-Sospedra, J. (Contributor), Richter, P. (Contributor), Leppäkoski, H. (Contributor), Huerta, J. (Contributor), Cramariuc, A. (Contributor)

Publisher: Tampere University of Technology

Date made available: 12 Sep 2017

DOI: 10.5281/zenodo.889798

Dataset

### **WLAN RSS fingerprint database**

Three RSS data sets from adjacent indoor and outdoor areas. The databases together contain 857 fingerprints.

Fingerprint coordinates are two-dimensional in WGS 84 / Pseudo-Mercator (EPSG 3857) reference system. Arithmetic average of RSS and other statistics are included.

Organisations: Electronics and Communications Engineering, Research group: Wireless Communications and Positioning

People: Richter, P. (Creator)

Publisher: Multidisciplinary Digital Publishing Institute (MDPI)

Date made available: 8 Sep 2015

DOI: 10.3390/s150922587

Links: <http://www.mdpi.com/1424-8220/15/9/22587/s1>

Dataset

### Supplementary data for "A chirped photonic-crystal fibre"

Photonic crystals have widely increased the facility to guide and confine light at wavelengths close to the optical wavelength<sup>1, 2, 3</sup>. Because they can include extremely sharp bends, photonic-crystal waveguides are a key element in future integrated optical devices<sup>4</sup>. Moreover, they enable the manipulation of the spontaneous emission properties of luminescent devices<sup>5</sup>, the localization of light in microcavities<sup>6</sup>, and they may serve to generate negative refraction<sup>7, 8</sup>. A special class of these devices are the hollow-core photonic-crystal fibres<sup>9, 10, 11</sup>, which confine the light by means of a periodic cladding, consisting of several layers of identical cells. This design resonantly decreases the transmission losses of such fibres to values of a few dB km<sup>-1</sup> in a narrow wavelength range. However, the rather narrowband transmission bands and the detrimental third-order dispersion characteristics of this single-cell design generally render application of such hollow-core fibres difficult in the femtosecond range<sup>12</sup>. Therefore, no fibre-based concept can currently provide guiding of sub-100 fs pulses over extended distances. By introducing a radial chirp into the photonic crystal, we here demonstrate a novel concept for photonic-crystal fibres that breaks with the paradigm of lattice homogeneity and enables a new degree of freedom in photonic-crystal-fibre design, eliminating much of the pulse duration restriction of earlier approaches.

Organisations: Optoelectronics Research Centre

People: Skibina, J. (Creator), Iliw, R. (Creator), Bethge, J. (Creator), Bock, M. (Creator), Fischer, D. (Creator), Beloglasov, V. I. (Creator), Wedell, R. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://www.nature.com/nphoton/journal/v2/n11/supinfo/nphoton.2008.203\\_S1.html](https://www.nature.com/nphoton/journal/v2/n11/supinfo/nphoton.2008.203_S1.html)

Dataset

### Supplementary data for "Quantum size effects in ambient CO oxidation catalysed by ligand-protected gold clusters"

Finely dispersed nanometre-scale gold particles are known to catalyse several oxidation reactions in aerobic, ambient conditions. The catalytic activity has been explained by various complementary mechanisms, including support effects, particle-size-dependent metal-insulator transition, charging effects, frontier orbital interactions and geometric fluxionality. We show, by considering a series of robust and structurally well-characterized ligand-protected gold clusters with diameters between 1.2 and 2.4 nm, that electronic quantum size effects, particularly the magnitude of the so-called HOMO-LUMO energy gap, has a decisive role in binding oxygen to the nano-catalyst in an activated form. This can lead to the oxidation reaction  $2\text{CO} + \text{O}_2 \rightarrow 2\text{CO}_2$  with low activation barriers. Binding of dioxygen is significant only for the smallest particles with a metal core diameter clearly below 2 nm. Our results suggest a potentially viable route to practical applications using ligand-protected gold clusters for green chemistry.

Organisations: Department of Physics

People: Lopez-Acevedo, O. (Creator), Kacprzak, K. A. (Creator), Akola, J. (Creator), Häkkinen, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nchem/journal/v2/n4/full/nchem.589.html#supplementary-information>

Dataset

### Supplementary data for "High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity"

Prolonged culture of human embryonic stem cells (hESCs) can lead to adaptation and the acquisition of chromosomal abnormalities, underscoring the need for rigorous genetic analysis of these cells. Here we report the highest-resolution study of hESCs to date using an Affymetrix SNP 6.0 array containing 906,600 probes for single nucleotide polymorphisms (SNPs) and 946,000 probes for copy number variations (CNVs). Analysis of 17 different hESC lines maintained in different laboratories identified 843 CNVs of 50 kb–3 Mb in size. We identified, on average, 24% of the loss of heterozygosity (LOH) sites and 66% of the CNVs changed in culture between early and late passages of the same lines. Thirty percent of the genes detected within CNV sites had altered expression compared to samples with normal copy number states, of which >44% were functionally linked to cancer. Furthermore, LOH of the q arm of chromosome 16, which has not been observed previously in hESCs, was detected.

Organisations: Department of Signal Processing

People: Närvä, E. (Creator), , Rahkonen, N. (Creator), Rasool, O. (Creator), Lahesmaa, R. (Creator), , Yli-Harja, O. (Creator), Harrison, N. (Creator), Andrews, P. W. (Creator), Kitsberg, D. (Creator), Borghese, L. (Creator), Brüstle, O. (Creator), Itskovitz-Eldor, J. (Creator), Dvorak, P. (Creator), Hovatta, O. (Creator), Otonkoski, T. (Creator), Tuuri, T. (Creator), Cui, W. (Creator), Baker, D. (Creator), Maltby, E. (Creator), Moore, H. D. (Creator), Benvenisty, N. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nbt/journal/v28/n4/full/nbt.1615.html#supplementary-information>

Dataset

### Supplementary data for "From local structure to nanosecond recrystallization dynamics in AgInSbTe phase-change materials"

Phase-change optical memories are based on the astonishingly rapid nanosecond-scale crystallization of nanosized amorphous 'marks' in a polycrystalline layer. Models of crystallization exist for the commercially used phase-change alloy Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub> (GST), but not for the equally important class of Sb–Te-based alloys. We have combined X-ray diffraction, extended X-ray absorption fine structure and hard X-ray photoelectron spectroscopy experiments with density functional simulations to determine the crystalline and amorphous structures of Ag<sub>3.5</sub>In<sub>3.8</sub>Sb<sub>75.0</sub>Te<sub>17.7</sub> (AIST) and how they differ from GST. The structure of amorphous (a-) AIST shows a range of atomic ring sizes, whereas a-GST shows mainly small rings and cavities. The local environment of Sb in both forms of AIST is a distorted 3+3 octahedron. These structures suggest a bond-interchange model, where a sequence of small displacements of Sb atoms accompanied by interchanges of short and long bonds is the origin of the rapid crystallization of a-AIST. It differs profoundly from crystallization in a-GST.

Organisations: Department of Physics

People: Matsunaga, T. (Creator), Yamada, N. (Creator), Kojima, R. (Creator), Kohara, S. (Creator), Takata, M. (Creator), Akola, J. (Creator), Jones, R. O. (Creator), Honma, T. (Creator), Ikenaga, E. (Creator), Kobayashi, K. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nmat/journal/v10/n2/full/nmat2931.html#supplementary-information>

Dataset

### Supplementary data for "Cholesterol modulates glycolipid conformation and receptor activity"

We document a new dimension of surface recognition in which communication is controlled through the collective behavior of lipids. Membrane cholesterol induces a tilt in glycolipid receptor headgroup, resulting in loss of access for ligand binding. This property appears to organize erythrocyte blood group presentation and glycolipid receptor function during the activation of sperm fertility, suggesting that lipid 'allostery' is a means to regulate membrane recognition processes.

Organisations: Department of Physics

People: Lingwood, D. (Creator), Grzybek, M. (Creator), Coskun, Ü. (Creator), Simons, K. (Creator), Binnington, B. (Creator), Lingwood, C. A. (Creator), Rog, T. (Creator), Vattulainen, I. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nchembio/journal/v7/n5/full/nchembio.551.html#supplementary-information>

Dataset

### Supplementary data for "ESTOOLS Data@Hand: human stem cell gene expression resource"

We developed ESTOOLS Data@Hand, a resource to facilitate exploration of published gene expression array data in stem cell research. The resource, updated four times a year, offers efficient sample identification, preprocessing that enables cross-experiment comparisons and computational analysis.

Organisations: Department of Signal Processing, Department of Signal Processing

People: , , , Lund, R. (Creator), , Seppälä, J. (Creator), Gokhale, P. (Creator), , Hahne, L. (Creator), , , , Närvä, E. (Creator), Yli-Harja, O. (Creator), Andrews, P. W. (Creator), , Lahesmaa, R. (Creator), ,

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nmeth/journal/v10/n9/full/nmeth.2576.html#supplementary-information>

Dataset

### Supplementary data for "Acceleration of the rotation of asteroid 1862 Apollo by radiation torques"

The anisotropic reflection and thermal re-emission of sunlight from an asteroid's surface acts as a propulsion engine. The net propulsion force (Yarkovsky effect) changes the orbital dynamics of the body at a rate that depends on its physical properties; for irregularly shaped bodies, the propulsion causes a net torque (the Yarkovsky–O'Keefe–Radzievskii–Paddack or YORP effect) that can change the object's rotation period and the direction of its rotation axis. The Yarkovsky effect has been observed directly, and there is also indirect evidence of its role in the orbital evolution of asteroids over long time intervals. So far, however, only indirect evidence exists for the YORP effect through the clustering of the directions of rotation axes in asteroid families. Here we report a change in the rotation rate of the asteroid 1862 Apollo, which is best explained by the YORP mechanism. The change is fairly large and clearly visible in photometric lightcurves, amounting to one extra rotation cycle in just 40 years even though Apollo's size is well over one kilometre. This confirms the prediction that the YORP effect plays a significant part in the dynamical evolution of asteroids.

Organisations: Department of Mathematics

People: Kaasalainen, M. (Creator), Durech, J. (Creator), Warner, B. D. (Creator), Krugly, Y. N. (Creator), Gaftonyuk, N. M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nature/journal/v446/n7134/supinfo/nature05614.html>

Dataset

### Supplementary data for "New particle formation in forests inhibited by isoprene emissions"

It has been suggested that volatile organic compounds (VOCs) are involved in organic aerosol formation, which in turn affects radiative forcing and climate. The most abundant VOCs emitted by terrestrial vegetation are isoprene and its derivatives, such as monoterpenes and sesquiterpenes. New particle formation in boreal regions is related to monoterpene emissions and causes an estimated negative radiative forcing of about -0.2 to -0.9 W m<sup>-2</sup>. The annual variation in aerosol growth rates during particle nucleation events correlates with the seasonality of monoterpene emissions of the local vegetation, with a maximum during summer. The frequency of nucleation events peaks, however, in spring and autumn. Here we present evidence from simulation experiments conducted in a plant chamber that isoprene can significantly inhibit new particle formation. The process leading to the observed decrease in particle number concentration is linked to the high reactivity of isoprene with the hydroxyl radical (OH). The suppression is stronger with higher concentrations of isoprene, but with little dependence on the specific VOC mixture emitted by trees. A parameterization of the observed suppression factor as a function of isoprene concentration suggests that the number of new particles produced depends on the OH concentration and VOCs involved in the production of new particles undergo three to four steps of oxidation by OH. Our measurements simulate conditions that are typical for forested regions and may explain the observed seasonality in the frequency of aerosol nucleation events, with a lower number of nucleation events during summer compared to autumn and spring. Biogenic emissions of isoprene are controlled by temperature and light, and if the relative isoprene abundance of biogenic VOC emissions increases in response to climate change or land use change, the new particle formation potential may decrease, thus damping the aerosol negative radiative forcing effect.

Organisations: Department of Physics

People: Kiendler-Scharr, A. (Creator), Wildt, J. (Creator), Dal Maso, M. (Creator), Hohaus, T. (Creator), Kleist, E. (Creator), Mentel, T. F. (Creator), Tillmann, R. (Creator), Uerlings, R. (Creator), Schurr, U. (Creator), Wahner, A. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nature/journal/v461/n7262/supinfo/nature08292.html>

Dataset

### Supplementary data for "Atom manipulation on an insulating surface at room temperature"

Atomic manipulation enables us to fabricate a unique structure at the atomic scale. So far, many atomic manipulations have been reported on conductive surfaces, mainly at low temperature with scanning tunnelling microscopy, but atomic manipulation on an insulator at room temperature is still a long-standing challenge. Here we present a systematic atomic manipulation on an insulating surface by advanced atomic force microscopy, enabling construction of complex patterns such as a 'Swiss cross' of substitutional bromine ions in the sodium chloride surface.

Organisations: Department of Physics, Department of Physics

People: Kawai, S. (Creator), , , Onodera, H. (Creator), Kitamura, S. (Creator), Meyer, E. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/ncomms5403#supplementary-information>

Dataset

### Supplementary data for "Nitrous oxide as a function of oxygen and archaeal gene abundance in the North Pacific"

Oceanic oxygen minimum zones are strong sources of the potent greenhouse gas N<sub>2</sub>O but its microbial source is unclear. We characterized an exponential response in N<sub>2</sub>O production to decreasing oxygen between 1 and 30 μmol O<sub>2</sub> l<sup>-1</sup> within and below the oxycline using <sup>15</sup>NO<sub>2</sub><sup>-</sup>, a relationship that held along a 550 km offshore transect in the North Pacific. Differences in the overall magnitude of N<sub>2</sub>O production were accounted for by archaeal functional gene abundance. A one-dimensional (1D) model, parameterized with our experimentally derived exponential terms, accurately reproduces N<sub>2</sub>O profiles in the top 350 m of water column and, together with a strong 45N<sub>2</sub>O signature indicated neither canonical nor nitrifier-denitrification production while statistical modelling supported production by archaea, possibly via hybrid N<sub>2</sub>O formation. Further, with just archaeal N<sub>2</sub>O production, we could balance high-resolution estimates of sea-to-air N<sub>2</sub>O exchange. Hence, a significant source of N<sub>2</sub>O, previously described as leakage from bacterial ammonium oxidation, is better described by low-oxygen archaeal production at the oxygen minimum zone's margins.

Organisations: Chemistry and Bioengineering

People: Trimmer, M. (Creator), Chronopoulou, P. (Creator), Maanoja, S. (Creator), Upstill-Goddard, R. C. (Creator), Kitidis, V. (Creator), Purdy, K. J. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/ncomms13451#supplementary-information>

Dataset

### Supplementary data for "Structured light enables biomimetic swimming and versatile locomotion of photoresponsive soft microrobots"

Microorganisms move in challenging environments by periodic changes in body shape. In contrast, current artificial microrobots cannot actively deform, exhibiting at best passive bending under external fields. Here, by taking advantage of the wireless, scalable and spatiotemporally selective capabilities that light allows, we show that soft microrobots consisting

of photoactive liquid-crystal elastomers can be driven by structured monochromatic light to perform sophisticated biomimetic motions. We realize continuum yet selectively addressable artificial microswimmers that generate travelling-wave motions to self-propel without external forces or torques, as well as microrobots capable of versatile locomotion behaviours on demand. Both theoretical predictions and experimental results confirm that multiple gaits, mimicking either symplectic or antiplectic metachrony of ciliate protozoa, can be achieved with single microswimmers. The principle of using structured light can be extended to other applications that require microscale actuation with sophisticated spatiotemporal coordination for advanced microrobotic technologies.

Organisations: Chemistry and Bioengineering

People: Palagi, S. (Creator), Mark, A. G. (Creator), Melde, K. (Creator), Qiu, T. (Creator), Sanchez-Castillo, A. (Creator), Fischer, P. (Creator), Reigh, S. Y. (Creator), Lauga, E. (Creator), Zeng, H. (Creator), Parmeggiani, C. (Creator), Martella, D. (Creator), Wiersma, D. S. (Creator), Kapernaum, N. (Creator), Giesselmann, F. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nmat/journal/v15/n6/full/nmat4569.html#supplementary-information>

Dataset

#### **Supplementary data for "Sub-micron scale patterning of fluorescent silver nanoclusters using low-power laser"**

Supplementary information

Organisations: Photonics, Research area: Optics, Research group: Nonlinear Optics, Research group: Nonlinear Optics, Research group: Applied Optics

People: Bautista, G. (Creator), Toivonen, J. (Creator), Kunwar, P. (Creator)

Publisher: Tampere University of Technology

Date made available: 2016

Links: <https://www.nature.com/articles/srep23998#supplementary-information>

Dataset

#### **Supplementary data for "Polarized THG Microscopy Identifies Compositionally Different Lipid Droplets in Mammalian Cells"**

Supplementary information

Organisations: Photonics, Research area: Optics, Research group: Nonlinear Optics, Research group: Nonlinear Optics

People: Bautista, G. (Creator)

Publisher: Tampere University of Technology

Date made available: 2014

Links: <http://www.cell.com/cms/attachment/2092785532/2076620290/mmc1.pdf>

Dataset

#### **Supplementary data for "Direct Laser Writing of Photostable Fluorescent Silver Nanoclusters in Polymer Films"**

Supplementary files

Organisations: Photonics, Research area: Optics, Research group: Nonlinear Optics, Research group: Nonlinear Optics, Research group: Applied Optics

People: Bautista, G. (Creator), Kunwar, P. (Creator), Toivonen, J. (Creator)

Publisher: Tampere University of Technology

Date made available: 2014

Links: <http://pubs.acs.org/doi/suppl/10.1021/nn5059503>

Dataset

#### **Supplementary data for "Nonlinear microscopy using cylindrical vector beams: Applications to three-dimensional imaging of nanostructures"**

Supplementary files

Organisations: Photonics, Research area: Optics, Research group: Nonlinear Optics, Research group: Nonlinear Optics

People: Bautista, G. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.osapublishing.org/oe/abstract.cfm?uri=oe-25-11-12463#articleSupplMat>

Dataset

#### **Supplementary data for "Tailorable second-harmonic generation from an individual nanowire using spatially phase-shaped beams"**

Supplementary Information

Organisations: Photonics, Research area: Optics, Research group: Nonlinear Optics, Research group: Nonlinear Optics

People: Bautista, G. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <http://onlinelibrary.wiley.com/doi/10.1002/lpor.201600175/supinfo>

Dataset

### **Supplementary data for "Ghost imaging in the time domain"**

Ghost imaging is a novel technique that produces the image of an object by correlating the intensity of two light beams, neither of which independently carries information about the shape of the object. Ghost imaging has opened up new perspectives to obtain highly resolved images, even in the presence of noise and turbulence. Here, by exploiting the duality between light propagation in space and time, we demonstrate the temporal analogue of ghost imaging. We use a conventional fast detector that does not see the temporal 'object' to be characterized and a slow integrating 'bucket' detector that does see the object but without resolving its temporal structure. Our experiments achieve temporal resolution at the picosecond level and are insensitive to the temporal distortion that may occur after the object. The approach is scalable, can be integrated on-chip, and offers great promise for dynamic imaging of ultrafast waveforms.

Organisations: Photonics, Department of Physics

People: Ryczkowski, P. (Creator), Genty, G. (Creator), Friberg, A. T. (Creator), Dudley, J. M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nphoton/journal/v10/n3/full/nphoton.2015.274.html#supplementary-information>

Dataset

### **Supplementary data for "Guiding light via geometric phases"**

All known methods for transverse confinement and guidance of light rely on modification of the refractive index, that is, on the scalar properties of electromagnetic radiation. Here, we disclose the concept of a dielectric waveguide that exploits vectorial spin-orbit interactions of light and the resulting geometric phases. The approach relies on the use of anisotropic media with an optic axis that lies orthogonal to the propagation direction but is spatially modulated, so that the refractive index remains constant everywhere. A spin-controlled cumulative phase distortion is imposed on the beam, balancing diffraction for a specific polarization. As well as theoretical analysis, we present an experimental demonstration of the guidance using a series of discrete geometric-phase lenses made from liquid crystal. Our findings show that geometric phases may determine the optical guiding behaviour well beyond a Rayleigh length, paving the way to a new class of photonic devices. The concept is applicable to the whole electromagnetic spectrum.

Organisations: Department of Physics

People: Slussarenko, S. (Creator), Piccirillo, B. (Creator), Santamato, E. (Creator), Marrucci, L. (Creator), Assanto, G. (Creator), Jisha, C. P. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nphoton/journal/v10/n9/full/nphoton.2016.138.html#supplementary-information>

Dataset

### **Supplementary data for "Evaluation of methods for modeling transcription factor sequence specificity"**

Genomic analyses often involve scanning for potential transcription factor (TF) binding sites using models of the sequence specificity of DNA binding proteins. Many approaches have been developed to model and learn a protein's DNA-binding specificity, but these methods have not been systematically compared. Here we applied 26 such approaches to in vitro protein binding microarray data for 66 mouse TFs belonging to various families. For nine TFs, we also scored the resulting motif models on in vivo data, and found that the best in vitro-derived motifs performed similarly to motifs derived from the in vivo data. Our results indicate that simple models based on mononucleotide position weight matrices trained by the best methods perform similarly to more complex models for most TFs examined, but fall short in specific cases (<10% of the TFs examined here). In addition, the best-performing motifs typically have relatively low information content, consistent with widespread degeneracy in eukaryotic TF sequence preferences.

Organisations: Department of Signal Processing

People: Weirauch, M. T. (Creator), Cote, A. (Creator), Norel, R. (Creator), Zhao, Y. (Creator), Riley, T. R. (Creator), Saez-Rodriguez, J. (Creator), Cokelaer, T. (Creator), Vedenko, A. (Creator), Talukder, S. (Creator), Bussemaker, H. J. (Creator), Morris, Q. D. (Creator), Bulyk, M. L. (Creator), Stolovitzky, G. (Creator), Hughes, T. R. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nbt/journal/v31/n2/full/nbt.2486.html#supplementary-information>

Dataset

### **Supplementary data for "Gene-pair expression signatures reveal lineage control"**

The distinct cell types of multicellular organisms arise owing to constraints imposed by gene regulatory networks on the collective change of gene expression across the genome, creating self-stabilizing expression states, or attractors. We curated human expression data comprising 166 cell types and 2,602 transcription-regulating genes and developed a data-driven method for identifying putative determinants of cell fate built around the concept of expression reversal of gene pairs, such as those participating in toggle-switch circuits. This approach allows us to organize the cell types into their ontogenic lineage relationships. Our method identifies genes in regulatory circuits that control neuronal fate, pluripotency



and blood cell differentiation, and it may be useful for prioritizing candidate factors for direct conversion of cell fate.

Organisations: Department of Signal Processing

People: Heinänen, M. (Creator), Wienecke-Baldacchino, A. (Creator), Sinkkonen, L. (Creator), , Kramer, R. (Creator), Zhou, J. X. (Creator), Kreisberg, R. (Creator), Kauffman, S. (Creator), Huang, S. (Creator), Shmulevich, I. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nmeth/journal/v10/n6/full/nmeth.2445.html#supplementary-information>

Dataset

### **Supplementary data for "Measurement of the nucleation of atmospheric aerosol particles"**

The formation of new atmospheric aerosol particles and their subsequent growth have been observed frequently at various locations all over the world. The atmospheric nucleation rate (or formation rate) and growth rate (GR) are key parameters to characterize the phenomenon. Recent progress in measurement techniques enables us to measure atmospheric nucleation at the size (mobility diameter) of 1.5 ( $\pm$ 0.4) nm. The detection limit has decreased from 3 to 1 nm within the past 10 years. In this protocol, we describe the procedures for identifying new-particle-formation (NPF) events, and for determining the nucleation, formation and growth rates during such events under atmospheric conditions. We describe the present instrumentation, best practices and other tools used to investigate atmospheric nucleation and NPF at a certain mobility diameter (1.5, 2.0 or 3.0 nm). The key instruments comprise devices capable of measuring the number concentration of the formed nanoparticles and their size, such as a suite of modern condensation particle counters (CPCs) and air ion spectrometers, and devices for characterizing the pre-existing particle number concentration distribution, such as a differential mobility particle sizer (DMPS). We also discuss the reliability of the methods used and requirements for proper measurements and data analysis. The time scale for realizing this procedure is 1 year.

Organisations: Physics, Department of Physics

People: Kulmala, M. (Creator), Petäjä, T. (Creator), Nieminen, T. (Creator), Sipilä, M. (Creator), , Lehtipalo, K. (Creator), Dal Maso, M. (Creator), Aalto, P. P. (Creator), Junninen, H. (Creator), Paasonen, P. (Creator), Kerminen, V. (Creator), Riipinen, I. (Creator), Lehtinen, K. E. J. (Creator), Laaksonen, A. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nprot/journal/v7/n9/full/nprot.2012.091.html#supplementary-information>

Dataset

### **Supplementary data for "Construction of bispirooxindoles containing three quaternary stereocentres in a cascade using a single multifunctional organocatalyst"**

Single-step constructions of molecules with multiple quaternary carbon stereocentres are rare. The spirooxindole structural motif is common to a range of bioactive compounds; however, asymmetric synthesis of this motif is complicated due to the presence of multiple chiral centres. The development of organocatalytic cascade reactions has proven to be valuable for the construction of several chiral centres in one step. Here, we describe a newly designed organocatalytic asymmetric domino Michael–aldol reaction between 3-substituted oxindoles and methyleneindolinones that affords complex bispirooxindoles. This reaction was catalysed by a novel multifunctional organocatalyst that contains tertiary and primary amines and thiourea moieties to activate substrates simultaneously, providing extraordinary levels of stereocontrol over four stereocentres, three of which are quaternary carbon stereocentres. This new methodology provides facile access to a range of multisubstituted bispirocyclooxindole derivatives, and should be useful in medicinal chemistry and diversity-oriented syntheses of this intriguing class of compounds.

Organisations: Department of Chemistry and Bioengineering

People: Tan, B. (Creator), Rafael Candeias, N. (Creator), Barbas, C. F. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nchem/journal/v3/n6/full/nchem.1039.html#supplementary-information>

Dataset

### **Supplementary data for "Copy number variation and selection during reprogramming to pluripotency"**

The mechanisms underlying the low efficiency of reprogramming somatic cells into induced pluripotent stem (iPS) cells are poorly understood. There is a clear need to study whether the reprogramming process itself compromises genomic integrity and, through this, the efficiency of iPS cell establishment. Using a high-resolution single nucleotide polymorphism array, we compared copy number variations (CNVs) of different passages of human iPS cells with their fibroblast cell origins and with human embryonic stem (ES) cells. Here we show that significantly more CNVs are present in early-passage human iPS cells than intermediate passage human iPS cells, fibroblasts or human ES cells. Most CNVs are formed de novo and generate genetic mosaicism in early-passage human iPS cells. Most of these novel CNVs rendered the affected cells at a selective disadvantage. Remarkably, expansion of human iPS cells in culture selects rapidly against mutated cells, driving the lines towards a genetic state resembling human ES cells.

Organisations: Department of Signal Processing

People: Hussein, S. M. (Creator), Sourour, M. (Creator), Hämäläinen, R. (Creator), Nagy, A. (Creator), Vuoristo, S. (Creator), Olsson, C. (Creator), Lundin, K. (Creator), Mikkola, M. (Creator), Trokovic, R. (Creator), Otonkoski, T. (Creator), Batada, N. N. (Creator), Ng, S. (Creator), Ching, R. W. (Creator), Bazett-Jones, D. P. (Creator), , Närvä, E. (Creator), Lahesmaa, R. (Creator), Peitz, M. (Creator), Brüstle, O. (Creator), Alitalo, K. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nature/journal/v471/n7336/full/nature09871.html#supplementary-information>  
Dataset

#### **Supplementary data for "An amorphous solid state of biogenic secondary organic aerosol particles"**

Secondary organic aerosol (SOA) particles are formed in the atmosphere from condensable oxidation products of anthropogenic and biogenic volatile organic compounds (VOCs). On a global scale, biogenic VOCs account for about 90% of VOC emissions and of SOA formation (90 billion kilograms of carbon per year). SOA particles can scatter radiation and act as cloud condensation or ice nuclei, and thereby influence the Earth's radiation balance and climate. They consist of a myriad of different compounds with varying physicochemical properties, and little information is available on the phase state of SOA particles. Gas-particle partitioning models usually assume that SOA particles are liquid, but here we present experimental evidence that they can be solid under ambient conditions. We investigated biogenic SOA particles formed from oxidation products of VOCs in plant chamber experiments and in boreal forests within a few hours after atmospheric nucleation events. On the basis of observed particle bouncing in an aerosol impactor and of electron microscopy we conclude that biogenic SOA particles can adopt an amorphous solid—most probably glassy—state. This amorphous solid state should provoke a rethinking of SOA processes because it may influence the partitioning of semi-volatile compounds, reduce the rate of heterogeneous chemical reactions, affect the particles' ability to accommodate water and act as cloud condensation or ice nuclei, and change the atmospheric lifetime of the particles. Thus, the results of this study challenge traditional views of the kinetics and thermodynamics of SOA formation and transformation in the atmosphere and their implications for air quality and climate.

Organisations: Department of Physics, Department of Physics

People: , , Mäkelä, J. (Creator), Joutsensaari, J. (Creator), Worsnop, D. R. (Creator), Laaksonen, A. (Creator), Koop, T. (Creator), Yli-Pirilä, P. (Creator), Leskinen, J. (Creator), Holopainen, J. K. (Creator), Pöschl, U. (Creator), Kulmala, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nature/journal/v467/n7317/full/nature09455.html#supplementary-information>  
Dataset

#### **Supplementary data for "Prediction of disease-related mutations affecting protein localization"**

Background Eukaryotic cells contain numerous compartments, which have different protein constituents. Proteins are typically directed to compartments by short peptide sequences that act as targeting signals. Translocation to the proper compartment allows a protein to form the necessary interactions with its partners and take part in biological networks such as signalling and metabolic pathways. If a protein is not transported to the correct intracellular compartment either the reaction performed or information carried by the protein does not reach the proper site, causing either inactivation of central reactions or misregulation of signalling cascades, or the mislocalized active protein has harmful effects by acting in the wrong place. Results Numerous methods have been developed to predict protein subcellular localization with quite high accuracy. We applied bioinformatics methods to investigate the effects of known disease-related mutations on protein targeting and localization by analyzing over 22,000 missense mutations in more than 1,500 proteins with two complementary prediction approaches. Several hundred putative localization affecting mutations were identified and investigated statistically. Conclusion Although alterations to localization signals are rare, these effects should be taken into account when analyzing the consequences of disease-related mutations.

Organisations: Department of Signal Processing

People: , Vihinen, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-10-122#Declarations>  
Dataset

#### **Supplementary data for "Information propagation within the Genetic Network of *Saccharomyces cerevisiae*"**

Background A gene network's capacity to process information, so as to bind past events to future actions, depends on its structure and logic. From previous and new microarray measurements in *Saccharomyces cerevisiae* following gene deletions and overexpressions, we identify a core gene regulatory network (GRN) of functional interactions between 328 genes and the transfer functions of each gene. Inferred connections are verified by gene enrichment. Results We find that this core network has a generalized clustering coefficient that is much higher than chance. The inferred Boolean transfer functions have a mean p-bias of 0.41, and thus similar amounts of activation and repression interactions. However, the distribution of p-biases differs significantly from what is expected by chance that, along with the high mean connectivity, is found to cause the core GRN of *S. cerevisiae*'s to have an overall sensitivity similar to critical Boolean networks. In agreement, we find that the amount of information propagated between nodes in finite time series is much higher in the inferred core GRN of *S. cerevisiae* than what is expected by chance. Conclusions We suggest that *S. cerevisiae* is likely to have evolved a core GRN with enhanced information propagation among its genes.

Organisations: Research group: Laboratory of Biosystem Dynamics-LBD, Faculty of Biomedical Sciences and Engineering , Research group: Laboratory of Biosystem Dynamics-LBD

People: , , Baici, W. C. V. (Creator), Hughes, T. R. (Creator), Yli-Harja, O. (Creator), Chua, G. (Creator), Sanches Ribeiro, A. (Creator)  
Publisher: Tampere University of Technology  
Date made available: 2017  
Links: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-4-143#Declarations>  
Dataset

### **Supplementary data for "Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis"**

**Background** Approximately half of all human genes use alternative transcription start sites (TSSs) to control mRNA levels and broaden the transcriptional output in healthy tissues. Aberrant expression patterns promoting carcinogenesis, however, may arise from alternative promoter usage. **Results** By profiling 108 colorectal samples using exon arrays, we identified nine genes (TCF12, OSBPL1A, TRAK1, ANK3, CHEK1, UGP2, LMO7, ACSL5, and SCIN) showing tumor-specific alternative TSS usage in both adenoma and cancer samples relative to normal mucosa. Analysis of independent exon array data sets corroborated these findings. Additionally, we confirmed the observed patterns for selected mRNAs using quantitative real-time reverse-transcription PCR. Interestingly, for some of the genes, the tumor-specific TSS usage was not restricted to colorectal cancer. A comprehensive survey of the nine genes in lung, bladder, liver, prostate, gastric, and brain cancer revealed significantly altered mRNA isoform ratios for CHEK1, OSBPL1A, and TCF12 in a subset of these cancer types. To identify the mechanism responsible for the shift in alternative TSS usage, we antagonized the Wnt-signaling pathway in DLD1 and Ls174T colorectal cancer cell lines, which remarkably led to a shift in the preferred TSS for both OSBPL1A and TRAK1. This indicated a regulatory role of the Wnt pathway in selecting TSS, possibly also involving TP53 and SOX9, as their transcription binding sites were enriched in the promoters of the tumor preferred isoforms together with their mRNA levels being increased in tumor samples. Finally, to evaluate the prognostic impact of the altered TSS usage, immunohistochemistry was used to show deregulation of the total protein levels of both TCF12 and OSBPL1A, corresponding to the mRNA levels observed. Furthermore, the level of nuclear TCF12 had a significant correlation to progression free survival in a cohort of 248 stage II colorectal cancer samples. **Conclusions** Alternative TSS usage in colorectal adenoma and cancer samples has been shown for nine genes, and OSBPL1A and TRAK1 were found to be regulated in vitro by Wnt signaling. TCF12 protein expression was upregulated in cancer samples and correlated with progression free survival.

Organisations: Department of Signal Processing

People: Thorsen, K. (Creator), Schepeler, T. (Creator), Oster, B. (Creator), Rasmussen, M. H. (Creator), Vang, S. (Creator), Wang, K. (Creator), Hansen, K. Q. (Creator), Lamy, P. (Creator), Pedersen, J. S. (Creator), Eller, A. (Creator), Mansilla, F. (Creator), , Wiuf, C. (Creator), Laurberg, S. (Creator), Dyrskjot, L. (Creator), Orntoft, T. F. (Creator), Andersen, C. L. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://bmccgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-505#Declarations>

Dataset

### **Supplementary data for "In vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps"**

**Background** In *Escherichia coli* the mean and cell-to-cell diversity in RNA numbers of different genes vary widely. This is likely due to different kinetics of transcription initiation, a complex process with multiple rate-limiting steps that affect RNA production. **Results** We measured the in vivo kinetics of production of individual RNA molecules under the control of the lar promoter in *E. coli*. From the analysis of the distributions of intervals between transcription events in the regimes of weak and medium induction, we find that the process of transcription initiation of this promoter involves a sequential mechanism with two main rate-limiting steps, each lasting hundreds of seconds. Both steps become faster with increasing induction by IPTG and Arabinose. **Conclusions** The two rate-limiting steps in initiation are found to be important regulators of the dynamics of RNA production under the control of the lar promoter in the regimes of weak and medium induction. Variability in the intervals between consecutive RNA productions is much lower than if there was only one rate-limiting step with a duration following an exponential distribution. The methodology proposed here to analyze the in vivo dynamics of transcription may be applicable at a genome-wide scale and provide valuable insight into the dynamics of prokaryotic genetic networks.

Organisations: Department of Signal Processing

People: Kandhavelu, M. (Creator), , , , Yli-Harja, O. (Creator), Sanches Ribeiro, A. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-5-149#Declarations>

Dataset

### **Supplementary data for "A Beta-mixture model for dimensionality reduction, sample classification and analysis"**

**Background** Patterns of genome-wide methylation vary between tissue types. For example, cancer tissue shows markedly different patterns from those of normal tissue. In this paper we propose a beta-mixture model to describe genome-wide methylation patterns based on probe data from methylation microarrays. The model takes dependencies between neighbour probe pairs into account and assumes three broad categories of methylation, low, medium and high. The model

is described by 37 parameters, which reduces the dimensionality of a typical methylation microarray significantly. We used methylation microarray data from 42 colon cancer samples to assess the model. Results Based on data from colon cancer samples we show that our model captures genome-wide characteristics of methylation patterns. We estimate the parameters of the model and show that they vary between different tissue types. Further, for each methylation probe the posterior probability of a methylation state (low, medium or high) is calculated and the probability that the state is correctly predicted is assessed. We demonstrate that the model can be applied to classify cancer tissue types accurately and that the model provides accessible and easily interpretable data summaries. Conclusions We have developed a beta-mixture model for methylation microarray data. The model substantially reduces the dimensionality of the data. It can be used for further analysis, such as sample classification or to detect changes in methylation status between different samples and tissues.

Organisations: Department of Signal Processing

People: Laurila, K. (Creator), Oster, B. (Creator), Andersen, C. L. (Creator), Lamy, P. (Creator), Orntoft, T. (Creator), Yli-Harja, O. (Creator), Wiuf, C. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-215#Declarations>

Dataset

### **Supplementary data for "An integrative computational systems biology approach identifies differentially regulated dynamic transcriptome signatures which drive the initiation of human T helper cell differentiation"**

Background A proper balance between different T helper (Th) cell subsets is necessary for normal functioning of the adaptive immune system. Revealing key genes and pathways driving the differentiation to distinct Th cell lineages provides important insight into underlying molecular mechanisms and new opportunities for modulating the immune response. Previous computational methods to quantify and visualize kinetic differential expression data of three or more lineages to identify reciprocally regulated genes have relied on clustering approaches and regression methods which have time as a factor, but have lacked methods which explicitly model temporal behavior. Results We studied transcriptional dynamics of human umbilical cord blood T helper cells cultured in absence and presence of cytokines promoting Th1 or Th2 differentiation. To identify genes that exhibit distinct lineage commitment dynamics and are specific for initiating differentiation to different Th cell subsets, we developed a novel computational methodology (LIGAP) allowing integrative analysis and visualization of multiple lineages over whole time-course profiles. Applying LIGAP to time-course data from multiple Th cell lineages, we identified and experimentally validated several differentially regulated Th cell subset specific genes as well as reciprocally regulated genes. Combining differentially regulated transcriptional profiles with transcription factor binding site and pathway information, we identified previously known and new putative transcriptional mechanisms involved in Th cell subset differentiation. All differentially regulated genes among the lineages together with an implementation of LIGAP are provided as an open-source resource. Conclusions The LIGAP method is widely applicable to quantify differential time-course dynamics of many types of datasets and generalizes to any number of conditions. It summarizes all the time-course measurements together with the associated uncertainty for visualization and manual assessment purposes. Here we identified novel human Th subset specific transcripts as well as regulatory mechanisms important for the initiation of the Th cell subset differentiation.

Organisations: Department of Signal Processing, Department of Signal Processing

People: , Edelman, S. M. (Creator), Lönnberg, T. (Creator), , Kallionpää, H. (Creator), Tuomela, S. (Creator), Engström, E. (Creator), Lahesmaa, R. (Creator), Lähdesmäki, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://bmccgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-572#Declarations>

Dataset

### **Supplementary data for "Interfacing cellular networks of *S. cerevisiae* and *E. coli*: Connecting dynamic and genetic information"**

Background In recent years, various types of cellular networks have penetrated biology and are nowadays used omnipresently for studying eukaryote and prokaryote organisms. Still, the relation and the biological overlap among phenomenological and inferential gene networks, e.g., between the protein interaction network and the gene regulatory network inferred from large-scale transcriptomic data, is largely unexplored. Results We provide in this study an in-depth analysis of the structural, functional and chromosomal relationship between a protein-protein network, a transcriptional regulatory network and an inferred gene regulatory network, for *S. cerevisiae* and *E. coli*. Further, we study global and local aspects of these networks and their biological information overlap by comparing, e.g., the functional co-occurrence of Gene Ontology terms by exploiting the available interaction structure among the genes. Conclusions Although the individual networks represent different levels of cellular interactions with global structural and functional dissimilarities, we observe crucial functions of their network interfaces for the assembly of protein complexes, proteolysis, transcription, translation, metabolic and regulatory interactions. Overall, our results shed light on the integrability of these networks and their interfacing biological processes.

Organisations: Research group: Computational Medicine and Statistical Learning Laboratory (CMSL)

People: Matos Simoes, R. D. (Creator), Dehmer, M. (Creator), Emmert-Streib, F. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-324#Declarations>  
Dataset

#### **Supplementary data for "Bioprocess data mining using regularized regression and random forests"**

**Background** In bioprocess development, the needs of data analysis include (1) getting overview to existing data sets, (2) identifying primary control parameters, (3) determining a useful control direction, and (4) planning future experiments. In particular, the integration of multiple data sets causes that these needs cannot be properly addressed by regression models that assume linear input-output relationship or unimodality of the response function. Regularized regression and random forests, on the other hand, have several properties that may appear important in this context. They are capable, e.g., in handling small number of samples with respect to the number of variables, feature selection, and the visualization of response surfaces in order to present the prediction results in an illustrative way. **Results** In this work, the applicability of regularized regression (Lasso) and random forests (RF) in bioprocess data mining was examined, and their performance was benchmarked against multiple linear regression. As an example, we used data from a culture media optimization study for microbial hydrogen production. All the three methods were capable in providing a significant model when the five variables of the culture media optimization were linearly included in modeling. However, multiple linear regression failed when also the multiplications and squares of the variables were included in modeling. In this case, the modeling was still successful with Lasso (correlation between the observed and predicted yield was 0.69) and RF (0.91). **Conclusion** We found that both regularized regression and random forests were able to produce feasible models, and the latter was efficient in capturing the non-linearity in the data. In this kind of a data mining task of bioprocess data, both methods outperform multiple linear regression.

**Organisations:** Department of Signal Processing, Department of Chemistry and Bioengineering, Signal Processing, Chemistry and Bioengineering

**People:** , , Mangayil, R. (Creator), Huttunen, H. (Creator),

**Publisher:** Tampere University of Technology

**Date made available:** 2017

Links: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-7-S1-S5#Declarations>

Dataset

#### **Supplementary data for "Functional and genetic analysis of the colon cancer network"**

Cancer is a complex disease that has proven to be difficult to understand on the single-gene level. For this reason a functional elucidation needs to take interactions among genes on a systems-level into account. In this study, we infer a colon cancer network from a large-scale gene expression data set by using the method BC3Net. We provide a structural and a functional analysis of this network and also connect its molecular interaction structure with the chromosomal locations of the genes enabling the definition of cis- and trans-interactions. Furthermore, we investigate the interaction of genes that can be found in close neighborhoods on the chromosomes to gain insight into regulatory mechanisms. To our knowledge this is the first study analyzing the genome-scale colon cancer network.

**Organisations:** Research group: Computational Medicine and Statistical Learning Laboratory (CMSL)

**People:** Emmert-Streib, F. (Creator), Matos Simoes, R. D. (Creator), Glazko, G. V. (Creator), McDade, S. (Creator), Haibe-Kains, B. (Creator), Holzinger, A. (Creator), Dehmer, M. (Creator), Campbell, F. (Creator)

**Publisher:** Tampere University of Technology

**Date made available:** 2017

Links: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-S6-S6#Declarations>

Dataset

#### **Supplementary data for "Comparative evaluation of gene set analysis approaches for RNA-Seq data"**

**Background** Over the last few years transcriptome sequencing (RNA-Seq) has almost completely taken over microarrays for high-throughput studies of gene expression. Currently, the most popular use of RNA-Seq is to identify genes which are differentially expressed between two or more conditions. Despite the importance of Gene Set Analysis (GSA) in the interpretation of the results from RNA-Seq experiments, the limitations of GSA methods developed for microarrays in the context of RNA-Seq data are not well understood. **Results** We provide a thorough evaluation of popular multivariate and gene-level self-contained GSA approaches on simulated and real RNA-Seq data. The multivariate approach employs multivariate non-parametric tests combined with popular normalizations for RNA-Seq data. The gene-level approach utilizes univariate tests designed for the analysis of RNA-Seq data to find gene-specific P-values and combines them into a pathway P-value using classical statistical techniques. Our results demonstrate that the Type I error rate and the power of multivariate tests depend only on the test statistics and are insensitive to the different normalizations. In general standard multivariate GSA tests detect pathways that do not have any bias in terms of pathways size, percentage of differentially expressed genes, or average gene length in a pathway. In contrast the Type I error rate and the power of gene-level GSA tests are heavily affected by the methods for combining P-values, and all aforementioned biases are present in detected pathways. **Conclusions** Our result emphasizes the importance of using self-contained non-parametric multivariate tests for detecting differentially expressed pathways for RNA-Seq data and warns against applying gene-level GSA tests, especially because of their high level of Type I error rates for both, simulated and real data.

**Organisations:** Faculty of Biomedical Sciences and Engineering

**People:** Rahmatallah, Y. (Creator), Emmert-Streib, F. (Creator), Glazko, G. (Creator)

**Publisher:** Tampere University of Technology

Date made available: 2017

Links: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-014-0397-8#Declarations>  
Dataset

### **Supplementary data for "Comprehensive molecular portraits of human breast tumours"**

We analysed primary breast cancers by genomic DNA copy number arrays, DNA methylation, exome sequencing, messenger RNA arrays, microRNA sequencing and reverse-phase protein arrays. Our ability to integrate information across platforms provided key insights into previously defined gene expression subtypes and demonstrated the existence of four main breast cancer classes when combining data from five platforms, each of which shows significant molecular heterogeneity. Somatic mutations in only three genes (TP53, PIK3CA and GATA3) occurred at >10% incidence across all breast cancers; however, there were numerous subtype-associated and novel gene mutations including the enrichment of specific mutations in GATA3, PIK3CA and MAP3K1 with the luminal A subtype. We identified two novel protein-expression-defined subgroups, possibly produced by stromal/microenvironmental elements, and integrated analyses identified specific signalling pathways dominant in each molecular subtype including a HER2/phosphorylated HER2/EGFR/phosphorylated EGFR signature within the HER2-enriched expression subtype. Comparison of basal-like breast tumours with high-grade serous ovarian tumours showed many molecular commonalities, indicating a related aetiology and similar therapeutic opportunities. The biological finding of the four main breast cancer subtypes caused by different subsets of genetic and epigenetic abnormalities raises the hypothesis that much of the clinically observable plasticity and heterogeneity occurs within, and not across, these major biological subtypes of breast cancer.

Organisations: Department of Signal Processing

People: Koboldt, D. C. (Creator), Fulton, R. S. (Creator), McLellan, M. D. (Creator), Schmidt, H. (Creator), Kalicki-Veizer, J. (Creator), McMichael, J. F. (Creator), Fulton, L. L. (Creator), Dooling, D. J. (Creator), Ding, L. (Creator), Mardis, E. R. (Creator), Wilson, R. K. (Creator), Ally, A. (Creator), Balasundaram, M. (Creator), Butterfield, Y. S. N. (Creator), Carlsen, R. (Creator), Carter, C. (Creator), Chu, A. (Creator), Chuah, E. (Creator), Chun, H. E. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nature/journal/v490/n7418/full/nature11412.html#supplementary-information>  
Dataset

### **Supplementary data for "Integrated genomic characterization of endometrial carcinoma"**

We performed an integrated genomic, transcriptomic and proteomic characterization of 373 endometrial carcinomas using array- and sequencing-based technologies. Uterine serous tumours and ~25% of high-grade endometrioid tumours had extensive copy number alterations, few DNA methylation changes, low oestrogen receptor/progesterone receptor levels, and frequent TP53 mutations. Most endometrioid tumours had few copy number alterations or TP53 mutations, but frequent mutations in PTEN, CTNNB1, PIK3CA, ARID1A and KRAS and novel mutations in the SWI/SNF chromatin remodelling complex gene ARID5B. A subset of endometrioid tumours that we identified had a markedly increased transversion mutation frequency and newly identified hotspot mutations in POLE. Our results classified endometrial cancers into four categories: POLE ultramutated, microsatellite instability hypermutated, copy-number low, and copy-number high. Uterine serous carcinomas share genomic features with ovarian serous and basal-like breast carcinomas. We demonstrated that the genomic features of endometrial carcinomas permit a reclassification that may affect post-surgical adjuvant treatment for women with aggressive tumours.

Organisations: Department of Signal Processing

People: Getz, G. (Creator), Gabriel, S. B. (Creator), Cibulskis, K. (Creator), Lander, E. (Creator), Sivachenko, A. (Creator), Sougnez, C. (Creator), Lawrence, M. (Creator), Kandoth, C. (Creator), Dooling, D. (Creator), Fulton, R. (Creator), Fulton, L. (Creator), Kalicki-Veizer, J. (Creator), McLellan, M. D. (Creator), O'Laughlin, M. (Creator), Schmidt, H. (Creator), Wilson, R. K. (Creator), Ye, K. (Creator), Li, D. (Creator), Ally, A. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nature/journal/v497/n7447/full/nature12113.html#supplementary-information>  
Dataset

### **Supplementary data for "Critical assessment of automated flow cytometry data analysis techniques"**

Traditional methods for flow cytometry (FCM) data processing rely on subjective manual gating. Recently, several groups have developed computational methods for identifying cell populations in multidimensional FCM data. The Flow Cytometry: Critical Assessment of Population Identification Methods (FlowCAP) challenges were established to compare the performance of these methods on two tasks: (i) mammalian cell population identification, to determine whether automated algorithms can reproduce expert manual gating and (ii) sample classification, to determine whether analysis pipelines can identify characteristics that correlate with external variables (such as clinical outcome). This analysis presents the results of the first FlowCAP challenges. Several methods performed well as compared to manual gating or external variables using statistical performance measures, which suggests that automated methods have reached a sufficient level of maturity and accuracy for reliable use in FCM data analysis.

Organisations: Signal Processing, Pori

People: Aghaeepour, N. (Creator), Brinkman, R. (Creator), Finak, G. (Creator), Gottardo, R. (Creator), Hoos, H. (Creator), Mosmann, T. R. (Creator), Scheuermann, R. H. (Creator), Huttunen, H. (Creator), Kollman, T. R. (Creator), Ruusuvoori, P. (Creator), Seifert, M. (Creator), Guex, N. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/nmeth/journal/v10/n3/full/nmeth.2365.html#supplementary-information>

Dataset

#### **Supplementary data for "Atomic and electronic structures of an extremely fragile liquid"**

The structure of high-temperature liquids is an important topic for understanding the fragility of liquids. Here we report the structure of a high-temperature non-glass-forming oxide liquid, ZrO<sub>2</sub>, at an atomistic and electronic level. The Bhatia–Thornton number–number structure factor of ZrO<sub>2</sub> does not show a first sharp diffraction peak. The atomic structure comprises ZrO<sub>5</sub>, ZrO<sub>6</sub> and ZrO<sub>7</sub> polyhedra with a significant contribution of edge sharing of oxygen in addition to corner sharing. The variety of large oxygen coordination and polyhedral connections with short Zr–O bond lifetimes, induced by the relatively large ionic radius of zirconium, disturbs the evolution of intermediate-range ordering, which leads to a reduced electronic band gap and increased delocalization in the ionic Zr–O bonding. The details of the chemical bonding explain the extremely low viscosity of the liquid and the absence of a first sharp diffraction peak, and indicate that liquid ZrO<sub>2</sub> is an extremely fragile liquid.

Organisations: Physics

People: Kohara, S. (Creator), Akola, J. (Creator), Patrikeev, L. (Creator), , Ohara, K. (Creator), Itou, M. (Creator), Fujiwara, A. (Creator), Yahiro, J. (Creator), Okada, J. T. (Creator), Ishikawa, T. (Creator), Mizuno, A. (Creator), Masuno, A. (Creator), Watanabe, Y. (Creator), Usuki, T. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/ncomms6892#supplementary-information>

Dataset

#### **Supplementary data for "Molecular interactions on single-walled carbon nanotubes revealed by high-resolution transmission microscopy"**

The close solid-state structure–property relationships of organic  $\pi$ -aromatic molecules have attracted interest due to their implications for the design of organic functional materials. In particular, a dimeric structure, that is, a unit consisting of two molecules, is required for precisely evaluating intermolecular interactions. Here, we show that the sidewall of a single-walled carbon nanotube (SWNT) represents a unique molecular dimer platform that can be directly visualized using high-resolution transmission electron microscopy. Pyrene is chosen as the  $\pi$ -aromatic molecule; its dimer is covalently linked to the SWNT sidewalls by aryl addition. Reflecting the orientation and separation of the two molecules, the pyrene dimer on the SWNT exhibits characteristic optical and photophysical properties. The methodology discussed here—form and probe molecular dimers—is highly promising for the creation of unique models and provides indispensable and fundamental information regarding molecular interactions.

Organisations: Department of Chemistry and Bioengineering, Chemistry and Bioengineering, Research group: Chemistry & Advanced Materials

People: Umeyama, T. (Creator), Baek, J. (Creator), Sato, Y. (Creator), Suenaga, K. (Creator), , Tkachenko, N. (Creator), , Imahori, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/ncomms8732#supplementary-information>

Dataset

#### **Supplementary data for "Controlling the motion of multiple objects on a Chladni plate"**

The origin of the idea of moving objects by acoustic vibration can be traced back to 1787, when Ernst Chladni reported the first detailed studies on the aggregation of sand onto nodal lines of a vibrating plate. Since then and to this date, the prevailing view has been that the particle motion out of nodal lines is random, implying uncontrollability. But how random really is the out-of-nodal-lines motion on a Chladni plate? Here we show that the motion is sufficiently regular to be statistically modelled, predicted and controlled. By playing carefully selected musical notes, we can control the position of multiple objects simultaneously and independently using a single acoustic actuator. Our method allows independent trajectory following, pattern transformation and sorting of multiple miniature objects in a wide range of materials, including electronic components, water droplets loaded on solid carriers, plant seeds, candy balls and metal parts.

Organisations: Department of Automation Science and Engineering

People: , Sariola, V. (Creator), , Liimatainen, V. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/ncomms12764#supplementary-information>

Dataset

#### **Supplementary data for "The complex nature of calcium cation interactions with phospholipid bilayers"**

Understanding interactions of calcium with lipid membranes at the molecular level is of great importance in light of their involvement in calcium signaling, association of proteins with cellular membranes, and membrane fusion. We quantify these interactions in detail by employing a combination of spectroscopic methods with atomistic molecular dynamics

simulations. Namely, time-resolved fluorescent spectroscopy of lipid vesicles and vibrational sum frequency spectroscopy of lipid monolayers are used to characterize local binding sites of calcium in zwitterionic and anionic model lipid assemblies, while dynamic light scattering and zeta potential measurements are employed for macroscopic characterization of lipid vesicles in calcium-containing environments. To gain additional atomic-level information, the experiments are complemented by molecular simulations that utilize an accurate force field for calcium ions with scaled charges effectively accounting for electronic polarization effects. We demonstrate that lipid membranes have substantial calcium-binding capacity, with several types of binding sites present. Significantly, the binding mode depends on calcium concentration with important implications for calcium buffering, synaptic plasticity, and protein-membrane association.

Organisations: Physics

People: Melcrová, A. (Creator), Pokorna, S. (Creator), Pullanchery, S. (Creator), Kohagen, M. (Creator), Jurkiewicz, P. (Creator), Hof, M. (Creator), , Cremer, P. S. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/srep38035#supplementary-information>

Dataset

### **Supplementary data for "Membrane cholesterol access into a G-protein-coupled receptor"**

Cholesterol is a key component of cell membranes with a proven modulatory role on the function and ligand-binding properties of G-protein-coupled receptors (GPCRs). Crystal structures of prototypical GPCRs such as the adenosine A2A receptor (A2AR) have confirmed that cholesterol finds stable binding sites at the receptor surface suggesting an allosteric role of this lipid. Here we combine experimental and computational approaches to show that cholesterol can spontaneously enter the A2AR-binding pocket from the membrane milieu using the same portal gate previously suggested for opsin ligands. We confirm the presence of cholesterol inside the receptor by chemical modification of the A2AR interior in a biotinylation assay. Overall, we show that cholesterol's impact on A2AR-binding affinity goes beyond pure allosteric modulation and unveils a new interaction mode between cholesterol and the A2AR that could potentially apply to other GPCRs.

Organisations: Department of Physics

People: Guixa-Gonzalez, R. (Creator), Albasanz, J. L. (Creator), Rodriguez-Espigares, I. (Creator), Pastor, M. (Creator), Sanz, F. (Creator), Martí-Solano, M. (Creator), , Martinez-Seara, H. (Creator), Hildebrand, P. W. (Creator), Martín, M. (Creator), Selent, J. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://www.nature.com/articles/ncomms14505#supplementary-information>

Dataset

### **Supplementary data for "Changes in global gene expression of *Vibrio parahaemolyticus* induced by cold- and heat-stress"**

Background *Vibrio* (*V.*) *parahaemolyticus* causes seafood-borne gastro-intestinal bacterial infections in humans worldwide. It is widely found in marine environments and is isolated frequently from seawater, estuarine waters, sediments and raw or insufficiently cooked seafood. Throughout the food chain, *V. parahaemolyticus* encounters different temperature conditions that might alter metabolism and pathogenicity of the bacterium. In this study, we performed gene expression profiling of *V. parahaemolyticus* RIMD 2210633 after exposure to 4, 15, 20, 37 and 42 °C to describe the cold and heat shock response. Methods Gene expression profiles of *V. parahaemolyticus* RIMD 2210633 after exposure to 4, 15, 20, 37 and 42 °C were investigated via microarray. Gene expression values and RT-qPCR experiments were compared by plotting the log<sub>2</sub> values. Moreover, volcano plots of microarray data were calculated to visualize the distribution of differentially expressed genes at individual temperatures and to assess hybridization qualities and comparability of data. Finally, enriched terms were searched in annotations as well as functional-related gene categories using the Database for Annotation, Visualization and Integrated Discovery. Results Analysis of 37 °C normalised transcriptomics data resulted in differential expression of 19 genes at 20 °C, 193 genes at 4 °C, 625 genes at 42 °C and 638 genes at 15 °C. Thus, the largest number of significantly expressed genes was observed at 15 and 42 °C with 13.3 and 13 %, respectively. Genes of many functional categories were highly regulated even at lower temperatures. Virulence associated genes (*tdh1*, *tdh2*, *toxR*, *toxS*, *vopC*, *T6SS-1*, *T6SS-2*) remained mostly unaffected by heat or cold stress. Conclusion Along with folding and temperature shock depending systems, an overall temperature-dependent regulation of expression could be shown. Particularly the energy metabolism was affected by changed temperatures. Whole-genome gene expression studies of food related pathogens such as *V. parahaemolyticus* reveal how these pathogens react to stress impacts to predict its behaviour under conditions like storage and transport. Additional files include DAVID data for all temperatures and the homolog and antagonistic reacting genes.

Organisations: Department of Chemistry and Bioengineering, Department of Signal Processing

People: Urmersbach, S. (Creator), , Alter, T. (Creator), , Huehn, S. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.c.3607616

Dataset



### **Supplementary data for "CytoSpectre: a tool for spectral analysis of oriented structures on cellular and subcellular levels"**

Background Orientation and the degree of isotropy are important in many biological systems such as the sarcomeres of cardiomyocytes and other fibrillar structures of the cytoskeleton. Image based analysis of such structures is often limited to qualitative evaluation by human experts, hampering the throughput, repeatability and reliability of the analyses. Software tools are not readily available for this purpose and the existing methods typically rely at least partly on manual operation. Results We developed CytoSpectre, an automated tool based on spectral analysis, allowing the quantification of orientation and also size distributions of structures in microscopy images. CytoSpectre utilizes the Fourier transform to estimate the power spectrum of an image and based on the spectrum, computes parameter values describing, among others, the mean orientation, isotropy and size of target structures. The analysis can be further tuned to focus on targets of particular size at cellular or subcellular scales. The software can be operated via a graphical user interface without any programming expertise. We analyzed the performance of CytoSpectre by extensive simulations using artificial images, by benchmarking against FibrilTool and by comparisons with manual measurements performed for real images by a panel of human experts. The software was found to be tolerant against noise and blurring and superior to FibrilTool when analyzing realistic targets with degraded image quality. The analysis of real images indicated general good agreement between computational and manual results while also revealing notable expert-to-expert variation. Moreover, the experiment showed that CytoSpectre can handle images obtained of different cell types using different microscopy techniques. Finally, we studied the effect of mechanical stretching on cardiomyocytes to demonstrate the software in an actual experiment and observed changes in cellular orientation in response to stretching. Conclusions CytoSpectre, a versatile, easy-to-use software tool for spectral analysis of microscopy images was developed. The tool is compatible with most 2D images and can be used to analyze targets at different scales. We expect the tool to be useful in diverse applications dealing with structures whose orientation and size distributions are of interest. While designed for the biological field, the software could also be useful in non-biological applications.

Organisations: Signal Processing

People: , Pölonen, R. P. (Creator), Ojala, M. (Creator), Rasku, J. (Creator), Lekkala, J. (Creator), Aalto-Setälä, K. (Creator), Kallio, P. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.c.3598943

Dataset

### **Supplementary data for "Health figures: an open source JavaScript library for health data visualization"**

Background The way we look at data has a great impact on how we can understand it, particularly when the data is related to health and wellness. Due to the increased use of self-tracking devices and the ongoing shift towards preventive medicine, better understanding of our health data is an important part of improving the general welfare of the citizens. Electronic Health Records, self-tracking devices and mobile applications provide a rich variety of data but it often becomes difficult to understand. We implemented the hFigures library inspired on the hGraph visualization with additional improvements. The purpose of the library is to provide a visual representation of the evolution of health measurements in a complete and useful manner. Results We researched the usefulness and usability of the library by building an application for health data visualization in a health coaching program. We performed a user evaluation with Heuristic Evaluation, Controlled User Testing and Usability Questionnaires. In the Heuristics Evaluation the average response was 6.3 out of 7 points and the Cognitive Walkthrough done by usability experts indicated no design or mismatch errors. In the CSUQ usability test the system obtained an average score of 6.13 out of 7, and in the ASQ usability test the overall satisfaction score was 6.64 out of 7. Conclusions We developed hFigures, an open source library for visualizing a complete, accurate and normalized graphical representation of health data. The idea is based on the concept of the hGraph but it provides additional key features, including a comparison of multiple health measurements over time. We conducted a usability evaluation of the library as a key component of an application for health and wellness monitoring. The results indicate that the data visualization library was helpful in assisting users in understanding health data and its evolution over time.

Organisations: Department of Signal Processing

People: Ledesma Figueroa, A. (Creator), Al-Musawi, M. (Creator), Nieminen, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.c.3621374

Dataset

### **Supplementary data for "Physical activity, body mass index and heart rate variability-based stress and recovery in 16 275 Finnish employees: a cross-sectional study"**

Background Physical inactivity, overweight, and work-related stress are major concerns today. Psychological stress causes physiological responses such as reduced heart rate variability (HRV), owing to attenuated parasympathetic and/or increased sympathetic activity in cardiac autonomic control. This study's purpose was to investigate the relationships between physical activity (PA), body mass index (BMI), and HRV-based stress and recovery on workdays, among Finnish employees. Methods The participants in this cross-sectional study were 16 275 individuals (6863 men and 9412 women; age 18–65 years; BMI 18.5–40.0 kg/m<sup>2</sup>). Assessments of stress, recovery and PA were based on HRV data from beat-to-beat R-R interval recording (mainly over 3 days). The validated HRV-derived variables took into account the dynamics and individuality of HRV. Stress percentage (the proportion of stress reactions, workday and working hours), and stress balance (ratio between recovery and stress reactions, sleep) describe the amount of physiological stress and recovery,

respectively. Variables describing the intensity (i.e. magnitude of recognized reactions) of physiological stress and recovery were stress index (workday) and recovery index (sleep), respectively. Moderate to vigorous PA was measured and participants divided into the following groups, based on calculated weekly PA: inactive (0 min), low (0 < 150 min), medium (150–300 min), and high (>300 min). BMI was calculated from self-reported weight and height. Linear models were employed in the main analyses. Results High PA was associated with lower stress percentages (during workdays and working hours) and stress balance. Higher BMI was associated with higher stress index, and lower stress balance and recovery index. These results were similar for men and women ( $P < 0.001$  for all). Conclusion Independent of age and sex, high PA was associated with a lower amount of stress on workdays. Additionally, lower BMI was associated with better recovery during sleep, expressed by a greater amount and magnitude of recovery reactions, which suggests that PA in the long term resulting in improved fitness has a positive effect on recovery, even though high PA may disturb recovery during the following night. Obviously, several factors outside of the study could also affect HRV-based stress.

Organisations: Department of Signal Processing

People: Föhr, T. (Creator), Pietilä, J. (Creator), Helander, E. (Creator), Myllymäki, T. (Creator), Lindholm, H. (Creator), Rusko, H. (Creator), Kujala, U. M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.c.3598376

Dataset

### **DOPC lipid bilayer with PYR8 simulation trajectory and files from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations"**

Simulation files for DOPC with PYR8 simulation from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations" Franova et al. *Biochimica et Biophysica Acta* 1838 (2014) 1406–1411

Organisations: Physics

People: Franova, M. D. (Creator), Vattulainen, I. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.33331

Dataset

### **DOPC lipid bilayer with PYR6 simulation trajectory and files from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations"**

Simulation files for DOPC with PYR6 simulation from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations" Franova et al. *Biochimica et Biophysica Acta* 1838 (2014) 1406–1411

Organisations: Physics

People: Franova, M. D. (Creator), Vattulainen, I. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.33382

Dataset

### **DOPC lipid bilayer with PYR10 simulation trajectory and files from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations"**

Simulation files for DOPC with PYR10 simulation from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations" Franova et al. *Biochimica et Biophysica Acta* 1838 (2014) 1406–1411

Organisations: Physics

People: Franova, M. D. (Creator), Vattulainen, I. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.33394

Dataset

### **Supplementary data for "Molecular electrometer and binding of cations to phospholipid bilayers"**

Collection of molecular dynamics simulation trajectories previously published in the NMRlipids community of Zenodo, and the related input files needed for their analysis. These centered (origin at the lipid bilayer center) trajectories were used in the NMRlipids II project. In addition to trajectories, files containing lipidwise C-H order parameters (for the beta and alpha segments of the PC-lipid headgroup) for each frame are provided.

Organisations: Physics

People: , , Melcr, J. (Creator), Monticelli, L. (Creator), Määttä, J. (Creator), Ollila, O. H. S. (Creator), Tynkkynen, J. (Creator)

Publisher: Tampere University of Technology  
Date made available: 2017  
DOI: 10.5281/zenodo.167336  
Dataset

### **DOPC lipid bilayer simulation trajectory and files from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations"**

Simulation files for pure DOPC (no pyrene present) simulation from "Can pyrene probes be used to measure lateral pressure profiles of lipid membranes? Perspective through atomistic simulations" Franova et al. *Biochimica et Biophysica Acta* 1838 (2014) 1406–1411

Organisations: Physics

People: Franova, M. D. (Creator), Vattulainen, I. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.33055

Dataset

### **TUT Acoustic scenes 2017, Development dataset**

TUT Acoustic Scenes 2017, development dataset consists of 10-seconds audio segments from 15 acoustic scenes: Bus - traveling by bus in the city (vehicle) Cafe / Restaurant - small cafe/restaurant (indoor) Car - driving or traveling as a passenger, in the city (vehicle) City center (outdoor) Forest path (outdoor) Grocery store - medium size grocery store (indoor) Home (indoor) Lakeside beach (outdoor) Library (indoor) Metro station (indoor) Office - multiple persons, typical work day (indoor) Residential area (outdoor) Train (traveling, vehicle) Tram (traveling, vehicle) Urban park (outdoor) Each acoustic scene has 312 segments totaling 52 minutes of audio.

Organisations: Signal Processing

People: Mesaros, A. (Creator), Virtanen, T. (Creator), Heittola, T. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.400515

Dataset

### **TUT Acoustic scenes 2016, Evaluation dataset**

TUT Acoustic Scenes 2016, evaluation dataset consists of 30-seconds audio segments from 15 acoustic scenes: Bus - traveling by bus in the city (vehicle) Cafe / Restaurant - small cafe/restaurant (indoor) Car - driving or traveling as a passenger, in the city (vehicle) City center (outdoor) Forest path (outdoor) Grocery store - medium size grocery store (indoor) Home (indoor) Lakeside beach (outdoor) Library (indoor) Metro station (indoor) Office - multiple persons, typical work day (indoor) Residential area (outdoor) Train (traveling, vehicle) Tram (traveling, vehicle) Urban park (outdoor) Each acoustic scene has 26 segments totaling 13 minutes of audio.

Organisations: Signal Processing

People: Mesaros, A. (Creator), Heittola, T. (Creator), Virtanen, T. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.165995

Dataset

### **TUT Sound events 2017, Development dataset**

TUT Sound events 2017, development dataset consists of 24 audio recordings from a single acoustic scene: Street (outdoor), totaling 1:32:08

Organisations: Signal Processing

People: Mesaros, A. (Creator), Heittola, T. (Creator), Virtanen, T. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5281/zenodo.400516

Dataset

### **Topology and structure of Au<sub>144</sub>(SR<sub>COO</sub>)<sub>60</sub> from "Atomistic Simulations of Functional Au<sub>144</sub>(SR)<sub>60</sub> Gold Nanoparticles in Aqueous Environment"**

Negatively charged monolayer-protected gold nanoparticles (AuNPs) structure and topology files for GROMACS used in DOI: 10.1021/jp301094m. The final structure of the simulation reported in DOI: 10.1021/jp301094m for the neutral case is provided. The gold nanoparticle contain a core of 144 Au atoms and 60 functionalized alkanethiol side groups (undecanyl chain, R = C<sub>11</sub>H<sub>22</sub>), each possessing a negatively charged carboxylic terminal group.

Organisations: Department of Physics

People: , Gurtovenko, A. A. (Creator), Martinez-Seara, H. (Creator), Häkkinen, H. (Creator), Vattulainen, I. (Creator), Akola, J. (Creator)  
Publisher: Tampere University of Technology  
Date made available: 2017  
DOI: 10.5281/zenodo.16154  
Dataset

### **Topology and structure of Au144(SRNH3+)60 from "Atomistic Simulations of Functional Au144(SR)60 Gold Nanoparticles in Aqueous Environment"**

Positively charged monolayer-protected gold nanoparticles (AuNPs) structure and topology files for GROMACS used in DOI: 10.1021/jp301094m. The final structure of the simulation reported in DOI: 10.1021/jp301094m for the neutral case is provided. The gold nanoparticle contain a core of 144 Au atoms and 60 functionalized alkanethiol side groups (undecanyl chain, R = C11H22), each possessing a positively charged ammonium terminal group.

Organisations: Physics

People: , Gurtovenko, A. A. (Creator), Martinez-Seara, H. (Creator), Häkkinen, H. (Creator), Vattulainen, I. (Creator), Akola, J. (Creator)  
Publisher: Tampere University of Technology  
Date made available: 2017  
DOI: 10.5281/zenodo.16155  
Dataset

### **Supplementary data for "GSAR: Bioconductor package for Gene Set analysis in R"**

Background Gene set analysis (in a form of functionally related genes or pathways) has become the method of choice for analyzing omics data in general and gene expression data in particular. There are many statistical methods that either summarize gene-level statistics for a gene set or apply a multivariate statistic that accounts for intergene correlations. Most available methods detect complex departures from the null hypothesis but lack the ability to identify the specific alternative hypothesis that rejects the null. Results GSAR (Gene Set Analysis in R) is an open-source R/Bioconductor software package for gene set analysis (GSA). It implements self-contained multivariate non-parametric statistical methods testing a complex null hypothesis against specific alternatives, such as differences in mean (shift), variance (scale), or net correlation structure. The package also provides a graphical visualization tool, based on the union of two minimum spanning trees, for correlation networks to examine the change in the correlation structures of a gene set between two conditions and highlight influential genes (hubs). Conclusions Package GSAR provides a set of multivariate non-parametric statistical methods that test a complex null hypothesis against specific alternatives. The methods in package GSAR are applicable to any type of omics data that can be represented in a matrix format. Supplementary data includes additional document presenting computational considerations and uniqueness of package GSAR.

Organisations: Faculty of Biomedical Sciences and Engineering

People: Rahmatallah, Y. (Creator), Zybailov, B. (Creator), Emmert-Streib, F. (Creator), Glazko, G. (Creator)  
Publisher: Tampere University of Technology  
Date made available: 2017  
DOI: 10.6084/m9.figshare.c.3673834  
Dataset

### **Supplementary data for "The type of the functional cardiovascular response to upright posture is associated with arterial stiffness: a cross-sectional study in 470 volunteers"**

Background In a cross-sectional study we examined whether the haemodynamic response to upright posture could be divided into different functional phenotypes, and whether the observed phenotypes were associated with known determinants of cardiovascular risk. Methods Volunteers (n=470) without medication with cardiovascular effects were examined using radial pulse wave analysis, whole-body impedance cardiography, and heart rate variability analysis. Based on the passive head-up tilt induced changes in systemic vascular resistance and cardiac output, the principal determinants of blood pressure, a cluster analysis was performed. Results The haemodynamic response could be clustered into 3 categories: upright increase in vascular resistance and decrease in cardiac output were greatest in the first (+45 % and -27 %, respectively), smallest in the second (+2 % and -2 %, respectively), and intermediate (+22 % and -13 %, respectively) in the third group. These groups were named as 'constrictor' (n=109), 'sustainer' (n=222), and 'intermediate' (n=139) phenotypes, respectively. The sustainers were characterized by male predominance, higher body mass index, blood pressure, and also by higher pulse wave velocity, an index of large arterial stiffness, than the other groups (p<0.01 for all). Heart rate variability analysis showed higher supine and upright low frequency/high frequency (LF/HF) ratio in the sustainers than constrictors, indicating increased sympathovagal balance. Upright LF/HF ratio was also higher in the sustainer than intermediate group. In multivariate analysis, independent explanatory factors for higher pulse wave velocity were the sustainer (p<0.022) and intermediate phenotypes (p<0.046), age (p<0.001), body mass index (p<0.001), and hypertension (p<0.001). Conclusions The response to upright posture could be clustered to 3 functional phenotypes. The sustainer phenotype, with smallest upright decrease in cardiac output and highest sympathovagal balance, was independently associated with increased large arterial stiffness. These results indicate an association of the functional haemodynamic phenotype with an acknowledged marker of cardiovascular risk.

Organisations: Department of Electronics and Communications Engineering

People: Tahvanainen, A. M. (Creator), Tikkakoski, A. J. (Creator), Koskela, J. K. (Creator), Nordhausen, K. (Creator), Viitala, J. M. (Creator), Leskinen, M. H. (Creator), Kähönen, M. A. P. (Creator), Kööbi, T. (Creator), Viik, J. (Creator), Mustonen, J. T. (Creator), Pörsti, I. H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.c.3631067

Dataset

### **Supplementary data for "Clinical association analysis of ependymomas and pilocytic astrocytomas reveals elevated FGFR3 and FGFR1 expression in aggressive ependymomas"**

Background Fibroblast growth factor receptors (FGFRs) are well-known proto-oncogenes in several human malignancies and are currently therapeutically targeted in clinical trials. Among glioma subtypes, activating FGFR1 alterations have been observed in a subpopulation of pilocytic astrocytomas while FGFR3 fusions occur in IDH wild-type diffuse gliomas, resulting in high FGFR3 protein expression. The purpose of this study was to associate FGFR1 and FGFR3 protein levels with clinical features and genetic alterations in ependymoma and pilocytic astrocytoma. Methods FGFR1 and FGFR3 expression levels were detected in ependymoma and pilocytic astrocytoma tissues using immunohistochemistry. Selected cases were further analyzed using targeted sequencing. Results Expression of both FGFR1 and FGFR3 varied within all tumor types. In ependymomas, increased FGFR3 or FGFR1 expression was associated with high tumor grade, cerebral location, young patient age, and poor prognosis. Moderate-to-strong expression of FGFR1 and/or FGFR3 was observed in 76% of cerebral ependymomas. Cases with moderate-to-strong expression of both proteins had poor clinical prognosis. In pilocytic astrocytomas, moderate-to-strong FGFR3 expression was detected predominantly in non-pediatric patients. Targeted sequencing of 12 tumors found no protein-altering mutations or fusions in FGFR1 or FGFR3. Conclusions Elevated FGFR3 and FGFR1 protein expression is common in aggressive ependymomas but likely not driven by genetic alterations. Further studies are warranted to evaluate whether ependymoma patients with high FGFR3 and/or FGFR1 expression could benefit from treatment with FGFR inhibitor based therapeutic approaches currently under evaluation in clinical trials.

Organisations: Department of Signal Processing

People: Lehtinen, B. (Creator), Raita, A. (Creator), Kesseli, J. (Creator), Annala, M. (Creator), Nordfors, K. (Creator), Yli-Harja, O. (Creator), Visakorpi, T. (Creator), Nykter, M. (Creator), Haapasalo, H. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.c.3770810

Dataset

### **Supplementary data for "Ultrathin Polyimide Membrane as Cell Carrier for Subretinal Transplantation of Human Embryonic Stem Cell Derived Retinal Pigment Epithelium"**

In this study, we investigated the suitability of ultrathin and porous polyimide (PI) membrane as a carrier for subretinal transplantation of human embryonic stem cell (hESC) -derived retinal pigment epithelial (RPE) cells in rabbits. The in vivo effects of hESC-RPE cells were analyzed by subretinal suspension injection into Royal College of Surgeons (RCS) rats. Rat eyes were analyzed with electroretinography (ERG) and histology. After analyzing the surface and permeability properties of PI, subretinal PI membrane transplantations with and without hESC-RPE were performed in rabbits. The rabbits were followed for three months and eyes analyzed with fundus photography, ERG, optical coherence tomography (OCT), and histology. Animals were immunosuppressed with cyclosporine the entire follow-up time. In dystrophic RCS rats, ERG and outer nuclear layer (ONL) thickness showed some rescue after hESC-RPE injection. Cells positive for human antigen were found in clusters under the retina 41 days post-injection but not anymore after 105 days. In rabbits, OCT showed good placement of the PI. However, there was loss of pigmentation on the hESC-RPE-PI over time. In the eyes with PI alone, no obvious signs of inflammation or retinal atrophy were observed. In the presence of hESC-RPE, mononuclear cell infiltration and retinal atrophy were observed around the membranes. The porous ultrathin PI membrane was well-tolerated in the subretinal space and is a promising scaffold for RPE transplantation. However, the rejection of the transplanted cells seems to be a major problem and the given immunosuppression was insufficient for reduction of xenograft induced inflammation

Organisations: Department of Electronics and Communications Engineering

People: Ilmarinen, T. (Creator), Hiidenmaa, H. (Creator), Kööbi, P. (Creator), Nymark, S. (Creator), Sorkio, A. (Creator), Wang, J. (Creator), Stanzel, B. V. (Creator), Thielges, F. (Creator), Alajuuma, P. T. (Creator), Oksala, O. E. (Creator), Kataja, M. (Creator), Uusitalo, H. (Creator), Skottman, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Ultrathin/1612754>

Dataset

### **Supplementary data for "Robustness and Information Propagation in Attractors of Random Boolean Networks"**

Attractors represent the long-term behaviors of Random Boolean Networks. We study how the amount of information propagated between the nodes when on an attractor, as quantified by the average pairwise mutual information ( $I(A)$ ), relates to the robustness of the attractor to perturbations ( $R(A)$ ). We find that the dynamical regime of the network affects the relationship between  $I(A)$  and  $R(A)$ . In the ordered and chaotic regimes,  $I(A)$  is anti-correlated with  $R(A)$ , implying that

attractors that are highly robust to perturbations have necessarily limited information propagation. Between order and chaos (for so-called "critical" networks) these quantities are uncorrelated. Finite size effects cause this behavior to be visible for a range of networks, from having a sensitivity of 1 to the point where  $I(A)$  is maximized. In this region, the two quantities are weakly correlated and attractors can be almost arbitrarily robust to perturbations without restricting the propagation of information in the network.

Organisations: Department of Signal Processing

People: , , Ribeiro, A. S. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Robustness/121814>

Dataset

### **Supplementary data for "Thermal Transport Characteristics of Human Skin Measured In Vivo Using Ultrathin Conformal Arrays of Thermal Sensors and Actuators"**

Measurements of the thermal transport properties of the skin can reveal changes in physical and chemical states of relevance to dermatological health, skin structure and activity, thermoregulation and other aspects of human physiology. Existing methods for in vivo evaluations demand complex systems for laser heating and infrared thermography, or they require rigid, invasive probes; neither can apply to arbitrary regions of the body, offers modes for rapid spatial mapping, or enables continuous monitoring outside of laboratory settings. Here we describe human clinical studies using mechanically soft arrays of thermal actuators and sensors that laminate onto the skin to provide rapid, quantitative in vivo determination of both the thermal conductivity and thermal diffusivity, in a completely non-invasive manner. Comprehensive analysis of measurements on six different body locations of each of twenty-five human subjects reveal systematic variations and directional anisotropies in the characteristics, with correlations to the thicknesses of the epidermis (EP) and stratum corneum (SC) determined by optical coherence tomography, and to the water content assessed by electrical impedance based measurements. Multivariate statistical analysis establishes four distinct locations across the body that exhibit different physical properties: heel, cheek, palm, and wrist/volar forearm/dorsal forearm. The data also demonstrate that thermal transport correlates negatively with SC and EP thickness and positively with water content, with a strength of correlation that varies from region to region, e.g., stronger in the palmar than in the follicular regions. Supplementary Notes 1–6: Supporting text, figures and tables.

Organisations: Department of Electronics and Communications Engineering

People: Webb, R. C. (Creator), Pielak, R. M. (Creator), Bastien, P. (Creator), Ayers, J. (Creator), , Kurniawan, J. (Creator), Manco, M. (Creator), Lin, A. (Creator), Cho, N. H. (Creator), Malyrchuk, V. (Creator), Balooch, G. (Creator), Rogers, J. A. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: [10.1371/journal.pone.0118131.s001](https://doi.org/10.1371/journal.pone.0118131.s001)

Dataset

### **Supplementary data for "In Vivo Facilitated Diffusion Model"**

Under dilute in vitro conditions transcription factors rapidly locate their target sequence on DNA by using the facilitated diffusion mechanism. However, whether this strategy of alternating between three-dimensional bulk diffusion and one-dimensional sliding along the DNA contour is still beneficial in the crowded interior of cells is highly disputed. Here we use a simple model for the bacterial genome inside the cell and present a semi-analytical model for the in vivo target search of transcription factors within the facilitated diffusion framework. Without having to resort to extensive simulations we determine the mean search time of a lac repressor in a living *E. coli* cell by including parameters deduced from experimental measurements. The results agree very well with experimental findings, and thus the facilitated diffusion picture emerges as a quantitative approach to gene regulation in living bacteria cells. Furthermore we see that the search time is not very sensitive to the parameters characterizing the DNA configuration and that the cell seems to operate very close to optimal conditions for target localization. Local searches as implied by the colocalization mechanism are only found to mildly accelerate the mean search time within our model. In this supporting information we detail the explicit calculations which are beyond the scope of the main text.

Organisations: Department of Physics

People: Bauer, M. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

DOI: [10.1371/journal.pone.0053956.s002](https://doi.org/10.1371/journal.pone.0053956.s002)

Dataset

### **Supplementary data for "Gene Expression Profiling of Immune-Competent Human Cells Exposed to Engineered Zinc Oxide or Titanium Dioxide Nanoparticles"**

A comprehensive in vitro assessment of two commercial metal oxide nanoparticles, TiO<sub>2</sub> and ZnO, was performed using human monocyte-derived macrophages (HMDM), monocyte-derived dendritic cells (MDDC), and Jurkat T cell leukemia-derived cell line. TiO<sub>2</sub> nanoparticles were found to be non-toxic whereas ZnO nanoparticles caused dose-dependent cell death. Subsequently, global gene expression profiling was performed to identify transcriptional response underlying the cytotoxicity caused by ZnO nanoparticles. Analysis was done with doses 1 µg/ml and 10 µg/ml after 6 and 24 h of

exposure. Interestingly, 2703 genes were significantly differentially expressed in HMDM upon exposure to 10 µg/ml ZnO nanoparticles, while in MDDCs only 12 genes were affected. In Jurkat cells, 980 genes were differentially expressed. It is noteworthy that only the gene expression of metallothioneins was upregulated in all the three cell types and a notable proportion of the genes were regulated in a cell type-specific manner. Gene ontology analysis revealed that the top biological processes disturbed in HMDM and Jurkat cells were regulating cell death and growth. In addition, genes controlling immune system development were affected. Using a panel of modified ZnO nanoparticles, we obtained an additional support that the cellular response to ZnO nanoparticles is largely dependent on particle dissolution and show that the ligand used to modify ZnO nanoparticles modulates Zn<sup>2+</sup> leaching. Overall, the study provides an extensive resource of transcriptional markers for mediating ZnO nanoparticle-induced toxicity for further mechanistic studies, and demonstrates the value of assessing nanoparticle responses through a combined transcriptomics and bioinformatics approach.

Organisations: Department of Signal Processing

People: Tuomela, S. (Creator), , Buerki-Thurnherr, T. (Creator), Arslan, O. (Creator), Kunzmann, A. (Creator), Andersson-Willman, B. (Creator), Wick, P. (Creator), Mathur, S. (Creator), Scheynius, A. (Creator), Krug, H. F. (Creator), Fadeel, B. (Creator), Lahesmaa, R. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/Gene\\_Expression/750110](https://figshare.com/articles/Gene_Expression/750110)

Dataset

### Supplementary data for "Unidirectional P-Body Transport during the Yeast Cell Cycle"

P-bodies belong to a large family of RNA granules that are associated with post-transcriptional gene regulation, conserved from yeast to mammals, and influence biological processes ranging from germ cell development to neuronal plasticity. RNA granules can also transport RNAs to specific locations. Germ granules transport maternal RNAs to the embryo, and neuronal granules transport RNAs long distances to the synaptic dendrites. Here we combine microfluidic-based fluorescent microscopy of single cells and automated image analysis to follow p-body dynamics during cell division in yeast. Our results demonstrate that these highly dynamic granules undergo a unidirectional transport from the mother to the daughter cell during mitosis as well as a constrained "hovering" near the bud site half an hour before the bud is observable. Both behaviors are dependent on the Myo4p/She2p RNA transport machinery. Furthermore, single cell analysis of cell size suggests that PBs play an important role in daughter cell growth under nutrient limiting conditions.

Organisations: Pori

People: Garmendia-Torres, C. (Creator), Skupin, A. (Creator), Michael, S. A. (Creator), Ruusuvoori, P. (Creator), Kuwada, N. J. (Creator), Falconnet, D. (Creator), Cary, G. A. (Creator), Hansen, C. (Creator), Wiggins, P. A. (Creator), Dudley, A. M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Unidirectional/1053452>

Dataset

### Supplementary data for "Conformational Changes and Slow Dynamics through Microsecond Polarized Atomistic Molecular Simulation of an Integral Kv1.2 Ion Channel"

Structure and dynamics of voltage-gated ion channels, in particular the motion of the S4 helix, is a highly interesting and hotly debated topic in current membrane protein research. It has critical implications for insertion and stabilization of membrane proteins as well as for finding how transitions occur in membrane proteins—not to mention numerous applications in drug design. Here, we present a full 1 µs atomic-detail molecular dynamics simulation of an integral Kv1.2 ion channel, comprising 120,000 atoms. By applying 0.052 V/nm of hyperpolarization, we observe structural rearrangements, including up to 120° rotation of the S4 segment, changes in hydrogen-bonding patterns, but only low amounts of translation. A smaller rotation (~35°) of the extracellular end of all S4 segments is present also in a reference 0.5 µs simulation without applied field, which indicates that the crystal structure might be slightly different from the natural state of the voltage sensor. The conformation change upon hyperpolarization is closely coupled to an increase in 310 helix contents in S4, starting from the intracellular side. This could support a model for transition from the crystal structure where the hyperpolarization destabilizes S4–lipid hydrogen bonds, which leads to the helix rotating to keep the arginine side chains away from the hydrophobic phase, and the driving force for final relaxation by downward translation is partly entropic, which would explain the slow process. The coordinates of the transmembrane part of the simulated channel actually stay closer to the recently determined higher-resolution Kv1.2 chimera channel than the starting structure for the entire second half of the simulation (0.5–1 µs). Together with lipids binding in matching positions and significant thinning of the membrane also observed in experiments, this provides additional support for the predictive power of microsecond-scale membrane protein simulations.

Organisations: Department of Physics

People: Bjelkmar, P. (Creator), Niemelä, P. S. (Creator), Vattulainen, I. (Creator), Lindahl, E. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Conformational/148493>

Dataset

### **Supplementary data for "Quantification of Dynamic Morphological Drug Responses in 3D Organotypic Cell Cultures by Automated Image Analysis"**

Glandular epithelial cells differentiate into complex multicellular or acinar structures, when embedded in three-dimensional (3D) extracellular matrix. The spectrum of different multicellular morphologies formed in 3D is a sensitive indicator for the differentiation potential of normal, non-transformed cells compared to different stages of malignant progression. In addition, single cells or cell aggregates may actively invade the matrix, utilizing epithelial, mesenchymal or mixed modes of motility. Dynamic phenotypic changes involved in 3D tumor cell invasion are sensitive to specific small-molecule inhibitors that target the actin cytoskeleton. We have used a panel of inhibitors to demonstrate the power of automated image analysis as a phenotypic or morphometric readout in cell-based assays. We introduce a streamlined stand-alone software solution that supports large-scale high-content screens, based on complex and organotypic cultures. AMIDA (Automated Morphometric Image Data Analysis) allows quantitative measurements of large numbers of images and structures, with a multitude of different spheroid shapes, sizes, and textures. AMIDA supports an automated workflow, and can be combined with quality control and statistical tools for data interpretation and visualization. We have used a representative panel of 12 prostate and breast cancer lines that display a broad spectrum of different spheroid morphologies and modes of invasion, challenged by a library of 19 direct or indirect modulators of the actin cytoskeleton which induce systematic changes in spheroid morphology and differentiation versus invasion. These results were independently validated by 2D proliferation, apoptosis and cell motility assays. We identified three drugs that primarily attenuated the invasion and formation of invasive processes in 3D, without affecting proliferation or apoptosis. Two of these compounds block Rac signalling, one affects cellular cAMP/cGMP accumulation. Our approach supports the growing needs for user-friendly, straightforward solutions that facilitate large-scale, cell-based 3D assays in basic research, drug discovery, and target validation.

Organisations: Department of Signal Processing

People: Härmä, V. (Creator), Schukov, H. (Creator), , Ahonen, I. (Creator), Virtanen, J. (Creator), Siitari, H. (Creator), Åkerfelt, M. (Creator), Lötjönen, J. (Creator), Nees, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Quantification/1021263>

Dataset

### **Supplementary data for "Association of Lipidome Remodeling in the Adipocyte Membrane with Acquired Obesity in Humans"**

Identification of early mechanisms that may lead from obesity towards complications such as metabolic syndrome is of great interest. Here we performed lipidomic analyses of adipose tissue in twin pairs discordant for obesity but still metabolically compensated. In parallel we studied more evolved states of obesity by investigating a separated set of individuals considered to be morbidly obese. Despite lower dietary polyunsaturated fatty acid intake, the obese twin individuals had increased proportions of palmitoleic and arachidonic acids in their adipose tissue, including increased levels of ethanolamine plasmalogens containing arachidonic acid. Information gathered from these experimental groups was used for molecular dynamics simulations of lipid bilayers combined with dependency network analysis of combined clinical, lipidomics, and gene expression data. The simulations suggested that the observed lipid remodeling maintains the biophysical properties of lipid membranes, at the price, however, of increasing their vulnerability to inflammation. Conversely, in morbidly obese subjects, the proportion of plasmalogens containing arachidonic acid in the adipose tissue was markedly decreased. We also show by in vitro Elov16 knockdown that the lipid network regulating the observed remodeling may be amenable to genetic modulation. Together, our novel approach suggests a physiological mechanism by which adaptation of adipocyte membranes to adipose tissue expansion associates with positive energy balance, potentially leading to higher vulnerability to inflammation in acquired obesity. Further studies will be needed to determine the cause of this effect.

Organisations: Department of Physics

People: Pietiläinen, K. (Creator), Rog, T. (Creator), Seppänen-Laakso, T. (Creator), Virtue, S. (Creator), Gopalacharyulu, P. (Creator), Tang, J. (Creator), Rodriguez-Cuenca, S. (Creator), , Naukkarinen, J. (Creator), Ruskeepää, A. (Creator), Niemelä, P. S. (Creator), Yetukuri, L. (Creator), Tan, C. Y. (Creator), Velagapudi, V. (Creator), Castillo, S. (Creator), Nygren, H. (Creator), Hyötyläinen, T. (Creator), Rissanen, A. (Creator), Kaprio, J. (Creator), Yki-Järvinen, H. (Creator), Vattulainen, I. (Creator), Vidal-Puig, A. (Creator), Oresic, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Association/136152>

Dataset

### **Supplementary data for "Cholesterol Induces Specific Spatial and Orientational Order in Cholesterol/Phospholipid Membranes"**

Background In lipid bilayers, cholesterol facilitates the formation of the liquid-ordered phase and enables the formation of laterally ordered structures such as lipid rafts. While these domains have an important role in a variety of cellular processes, the precise atomic-level mechanisms responsible for cholesterol's specific ordering and packing capability have remained unresolved. Methodology/Principal Findings Our atomic-scale molecular dynamics simulations reveal that this ordering and the associated packing effects in membranes largely result from cholesterol's molecular structure, which



differentiates cholesterol from other sterols. We find that cholesterol molecules prefer to be located in the second coordination shell, avoiding direct cholesterol-cholesterol contacts, and form a three-fold symmetric arrangement with proximal cholesterol molecules. At larger distances, the lateral three-fold organization is broken by thermal fluctuations. For other sterols having less structural asymmetry, the three-fold arrangement is considerably lost.

**Conclusions/Significance** We conclude that cholesterol molecules act collectively in lipid membranes. This is the main reason why the liquid-ordered phase only emerges for Chol concentrations well above 10 mol% where the collective self-organization of Chol molecules emerges spontaneously. The collective ordering process requires specific molecular-scale features that explain why different sterols have very different membrane ordering properties: the three-fold symmetry in the Chol-Chol organization arises from the cholesterol off-plane methyl groups allowing the identification of raft-promoting sterols from those that do not promote rafts.

Organisations: Department of Physics

People: Martinez-Seara, H. (Creator), Rog, T. (Creator), Karttunen, M. (Creator), Vattulainen, I. (Creator), Reigada, R. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Cholesterol/143019>

Dataset

### **Supplementary data for "Adipose Tissue Dysfunction and Altered Systemic Amino Acid Metabolism Are Associated with Non-Alcoholic Fatty Liver Disease"**

**Background** Fatty liver is a major cause of obesity-related morbidity and mortality. The aim of this study was to identify early metabolic alterations associated with liver fat accumulation in 50- to 55-year-old men (n = 49) and women (n = 52) with and without NAFLD. **Methods** Hepatic fat content was measured using proton magnetic resonance spectroscopy (1H MRS). Serum samples were analyzed using a nuclear magnetic resonance (NMR) metabolomics platform. Global gene expression profiles of adipose tissues and skeletal muscle were analyzed using Affymetrix microarrays and quantitative PCR. Muscle protein expression was analyzed by Western blot. **Results** Increased branched-chain amino acid (BCAA), aromatic amino acid (AAA) and orosomucoid were associated with liver fat accumulation already in its early stage, independent of sex, obesity or insulin resistance (p<0.05 for all). Significant down-regulation of BCAA catabolism and fatty acid and energy metabolism was observed in the adipose tissue of the NAFLD group (p<0.001 for all), whereas no aberrant gene expression in the skeletal muscle was found. Reduced BCAA catabolic activity was inversely associated with serum BCAA and liver fat content (p<0.05 for all). **Conclusions** Liver fat accumulation, already in its early stage, is associated with increased serum branched-chain and aromatic amino acids. The observed associations of decreased BCAA catabolism activity, mitochondrial energy metabolism and serum BCAA concentration with liver fat content suggest that adipose tissue dysfunction may have a key role in the systemic nature of NAFLD pathogenesis. E-MTAB-3616 - Transcription profiling by array of skeletal muscle tissue of several families to study the role of adiposity-related low-grade inflammation on interactions between adipose tissue, muscle, and bone

Organisations: Department of Signal Processing

People: Cheng, S. (Creator), Wiklund, P. (Creator), Borra, R. (Creator), Ojanen, X. (Creator), Xu, L. (Creator), Törmäkangas, T. (Creator), Alen, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-3616/>,

[https://figshare.com/articles/Adipose\\_Tissue/1565858](https://figshare.com/articles/Adipose_Tissue/1565858)

Dataset

### **Supplementary data for "Inter-Subject Correlation in fMRI: Method Validation against Stimulus-Model Based Analysis"**

Within functional magnetic resonance imaging (fMRI), the use of the traditional general linear model (GLM) based analysis methods is often restricted to strictly controlled research setups requiring a parametric activation model. Instead, Inter-Subject Correlation (ISC) method is based on voxel-wise correlation between the time series of the subjects, which makes it completely non-parametric and thus suitable for naturalistic stimulus paradigms such as movie watching. In this study, we compared an ISC based analysis results with those of a GLM based in five distinct controlled research setups. We used International Consortium for Brain Mapping functional reference battery (FRB) fMRI data available from the Laboratory of Neuro Imaging image data archive. The selected data included measurements from 37 right-handed subjects, who all had performed the same five tasks from FRB. The GLM was expected to locate activations accurately in FRB data and thus provide good grounds for investigating relationship between ISC and stimulus induced fMRI activation. The statistical maps of ISC and GLM were compared with two measures. The first measure was the Pearson's correlation between the non-thresholded ISC test-statistics and absolute values of the GLM Z-statistics. The average correlation value over five tasks was 0.74. The second was the Dice index between the activation regions of the methods. The average Dice value over the tasks and three threshold levels was 0.73. The results of this study indicated how the data driven ISC analysis found the same foci as the model-based GLM analysis. The agreement of the results is highly interesting, because ISC is applicable in situations where GLM is not suitable, for example, when analyzing data from a naturalistic stimuli experiment.

Organisations: Department of Signal Processing

People: , Kauppi, J. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/Inter\\_Subject/121456](https://figshare.com/articles/Inter_Subject/121456)

Dataset

### **Supplementary data for "Information-Theoretic Analysis of the Dynamics of an Executable Biological Model"**

To facilitate analysis and understanding of biological systems, large-scale data are often integrated into models using a variety of mathematical and computational approaches. Such models describe the dynamics of the biological system and can be used to study the changes in the state of the system over time. For many model classes, such as discrete or continuous dynamical systems, there exist appropriate frameworks and tools for analyzing system dynamics. However, the heterogeneous information that encodes and bridges molecular and cellular dynamics, inherent to fine-grained molecular simulation models, presents significant challenges to the study of system dynamics. In this paper, we present an algorithmic information theory based approach for the analysis and interpretation of the dynamics of such executable models of biological systems. We apply a normalized compression distance (NCD) analysis to the state representations of a model that simulates the immune decision making and immune cell behavior. We show that this analysis successfully captures the essential information in the dynamics of the system, which results from a variety of events including proliferation, differentiation, or perturbations such as gene knock-outs. We demonstrate that this approach can be used for the analysis of executable models, regardless of the modeling framework, and for making experimentally quantifiable predictions.

Organisations: Department of Signal Processing

People: Sadot, A. (Creator), , Amir-Kroll, H. (Creator), Zhang, W. (Creator), Nykter, M. (Creator), Shmulevich, I. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/Information\\_Theoretic/654861](https://figshare.com/articles/Information_Theoretic/654861)

Dataset

### **Supplementary data for "A Linear Model for Transcription Factor Binding Affinity Prediction in Protein Binding Microarrays"**

Protein binding microarrays (PBM) are a high throughput technology used to characterize protein-DNA binding. The arrays measure a protein's affinity toward thousands of double-stranded DNA sequences at once, producing a comprehensive binding specificity catalog. We present a linear model for predicting the binding affinity of a protein toward DNA sequences based on PBM data. Our model represents the measured intensity of an individual probe as a sum of the binding affinity contributions of the probe's subsequences. These subsequences characterize a DNA binding motif and can be used to predict the intensity of protein binding against arbitrary DNA sequences. Our method was the best performer in the Dialogue for Reverse Engineering Assessments and Methods 5 (DREAM5) transcription factor/DNA motif recognition challenge. For the DREAM5 bonus challenge, we also developed an approach for the identification of transcription factors based on their PBM binding profiles. Our approach for TF identification achieved the best performance in the bonus challenge.

Organisations: Department of Signal Processing

People: , Laurila, K. (Creator), Lähdesmäki, H. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/A\\_Linear\\_Model/136446](https://figshare.com/articles/A_Linear_Model/136446)

Dataset

### **Supplementary data for "NanoMiner — Integrative Human Transcriptomics Data Resource for Nanoparticle Research"**

The potential impact of nanoparticles on the environment and on human health has attracted considerable interest worldwide. The amount of transcriptomics data, in which tissues and cell lines are exposed to nanoparticles, increases year by year. In addition to the importance of the original findings, this data can have value in broader context when combined with other previously acquired and published results. In order to facilitate the efficient usage of the data, we have developed the NanoMiner web resource (<http://nanominer.cs.tut.fi/>), which contains 404 human transcriptome samples exposed to various types of nanoparticles. All the samples in NanoMiner have been annotated, preprocessed and normalized using standard methods that ensure the quality of the data analyses and enable the users to utilize the database systematically across the different experimental setups and platforms. With NanoMiner it is possible to 1) search and plot the expression profiles of one or several genes of interest, 2) cluster the samples within the datasets, 3) find differentially expressed genes in various nanoparticle studies, 4) detect the nanoparticles causing differential expression of selected genes, 5) analyze enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) terms for the detected genes and 6) search the expression values and differential expressions of the genes belonging to a specific KEGG pathway or Gene Ontology. In sum, NanoMiner database is a valuable collection of microarray data which can be also used as a data repository for future analyses.

Organisations: Department of Signal Processing

People: , Tuomela, S. (Creator), Hahne, L. (Creator), Ahlfors, H. (Creator), Yli-Harja, O. (Creator), Fadeel, B. (Creator), Lahesmaa, R. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/NanoMiner/744959>

### Supplementary data for "Assessing the Nature of Lipid Raft Membranes"

The paradigm of biological membranes has recently gone through a major update. Instead of being fluid and homogeneous, recent studies suggest that membranes are characterized by transient domains with varying fluidity. In particular, a number of experimental studies have revealed the existence of highly ordered lateral domains rich in sphingomyelin and cholesterol (CHOL). These domains, called functional lipid rafts, have been suggested to take part in a variety of dynamic cellular processes such as membrane trafficking, signal transduction, and regulation of the activity of membrane proteins. However, despite the proposed importance of these domains, their properties, and even the precise nature of the lipid phases, have remained open issues mainly because the associated short time and length scales have posed a major challenge to experiments. In this work, we employ extensive atom-scale simulations to elucidate the properties of ternary raft mixtures with CHOL, palmitoylsphingomyelin (PSM), and palmitoylcholine. We simulate two bilayers of 1,024 lipids for 100 ns in the liquid-ordered phase and one system of the same size in the liquid-disordered phase. The studies provide evidence that the presence of PSM and CHOL in raft-like membranes leads to strongly packed and rigid bilayers. We also find that the simulated raft bilayers are characterized by nanoscale lateral heterogeneity, though the slow lateral diffusion renders the interpretation of the observed lateral heterogeneity more difficult. The findings reveal aspects of the role of favored (specific) lipid–lipid interactions within rafts and clarify the prominent role of CHOL in altering the properties of the membrane locally in its neighborhood. Also, we show that the presence of PSM and CHOL in rafts leads to intriguing lateral pressure profiles that are distinctly different from corresponding profiles in nonraft-like membranes. The results propose that the functioning of certain classes of membrane proteins is regulated by changes in the lateral pressure profile, which can be altered by a change in lipid content.

Organisations: Department of Physics

People: Niemelä, P. S. (Creator), Hyvönen, M. T. (Creator), , Karttunen, M. (Creator), Vattulainen, I. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Assessing/152529>

Dataset

### Supplementary data for "Efflux Protein Expression in Human Stem Cell-Derived Retinal Pigment Epithelial Cells"

Retinal pigment epithelial (RPE) cells in the back of the eye nourish photoreceptor cells and form a selective barrier that influences drug transport from the blood to the photoreceptor cells. At the molecular level, ATP-dependent efflux transporters have a major role in drug delivery in human RPE. In this study, we assessed the relative expression of several ATP-dependent efflux transporter genes (MRP1, -2, -3, -4, -5, -6, p-gp, and BCRP), the protein expression and localization of MRP1, MRP4, and MRP5, and the functionality of MRP1 efflux pumps at different maturation stages of undifferentiated human embryonic stem cells (hESC) and RPE derived from the hESC (hESC-RPE). Our findings revealed that the gene expression of ATP-dependent efflux transporters MRP1, -3, -4, -5, and p-gp fluctuated during hESC-RPE maturation from undifferentiated hESC to fusiform, epithelioid, and finally to cobblestone hESC-RPE. Epithelioid hESC-RPE had the highest expression of MRP1, -3, -4, and P-gp, whereas the most mature cobblestone hESC-RPE had the highest expression of MRP5 and MRP6. These findings indicate that a similar efflux protein profile is shared between hESC-RPE and the human RPE cell line, ARPE-19, and suggest that hESC-RPE cells are suitable in vitro RPE models for drug transport studies. Embryonic stem cell model might provide a novel tool to study retinal cell differentiation, mechanisms of RPE -derived diseases, drug testing and targeted drug therapy. Table S1: Gene expression data with standard deviations (SD) and calculations of statistical significance (p).

Organisations: BioMediTech

People: Juuti-Uusitalo, K. (Creator), Vaajasaari, H. (Creator), Ryhänen, T. (Creator), Narkilahti, S. (Creator), Suuronen, R. (Creator), Mannermaa, E. (Creator), Kaarniranta, K. (Creator), Skottman, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.1371/journal.pone.0030089.s001

Dataset

### Supplementary data for "Mutually Exclusive Roles of SHARPIN in Integrin Inactivation and NF-κB Signaling"

SHANK-associated RH domain interactor (SHARPIN) inhibits integrins through interaction with the integrin  $\alpha$ -subunit. In addition, SHARPIN enhances nuclear factor- $\kappa$ B (NF- $\kappa$ B) activity as a component of the linear ubiquitin chain assembly complex (LUBAC). However, it is currently unclear how regulation of these seemingly different roles is coordinated. Here, we show that SHARPIN binds integrin and LUBAC in a mutually exclusive manner. We map the integrin binding site on SHARPIN to the ubiquitin-like (UBL) domain, the same domain implicated in SHARPIN interaction with LUBAC component RNF31 (ring finger protein 31), and identify two SHARPIN residues (V267, L276) required for both integrin and RNF31 regulation. Accordingly, the integrin  $\alpha$ -tail is capable of competing with RNF31 for SHARPIN binding in vitro. Importantly, the full SHARPIN RNF31-binding site contains residues (F263A/I272A) that are dispensable for SHARPIN-integrin interaction. Importantly, disrupting SHARPIN interaction with integrin or RNF31 abolishes SHARPIN-mediated regulation of integrin or NF- $\kappa$ B activity, respectively. Altogether these data suggest that the roles of SHARPIN in inhibiting integrin activity and supporting linear ubiquitination are (molecularly) distinct. S1 Fig: Expression levels of WT or mutant GFP-SHARPIN. (A) Western blot analysis of WT or mutant GFP-SHARPIN in HeLa cells. (B) Western blot analysis of GFP alone or WT or mutant GFP-SHARPIN in CHO cells. Also the levels of RFP-TALIN head were determined. Non-

transfected CHO cells were used as control.

Organisations: Department of Physics

People: Franceschi, N. D. (Creator), Peuhu, E. (Creator), Parsons, M. (Creator), , Vattulainen, I. (Creator), Salmi, M. (Creator), Ivaska, J. (Creator), Pouwels, J. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.1371/journal.pone.0143423.s001

Dataset

### Supplementary data for "Reconstruction and Validation of RefRec: A Global Model for the Yeast Molecular Interaction Network"

Molecular interaction networks establish all cell biological processes. The networks are under intensive research that is facilitated by new high-throughput measurement techniques for the detection, quantification, and characterization of molecules and their physical interactions. For the common model organism yeast *Saccharomyces cerevisiae*, public databases store a significant part of the accumulated information and, on the way to better understanding of the cellular processes, there is a need to integrate this information into a consistent reconstruction of the molecular interaction network. This work presents and validates RefRec, the most comprehensive molecular interaction network reconstruction currently available for yeast. The reconstruction integrates protein synthesis pathways, a metabolic network, and a protein-protein interaction network from major biological databases. The core of the reconstruction is based on a reference object approach in which genes, transcripts, and proteins are identified using their primary sequences. This enables their unambiguous identification and non-redundant integration. The obtained total number of different molecular species and their connecting interactions is ~67,000. In order to demonstrate the capacity of RefRec for functional predictions, it was used for simulating the gene knockout damage propagation in the molecular interaction network in ~590,000 experimentally validated mutant strains. Based on the simulation results, a statistical classifier was subsequently able to correctly predict the viability of most of the strains. The results also showed that the usage of different types of molecular species in the reconstruction is important for accurate phenotype prediction. In general, the findings demonstrate the benefits of global reconstructions of molecular interaction networks. With all the molecular species and their physical interactions explicitly modeled, our reconstruction is able to serve as a valuable resource in additional analyses involving objects from multiple molecular -omes. For that purpose, RefRec is freely available in the Systems Biology Markup Language format. Table S1: Unconditionally essential genes and their enrichments in GO Biological Processes.

Organisations: Department of Signal Processing, Department of Signal Processing

People: , Almusa, H. (Creator), Matilainen, J. (Creator), , Ruusuvaori, P. (Creator), , Wilhelm, T. (Creator), , Beyer, A. (Creator), Harju, M. (Creator), , , Roos, C. (Creator), Yli-Harja, O. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.1371/journal.pone.0010662.s001

Dataset

### Supplementary data for "Role of Lipids in Spheroidal High Density Lipoproteins"

We study the structure and dynamics of spherical high density lipoprotein (HDL) particles through coarse-grained multi-microsecond molecular dynamics simulations. We simulate both a lipid droplet without the apolipoprotein A-I (apoA-I) and the full HDL particle including two apoA-I molecules surrounding the lipid compartment. The present models are the first ones among computational studies where the size and lipid composition of HDL are realistic, corresponding to human serum HDL. We focus on the role of lipids in HDL structure and dynamics. Particular attention is paid to the assembly of lipids and the influence of lipid-protein interactions on HDL properties. We find that the properties of lipids depend significantly on their location in the particle (core, intermediate region, surface). Unlike the hydrophobic core, the intermediate and surface regions are characterized by prominent conformational lipid order. Yet, not only the conformations but also the dynamics of lipids are found to be distinctly different in the different regions of HDL, highlighting the importance of dynamics in considering the functionalization of HDL. The structure of the lipid droplet close to the HDL-water interface is altered by the presence of apoA-I, with most prominent changes being observed for cholesterol and polar lipids. For cholesterol, slow trafficking between the surface layer and the regimes underneath is observed. The lipid-protein interactions are strongest for cholesterol, in particular its interaction with hydrophobic residues of apoA-I. Our results reveal that not only hydrophobicity but also conformational entropy of the molecules are the driving forces in the formation of HDL structure. The results provide the first detailed structural model for HDL and its dynamics with and without apoA-I, and indicate how the interplay and competition between entropy and detailed interactions may be used in nanoparticle and drug design through self-assembly.

Organisations: Department of Physics, Department of Physics

People: , , Niemelä, P. S. (Creator), , Hyvönen, M. T. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/Role\\_of\\_Lipids/140869](https://figshare.com/articles/Role_of_Lipids/140869)

Dataset

### **Supplementary data for "Structure-Dynamics Relationships in Bursting Neuronal Networks Revealed Using a Prediction Framework"**

The question of how the structure of a neuronal network affects its functionality has gained a lot of attention in neuroscience. However, the vast majority of the studies on structure-dynamics relationships consider few types of network structures and assess limited numbers of structural measures. In this *in silico* study, we employ a wide diversity of network topologies and search among many possibilities the aspects of structure that have the greatest effect on the network excitability. The network activity is simulated using two point-neuron models, where the neurons are activated by noisy fluctuation of the membrane potential and their connections are described by chemical synapse models, and statistics on the number and quality of the emergent network bursts are collected for each network type. We apply a prediction framework to the obtained data in order to find out the most relevant aspects of network structure. In this framework, predictors that use different sets of graph-theoretic measures are trained to estimate the activity properties, such as burst count or burst length, of the networks. The performances of these predictors are compared with each other. We show that the best performance in prediction of activity properties for networks with sharp in-degree distribution is obtained when the prediction is based on clustering coefficient. By contrast, for networks with broad in-degree distribution, the maximum eigenvalue of the connectivity graph gives the most accurate prediction. The results shown for small ( $N < 100$ ) networks hold with few exceptions when different neuron models, different choices of neuron population and different average degrees are applied. We confirm our conclusions using larger ( $N < 1000$ ) networks as well. Our findings reveal the relevance of different aspects of network structure from the viewpoint of network excitability, and our integrative method could serve as a general framework for structure-dynamics studies in biosciences.

Organisations: Department of Signal Processing, Department of Mathematics

People: , Acimovic, J. (Creator), , Linne, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Structure/754895>

Dataset

### **Supplementary data for "Fluctuations of Hi-Hat Timing and Dynamics in a Virtuoso Drum Track of a Popular Music Recording"**

Long-range correlated temporal fluctuations in the beats of musical rhythms are an inevitable consequence of human action. According to recent studies, such fluctuations also lead to a favored listening experience. The scaling laws of amplitude variations in rhythms, however, are widely unknown. Here we use highly sensitive onset detection and time series analysis to study the amplitude and temporal fluctuations of Jeff Porcaro's one-handed hi-hat pattern in "I Keep Forgetting"—one of the most renowned 16th note patterns in modern drumming. We show that fluctuations of hi-hat amplitudes and interbeat intervals (times between hits) have clear long-range correlations and short-range anticorrelations separated by a characteristic time scale. In addition, we detect subtle features in Porcaro's drumming such as small drifts in the 16th note pulse and non-trivial periodic two-bar patterns in both hi-hat amplitudes and intervals. Through this investigation we introduce a step towards statistical studies of the 20th and 21st century music recordings in the framework of complex systems. Our analysis has direct applications to the development of drum machines and to drumming pedagogy. The file contains all the detected onsets in "I Keep Forgetting". The onset times (in seconds) and the corresponding amplitudes (in arbitrary units) are given in the first and second column, respectively.

Organisations: Department of Physics, Department of Signal Processing

People: Räsänen, E. (Creator), , Virtanen, T. (Creator), Zollner, M. (Creator), Hennig, H. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.1371/journal.pone.0127902.s001

Dataset

### **Supplementary data for "Fluorescent Protein Based FRET Pairs with Improved Dynamic Range for Fluorescence Lifetime Measurements"**

Fluorescence Resonance Energy Transfer (FRET) using fluorescent protein variants is widely used to study biochemical processes in living cells. FRET detection by fluorescence lifetime measurements is the most direct and robust method to measure FRET. The traditional cyan-yellow fluorescent protein based FRET pairs are getting replaced by green-red fluorescent protein variants. The green-red pair enables excitation at a longer wavelength which reduces cellular autofluorescence and phototoxicity while monitoring FRET. Despite the advances in FRET based sensors, the low FRET efficiency and dynamic range still complicates their use in cell biology and high throughput screening. In this paper, we utilized the higher lifetime of NowGFP and screened red fluorescent protein variants to develop FRET pairs with high dynamic range and FRET efficiency. The FRET variations were analyzed by proteolytic activity and detected by steady-state and time-resolved measurements. Based on the results, NowGFP-tdTomato and NowGFP-mRuby2 have shown high potentials as FRET pairs with large fluorescence lifetime dynamic range. The *in vitro* measurements revealed that the NowGFP-tdTomato has the highest Förster radius for any fluorescent protein based FRET pairs yet used in biological studies. The developed FRET pairs will be useful for designing FRET based sensors and studies employing Fluorescence Lifetime Imaging Microscopy (FLIM).

Organisations: Department of Chemistry and Bioengineering

People: , Sarkisyan, K. S. (Creator), Mishin, A. S. (Creator), Santala, V. (Creator), Tkachenko, N. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Fluorescent/1501424>

Dataset

### Supplementary data for "Co-Exposure with Fullerene May Strengthen Health Effects of Organic Industrial Chemicals"

In vitro toxicological studies together with atomistic molecular dynamics simulations show that occupational co-exposure with C60 fullerene may strengthen the health effects of organic industrial chemicals. The chemicals studied are acetophenone, benzaldehyde, benzyl alcohol, m-cresol, and toluene which can be used with fullerene as reagents or solvents in industrial processes. Potential co-exposure scenarios include a fullerene dust and organic chemical vapor, or a fullerene solution aerosolized in workplace air. Unfiltered and filtered mixtures of C60 and organic chemicals represent different co-exposure scenarios in in vitro studies where acute cytotoxicity and immunotoxicity of C60 and organic chemicals are tested together and alone by using human THP-1-derived macrophages. Statistically significant co-effects are observed for an unfiltered mixture of benzaldehyde and C60 that is more cytotoxic than benzaldehyde alone, and for a filtered mixture of m-cresol and C60 that is slightly less cytotoxic than m-cresol. Hydrophobicity of chemicals correlates with co-effects when secretion of pro-inflammatory cytokines IL-1 $\beta$  and TNF- $\alpha$  is considered. Complementary atomistic molecular dynamics simulations reveal that C60 co-aggregates with all chemicals in aqueous environment. Stable aggregates have a fullerene-rich core and a chemical-rich surface layer, and while essentially all C60 molecules aggregate together, a portion of organic molecules remains in water.

Organisations: Department of Physics

People: Lehto, M. (Creator), Rog, T. (Creator), Vanhala, E. (Creator), Tornaueus, J. (Creator), Taberman, H. (Creator), Jänis, J. (Creator), Alenius, H. (Creator), Vattulainen, I. (Creator), Laine, O. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/Co\\_Exposure/1260383](https://figshare.com/articles/Co_Exposure/1260383)

Dataset

### Supplementary data for "In Vivo Transcription Kinetics of a Synthetic Gene Uninvolved in Stress-Response Pathways in Stressed Escherichia coli Cells"

The fast adaptation of *Escherichia coli* to stressful environments includes the regulation of gene expression rates, mainly of transcription, by specific and global stress-response mechanisms. To study the effects of mechanisms acting on a global level, we observed with single molecule sensitivity the effects of mild acidic shift and oxidative stress on the in vivo transcription dynamics of a probe gene encoding an RNA target for MS2d-GFP, under the control of a synthetic promoter. After showing that this promoter is uninvolved in fast stress-response pathways, we compared its kinetics of transcript production under stress and in optimal conditions. We find that, following the application of either stress, the mean rates of transcription activation and of subsequent RNA production of the probe gene are reduced, particularly under oxidative stress. Meanwhile, the noise in RNA production decreases under oxidative stress, but not under acidic shift. From distributions of intervals between consecutive RNA productions, we infer that the number and duration of the rate-limiting steps in transcription initiation change, following the application of stress. These changes differ in the two stress conditions and are consistent with the changes in noise in RNA production. Overall, our measurements of the transcription initiation kinetics of the probe gene indicate that, following sub-lethal stresses, there are stress-specific changes in the dynamics of transcription initiation of the probe gene that affect its mean rate and noise of transcript production. Given the non-involvement of the probe gene in stress-response pathways, we suggest that these changes are caused by global response mechanisms of *E. coli* to stress. The data contains, for each condition, the total intensity of spots in each cell after subtraction of background fluorescence, the values of the function that is fitted to these values to detect the production events, and the count of RNA molecules produced so far in each cell, at each time point.

Organisations: Department of Signal Processing

People: , Martikainen, A. (Creator), , Sanches Ribeiro, A. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: [https://figshare.com/articles/In\\_Vivo\\_Transcription/1187355](https://figshare.com/articles/In_Vivo_Transcription/1187355)

Dataset

### Supplementary data for "Enzymatic Oxidation of Cholesterol: Properties and Functional Effects of Cholestenone in Cell Membranes"

Bacterial cholesterol oxidase is commonly used as an experimental tool to reduce cellular cholesterol content. That the treatment also generates the poorly degradable metabolite 4-cholesten-3-one (cholestenone) has received less attention. Here, we investigated the membrane partitioning of cholestenone using simulations and cell biological experiments and assessed the functional effects of cholestenone in human cells. Atomistic simulations predicted that cholestenone reduces membrane order, undergoes faster flip-flop and desorbs more readily from membranes than cholesterol. In primary human fibroblasts, cholestenone was released from membranes to physiological extracellular acceptors more avidly than cholesterol, but without acceptors it remained in cells over a day. To address the functional effects of cholestenone, we studied fibroblast migration during wound healing. When cells were either cholesterol oxidase treated or part of cellular cholesterol was exchanged for cholestenone with cyclodextrin, cell migration during 22 h was markedly inhibited. Instead, when a similar fraction of cholesterol was removed using cyclodextrin, cells replenished their cholesterol content in 3 h and migrated similarly to control cells. Thus, cholesterol oxidation produces long-term functional effects in cells and these

are in part due to the generated membrane active cholesterol.

Organisations: Department of Physics

People: Neuvonen, M. (Creator), , Mokka, S. (Creator), , Rog, T. (Creator), Liu, Z. (Creator), Bittman, R. (Creator), Vattulainen, I. (Creator), Ikonen, E. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Enzymatic/1151995>

Dataset

### **Supplementary data for "Compensation of Missing Wedge Effects with Sequential Statistical Reconstruction in Electron Tomography"**

Electron tomography (ET) of biological samples is used to study the organization and the structure of the whole cell and subcellular complexes in great detail. However, projections cannot be acquired over full tilt angle range with biological samples in electron microscopy. ET image reconstruction can be considered an ill-posed problem because of this missing information. This results in artifacts, seen as the loss of three-dimensional (3D) resolution in the reconstructed images. The goal of this study was to achieve isotropic resolution with a statistical reconstruction method, sequential maximum a posteriori expectation maximization (sMAP-EM), using no prior morphological knowledge about the specimen. The missing wedge effects on sMAP-EM were examined with a synthetic cell phantom to assess the effects of noise. An experimental dataset of a multivesicular body was evaluated with a number of gold particles. An ellipsoid fitting based method was developed to realize the quantitative measures elongation and contrast in an automated, objective, and reliable way. The method statistically evaluates the sub-volumes containing gold particles randomly located in various parts of the whole volume, thus giving information about the robustness of the volume reconstruction. The quantitative results were also compared with reconstructions made with widely-used weighted backprojection and simultaneous iterative reconstruction technique methods. The results showed that the proposed sMAP-EM method significantly suppresses the effects of the missing information producing isotropic resolution. Furthermore, this method improves the contrast ratio, enhancing the applicability of further automatic and semi-automatic analysis. These improvements in ET reconstruction by sMAP-EM enable analysis of subcellular structures with higher three-dimensional resolution and contrast than conventional methods.

Organisations: Department of Signal Processing

People: Paavolainen, L. (Creator), , , Peltonen, S. (Creator), , Soonsawad, P. (Creator), Marjomäki, V. (Creator), Cheng, R. H. (Creator),

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Compensation/1193688>

Dataset

### **Supplementary data for "Lipid Exchange Mechanism of the Cholesteryl Ester Transfer Protein Clarified by Atomistic and Coarse-grained Simulations"**

Cholesteryl ester transfer protein (CETP) transports cholesteryl esters, triglycerides, and phospholipids between different lipoprotein fractions in blood plasma. The inhibition of CETP has been shown to be a sound strategy to prevent and treat the development of coronary heart disease. We employed molecular dynamics simulations to unravel the mechanisms associated with the CETP-mediated lipid exchange. To this end we used both atomistic and coarse-grained models whose results were consistent with each other. We found CETP to bind to the surface of high density lipoprotein (HDL)-like lipid droplets through its charged and tryptophan residues. Upon binding, CETP rapidly (in about 10 ns) induced the formation of a small hydrophobic patch to the phospholipid surface of the droplet, opening a route from the core of the lipid droplet to the binding pocket of CETP. This was followed by a conformational change of helix X of CETP to an open state, in which we found the accessibility of cholesteryl esters to the C-terminal tunnel opening of CETP to increase. Furthermore, in the absence of helix X, cholesteryl esters rapidly diffused into CETP through the C-terminal opening. The results provide compelling evidence that helix X acts as a lid which conducts lipid exchange by alternating the open and closed states. The findings have potential for the design of novel molecular agents to inhibit the activity of CETP. Figure S1: Spatial density maps for one of the atomistic systems and for two additional coarse-grained models. Figure S2: Number of intrinsic contacts of CETP as a function of time.

Organisations: Department of Physics, Department of Physics

People: , , Kovanen, P. T. (Creator), Vattulainen, I. (Creator), Hyvönen, M. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Lipid/129702>

Dataset

### **Supplementary data for "How Anacetrapib Inhibits the Activity of the Cholesteryl Ester Transfer Protein? Perspective through Atomistic Simulations"**

Cholesteryl ester transfer protein (CETP) mediates the reciprocal transfer of neutral lipids (cholesteryl esters, triglycerides) and phospholipids between different lipoprotein fractions in human blood plasma. A novel molecular agent known as anacetrapib has been shown to inhibit CETP activity and thereby raise high density lipoprotein (HDL)-cholesterol and decrease low density lipoprotein (LDL)-cholesterol, thus rendering CETP inhibition an attractive target to prevent and treat

the development of various cardiovascular diseases. Our objective in this work is to use atomistic molecular dynamics simulations to shed light on the inhibitory mechanism of anacetrapib and unlock the interactions between the drug and CETP. The results show an evident affinity of anacetrapib towards the concave surface of CETP, and especially towards the region of the N-terminal tunnel opening. The primary binding site of anacetrapib turns out to reside in the tunnel inside CETP, near the residues surrounding the N-terminal opening. Free energy calculations show that when anacetrapib resides in this area, it hinders the ability of cholesteryl ester to diffuse out from CETP. The simulations further bring out the ability of anacetrapib to regulate the structure-function relationships of phospholipids and helix X, the latter representing the structural region of CETP important to the process of neutral lipid exchange with lipoproteins. Altogether, the simulations propose CETP inhibition to be realized when anacetrapib is transferred into the lipid binding pocket. The novel insight gained in this study has potential use in the development of new molecular agents capable of preventing the progression of cardiovascular diseases.

Organisations: Department of Physics

People: Äijänen, T. (Creator), Koivuniemi, A. (Creator), , Rog, T. (Creator), Vattulainen, I. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

Links: <https://figshare.com/articles/Anacetrapib/1247932>

Dataset

### **Supplementary data for "Texture Descriptors Ensembles Enable Image-Based Classification of Maturation of Human Stem Cell-Derived Retinal Pigmented Epithelium"**

The RPE dataset contains 1862 subwindows from 195 phase contrast images. Each RPE cell culture was imaged with the same settings. The microscope used was a Nikon Eclipse TE200S phase-contrast microscope (Nikon Instruments Europe B.V., Amstelveen, Netherlands) with the 20x objective and Ph1 phase contrast.

Organisations: Department of Electronics and Communications Engineering

People: Nanni, L. (Creator), Paci, M. (Creator), , Skottman, H. (Creator), Juuti-Uusitalo, K. (Creator), Hyttinen, J. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.6084/m9.figshare.2070109

Dataset

### **Supplementary data for "Rapid and accurate detection of urinary pathogens by mobile IMS-based electronic nose: a proof-of-principle study"**

Contains the raw data of the measurements and logbook. UT and UU -codes indicate the individual sample plates; they are individual samples of the bacteria culture described in the header row.

Organisations: Department of Automation Science and Engineering

People: Roine, A. (Creator), Saviak, T. (Creator), , Tuokko, A. (Creator), Aittoniemi, J. (Creator), Vuento, R. (Creator), Lekkala, J. (Creator), Lehtimäki, T. (Creator), Tammela, T. L. (Creator), Oksala, N. K. J. (Creator)

Publisher: Tampere University of Technology

Date made available: 2017

DOI: 10.5061/dryad.v3g17

Dataset

### **Supplementary data for "Temperature-Dependent Model of Multi-step Transcription Initiation in Escherichia coli Based on Live Single-Cell Measurements"**

S1 Appendix. Extended Methods and Materials. <https://doi.org/10.1371/journal.pcbi.1005174.s001> (PDF) S1 Table. Mean, standard deviation (sd), and coefficient of variation (cv) of the transcription intervals for the Plac/ara-1 promoter. The table shows the condition, number of exact intervals (samples), number of right-censored intervals (R-samples), and the estimated mean, sd, and cv of the intervals (considering exact and right-censored ones).

<https://doi.org/10.1371/journal.pcbi.1005174.s002> (CSV) S2 Table. Presence or absence of on/off switching and number of identified steps in Plac/ara-1 promoter. The table shows the condition, the existence or absence (yes/no) of on/off switching, the durations of significant steps (Steps 1 to 3), the second-best fitting model (Alt model), and the difference in BICs ( $\Delta$ BIC) of the second-best and the best fitting models and its lower bound.

<https://doi.org/10.1371/journal.pcbi.1005174.s003> (CSV) S3 Table. Estimated model parameters for Plac/ara-1 promoter.

The table shows the induction condition and the estimated parameter values. Here,  $T$  is the temperature in degrees Celsius. <https://doi.org/10.1371/journal.pcbi.1005174.s004> (CSV) S4 Table. Statistics of the on/off switching for Plac/ara-1 promoter. The table shows the condition, the fraction of time the gene is on (Duty cycle) and the average burst size and interval. <https://doi.org/10.1371/journal.pcbi.1005174.s005> (CSV) S5 Table. Mean, standard deviation (sd), and coefficient of variation (cv) of the transcription intervals for the PtetA promoter. The table shows the temperature condition, number of exact intervals (samples), number of right-censored intervals (R-samples), and the estimated mean, sd, and cv of the intervals.

<https://doi.org/10.1371/journal.pcbi.1005174.s006> (CSV) S6 Table. Presence or absence of on/off switching and number of identified steps in the PtetA promoter. The table shows the temperature condition, the presence or absence of on/off switching (On/off), the durations of the significant steps (Steps 1 to 3), the second-best fitting model (Alt model), the difference in BICs ( $\Delta$ BIC) of the second-best and the best fitting model, and its lower bound (LB).

<https://doi.org/10.1371/journal.pcbi.1005174.s007> (CSV) S7 Table. List of promoters and measurement conditions used in the manuscript. The table shows an arbitrary condition ID, the promoter, concentrations of the inducers (IPTG (mM), Ara



(L-arabinose; %), and aTc (anhydrotetracycline; ng/mL)), and the temperature (°C).  
<https://doi.org/10.1371/journal.pcbi.1005174.s008> (CSV) S8 Table. List of all intervals between the production of consecutive transcripts used in the manuscript. The table shows an arbitrary interval ID, condition ID (a foreign key to), the lower (Interval LB) and upper bounds (Interval UB), and the ID of the interval that precedes in time (Previous Interval ID) for each interval observed in the cells under both promoters in each condition. The order of the rows has no particular meaning. <https://doi.org/10.1371/journal.pcbi.1005174.s009> (CSV) S1 Fig. Mean and standard deviation of transcription intervals generated using a Monte Carlo simulation as a function of temperature for the Plac/ara-1 promoter. The dashed curves represent the means and standard deviations of the best fitting models.  
<https://doi.org/10.1371/journal.pcbi.1005174.s010> (EPS) S2 Fig. Mean and standard deviation (sd) of transcription intervals as a function of the relative change in  $\mu$ ,  $\sigma$ , and  $\tau$ . The horizontal axis corresponds to the relative value in the specified parameter. The dashed curves are the asymptotes discussed in the text. The pivot parameters result in a mean of 2 and sd of  $\sqrt{2}$ . <https://doi.org/10.1371/journal.pcbi.1005174.s011> (EPS) S3 Fig. Lower bound for the coefficient of variation of transcription intervals as a function of the relative durations between pre- and post-commit stages. Blue points represent instances of random models. Green curves are the cv lower bounds for different number of steps n in the post-commit stage: from top to bottom: 1, 2 (solid curve; the model in Eq (1)), 3, 4, and 5. The dotted black line is the absolute lower bound attained with a constant-duration post-commit stage ( $n = \infty$ ).  
<https://doi.org/10.1371/journal.pcbi.1005174.s012> (EPS) S4 Fig. Mean, standard deviation (sd), and coefficient of variation (cv) of the transcription intervals for the Plac/ara-1 promoter under full induction, and the corresponding best fitting model. The data points correspond to those shown in S1 Table. Error bars denote one sd of the estimator uncertainty.  
<https://doi.org/10.1371/journal.pcbi.1005174.s013> (EPS) S5 Fig. Mean, standard deviation (sd), and coefficient of variation (cv) of the transcription intervals for the PtetA promoter under full induction, and the corresponding best fitting model. The data points correspond to those shown in S5 Table. Error bars denote one sd of the estimator uncertainty.  
<https://doi.org/10.1371/journal.pcbi.1005174.s014> (EPS) S6 Fig. Mean, standard deviation (sd), and coefficient of variation (cv) of the transcription intervals when varying each parameter independently around the parameter values found for the PtetA promoter, full induction. The numeric labels for the crosses indicate temperature (°C). The data points are derived from the best fitting model. <https://doi.org/10.1371/journal.pcbi.1005174.s015> (EPS)  
Organisations: Faculty of Biomedical Sciences and Engineering, Research group: Laboratory of Biosystem Dynamics-LBD , Research group: Laboratory of Biosystem Dynamics-LBD  
People: , , , , Sanches Ribeiro, A. (Creator)  
Publisher: Tampere University of Technology  
Date made available: 2016  
DOI: 10.1371/journal.pcbi.1005174#sec012  
Links: <https://figshare.com/articles/1/4128594>  
Dataset

### Digital hydraulic boom with valve faults

Organisations: Automation and Hydraulic Engineering  
People: Linjama, M. (Creator), , Ersfolk, J. (Creator)  
Publisher: Tampere University of Technology  
Date made available: 8 May 2017  
Links: <http://urn.fi/urn:nbn:fi:csc-kata20170508154319793705>  
Dataset

### Mimetic Interfaces: Facial Surface EMG Dataset 2015

Mimetic Interfaces: Facial Surface EMG Dataset 2015 (c) by Ville Rantanen, Mirja Ilves, Antti Vehkaoja, Anton Kontunen, Jani Lylykangas, Eeva Mäkelä, Markus Rautiainen, Veikko Surakka, and Jukka Lekkala Mimetic Interfaces: Facial Surface EMG Dataset 2015 is licensed under a Creative Commons Attribution 4.0 International License. You should have received a copy of the license along with this work. If not, see <http://creativecommons.org/licenses/by/4.0/>. The dataset includes Matlab script and functions for processing the data. These are licenced under the MIT license. Data Description The data consists of facial surface EMG signals from the muscles corrugator supercilii, zygomaticus major, orbicularis oris, orbicularis oculi, and masseter. Fifteen participants (8 females, 7 males) participated in the experiments for data collection. Their age was between 26 and 57 years (mean 40.7, standard deviation 9.6). They all had normal or corrected-to-normal vision and normal hearing by their own report. The experiments were conducted in two phases. In the A phase, the participant performed voluntary smile, lip pucker, and frown movement tasks while EMG signals from zygomaticus major, orbicularis oris, orbicularis oculi, and corrugator supercilii were measured from the right side of the face. B phase included the measurement of zygomaticus major, orbicularis oris, orbicularis oculi, and masseter from the right side of the face while the participant performed smile and pucker movement tasks while chewing a gum. Both phases had resting tasks between the movement tasks. The resting task was a neutral expression in the first phase, but included chewing in the second one. Phases started with a 1-minute-long resting task. Then 10 repetitions of each movement tasks were performed in randomized order. Each task lasted for 6 seconds. The movements were instructed to be performed as naturally as possible for the time that an on-screen instruction was visible. Instructions regarding movement intensities were not given, and the participants were not informed that their eye blinks will be monitored to avoid causing abnormal blinking. The used experimental software was E-Prime [1] stimulation software. The facial surface EMG signals were measured with a NeXus-10 physiological monitoring device by Mind Media BV. The sampling rate was 2048 Hz. Filtering was not set from the measurement software but the hardware is to be expected to have antialiasing filters. The measurements were bipolar using pre-gelled, sintered Ag–AgCl electrodes. A separate grounding electrode was used on

the forehead, and the electrodes were placed according to the guidelines of Fridlund and Cacioppo [2] as shown in the images accompanied with the dataset. Corrugator supercilii was left out from the second phase of the experiments because the measurement device only had 4 channels, and the muscle is not as important in facial pacing as the others are. The experiments were recorded with a digital video camera at HD quality at 25 fps. The recordings are not published due to privacy issues. They were visually inspected to find out the onset and termination of each eye blink. The beginning each movement task was also determined from the video where the instructions shown to the participant were visible. Screenshots of the videos are included to illustrate the experimental setup. Eye blinks were classified to three categories: ones with a small eyelid movement where the pupil wasn't fully covered, ones where the pupil was fully covered and once where the eye lids was fully closed. Some participants performed multipart blinks where the previous one hadn't ended before the second one started. These are annotated separately in the data. Schneider, W., Eschman, A., Zuccolotto, A. (2002). E-Prime User's Guide. Psychology software Tools Inc., Pittsburgh. A. J. Fridlund and J. T. Cacioppo, "Guidelines for human electromyographic research," Psychophysiology, vol. 23, no. 5, pp. 567–589, Sep. 1986. The log files from E-Prime are not published in raw form but they have been read to Matlab and saved as mat-files along with metadata and data from the visual inspection. The log files are named according to the scheme 01A.mat where the number (01-15) is the number of the participant and the letter (A/B) is the experimental phase. The dataset is accompanied with a Matlab script and functions needed to produce the results for a publication titled "A Survey on the Feasibility of Surface EMG in Facial Pacing" that is to be published at IEEE Engineering in Medicine and Biology Society's EMBC '16 conference. The included files are: - Data/\*.mat - Matlab data files - Results/\* - Result figures and table as LaTeX tabular - Screenshots/\*.jpg - Images from the experiments - helper\_functions/\* - Helper functions for Matlab - CHANGELOG.txt - Change log to document possible updates - \*\_LICENSE.txt - License files for data, metadata, and the Matlab scripts (software) - README.txt - This document - Participants.csv - Table with the participants' ages and genders - dataprocessing.m - The Matlab script for the data processing and outputting the results - metadata.mat - Metadata with some variables used in the Matlab script

Organisations: Department of Automation Science and Engineering, Research area: Microsystems, Research area: Measurement Technology and Process Control

People: , Ilves, M. (Contributor), Vehkaoja, A. (Contributor), Kontunen, A. (Contributor), Lylykangas, J. (Contributor), Mäkelä, E. (Contributor), Rautiainen, M. (Contributor), Surakka, V. (Contributor), Leikkala, J. (Contributor)

Publisher: Tampere University of Technology

Date made available: 2016

Links: <https://etsin.avointiede.fi/dataset/urn-nbn-fi-csc-kata20160519233254366793>,

<http://avaa.tdata.fi/openida/dl.jsp?pid=urn:nbn:fi:csc-ida-9x201605192015015521963s>,

<http://uta.fi/sis/tauchi/projects/mimeface.html> and one more link, <http://urn.fi/URN:NBN:fi:tty-201611044685> (show fewer)

Dataset

### **TUT Acoustic scenes 2016, Development dataset**

TUT Acoustic Scenes 2016, development dataset consists of 30-seconds audio segments from 15 acoustic scenes: Bus - traveling by bus in the city (vehicle) Cafe / Restaurant - small cafe/restaurant (indoor) Car - driving or traveling as a passenger, in the city (vehicle) City center (outdoor) Forest path (outdoor) Grocery store - medium size grocery store (indoor) Home (indoor) Lakeside beach (outdoor) Library (indoor) Metro station (indoor) Office - multiple persons, typical work day (indoor) Residential area (outdoor) Train (traveling, vehicle) Tram (traveling, vehicle) Urban park (outdoor) Each acoustic scene has 78 segments totaling 39 minutes of audio.

Organisations: Department of Signal Processing, Research group: Audio research group

People: Mesaros, A. (Creator), Heittola, T. (Creator), Virtanen, T. (Creator), Fagerlund, E. (Data Collector), Hiltunen, A. (Data Collector)

Publisher: Tampere University of Technology

Date made available: 8 Feb 2016

DOI: 10.5281/zenodo.45739

Links: <http://urn.fi/URN:NBN:fi:tty-201611044680>

Dataset

### **TUT Sound events 2016, Development dataset**

TUT Sound events 2016, development dataset consists of 22 audio recordings from two acoustic scenes: Home (indoor), 10 recordings, totaling 36:16 Residential area (outdoor), 12 recordings, totalling 42:00

Organisations: Department of Signal Processing, Research group: Audio research group

People: Mesaros, A. (Creator), Heittola, T. (Creator), Virtanen, T. (Creator), Fagerlund, E. (Data Collector), Hiltunen, A. (Data Collector)

Publisher: Tampere University of Technology

Date made available: 8 Feb 2016

DOI: 10.5281/zenodo.45759

Links: <http://urn.fi/URN:NBN:fi:tty-201611044686>

Dataset

### **WLAN RSS indoor measurements**

Organisations: Research group: Wireless Communications and Positioning, Department of Electronics and Communications Engineering

People: Lohan, E. (Creator), Talvitie, J. (Creator)

Publisher: [http://www.cs.tut.fi/tit/pos/MEASUREMENTS\\_WLAN\\_FOR\\_WEB.zip](http://www.cs.tut.fi/tit/pos/MEASUREMENTS_WLAN_FOR_WEB.zip)

Date made available: 2015

Links: [http://www.cs.tut.fi/tit/pos/MEASUREMENTS\\_WLAN\\_FOR\\_WEB.zip](http://www.cs.tut.fi/tit/pos/MEASUREMENTS_WLAN_FOR_WEB.zip), <http://urn.fi/URN:NBN:fi:tty-201611044687>  
Dataset

### **Pernio failure experiment data**

This page contains measured data from the Perniö embankment failure experiment, described in Lehtonen et al (201X).

Organisations: Research group: Foundation Structures, Department of Civil Engineering

People: Länsivaara, T. (Owner),

Publisher: Tampere University of Technology

Date made available: 2015

Links: <http://urn.fi/urn:nbn:fi:csc-kata20150507094517328502>, <http://urn.fi/URN:NBN:fi:tty-201611044688>

Dataset