Appendix A

Cell cycle

This appendix contains a summary of the key features of the cell cycle that occurs in dividing, non-germline cells in multi-cellular organisms; mainly summarised from material in Effler *et al.* (2006) and Alberts *et al.* (2002), which the reader is encouraged to turn to if they seek more information on the cell cycle. The cell cycle is an important cycle that all dividing cells go through. It consists of four main phases, or states:

1. G_1 phase. This is a **g**rowth phase, in which the cell grows and builds up materials ready for S phase.
2. S phase, in which the new copy of the DNA is **s**ynthesised.
3. G_2, a second **g**rowth phase in which the cell increases in size.
4. **M**itotic phase, during which mitosis occurs. The M phase consists of two distinct parts: mitosis and cytokinesis. In mitosis the two sets of chromosomes separate, ready for cytokinesis, where the cytoplasm divides in two.

These four phases are illustrated in Figure A.1. There is also a G_0 (resting) phase which the cells can enter into post-mitosis and stay there, or move into G_1 if they commit to replicate further. The majority of mammalian cells are in this phase. Alternatively, cells can go straight from the M phase to G_1. The distinction between G_0 and G_1 is not always clear, and sometimes a combined G_0/G_1 phase is referred to. The M phase is described in more detail below, since this is a key area where genetic changes can influence gross cell abnormalities during cancer. The M phase is divided into two parts, mitosis and cytokinesis, and mitosis can be further subdivided into a number of discrete phases:

1. Prophase, in which the chromatids (sets of chromosomes) condense (contract) in a region (of the cell) called the centromere. In most cases of mitosis, the nuclear

NOTE: This figure is included on page 178 of the print copy of the thesis held in the University of Adelaide Library.

Figure A.1. Four main stages of the cell cycle. This figure, from Alberts et al. (2002), shows the four main stages of the cell cycle. These are the two growth phases, G_1 and G_2 in which the cell prepares for the succeeding stages of S (DNA synthesis) and M (mitosis and cytokinesis).

envelope breaks down. An intra-cellular structure known as the spindles form. These govern the separation of the chromatids in the later phases of mitosis.

2. Prometaphase and metaphase, in which the chromosomes congregate at a plane halfway between the two ends. The spindle tapers to these ends. The chromatids attach to this spindle. This plane is called the equatorial plane and is the plane at which the whole cell will divide when nuclear division is complete.

3. Anaphase, in which the two chromatids separate and move to opposite poles, following the line of the spindles.

4. Telophase, where new nuclear envelopes form around the two groups of daughter chromosomes (as they are now called), and the spindles eventually disappear as the cell then divides in two during cytokinesis.

These phases are illustrated in Figure A.2. The cell cycle is a highly regulated process, by a large number of proteins which act to “checkpoint” the cell, and can halt the cell cycle to allow the cell to recover (for example, to repair errors in DNA) or even call for the programmed death (apoptosis) of the cell if the damage is irrecoverable. A number of checkpoints occur. These are:

• One that occurs at the end ofG1 where the cell decides if conditions are favourable to divide. This decision depends on the internal resources of the cell, the external

NOTE: This figure is included on page 179 of the print copy of the thesis held in the University of Adelaide Library.

Figure A.2. Mitotic phase of the cell cycle. This figure, from Alberts et al. (2002), shows the M (mitotic) phase of the cell cycle. This is divided into two main parts: mitosis, and cytokinesis. In mitosos there a number of stages in which the two sets of chromosomes condense and then separate to two poles (ends) of the cell, in preparation for cytokinesis, in which the cell cytoplasm (and set of organelles) divides in two. The cytoskeleton furrows and then cleaves along a line, which eventually forms a bridge between the original parent cell and the new daughter cell. This bridge is then dismantled over a process taking minutes to hours, and the cytoskeletal material is then recycled and used in the next M phase. The interphase period shown is simply the set of other stages: G_1, S, and G_2, in that order

resources available for the next phases, and also signals from other cells on space available and if cell division should be occurring in the location in an organ in which the cell resides.

• Several DNA damage checkpoints, that sense damage:

1. Before the cell enters S phase (during G1).
2. During S phase.
3. After DNA replication (during G2).

• Spindle checkpoints. These detect any problems that have developed during the mitosis part of M phase and before cytokinesis occurs.

Appendix B

SVM mathematical description

B.1 Introduction

This appendix contains an overview of the mathematics behind support vector machines (SVMs). (Binary) SVMs attempt to fit a surface in an n-dimensional space, such that it groups the (training) samples of one class on one side, and the samples of another class on the other side. This then gives a function that can classify samples of an unknown class into one of the two classes, hence why this is called a binary SVM. Multiclass SVMs can be built, typically using multiple binary SVMs, and are not discussed here. One would first test the classifier function using samples of known class that are fed into a trained SVM to see if the SVM would perform with a required level of accuracy.

In order to explain SVMs properly, one must first introduce generalised optimisation problems as these are a key component of SVMs, being used to fit surfaces to separate the samples as best as possible. A simple type of SVM is the C-SVM, where the C is a classification parameter used in the algorithm. There is a newer type of SVM known as the ν-SVM, which is essentially the same as the C-SVM, but with the advantage that the parameter ν has two useful interpretations as both the upper bound on the fraction of training errors and a lower bound of the fraction of support vectors that would be classified correctly.

B.2 Constrained optimisation problems

SVMs are formulated as a particular constrained optimisation problem. In this section an overview of a generalised constrained optimisation problem is given, based on

Boyd and Vendenberghe (2004). A constrained optimisation problem can be written as,

$$\min_x f_0(\boldsymbol{x}), \tag{B.1}$$

subject to,

$$f_i, h_j : \mathbb{R}^n \to \mathbb{R},\ f_i(\boldsymbol{x}) \leq 0 \forall i,\ h_j(\boldsymbol{x}) = 0 \forall j, \tag{B.2}$$

that is, in general, a mix of inequality constraints $f_i(\boldsymbol{x})$, and equality constraints $h_j(\boldsymbol{x})$. The primal Lagrangian of the problem is

$$\mathcal{L}_P(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_i \lambda_i f_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x}),\ \lambda_i \geq 0 \forall i. \tag{B.3}$$

The dual Lagrangian is then

$$\mathcal{L}_D(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_x \mathcal{L}_P(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}). \tag{B.4}$$

Since the dual Lagrangian is the infimum of the primal Lagrangians, it can become $-\infty$ for certain values of its arguments. The parameters $\boldsymbol{\lambda} : \lambda_i \geq 0$ and ν for which $\mathcal{L}_D > -\infty$ are called dual feasible. Further, since the dual Lagrangian is a point-wise infimum of a function, so it is concave, even if the primal is not convex.

If we denote the primal optimal point p^*, then because (using Eq. B.3)

$$\sum_i \lambda_i f_i(\boldsymbol{x}) + \sum_j \nu_j h_j(\boldsymbol{x}) \leq 0, \tag{B.5}$$

for a primal feasible point x^*, then

$$\mathcal{L}_D(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*. \tag{B.6}$$

Thus the dual Lagrangian (and hence dual problem) always provides a lower bound to the primal problem. This optimal lower bound can be found by solving the dual problem,

$$\max_{\boldsymbol{\lambda}, \boldsymbol{\nu}} \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\nu}) \tag{B.7}$$

If we denote the dual optimal point by d^*, the condition $d^* \leq p^*$ always holds, which is weak duality, strong duality being obtained when $d^* = p^*$. Strong duality is useful because if it holds, then the dual problem can be solved (which is often easier) and the answer can be converted to the primal domain since we know the solution must

be optimal. In general, strong duality holds if the primal problem is convex and the equality constraints are linear. . That is if both of the following conditions hold:

$$f_i \text{ is convex for } i = 0, \ldots, l \text{ ,}$$
$$\text{and}$$
$$h_j(\boldsymbol{x}) = A\boldsymbol{x} + b.$$

This is true for both types of SVMs that present in this appendix.

As an aside, the primal problem and dual problem can be written more concisely, by noting that

$$\sup_{\lambda \geq 0, \nu} \inf_{x} \mathcal{L}_P(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} f_0(\boldsymbol{x}), & \boldsymbol{x} \text{ is feasible,} \\ \infty, & \text{otherwise.} \end{cases} \tag{B.8}$$

Then the primal problem can be written as

$$p^* = \inf_{\boldsymbol{x}} \sup_{\lambda \geq 0, \nu} \mathcal{L}_P(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}), \tag{B.9}$$

and the dual problem as

$$d^* = \sup_{\lambda \geq 0, \nu} \inf_{\boldsymbol{x}} \mathcal{L}_P(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}). \tag{B.10}$$

B.3 Goal of SVMs

In SVMs, we wish to find a function $f(\boldsymbol{x}) : \mathbb{R}^n \rightarrow \{-1, 1\}$ such that

$$g(\boldsymbol{x}_i) = y_i = \begin{cases} -1, & x_i \in \mathcal{S}_1, \\ 1, & x_i \in \mathcal{S}_2, \end{cases} \tag{B.11}$$

where $\mathcal{S}_1$ and $\mathcal{S}_2$ are two classes that the samples are from. We would then prefer that this function is reasonably accurate (in terms of low numbers of false positives and false negatives). Depending on the application, it may be that we desire a low number of false negatives (since we might want to detect all cancers) but that we are less concerned about false positives (removing a benign lump only has body image implications). The two types of SVM considered are detailed below.

B.4 C-SVMs

The main algorithm tried in order to train a support vector machine is the C-support vector classification (Cortes and Vapnik 1995, Vapnik 1998). This solves the problem,

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i, \tag{B.12}$$

where $x_i \in \mathbb{R}^n$ subject to

$$\begin{aligned} y_i \left(w^T \phi(x_i) + b \right) &\geq 1 - \xi_i, \\ \xi_i &\geq 0, \\ i &= 1, \ldots, l. \end{aligned}$$

The dual problem is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e, \; 0 \leq \alpha_i \leq C, \; i = 1, \ldots, l, \tag{B.13}$$

subject to

$$y^T \alpha = 0,$$

where e is a vector with 1s in all its entries, $C > 0$ is the upper bound, Q is an $(l \times l)$ positive semidefinite matrix $Q_{ij} = y_i y_j K(x_i, x_j)$ for a kernel $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ (see below for more on kernel functions). Once the system is trained, this results in a particular set of parameters $y_i, \alpha_i, b, \; i = 1, \ldots, l.$. These are then used in a decision function:

$$g(x) = \operatorname{sgn} \left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b \right), \tag{B.14}$$

where x is the test/unknown sample vector $x \notin \{x_1, \ldots, x_i, \ldots, x_l\}$ to be classified. The set $\{x_1, \ldots, x_i, \ldots, x_l\}$ is the set of training samples.

B.5 ν-SVMs

ν-SVMs (Schölkopf *et al.* 2000) uses a different parameter than C, denoted $\nu \in (0, 1]$ which is an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors.

The primal optimisation problem is

$$\min_{w,b,\xi,\rho} \frac{1}{2} w^T w - \nu\rho + \frac{1}{l} \sum_{i=1}^{l} \xi_i, \tag{B.15}$$

subject to

$$\begin{aligned} y_i \left(\boldsymbol{w}^T \boldsymbol{\phi}(\boldsymbol{x}_i) + b \right) &\geq \rho - \xi_i, \\ \xi_i &\geq 0, \\ i &= 1, \ldots, l, \\ \rho &\geq 0 \end{aligned}$$

The dual is:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \boldsymbol{\alpha}^T Q \boldsymbol{\alpha}, \tag{B.16}$$

where Q is a matrix as defined above for C-SVMs. subject to,

$$\begin{aligned} 0 \leq \alpha_i &\leq \frac{1}{l}, \\ i &= 1, \ldots, l, \\ \boldsymbol{e}^T \boldsymbol{\alpha} &\geq \nu, \\ \boldsymbol{y}^T \boldsymbol{\alpha} &= 0. \end{aligned}$$

The decision function is the same as for C-SVMs.

B.6 Kernels

Four kernels can be used in the software package we used, and these are the most common choice of kernels:

1. linear: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$,
2. polynomial: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left(\gamma \boldsymbol{x}_i^T \boldsymbol{x}_j + r \right)^d$, $\gamma > 0$,
3. radial basis function (RBF): $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma \| \boldsymbol{x}_i - \boldsymbol{x}_j \| \right)$, $\gamma > 0$, and
4. sigmoid: $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh\left(\gamma \boldsymbol{x}_i^T \boldsymbol{x}_j + r \right)$.

Note that γ is another parameter used in both models. A range of $(C, \gamma) \in \mathbb{R}^2$ or $(\nu, \gamma) \in \mathbb{R}^2$ values are typically tried for the nonlinear kernels. If a linear kernel is successful, it suggests the data points are linear separable, which means when plotted in $\mathbb{R}^n$, they fall into two distinct clusters, and hence we would also expect clustering algorithms to cluster the data well.

Bibliography

ACCARDO-A., AFFINITO-M., CARROZZI-M., AND BOUQUET-F. (1997). Use of the fractal dimension for the analysis of electroencephalographic time series, *Biological Cybernetics*, **77**, pp. 339–350.

ALBERTS-B., JOHNSON-A., LEWIS-J., RAFF-M., ROBERTS-K., AND WALTER-P. (2002). *Molecular Biology of the Cell, 4th Ed.*, Garland Science.

ALM-R. A., LING., MOIR-D. T., KING-B. L., BROWN-E. D., AND OTHERS. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen helicobacter pylori, *Nature*, **397**, pp. 176–180.

ANASTASSIOU-D. (2000). Frequency-domain analysis of biomolecular sequences, *Bioinformatics*, **16**(12), pp. 1073–1081.

ANASTASSIOU-D. (2001). Genomic signal processing, *IEEE Signal Processing Magazine*, **18**, pp. 8–20.

ANDERER-P., GRUBER-G., KLÖSCH-G., KLIMESCH-W., SALETU-B., AND ZEITLHOFER-J. (2002). Sleep and memory consolidation: the role of electrophysiological neuroimaging, *Somnologie*, **6**(2), pp. 54–62.

ANDERSON-P. W. (1972). More is different, *Science*, **177**(4047), pp. 393–396.

ANH-V., LAU-K., AND YU-Z. (2001). Multifractal characterisation of complete genomes, *Journal of Physics A*, **34**(36), pp. 7127–7139.

ARENDS-J. W. (2000). Molecular interactions in the Vogelstein model of colorectal carcinoma, *Journal of Pathology*, **190**, pp. 412–416.

ARMITAGE-P., AND DOLL-R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis, *British Journal of Cancer*, **8**, pp. 1–12.

ARNEODO-A., BACRY-E., GRAVES-P. V., AND MUZY-J. F. (1995). Characterizing long-range correlation in dna sequences from wavelet analysis, *Physical Review Letters*, **74**, pp. 823–837.

ARNOSTI-D. N., GRAY-S., BAROLO-S., ZHOU-J., AND LEVINE-M. (1996). The gap protein knirps mediates both quenching and direct repression in the *Drosophila* embro, , **15**(14), pp. 3659–3666.

ATTARDI-L. D., AND JACKS-T. (1999). The role of p53 in tumor supression: lessons from mouse models, *Cell Mol. Life. Sci.*, **55**(1), pp. 48–63.

AVNER-P., AND HEARD-E. (2001). X-chromosome inactivation: counting, choice and initiation, *Nature Reviews Genetics*, **2**, pp. 59–67.

BALL-F. (1999). Stochastic and deterministic models for SIS epidemics among a population partitioned into households, *Mathematical Biosciences*, **156**, pp. 41–67.

BALL-F. G. (1996). Threshold behaviour in stochastic epidemics among households, *in* C. C. Heyde., Y. V. Prohorov., R. Pyke., and S. T. Rachev. (eds.), *Athens Conf. Appl. Prob. and Time Series, Vol. 1: Applied Probability (Lecture Notes Statist. 114)*, pp. 253–266.

BALL-F. G. (2001). Stochastic multitype SIR epidemics among a population partitioned into households, *Adv. Appl. Prob.*, **33**, pp. 99–123.

BARABÁSI-A. L., AND OLTVAI-Z. N. (2004). Network biology: understanding the cell's functional organization, *Nature Reviews Genetics*, **5**, pp. 101–113.

BELL-M. V., COWPER-A. E., LEFRANC-M. P., BELL-J. I., AND SCREATON-G. R. (1998). Influence of intron length on alternative splicing of cd44, *Molecular and Cellular Biology*, **18**(10), pp. 5930–5941.

BELOUCHRANI-A., MERAIM-K. A., CARDOSO-J. F., AND MOULINES-E. (1997). A blind source separation technique using second order statistics, *IEEE Trans. on Sig. Proc.*, **45**, pp. 434–444.

BENSON-D. A., MIZRACHI-K. I., LIPMAN-D. J., OSTELL-J., AND WHEELER-D. L. (2003). Genbank, *Nucleic Acids Research*, **31**(1), pp. 23–27.

BERRYMAN-M. J., ALLISON-A., AND ABBOTT-D. (2003a). Statistical techniques for text classification based on word recurrence intervals, *Fluctuations and Noise Letters*, **3**(1), pp. L1–L10.

BERRYMAN-M. J., ALLISON-A., AND ABBOTT-D. (2003b). Stochastic evolution and multifractal classification of prokaryotes, *in* S. M. Bezrukov. (ed.), *Proc. of the SPIE: Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems*, Vol. 5110, pp. 192–200.

BERTRAND-O., BOHORQUEZ-J., AND PERNIER-J. (1994). Time-frequency digital filtering based on an invertable wavelet transform: an application to evoked potentials, *IEEE Transactions on Biomedical Engineering*, **41**(1), pp. 77–88.

BINO-R. J., HALL-R. D., FIEHN-O., KOPKA-J., SAITO-K., DRAPER-J., NIKOLAU-B. J., MENDES-P., TUNALI-R. U., BEALE-M. H., TRETHEWEY-R. N., LANGE-B. M., WURTELE-E. S., AND SUMNER-L. W. (2004). Potential of metabolomics as a functional genomics tool, *Trends in Plant Science*, **9**(9), pp. 1360–1385.

BISHOP-C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press.

BOCCARA-N. (2003). *Modeling Complex Systems*, Springer-Verlag.

BON-E., CASAREGOL-S., BLANDIN-G., LLORENT-B., NEUVÉGLISE-C., AND OTHERS. (2003). Molecular evolution of eukaryotic genomes: hemiascomycetous yeast splieosomal introns, *Nucleic Acids Research*, **31**(4), pp. 1121–1135.

BOYD-D. C., POMBO-A., AND MURPHY-S. (2003). Interaction of proteins with promoter elements the human U2 snRNA genes *in vivo*, *Gene*, **315**, pp. 103–112.

BOYD-S. P., AND VENDENBERGHE-L. (2004). *Convex Optimisation*, Cambridge University Press, New York, NY, USA.

BRODEUR-W. R. R., AND LOWE-S. W. (1999). Clinical implications of p53 mutations, *Cell Mol. Life Sci.*, **55**, pp. 64–75.

BROWN-J. R., DOUADY-C. J., ITALIA-M. J., MARSHALL-W. E., AND STANHOPE-M. J. (2001). Universal trees based on large combined protein sequence data sets, *Nature Genetics*, **28**, pp. 281–285.

BROWN-K. A., PLATT-R., AND BATES-J. H. T. (2002). Automated analysis of paradoxical ribcage motion during sleep in infants, *Pediatric Pulmonology*, **33**, pp. 33–46.

BURGES-C. (1998). A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, **2**, pp. 121–167.

CAHILL-D. P., KINZLER-K. W., VOGELSTEIN-B., AND LENGAUER-C. (1999). Genetic instability and darwinian selection in tumours, *Trends Cell Biology*, **9**, pp. M57–M60.

CAPORALE-L. H. (2003). *Darwin in the Genome: Molecular Strategies in Biological Evolution*, McGraw-Hill.

CARDOSO-J. F. (1999). High-order contrasts for independent component analysis, *Neural Computation*, **11**, pp. 157–192.

CARDOSO-J. F., AND SOULOUMIAC-A. (1993). An efficient technique for blind separation of complex source, *Proc. IEEE Sig. Proc. Workshop on Higher-Order Stat.*, pp. 275–279.

CASPARI-T. (2000). Checkpoints: how to activate p53, *Curr. Bio.*, **10**(8), pp. R315–R317.

CHRISTIANINI-N., AND TAYLOR-S. J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press.

CLAUSET-A., AND MOORE-C. (2003). How do networks become navigable?, *ArXiv cond-mat/0309415*.

CLERC-P., AND AVNER-P. (1998). Role of the region 3′ to *xist* exon 6 in the counting process of X-chromosome inactivation, *Nature Genetics*, **19**(3), pp. 249–253.

CONOVER-W. J. (1971). *Practical nonparametric statistics*, Wiley.

CORTES-C., AND VAPNIK-V. N. (1995). Support-vector networks, *Machine Learning*, **20**, pp. 273–297.

COWARD-E. (1997). Equivalence of two Fourier methods for biological sequences, *Journal of Mathematical Biology*, **36**, pp. 64–70.

DAS-A., AND DAS-P. (2004). Nonlinearity in EEG: investigation by surrogate data analysis, *Complexity International*.

DASGUPTA-J., SEN-U., AND DATTAGUPTA-J. K. (2003). *In silico* mutations and molecular dynamics studies on a winged bean chymotrypsin inhibitor protein, *Protein Engineering*, **16**, pp. 489–496.

DA SILVA-A. C. R., FERRO-J. A., REINACH-F. C., FARAH-C. S., AND FURLAN-L. R. (2002). Comparison of the genomes of two xanthomonas pathogens with differing host specificities, *Nature*, **417**, pp. 459–463.

DAWKINS-R. (1989). *The Selfish Gene*, Oxford University Press.

DECKERT-G., WARREN-P. V., GAASTERLAND-T., YOUNG-W. G., LENOX-A. L., AND OTHERS. (1998). The complete genome of the hyperthermophilic bacterium aquifex aeolicus, *Nature*, **392**, pp. 353–385.

DEHNERT-M., HELM-W. E., AND HÜTT-M. T. (2005). Information theory reveals large-scale synchronisation of statistical correlations in eukaryote genomes, *Gene*, **345**, pp. 81–90.

DE JONG-H., GEISELMANN-J., BATT-G., HERNANDEZ-C., AND PAGE-M. (2004a). Qualitative simulation of the initiation of sporulation in *bacillus subtilis*, *Bulletin of Mathematical Biology*, **66**, pp. 261–299.

DE JONG-H., GOUZÉ-J. L., HERNANDEZ-C., PAGE-M., SARI-T., AND GEISELMANN-J. (2004b). Quantitative simulation of genetic regulatory networks using piecewise-linear models, *Bulletin of Mathematical Biology*, **66**, pp. 301–340.

DELBRIDGE-A. (2005). *Macquarie Dictionary*, 4th edn, Macquarie Library.

DELOUKAS-P., MATTHEWS-L. H., ASHURST-J., BURTON-J., GILBERT-J. G. R., AND OTHERS. (2001). The dna sequence and comparative analysis of human chromosome 20, *Nature*, **414**, pp. 865–871.

DEMEREC-M., ADELBERG-E. A., CLARK-A. J., AND HARTMAN-P. E. (1966). A proposal for a uniform nomenclature in bacterial genetics, *Genetics*, **54**, pp. 61–76.

DEPINHO-R. A. (2000). The age of cancer, *Nature*, **408**, pp. 248–254.

DOVER-G. (2000). *Dear Mr Darwin: Letters on the Evolution of Life and Human Nature*, Weidenfeld & Nicolson.

DRINNAN-M. J., MURRAY-A., WHITE-J. E. S., SMITHSON-A. J., GIBSON-G. J., AND GRIFFITHS-C. J. (1996). Evaluation of activity-based techniques to identify transient arousal in respiratory sleep disorders, *Journal of Sleep Research*, **5**, pp. 173–180.

DRUMMOND-S. P. A., BROWN-G. G., GILLIN-J. C., STRICKER-J. L., WONG-E. C., AND BUXTON-R. B. (2000). Altered brain response to verbal learning following sleep deprivation, *Nature*, **403**, pp. 655–657.

DURBIN-R., EDDY-S., KROGH-A., AND MITCHISON-G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*, Cambridge University Press.

EARN-D. J. D., ROHANI-P., BOLKER-B. M., AND GRENFELL. (2000). A simple model for complex dynamical transitions in epidemics, *Science*, **287**, pp. 667–670.

EDWARDS-R. (2000). Analysis of continuous-time switching networks, *Physica D*, **146**, pp. 165–199.

EDWARDS-R., AND GLASS-L. (2000). Combinatorial explosion in model gene networks, *Chaos*, **10**, pp. 691–704.

EFFLER-J. C., IGLESIAS-P. A., AND ROBINSON-D. N. (2006). *Control of Uncertain Systems: Modelling, Approximation and Design*, Lecture Notes in Control and Information Science, Springer-Verlag, chapter Regulating cell shape during cytokinesis, pp. 203–224.

ELBERT-T., RAY-W. J., KOWALIK-Z. J., SKINNER-J. E., GRAF-K. E., AND BIRBAUMER-N. (1994). Chaos and physiology: deterministic chaos in excitable cell assemblies, *Physiological Reviews*, **74**(1), pp. 1–47.

ELDON-E. D., AND PIRROTTA-V. (1991). Interactions of the *Drosophila* gap gene *giant* with maternal and zygotic pattern-forming genes, *Development*, **111**, pp. 367–378.

ELLEDGE-S. J. (1996). Cell cycle checkpoints: preventing an identity crisis, *Science*, **274**, pp. 1664–1672.

ELLIOTT-W. H., AND ELLIOTT-D. C. (1997). *Biochemistry and Molecular Biology*, 2nd edn, Oxford University Press.

FELLER-W. (1945). The fundamental limit theorems in probability, *Bull. Amer. Math. Soc.*, **51**, pp. 800–832.

FELL-J., ELFADIL-H., KLAVER-P., RÖSCHKE-J., ELGER-C. E., AND FERNÁNDEZ-G. (2002). Covariation of spectral and nonlinear eeg measures with alpha biofeedback., *Int J Neurosci*, **112**(9), pp. 1047–1057.

FERRI-R., CHIARAMONTI-R., ELIA-M., MUSUMECI-S., RAGAZZONI-A., AND STAM-C. J. (2003). Non-linear EEG analysis during sleep in premature and full-term newborns, *Clinical Neurophysiology*, **114**, pp. 1176–1180.

FERRI-R., PARRINO-L., SMERIERI-A., TERZANO-M. G., ELIA-M., MUSUMECI-A., PETTINATO-S., AND STAM-C. J. (2002). Non-linear EEG measures during sleep: effects of the different sleep stages and cyclic alternating pattern, *International Journal of Psychophysiology*, **43**, pp. 273–286.

FICCA-G., AND SALZARULO-P. (2004). What in sleep is for memory, *Sleep Medicine*, **5**, pp. 225–230.

FILIPPOV-A. F. (1988). *Differential equations with discontinuous righthand sides*, Kluwer Academic Publishers.

FITCH-J. P., AND SOKHANSANJ-B. (2000). Genomic engineering: moving beyond DNA sequence to function, *Proc. IEEE*, **88**(12), pp. 1949–1971.

FOLKMAN-J. (1990). What is the evidence that tumors are angiogenesis dependent?, *Journal of the National Cancer Institute*, **82**, pp. 4–6.

FRANK-J. (1999). Applying memetics to financial markets: do markets evolve towards efficiency?, *Journal of Memetics—Evolutionary Models of Information Transmission*, **3**, pp. 87–99.

FRANKLIN-R. E., AND GOSLING-R. G. (1953). Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate, *Nature*, **172**, pp. 156–157.

FRASCH-M., AND LEVINE-M. (1987). Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in *Drosophila*, *Genes and development*, **1**, pp. 981–995.

FRIBERG-S., AND MATTSON-S. (1997). On the growth rates of human malignant tumors: Implications for medical decision making, *Journal of Surgical Oncology*, **65**, pp. 284–297.

FRONHÖFER-H. G., AND VOLHARD-N. C. (1987). Maternal genes required for the anterior localization of *bicoid* activity in the embryo of *Drosophila* embryo, *Genes and Development*, **1**, pp. 880–890.

FUTREAL-P. A., COIN-L., MARSHALL-M., DOWN-T., HUBBARD-T., WOOSTER-R., RAHMAN-N., AND STRATTON-M. R. (2004). A census of human cancer genes, *Nature Reviews Cancer*, **4**, pp. 177–183.

GAIMAN-N. (2005). untitled, Private communication.

GALVÁN-B. P., GROSSE-I., CARPENA-P., OLIVER-J. L., ROLDÁN-R. R., AND STANLEY-H. E. (2000). Finding borders between coding and noncoding dna regions by an entropic segmentation method, *Physical Review Letters*, **85**(6), pp. 1342–1345.

GHOSHAL-G., AND SANDER-L. M. (2004). SIS epidemics with household structure: the self-consistent field method, *Mathematical Biosciences*, **190**, pp. 71–85.

GLASS-L. (1975). Classification of biological networks by their qualitative dynamics, *Journal of Theoretical Biology*, **54**, pp. 85–107.

GOLD-D., AND SOKOLOWSKI-D. (2004). p53 protein-DNA complex.

GOLDMANN-D. A. (2001). Respiratory infections in health-care institution, *Emerging Infectious Diseases*, **7**(2), pp. 249–253.

GORODNITSKY-I. F., AND BELOUCHRANI-A. (2001). Joint cumulant and correlation based signal separation with application to EEG data analysis, *ICA 2001*.

GOUKASSIAN-D., GAD-F., YAAR-M., ELLER-M. S., NEHAL-U. S., AND GILCHREST-B. A. (2000). Mechanisms and implications of the age associated decrease in DNA repair capacity, *The Faseb Journal*, **14**, pp. 1325–1334.

GREGORY-S. G., SEKHON-M., SCHEIN., ZHAO-S., OSOEGAWA-K., AND OTHERS. (2002). A physical map of the mouse genome, *Nature*, **418**, pp. 743–750.

GRIFFTHS-A. J. F., MILLER-J. H., SUZUKI., LEWONTIN-R. C., AND GELBART-W. M. (1996). *An Introduction to Genetic Analysis*, 6th edn, W. H. Freeman & Company.

GUERMUER-Y. (2002). Combining discriminant models with new multi-class SVMs, *Pattern Analysis & Applications*, **5**, pp. 168–179.

GWALTNEY-J. M. (2000). *The Common Cold*, 5th ed. edn, Churchill Livingstone, pp. 651–656.

GWALTNEY-J. M., HENDLEY-J., SIMON-G., AND JORDON-W. S. J. (1967). Rhinovirus infections in an industrial population. II. Characteristics of illness and antibody response, *JAMA*, **202**, pp. 494–500.

HANAHAN-D., AND WEINBERG-R. A. (2000). The hallmarks of cancer, *Cell*, **100**, pp. 57–70.

HARMER-G. P., ABBOTT-D., TAYLOR-P. G., AND PARRONDO-J. M. R. (2001). Brownian ratchets and Parrondo's games, *Chaos*, **11**(3), pp. 705–714.

HARRIS-J. M., AND GWALTNEY-J. M. (1996). Incubation periods of experimental rhinovirus infection and illness, *Clinical Infectious Diseases*, **23**, pp. 1287–1290.

HASTY-J., MCMILLEN-D., AND COLLINS-J. J. (2002). Engineered gene circuits, *Nature*, **420**, pp. 224–230.

HAYASHI-T., MAKINO-K., OHNISHI-M., KUROKAWA-K., ISHII-K., YOKOYAMA-K., AND OTHERS. (2001). Complete genome sequence of enterohemorrhagic *escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12, *DNA Research*, **8**, pp. 11–22.

HEIKKINEN-T., AND JÄRVINEN-A. (2003). The common cold, *Lancet*, **361**, pp. 51–59.

HIGUCHI-T. (1988). Approach to an irregular time series on the basis of the fractal theory, *Physica D*, **31**, pp. 277–283.

HOCH-M., GERWIN-N., TAUBERT-H., AND JÄCKLE-H. (1992). Competition for overlapping sites in the regulatory region of the *Drosophila* gene *Krüppel*, *Science*, **256**, pp. 94–97.

HOCH-M., SEIFERT-E., AND JÄCKLE-H. (1991). Gene expression mediated by *cis*-acting sequences of the Krüppel gene in response to the *Drosophila* morphogens *bicoid* and *hunchback*, *The EMBO Journal*, **10**(8), pp. 2267–2278.

HOLLAND-J. H. (1995). *Hidden Order: How Adaptation Builds Complexity*, Perseus Books.

HOLLAND-J. H. (1998). *Emergence: From Chaos to Order*, Oxford University Press.

HOLLOWAY-D. M., HARRISON-L. G., AND SPIROV-A. V. (2003). Noise in the segmentation gene network of *Drosophila*, with implications for mechanisms of body axis specification, *in* S. M. Bezrukov., H. Fraunfelder., and F. Moss. (eds.), *Proc. SPIE: Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems*, Vol. 5110, pp. 180–191.

HOUCHMANDZADEH-B., WIESCHAUS-E., AND LEIBLER-S. (2002). Establishment of developmental precision and proportions in the early *Drosophila* embryo, *Nature*, **415**, pp. 798–802.

HUBER-R., GHILARDI-M. F., MASSIMINI-M., AND TOMONI-G. (2004). Local sleep and learning, *Nature*, **430**, pp. 78–81.

HWA-R. C., AND FERREE-T. C. (2001). Fluctuation analysis of human electroencephalogram, *arXiv physics/0105029*.

INTERNATIONAL. (2001). Initial sequencing and analysis of the human genome, *Nature*, **409**, pp. 860–921.

IRISH-V., LEHMANN-R., AND AKAM-M. (1985). The *Drosophila* posterior-group gene *nanos* functions by repressing *hunchback* activity, *Nature*, **338**, pp. 646–648.

JACKSON-A. L., AND LOEB-L. A. (1998). The mutation rate and cancer, *Genetics*, **148**, pp. 1483–1490.

JAIN-A. K., AND DUBES-R. C. (1990). *Algorithms for Clustering Data*, Prentice Hall.

JANUS-F., ALBRECHTSEN-N., DORNREITER-I., WIESMULLER-L., GROSSE-F., AND DEPPERT-W. (1999). The dual role model for p53 in maintaining genomic integrity, *Cell Mol. Life Sci.*, **55**(1), pp. 12–27.

JONES-J. H., AND HANDCOCK-M. S. (2003). Sexual contacts and epidemic thresholds, *Nature*, **423**, pp. 605–606.

JONSSON-P., GULLBERG-J., NORDSTRÖM-A., AND OTHERS. (2004). A strategy for identifying differences in large series of metabolomic samples analyzed by gc/ms, *Anal. Chem.*, **76**, pp. 1738–1745.

JOSET-F., AND MICHEL-G. J. (1993). *Prokaryotic Genetics: Genome Organisation, Transfer and Plasticity*, Blackwell Scientific Publications.

KAEMING-K. L., PASVOGEL-A. E., GOODWIN-J. L., MULVANEY-S. A., MARTINEZ-F., ENRIGHT-P. L., ROSEN-G. M., MORGAN-W. J., FREGOSI-R. F., AND QUAN-S. F. (2003). Learning in children and sleep disordered breathing: findings of the Tucson children's assessment of sleep apnea (TuCASA) prospective cohort study, *Journal of the International Neuropsychological Society*, **9**, pp. 1016–1026.

KÆRN-M., BLAKE-W. J., AND COLLINS-J. J. (2003). The engineering of gene regulatory networks, *Ann. Rev. Biomed. Eng.*, **5**, pp. 179–206.

KANTZ-H., AND SCHREIBER-T. (1997). *Nonlinear Time Series Analysis*, Cambridge University Press.

KANUNGO-T., MOUNT-D. M., NETANYAHU-N. S., PIATKO-C. D., SILVERMAN-R., AND WU-A. Y. (2002). An efficient *k*-means clustering algorithm: analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

KARCHIN-R., DIEKHANS-M., LIBUSHA-K., THOMAS-D. J., PIEPER-U., ESWAR-N., HAUSSLER-D., AND SALI-A. (2005). Ls-snp: large-scale annotation of coding non-synonymous snps based on multiple information sources, *Bioinformatics*, **21**, pp. 2814–2820.

KAUFFMAN-S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press.

KAUFMAN-L., AND ROUSSEEUW-P. J. (1990). *Finding Groups in Data: An Introduction*, John Wiley and Sons Inc.

KEELING-M. J., AND RAND-D. A. (2001). *Spatial Correlations and Local Fluctuations in Host Parasite Systems*, Springer, pp. 5–57.

KHOLODENKO-B. N., KIYATKIN-A., BRUGGEMAN-F. J., SONTAG-E., WESTERHOFF-H. V., AND HOEK-J. B. (2002). Untangling the wires: A strategy to trace functional interactions in signaling and gene networks, *Proc. Natl. Acad. Sci. U.S.A.*, **99**(20), pp. 12841–12846.

KNUDSON-A. G. (2002). Cancer genetics, *American Journal of Medical Genetics*, **111**, pp. 96–102.

KOBAYASHI-H., KÆRN-M., ARAKI-M., CHUNG-K., GARDNER-T. S., CANTOR-C. R., AND COLLINS-J. J. (2004). Programmable cells: Interfacing natural and engineered gene networks, *Proc. Natl. Acad. Sci. U.S.A.*, **101**(22), pp. 8414–8419.

KOPLIK-J., REDNER-S., AND WILKINSON-D. (1988). Transport and dispersion in random networks with percolation disorder, *Physical Review A*, **37**(7), pp. 2619+.

KOSCIELNY-S., TUBIANA-M., AND VALLERON-A. J. (1985). A simulation model of the natural history of human breast cancer, *Br. J. Cancer*, **52**, pp. 515–524.

KRASKOV-A., STÖGBAUER-H., AND GRASSBERGER-P. (2003). Estimating mutual information, *arXiv cond-mat/0305641*.

KRISHNAN-S., AND RANGAYYAN-R. M. (2000). Automatic denoising of knee-joint vibration signals using adaptive time-frequency representations, *Medical and Biological Engineering and Computation*, **38**(1), pp. 2–8.

KULLBACK-S. (1968). *Information Theory and Statistics*, Dover Publications.

KURATA-H., INOUE-T., SUMIDA-Y., TANAKA-S., AND OHASHI-T. (2001). Simulation and system analysis of gene regulatory networks using the two-phase partition method, *Genomes Informatics*, **12**, pp. 286–287.

KURODA-M., OHTA-T., UCHIYAMA-I., BABA-T., YUZAWA-H., AND OTHERS. (2001). Whole genome sequencing of meticillin-resistant staphylococcus aureus, *Lancet*, **357**(9264), pp. 1225–1240.

LAKIN-N. D., AND JACKSON-S. P. (1999). Regulation of p53 in response to DNA damage, *Oncogene*, **18**(54), pp. 7644–7655.

LEHMANN-R., AND VOLHARD-N. C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of *oskar*, a maternal gene in *Drosophila*., *Cell*, **47**, pp. 141–152.

LEHMANN-R., AND VOLHARD-N. C. (1991). The maternal gene *nanos* has a central role in posterior pattern formation of the *Drosophila* embryo, *Development*, **112**, pp. 679–693.

LEVEILLE-J. (2002). *Epidemic spreading in technological networks*, Master's thesis, School of Cognitive and Computing Sciences, University of Sussex, UK.

LEVINE-M., AND TJIAN-R. (2003). Transcription regulation and animal diversity, *Nature*, **424**, pp. 147–151.

LILJEROS-F. (2004). Sexual networks in contemporary western societies, *Physics A*, **338**, pp. 238–245.

LIM-L. M., AKAY-M., AND DAUBENSPECK-J. A. (1995). Identifying respiratory-related evoked potentials, *IEEE Engineering in Medicine and Biology*, **14**(2), pp. 174–178.

LINDSLEY-D. L., AND ZIMM-G. G. (1992). *The Genome of Drosophila melanogaster*, Academic Press Inc.

LI-W. (1990). Mutual information functions versus correlation functions, *Journal of Statistical Physics*, **60**, pp. 823–837.

LI-W. (1992). Generating nontrivial long-range correlations and 1/f spectra by replication and mutation, *International Journal of Bifurcation and Chaos*, **2**(1), pp. 137–154.

LI-W., AND KANEKO-K. (1992). Long-range correlation and partial $1/f^{\alpha}$ spectrum in a noncoding DNA sequence, *Europhysics Letters*, **17**, pp. 655–660.

LOEB-L. A. (1991). Mutator phenotype may be required for multistage carcinogenesis, *Cancer Research*, **51**, pp. 3075–3079.

LOWE-S. W., AND LIN-A. W. (2000). Apoptosis in cancer, *Carcinogenesis*, **21**(3), pp. 485–495.

LUEBECK-E. G., AND MOOLGAVKAR-S. H. (2002). Multistage carcinogenesis and the incidence of colorectal cancer, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, pp. 15095–15100.

LUKASHIN-A. V., AND BORODOVSKY-M. (1998). GeneMark.hmm: new solutions for gene finding, *Nucleic Acids Research*, **26**(4), pp. 1107–1115.

LU-X., SUN-Z., CHEN-H., AND LI-Y. (1998). Characterizing self-similarity in bacteria DNA sequences, *Physical Review E*, **58**, pp. 3578–3584.

MALLAT-S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7), pp. 674–693.

MALLAT-S. G. (1999). *A Wavelet Tour of Signal Processing*, Academic Press.

MALLAT-S. G., AND ZHANG-Z. (1993). Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing*, **41**(12), pp. 3397–3415.

MANSEAU-L. J., AND SCHÜPBACH-T. (1989). *cappucino* and *spire*: two unique maternal-effect loci required for both the anteroposterior and dorsoventral patterns of the *Drosophila embryo*, *Nature*, **329**, pp. 167–170.

MARCUS-C. L., OMLIN-K. J., BASINKI-D. J., BAILEY-S. L., RACHAL-A. B., VON PECHMANN-W. S., KEENS-T. G., AND WARD-S. L. (1992). Normal polysomnographic values for children and adolescents, *Am. Rev. Respir. Dir.*, **146**, pp. 1235–1239.

MATALGAH-M., AND KNOPP-J. (1994). Improved noise stability multiresolution filters based on a class of wavelets, *Proceedings of the 37th Midwest Symposium on Circuits and Systems: Vol 2.*, IEEE, pp. 817–820.

MATTICK-J. S. (2001). Non-coding RNAs: the architects of eukaryotic complexity, *EMBO Reports*, **2**(11), pp. 986–991.

MENON-V. J., AND AGRAWAL-D. C. (2003). Physics of quiet and suffocative breathing, *American Journal of Physics*, **71**, pp. 474–478.

MEZA-A. B., YU-Y., SCHÜTTLER-H. B., ARNOLD-J., AND TAHA-T. R. (2004). KINSOLVER: A simulator for compting large ensembles of biochemical and gene regulatory networks, *http://lsdis.cs.uga.edu/∼aleman/kinsolver/*.

MONTELONE-B. A. (1998). Mutation, mutagens & DNA repair, *http://www-personal.ksu.edu/∼bethmont/mutdes.html*.

MOREAU-Y., DE SMET-F., THIJS-G., MARCHAL-K., AND DE MOOR-B. (2002). Functional bioinformatics of microarray data: from expression to regulation, *Proceedings of the IEEE*, **90**(11), pp. 1722–1743.

MOUSE. (2002). Initial sequencing and comparative analysis of the mouse genome, *Nature*, **420**, pp. 520–562.

MURRAY-J. D. (2002). *Mathematical Biology*, 3rd edn, Springer.

MUZUR-A., SCHOTT-P. E. F., AND HOBSON-J. A. (2002). The prefrontal cortex in sleep, *Trends in Cognitive Sciences*, **6**, pp. 475–481.

NAKADE-K. (2004). The p53 network, *http://www.brc.riken.jp/lab/dna/en/GENESETBANK/201p53_network.html*.

NATURE. (2002). Wag the dogma, *Nature Genetics*, **30**(4), pp. 343–344.

NELSON-K. E., CLAYTON-R. A., GILL-S. R., GWINN-M. L., DODSON-R. J., AND OTHERS. (1999). Evidence for lateral gene transfer between archaea and bacteria from genome sequence of thermotoga maritima, *Nature*, **399**, pp. 323–329.

NIJHOUT-H. F. (2002). The nature of robustness in development, *BioEssays*, **24**(6), pp. 553–563.

NIJHOUT-H. F. (2003). The importance of context in genetics, *American Scientist*, **91**(5), pp. 416–423.

NOWAK-M. A., MICHOR-F., KOMAROVA-N. L., AND IWASA-Y. (2004). Evolutionary dynamics of tumor suppressor gene inactivation, *Proceedings of the National Academy of Sciences*, **101**(29), pp. 10635–10638.

NOWELL-P. C. (1976). The clonal evolution of tumor cell populations, *Science*, **194**, pp. 23–28.

OZBUDAK-E. M., THATTAI-M., KURSTER-I., GROSSMAN-A. D., AND VAN OUDENAARDEN-A. (2002). Regulation of noise in the expression of a single gene, *Nature Genetics*, **31**, pp. 69–73.

PANKRATZ-M. J., SEIFERT-E., GERWIN-N., BILLI-B., NAUBER-U., AND JÄCKLE-H. (1990). Gradients of *Krüppel* and *knirps* gene products direct pair-rule gene stripe patterning in the posterior region of the *Drosophila* embryo, *Cell*, **61**, pp. 309–317.

PARKHILL-J., WREN-B. W., MUNGALL-K., KETLEY., AND CHURCHER-C. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences, *Nature*, **403**, pp. 665–668.

PENG-C. K., BULDYREV-S. V., GOLDBERGER-A. L., HAVLIN-S., SCIORTINO-F., SIMON-M., AND STANLEY-H. E. (1992). Long-range correlations in nucleotide sequences, *Nature*, **356**, pp. 168–170.

PINTO-A. L., TEIXEIRA-S. M. R., PENA-S. D. J., AND MACHADO-C. R. (2003). Single-nucleotide polymorphims of the *Trypanosoma cruzi MSH2* gene support the existence of three phylogenetic lineages presenting differences in mismatch-repair efficiency, *Genetics*, **164**, pp. 117–126.

PRESS-W. H., FLANNER-B. P., TEUKOLSKY-S. A., AND VETTERLING-W. T. (1986). *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press.

PRISK-G. K., HAMMER-J., AND NEWTH-C. J. L. (2002). Techniques for measurement of thoracoabdominal asynchrony, *Pediatric Pulmonology*, **34**, pp. 462–472.

PUROHIT-P. V., AND ROCKE-D. M. (2003). Discriminant models for high-throughput proteomics mass spectrometer data, *Proteomics*, **3**, pp. 1699–1703.

RAJAGOPALAN-H., NOWAK-M. A., VOGELSTEIN-B., AND LENGAUER-C. (2003). The significance of unstable chromosomes in colorectal cancer, *Nature Reviews Cancer*, **3**, pp. 695–701.

RAND-D. A. (1999). Correlation equations and pair approximations for spatial ecologies, *CWI Quarterly*, **12**(3-4), pp. 329–368.

RAY-R. P., AND SCHÜPBACH-T. (1996). Intercellular signaling and the polarization of body axes during *Drosophila* oogenesis, *Genes and Development*, **10**, pp. 1711–1723.

RECHTSCHAFFEN-A., AND KALES-A. (1968). *A Manual of Standardized Terminology, Techniques, and Scoring Systems for Sleep Stages of Human Subjects*, Brain Information/Brain Research Institute.

RENAN-M. J. (1993). How many mutations are required for tumorigenesis? implications from human cancer data, *Molecular Carcinogenesis*, **7**, pp. 139–146.

RÉNYI. (1960). On measures of information and entropy, *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561.

REW-D. A., AND WILSON-G. D. (2000). Cell production rates in human tissues and tumours and their significance. Part II: clinical data., *European Journal of Surgical Oncology*, **26**, pp. 405–417.

ROTHE-M., WIMMER-E. A., PANKRATZ-M. J., GAITAN-G. M., AND JÄCKLE-H. (1994). Identical transacting factor requirement for knirps and knirps-related gene expression in the anterior but not the posterior region of the *Drosophila* embryo, *Mechanisms of Development*, **46**(3), pp. 169–181.

RUBINSZTEIN-D. C., AMOS-W., LEGGO-J., GOODBURN-S., RAMESAR-R. S., OLD-J., BONTROP-R., MCMAHON-R., BARTON-D. E., AND SMITH-F. M. A. (1994). Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalance, *Nature Genetics*, **7**, pp. 525–530.

SADOVSKY-M. G. (2006). Information capacity of nucleotide sequences and its applications, *Bulletin of Mathematical Biology*, **68**, pp. 785–806.

SAITOU-N., AND NEI-M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution*, **4**, pp. 406–425.

SAUER-F., AND JÄCKLE-H. (1995). Heterodimeric *Drosophila* gap gene protein complexes acting as transcriptional repressors, *The EMBO Journal*, **14**, pp. 4773–4780.

SCHÖLKOPF-B., SMOLA-A., WILLIAMSON-R. C., AND BARTLETT-P. L. (2000). New support vector algorithms, *Neural Computation*, **12**, pp. 1207–1245.

SCHREIBER-T., AND SCHMITZ-A. (1996). Improved surrogate data for nonlinearity tests, *Physical Review Letters*, **77**(4), pp. 635–638.

SCHREIBER-T., AND SCHMITZ-A. (1997). Discrimination power of measures for nonlinearity in a time series, *Physical Review E*, **55**(5), pp. 5443–5447.

SFORZA-E., GRANDIN-S., JOUNY-C., ROCHAT-T., AND IBANEZ-V. (2002). Is waking electroencephalographic activity a predictor of daytime sleepiness in sleep-related breathing disorders?, *European Respiratory Journal*, **19**(4), pp. 645–652.

SHANNON-C. E. (1993). Claude elwood shannon: Collected papers.

SHERR-C. J. (1998). Tumor surveillance via the ARF-p53 pathway, *Genes and Development*, **12**(19), pp. 2984–2991.

SHOEMAKER-D. D., SCHADT-E. E., ARMOUR-C. D., HE-Y. D., ENGELE-G. P., MCDONAGH-P. D., LOERCH-P. M., AND OTHERS. (2001). Experimental annotation of the human genome using microarray technology, *Nature*, **409**, pp. 922–927.

SIEBER-O. M., HEINIMANN-K., AND TOMLINSON-I. P. M. (2003). Genomic instability – the engine of tumorigenesis?, *Nature Reviews Cancer*, **3**, pp. 701–708.

SILVERMAN-B. D., AND LINSKER-R. (1986). A measure of DNA periodicity, *Journal of Theoretical Biology*, **118**, pp. 295–300.

SMALL-S., BLAIR-A., AND LEVINE-M. (1992). Regulation of even-skipped stripe 2 in the *Drosophila* embryo, *The EMBO Journal*, **11**(11), pp. 4047–4057.

SNEL-B., BORK-P., AND HUYNEN-M. A. (1999). Genome phylogeny based on gene content, *Nature Genetics*, **21**, pp. 108–110.

SOCIETY-A. T. (1996). Standards and indications for cardiopulmonary sleep studies in children, *Am. J. Respir. Crit. Care Med.*, **153**, pp. 866–878.

SPEER-J. F., PETROSKY-V. E., RETSKY-M. W., AND WARDEWLL-R. H. (1984). A stochastic numerical model of breast cancer growth that simulates clinical data, *Cancer Res.*, **44**, pp. 4124–4130.

SPENCER-S. L., BERRYMAN-M. J., GARCÍA-J. A., AND ABBOTT-D. (2004). An ordinary differential equation model for the multistep transformation to cancer, *Journal of Theoretical Biology*, **231**, pp. 515–524.

SPENCER-S. L. L., GERETY-R. A. A., PIENTA-K. J. J., AND FORREST-S. (2006). Modeling somatic evolution in tumorigenesis., *PLoS Comput Biol*.

STEPHENSON-E. C., CHAO-Y., AND FACKENTHAL-J. D. (1988). Molecular analysis of the *swallow* gene of *Drosophila melanogaster*, *Genes and Development*, **2**, pp. 1655–1665.

STEPIEŃ-R. A. (2002). Testing for non-linearity in EEG signal of healthy subjects, *Acta Neurobiological Experiments*, **62**, pp. 277–281.

ST JOHNSTON-D., BEUCHLE-D., AND VOLHARD-N. C. (1991). *staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg, *Cell*.

ST. JOHNSTON-D., DRIEVER-W., BERLETH-T., RICHSTEIN-S., AND VOLHARD-N. C. (1989). Multiple steps in the localization of *bicoid* RNA to the anterior pole of the *Drosophila* oocyte, *Development*, **107**, pp. 13–19.

STUDIER-J. A., AND KEPPLER-K. J. (1988). A note on the neighbour-joining algorithm of saitou and nei, *Molecular Biology and Evolution*, **5**, pp. 729–731.

SUSKO-E., INAGAKI-Y., FIELD-C., HOLDER-M. E., AND ROGER-A. J. (2002). Testing for difference in rates-across-sites distribution in phylogenetic subtrees, *Molecular Biology and Evolution*, **19**, pp. 1514–1523.

TAVARÉ-S., AND GIDDINGS-B. W. (1989). *Mathematical Methods for DNA Sequences*, CRC Press, pp. 117–131.

TEPLAN-M. (2002). Fundamentals of EEG measurement, *Measurement Science Review*, **2**, pp. 1–11.

THATTAI-M., AND VAN OUDENAARDEN-A. (2001). Intrinsic noise in gene regulatory networks, *Proc. Natl. Acad. Sci. U.S.A.*, **98**(15), pp. 8614–8619.

TOMLINSON-I., AND BODMER-W. (1999). Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog, *Nature Medicine*, **5**, pp. 11–12.

TOMLINSON-I. P. M., AND BODMER-W. F. (1995). Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models, *Proc. Natl. Acad. Sci. U.S.A.*, **92**, pp. 11130–11134.

TOMLINSON-I. P. M., NOVELLI-M. R., AND BODMER-W. F. (1996). The mutation rate and cancer, *Proc. Natl. Acad. Sci. U.S.A.*, **93**, pp. 14800–14803.

TROTTER-H. F. (1959). An elementary proof of the central limit theorem, *Arch. Math.*, **10**, pp. 226–234.

TURING-A. M. (1952). The chemical basis of morphogenesis, *Philosophical Transactions of the Royal Society of London*, **B 237**, pp. 37–72.

TYSON-J. T., CHEN-K. C., AND NOVAK-B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell, *Current Opinion in Cell Biology*, **15**, pp. 221–231.

VANDEMOORTELE., AND DELAMONICA-E. (2000). The "education" vaccine against HIV, *Current Issues in Comparative Education*, **3**(1), pp. 1–6.

VAPNIK-V. N. (1998). *Statistical Learning Theory*, Wiley.

VOGELSTEIN-B., LANE-D., AND LEVINE-A. J. (2000). Surfing the p53 network, *Nature*, **408**, pp. 307–310.

WAIN-H. M., BRUFORD-E. A., LOVERING-R. C., LUSH-M. J., WRIGHT-M. W., AND POVEY-S. (2002). Guidelines for human gene nomenclature, *Genomics*, **79**, pp. 464–470.

WANG-C., AND LEHMANN-R. (1991). Nanos is the localized posterior determinant in *Drosophila*, *Cell*, **66**, pp. 637–647.

WATSON-J. D., AND CRICK-F. H. C. (1953). Molecular structure of nucleic acids. a structure for deoxyribose nucleic acid., *Nature*, **171**, pp. 737–738.

WIGNER-E. (1960). The unreasonable effectiveness of mathematics in the natural sciences, *Communications in Pure and Applied Mathematics*, **13**, pp. 1–14.

WINTER-P. (1987). Steiner problem in networks: a survey, *Networks*, **17**(2), pp. 129–167.

WOLPERT-W., BEDDINGTON-R., BROCKES-J., JESSELL-T., LAWRENCE-P., AND MEYEROWITZ-E. (1998). *Principles of Development*, Oxford University Press.

WOOD-R. D., MITCHEL-M., SGOUROS-J., AND LINDAHL-T. (2001). Human DNA repair genes, *Science*, **291**(5507), pp. 1284–1289.

WORNING-P., JENSEN-L. J., NELSON-K. E., BRUNAK-S., AND USSERY-D. W. (2000). Structural analysis of DNA sequence: evidence for lateral gene transfer in *Thermotoga maritima*, *Nucleic Acids Res.*, **28**(3), pp. 706–709.

WU-X., VASISHT-V., KOSMAN-D., REINITZ-J., AND SMALL-S. (2001). Thoracic patterning by the *Drosophila* gap gene *hunchback*, *Developmental Biology*, **237**, pp. 79–92.

YEUNG-K. Y., AND RUZZO-W. L. (2001). Principal component analysis for clustering gene expression data, *Bioinformatics*, **17**, pp. 764–774.

YUAN-J., YEASKY-T. M., HAVRE-P. A., AND GLAZER-P. M. (1995). Induction of p53 in mouse cells decreases mutagenesis by UV radiation, *Carcinogenesis*, **16**, pp. 2295–2300.

YU-Z., ANH-V., AND LAU-K. (2001). Measure representation and multifractal analysis of complete genomes, *Physical Review E*.

YU-Z. G., AND ANH-V. (2004). Phylogenetic tree of prokaryotes based on complete genomes using fractal and correlation analyses, *Proceedings of the second conference on Asia-Pacific bioinformatics*, Vol. 29 of *ACM International Conference Proceeding Series*, pp. 321–326.

ZHANG-X. Y., CHEN-F., ZHANG-Y. T., AGNER-S. C., AKAY-M., LU-Z. H., WAYE-M. M. Y., AND TSUI. (2002). Signal processing techniques in genomic engineering, *Proc. IEEE*, **90**(12), pp. 1822–1833.

ZHOU-B. B., AND ELLEDGE-S. J. (2000). The DNA damage response: putting checkpoints in perspective, *Nature*, **408**, pp. 433–439.

Biography

Matthew Berryman received a B.Sc. in Mathematical and Computer Sciences and a B.E. (Hons.) in Computer Systems from The University of Adelaide, Australia. In his Ph.D. candidature at The University of Adelaide at the Centre for Biomedical Engineering, he worked on several areas in complex systems including signal processing of DNA and analysis of EEG signals. In 2003 he won a Santa Fe Institute CSSS scholarship and was also a visiting scholar working with Cosma Shalizi at the Center for the Study of Complex Systems at the University of Michigan, Ann Arbor. In 2004 he returned to spend another two and a half months working with Cosma, and in 2005 spent two months working with Neil Johnson at Oxford University. He was a research assistant at the University of South Australia at the Centre for Pharmaceutical Research, applying machine learning techniques to biomarkers and drug discovery. He was the treasurer of the Electrical and Electronic Engineering Society of Adelaide University (EEESAU), an IEEE affiliated student group. Prizes won include the Seimens Prize for Innovation (South Australian winner) in 2003, and the University of Adelaide Research Abroad Scholarship and Society for Mathematical Biology Landahl Scholarship in 2004.

Disclaimer

NOTE: This cartoon strip is included on page 203 of the print copy of the thesis held in the University of Adelaide Library.

Figure B.1. Disclaimer. Used with permission from"Piled Higher and Deeper" by Jorge Cham
www.phdcomics.com