SHORT COMMUNICATION

Do Cells use Passwords? Do they Encrypt Information?

Alex Root [1,*]

[1] Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

AR: Alex Root 0000-0003-1206-7010

* Correspondence: Alex Root

Molecular Biology Program

Memorial Sloan Kettering Cancer Center

New York, NY, USA

Tel 646-286-3612

Email roota@mskcc.org

Abstract

Living organisms must maintain proper regulation including defense and healing. Life-threatening problems may be caused by pathogens or an organism's own cells' deficiency or hyperactivity, in cancer or auto-immunity. Life evolved solutions to these problems that can be conceptualized through the lens of information security, which is a well-developed field in computer science. Here I argue that taking an information security view of cell biology is not merely semantics, but useful to explain features of cell signaling and regulation. It also offers a conduit for cross-fertilization of advanced ideas from computer science, and the potential for biology to inform computer science. First, I consider whether cells use passwords, i.e., precise initiation sequences that are required for subsequent signals to have any effect, by analyzing chromatin regulation and cellular reprogramming. Second, I consider whether cells use the more advanced security feature of encryption. Encryption could benefit cells by making it more difficult for pathogens to hijack cell networks. Because the 'language' of cell signaling is unknown, i.e., similar to an alien language detected by SETI, I use information theory to consider the general case of how non-randomness filters can be used to recognize (1) that a data stream encodes a language, rather than noise, and (2) quantitative criteria for whether an unknown language is encrypted. This leads to the result that an unknown language is encrypted if efforts at decryption produce sharp decreases in entropy and increases in mutual information. A fully decrypted language should have minimum entropy and maximum mutual information. The magnitude of which should scale with language complexity. I demonstrate this with a simple numerical experiment on English language text encrypted with a basic polyalphabetic cipher. I conclude with unanswered questions for future research.

Keywords: cell signaling; reprogramming; host-pathogen interactions; information security; biochemical passwords; biochemical encryption; exobiology; SETI

Main

Cell signaling and regulatory networks transmit, receive, and process information resulting in decision making concerning growth, defense, differentiation, migration, apoptosis, metabolism, and other processes[1,2]. Groundbreaking studies over the last several decades have elucidated properties of cellular biochemical signaling and regulatory networks, including scale-free, robustness, fragility, noise-filtering, bistability, controllability, ultrasensitivity, signal dissipation,

amplification, memory, modularity, feedfoward and other motifs, which are reviewed by Krakauer and colleagues[2], Uda and Kuroda[3], Mousavian and colleagues[4,5], Walterman and Klipp[6], Azeloglu and Iyengar[1], and Antebi and colleagues[7]. Cell networks can become dysfunctional through somatic mutation, chemical injury, infection, or other processes, that achieve varying degrees of control over the network[8]. Here, I begin to consider these processes through the lens of information security, which as far as I can determine is not common. This is notable for its stark contrast to human telecommunications, where cybersecurity is of paramount importance[9]. In an elegant and trenchant examination of theoretical biology, Krakauer and colleagues argue "before we can look for patterns, we often need to know what kinds of patterns to look for, which requires some fragments of theory to begin with[10]." Therefore, I propose fragments of theory for information security in cells for the community to begin to hunt for patterns and test predictions.

By explicitly incorporating information security concepts into thinking about biological systems, several outcomes are possible in general: (1) distinctions without differences: rephrasing familiar concepts of immunity and regulation in terms of information security adds no value; (2) cross-disciplinary fertilization occurs as information security concepts are imported into biological theory; (3) new information security knowledge arises from examination of biological systems. Recent studies on network controllability provide one framework for examining information security in biochemical networks[11-15]. In this essay, a different perspective is taken to analyze whether cells use passwords and encrypt information.

Immune Systems and biological security

The evolution of immune systems and self-defense against injury and mutation are major innovations in the history of life on earth[16-18]. By total volume, life on earth has its largest habitat in the deep ocean with an abundance of bacteriophages, suggesting that evolution leads to a proliferation of simple life forms, with consciousness as a kind of statistical accident[19]. Single-celled and multi-cellular organisms evolved a wide-variety of defense systems, often dichotomized into innate and adaptive systems[20]. These systems can be conceptualized more

generally to include protective mechanisms against both external and internal damage. The connection between external and internal injury is seen in the study of viruses, which led to insights in cancer biology and the discovery of oncogenes[17]. Organisms developed the ability to recognize self from non-self and destroy xenobiotic material. However, not all foreign genetic material is completely destroyed, because it can increase fitness, e.g., antibiotic resistance plasmids[20,21]. On the intracellular level, bacterial defense mechanisms include blocking receptor binding (surface modification), genome injection (superinfection exclusion), viral replication (restriction modification, CRISPR-Cas, and prokaryotic Argonaute), and abortive infection (programmed cell death)[21]. Similar mechanisms exist in eukaryotic cells, including, RIG-like receptor proteins that recognize RNA[16], xenophagy[22], advanced intracellular nucleic acid recognition systems and other cell-autonomous mechanisms[23]. In plants, sophisticated DICERs defend against retroviruses[24]. Similarly, pathogens use a variety of mechanisms to co-opt, hijack, and counteract host defenses[25-28]. Mutations leading to oncogenes reprogram signaling networks[29]. All of these attacks and counter-attacks involve changes in signaling and regulatory networks, and therefore, changes in information.

Information security in computer science

Information security has been critically important for millennia, with the Caesar substitution cipher being a prominent early example[30]. (The cipher works by shifting each letter of the English alphabet by 3, i.e., A->C, B->D,...,X->A[30]) Computer viruses achieved notoriety in 1987 when the Brain, Lehigh, and April Fool viruses came to worldwide attention[31]. Hackers achieved infamy and also contributed to the advancement of information technology[32]. Information security depends on the use of passwords for system access and encryption to alter information so that its meaning is obfuscated[33]. Development of secure encryption systems, e.g., the RSA asymmetric public key cryptography, was an essential innovation in the history of the internet[33] and must constantly evolve to meet new threats[9]. Steganography is an altogether different approach that conceals the existence of information, e.g., writing with invisible ink, and appears to have had played less importance in the history of information technology than cryptography[33]. Attacks on encrypted

systems can involve interception, modification, fabrication, or interruption of information[33]. There has been considerable work in adapting biomolecules for use in information security in human telecommunications using biosteganography[34] where information is invisible and molecular cryptography, where synthetic biology is used to re-engineer molecules to decode and encode information[35]. Despite obvious parallels in the world of computers, less explicit attention appears to have been paid to theoretical descriptions of cells in terms of their native information security systems, prompting me to ask: Do cells use passwords? Do they encrypt information?

Information systems in cells

Individual cells have a variety of sophisticated information systems. They encode information through the genetic code, which utilizes double-stranded complementary base pairing to provide built-in error correction, which is a type of backup or repair security system. At the proteome level, cells can greatly expand on the genetic code with a few hundred different post-translational modifications in various combinations, that give rise to numerous proteoforms[36], which form components of signaling and regulatory networks. Somatic recombination in immunoglobulins and T-cell receptors can vastly increase protein variants in certain cell types[37]. Interactions of these macromolecules form networks that store and transmit information[6]. There is a context specificity to many signaling pathways, including TGF-beta and AKT, which means that cells respond differently to pathway activation depending on the cell type[38,39]. Many intracellular signaling pathways do not match one receptor to a single ligand, but instead use multiple receptors and ligands that interact combinatorially[40], or use combinations of numerous nuclear-receptor cofactors to regulate activity[41]. Therefore, genetic, epigenetic, transcriptomic and proteomic variation gives rise to a large repertoire of interacting components. These mechanisms are present in complex multicellular organisms, where advanced regulation is needed to control differentiation[42] and also in bacteria for quorum sensing[2].

Cancer has been shown to involve rewiring cellular networks by oncogenes and therefore, in some sense, these represent alterations in information transmission and compromised

security[29,43]. Cells can be reprogrammed through microRNAs and gene regulatory networks in cancer to oncogenic states with distinct metabolism[44]. Similarly, viruses can substantially rewire signaling and regulatory networks to hijack cellular machinery for viral benefit[45]. In the early days of cancer research, similarities between the two systems caused the scientific community to think that viruses cause cancer, and studies into viral biology provided insights into cancer[17,46]. Both pathogenic and pathological processes involve hijacking cellular networks.

In multicellular organisms, combinations of histone modifications give rise to varying chromosomal accessibility and epigenetic states, which are read, written, and erased by chromatin modifiers[47,48]. This epigenetic regulation is capable of encoding memory at the single-cell level[49]. Redundancy and correlation among epigenetic marks, transcription factors, and co-regulators provides a system of information compression to specific cell state[50]. For example, ligand identity can be encoded as pulsatile (DLL1-Notch1) or sustained (DLL4-Notch1) to induce opposite cell fates. In the adult human body, several hundred distinct cell types exist in "cell states", some of which can be dynamically reprogrammed from one state to the next using sophisticated perturbations[51-53]. The language used to describe these cellular properties (code, encode, read, write, memory, erase, reprogram, compression, rewire) points to their aspects as information systems.

Do cells use passwords?

Password authorization systems allow access based upon entry of a correct code out of many possible entries. They can be viewed conceptually as an initiation sequence of signals without which the system will not respond to subsequent signals. Typically, passwords function as a logical AND operation, i.e., each character must be entered correctly to allow system access. However, a logical AND gate is not strictly required. For example, a bouncer at a nightclub may listen for the password "more cheese" but accept partial matches, such as "more these" or "Moishe's". I consider whether there is an evidence for the existence of passwords, i.e., an

initiation sequence of signals without which the system will not respond to subsequent signals using the example of transcription factor-chromatin accessibility.

Organization of chromatin into highly compact, inaccessible regions, and open, accessible regions appears on its face to be a form of cellular information security because some genes are "locked" and therefore, cannot be transcribed. Chromatin is frequently characterized as being in "open" and "closed" states that must be unlocked for cell differentiation by pioneer transcription factors. This appears to be a potential case where cells use passwords. There are multiple algorithms to predict combinations of transcription factors to reprogram human cells from one type to another with the number of successful conversions being relatively low, reviewed by Kamaraj and colleagues[51]. The systems work by engineering overexpression of transcription factors, rather than as it happens normally in development through extracellular signaling molecules that signal to transcription factors to achieve the rewiring. To our knowledge, no one has attempted to predict upstream combinations of signals, e.g., growth factors, hormones, adhesion contacts, etc. that would trigger the right combinations of transcription factors. Sampattavanich and colleagues demonstrated that FOXO3 dynamics can code for different growth factors and their concentrations, which are under combinatorial control of ERK and AKT pathways[54]. One simple way to conceptualize this is that it takes the right combination of transcription factors to unlock the epigenetic code to transdifferentiate cells, i.e., it might require an initiation sequence and therefore, a password. This is distinct from simply requiring a series of events. If the reprogramming transcription factors are active during the entire reprogramming process, then they are not performing an initiation sequence and therefore, not entering a password. Similarly, if only one member of the combination can partially reprogram cells then it would seem inappropriate to conceptualize the mechanism as a password. I predict that password-length, i.e., the complexity of the reprogramming initiation is directly proportional the fitness cost posed to the organism from the conversion. For example, because stem cells have greater replicative potential, they might pose greater risk to develop into cancer and consequently, require a more complex password for reprogramming. Detailed time course

measurements are necessary to resolve whether there is a distinct initiation sequence during cell reprogramming.

Do cells encrypt information?

If cell signaling networks use encryption, how might we know? Put another way, if we do not know the underlying language, i.e., the *unencrypted information*, how can we recognize *encrypted information*? To explore this question, several concepts from information theory are useful. The Shannon entropy is defined as[55]:

$$H = E[I(X)] = -\sum_{x \ni X} p(x)\log_2(x) \quad \textbf{(1)}$$

where *H* is the entropy in bits, defined as the expected information of a distribution of random variables *X*. The entropy can be thought of as how predictable the next character in a transmitted message is. A message that is purely random characters and therefore, not meaningful language, will have the highest entropy[55]. Considering only the 26 letters in the English alphabet, the maximum entropy is $\log_2(26)=4.7$ bits. Shannon analyzed words of size *N* up to 8 letters and found the entropy of the English language to be roughly 2.3 bits per letter, a 50% reduction over random[56]. The English alphabet could eliminate the letter *c* with either *k* or *s* without any meaningful effects. Moreover, English text can be re-coded and stored in smaller file sizes without loss of information (lossless compression) using sophisticated algorithms[55]. Entropy provides a limit on lossless compression[55].

A related concept to entropy is Zipf's law, which states that a word's probability is inversely proportional to its rank and has been found in English language phrases, and also other fields, e.g., city sizes, firm sizes, and neural activity[57].

$$Frequency \propto \frac{1}{Rank} \quad \textbf{(2)}$$

A large number of explanations has been proposed for why Zipf's law exists, which are reviewed by Piantadosi[58]. Purely random texts do not follow Zipf's law[59]. Salge and colleagues found that Zipf's law emerges through minimization of communication inefficiency and direct signal cost[60].

Williams and colleagues found that Zipf's law held more generally for phrases in English than words, which is intriguing because phrases are "the most coherent units of meaning in language[61]."

Language has additional structure that can be captured through analysis of pairwise and higher-order interactions[62]. One measure of association is mutual information[6]. It can be defined between two sets of variables *X* and *Y*, e.g., adjacent letters in the English alphabet as

$$MI(X,Y) = H(X) + H(Y) - H(X,Y), \quad \textbf{(3)}$$

where *H(X,Y)* is the joint entropy between the *X* and *Y*, which is defined as

$$H(X,Y) = -\sum_{x \ni X} \sum_{y \ni Y} p(x,y) \log_2 p(x,y). \quad \textbf{(4)}$$

When *X* and *Y* are statistically dependent, the joint entropy *H(X,Y)* is lowest and the mutual information is maximized.

Doyle and colleagues describe the search for extraterrestrial intelligence (SETI) as fundamentally applying Zipf's law and higher-order information-entropic filters to received sources of electromagnetic radiation[63]. Cell signaling and gene expression have been shown to pass both of these non-randomness filters[6,64]. These non-random filters can also be applied to any sort of data stream to check if it is non-random.

If a simple substitution cipher is applied to an unknown language, the frequency distributions of letters, words, and phrases do not change, and therefore, given enough text would be recognizable as language, although perhaps untranslatable. For a more complex cipher, e.g., a polyalphabetic cipher, the entropy will increase and frequency distributions will deviate from Zipf's law. In other words, if SETI receives a long stream of an alien communication that is encrypted by relatively simple methods, its non-randomness filters should recognize it as a language. If the alien language is encrypted with a polyalphabetic cipher, which was subsequently decrypted, the plaintext would have lower but non-trivial entropy.

A quantitative test for whether a text is encrypted is whether there is a decryption, such that:

$$\underset{d \ni D}{\operatorname{argmax}}\left(MI(E), -H(E)\right) \quad \textbf{(3)}$$

Where *d* is a decryption out the set of all possible decryptions *D, E* is the decrypted plaintext, and *MI* is the mutual information in the decrypted plaintext, e.g., the mutual information in adjacent letters, and *H* is the entropy of the decrypted plaintext, e.g., per letter. In other words, a signal stream is encrypted if a decryption can be found, such that the entropy is minimized and the mutual information is maximized.

To demonstrate this, I provide a simple numerical example. The text of Jane Austen's novel Pride and Prejudice was downloaded from the Gutenberg project[65], processed and cleaned of special characters in the R programming language using the textclean package[66], and encrypted with a simple polyalphabetic substitution cipher of 0,+1,+2. Figure 1A shows the frequency distributions of adjacent letters in the plaintext. Figure1B shows how the frequency distributions of adjacent letters in the encrypted text result in an increase in entropy. The Entropy R package was used to compute entropy per letter and mutual information for adjacent letters[67]. Figure 1C shows how applying varying levels of decryption using several different methods results in changing entropy per letter and mutual information of adjacent letters. As the text is decrypted more completely, the entropy per letter decreases and the mutual information per pair of adjacent letters increases. Complete decryption produces a maximum of this mutual information and a minimum of entropy. Therefore, we can begin to look for patterns that may involve encryption in very rich data of cell signaling by applying this quantitative criterion.

Conclusions and open questions

Evolutionary potential is vast and a complex interplay among environmental change, ecosystems, speciation, niche diversification, extinctions, and innovation have shaped life on earth[68,69]. Considering how rapidly passwords and encryption evolved in human telecommunications, it is natural to ask whether they are used in nature by cells. This theoretical exploration suggests that

cells may use passwords to lock-in cell state, which must be unlocked through the right combination of transcription factors. Open questions include if cells use passwords to initiate cell signaling cascades, programmed cell death, neuron-to-neuron transmissions, or other areas. Also, is it the case that password-length, i.e., the complexity of an initiation sequence is directly proportional the fitness cost? When there is selection pressure due to co-evolution of pathogens are there more complex initiation sequences, i.e., harder-to-crack passwords? Do these have greater complexity in their molecular mechanisms? Another open question is do pathogens launch attacks similar to those seen in computer science, e.g., denial-of-service attacks?

While I have not presented direct evidence for cell passwords, encryption, or other security measures, I suggest that they may exist and provide fragments of theory and criteria that the community can use to look for patterns that may demonstrate their existence. This framework does not address how to encode the information present in cell signaling, nor which decryption strategies to try. I also have not addressed the critical question of noise in biological systems and measurements, which add considerable complexity to information theoretic analysis of biological systems. If encryption does exist, it would seem to point towards both greater complexity because of the existence of encoders/decoders, but perhaps also greater simplicity, because if a message is encrypted, it may become intelligible once decrypted. If there is evidence for encryption, identifying the molecular mechanisms by which it occurs could yield new and powerful insights into signaling, pathogens, and pathologies.

Figures

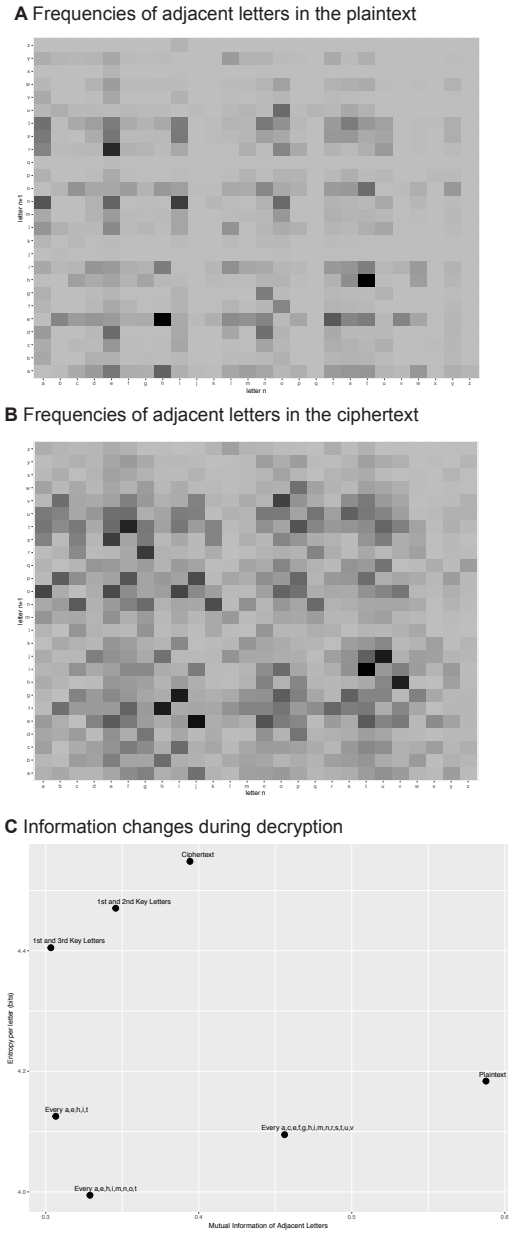

Figure 1. How entropy and mutual information change during decryption of a complex text *Pride and Prejudice*. **A** Frequencies of adjacent letters in the un-encrypted text (plaintext). **B** Frequencies of adjacent letters in the encrypted (ciphertext). **C** Changes in entropy and mutual information during decryption of Pride and Prejudice. The entropy is maximal in the ciphertext and decreases during decryption. The mutual information follows a more complex trajectory before becoming maximal at complete decryption.

Declarations

Author contributions

A.R. wrote the paper.

Acknowledgments

The author thanks H. Alexander Ebhardt for critical comments and discussion.

Consent for publication

No humans were actively recruited to the work presented here.

Competing interests

The author declares that he has no competing interests.

Ethics approval and consent to participate

No humans were actively recruited to the work presented here.

Funding

U.S. National Cancer Institute (NCI) P30 Cancer Center Support Grant (CCSG) P30 CA008748 to A.R. The funding covers general support for the research center.

Publisher's Note

Springer remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Azeloglu, E.U. & Iyengar, R. Signaling networks: information flow, computation, and decision making. *Cold Spring Harb Perspect Biol* **7**, a005934 (2015).
2. Krakauer, D.C., Muller, L., Prohaska, S.J. & Stadler, P.F. Design specifications for cellular regulation. *Theory Biosci* **135**, 231-240 (2016).
3. Uda, S. & Kuroda, S. Analysis of cellular signal transduction from an information theoretic approach. *Semin Cell Dev Biol* **51**, 24-31 (2016).
4. Mousavian, Z., Kavousi, K. & Masoudi-Nejad, A. Information theory in systems biology. Part I: Gene regulatory and metabolic networks. *Semin Cell Dev Biol* **51**, 3-13 (2016).
5. Mousavian, Z., Diaz, J. & Masoudi-Nejad, A. Information theory in systems biology. Part II: protein-protein interaction and signaling networks. *Semin Cell Dev Biol* **51**, 14-23 (2016).
6. Waltermann, C. & Klipp, E. Information theory based approaches to cellular signaling. *Biochim Biophys Acta* **1810**, 924-32 (2011).
7. Antebi, Y.E., Nandagopal, N. & Elowitz, M.B. An operational view of intercellular signaling pathways. *Curr Opin Syst Biol* **1**, 16-24 (2017).
8. Vidal, M., Cusick, M.E. & Barabasi, A.L. Interactome Networks and Human Disease. *Cell* **144**, 986-998 (2011).
9. Bernstein, D.J. & Lange, T. Post-quantum cryptography. *Nature* **549**, 188-194 (2017).
10. Krakauer, D.C. *et al.* The challenges and scope of theoretical biology. *J Theor Biol* **276**, 269-76 (2011).
11. Liu, Y.Y., Slotine, J.J. & Barabasi, A.L. Controllability of complex networks. *Nature* **473**, 167-73 (2011).
12. Gates, A.J. & Rocha, L.M. Control of complex networks requires both structure and dynamics. *Sci Rep* **6**, 24456 (2016).
13. Uhart, M., Flores, G. & Bustos, D.M. Controllability of protein-protein interaction phosphorylation-based networks: Participation of the hub 14-3-3 protein family. *Sci Rep* **6**, 26234 (2016).
14. Li, A., Hu, Q., Liu, J. & Pan, Y. Resistance and Security Index of Networks: Structural Information Perspective of Network Security. *Sci Rep* **6**, 26810 (2016).
15. Albert, R., Jeong, H. & Barabasi, A.L. Error and attack tolerance of complex networks. *Nature* **406**, 378-82 (2000).
16. Mukherjee, K., Korithoski, B. & Kolaczkowski, B. Ancient origins of vertebrate-specific innate antiviral immunity. *Mol Biol Evol* **31**, 140-53 (2014).
17. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
18. Ameisen, J.C. On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death Differ* **9**, 367-93 (2002).
19. Lara, E. *et al.* Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci Adv* **3**, e1602565 (2017).
20. Sirisinha, S. Evolutionary insights into the origin of innate and adaptive immune systems: different shades of grey. *Asian Pac J Allergy Immunol* **32**, 3-15 (2014).
21. van Houte, S., Buckling, A. & Westra, E.R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol Mol Biol Rev* **80**, 745-63 (2016).
22. Miller, C. & Celli, J. Avoidance and Subversion of Eukaryotic Homeostatic Autophagy Mechanisms by Bacterial Pathogens. *J Mol Biol* **428**, 3387-98 (2016).
23. Wu, J. & Chen, Z.J. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* **32**, 461-88 (2014).
24. Ebhardt, H.A., Thi, E.P., Wang, M.B. & Unrau, P.J. Extensive 3' modification of plant small RNAs is modulated by helper component-proteinase expression. *Proc Natl Acad Sci U S A* **102**, 13398-403 (2005).
25. Alto, N.M. & Orth, K. Subversion of cell signaling by pathogens. *Cold Spring Harb Perspect Biol* **4**, a006114 (2012).

26. Moffatt, J.H., Newton, P. & Newton, H.J. Coxiella burnetii: turning hostility into a home. *Cell Microbiol* **17**, 621-31 (2015).
27. Alcami, A. Viral mimicry of cytokines, chemokines and their receptors. *Nat Rev Immunol* **3**, 36-50 (2003).
28. Agol, V.I. & Gmyl, A.P. Viral security proteins: counteracting host defences. *Nat Rev Microbiol* **8**, 867-78 (2010).
29. Stuhlmiller, T.J., Earp, H.S. & Johnson, G.L. Adaptive reprogramming of the breast cancer kinome. *Clin Pharmacol Ther* **95**, 413-5 (2014).
30. Gardner, M. *Codes, Ciphers, and Secret Writing*, (Dover Publications, Inc., New York, NY, USA, 1972).
31. Highland, H.J. History of Computer Viruses. *Computers & Security* **16**, 416-429 (1997).
32. Levy, S. *Hackers*, (O'Reilly Media, Sebastopol, 2010).
33. Tipton, H.F. & Krause, M. *Information Security Management Handbook*, 3280 (CRC Press, Boca Raton, FL, 2007).
34. Brunet, T.D. Aims and methods of biosteganography. *J Biotechnol* **226**, 56-64 (2016).
35. Lustgarten, O., Motiei, L. & Margulies, D. User Authorization at the Molecular Scale. *Chemphyschem* **18**, 1678-1687 (2017).
36. Aebersold, R. *et al.* How many human proteoforms are there? *Nat Chem Biol* **14**, 206-214 (2018).
37. Hood, L., Kronenberg, M. & Hunkapiller, T. T cell antigen receptors and the immunoglobulin supergene family. *Cell* **40**, 225-9 (1985).
38. Feng, X.H. & Derynck, R. Specificity and versatility in tgf-beta signaling through Smads. *Annu Rev Cell Dev Biol* **21**, 659-93 (2005).
39. Toker, A. & Marmiroli, S. Signaling specificity in the Akt pathway in biology and disease. *Adv Biol Regul* **55**, 28-38 (2014).
40. Antebi, Y.E. *et al.* Combinatorial Signal Perception in the BMP Pathway. *Cell* **170**, 1184-1196 e24 (2017).
41. Broekema, M.F. *et al.* Profiling of 3696 Nuclear Receptor-Coregulator Interactions: A Resource for Biological and Clinical Discovery. *Endocrinology* **159**, 2397-2407 (2018).
42. Spickard, E.A., Joshi, P.M. & Rothman, J.H. The multipotency-to-commitment transition in Caenorhabditis elegans-implications for reprogramming from cells to organs. *FEBS Lett* **592**, 838-851 (2018).
43. Drake, J.M. *et al.* Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell* **166**, 1041-1054 (2016).
44. Pinweha, P., Rattanapornsompong, K., Charoensawan, V. & Jitrapakdee, S. MicroRNAs and oncogenic transcriptional regulatory networks controlling metabolic reprogramming in cancers. *Comput Struct Biotechnol J* **14**, 223-33 (2016).
45. McKinney, C. *et al.* Global reprogramming of the cellular translational landscape facilitates cytomegalovirus replication. *Cell Rep* **6**, 9-17 (2014).
46. Koonin, E.V., Dolja, V.V. & Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **479-480**, 2-25 (2015).
47. Voigt, P. *et al.* Asymmetrically modified nucleosomes. *Cell* **151**, 181-93 (2012).
48. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41-5 (2000).
49. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720-4 (2016).
50. Ahsendorf, T., Muller, F.J., Topkar, V., Gunawardena, J. & Eils, R. Transcription factors, coregulators, and epigenetic marks are linearly correlated and highly redundant. *PLoS One* **12**, e0186324 (2017).
51. Kamaraj, U.S., Gough, J., Polo, J.M., Petretto, E. & Rackham, O.J. Computational methods for direct cell conversion. *Cell Cycle* **15**, 3343-3354 (2016).
52. Li, C. & Wang, J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS Comput Biol* **9**, e1003165 (2013).

53. Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci U S A* **108**, 8257-62 (2011).
54. Sampattavanich, S. *et al.* Encoding Growth Factor Identity in the Temporal Dynamics of FOXO3 under the Combinatorial Control of ERK and AKT Kinases. *Cell Syst* (2018).
55. Cover, T.M. & Thomas, J.A. *Elements of information theory*, 748 (John Wiley & Sons, Hoboken, NJ, USA, 2006).
56. Shannon, C.E. Prediction and entropy of printed English. *The Bell System Technical Journal* **January**, 50-64 (1951).
57. Aitchison, L., Corradi, N. & Latham, P.E. Zipf's Law Arises Naturally When There Are Underlying, Unobserved Variables. *PLoS Comput Biol* **12**, e1005110 (2016).
58. Piantadosi, S.T. Zipf's word frequency law in natural language: a critical review and future directions. *Psychon Bull Rev* **21**, 1112-30 (2014).
59. Ferrer, I.C.R. & Elvevag, B. Random texts do not exhibit the real Zipf's law-like rank distribution. *PLoS One* **5**, e9411 (2010).
60. Salge, C., Ay, N., Polani, D. & Prokopenko, M. Zipf's Law: Balancing Signal Usage Cost and Communication Efficiency. *PLoS One* **10**, e0139475 (2015).
61. Williams, J.R. *et al.* Zipf's law holds for phrases, not words. *Sci Rep* **5**, 12209 (2015).
62. Stephens, G.J. & Bialek, W. Statistical mechanics of letters in words. *Phys Rev E Stat Nonlin Soft Matter Phys* **81**, 066119 (2010).
63. Doyle, L.R., McCowan, B., Johston, S. & Hanser, S.F. Information theory, animal communication, and the search for extraterrestrial intelligence. *Acta Astronautica* **68**, 406-417 (2011).
64. Furusawa, C. & Kaneko, K. Zipf's law in gene expression. *Phys Rev Lett* **90**, 088102 (2003).
65. Austen, J. Pride and Prejudice. (Project Gutenberg, https://www.gutenberg.org/files/1342/1342-h/1342-h.htm, 2008).
66. Rinker, T.W. textclean: Text Cleaning Tools. 0.9.2 edn (Buffalo, NY, USA, 2018).
67. Hausser, J. & Strimmer, K. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learn. Res.* **10**, 1469-1484 (2009).
68. Knoll, A.H. & Nowak, M.A. The timetable of evolution. *Sci Adv* **3**, e1603076 (2017).
69. Cazzolla Gatti, R., Fath, B., Hordijk, W., Kauffman, S. & Ulanowicz, R. Niche emergence as an autocatalytic process in the evolution of ecosystems. *J Theor Biol* **454**, 110-117 (2018).