Chapter 1

Introduction

Rod C. Scott and J. Matthew Mahoney

This book takes a view of the brain as a *complex adaptive system* and seeks to identify mechanisms underlying the clinical outcomes as well as the therapeutic opportunities for epilepsy using this framework.

Complex systems theory is a nebulous field whose overarching goal is to understand the dynamical behavior of systems consisting of many interconnected component parts. It has attracted widespread interest from many domains that study examples of such systems, including ecologists, sociologists, engineers, artificial intelligence researchers, condensed matter physicists, neuroscientists, and many others. The results of these collected, multi-disciplinary efforts have not been so much a comprehensive theory of Complex Systems (capital-C, capital-S), but rather a set of techniques, analogies, and attitudes toward problem solving that emphasize interactions and dynamics over individual components and their functions. The chapters are written in a complex adaptive systems frame and therefore it is useful to provide a provisional theoretical description of such systems. Following Holland [1], a generalizable description of complex adaptive systems is that they are collections of relatively simple *agents* that have the property that they can *aggregate*, so that collections of agents can form meta-agents (and meta-meta-agents etc.) with higher-order structure. These aggregates interact *nonlinearly*, so that the aggregate behavior of a collection of agents is qualitatively different from the behavior of the individual agents. The interactions among agents mediate *flows* of materials or information. Finally, the agents are typically *diverse* with distinct specialties that are optimized through adaptation to selective pressures in their environments.

To manifest these properties, complex adaptive systems have mechanisms that underpin the formation and function of the whole system. In full generality, these mechanisms may seem unnecessarily abstract or obscure for application to a specific system, like the neural circuits of the brain. Nevertheless, the abstraction is precisely what accounts for the cross-disciplinarity of complex systems theory, and the applicability of its approaches across biological length scales from subcellular structures to whole brains. The first mechanism is *tagging*, which allows diverse agents in the system to signal their identities to other agents thus enabling complex self-organization into aggregates. The second mechanism is the ability to generate *internal models* that approximate and anticipate the world external to the system, which enables adaptive behavior by the aggregate system.

From the above description, brains are clearly complex adaptive systems par excellence. There are several hierarchical layers of agents. A diversity of genes aggregates into gene networks that form a diversity of proteins that aggregate from a diversity of cells (e.g., neurons and glia) that aggregate and form a diversity of brain regions that aggregate and form the brain with a diversity of emergent phenomena. Indeed, individual cells themselves are complex adaptive systems, where biomolecules as agents interact through electrostatic fields generated by patterns of charges (tags) that facilitate aggregation into complexes and structures. These structures implicitly compute models of the world outside the cell and generate an appropriate transcriptional response. For example, the presence of a phosphorylated signaling molecule inside a cell carries information about the concentration of particular ligands outside the cell. This organization is approximately repeated at the level of neural networks. Neurons as agents use a variety of biochemical and electrical cues (tags) to form into circuits that mediate the flow of sensory information into motor output, memory etc., through massively parallel nonlinear dynamics. These dynamics implicitly compute internal models of the external world to generate adaptive behavioral responses.

The brain is one of the guiding metaphors of complex systems science, so that other examples – economies, ecological systems, social networks, transportation networks – are often conceptualized as "brain-like" in one way or another within complex systems theory. However, these other systems repay the favor and invite tantalizing metaphors of their own. For example, the synchronous blinking of fireflies has long fascinated mathematical biologists [2]. In this system, nonlinear interactions among blinking fireflies causes a spontaneous synchronized blinking that spans a whole swarm. Intriguingly, a lone firefly does not even display periodic blinking, so the drive to synchronous blinking is fully mediated by the network of interactions. In the 1990s, as mathematical tools and computer simulations began to clarify these dynamics, the potential connection to synchronous brain activity, and specifically epilepsy, began to be seriously considered [3]. One of the major discoveries in complex systems theory over the last few decades was the "small-world" phenomenon in many real-world networks [4]. Small-world networks have the property that most nodes are not directly connected to each other, but nevertheless most pairs of nodes can be connected by short paths. In their seminal paper on small-world networks, Watts and Strogatz showed that the synchronizability of a network is highly sensitive to the structure of connections – the *topology* – of the network, where small-world networks synchronize more readily than other patterns of connections, and they speculated that this may underlie the synchronizability of physically distant pairs of neurons in the visual cortex. There has since been a wealth of research on small-world and other topological properties of many kinds of brain networks in health and disease (see, for example, Chapters 9 and 10). It is interesting from the "complex systems perspective" that the early luminaries in the mathematics of synchronization were inspired as much by brains as by firefly swarms.

The example of synchronizing fireflies highlights a dictum in complex systems coined by the physicist Philip Anderson in the title of a classic essay "More is Different" [5]. The essential point of that essay, beyond the particular physical examples given, is that aggregates of many things can have qualitatively distinct collective behavior from any of the parts (whole brains do not behave like big neurons). For the fireflies, a network of interacting, asynchronous fireflies becomes a wave of synchronous blinking over length scales many orders of magnitude larger than an individual firefly blinking. This *emergence* of new phenomena has achieved highly refined mathematical description in condensed matter physics, but has echoes across many disciplines, and forms an organizing metaphor in complex systems thinking [6–8].

But how can we put these ideas to work in understanding clinical phenomena and designing new treatments for epilepsy? Said more stridently, what is the added value of taking this abstract, complicated, and potentially sterile perspective? Or more sympathetically, how does complex systems theory help us understand clinical variability and design new interventions in the brain to produce desired outcomes?

An interesting observation among genetic epilepsies is that mutations in a single gene can result in vastly different phenotypes. Specific examples include variability in outcomes in tuberous sclerosis even within single families [9], and the wide clinical variability associated with sodium channel mutations [10]. Given that patients with identical mutations can have outcomes ranging from cognitively normal and medically tractable epilepsy to developmental delay, intellectual disability, and intractable epilepsy, within a complex systems framework it is clear that the individually variable adaptation of the whole brain system to the same genetic perturbation is a critical driver of outcomes. Understanding the nature of the adapted network that predicts good vs. poor outcomes will provide extremely important pathophysiological information that cannot be inferred from the mutation per se. The same ideas can be applied to acquired epilepsies. For example, the variability of outcomes following traumatic brain injury [11] is partly a function of the injury itself but also a function of network adaptation that is likely to be influenced by the nature of the individual pre-injury networks.

In terms of treatment, a few analogies help emphasize the perils of ignoring complexity and the promise of embracing it. The networks in which humans intervene most deliberately and totally are traffic networks. The purpose of any traffic network is to facilitate the efficient transfer of people and goods in space. All else being equal, we would expect that adding more roads to a network would necessarily add efficiency – there is more room for cars to drive, more possible paths from point A to point B. Alas, this is not so, as

described in what is now known as Braess' "paradox." This classic argument shows that adding roads (under reasonable assumptions about driver behavior) can cause the overall traffic within the network to slow down. Conversely, there have been several real-world examples in which a temporary shutdown of major roads in cities has actually improved traffic flow [12]. The key point is that the overall traffic flow is a function of the whole network's topology. Thus, local heuristics, like "adding an expressway between two popular points will improve traffic flow," can have highly counterintuitive, negative consequences. A "toy" example of this effect can be seen in the ancient Hindu game Snakes and Ladders, where the addition of some ladders can lengthen the expected game length, while the strategic addition of snakes can actually shorten the expected length [13].

Now let us operationalize this analogy for epilepsy. Instead of cars on roads, the brain transports information along connectomes. Among the major therapeutic decisions in epilepsy is the strategic resection of some brain tissue or, more recently, the implantation of a neurostimulator device. However, if we take the traffic network analogies seriously, we must accept that local heuristics can lead us badly awry. If the emergent dynamics of the brain are determined by the whole connectome, then we must treat the whole connectome. Like adding or shutting down roads in a city center, adding or removing electrical pathways in the brain can have potent positive effects on whole brain function, but only if the rest of the brain is considered. Advances in imaging, machine learning, and dynamical modeling are facilitating such a holistic view, where virtual surgeries can be used to predict outcomes based on patient-specific network data (see Chapter 4).

Considering drug interventions, we can again consult far flung metaphors. The purpose of a drug in epilepsy is to suppress seizures. Medications do not directly influence the emergent phenomenon of seizures, but rather interact with a set of target molecules within cells and tissues in the body. In response, cells change their physiology, ideally toward a non-seizure-prone state. As is well known, however, the fraction of patients who are seizure free on any medication has remained stuck at around two-thirds for decades [14], and existing medications can have debilitating side effects, particularly when multiple treatments are prescribed simultaneously. The ability to predict what kinds of novel molecules will interact in just the right ways to normalize and stabilize the ceaseless molecular activity of the brain to prevent seizures is a goal of therapy development in a complex systems framework.

This problem is at least as hard as intervening in an ecosystem to normalize and stabilize population dynamics. Analogous to molecules within cells, organisms in ecosystems have diverse interactions forming a trophic network defining energy and material flows. There is an ignoble history of abject failures and a few instructive successes of human intervention into ecological systems. Canonical among the failures is the introduction of cane toads to Australia to control cane beetles; a strategy that had broad scientific consensus at the time. Not only did the toads fail to control cane beetles, they also destabilized the native ecosystem, endangering several species that did not coevolve with them [15]. In contrast, the reintroduction of wolves to Yellowstone National Park in the United States was successful beyond expectations [16]. Unlike the cane toads, the Yellowstone ecosystem evolved with wolves as an apex predator, who were extirpated by human activities. The reintroduced wolves had a number of salutary effects. Principally, as apex predators, they induced significant changes in behavior in their main prey species, elk, who no longer ventured out into the open to graze exclusively on the most desirable plants. This change in behavior had the downstream effect of allowing multiple plant populations to recover from overgrazing, which in turn allowed their roots to stabilize the soil, which arrested the erosion that was causing rivers to change course and further disrupt other niches. Furthermore, the availability of elk carcasses helped restore other scavenger species. Overall, biodiversity and population stability are both markedly improved.

The critical point to take away from the toads versus the wolves is that the wolves succeeded and the toads failed because of where they each sat within the trophic network. The wolves had an evolved function and a critical topological location within the trophic network as the apex predator. In contrast, the toads were speculatively introduced as a totally new node within a network. Importantly, both interventions had the proximal goal of controlling a target species (elk for the wolves, beetles for the toads), but it was network effects that determined success. In epilepsy terms,

these examples ask us to think deeply about how and why we choose molecular targets for anti-seizure drugs, and our strategies for targeting them. It is interesting to speculate at this level of generality whether we think any of our modern anti-seizure medications are wolves or toads. Like the traffic and ecological examples, the effects of introducing a molecule depends in highly nontrivial ways on the dynamics of the whole network of interactions. Systems biological approaches to genetic risk prediction and drug discovery, therefore, treat molecular networks and their emergent functions as fundamental, alongside individual molecule-trait associations (see Chapters 2 and 3).

The foregoing discussion has briefly highlighted the character of complex systems theory and sought preliminary connections to the main topic of this book. We hope this inspires interested readers to seek out comprehensive treatments of complex systems theory (as can be found in [1,8,17]), and keep these analogies and principles in mind as they go through the chapters. Overall, we have chosen to organize the book by physical scale within the brain, starting with genes and ending on whole brains. It should be stressed, however, that each chapter is a self-contained treatment of a topic. Each chapter in its own way, and to the extent possible for each data domain within neuroscience, discusses the promise of networked, dynamical thinking for epilepsy research and practice.

References

1. Holland, J. H. *Hidden Order: How Adaptation Builds Complexity*. New York: Basic Books. 1996.
2. Buck, J., and Buck, E. Synchronous fireflies. *Sci Am.*, 234(5), 74–9, 82–5 (1976). doi:10.1038/scientificamerican0576-74
3. Sanger, G. F. Beyond isolation: Preferred rates of oscillation, from fireflies to epilepsy. *Med Hypotheses*, 41(3), 211–14 (1993). doi:10.1016/0306-9877(93)90232-f
4. Watts, D. J., and Strogatz, S. H. Collective dynamics of "small-world" networks. *Nature*, 393 (6684), 440–2 (1998). doi:10.1038/30918
5. Anderson, P. W. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047), 393–6 (1972). www.science.org/doi/10.1126/science.177.4047.393
6. Kauffman, S. A. *The Origins of Order: Self Organization and Selection in Evolution*. New York: Oxford University Press. (1993).
7. Meadows, D. H. *Thinking in Systems: A Primer*. Vermont: Chelsea Green Publishing. (2008).
8. Bar-Tam, Y. *Dynamics of Complex Systems*. Boca Raton, FL: CRC Press. (1999). https://necsi.edu/dynamics-of-complex-systems
9. Wang, F., Xiong, S., Wu, L., et al. A novel TSC2 missense variant associated with a variable phenotype of tuberous sclerosis complex: case report of a Chinese family. *BMC Med Genet.*, 19(1), 90 (2018). Doi:10.1186/s12881-018-0611-z
10. Guerrini, R., Cellini, E., Mei, D., et al. Variable epilepsy phenotypes associated with a familial intragenic deletion of the SCN1A gene. *Epilepsia*, 51 (12), 2474–7 (2010). doi:10.1111/j.1528-1167.2010.02790.x
11. Forslund, M. V., Perrin, P. B., Røe, C., et al. Global outcome trajectories up to 10 years after moderate to severe traumatic brain injury. *Front Neurol.*, 10, 219, (2019). doi:10.3389/fneur.2019.00219
12. Easley, D., and Kleiburg, J. *Networks, Crowds, and Markets*. Cambridge: Cambridge University Press. 2021.
13. Althoen, S. C., King, L., and Schilling, K. How long is a game of snakes and ladders? *Math. Gaz.*, 77(478), 71–6, (1993). doi:10.2307/3619261
14. Brodie, M. J. Outcomes in newly diagnosed epilepsy in adolescents and adults: Insights across a generation in Scotland. *Seizure*, 44, 206–10, (2017). doi:10.1016/j.seizure.2016.08.010
15. Turvey, N. Everyone agreed: Cane toads would be a winner for Australia. The Conversation. November 7, 2013. http://theconversation.com/everyone-agreed-cane-toads-would-be-a-winner-for-australia-19881
16. Peterson, Christine. 25 years after returning to Yellowstone, wolves have helped stabilize the ecosystem. National Geographic. July 10, 2020. www.nationalgeographic.com/animals/article/yellowstone-wolves-reintroduction-helped-stabilize-ecosystem
17. Newman, M. E. J. *Networks: An Introduction*, Oxford: Oxford University Press. 2010.

Chapter 2

Systems Biology Approaches to the Genetic Complexity of Epilepsy

Jeffrey L. Brabec, Montana Kay Lara, Anna L. Tyler, and J. Matthew Mahoney

2.1 The Epilepsy Genetic Revolution

The genetic underpinnings of epilepsy have come into much clearer focus over the past two decades. Advances in high-throughput molecular techniques have markedly improved our ability to identify potential therapeutic targets in epilepsy. Many of the monogenic effects identified through these methods have resulted in effective therapeutic targets for seizure amelioration [1,2,3]. Currently, around 200 definitively annotated epilepsy genes causing a range of seizure disorders and phenotypes have been identified [4]. Many more genes with putative associations with epilepsy pathways require further study [5]. The expansion of known genetic mechanisms and risk factors presents us with several benefits, including an increased pool of possible drug targets [6], genetic subtyping of seizure disorders [7], and the possibility for integrative analysis across different disorders [8,9]. However, the increasingly rich collection of genetic associations has also revealed the complexity of seizure disorders. Many mutations in different genes can converge on a similar clinical presentation [10], while different mutations in the same gene can have radically divergent outcomes [11,12]. Moreover, while robust data from twin and family studies demonstrate that common epilepsies are highly heritable [13,14], association studies have only detected risk factors that account for a small fraction of risk [15]. Thus, the data on epilepsy suggests a dichotomy. On one side, genetics is critical for describing etiology [16]. On the other side, using this information for prognosis or therapeutic development is limited by our current understanding of the complex genetic underpinnings of the disease and our analytic tools [10,17]. As a response to this complexity, researchers have started to shift toward complex systems approaches to genetics, which changes the focus from individual mutations to interactions among many mutations. The purpose of this chapter is to elaborate this ethos and present examples of this approach.

It is important to stress that genetic associations in epilepsy come from two essentially distinct sources. One source is rare variant detection via whole-exome or whole-genome sequencing. In these studies, a patient's and family members' DNA is sequenced to identify putatively deleterious rare mutations [18]. These studies are typically undertaken in patients with epileptic encephalopathies for which no known genetic etiology is implicated. The other source of associations are genome-wide association studies (GWAS), in which large populations of cases and controls are genotyped at a set of common genetic variants, which are then statistically associated with disease status. The GWAS approach is used for common epilepsies such as temporal lobe epilepsy (TLE) and idiopathic generalized epilepsies (IGE). As will be discussed in the next section, these two approaches typically fall on two ends of a spectrum, on one end of which reside the monogenic disorders that are caused by mutation of a single gene and on the other end the polygenic disorders that arise through the combined effects of many genes.

The International League Against Epilepsy (ILAE) recently published a GWAS "mega-analysis" for several common epilepsies, including focal and generalized epilepsies [15]. Their analysis revealed 11 novel loci associated with common epilepsies, which implicated diverse biological mechanisms across epilepsy subtypes. Despite the statistical significance of these associations, the risk conferred by the newly associated variants was low (1.5–3.3 odds ratios). This is typical of GWAS for many complex diseases, and not a feature of epilepsy GWAS per se [19,20]. For monogenic epilepsies, there have been concerted bioinformatic efforts to collate rare variant data into searchable public databases. Resources such as the Online Mendelian Inheritance in Man (OMIM) database [21] collect validated mutations, while research databases such as ClinVar [22] enable

investigators to share variants of unknown significance detected in their patients alongside corresponding clinical presentations.

Despite this wealth of publicly available information, we are still far from precision medicine for patients with epilepsy. Indeed, few new therapies have been developed for specific genetic targets. A notable exception is Everolimus, which targets overactivation of mTOR signaling caused by mutations in the mTOR pathway. Unfortunately, Everolimus has only modest effects on epilepsy symptoms and does not appear to be more efficacious than other, nontargeted anti-seizure drugs at controlling seizures [23]. Why has the genetic revolution not resulted in a treatment revolution? To begin to answer this question, it helps to take a theoretical perspective and survey the genetic architecture of complex traits.

2.2 The Genetic Architecture of Complex Traits

The goal of genetic analysis is to identify a mathematical function that predicts an individual's phenotype from their genotype: a *genotype–phenotype map*. A phenotype can be a quantitative measure, such as body mass index, or a discrete category, such as disease status. In the latter case, which is most common for GWAS, the genotype–phenotype map predicts a risk for belonging to a given category. Assuming such a function can be found, it would appear to accomplish multiple objectives simultaneously. First, it would allow rigorous prediction of an individual's phenotype. Second, it would establish the individual contributions of genetic variants to overall risk. This latter property, more than just quantifying risk, should aid in developing a mechanistic understanding of the disease. The idea of precision medicine is predicated, in part, on using such predictive models for prognosis and treatment selection [24]. Of course, how predictive the genome is for a given trait and just what mathematical form a genotype–phenotype map should take are nuanced questions that require careful consideration.

The space of all possible genotype–phenotype maps is truly vast. Across the human population, the number of genetic variants in the genome is so large – and our sample sizes so minuscule by comparison – that we must be guided by theoretical considerations and be willing to accept reasonable approximations to make any headway. There are on the order of 10^7 single nucleotide polymorphisms (SNPs; i.e., common point mutations) in the human population. This is in addition to insertion-deletions (indels), structural variants, such as copy number variations, and so-called rare variants, which can be point mutations or larger structural variants [25–27]. A genotype–phenotype map would therefore include the effects of tens of millions of variables and all their possible interactions to predict a trait. To get a feel for the magnitude of combinatorial possibilities, note that for 10 million SNPs there are on the order of 100 trillion pairs of SNPs, to say nothing of higher-order interactions. This is colloquially known as combinatorial explosion. (To make a scale comparison, the number of possible SNP pairs is greater than the number of stars in the Milky Way.) To make matters yet more complicated, many genetic factors only become relevant through interactions with specific environments, such as an in utero exposure, so a generic genotype–phenotype map also requires terms for all possible interactions among genetic variants and variables representing environmental factors [28].

The above considerations paint a dire picture of our ability to estimate a genotype–phenotype map in general. Fortunately, the situation is not nearly so bleak in practice. For most diseases, we typically assume that only a tiny fraction of the putative predictors (e.g., SNPs) are in fact predictive. Thus, using an association study design we can statistically screen for a small number of relatively strong effects for further consideration. The association studies discussed in the previous section fit this pattern and have made significant advances in identifying genetic risk factors for common epilepsies.

It is worth noting, however, that the existence of a small number of relatively strong effects is not a biological imperative and is, in fact, quite rare for common diseases [25–27]. In principle, the variants influencing a trait could be diffuse throughout the genome, where a large number of extremely weak effects conspire to produce a given phenotype. Indeed, this was the model Ronald Fisher had in mind when developing the early tools of statistical genetics [29–31]. In that model, developed decades before the discovery of DNA as a store of heritable variation, Fisher conceptualized an infinite number of genes (operationally defined as units of inherence), each

making an infinitesimal contribution to the overall trait. Fisher was motivated, in part, by the observation that many heritable traits are normally distributed, i.e., have bell curve distributions, which follows naturally from a theoretical analysis of his infinitesimal model. Fisher's model has had a strong influence on many fields of genetics, including selective breeding and evolutionary and population genetics [32], and has received modern resonance in Boyle *et al.*'s concept of omnigenic inheritance. In the omnigenic model, essentially every gene contributes to every phenotype [33].

Paradigmatic among polygenic traits is height. Height is highly heritable – you are roughly as tall as your parents – but there is far from only one "height gene." On the contrary, the most recent meta-analysis of height has identified tens of thousands of SNPs that significantly influence height, but each contributes only a tiny fraction of a millimeter one way or the other [34]. One's height is largely determined by the complement of these tiny nudges inherited from one's parents. For any phenotype of interest, however, the question of whether a small number of strong effects, a large number of weak effects, or some heterogeneous combination of both, is a question that will need to be resolved with experiments. The answer to this question is often referred to as the *genetic architecture* of the trait.

In the present case of epilepsy genetics, we see a spectrum of genetic architectures. In rare, sporadic epileptic encephalopathies, exome sequencing has revealed a relatively simple architecture of a small number of large effects [35]. (Note that the simplicity of the architecture does not imply the simplicity of identifying these effects!) In contrast, many common seizure disorders, e.g., TLE and IGE, are expected to have the latter kind of inheritance [36]. This has significant ramifications for how to identify and, more importantly, how to use genetic associations that arise from association studies. We will discuss strategies for coping with this genetic complexity in the next section.

Before discussing how to approach genetic complexity, it is worth asking why individual genetic effects are often so weak. Part of the answer is that in GWAS the associations are made to SNPs, which are by definition common variants. No SNP with minor allele frequency of, say, 20% can be a complete causal explanation of a disease that afflicts at most a few percent of the population. At best, SNPs can reveal modifiers of an underlying pathology that alter disease risk. This is reflected in the odds ratios for disease risk for individuals SNPs, which are often in the range of 2 to 10. Alternatively, there could be cryptic causal variants that are not SNPs but are correlated with them (known as linkage disequilibrium). Finding a preponderance of SNP associations at some location, each with weak individual effects, can often signal the presence of a hidden strong variant. Thus, SNPs are a blunt instrument. However, this is unlikely to be the whole answer. Consider tuberous sclerosis complex (TSC), which is caused by loss-of-function mutations in either the TSC1 or the TSC2 genes. Despite the proximate cause of TSC being such a mutation, TSC is still highly heterogeneous. Indeed, even siblings who inherit identical mutations and have essentially identical environments can have markedly different outcomes [37]. While such case reports cannot rule out unmeasured environmental insults or rare mutations as second genetic hits, they do suggest hypothetical genetic modifiers of disease outcomes that interact with the primary mutation to push outcomes one way or another. This latter hypothesis implicates network-level effects even in putatively monogenic disorders, and all the more so in complex traits.

No gene operates in isolation. Indeed, genes are regulated by interconnected transcriptional networks and their products take part in overlapping signaling cascades and binding interactions [38–40]. There have been multiple attempts to organize our models of molecular networks that have variously emphasized the *computational* aspects, the *physical interactions*, and the *circuit-like* aspects of molecular systems (Fig. 2.1). In the 1990s, Denis Bray showed that the basic enzyme kinetics equations of signaling cascades in c ells are formally mathematically equivalent to multilayer perceptrons, a form of artificial neural network [41] (Fig. 2.1A). He posited that cells are biologically instantiated classification devices for transforming external stimuli into transcriptional responses [42]. Multilayer perceptrons have several appealing properties as computational architectures. They can implement highly nonlinear input–out relationships. Thus, signaling cascades in cells can make complex calculations on stimuli and respond with a vast repertoire of responses. Furthermore, multilayer perceptrons have a "graceful degradation" property that corrupted

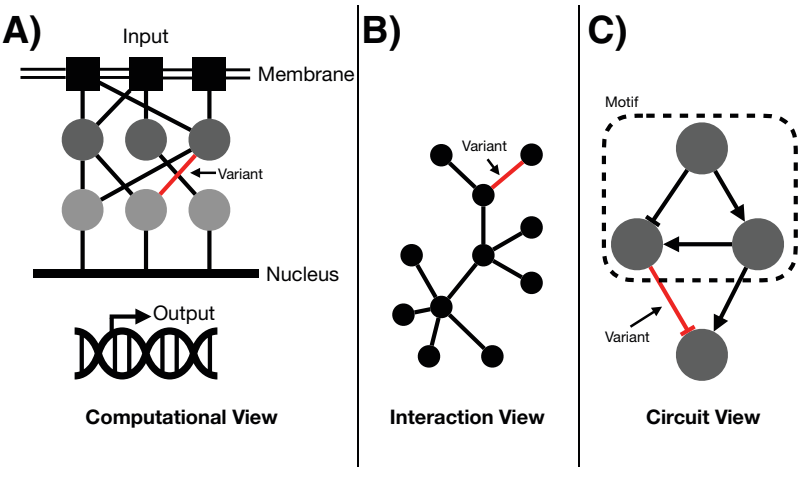

Figure 2.1 Schematic of network views common in systems biology. The mathematical language of networks is a useful tool in systems biology to capture the organization of molecular systems. A) An early view in systems biology was the *computational view* that used ideas from machine learning to analogize cell signaling interactions as a multilayer perceptron, which is a form of artificial neural network. In this view, the concentrations of molecules outside the cells are *input* and the transcriptional response is *output*. The signaling cascade itself instantiates a nonlinear relationship between input and output. B) As high-throughput data on molecular interactions, such as protein–protein interactions, became available, a purely topological *interaction view* became popular. In this view, one effectively ignores the dynamics of the system and studies the patterns of interactions and correlates them to other molecular properties, such as the effects of mutations. C) A *circuit view* of a molecular network attempts to decompose a topologically complex network into commonly repeated subunits, called motifs. The dynamics of motifs in isolation can provide some insight into how those motifs function within the full system. In each view, the effect of a genetic variant is to alter an interaction in the system, e.g., by altering the strength of a binding interaction.

networks have decreased performance in proportion to the amount of corruption, rather than catastrophic failure after a few components are removed [43]. This *computational view* hypothesized that it is the computational capabilities of cells that are positively selected by evolution, and these systems have converged on networks that are robust to mutations. Therefore, the convoluted interconnectivity of molecular networks appears as a property that mitigates the effects of individual mutations. Thus, the complex organization of signaling pathways is not a "bug" due to historical contingencies of evolution, but an essential "feature" of the computational architecture of cells.

Network-theoretic investigations have established that biological networks are indeed largely robust to random perturbations. For example, Barabási and collaborators have argued that protein–protein interaction networks, i.e., networks of physical binding between proteins, have evolved into so-called scale-free topologies that are robust to random mutations that knock out a random protein [44]. This *interaction view* abstracts away from the complexity of the dynamics of molecular systems and emphasizes the topology of those interactions. The scale-free network concept of Barabási & Albert has come under significant scrutiny [45]. However, the general point that protein–protein interaction networks have highly nonrandom structure that makes them robust to randomly deleting a node appears sound [46] (Fig. 2.1B). Likewise, Alon and collaborators have identified network motifs in transcriptional regulation networks, which confer robustness to transcription as a function of biophysical parameters that could be altered by mutations (Fig. 2.1C) [39,47]. In this *circuit view*, the complex interactions and dynamics of molecular systems can be approximately decomposed into subsystems, each themselves nonlinear components, that perform functions that are robust to perturbations of their parameters.

Through collecting these insights, a provisional explanation of the polygenic nature of many traits takes form. Common mutations provide small perturbations to the function of one component or interaction in a system, for example, the efficacy of a signaling protein (Fig. 2.1A), the binding affinity of a protein to a partner (Fig. 2.1B), or of a transcription factor to the binding site of another transcription factor (Fig. 2.1C). Network robustness attenuates the effect of these perturbations to maintain overall function. The tortuous path from any mutation to

an observable phenotype is mediated through hundreds/thousands/millions of molecular interactions, many of which have been perturbed by other mutations, each contributing additional variance to the subpopulation carrying the first mutation. Hence, at the population level, most variants have a small effect.

So far, we have only considered the effect of one variant at a time within an association-analysis framework. This ignores the putative interaction terms described in the generic genotype–phenotype map. Given the preceding discussion of networks and interacting mutations, it seems reasonable to model these interactions. Statistical interactions among variants, i.e., deviation from additivity in a linear model, is called *statistical epistasis* [48,49]. Epistasis has a long history in genetics going back to Sewall Wright [50]. The existence of significant statistical epistasis in human populations is an ongoing point of controversy. Some argue that epistasis is both statistically detectable with reasonable sample sizes (despite the combinatorial explosion of the number of interactions) and identifies biologically relevant networks [51–54]. Others argue from theoretical evolutionary genetics models and empirical considerations that nearly all population-level genetic variance in humans is captured by additive models, despite the underlying nonlinear molecular interactions among gene products [55,56]. In model systems, where experimental crosses between evolutionarily diverged lines are possible, epistasis is not controversial. It has been observed in multiple model species that epistasis is particularly important for predicting extreme phenotypes [52,54,57], indicating the potential relevance of epistasis for predicting individual disease risk. With these caveats, the present authors are sympathetic to epistasis analysis in general and look to promising computational [58,59] and theoretical [60,61] advances that will potentially make epistasis modeling impactful for epilepsy GWAS.

2.3 Overcoming Genetic Complexity for New Insights

The discussion in Section 2.2 argued that it may be impossible to completely enumerate all genetic risk factors for epilepsy. Despite heritability, some amount of risk may be so diffusely embedded in molecular networks that we will never observe it in association analyses. This does not prevent our ability to make progress, but it does require that we modify our approach. In the following sections, we describe examples of approaches that confront this complexity directly.

2.3.1 Genetics of Gene Expression Networks: The Case of SESN3

One straightforward solution to the inadequacy of GWAS data to resolve all risk genes is to augment the genetic data with additional information. The most obvious choice is gene expression data, which provides a functional readout of the genome and can be measured in tissue from patients who undergo epilepsy surgery. Just as "omics" approaches, which study high-throughput cross-sections of molecular systems, are called *systems biology*, the combined analysis of genetics and gene expression is called *systems genetics*. Recently, Johnson *et al.* performed a systems genetic analysis of TLE [62]. Starting with gene expression data from resected hippocampal tissue from patients with TLE, they modeled gene co-expression using a gaussian graphical model, which captures the partial correlations among all gene pairs to estimate direct gene interactions. This allowed them to build a hippocampus-specific gene interaction network whose structure encodes the pathways that connect genes to each other. To ascertain whether this network was directly linked to underlying genetic risk for TLE, they performed a de novo analysis of TLE GWAS data and used a relatively liberal false discovery rate-based correction for multiple hypothesis testing. Within their hippocampus network, they identified two gene expression modules that were highly enriched for TLE GWAS risk genes. Each module was then run through pathway analysis to identify significantly enriched pathways. By accepting a certain amount of statistical noise in the gene associations, they were able to get a robust pathway-level signal for TLE risk genes.

One of the modules was highly enriched for pro-inflammatory cytokine signaling that was conserved across humans and mice. The expression of the genes in this module, therefore, represents an endophenotype for TLE risk, i.e., an intermediate phenotype with a clearer connection to genetics [63]. They then used the module expression as a phenotype for genetic mapping and identified one significant genomic locus

altering module expression. Through gene prioritization of the candidates within the locus, they were able to identify and validate the gene *SESN3* as a novel regulator of a proconvulsant gene network. Interestingly, *SESN3* was not present in the original module, suggesting that *SESN3* is a *trans*-acting regulator of expression.

The study by Johnson *et al.* highlights several key features of systems genetics analysis [62]. First, there is always a tension between false positives and false negatives in any high-throughput screen such as GWAS or transcriptomics. In this study, rather than requiring stringent evidence in each layer of data, evidence accumulated across multiple layers, ultimately implicating a novel gene that would have been missed by stringent statistical criteria. Second, by treating gene expression as a network phenomenon, they were able to resolve biologically specific co-expression modules, one of which was sufficiently genetically regulated to enable novel gene discovery. Third, the use of gene expression as an endophenotype is a powerful approach to resolving the complexity of genotype–phenotype maps. Gene expression more closely reflects the underlying genetic sequence than a subject-level outcome.

In follow-up work, Delahaye-Duriez *et al.* further elaborated the convergence of genes for rare and common epilepsies on shared functional networks [10]. Similarly, Johnson *et al.* showed that cognitive and neurodevelopmental disorders also share common genetic networks [64]. These studies suggest that, despite overt differences in genetic architecture and clinical presentation, there is some mechanistic convergence underlying epilepsy and comorbidities, and regulators of these networks, such as *SESN3*, can be identified through systems genetics and therapeutically targeted to improve outcomes.

2.3.2 Augmenting Statistical Genetics with Functional Networks

The empirical finding that genetic risk alleles for epilepsy are enriched in specific gene networks opens the possibility that we could search for those functional gene networks directly. In the systems biology field, several tools have been developed to combine genomic data with bioinformatic gene interaction networks to rigorously circumscribe disease gene networks [65–70]. For example, the Network-wide Association Study (NetWAS) tool uses GWAS summary statistics to identify tissue-specific disease gene networks that are enriched for risk alleles. In the original article on NetWAS, Greene *et al.* used NetWAS to reprioritize genes from a hypertension GWAS study using a blood vessel tissue-specific functional gene interaction network [68]. The top genes from their model were localized to an IL-1β inflammatory response network, a known disease pathway in hypertension. Moreover, the network-based gene rankings dramatically outperformed GWAS summary statistics at identifying drug targets for antihypertensive medications. Importantly, NetWAS and similar tools are designed to work with liberal statistical cutoffs for GWAS associations, using bioinformatic prior knowledge to "de-noise" the underlying signal. In their proof of concept on hypertension, Greene *et al.* were able to show that the network-based signals for hypertension were indeed highly enriched for biologically actionable information, including drug targets, despite using liberal GWAS cutoffs for gene associations.

Since the original publication, NetWAS has been cited 508 times, and applied to numerous complex diseases. While it has not yet been used for epilepsy, it has been applied to several neurological disorders, including Alzheimer's disease (AD) [71]. The present authors used a NetWAS-like approach to identify genes involved with amygdalar and hippocampal atrophy in AD, implicating genes involved in actin regulation whose dysfunction leads to the collapse of the tripartite synapse and excitotoxic neuron death [72]. Chang *et al.* applied similar network techniques to rank genes for association to schizophrenia, another heterogeneous and genetically complex neurological disorder [73]. Additionally, Krishnan *et al.* functionally characterized genes in autism spectrum disorder using network-based methods [74]. While AD, autism, schizophrenia, and hypertension are each biologically distinct, and different from epilepsy, they share many similar features in their genetic architecture. Bioinformatic network-based techniques are a promising avenue for detecting epilepsy risk gene pathways from faint genome-wide signals.

2.3.3 Model System Studies of Risk Factors and Modifiers

While genetic complexity is a hindrance to statistics in observational studies like GWAS, we can