Abstract
Massive investment and technological advances in the collection of extensive and longitudinal information on thousands of Alzheimer patients results in large amounts of data. These “big-data” databases can potentially advance CNS research and drug development. However, although necessary, they are not sufficient, and we posit that they must be matched with analytical methods that go beyond retrospective data-driven associations with various clinical phenotypes. Although these empirically derived associations can generate novel and useful hypotheses, they need to be organically integrated in a quantitative understanding of the pathology that can be actionable for drug discovery and development. We argue that mechanism-based modeling and simulation approaches, where existing domain knowledge is formally integrated using complexity science and quantitative systems pharmacology can be combined with data-driven analytics to generate predictive actionable knowledge for drug discovery programs, target validation, and optimization of clinical development.

Keywords: Brain disorders; Alzheimer’s dementia; Complexity theory; Systems pharmacology; Systems biology; Drug discovery and development

1. Introduction
Among all the enduring disabling conditions of our increasingly aged society, chronic brain disorders, such as dementia/Alzheimer’s disease (AD), are the leading contributors to spiraling health care costs as well as individual and caregiver burden. Leading health organizations across the world have estimated that brain disorders (neurological, psychiatric, brain injury, or pain) will affect one in five individuals in their lifetime with an associated cost of more than 2 trillion US dollars annually in the United States and Europe alone [1]. This exceeds the annual combined burden of cardiovascular disease, cancer, and diabetes, and it is expected to rise with increasing life expectancy. Among these brain disorders, dementia represents one of the largest burdens to our aging societies [2], afflicting more than 35 million people worldwide [3]. Today, there are no effective therapies for these conditions, despite enormous financial and research investments. This reality has galvanized a global effort launched by the G8 Summit on Dementia in 2014; however, we suggest that the large
investment in research and development can substantially benefit from integrative predictive modeling.

From 2002 to 2012, 99.6% of the clinical trials of disease-modifying treatments for AD have failed [4]. In stark contrast, between 1995 and 2010, approximately 300 interventions were reported to reduce pathology and/or improve behavior in transgenic AD mouse models [5]. Although changes in preclinical animal research design might somewhat improve the predictive value of these animal models [6], rodents will continue to be fundamentally different from humans [7].

AD is inherently a multifactorial syndrome, and individual patients present with a wide variety of pathologies, as a consequence of comorbidities, life history, and genotypes (Fig 1). In fact, neuropathologic evidence suggests that different pathologies converge in the brains of elderly people with dementia [8,9]. This suggests that the biological processes driving the clinical phenotype can differ markedly from patient to patient. In addition, up to one-third of nondemented, high-functioning seniors may harbor underlying pathology to an extent that would be expected to cause dementia. So far reductionist molecular biological approaches have failed to explain this phenomenon [9].

The complexity of clinical trials for AD has also contributed to the therapeutic failure rate. The clinical outcome metrics related to cognition and function are highly variable, not only due to the inherent variability in the pathological processes (see above), but also the impact of co-medications and genotypes both within and across patient groups, necessitating large sample size and treatment duration to detect remediation. New modeling efforts such as the precompetitive consortium, the Coalition Against Major Diseases, can help develop tools to optimize the efficiency of clinical trial design [10]. Biomarkers can quantify neuropathology and its progression but the use of single molecular biomarkers in isolation has unfortunately not successfully predicted the functional and cognitive outcomes relevant to patients.

2. From reductionism to integration

The prevailing paradigm for scientific inquiry in the neuroscience field has used classical reductionism, an approach wherein explanation of entire systems is predicated in terms of their individual, constituent parts, and their interactions. This molecular biology approach, often based on data-driven correlation analysis, is basically a bottom-up strategy, where the resulting outcome is defined usually as a consequence of a single set of linear assumptions. This often negates the many nonlinear interactions between subsystems and the appearance of emergent properties that cannot be reduced to a single target.

The case of beta-amyloid modulation as a therapeutic approach for AD illustrates the problems associated with a statistical approach that correlates a clinical phenotype with genetic information. The most optimistic perspective on the failure of this approach is that these trials have been conducted too late in the course of the disease, a failure in the trial design rather than the targets, and that the solution is to conduct trials in prodromal conditions [11]. However, the assumption that “reducing beta-amyloid load” leads to cognitive improvement is probably a major oversimplification of the complex biology of beta-amyloid in the human AD brain that we are gradually starting to understand. Recent studies indeed document different aggregation dynamics [12], different formation and clearance in the human brain [13], different neuroprotective versus neurotoxic properties of the shorter versus longer amyloid peptides [14], and the complex nonlinear interaction of co-medications and genotypes on clinical cognitive readouts [15]. In other words, even if beta-amyloid is the correct therapeutic target, successful drug development will likely require a more sophisticated understanding of its complex dynamics. In addition, nonamyloid processes such as tau pathology, neuroinflammation, and oxidative stress interact with beta-amyloid physiology resulting most likely in an idiosyncratic cognitive trajectory for each AD patient.

With the development of systems biology, the concept of circuit and network insights was combined with multivariate analyses, resulting in an integrative approach that starts with the patient. It “is about putting together rather than taking apart, integration rather than reduction. It requires that we develop ways of thinking about integration that are as rigorous as our reductionist programs, but different...It means changing our philosophy, in the full sense of the term” [16]. Quantitative systems pharmacology (QSP) goes one step further by adding formalized domain expertise about the biological nature such as enzyme kinetics and interaction between drugs and targets of the different parts of key circuits and pathways. In this way, causation is explicitly integrated in the modeling. In the case of central nervous system (CNS) disorders, QSP also integrates this information into biophysically realistic neuronal networks, the firing properties of which can be associated with a clinical phenotype [17].

3. From big data to smart data

As part of the new approaches to reduce clinical trial failure rates, global efforts are now shifting toward a focus on gathering “big data” [18]. The integration of large clinical data sets is viewed as a potentially powerful approach to expedite medical discovery, and there is justifiable enthusiasm based on results of global studies of disease progression and large-scale genomics efforts [19]. Advanced deep analytical approaches have been developed and are covered by other publications in the field of bioinformatics [20,21] and pharmacology [22]; however, specific case studies for brain disease are limited and these publications are typically written for a very narrow specialty audience.

These large-scale data collection efforts, such as the Alzheimer’s Disease Neuroimaging Initiative (ADNI) consortium, and the European initiatives (http://www.alzheimers.org.uk/site/scripts/download_info.php?fileID=2273) will yield
maximal impact if they are combined with advanced predictive modeling approaches where existing domain expertise is formally integrated. We call this approach “turning big data into smart data” with the idea of generating actionable knowledge that could help us in developing new treatment paradigms. This provocative title is intended to highlight the link between the world of “data-information” and the realm of actionable “knowledge”—thus the concept of smart data.

4. Modeling and simulation in pharma and other industries

Across many fields of science and physics, modeling and simulation have come to complement theory and experiment as a key component of the scientific method. Computer aided design is a technology commonly used in many engineering disciplines to generate in silico prototypes that are extensively simulated before the actual prototype is built and lead to much higher success rates and shorter cycle development times in automotive, aeronautics, and micro-electronic industries. High-energy physicists use computer models such as Geant4 (https://geant4.web.cern.ch/geant4/), based on physical principles to generate knowledge from the massive amount of information (100 terabytes/day) gathered in large particle accelerators to better understand the complex nature of the universe.

In pharmaceutical industry, mechanism-based modeling and simulation have become mainstream in other disease indications such as metabolic disorders, toxicology, and oncology. Physiology-based pharmacokinetic modeling, which uses mechanism-based equations rather than empirical methods [23], is now part of the growing toolbox of the pharmaceutical industry and is increasingly being accepted by regulatory agencies [24,25]. As an example, DILISym, a complex physiology-based computer model or QSP model of liver injury [26], has been tested extensively and is used by pharmaceutical companies to detect early signs of liver toxicity. Although hard numbers are difficult to come by, informal testimonials reveal that this QSP approach can accelerate clinical development by several months while substantially increasing the probabilities of success. In another example, a mechanism-based QSP model of cognitive impairment correctly predicted an unexpected clinical proof-of-concept for a clinical candidate for AD and identified a back-up compound that could
possibly have saved a complete clinical development program [27].

The fact that large pharmaceutical companies see value in computer-based mechanistic modeling is further underscored by the recent announcement that Rosa, a company providing modeling and simulation capabilities, has signed a multiyear worldwide multi-indication contract with Sanofi Global R&D (http://www.prnewswire.com/news-releases/rosa-co-announces-multi-year-world-wide-physiopd-research-agreement-with-sanofi-300089963.html).

5. The brain and complexity theory

The examples of CAD and high-energy physics illustrate the power of applying first principles based on the understanding of the nature of the processes and formalized in a mathematical manner to elucidate emergent properties that are difficult to relate to one specific and unique process. This form of complexity theory includes tipping points (moments when unique phenomena become commonplace) and emergence (the idea that new structures can emerge unexpectedly from complex systems).

For example, experimental electrophysiological data on pharmacological manipulations of working memory [28] can best be explained by attractor dynamics [29], where the system dynamically settles into one of several states, and “escape” rates can be affected by disease state (such as “perseverance”). These studies suggest a novel hypothesis to explain some of the nonlinear dose-responses observed in the clinic, especially around dopamine modulation [30]. On the other hand, the clinical observation from deep brain recordings in Parkinson’s disease patients that motor symptoms are not related to firing frequencies, but to complex oscillatory behavior of local field potentials [31] is an example of an unexpected emergent property that cannot be reduced to a single target. This has led to a better understanding and optimization of deep-brain stimulation protocols.

Other examples of this nascent field include the association of brain health with fractality of the alpha rhythm of human EEG [32,33], the realization that chaotic behavior of neuronal networks is associated with healthy functioning [34,35], and the global emergence of cognition from social and neural cooperation [36].

Complexity science can help researchers to understand the causal relationships between molecular events (lower levels of complexity) and clinical expression of neurodegeneration (higher level of complexity; see for example Fig 2), by developing simulation platforms that link the pathogenesis of dementia with the clinical features of the disease such as cognitive impairment.

Fig. 2. Illustration of the multimodal processes that describe the complexity of going from a single gene (in this case, the huntingtin gene) to fully understand the pathology that leads to the multiple clinical phenotypes in Huntington’s disease patients. The increasing complexity when going from one level to the next necessitates the introduction of advanced mathematical modeling and simulation approaches that fully embraces nonlinear and stochastic descriptions of the neurophysiological processes that ultimately leads to clinical phenotypes. For example, although the basic driver of the pathology is the mutated huntingtin gene, its effects on behavior are related in complex nonlinear ways to other processes to the point that is not clear what the optimal target modulation approach would be. In addition, environmental factors or other genotypes likely affect the relative contribution of these pathologic processes to the clinical phenotype.
6. Defining the questions to be asked in AD

It is also important to realize that several scientific questions need to be addressed with different levels of detail in various models. For purely empirical relationships (e.g., what genotypes drive a specific phenotype), deep analytical approaches based on sound statistical modeling and machine-learning principles are probably sufficient. For more predictive and biological questions that are not readily apparent from existing databases (e.g., how to therapeutically affect biological processes that will impact the clinical phenotype in a specific patient population), extensive neurobiological domain expertise needs to be fully and formally integrated. Fig 1 illustrates a possible strategy on how the information from big data analytics can be combined with mechanism-based modeling to generate an advanced and realistic computational platform for AD. The purpose of the modeling is to bring the arrows (relations) between different agents in the pathology to “life,” by using first principles about the underlying biology and explicitly simulating their time-dependent and concentration-dependent impact on the whole network. This will help to better define the causality between different biological processes and in relation to the different clinical phenotypes. Further constraining the parameters with clinical data will improve the predictability of such models.

The heterogeneity of AD suggests that interventions to treat or prevent it will be effective in some but not all patient subpopulations. For example, the omega-3 fatty acid docosahexaenoic acid (DHA) has been tested as a therapy to treat or prevent AD. Trials have generally failed, but secondary analysis of a clinical trial in AD patients reported that *APOE ε2/ε2, APOE ε2/ε3* carriers benefited, whereas *APOE ε2/ε4, APOE ε3/ε4* and *APOE ε4/ε4* carriers did not [37]. Additional evidence from observational studies [38], animal studies [39], and DHA pharmacokinetic experiments in humans [40,41] indicate that the *APOE* genotype alters the response to DHA supplementation and should be considered in clinical trials testing the efficacy of DHA for AD treatment or prevention. Although this finding was based on a statistical post hoc modeling of clinical results; the challenge is to turn this into a better understanding of the biological processes driving this outcome, so that possibly new targets can be identified and validated.

Below are some concrete objectives for which the integration of bioinformatics, statistical modeling, complexity science, and deep analytics could revolutionize therapy development.

1. What biomarkers or combinations of biomarkers hold most promise for tracking with clinically meaningful change in AD trials? For example, what changes could be detected in a 3-month period in patients and how can this help us better understand the longer term clinical trajectory of AD patient subgroups, develop therapeutics, and design clinical trials?

2. What biomarkers can be used to help identify patient subgroups, differential diagnosis, and disease progression? Can the integration of complexity models with other analytical methods help identify biomarkers that will distinguish between subgroups to differentially predict disease progression, predict disease risk in asymptomatic populations, design more robust clinical trials with relatively homogeneous and clinically more relevant populations, and identify patient subpopulations most likely to benefit from a given therapeutic?

3. How can we support the discovery of new therapeutics that appropriately modulate key pathways associated with specific biomarkers that define a patient subpopulation and therefore will most likely be of benefit? In what direction and to what degree do we need to affect specific pathway(s) to generate a clinically significant improvement? In view of the limited efficacy of highly selective therapeutic interventions, how can we optimize combination therapies that have a substantial impact on the cognitive trajectory?

4. How can we conduct trials on virtual human patients to improve development of drugs for AD patients with different co-medications and genotypes?

7. Brain Health Modeling Initiative: Rationale

We strongly believe that generating actionable knowledge from these data sets is essential to develop new and effective treatment paradigms. To this end, we have established the Brain Health Modeling Initiative (BHMI). The objective of this initiative is to accelerate biomarker and therapeutic development by raising the awareness of integrative analytics and mechanism-based computational approaches that optimize the use of complex big data to generate a more accurate and actionable understanding of the disease. This may in turn lead to the development of more effective therapies or more effectively screen for patients with AD-specific pathology or an improved match between therapeutic targets and biomarkers that might help effect the promise of “precision medicine” [42] and therefore help deliver the right drug to the right patients.

We propose to address these issues by rationally integrating advanced modeling and simulation approaches with analytical algorithms from big data studies. Deep analytic approaches can identify complex relationships from mining existing clinical data, whereas mechanism-based disease modeling using complexity science can simulate how the emergent properties of a system (e.g., the clinical syndrome of dementia) emerge from the interaction of these diverse-related variables so that the right target for the patient population with that specific biomarker can be identified, validated, and optimally modulated.

As an example whereby mechanism-based modeling approaches can be combined with insights from bioinformatics, a computer model to identify vulnerability nodes of
Building Predictive Computer Simulations of Brain Disorders

1. Assemble and Integrate Diverse Data
2. Look for Connections
3. Validate and Confirm Hypotheses Through Biology Experiments
4. Build Predictive Model

Fig. 3. Steps for building predictive models. Starting from integrated databases, causal relationships can be identified using not only statistical analysis but also approaches where domain expertise is formalized. These relationships can be tested in biological experiments, together with clinical neuroimaging and neuropathology data and quantitative complex computer models can be developed. Parameters of this model are constrained by clinical data, and predictions can then be tested against actual clinical outcomes. We anticipate a series of interactive steps that will ultimately result in more complex and predictive models.

The analytics toolbox aims to provide high-level explanations of various computational algorithmic approaches available today.

Case studies will highlight valuable examples in which predictive modeling tools ranging from statistical modeling and deep analytics over pathway-guided correlation analysis to mechanism-based modeling have already been applied for the identification of prognostic and diagnostic biomarkers and clinical outcomes of new therapies in CNS disorders.

Operational challenges, such as data standardization and quality control, will address actionability with the aim of defining the big data requirements of clinical trials (experimental data) and epidemiology/demography/public health (observational data), outline the respective resource-infrastructure requirement of these two worlds, and suggest a roadmap to bridge the gap between these related endeavors.

Modeling applications will illustrate the viewpoints of pharmaceutical companies, regulatory agencies, and other major stakeholders regarding the use of predictive modeling in rational drug discovery and development, combination therapies, and the repurposing of old drugs.

8. Conclusion

Around the world, unprecedented amounts of data are being collected with diverse content ranging from the genetic tau neurophysiology was developed [43] based on biological principles, but where the constraints for the transition rates between the different tau states are informed by an a priori identifiability approach [43,44] commonly used in bioinformatics studies.

Fig 3 shows a general overview on how to bridge the world of experimental observation to deep analytics and insight-providing modeling and simulation, in an iterative cycle to generate better knowledge.

The mathematical toolbox for applying this network systems biology conceptual framework to brain diseases is still nascent and requires investment to realize its full potential.

In a series of perspective articles, we will explore the scientific, operational, and computational challenges and opportunities to use these approaches for brain-related data and chart a path to implement these novel technologies to accelerate the discovery of new and effective therapeutic interventions. The ultimate goal is to combine the best tools of the various computational approaches to develop a multiscale complex predictive modeling platform.

These position articles will discuss the rationale for generating actionable knowledge from the many databases that are currently being collected, the unique data challenges, as well as new opportunities.
and molecular "omics" to the clinical phenotypes of patients in their doctor’s office. Big data could revolutionize the development of effective treatments for AD but only if such data are turned into actionable knowledge. Integrative mechanism-based predictive platforms using complexity science have successfully led to scientific advances in other fields. Such advanced algorithms when combined with big data information could similarly advance AD research and development by creating a system-based understanding of this heterogeneous disease to predict which molecular targets (and corresponding drugs) will yield clinical benefit in which patients and to improve the clinical development success rate. The BHMI is an open call to action to share information on analytic approaches, address operational challenges, and develop new modeling applications that can transform “big data” into “smart data” and help bring the right drugs to the right patients in the shortest amount of time.

RESEARCH IN CONTEXT

1. Systematic review: In this invited review, we challenge current approaches for understanding complex disease conditions, such as Alzheimer’s disease (AD). Specifically, we explore the potential for modeling and more advanced computational methodologies to generate actionable knowledge that will support AD drug research and development. This article integrates multiple views by clinical, systems biology, and computational modeling experts on the contentious question of whether reductionist scientific methods alone can solve the challenges we face in developing therapeutics.

2. Interpretation: Given the inherent complexity of AD pathology, and challenges in working with big data, we propose that bringing in formalized domain expertise and novel analytic approaches is a possible solution to this problem.

3. Future directions: This article is intended to raise awareness of the problems and issues associated with translating insights from big data into actionable outcomes and outlines the principles on which we will develop a series of articles with the objective of informing and educating the scientific community about various computational methods that are not widely used today.

References


