nature medicine

The search for imaging biomarkers in psychiatric disorders

Anissa Abi-Dargham[1,2] & Guillermo Horga[1]

The field of medicine is moving toward the use of biomarkers for the optimization of individualized care. This is a particular challenge for the field of psychiatry, in which diagnosis is based on a descriptive collection of behaviors without the availability of any objective test to stratify patients. Neuroimaging techniques such as molecular imaging with positron-emission tomography (PET) or structural and functional magnetic resonance imaging (MRI) provide an opportunity to bring psychiatry from an era of subjective descriptive classification into objective and tangible brain-based measures. Here we provide steps toward the development of robust, reliable and valid biomarkers. The success of such development is crucial because it will enable the field of psychiatry to move forward into the era of modern medicine.

The search for biomarkers—objective biological measures that can predict clinical outcomes—is consistent with the precision-medicine initiative, which gained official support from the White House in January 2015 (ref. 1). The approach, in which treatment and prevention for each person is carried out by taking into account the individual's genes, environment and lifestyle, relies heavily on biomarkers (genetic or otherwise). This personalized 'stratification' approach is already revolutionizing cancer treatment, wherein novel drugs that target specific molecular-signaling pathways related to genetic mutations are currently under development and in testing[2], thereby enabling treatments to be tailored to a patient's genomic profile[3]. Treatment of some neurological disorders, such as epilepsy, has been carried out with some success by using similar strategies[4]. The application of precision medicine to psychiatry, however, is more challenging, because the path from genes to behaviors is influenced by a series of complex interactive links that have yet to be fully understood. Yet, psychiatric disorders are responsible for immense personal, social and financial burden. Medical costs in the USA alone were estimated at $57 billion in 2006 (ref. 5). More importantly, indirect costs resulting from lost earnings, in particular income lost owing to severe mental illness, has been estimated at $193 billion annually in the USA[6]. Biomarkers can help to reduce these staggering costs by enabling better and earlier detection and improved treatment.

Biomarkers are either diagnostic biomarkers that index a biological process associated with health or disease, or predictive biomarkers that reflect a process associated with the therapeutic response and are used in clinical stratification. Neuroimaging could satisfy both of these goals, because molecular imaging with PET or structural and functional MRI can be used to measure phenotypic variations in molecular and cellular disease targets, or in specific brain circuits that are a unique representation of the interaction between genes and environment and are associated with specific alterations in behavior (**Fig. 1**), respectively. In the future, if these imaging measures demonstrate sufficient precision and reliability and can predict a clinical diagnosis or outcome, then they will become imaging biomarkers. Here we will discuss the main challenges of developing imaging biomarkers for psychiatric disorders and outline crucial benchmarks for the translation of neuroimaging findings into clinically useful biomarkers.

Challenges of developing biomarkers for psychiatric disorders

Lacking gold standards for psychiatric diagnoses. The first challenge is the definition of psychiatric disorders, which by standard nosology (as reflected in diagnostic manuals such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) or the International Classification of Diseases (ICD)) is based on combinations of symptoms alone. In general, there are no existing gold standards that are based on biological tests such as postmortem histological or other objective tests that can be used for definitive validation of psychiatric diagnoses. The hope is that neuroimaging could provide biomarkers that would ultimately support any nosological diagnostic classification on the basis of objective tests, as in other areas of medicine. These might be part of a panel of tests that include nonimaging modalities, for example, genetic, peripheral blood-based or cognitive tests. Here we focus on neuroimaging biomarkers as one subset of potential biomarkers.

A new classification scheme recently proposed by the US National Institute of Mental Health (NIMH) might aid in the development of neuroimaging biomarkers for psychiatric disorders. This is based on the Research Domain Criteria (RDoC) approach, a "new way of classifying mental disorders based on dimensions of observable behavior and neurobiological measures" (http://www.nimh.nih.gov/research-priorities/rdoc/index.shtml). This approach is directly relevant to the search for biomarkers because it aims to identify valid elements, such as genes, molecules, cells, circuits, physiological measures or behavior, that are associated with specific cognitive constructs across different systems. Neuroimaging (such as PET and single photon emission computed tomography (SPECT) radiotracer studies and magnetic resonance spectroscopy (MRS)) fits well within this approach because it can identify biomarkers related to the 'cells' and 'circuits,' although the results of these studies are not always direct or easily interpretable from a biological perspective. Diffusion tensor imaging (DTI) techniques can produce images of anatomical

[1]Department of Psychiatry, Columbia University & New York State Psychiatric Institute, New York, New York, USA. [2]Department of Radiology, Columbia University, New York, New York, USA. Correspondence should be addressed to A.A.-D. (aa324@cumc.columbia.edu).

Received 19 May; accepted 26 August; published online 26 October 2016; doi:10.1038/nm.4190

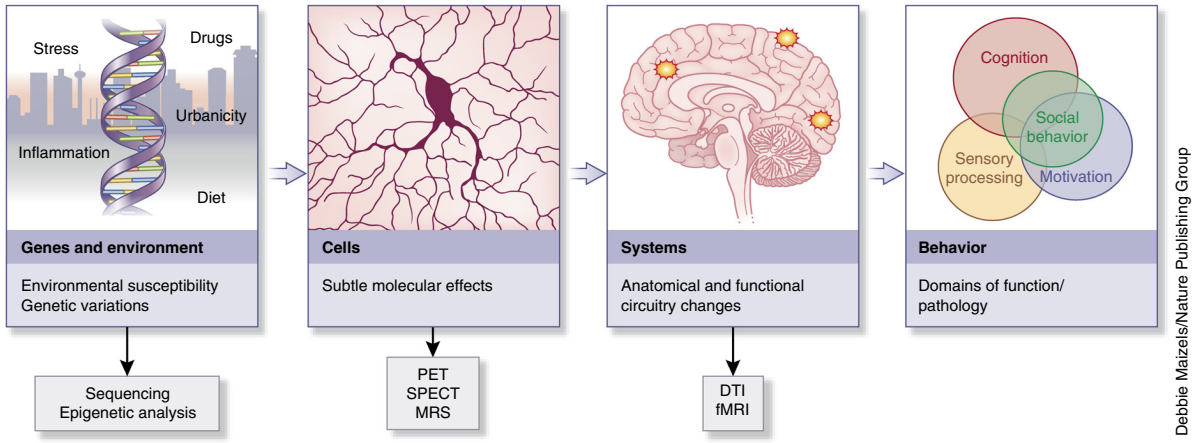

Figure 1 Genetic and imaging biomarkers. Imaging biomarkers can be cell- or circuit-based, imaged with MRI or PET modalities and represent intermediate biomarkers between genes and various domains of behavior. Adapted, with permission, from **Figure 1** in Birnbaum, R. & Weinberger, D.R. Functional neuroimaging and schizophrenia: a view towards effective connectivity modeling and polygenic risk. *Dialogues Clin Neurosci.* **15**, 279–289 (2013). (*Dialogues in Clinical Neuroscience*, © Institut La Conférence Hippocrate, Suresnes, France.)

pathways and circuits, whereas fMRI during rest or task performance affords characterization of functional circuits.

Although the RDoC is unlikely to replace current categorical classifications of mental illness in the short term, it can aid in the search for biomarkers. The RDoC approach might lead to the identification of neuroimaging markers of cognitive functions, such as reward learning or working memory, that might be impaired across psychiatric conditions. For instance, we recently described distinct neural patterns of activation in the dorsolateral and medial prefrontal cortex during a working-memory task that were disrupted in patients with schizophrenia in proportion to reductions in working-memory capacity[7]. Distinct markers of a working-memory deficit in psychiatric disorders (including but not limited to schizophrenia or any given DSM category) could represent different pathophysiological pathways that lead to a common behavioral deficit, and could thus help with the therapeutic selection of treatments that target specific pathways. This strategy is most useful for cases in which the treatment goals are to improve the cognitive dysfunction *per se*, rather than to treat the disorder as a whole.

Identification of pathological features. The second major challenge for the identification of imaging biomarkers for psychiatric disease is that many of the pathological features of psychiatric disease may be subtle and hence elusive to neuroimaging. It might be that the brains of individuals with psychiatric disorders appear typical at examination, but exhibit pathological phenotypes when 'at work,' which makes task-based assessment or other challenge paradigms an essential tool for revealing the characteristic patterns that could lead to the development of biomarkers.

The discovery of biomarkers is also limited by the availability of imaging tools, including the development of novel tracers for uncharted molecular targets. For instance, imaging of neuroinflammation became possible only recently, following the development of specific tracers for inflammation, after which reports of changes in these markers in major depressive disorder (MDD)[8] and schizophrenia[9,10] (but see ref. 11) rapidly started to emerge in the literature.

However, the paucity of replicated imaging findings is an important issue to consider here. One explanation is that attempts at replication are seldom the focus of imaging studies. The field rewards novelty over replication. Second, the neuroimaging field remains in the mechanistic discovery phase, wherein more effort is focused on uncovering alterations in imaging measures than on pursuing promising biomarkers, and so studies are not typically designed for the latter. Finally, a lack of statistical power is often an issue. To generate a larger pool of potential biomarkers, it may be essential that the need for replication using identical paradigms in well-powered studies be accepted as the norm.

In general, the field of psychiatric research has few potential biomarkers, and a more systematic search is needed to uncover additional candidates. This could emerge from initiatives similar to the precision-medicine study of 1,000 volunteers[1]. Systematic imaging of a critical number of patients with a specific DSM diagnosis with several imaging modalities across sites, for instance, could be a promising strategy to spur the search for and development of imaging biomarkers.

Validation of biomarkers. A third challenge of developing imaging biomarkers for psychiatric disease relates to the validation of biomarkers, which conventionally entails the comparison of a prediction with an actual outcome, be it diagnostic, histologic or therapeutic. Validation requires correlative analyses of *in vivo* imaging measures against *in vitro* measures such as postmortem histological examinations of brain tissue; although, as discussed above, the latter is typically unavailable for psychiatric illnesses. Longitudinal follow-up—which in psychiatry tends to stand in for an objective gold standard of outcome, be it diagnostic, therapeutic or functional—may be needed to establish a final diagnosis or observed clinical outcome that can be compared to the biomarker prediction. Moreover, for a biomarker to be useful for clinical practice, it needs to have an acceptable level of sensitivity, specificity and predictive value (see **Box 1**), to be easily accessible and practically feasible, easily quantifiable and cost effective.

Standardization of neuroimaging methods

To advance the discovery and the validation of imaging biomarkers, the field needs to adopt methods that are simple, reliable and easy to implement. Regardless of the specific method used, a database of healthy individuals who do not have psychiatric disorders may be first collected from each scanner to provide normative values and derive thresholds that separate health from illness; these thresholds can be

Box 1 Definition of relevant terms for biomarker development

Sensitivity (true positive rate) refers to the proportion of individuals who test positive among those who have the outcome of interest (i.e., test +/outcome +).

Specificity (true negative rate) refers to the proportion of individuals who test negative among those who do not have the outcome of interest (i.e., test –/outcome –).

Positive predictive value (PPV) refers to the proportion of individuals who have the outcome of interest among those who tested positive (i.e., outcome +/test +).

Negative predictive value (NPV) refers to the proportion of individuals who do not have the outcome of interest among those who tested negative (i.e., outcome –/test–).

Internal validity refers to the ability to claim that the measure of interest in a study measures the intended feature in an unbiased way and without the influence of confounding third variables (for example, that brain volumes are related to a particular diagnosis rather than to the treatment associated with that diagnosis).

External validity refers to the ability to extrapolate the results of a study to the general population of interest (that is, to real-life clinical situations). Demonstrating external validity typically requires the replication of study results in naturalistic samples independent from the original study sample.

Reliability refers to the consistency of a measure, either with itself when administered in several occasions (test–retest reliability) or in the appraisals of this measure across several raters (inter-rater reliability). Note that a reliable measure could still be invalid, for instance, when measuring consistently an unintended feature (for example, treatment) associated with the intended one (for example, diagnosis).

used for new scans to assign a likelihood of a diagnosis or outcome. To optimize the signal-to-noise ratios on each scanner and test the reproducibility across sites, this approach can then be implemented across sites and tested with the aid of imaging phantoms, which are models of brain anatomy or chemical composition, as well as traveling healthy controls for cross-site comparison purposes. To ensure standardized procedures, there should be a required protocol that includes performance of quality-assurance tests on the scanner before data acquisition is carried out, as well as specific analysis software for data processing. Specific protocols for subject preparation might be required—for instance, fasting for a certain period of time before scanning. The final data sets may be de-identified and shared across sites for centralized analyses to confirm measurements obtained by different raters or sites, an operation that may be implemented on a regular basis. See http://adni.loni.usc.edu/methods/pet-analysis/pre-processing/ for an illustration of the steps taken in the Alzheimer's Disease Neuroimaging Initiative (ADNI) study to provide homogeneous data acquisition and analysis across sites. Finally, biomarker development might require studies designed *a priori* as large-scale, multisite studies with a coordinated analysis plan that includes independent validation, rather than *post hoc* sharing of data sets acquired from multiple studies with differing goals.

Molecular imaging (PET) methods. Molecular PET imaging consists of the use of radiotracers to image specific targets in the brain by administering these radiotracers to subjects and collecting the radioactive signal produced by the radiotracers with a PET scanner. Radiotracers are designed to bind to target molecules such as neuroreceptors, re-uptake transporters or intracellular enzymes or act as substrates for the metabolic pathways of endogenous substances. The radiotracer is injected into a subject lying in the PET scanner. The concentration of the radiotracer in local brain volumes is then inferred and statistically fitted to mathematical models of the observed biological processes. Parameter estimates can then be derived to characterize the processes quantitatively. The imaging data can range from single, static snapshots of the tracer distribution, obtained from 10 to 20 min of scanning, to sequential dynamic images that record the kinetic profile of the tracer over a course of hours. Some mathematical models of the imaging data (comprehensive models) require the drawing of blood samples from the subject to estimate the radioactivity that relates only to the parent tracer crossing the blood–brain barrier (BBB) over time. For a more comprehensive discussion, see ref. 12.

In the case of PET biomarker development, ideally, a simplified bloodless method with a short scanning period—under 30 min—would first be validated against a gold-standard method on the basis of a comprehensive model that includes arterial sampling and full kinetic analysis. Bloodless methods have advantages because they bypass many sources of error and noise that relate to plasma and metabolite analysis needed for comprehensive models and forgo the need for expensive equipment to measure these. However, a key prerequisite for bloodless radiotracer methods is the presence of a reference brain region that is devoid of receptors or other targets of interest, which serves as a 'surrogate measure' of radiotracer availability. Selectivity of the tracer in the region of interest for the specific target and the feasibility of identifying a region with precise contours by visual inspection or the application of anatomical masks would also be an ideal scenario for molecular imaging of biomarkers.

Functional and structural MRI methods. MRI refers to a family of imaging techniques that use magnetic fields and radio waves to excite hydrogen atoms in water-containing tissues and read out signals that these atoms emit back, depending on the magnetic properties of the tissue. The changes in the signals are a result of tissue characteristics (such as proton density or changes in magnetic susceptibility that result from changes in blood oxygenation), and they allow for interpretations about brain morphology, integrity and neural activity. MRI techniques are commonly noninvasive and do not involve radioactivity. Structural MRI provides an anatomical image that delineates different tissues (for example, gray matter, white matter and cerebrospinal fluid) and brain structures. Functional MRI measures fluctuations in blood flow or oxygen level at rest or during a challenge, cognitive (for example, a memory test) or otherwise (for example, a drug challenge).

Validation steps to ensure that MRI techniques measure the intended anatomical or functional features are always required. These may be as varied as prediction of neuronal counts from structural techniques or capturing of individual performance or preferences

from a task-based functional technique. Regardless of the method, reliability is crucial for standardization. New initiatives of collaboration across academia and the pharmaceutical industry to accelerate the development of new therapeutics for severe mental illness, such as the NEWMEDS consortium, have emphasized the development of reliable fMRI paradigms[13]. Fully automated, standardized methods using a unified software platform are typically easier to implement across different sites, although manual methods might be better suited for certain purposes. Although standardized pipelines for preprocessing of MRI data exist and continue to become increasingly sophisticated as improved sequences are developed (for example, the recent advances linked to the Human Connectome Project)[14], no single pipeline is likely to become a one-size-fits-all approach that covers all needs for biomarker development. Instead, any given biomarker should include a detailed protocol that specifies everything from sequence parameters and data-collection protocol to preprocessing protocol and analysis to derive robust outcome measures. Postprocessing methods to minimize motion artifacts are also crucial, especially for resting-state fMRI, whereby design-predicted time-courses associated with task manipulations are not available for explicitly modeling signal dynamics.

Practical considerations. Imaging biomarkers require the availability of well-functioning imaging centers and could cost up to several thousand dollars per scan. Currently, an MRI scan costs around $600 per hour in academic centers and might be $1,000 per hour in commercial centers in the United States of America. PET scans range from $3,000 to $5,000 per scan (including radiotracer production and scanning costs), although this varies across centers. Although implementation feasibility is a relevant concern, it is hard to predict the price and availability of neuroimaging procedures in the near future. For instance, MRI scanning technologies are becoming accessible at most medical centers. Furthermore, the rapid development of technology suggests that automated analysis methods that might seem to be an infeasible addition to clinical practice today could soon be available through internet-based offsite systems. Thus, the field should focus on the development of robust biomarkers, even if their application is limited initially to a handful of specialized centers, because these might give way to a second generation of more affordable and accessible biomarkers that rely on new technological developments.

Finally, cost effectiveness, rather than cost, should be the priority. Cost effectiveness will be determined by the clinical usefulness of the biomarker in avoiding additional expenses related to misdiagnosis, unnecessary procedures or hospitalizations and other disease-related burden.

Current potential biomarkers in psychiatry

Reports of neural abnormalities in psychiatric disorders abound in the clinical neuroimaging literature, a number of which have been replicated by independent research groups. Typically, these neural abnormalities consist of statistically significant deviations from normality in a given neural feature or phenotype. For instance, an average thinning of the cortical mantle in the prefrontal cortex, a part of the brain associated with decision-making, self-control and other higher-order cognitive functions, was reported in a group of adult patients with MDD, as compared to a sociodemographically matched control group of healthy individuals[15]. Such a finding means that patients with MDD tend to display cortical thinning of the prefrontal cortex. However, in and of itself, this finding might be completely uninformative regarding clinical outcomes of individual patients. A neural finding in a clinical population becomes a true biomarker only if it can be used as an accurate proxy for some clinically relevant outcome, such as diagnosis, prognosis or treatment response, i.e., if it has sufficient clinical predictive value. The majority of clinical neuroimaging reports of neural abnormalities in clinical populations do not assess predictive value, and as a consequence, very few of the replicable abnormalities in psychiatric conditions can be regarded as actual biomarkers with potential clinical utility.

That said, the neuroimaging literature has provided important insights into the pathophysiology of mental disorders. Meta-analyses of patients with schizophrenia provide evidence that supports a pathological increase in dopamine storage and release capacity in presynaptic dopamine neurons[16], an increase that is correlated with the severity of psychotic symptoms. Furthermore, this was shown to predict response to treatment of psychotic symptoms, a finding replicated by examining responders in comparison to nonresponders[17]. This example illustrates a potential approach for screening patients who could benefit from drugs that target the dopaminergic system. This approach might become useful in the future once we better understand the biology and have treatments that target nondopaminergic aspects of psychosis.

Other replicated findings from patients with this disorder include an increase in blood volume in the cornus ammonis region 1 (CA1 region of the hippocampus[18,19]), a key region involved in memory, which has been shown to predict conversion to schizophrenia and could thus potentially serve as an indicator of risk. Studies of schizophrenia have also revealed deficient increases in hemodynamic responses during reward anticipation in the ventral striatum[20], which may be relevant for some symptoms of the disorder, such as a lack of motivation. In MDD, treatment resistance has been linked to increased blood flow in the subgenual anterior cingulate[21,22], a neural phenotype that has been targeted using deep brain stimulation (DBS), with some promising results[23]. Among other findings, dopamine-receptor availability in the striatum and gray matter volume in the ventromedial prefrontal cortex, both of which are regions involved in learning and decision-making, are reduced in patients with drug addiction[24–27], as compared to healthy individuals, and may predict treatment failure[28]. Patients with obsessive-compulsive disorder (OCD) consistently show increased volume of the striatum[29]. Patients with various anxiety disorders display increased hemodynamic responsiveness of the amygdala to negative emotional stimuli, whereas patients with post-traumatic stress disorder show a specific decrease of activity in various prefrontal regions associated with the regulation of emotion[30].

Thus, progress in understanding the neurobiological mechanisms of psychiatric illness afforded by modern neuroimaging techniques is unquestionable and its potential ever growing. However, we would argue that none of the aforementioned findings should yet be considered a biomarker proper, and that mechanistic and biomarker-oriented research should not be considered as interchangeable lines of research. Although, ideally, biomarkers would derive from neurobiologically and mechanistically interpretable findings, this might not always be necessary at first, as long as biomarkers are rigorously validated by following the steps discussed below (to draw a parallel with drug development, serendipitously discovered drugs with proven clinical effectiveness may be incorporated into clinical practice before their biological mechanisms are fully understood).

A framework for the development of neuroimaging biomarkers in psychiatry

The ultimate goal of developing neuroimaging biomarkers is to aid clinical practice in real-world settings. A hypothetical example would

be using a structural MRI scan for a patient with OCD to decide whether to initiate pharmacotherapy with fluvoxamine or cognitive-behavioral therapy as a first line of treatment—an important decision, given that prompt treatment with the most effective and best-tolerated option can carry substantial sociopersonal benefits and reduce morbi-mortality. In this example, the treatment decision would not necessarily be based on a subjective reading by a trained radiologist, but it could instead rely on a quantitative readout from a computer algorithm that exploits multiple scan features (for example, gray matter volumes in amygdala, orbitofrontal cortex and striatum) to maximize its predictive accuracy.

Identifying a clinically relevant question. Similarly to the widely accepted standards for drug discovery, the development of clinically useful neuroimaging biomarkers in psychiatry will probably require multiple steps (**Fig. 2**). A first crucial step is to define a clinically relevant question that, if addressed, could potentially lead to improvements in patients' long-term functioning and quality of life. That is, biomarkers should be developed for clinical tests that are worth ordering by clinicians by virtue of their ability to change clinical practice—biomarkers eventually need to translate into actionable tests[31]. Examples of this would be prediction of conversion to full-blown psychosis in individuals at clinical high risk (such as those who have family history of psychotic disorders and present with attenuated psychotic symptoms, such as unusual beliefs), which would be important in eventually developing prevention strategies for those individuals with the highest conversion risk while sparing those with lower conversion risk to avoid unnecessary treatment side effects, and virtually all questions related to treatment selection. Whereas relevant questions will probably be related to clinical outcome, differential diagnosis and treatment selection, diagnosis of a major psychiatric disorder such as schizophrenia or MDD (versus health) might be less relevant in a clinical setting, where treatment-seeking patients presenting with disruptive psychotic or depressive symptoms will, by definition, be ill. By the same token, differential diagnosis of treated patients, rather than of untreated patients, upon their clinical presentation is less likely to be a priority for biomarker development.

Few reports on imaging biomarkers under development focus on clinically relevant questions. Among those, a landmark MRI study aimed to predict longitudinal clinical outcome in patients at high clinical risk for psychosis[32]. In this study, a machine-learning algorithm using morphometric gray matter features from a structural MRI scan was able to predict conversion to psychotic disorders with positive and negative predictive values over 80%. This and other hypothetical examples, such as the stratification of patients who have a certain cognitive-dysfunction profile or receiving add-on cognitive training independently of diagnosis and concomitant treatments, make it apparent that clinical relevance does not require respecting current diagnostic boundaries, such as those based on the DSM diagnostic categories. Furthermore, the lack of histological gold standards is entirely circumvented by focusing on pragmatic questions such as clinical outcome and treatment stratification, the validity of which can be assessed via longitudinal designs.

To have potential clinical utility, biomarkers will need to demonstrate, even at early stages of development, that they provide useful information over and above clinical and sociodemographic data that are collected routinely in clinics. For instance, a combination of clinical factors improves prediction of conversion to psychosis in individuals at clinical high risk with approximately 80% positive predictive value[33]. This finding further raises the bar for new biomarkers

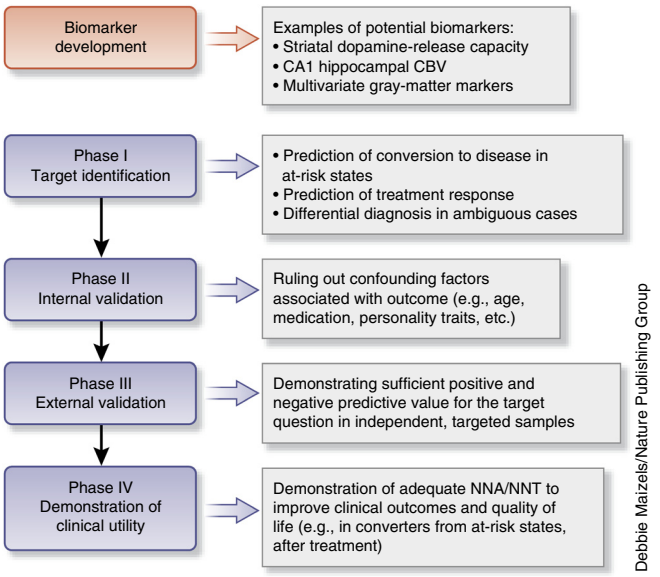

Figure 2 Steps for biomarker discovery. Left, flowchart outlining steps needed for biomarker development. Right, examples of potential biomarkers (see text for other examples) and goals associated with each step. CA1, cornus ammonis region (region I) of hippocampus. CBV, cerebral blood volume.

in this area, which would ideally need to demonstrate improved predictive power when added to such a combination of clinical factors, or at least show that the variance in the predicted outcome explained by the biomarker does not overlap with other clinical markers that are more easily accessible and just as reliable. A cautionary tale on this issue came from the ADHD-200 Global Competition: although the aim of this competition was to develop imaging biomarkers for diagnostic classification of attention-deficit hyperactivity disorder (ADHD) using a large functional and structural MRI data set, the best classification approach used solely personal characteristics (such as age, sex, handedness and IQ) and none of the available imaging data[34].

Ensuring the biomarker measures the intended biological process. The second step would involve ensuring that the biomarker is operating on brain-based phenotypes directly associated with the physiological mechanisms of interest, rather than on epiphenomenal consequences of the illness or its treatment[35]. In other words, biomarkers need to show internal validity (**Box 1**). For example, if a biomarker can accurately predict diagnosis of bipolar disorder versus unipolar depression by capitalizing on the neuroplastic changes associated with lithium—a first-line treatment for bipolar disorder—then its predictions would be equivocal when used to aid with differential diagnosis in an untreated patient with an unclear clinical presentation. Some evidence indeed suggests that classification algorithms are sensitive to treatment effects[36].

Demonstrating the biomarker's predictive value. The third step required to validate a biomarker externally is to show that it has sufficiently high predictive value to be clinically useful, beyond simple demonstration of statistically significant effects (**Fig. 3**). Whereas statistically significant differences in the distribution of a biological feature between health and disease mean that the probability of finding a difference of the observed size by chance is unlikely (for example,

less than 5%), the amount of overlap between the groups is clinically more relevant because it determines the percentage of individuals who are actually ill or healthy when they are classified as such—the positive and negative predictive value, respectively (**Box 1**).

Here, however, how high a 'sufficiently high' predictive value is depends on the specific clinical question; although a biomarker would typically be required to have positive and negative predictive values greater than 90%, in some scenarios in which the standard of care is based on an arbitrary decision between two comparable alternatives, a modestly predictive test might be clinically impactful[31]. This third step will necessarily involve external cross-validation in an independent clinical sample of adequate size. Specifically, this validation sample (or test set) needs to be fully independent from the discovery sample (or training set) used to develop the biomarker algorithm. This is necessary, particularly with multivariate classifiers, to avoid overfitting of the data on which the algorithm was trained and poor performance of this algorithm on new data (such as data from new patients in clinical settings where the established algorithms need to be most accurate), a risk that can also be decreased by using larger sample sizes for a given number of variables[37].

Existing proof-of-concept studies suggest that moderate-to-high predictive value for diagnostic classification in independent samples, not only between patients and healthy controls, but also between different patient groups (for example, schizophrenia versus bipolar disorder), can be achieved on the basis of morphometric features derived from structural MRI scans using supervised and semi-supervised machine-learning methods[38,39] (**Fig. 3**, for an example of a multivariate classification based on machine learning). We propose that only once a potential biomarker has been externally validated in an independent sample should it be considered as a biomarker.

Demonstration of clinical utility. Finally, a fourth step will involve longitudinal designs in which biomarkers are actually put to a definitive test to establish their clinical utility. Longitudinal designs allow for the establishment of a final diagnosis for patients with unclear presentation to confirm biomarker-based diagnoses. They can also be used to test whether biomarker-based treatment selection in randomized controlled trial (RCT) designs is superior to the assignment of treatment according to standard clinical practice. At that stage, the number needed to treat (NNT) or number needed to assess (NNA), which represent the number of individuals that would need to undergo a certain procedure to benefit one individual, can be useful in synthesizing the potential utility of the biomarker. Only at that point will a clinically useful biomarker (for example, one with low NNT or NNA) be able to make its way into real-world clinical settings. Even after this bench-to-bedside leap, however, studies in real-world settings will be required to definitively establish the utility and cost-effectiveness of biomarkers in the real world, for instance, by demonstrating that their use is associated with a reduction in morbidity and improvement in quality of life for the general population.

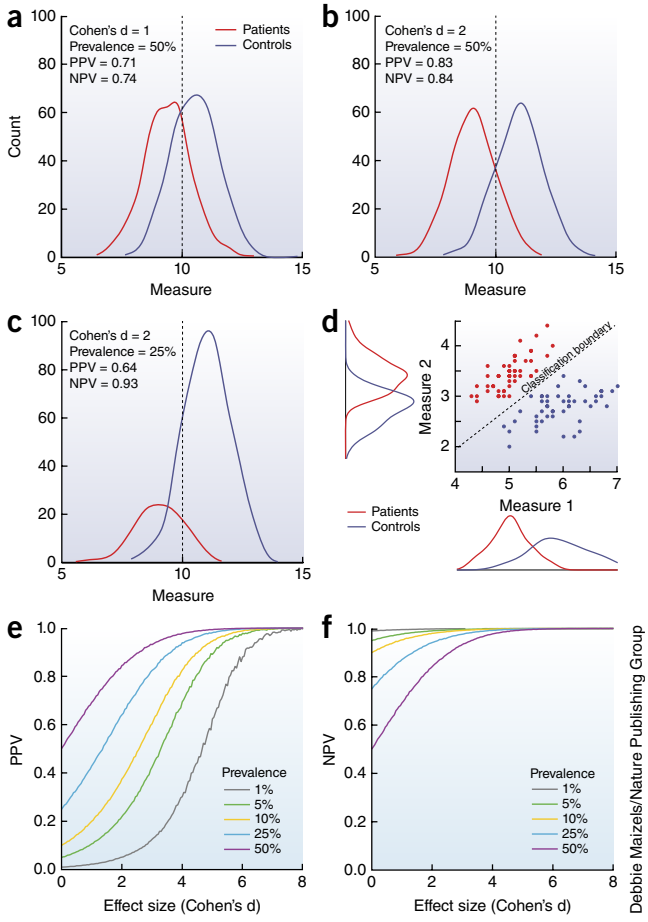

Figure 3 Statistically significant findings versus clinically useful biomarkers. (**a**–**c**) Histograms reflecting the distributions of a simulated imaging measure (*x* axis) among two groups (in this example, patients in red and healthy controls in blue). Illness prevalence of 50% (with effect sizes of 1 and 2, respectively) (**a**,**b**), and distributions corresponding to a prevalence of 25% (with an effect size of 2) (**c**). Positive and negative predictive values (PPV and NPV, respectively) are reported. There is substantial overlap of groups in **a**, with moderate PPV and NPV despite a large effect size of 1. There is decreased overlap, with an even larger effect size of 2, in **b** and **c**, but PPV and NPV depend on the prevalence of the predicted outcome. (**d**) A scatterplot of distributions of two simulated imaging measures (*x* and *y* axes) with marginal histograms corresponding to each of the two univariate distributions. A multivariate classifier (in this case, a linear support vector machine) was trained to maximally separate the two groups by simultaneously using both imaging measures. Whereas the univariate distributions for both measures overlap substantially between groups, the multivariate classifier is able to find a boundary that perfectly separates these groups. (**e**,**f**) PPV (**e**) and NPV (**f**) for simulated examples wherein the effect size and illness prevalence were varied systematically. Of note, whereas effect size has a monotonic effect on PPV and NPV, prevalence modifies the midpoint and slope of the sigmoidal relationship between effect size and these measures.

Future directions for biomarker development in psychiatry

The long-awaited arrival of neuroimaging-based biomarkers for psychiatric disorders would represent a historic paradigm shift in the biomedical sciences, one with several fundamental ramifications extending from the social perception of mental illness to the modernization of psychiatric practice. Thus, the development of clinically useful biomarkers should be a top priority of contemporary mental-health research, as the strategic objectives laid out by the NIMH already reflect (http://www.nimh.nih.gov/about/strategic-planning-reports/strategic-objective-2.shtml). In addition, this is congruent with the need recognized by the Food and Drug Administration (FDA) and the European Medical Agency (EMA) for biomarkers that predict clinical response or the emergence of side effects in drug trials[40]. The sheer scale of biomarker development calls for collaborative efforts across research sites and will require support by appropriate funding mechanisms designed *a priori* for this purpose. Along these lines, although publicly available databases of

imaging data might be a helpful start for method development and proof-of-concept studies, clinical samples for biomarker development should, first and foremost, be targeted to particular questions of clinical relevance (for example, a large data set of patients with chronic, treated schizophrenia would be unhelpful for developing a biomarker aimed at treatment selection or differential diagnosis for first psychotic episodes).

Biomarker development should employ a cost-effective platform that strikes a balance between hypothesis-driven development, aimed at exploiting well-replicated imaging phenotypes with the potential to result in useful biomarkers, and purely data-driven but nonetheless powerful approaches aimed at 'blindly' exploring multimodal data sets. The latter might include panels of multimodal imaging, genetic, clinical and other data. One timely example of the former is the emerging field of imaging neuroinflammation, which offers the prospect of a clinically useful and therapeutically actionable biomarker that might eventually prompt targeted immuno-modulatory treatments that cut across conventional diagnostic categories. Although specific neuroimaging modalities (such as resting-state versus task-based fMRI, or fMRI versus PET measures of blood flow) might be more feasible and accessible in general, and although some neurobiological processes (such as the dopamine system) might be better understood than others, biomarker discovery should be tailored to the clinical question under examination rather than abiding by generic prescriptions; it should exploit reliable leads from previous research while, at least initially, acknowledging gaps in knowledge and the need for further discovery.

To achieve useful biomarkers, we need mechanisms by which an expert consensus can be quickly reached in terms of optimal tracers, optimal cellular targets and quantification methods, and by which multisite testing across diagnostic groups is expedited. An example of the latter could be the use of multivariate tools that combine multiple measures within a modality (for example, combining PET measures of enhanced striatal dopamine release, a well-replicated marker of psychosis, and of extrastriatal dopamine release deficit[41] in a single classifier to increase the power to detect cases that have abnormal dopamine transmission) or across imaging modalities and nonimaging data, including, but not limited to, clinical, sociodemographic, genetic and biochemical data.

Once clinically targeted data sets of biomarker panels are collected, open scientific competitions could be a way to accelerate selection of the best performing algorithms[34]. Nationwide initiatives organized by the NIH, modeled after pioneering large-scale projects such as the North American Prodrome Longitudinal Study (NAPLS)[42], or international collaborative projects will thus probably be needed to spur the development of targeted biomarkers that have real potential to revolutionize psychiatric practice.

ACKNOWLEDGMENTS

G.H. is funded by the US National Institutes of Health (grant K23MH101637).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Collins, F.S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
2. Normanno, N. & Cree, I.A. Genomics driven-oncology: challenges and perspectives. *BMC Cancer* **15**, 141 (2015).
3. Kalia, M. Biomarkers for personalized oncology: recent advances and future challenges. *Metabolism* **64** (Suppl. 1), S16–S21 (2015).
4. El Achkar, C.M., Olson, H.E., Poduri, A. & Pearl, P.L. The genetics of the epilepsies. *Curr. Neurol. Neurosci. Rep.* **15**, 39 (2015).
5. Soni, A. in *The Five Most Costly Conditions, 1996 and 2006: Estimates for the U.S. Civilian Non-institutionalized Population* 1–3 (Agency for Healthcare Research and Quality, 2009).
6. Kessler, R.C. *et al.* Individual and societal effects of mental disorders on earnings in the United States: results from the national comorbidity survey replication. *Am. J. Psychiatry* **165**, 703–711 (2008).
7. Van Snellenberg, J.X. *et al.* Mechanisms of working memory impairment in schizophrenia. *Biol. Psychiatry* **80**, 617–626 (2016).
8. Setiawan, E. *et al.* Role of translocator protein density, a marker of neuroinflammation, in the brain during major depressive episodes. *JAMA Psychiatry* **72**, 268–275 (2015).
9. Doorduin, J. *et al.* Neuroinflammation in schizophrenia-related psychosis: a PET study. *J. Nucl. Med.* **50**, 1801–1807 (2009).
10. van Berckel, B.N. *et al.* Microglia activation in recent-onset schizophrenia: a quantitative (R)-[11C]PK11195 positron emission tomography study. *Biol. Psychiatry* **64**, 820–822 (2008).
11. Kenk, M. *et al.* Imaging neuroinflammation in gray and white matter in schizophrenia: an *in vivo* PET study with [18F]-FEPPA. *Schizophr. Bull.* **41**, 85–93 (2015).
12. Innis, R.B. *et al.* Consensus nomenclature for *in vivo* imaging of reversibly binding radioligands. *J. Cereb. Blood Flow Metab.* **27**, 1533–1539 (2007).
13. Plichta, M.M. *et al.* Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *Neuroimage* **60**, 1746–1758 (2012).
14. Glasser, M.F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
15. Schmaal, L. *et al.* Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* http://dx.doi.org/10.1038/mp.2016.60 (2016).
16. Howes, O.D. *et al.* The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Arch. Gen. Psychiatry* **69**, 776–786 (2012).
17. Demjaha, A., Murray, R.M., McGuire, P.K., Kapur, S. & Howes, O.D. Dopamine synthesis capacity in patients with treatment-resistant schizophrenia. *Am. J. Psychiatry* **169**, 1203–1210 (2012).
18. Schobel, S.A. *et al.* Imaging patients with psychosis and a mouse model establishes a spreading pattern of hippocampal dysfunction and implicates glutamate as a driver. *Neuron* **78**, 81–93 (2013).
19. Schobel, S.A. *et al.* Differential targeting of the CA1 subfield of the hippocampal formation by schizophrenia and related psychotic disorders. *Arch. Gen. Psychiatry* **66**, 938–946 (2009).
20. Radua, J. *et al.* Ventral striatal activation during reward processing in psychosis: A neurofunctional meta-analysis. *JAMA Psychiatry* **72**, 1243–1251 (2015).
21. Mayberg, H.S. *et al.* Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am. J. Psychiatry* **156**, 675–682 (1999).
22. Seminowicz, D.A. *et al.* Limbic-frontal circuitry in major depression: a path modeling metanalysis. *Neuroimage* **22**, 409–418 (2004).
23. Mayberg, H.S. *et al.* Deep brain stimulation for treatment-resistant depression. *Neuron* **45**, 651–660 (2005).
24. Fritz, H.C. *et al.* Current smoking and reduced gray matter volume—a voxel-based morphometry study. *Neuropsychopharmacology* **39**, 2594–2600 (2014).
25. Konova, A.B. *et al.* Structural and behavioral correlates of abnormal encoding of money value in the sensorimotor striatum in cocaine addiction. *Eur. J. Neurosci.* **36**, 2979–2988 (2012).
26. Martinez, D. *et al.* Cocaine dependence and d2 receptor availability in the functional subdivisions of the striatum: relationship with cocaine-seeking behavior. *Neuropsychopharmacology* **29**, 1190–1202 (2004).
27. Martinez, D. *et al.* Deficits in dopamine D(2) receptors and presynaptic dopamine in heroin dependence: commonalities and differences with other types of addiction. *Biol. Psychiatry* **71**, 192–198 (2012).
28. Martinez, D. *et al.* Imaging dopamine transmission in cocaine dependence: link between neurochemistry and response to treatment. *Am. J. Psychiatry* **168**, 634–641 (2011).
29. Radua, J., van den Heuvel, O.A., Surguladze, S. & Mataix-Cols, D. Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs other anxiety disorders. *Arch. Gen. Psychiatry* **67**, 701–711 (2010).
30. Etkin, A. & Wager, T.D. Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatry* **164**, 1476–1488 (2007).
31. Perlis, R.H. Translating biomarkers to clinical practice. *Mol. Psychiatry* **16**, 1076–1087 (2011).
32. Koutsouleris, N. *et al.* Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch. Gen. Psychiatry* **66**, 700–712 (2009).
33. Cannon, T.D. *et al.* Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Arch. Gen. Psychiatry* **65**, 28–37 (2008).
34. Brown, M.R. *et al.* ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Front. Syst. Neurosci.* **6**, 69 (2012).

35. Horga, G., Kaur, T. & Peterson, B.S. Annual research review: Current limitations and future directions in MRI studies of child- and adult-onset developmental psychopathologies. *J. Child Psychol. Psychiatry* **55**, 659–680 (2014).
36. Marquand, A.F. *et al.* Dissociable effects of methylphenidate, atomoxetine and placebo on regional cerebral blood flow in healthy volunteers at rest: a multi-class pattern recognition approach. *Neuroimage* **60**, 1015–1024 (2012).
37. Huys, Q.J., Maia, T.V. & Frank, M.J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **19**, 404–413 (2016).
38. Bansal, R. *et al.* Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. *PLoS One* **7**, e50698 (2012).
39. Schnack, H.G. *et al.* Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage* **84**, 299–306 (2014).
40. Laughren, T.P. What's next after 50 years of psychiatric drug development: an FDA perspective. *J. Clin. Psychiatry* **71**, 1196–1204 (2010).
41. Slifstein, M. *et al.* Deficits in prefrontal cortical and extrastriatal dopamine release in schizophrenia: a positron emission tomographic functional magnetic resonance imaging study. *JAMA Psychiatry* **72**, 316–324 (2015).
42. Addington, J. *et al.* North American Prodrome Longitudinal Study (NAPLS 2): The prodromal symptoms. *J. Nerv. Ment. Dis.* **203**, 328–335 (2015).