nature neuroscience

Building better biomarkers: brain models in translational neuroimaging

Choong-Wan Woo[1–4], Luke J Chang[5], Martin A Lindquist[6] & Tor D Wager[3,4]

Despite its great promise, neuroimaging has yet to substantially impact clinical practice and public health. However, a developing synergy between emerging analysis techniques and data-sharing initiatives has the potential to transform the role of neuroimaging in clinical applications. We review the state of translational neuroimaging and outline an approach to developing brain signatures that can be shared, tested in multiple contexts and applied in clinical settings. The approach rests on three pillars: (i) the use of multivariate pattern-recognition techniques to develop brain signatures for clinical outcomes and relevant mental processes; (ii) assessment and optimization of their diagnostic value; and (iii) a program of broad exploration followed by increasingly rigorous assessment of generalizability across samples, research contexts and populations. Increasingly sophisticated models based on these principles will help to overcome some of the obstacles on the road from basic neuroscience to better health and will ultimately serve both basic and applied goals.

Translational neuroscience is a field at the intersection of basic neuroscience and clinical applications. Basic neuroscience is concerned with understanding how brain activity gives rise to thoughts, feelings and behavior, whereas clinical applications are concerned with developing tools that are useful for clinical decision-making and therapeutic development. The advent of functional neuroimaging nearly 30 years ago generated great optimism about its potential for both revolutionizing our understanding of the physical basis of mind and delivering clinically useful tools. While much progress has been made on the former goal[1], few results and models from functional neuroimaging have been incorporated into clinical practice[2]. This paper explores some of the scientific reasons why this is the case and presents a framework for moving forward.

Over these decades, a wealth of translational neuroimaging studies have identified brain features—largely measures of activity in specific brain regions—that predict health-related outcomes. These outcomes include current diagnostic categories (for example, major depressive disorder[3]), as well as measures of symptoms (for example, anhedonia[4]), cognitive and affective component processes (for example, risk aversion[5]) and cognitive performance (for example, sustained attention[6]). Such outcomes are not currently considered 'disorders', but they are features that cut across disorders and influence healthy mental function[7]. Brain correlates of these outcomes could provide a basis for reconceptualizing diagnostic categories, identifying features of neuropathology and assessing healthy brain function beyond current clinical diagnostic categories. In this sense, cognitive neuroscience and translational neuroimaging share the common goals of establishing strong associations between brain measures and both subjective experience and objective behavior.

Early translational neuroimaging efforts were based on traditional brain mapping approaches. Built on a historical foundation of lesion studies[8] and theories of modularity[9], the fundamental goal of early neuroimaging studies was to understand what functions and processes are encoded in isolated, target brain regions of interest. Standard parametric mapping scales this approach up to a massive number (50,000–350,000) of separate tests in local regions or 'voxels' to create whole-brain maps (**Fig. 1a**). This is currently the most popular approach to neuroimaging. Early translational studies likewise identified isolated brain regions important for clinical disorders and symptoms. Researchers discovered relationships between subgenual anterior cingulate cortex and depression[3], thalamus and periaqueductal gray with chronic pain[10], basal ganglia with obsessive-compulsive disorder[11] and subthalamic nucleus with Parkinson's disease[12], among many others. Many contemporary studies are still aimed at identifying brain 'hot spots' predictive of health-related outcomes[13]. These associations can be useful, if the isolated features are diagnostic of and directly related to the underlying pathology. However, several clinical trials targeting these local regions with neuromodulation therapies, such as deep-brain stimulation and neurofeedback, have failed, particularly for depression and pain[10,14,15]. In addition, as we explain below, there is strong reason to suspect that these associations do not provide a sufficiently complete description of the relevant neuropathology to be clinically actionable.

A central problem is that local brain-mapping was not designed with translational goals in mind. Its main goal is not to provide a complete model of symptoms and behavior but to test hypotheses about structure–function associations. The focus is on whether there are any effects in one or more brain regions, rather than on whether the effects are large enough to have clinical utility[16]. This is in line with traditional goals of

[1]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea. [2]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Republic of Korea. [3]Department of Psychology and Neuroscience, University of Colorado, Boulder, Colorado, USA. [4]Institute of Cognitive Science, University of Colorado, Boulder, Colorado, USA. [5]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, New Hampshire, USA. [6]Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA. Correspondence should be addressed to T.D.W. (tor.wager@colorado.edu).

Received 12 April 2016; accepted 11 December 2016; published online 23 February 2017; doi:10.1038/nn.4478

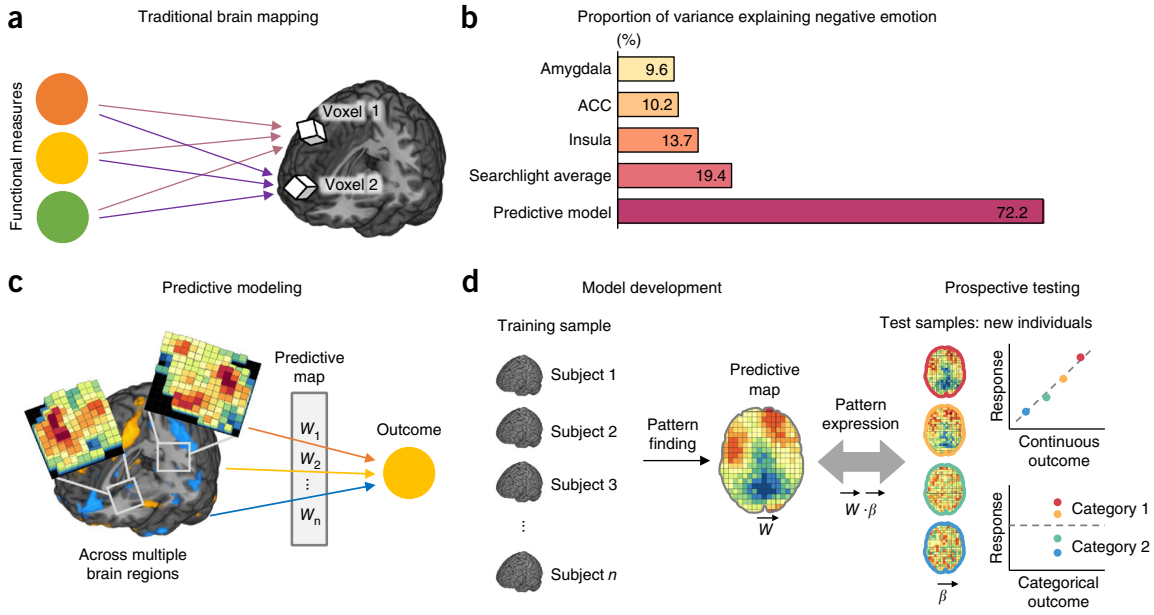

Figure 1 Standard mapping versus predictive modeling. (**a**) Traditional brain mapping, often called mass-univariate analysis or voxelwise encoding model. Brain maps are constructed by conducting massive number of tests on brain voxels one at a time. (**b**) An example showing small effect sizes (here, explained variance) when one brain region is considered in isolation and larger effect sizes for a multivariate model. Chang *et al.*[24] showed that local regions, including amygdala, anterior cingulate cortex (ACC), insula or searchlights, explained much less variance in experienced negative emotion than a whole-brain predictive model. (**c**) Predictive modeling explicitly aims to develop brain models that are tightly coupled with target outcomes. w_1, w_2, … w_n represent predictive weights across voxels. (**d**) Predictive model development and prospective testing. Here, a predictive map ($\vec{w}$) comprised of predictive weights across voxels is developed based on a training sample (i.e., a group of individuals) and tested on independent test samples (i.e., new individuals). The weights specify how to integrate brain data to produce a single prediction about the outcome, which could be continuous or categorical. In this example, calculating the dot product between the predictive map and the test images—a weighted sum of activity across the test image ($\vec{\beta}$), with the predictive map specifying the weights ($\vec{w}$)—generates a predicted outcome for each participant. The sensitivity, specificity and other properties of the predictive map are estimated from test samples. Data in **b** from Chang *et al.*[24].

understanding localized brain function but not with providing a sufficient brain-level description of a behavior. Another limitation is that a typical voxel—the smallest spatial unit of analysis—contains approximately 5.5 million neurons[17] with diverse properties and functions[18], particularly in the heteromodal brain areas most often associated with mental health disorders[19,20]. This lack of functional specificity creates problems in making inferences about mental processes, including symptoms, based on brain activity[21]. Brain mapping is designed to permit the inference that brain region B is activated conditionally on stimulus (or symptom) S and assesses the probability P(B|S). This does not allow one to make the reverse inference that stimulus S must have occurred given activation of region B, related to P(S|B) (ref. 22). The latter is what provides inferences about mental states and clinical conditions, but making such inferences requires a different analysis paradigm. Lastly, many features of neurologic and psychiatric disorders (e.g., pain, negative emotions, cognitive and social processes) are likely encoded in distributed neural systems involving networks of many regions[23,24] (**Fig. 1b**). Thus, many clinically relevant outcomes, including core features in the Research Domain Criteria (RDoC)[7], may not be measurable in isolated regions even in theory.

Recently, a new trend, predictive modeling, has emerged to address these issues. This approach uses pattern recognition techniques (or 'machine learning') to develop integrated models of activity across multiple brain regions (i.e., brain signatures) to predict clinical outcomes[25–27] (**Fig. 1c**). This approach might be referred to as 'translational neuroimaging 2.0', as it is qualitatively distinct from conventional brain mapping and has several important benefits. First, the direction of inference is reversed relative to conventional mapping: brain features (for example, structure, activity or connectivity) comprise a set of predictors, and behavioral or clinical variables comprise one or more outcomes. Second, predictive models integrate all available brain data into a single 'best guess' about the outcome, providing focused tests that avoid multiple comparisons and increase statistical power when evaluating their diagnostic utility[16]. Third, the diagnostic value of predictive models is typically tested by evaluating their performance in new, out-of-sample individuals[26] (**Fig. 1d**), providing valid estimates of effect size and clinical significance. Fourth, multivariate predictive models can capture information across multiple spatial scales[26,28,29], ranging from mesoscale information to large-scale information distributed across multiple brain systems. Much mesoscale information is encoded in functional structures smaller than a voxel, but in large neural populations distributed unevenly across voxels, allowing functional MRI (fMRI) patterns to be sensitive to this distribution (often called 'fMRI hyperacuity')[30,31]. This can result in large predictive effect sizes (and accuracy) in explaining outcomes, enhancing the models' clinical significance. This new approach is also synergistic with recent trends toward the aggregation and sharing of large-scale neuroimaging data sets[32].

In this paper, we review studies that use the predictive modeling approach to predict clinical outcomes and ask what is needed for the next generation of advances. Despite increasing interest in developing clinically useful biomarkers based on multiple neuroimaging modalities[33], many challenges still remain to be solved. These include developing sensitive and specific models that generalize across studies and heterogeneous populations; developing predictive models that build on and contribute to neuroscientific theory; and resolving issues

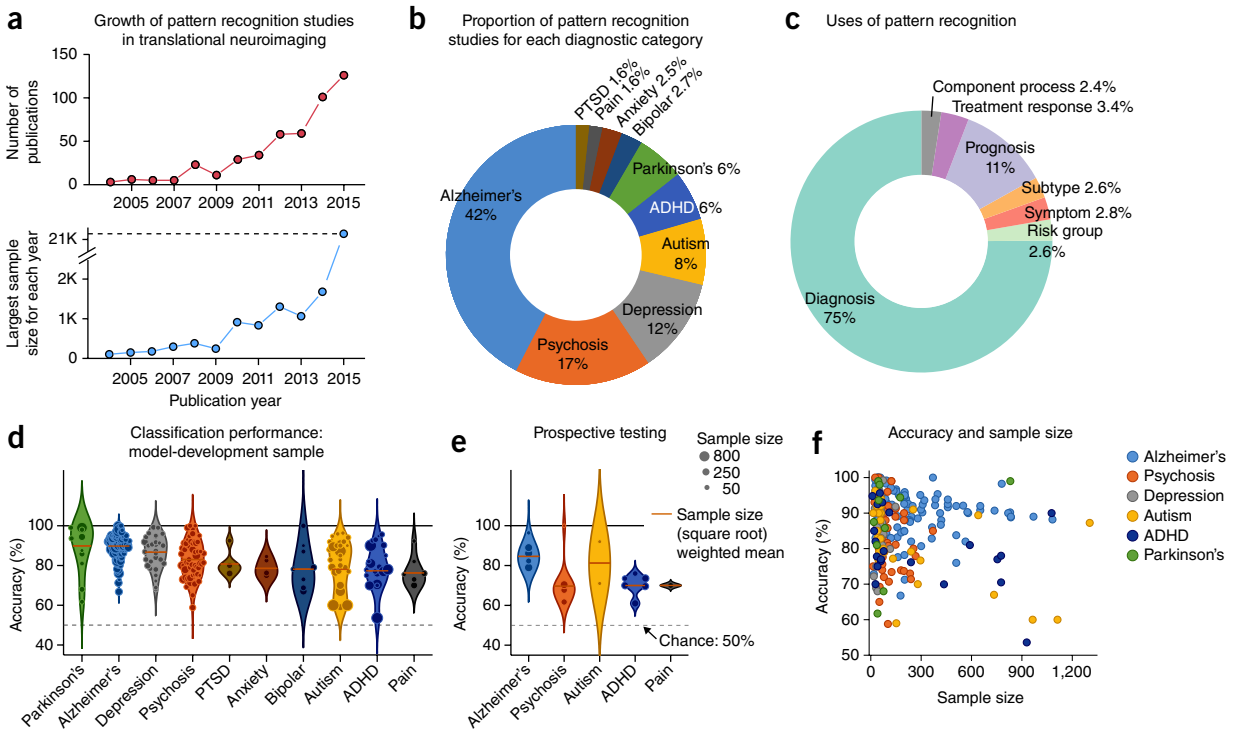

Figure 2 A snapshot of translational neuroimaging using multivariate predictive models. We searched PubMed for original neuroimaging research articles (including EEG, positron-emission topography (PET), MRI, diffusion tensor imaging (DTI) and arterial spin labeling (ASL)) published between 1983 and January 2016. The search terms can be found via this link: http://goo.gl/N7oh0i. Nonhuman and nonclinical studies were excluded, as well as those that did not employ multivariate pattern recognition. The initial search yielded 2,767 studies, of which 536 studies were selected based on review of their abstracts. Full-text review was used to select 475 studies that included 615 classification or predictive maps. (**a**) Top: growth of pattern recognition studies in translational neuroimaging since 2004. Bottom: growth of sample sizes in translational neuroimaging studies. The *y*-axis shows the largest sample size among studies published each year. (**b**) Breakdown of studies by diagnostic category. PTSD, post-traumatic stress disorder. ADHD, attention deficit hyperactivity disorder. (**c**) Uses of pattern recognition models. 'Diagnosis' refers to patient vs. control classification and 'risk group' to classification of groups at high risk (for example, relatives of people with disorders) vs. controls. 'Symptom' refers to prediction of continuous symptom scores. 'Subtype' refers to identification of subgroups of patients based on brain patterns. 'Prognosis' and 'treatment response' refer to predictions of individual differences in disease progression and response to an intervention, respectively. 'Component process' studies identify predictive models for basic cognitive or affective processes and apply those to classifying patient groups or to predicting symptoms in patients. (**d**) Precision-weighted accuracy, based on the square root of the sample size, for patient vs. control classification in model-development samples. Here we show classification accuracy only for patient vs. control classification, which was the most common use across disorders (75% of predictive models). The size of the circles shows the precision estimates, with larger circles indicating larger samples and more precise estimates. Accuracy was nearly always estimated using cross-validation. (**e**) Classification results from prospective testing on independent data sets. Only a small minority of studies report prospective tests. Lower accuracy in independent tests is indicative of bias in cross-validated accuracy estimates from training samples. Accuracy is lower in most cases reviewed here, with AD classification showing least evidence for bias. (**f**) Diagnostic classification accuracy as a function of sample size for six types of disorders. As the estimates from the largest studies are the most precise, they are most representative of the true accuracy. Across disorders, very high classification accuracy is reported in some small studies, but these have not been replicated in prospective tests. With a few exceptions, accuracy values for large-sample studies are much more modest. These observations point to the need for improvements in statistical model development, data aggregation and prospective testing of promising models across multiple, diverse samples.

with proper validation and quality control in multisite studies. Machine-learning algorithms and big data approaches will not, in themselves, be enough to address all these challenges. Rather, brain models should be developed and validated within a systematic biomarker development framework, treating brain models as sharable 'research products' that can be tested and annotated across research groups to demonstrate generalizability across samples, research contexts and populations. This new way of thinking about neuroimaging results integrates ideas from machine learning, big data, reproducible research and open science to bring translational goals within reach.

Clinical predictive modeling: state of the field

The value of predictive modeling aided by machine learning has grown rapidly for the last decades in translational neuroimaging, with over 500 papers using multivariate predictive models (**Fig. 2a**). This growth parallels a rapid and transformational development in the use of machine learning across diverse data-rich applications, including finance, genetics, security, marketing, games and information technology[34,35]. Some of the earliest work developing brain signatures for disease states came from neurology, particularly dementia studies[36,37], which emphasizes objective signs of pathology. Studies of Alzheimer's disease (AD) and related dementias remain the most prevalent, but the predictive mapping framework has been extended to other neurological disorders, such as Parkinson's disease and pain disorders, as well as to an impressive array of mental health disorders, including psychosis, depression, autism, attention deficit hyperactivity disorder (ADHD), bipolar disorder, anxiety disorders and post-traumatic stress disorder (**Fig. 2b**).

Sample sizes have also rapidly increased across the last decades, with studies on the order of 1,000 individuals appearing since 2010

Table 1 Research consortia for neuroimaging data sharing in translational neuroimaging

Resource	Initials	Clinical groups	Imaging modality	Web address
Data-sharing platforms				
International Data-sharing Initiative	INDI			http://fcon_1000.projects.nitrc.org/
Neuroimaging Informatics Tools and Resources Clearinghouse	NITRC			http://www.nitrc.org/
Laboratory of Neuro Imaging Image & Data Archive	LONI IDA			https://ida.loni.usc.edu/login.jsp
Collaborative Informatics and Neuroimaging Suite	COINS			http://coins.mrn.org/
National Alliance for Medical Image Computing	NA-MIC			http://wiki.na-mic.org/Wiki/index.php/Main_Page
Open functional Magnetic Resonance Imaging	OpenfMRI			https://openfmri.org/
Consortium or repository				
Autism Brain Imaging Data Exchange	ABIDE	Autism	rs-fMRI	http://fcon_1000.projects.nitrc.org/indi/abide/
Attention Deficit Hyperactivity Disorder-200	ADHD-200	ADHD	rs-fMRI	http://fcon_1000.projects.nitrc.org/indi/adhd200/
Alzheimer's Disease Neuroimaging Initiative	ADNI	Alzheimer's	Multimodal, including sMRI, fMRI, PET	http://adni.loni.usc.edu/; accessible through IDA
Australian Imaging, Biomarkers & Lifestyle Flagship Study of Aging	AIBL	Alzheimer's	Multimodal, including sMRI, PET	https://aibl.csiro.au/; accessible through IDA
Open Access Series of Imaging Studies	OASIS	Alzheimer's	sMRI	http://oasis-brains.org/
international Study to Predict Optimized Treatment for Depression	iSPOT-D	Depression	Multimodal, including EEG, dMRI, fMRI, sMRI	http://www.brainresource.com/home.html
MIND Clinical Imaging Consortium	MCIC	Schizophrenia	Multimodal, including dMRI, fMRI, sMRI	http://coins.mrn.org
Center of Biomedical Research Excellence	COBRE	Schizophrenia	sMRI, rs-fMRI	http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html
function Biomedical Informatics Research Network	fBIRN	Schizophrenia	fMRI, sMRI	http://www.schizconnect.org/
Parkinson's Progression Markers Initiative	PPMI	Parkinson's	Multimodal, including dMRI, fMRI, sMRI, PET	http://www.ppmi-info.org/
Pain and Interoception Imaging Network	PAIN	Pain	Multimodal, including dMRI, fMRI, sMRI	http://www.painrepository.org/
OpenPain	Openpain	Pain	fMRI	http://www.openpain.org

Note: dMRI, fMRI and sMRI refer to diffusion-weighted, functional and structural MRI, respectively; rs-fMRI refers to resting-state fMRI.

(**Fig. 2a**). These large-scale studies have been made possible by the development of research consortia committed to aggregation and sharing of data across research groups[32]. Such efforts include the Alzheimer's Disease Neuroimaging Initiative (ADNI), Autism Brain Imaging Data Exchange (ABIDE), Parkinson's Progression Markers Initiative (PPMI) and others (**Table 1**). This is a promising direction for the field, as it promotes (i) model development on large samples, which can increase statistical power; (ii) development of models based on multisite samples, which are more likely to generalize across scanners and commonly encountered variations in study procedures; and (iii) tests on independent data sets with different characteristics (for example, different population demographics).

Most studies have focused on diagnosis, identifying brain signatures that discriminate patients from healthy controls (75% of the 615 predictive models in our survey; **Fig. 2c**). These studies aim to establish objective signs of disease pathology. A common objection to these studies is that such models simply recapitulate existing diagnoses and therefore cannot move beyond them. However, the goal of such studies is not to replace existing diagnostic tools but rather to establish a meaningful neurobiological basis for the disorder of interest, supporting the development of new clinical measures and therapeutics. If a brain model does not differentiate patients from controls or predict symptoms, it is unclear whether it is a model of the relevant clinical pathology at all.

Importantly, some studies (25% of the models in our survey; **Fig. 2c**) have begun to develop brain models for more difficult classification and prediction problems not easily addressed using existing clinical measures. These include neuroimaging models for risk assessment, early detection, predicting conversion to full-scale disorders, differential diagnosis, subtyping of patients and predicting treatment response. Developing such models could be challenging but will potentially provide useful brain measures and new disorder categories that will eventually help treatments. Here we review some example studies.

Risk assessment, conversion prediction and early detection. Neuroimaging models can be used to assess who is at risk, predict who will later convert to a disease state in advance of its onset and detect patients in early stages of disease[38–42]. If successful, these models could provide a basis for early intervention, which can potentially prevent or even reverse the course of disease[43,44].

This type of research has been most active and successful in AD mainly because ADNI has collected longitudinal data from participants with mild cognitive impairment, which is a transitional state between AD and normal aging. The Spatial Pattern of Abnormality for Recognition of Early Alzheimer's Disease (SPARE-AD) index is one of the most promising models. SPARE-AD is a pattern classifier based on spatial patterns of brain atrophy measured by structural MRI[45] and indicates the presence of a brain atrophy pattern characteristic of AD. SPARE-AD scores predict subsequent cognitive decline[45,46] and transition to AD[47].

Predictive models have also been developed in some other disorders, including psychosis and depression. For example, pattern classifiers based on structural MRI were developed to discriminate individuals at risk for psychosis from healthy controls and to predict which individuals would transition to psychosis and which would not[40]. However, these models have yet to be independently validated.

Differential diagnosis and subtyping. Neuroimaging models can also be used for differential diagnosis[48–51] and patient subtyping (or 'stratification')[52–55]. These models can provide important information about the relationships among disorders and symptoms at the biological level, helping identify subgroup structures that are not reflected in current diagnostic categories but are potentially informative about treatment selection.

As many mental and neurologic disorders are comorbid, brain-based differential diagnosis can help identify distinguishing features of neuropathology and provide new ways of examining overlap across

Table 2 Named multivariate models in translational neuroimaging

Name	Initials	Predictive of	Features	Algorithms	Refs
Ordinal regression characteristic index of dementia	ORCHID	Progression of AD	sMRI	Ordinal regression with Gaussian process	38
Spatial pattern of abnormality for recognition of early Alzheimer's disease	SPARE-AD	Early detection of AD	sMRI	SVM with RAVENS methods	124
Alzheimer's disease pattern similarity	AD-PS	Risk assessment for AD	sMRI	Regularized logistic regression	125
Structural MRI-based brain amyloidosis score	sMRI-BAS	Amyloid beta positive	sMRI	Partial least squares	126
Structural abnormality index	STAND	Diagnosis of AD	sMRI	SVM	127
Parkinson's disease-related pattern	PDRP	PD status	FDG-PET	Spatial covariance analysis	95
Parkinson's disease-related cognitive pattern	PDCP	Cognitive dysfunction in PD	FDG-PET	Spatial covariance analysis	128
Parkinson's disease-related tremor pattern	PDTP	Tremor in PD	FDG-PET	Spatial covariance analysis	129
Multiple-system atrophy-related pattern	MSARP	MSA status	FDG-PET	Spatial covariance analysis	130
Corticobasal degeneration-related pattern	CBDRP	CBD status	FDG-PET	Spatial covariance analysis	131

Note: these models can be applied prospectively to new individuals and data sets to generate predictions about clinical status, symptoms and future outcomes. Their levels of diagnosticity, generalizability and ease of application vary, as do the amounts of prospective testing done to date. However, each of these constitutes a research product that can be shared and characterized across laboratories. CBD, corticobasal degeneration; MSA, multiple-system atrophy; PD, Parkinson's disease; RAVENS, regional analysis of volumes examined in normalized space; SVM, support vector machine.

disorders. For example, Tang *et al.* used fluorine-18-labeled fluorodeoxyglucose positron-emission tomography (FDG-PET) data to differentiate patients with idiopathic Parkinson's disease, multiple system atrophy and progressive supranuclear palsy[48]. Imaging-based classifiers achieved high accuracy in differential diagnosis (91–98% positive predictive value), whereas movement disorder specialists blinded to the imaging-based diagnosis reached the final clinical diagnosis only after two additional years of clinical follow-up.

Brain-based subtyping of patients can identify groups of patients, or 'biotypes', with differential disease course or treatment response[52,55]. This is an inherently difficult problem, as the 'ground truth' about how many subtypes are useful and who belongs in which group is unknown. This type of 'unsupervised' learning problem typically requires large data sets, preferably across multiple diagnoses; thus, there are only a few such studies. One study grouped psychosis patients—including 1,872 participants with schizophrenia, schizoaffective disorder, and bipolar disorder—into three transdiagnostic biotypes based on neuropsychology and electroencephalogram (EEG) data[53]. Another recent study used fMRI connectivity to group 458 depressed participants into four depression biotypes[55] that were differentially responsive to transcranial magnetic stimulation. Such models provide typologies for diagnosis and treatment that complement existing typologies based on clinical symptoms[2,52].

Predicting treatment outcome. Another use of brain models is the customization of treatment based on a patient's brain measures, an endeavor central to 'precision medicine'. In many areas of medicine, different kinds of pathology can give rise to the same clinical symptoms; the pathology, not the symptoms, guides effective treatment. For example, emerging cancer treatments are based on the cancer's molecular biotypes rather than standard typologies based on gross signs and location[56,57]. Likewise, studies using brain measures to predict who will respond to a particular treatment can identify brain biotypes useful for guiding treatment.

Most extant neuroimaging studies predicting treatment response—16 of the 22 studies in this category we surveyed—focused on depression and anxiety disorders[58–60]. One recent study, based on fMRI responses during fear conditioning, identified a distributed activity pattern with 82% accuracy in discriminating panic disorder patients who responded to cognitive behavioral therapy (CBT) from those who did not[58]. Another study used a regression-based method to predict individual differences in the magnitude of CBT effects[59]. A distributed pattern of brain responses to angry versus neutral faces explained 41% of the variance in CBT benefit. Only a handful of studies predict responses to other treatments, including electroconvulsive therapy[61], transcranial magnetic stimulation[55,62] and drugs[63,64]. Likewise, a small set of studies predict treatment responses in other disorders, including schizophrenia[63] and Parkinson's disease[64].

A critical evaluation of clinical predictive modeling

The studies surveyed above offer hope for understanding the neuropathology of brain disorders and providing useful clinical applications. However, there are some crucial ways in which modeling efforts must mature in order to realize this potential. In this section, we discuss challenges and recommendations with respect to four desirable characteristics[65]: diagnostic value, neuroscientific validity, deployability and scalability, and generalizability across contexts and populations.

Diagnostic value. For a brain measure to serve as a marker for an outcome, it must be diagnostic of that outcome at the individual-person level. Diagnostic models are sensitive and specific measures of the outcome. Sensitivity relates to how robustly the measure responds (i.e., is activated above some predefined threshold) when the outcome is present. Specificity relates to whether the measure responds only in the presence of the target outcome and not to others. When moving from assessing the diagnostic value of a brain measure in research settings to deploying it for clinical use, it is important to consider both the positive and negative predictive values. These predictive values depend on the base rates of the outcome as well as on sensitivity and specificity[33,66]. It is also important to consider the relative clinical and societal costs of false positive and negative errors when establishing thresholds for when brain markers are considered 'active'. These concepts can also be extended to continuous outcomes like symptom severity, for example, using effect sizes.

Formal evaluation of the sensitivity and specificity of brain models is a relatively new concept that emerged along with predictive modeling as part of translational neuroimaging 2.0. It is important not only for establishing clinical utility but also for identifying the construct[67]—the theoretical category of mental events, disorders, or performance—that particular brain patterns measure.

Potential biases in accuracy. Measures of diagnostic value in current studies are subject to potential pitfalls that need to be addressed as the field progresses. As **Figure 2d** shows, model accuracy varies substantially across published studies, with extremely high rates (near 100%) in each diagnostic category. This seems to provide cause for optimism, but there are reasons to be skeptical, and the pattern of results across studies points to the need for further rigorous testing. First, near-perfect accuracy is implausible, considering the low reliability

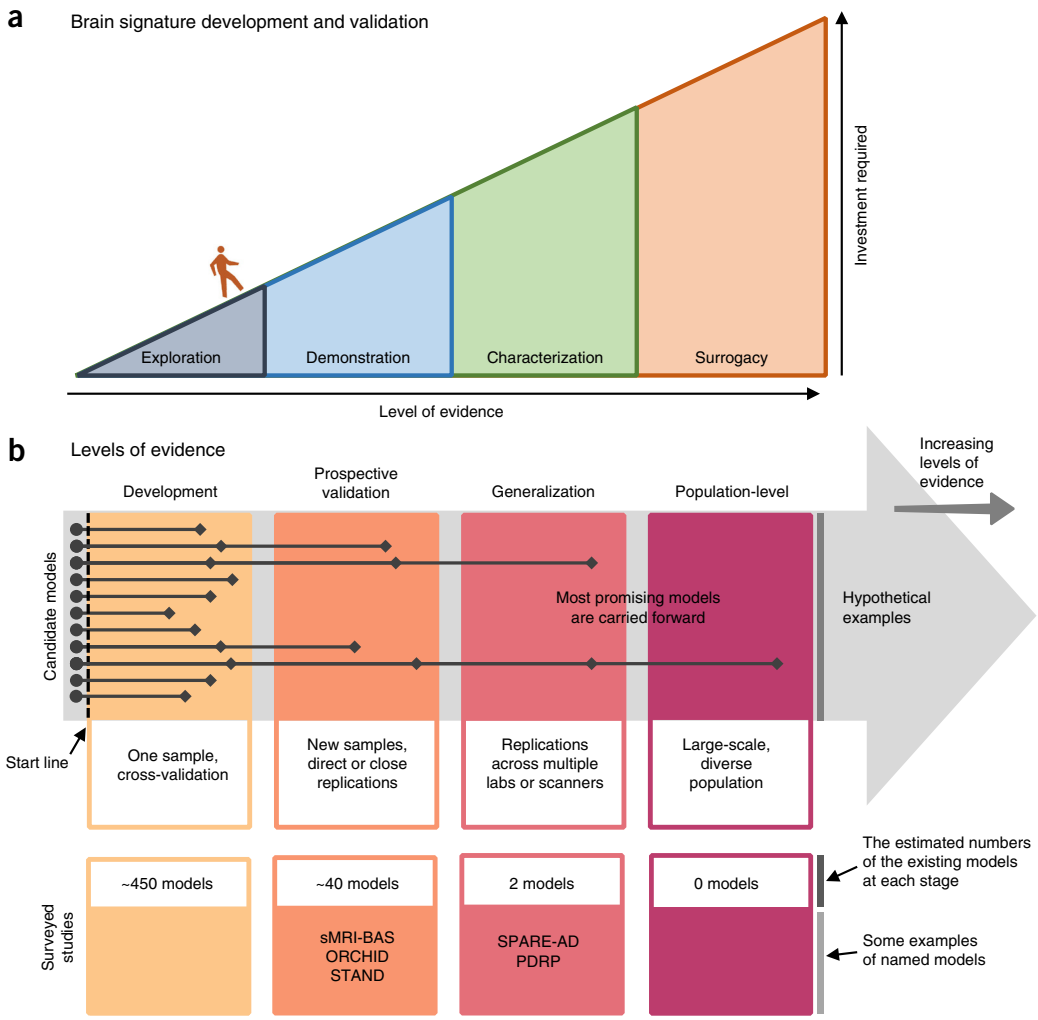

Figure 3 Brain signature development and validation. (**a**) In this process, broad exploration is the first step. Just as drug development involves screening many candidate drugs, exploring multiple approaches and models is important for identifying promising biomarkers. The most promising models must be tested in independent samples to demonstrate their diagnostic accuracy. During characterization, promising candidate biomarkers should show robust replications of findings (for example, high sensitivity and specificity) across multiple independent samples, laboratories, scanners and research settings. This requires tests in larger, more definitive studies, which can eventually promote identification of these biomarkers as surrogate measures and as endpoints in their own right. (**b**) Starting with many candidate models, the most promising ones garner support and are carried forward with increasing levels of evidence. In the development phase, models can be developed based on one study sample and model performance can be estimated using cross-validation. In the prospective validation phase, findings and model performance (i.e., sensitivity, specificity and predictive value) are replicated by applying models to new, independent samples of participants. In the generalization phase, findings and model performance are tested across multiple laboratories, scanners and variants of testing procedures to assess the models' robustness and boundary conditions. In the population-level phase, large-scale tests assess the model's performance when it is applied to diverse populations and test conditions, and additional moderators (i.e., age, race, culture, gender) and boundary conditions are identified in this phase. In this illustration, colored blocks denote the different phases, and black lines indicate hypothetical candidate models. Within each phase, a model can be more or less thoroughly evaluated and more or less successful at establishing utility. This variability is denoted graphically by the variable locations of dots for each model within each phase. A survey of empirical literature to date reveals that only 9% of neuroimaging-based models go beyond the initial development phase. Some notable exceptions include the named models shown here, which have been tested prospectively on new samples (see **Table 2** for details and abbreviations).

of many clinical diagnoses themselves[68]. For example, the inter-rater reliability of a diagnosis of major depression is very low, $\kappa = 0.28$ (ref. 68), which means that if one clinician diagnoses depression, a second clinician will agree only about 50% of the time. The average diagnostic accuracy over 30 brain models for depression included in our survey was high (86.7%). While it is possible that brain measures can be more stable and reliable than symptoms they are associated with[55], the overall pattern indicates likely bias.

Another standard way of assessing bias is to compare accuracy in small and large studies. As variability in accuracy estimates shrinks with sample size, the distribution of accuracy in small studies should form a 'funnel' distributed symmetrically around those of the large studies; asymmetries indicate bias. Our survey reveals evidence for such bias in predictive mapping studies (**Fig. 2f**). In AD, classification accuracy in large-scale studies (for example, $n > 500$) converges on ~90%, but in other areas, such as autism and ADHD, large-scale studies show substantially lower accuracy. Though there are some exceptional large-scale studies with very high accuracy[39,69–71], none of these models have been prospectively tested on independent data and thus await independent validation.

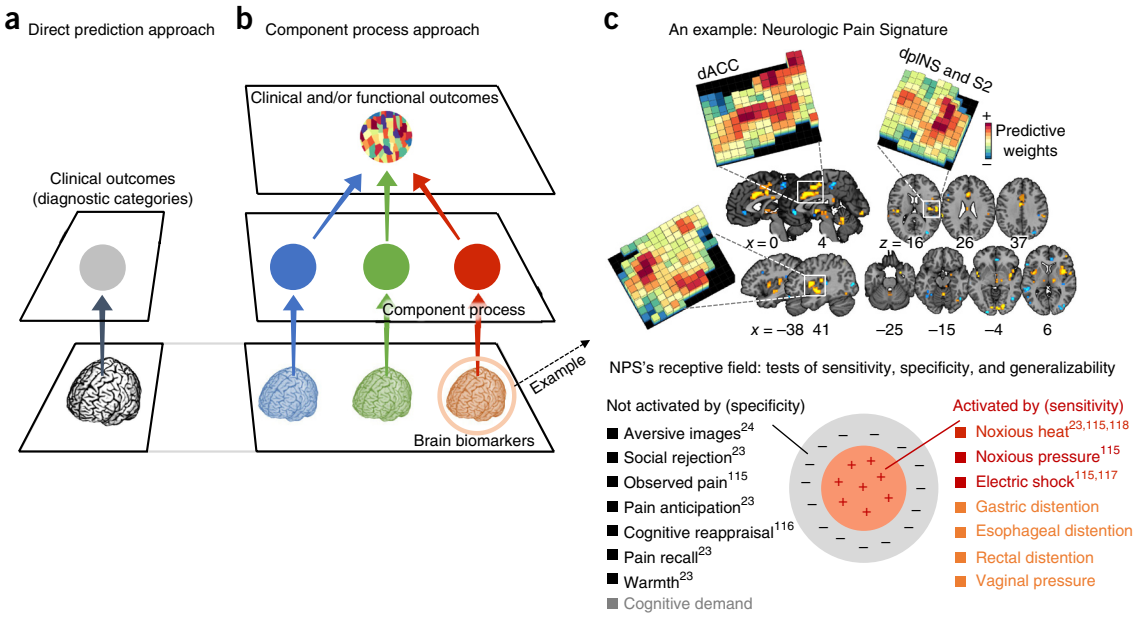

Figure 4 Future directions. (**a**) The direct prediction approach, in which brain features are mapped to clinical diagnostic categories or symptoms directly. (**b**) The component process approach, in which brain features are mapped to basic component processes, such as sustained attention load, memory load, positive or negative affect, pain and others. By introducing this additional layer, brain signatures can have implications for behavior and function beyond patient status or current diagnostic categories. This new level of analysis provides a way of understanding the nature of dysregulated brain processes, assessing risk factors for brain disorders, and understanding and predicting treatment responses. Rather than constructing one brain marker per disorder, component models provide a set of basis processes that are combined in different ways in different disorders. By analogy with color, three components (red, green and blue) can be combined in different ways to form a virtually infinite number of colors. (**c**) The NPS is a signature for one such component process, evoked somatic pain[23], that is potentially dysregulated in multiple disorders. The NPS is defined by brain-wide, mesoscale patterns of fMRI activity across multiple pain-related regions and can be prospectively tested on new individuals and datasets. This allows its properties to be characterized across studies, improving understanding of the types of mental processes and experiences it represents. Top: the NPS pattern map thresholded at $q < 0.05$, false discovery rate (FDR)-corrected for display purposes; the unthresholded patterns in selected regions (dACC, dorsal ACC; dorsal posterior insula, dpINS; secondary somatosensory cortex, S2) are visualized in the insets. Bottom: the NPS's 'psychological receptive field', which visualizes conditions that activate (sensitivity, in orange and red) or do not activate (specificity, in gray and black) the NPS. Dark colored conditions (in red and black) are from published results[23,24,115–118], and light colored conditions (in orange and gray) are from unpublished, preliminary results (data on cognitive demand were tested by C.-W.W.; visceral and vaginal pain data were tested by T.D.W.). Characterizing the NPS's sensitivity and specificity across these conditions and others aides in understanding what NPS alterations in clinical disorders mean from a psychological and functional perspective. In the future, we envision clinical biomarkers composed of sets of interpretable, well characterized models of basic cognitive and affective processes.

Such optimistic biases can be inadvertently introduced in every step of model fitting and testing. Among other problems, accuracy will be inflated if the data used to train the predictive model and test its accuracy are not truly independent, even when cross-validation is used to test nominally out-of-sample participants[26,72]. Some studies perform analysis procedures—e.g., denoising, scaling, component analyses, feature selection—across the entire data set before splitting it into training and testing data, creating dependence and thus optimistic biases in accuracy. Other studies test multiple learning algorithms on a data set and then pick the best one, which results in 'overfitting', an optimistic bias related to model flexibility.

A good way to reduce bias is to prospectively test a model on a new sample, without changing any of the model parameters. Testing the model on a completely independent sample eliminates data-dependence bias, and testing only one final model on the new data set eliminates model-flexibility bias. In our survey, only 9% of studies tested brain models on one or more independent data sets (**Fig. 2e**). These were most common in studies of AD; encouragingly, prospective tests in this field yielded comparable diagnostic accuracy, indicating relatively little bias. However, accuracy in prospective tests of psychosis, ADHD and pain were markedly lower than for tests on the model-development sample, indicating substantial optimistic biases in cross-validated results. Clearly, prospective tests with properly held-out data are critical.

In machine learning competitions, test data are typically held in escrow, and a team is only allowed to submit a single model for testing. Such data escrow practices would increase confidence in the diagnostic accuracy levels reported in published studies.

Recognizing the need for held-out samples tested only once also suggests a different use of consortium and multisite data sets. Of the multisite studies in our survey, 80% did not reserve hold-out test data for prospective testing. Thus, though these studies are large, they are not robust against overly optimistic biases caused by data dependence and overfitting. Considering prospective testing in the early stages of study design and analysis planning would move the field forward.

Testing specificity over multiple alternatives. Most current clinical studies evaluate brain models' sensitivity and specificity for one patient group relative to controls. However, specificity can and should be evaluated relative to a defined set of multiple alternatives[67], not to only one. For example, a model's specificity to depression can be high relative to schizophrenia or autism but low relative to other mood disorders. Therefore, testing sensitivity, specificity, positive predictive value, and so on, should be an open-ended process. There are many potential comparisons among diagnoses, comorbid conditions, symptoms and other outcomes[73]. This testing process will be greatly facilitated by large-scale data that include multiple alternative disease groups and conditions.

Neuroscientific validity. Neuroscientific validity relates to both a model's neurophysiological plausibility and what the model can contribute to advancing understanding of the neurophysiological basis of the outcome. Plausible models respect what is known about the physiological properties of the measures used and are corroborated by evidence from other sources, such as invasive animal or human studies that converge on the same brain regions. Plausible models are more likely to be valid and more likely to contribute to our cumulative understanding of brain health. For example, an award-winning model in the Pittsburgh Brain Imaging Analysis Competition was accurate but relied mainly on fMRI signal in the brain's ventricles[74]. As no meaningful fMRI activity is known to arise from the ventricles, these signals are implausible as measures of neural function and therefore do little to advance our understanding of the brain.

Models that can be understood and described by humans tend to be more neuroscientifically useful. This is a strength of the simple, single-region models used in traditional brain mapping approaches and a weakness of complex machine-learning-derived models with many features. The machine-learning algorithms used to train such models do not, in themselves, provide any constraints related to neuroscientific validity; these must be supplied by the analyst. Machine-learning techniques have, however, developed several heuristic techniques to simplify models; for example, least absolute shrinkage and selection operator (LASSO) and ridge-regularization methods are popular because they reduce the number of brain features in the model by imposing sparsity constraints[34]. Other recent studies improve interpretability by measuring the importance of input features[75]. These efforts can make the basis of model predictions more 'interpretable'—easier to understand and describe.

Developing plausible and interpretable models is important because such models are more likely to hold up to rigorous testing and generalize to new settings. For example, the model with the highest predictive accuracy in the ADHD-200 global competition[76] seemed to be based largely on in-scanner head motion[77]. Another group accurately predicted autism status from brain responses to auditory oddballs, but the model performed at chance after controlling for eye blinks[78]. These two models are not robust; they would likely perform poorly if better techniques were applied to reduce nuisance signals. More importantly, the models themselves tell us nothing about the neural basis of ADHD or autism. Essentially, if we do not know why a test succeeded, it is difficult to determine when it will fail and how meaningfully it contributes to our understanding of the disorder's pathophysiology.

Though converging evidence from previous findings can help validate a model, models that are not compatible with existing theories can also lead to new discoveries that advance theory. For example, though multiple sclerosis has long been thought of as a white matter disease, pathological signs have also been observed in gray matter, which is now known to play important roles in multiple sclerosis[79]. Model development need not be limited to currently available neurobiological mechanisms, if the data provide compelling evidence for the neuroscientific validity of new ones. Therefore, neuroimaging-based models can also play a role in discovery and theory-building.

A systematic approach to improving neuroscientific validity. One important challenge is that we have yet to establish systematic methods to evaluate and enhance multivariate models' interpretability and neuroscientific validity. Here, we propose three basic steps for evaluating and enhancing interpretability.

First, the models should be summarized and visualized in a human-readable way. In this step, statistical techniques for dimensionality reduction and feature selection can be helpful. Some studies reduce the dimensionality of complex models by first using principal or independent component decompositions and analyze relationships of outcomes with a small number of components, rather than a large number of features[51]. Others use a small set of graph theoretic features based on network topology[80,81]. Studies can also cluster voxels with similar predictive profiles[82] or use bootstrap-based significance tests for which voxels contribute most reliably[23].

Second, researchers should evaluate the neuroscientific plausibility of the predictive weights (or other parameters). For example, neuroimaging signatures for AD can be more confidently interpreted if they are validated with postmortem markers of pathology[83]. There is also a growing set of tools for rapidly comparing results to previous findings. Meta-analysis databases can provide comparisons with previous results across many tasks[84]. Large-scale analyses of activity across resting-state data[85] and multimodal data[86] from consortium studies can relate predictive models to established normative patterns. Meta-analytic databases of anatomical connectivity in nonhuman animals[87,88] can relate activity and connectivity across species, allowing mechanistic, invasive animal studies to be brought to bear in interpreting human findings.

Third, researchers should examine, to the degree possible, whether any confounding factors contribute to the model. There is a growing recognition of the complex and pervasive effects of head movement on models and increasing focus on mitigation[89]. There are also many other potential confounds, including physiological noise, eye movements, individual differences in vasculature and hemodynamics, medication use (and abstinence), age and others. Assessing whether machine-learning models based on patterns of confounding variables explain estimated outcomes could be helpful (for example, ref. 23). However, it might be impossible to definitively account for all potential confounds. This does not mean that brain models are not useful; current diagnostic procedures are also fraught with similar challenges. Rather, as we describe below, models can be provisionally trusted in proportion to their evidence and neuroscientific plausibility, and the most promising models should be scaled up to larger tests.

Deployability and scalability. Brain models that are useful for translation must be easily applicable to new individuals and shareable across laboratories, in the sense that testing procedures can be performed in new settings with minimum complexity and potential for error. In all but the simplest cases, this requires data file(s) that precisely specify which brain locations and/or connections are involved in the model and all relevant parameter estimates. Avoiding errors and standardizing procedures will require standardized data formats and software that produce identical results across different computing environments[90]. Scalable models and procedures can be cost-effectively deployed across different groups, supporting large-scale testing and application.

Named neuroimaging models. An encouraging recent trend is the development of named signatures for clinical disorders. Naming is important because it implies that a signature is a defined 'research product' that can be shared and annotated by many people and groups. Named signatures can facilitate subsequent model-sharing and prospective testing. Examples include the SPARE-AD for Alzheimer's and the Parkinson's Disease-Related Pattern (PDRP; based on FDG-PET), among others (**Table 2**). The SPARE-AD was developed in 2008 (ref. 46) and subsequently tested for prospective prediction of disease progression on multiple data sets[45,91]. Later, it was tested for sensitivity to cognitive impairment across multiple neurodegenerative disorders[92,93] and multiple study sites[94]. Likewise, the PDRP signature developed in 2006 (refs. 95,96) has been shared and tested with other groups,

Box 1 Varieties of predictive models

Just as machine learning comprises a large family of algorithms[34,35], there is a large family of models to which they can be applied. For our purposes here, models are ways of structuring variables to provide theoretically meaningful and practically useful representations of brain–outcome relationships. All models make assumptions, and if the assumptions fit the underlying nature of the brain representations involved, they are likely to be more predictive and more theoretically meaningful[132]. Here, we consider five important distinctions that characterize different model classes.

One fundamental choice when building predictive brain models is the choice of spatial scope (**Fig. 5a**). Many early machine-learning applications focused on understanding representations in isolated brain regions[30,133], which is implicitly a model of local representation. Searchlight analyses conduct such local tests across the brain and have become popular as a way of mapping which local regions accurately predict or 'decode' stimuli or outcomes[134]. If an outcome is truly encoded in a single brain area, local decoding analyses are appropriate. However, these searchlights are not typically integrated into unified predictive models that provide a single best prediction. In addition, brain representations relevant for performance and clinical outcomes may often be distributed across multiple regions and networks. If so, models that integrate contributions from different brain areas will likely be required for accurate prediction[24,58,135].

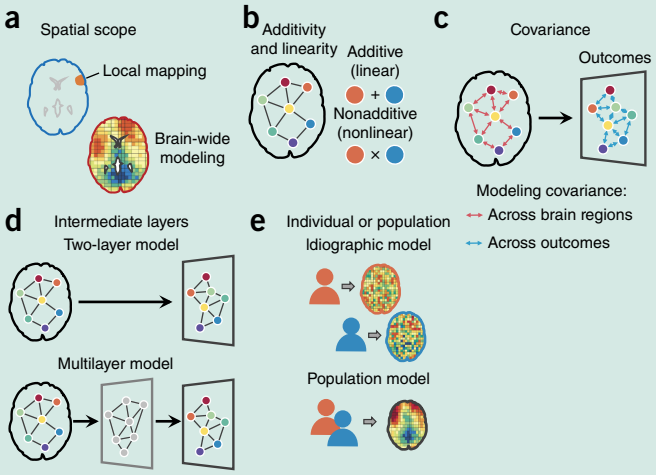

Figure 5 Varieties of predictive models. Developing a predictive model entails making choices about its input data, structural properties, and level of analysis. Five of the most important choices are discussed in **Box 1**.

Another aspect of spatial scope concerns how information is combined across voxels and regions (**Fig. 5b**). Linear models specify patterns of weights on voxels or distributed components that combine additively. They can also include nonadditive interactions and connectivity among regions that are closely related to one another. Nonlinear models can capture more complex, often nonmonotonic relationships between brains and outcomes. Modeling interactions and nonlinearities can sometimes improve accuracy, but they come at a cost in interpretability.

Third, models make different assumptions about how, and whether, to model covariance (i.e., relationships) across brain voxels and/or outcomes (**Fig. 5c**). Standard decoding models (for example, using an SVM to predict an outcome) considers the covariance across voxels when estimating predictive weights—that is, the influence of some voxels is assessed when controlling for other voxels. They do not, however, typically control for other correlated outcome variables (for example, age, task performance or behavior; but cf. refs. 136,137). Thus, they are multivariate in brain space and univariate in outcome space. These models are well suited for understanding how much variance of a stimulus or psychological state can be explained by a brain pattern. Encoding models[138], in contrast, specifically model covariance across psychological and behavioral outcome variables for a single voxel and ignore covariance between voxels. Thus, these models are multivariate in outcome space and univariate in brain space and are useful for assessing the amount of explained variance in a brain region by various processes associated with the stimulus feature space. Finally, some approaches are fully multivariate; they incorporate both encoding and decoding models to model the covariance between both outcomes and brain voxels[139,140].

Fourth, most models currently applied to neuroimaging are two-layer, modeling only relationships among brain variables and outcomes. A very promising class, prominent in models of vision and language, is the construction of multiple-layer models that include intermediate brain representations (**Fig. 5d**). Intermediate 'layers' can include nonlinear transformations of stimulus inputs[141,142], theoretically motivated basis sets of complex features[29,121,139], transformations into temporal or spatial frequencies[143] composite components or 'networks'[23,51], or composite features learned from training data, as in 'deep learning' networks[144,145]. Models can also combine multiple submodels with different operating rules and computations[146].

Finally, models can be trained within a single subject, to identify idiographic brain patterns that predict outcomes for one specific individual[30,133], or across subjects, to identify patterns that generalize across individuals[23,147] (**Fig. 5e**). Advantages of within-subject (idiographic) models include improved accuracy when patterns differ across individuals and have the ability to capture representations at finer spatial scales. Advantages of between-subject (population) models include generalizability, prognostic utility in clinical settings and greater robustness to confounds that can plague within-subject analyses[148,149]. New developments in functional alignment techniques such as hyperalignment might be able to benefit from the strengths of both within- and between-subject models[150].

Across all these choices, model comparisons can help understand which aspects of a model are critical for accurate prediction: local or distributed, additive or nonadditive, linear or nonlinear, two-layer or multilayer, idiographic or generalizable. Models are about much more than prediction: understanding which models work best for a given behavior can help us understand the necessary and sufficient representational basis in terms of brain function.

providing prospective validation beyond one laboratory and study site[97,98]. The ability to deploy and share these signatures across laboratories is a critical part of their widespread validation and testing.

Generalizability across contexts and populations. Generalizability of brain models is another frontier. Models useful for translation must be generalizable in several ways. First, they must generalize to new individuals. Their diagnostic accuracy should hold up in samples tested after the model has been developed and finalized. Second, the most useful signatures will generalize across laboratories, scanners and minor variants in testing conditions. The more robust a brain signature is to variation in test conditions, the more applicable and useful it will be. Third, useful signatures will generalize meaningfully to other outcomes related to the same construct[67]. For example, in addition to sensitivity and specificity to math performance, a signature for 'math ability' should predict performance across multiple types of math

Box 2 Recommendations for future efforts

Model development:
- Increase focus on classification and prediction problems that cannot be easily achieved with existing clinical measures. Problems include early detection, prognosis, differential diagnosis, patient stratification and predicting treatment response (**Fig. 2c**).
- Increase focus on process-based predictive models and intermediate basic processes that may map more closely onto patterns of brain activity than clinical categories themselves and may reveal patterns of dysfunction and neuropathology across disorders (**Fig. 4b**).
- Homogeneous samples can be used for discovery, but the models should eventually be tested on more ecologically valid (i.e., more heterogeneous) samples.

Model validation:
- Plan proper prospective tests with independent test data from the early stages of study design and analysis planning (**Fig. 2e**).
- Test model specificity over multiple alternative conditions (for example, differential diagnoses, multiple cognitive and affective processes).
- Demonstrate models' neuroscientific validity (see "A systematic approach to improving neuroscientific validity").

Cumulative science:
- Treat brain models as sharable research products that can be tested and annotated across different laboratories.
- Name newly developed predictive models to facilitate subsequent model-sharing and prospective testing (**Table 2**).
- Identify promising models and test them in increasingly broad and rigorous ways.

Big data approaches:
- Include multiple disease groups and task conditions in large-scale data initiatives. Important problems such as patient stratification and specificity testing can only be achieved with data that cut across multiple conditions and diagnoses.
- Establish quality-control standards and abide by established ones.
- When developing models on multisite data, carefully consider issues of variables that may be unbalanced across study sites (for example, patient/control ratios and measurement variances), and thus create confounds. Where such confounds are unavoidable, consider a strategy of developing models on one sample and then testing generalizability to other samples, rather than pooling data across sites.

tests[99]. Iterative testing of what a signature does and does not generalize to can help identify what construct a signature actually measures.

Ecologically valid data sets. One important challenge for generalizability comes from the fact that most clinical studies collect data from homogeneous patient samples carefully selected on the basis of a specific set of inclusion and exclusion criteria. Homogeneous samples can increase statistical power to discover differences between groups and rule out a number of potential confounds. However, such samples are typically not representative of the broader population, and models based on them are thus less likely to generalize to real-life clinical settings, where patient groups are highly heterogeneous. Some studies have attempted to overcome this problem by collecting data from all comers to the clinic[100], conducting large-scale community cohort studies[101] or using multisite consortium data sets[102]. Clearly, there is a tradeoff between tighter control and broader generalizability, and both approaches are needed. Managing this tradeoff remains an open challenge.

Big data approaches. Though explicit evaluations of generalizability across different populations, sites and scanners are still very rare, studies have begun to explicitly examine generalizability using large-scale data sets based on research consortia or multisite collaborations[32]. However, the big data approach also has many challenges. A central one is related to variability in data quality, acquisition parameters and procedures, clinical assessments, missing data, and study populations. Variability in data acquisition leads to differences in data scaling and to different patterns of artifacts and signal dropout. Variability in study populations and clinical assessments can cause clinical severity to be confounded with study site (among other confounds). Such confounds are difficult to fully account for in analysis. Another key issue is that many consortium-based datasets include a limited number of functional tasks, with heavy reliance on resting-state data. This limits the ability to assess a wide range of relevant cognitive and affective functions. In addition, the uncontrolled psychological nature of the resting-state 'task' can increase the possibility of confounds[103,104].

Despite these challenges, specific recommendations and massive efforts to resolve these issues have already been made for some data sets (for example, ADNI[102], the function Biomedical Informatics Research Network (fBIRN)[105], and Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP)[106]). The efforts include standardizing scan parameters and clinical measures, calibrating scanners using standard phantoms, developing new task models, using a traveling expert and centralized monitoring of data, providing multisite training for local staff, elaborating documentation, and others. Such solutions are mainly available for prospective data collection projects, not for retrospective data-sharing projects (such as the Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) project[107]), which require different solutions (for example, using meta-analytic approaches).

Future directions: toward a next generation of translational studies

Building a cumulative science of neurotranslation. The 475 studies summarized in **Figure 2** reveal a wilderness of different algorithms, models, methods and study populations. This is exactly as it should be. Discovering the most promising tasks and models requires broad exploration at first. What is needed now is to identify the most successful approaches and build on them, testing them in increasingly broad and rigorous ways. As Borsook, Hargreaves and colleagues have noted[108,109], the neuroimaging model development process should be similar to the process of developing pharmaceuticals and bringing them to market (**Fig. 3a**). As in the drug discovery process, neuroimaging model development begins with a broad search for potential predictive models. Those with good diagnostic performances should be further tested for accuracy and generalizability across multiple studies, laboratories and populations. As shown in **Figure 3b**, in the initial phases, it is advantageous to pursue many alternatives and to progressively scale up to large-scale testing in proportion to the model's utility. Importantly, most models (~90%) are still in early stages of development and have yet to be tested beyond an initial development study.

This model development and validation process can be greatly aided by the sharing of data, but arguably it can be aided even more by sharing the models themselves, including model specifications and parameter estimates. Biology is replete with assays and standardized protocols that are repeated daily in thousands of laboratories across the world. Such assays constitute shareable research products and routinely make the transition from scientific development to widespread use. In neuroimaging, defined signatures can be tested and annotated across laboratories and study cohorts, their utility validated and boundary conditions—what they measure and what they do not measure—characterized. Researchers are becoming increasingly comfortable with the idea of using other people's data. However, to advance the field, we must also begin to use others' models of brain disorders.

Process-based predictive models. Most current translational models are comprised of spatial patterns that map brain structure or function directly onto clinical outcomes (**Fig. 4a**). These models do not consider intermediate features or processes and are often minimally constrained by theories of brain function. They provide little description of the division of labor across brain regions and the dynamics of information flow through the regions included in the model. Considering these aspects affords opportunities to develop more sophisticated brain models of behavioral outcomes and disorders (**Fig. 5** and **Box 1**). Models that address these limitations might prove to be both more accurate and more neuroscientifically useful, providing greater insight into the nature of the mental processes that are disrupted in the course of brain disease.

One promising direction is the development of signatures for basic mental processes, which can then serve as intermediate features that are altered in various combinations in different disorders[6,110,111] (**Fig. 4b**). Such process-based predictive models are essential for moving beyond current diagnostic categories and establishing specific forms of neuropathology that lead to specific functional problems across disorders, as described in the RDoC (ref. 7). In one example, Lopez-Sola *et al.* developed a model to discriminate patients with fibromyalgia from healthy controls based on brain patterns related to several distinct component processes[110]. One process was evoked pain sensitivity, which is a feature of multiple pain-related disorders[112–114]. Wager *et al.* developed an fMRI-based signature for evoked pain, called the Neurologic Pain Signature (NPS)[23], which is sensitive and specific to pain across a number of conditions[23,24,115,116] and which generalizes to multiple types of acute pain across studies and diverse populations[115,117,118] (**Fig. 4c**). This signature does not measure a disorder but rather a negative sensory and affective process that cuts across disorders. Lopez-Sola *et al.*[110] found that enhanced NPS responses, combined with another brain signature related to nonpainful sensory processing, discriminated fibromyalgia from pain-free controls with 93% accuracy.

In another example, Wiecki *et al.* used computational model-derived parameters combined with EEG data to classify deep-brain stimulation (DBS) state (i.e., on versus off) in patients with Parkinson's disease[119]. DBS of the subthalamic nucleus often induces impulsive behaviors, which Wiecki *et al.*[119] modeled using a drift-diffusion model. EEG measures, combined with drift-diffusion model of impulsivity classified individuals' DBS status with 0.81 area under the curve (AUC), whereas combining EEG and response times—a simpler measure of impulsivity—yielded only 0.67 AUC. This study illustrates how computational models can aid in developing intermediate features, highlighting the promise of computational psychiatry[120].

Finally, another approach is to use dynamic process models as intermediate features, linking diagnostic performance with a process-level description of information flow among brain regions. In one example, Brodersen *et al.*[121] embedded a dynamic causal model[122] of speech processing within a classification framework. The model estimated which connections in a structured model of auditory system connectivity were disrupted in aphasia. This model discriminated aphasics from controls with high (98%) accuracy and outperformed standard support vector machine pattern classification. Recently, this approach has been extended to other domains, such as predicting hidden motives for social behavior[123], that may form core components of dysregulated behavior across mental disorders.

Conclusions

The widespread availability of pattern recognition techniques, combined with large multisite neuroimaging data sets, affords unprecedented opportunities to close the gap between basic and translational neuroscience. However, major advances will require specific ways of combining pattern recognition and aggregated neuroimaging data that are not yet the norm (**Box 2**). Predictive models should focus on both clinical outcomes and basic processes that may be dysregulated across multiple disorders. The models must be precisely defined, applicable to individual persons and neuroscientifically plausible and interpretable. These modeling efforts go hand in hand with increasingly systematic assessment of the diagnostic value of brain markers across diverse samples. Models should be generalizable in multiple ways: across individuals, assessment methods, experimental settings and populations. Testing a model's diagnostic value and generalizability is an open-ended process that requires participation from multiple laboratories. Therefore, the models themselves must be easily deployable and shareable. Increased collaborative efforts to share predictive models as well as data will allow the models to be rigorously and prospectively tested, helping to bring translational goals within reach.

ACKNOWLEDGMENTS

We thank our colleagues for discussion of issues surrounding biomarker development and consortium data, including V. Apkarian, M. Banich, D. Barch, P. Bellec, R. Casanova, C. Davatzikos, O. Doyle, D. Eidelberg, G. Glover, S. Mackey, E. Mayer, R. Poldrack, V. Prashanthi, M. Rosenberg, S. Smith, I. Tracey and others. We also thank J. Buhle, L. Van Oudenhove, M. Kano, P. Kragel, H. Giao Ly, P. Dupont, A. Rubio, C. Delon-Martin and B.L. Bonaz for contributing to work discussed in **Figure 4** and the authors of published manuscripts using the Neurologic Pain Signature. This work was funded by NIH R01DA035484 and R01MH076136 (T.D.W., PI).

1. Mather, M., Cacioppo, J.T. & Kanwisher, N. Introduction to the special section: 20 years of fMRI-what has it done for understanding cognition? *Perspect. Psychol. Sci.* **8**, 41–43 (2013).
2. Kapur, S., Phillips, A.G. & Insel, T.R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).
3. 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).
4. Keedwell, P.A., Andrew, C., Williams, S.C., Brammer, M.J. & Phillips, M.L. The neural correlates of anhedonia in major depressive disorder. *Biol. Psychiatry* **58**, 843–853 (2005).
5. Tom, S.M., Fox, C.R., Trepel, C. & Poldrack, R.A. The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518 (2007).
6. Rosenberg, M.D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19**, 165–171 (2016).
7. Sanislow, C.A. *et al.* Developing constructs for psychopathology research: research domain criteria. *J. Abnorm. Psychol.* **119**, 631–639 (2010).
8. Scoville, W.B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
9. Fodor, J.A. *The Modularity of Mind* (MIT Press, 1983).
10. Hamani, C. *et al.* Deep brain stimulation for chronic neuropathic pain: long-term outcome and the incidence of insertional effect. *Pain* **125**, 188–196 (2006).
11. Welter, M.L. *et al.* Basal ganglia dysfunction in OCD: subthalamic neuronal activity correlates with symptoms severity and predicts high-frequency stimulation efficacy. *Transl. Psychiatry* **1**, e5 (2011).
12. Krack, P. *et al.* Five-year follow-up of bilateral stimulation of the subthalamic nucleus in advanced Parkinson's disease. *N. Engl. J. Med.* **349**, 1925–1934 (2003).

13. Swartz, J.R., Knodt, A.R., Radtke, S.R. & Hariri, A.R. A neural biomarker of psychological vulnerability to future life stress. *Neuron* **85**, 505–511 (2015).
14. Dougherty, D.D. *et al.* A randomized sham-controlled trial of deep brain stimulation of the ventral capsule/ventral striatum for chronic treatment-resistant depression. *Biol. Psychiatry* **78**, 240–248 (2015).
15. Morishita, T., Fayad, S.M., Higuchi, M.A., Nestor, K.A. & Foote, K.D. Deep brain stimulation for treatment-resistant depression: systematic review of clinical outcomes. *Neurotherapeutics* **11**, 475–484 (2014).
16. Reddan, M.C., Lindquist, M.A. & Wager, T.D. Effect size estimation in neuroimaging. *JAMA Psychiatry* http://dx.doi.org/10.1001/jamapsychiatry.2016.3356 (2017).
17. Logothetis, N.K. What we can do and what we cannot do with fMRI. *Nature* **453**, 869–878 (2008).
18. Kvitsiani, D. *et al.* Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).
19. Price, J.L. & Drevets, W.C. Neural circuits underlying the pathophysiology of mood disorders. *Trends Cogn. Sci.* **16**, 61–71 (2012).
20. Roy, M., Shohamy, D. & Wager, T.D. Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends Cogn. Sci.* **16**, 147–156 (2012).
21. Wager, T.D. *et al.* Pain in the ACC? *Proc. Natl. Acad. Sci. USA* **113**, E2474–E2475 (2016).
22. Poldrack, R.A. Can cognitive processes be inferred from neuroimaging data? *Trends Cogn. Sci.* **10**, 59–63 (2006).
23. Wager, T.D. *et al.* An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).
24. Chang, L.J., Gianaros, P.J., Manuck, S.B., Krishnan, A. & Wager, T.D. A sensitive and specific neural signature for picture-induced negative affect. *PLoS Biol.* **13**, e1002180 (2015).
25. Doyle, O.M., Mehta, M.A. & Brammer, M.J. The role of machine learning in neuroimaging for drug discovery and development. *Psychopharmacology (Berl.)* **232**, 4179–4189 (2015).
26. Haynes, J.D. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* **87**, 257–270 (2015).
27. Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G. & Mechelli, A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* **36**, 1140–1152 (2012).
28. Hackmack, K., Paul, F., Weygandt, M., Allefeld, C. & Haynes, J.D. Multi-scale classification of disease using structural MRI and wavelet transform. *Neuroimage* **62**, 48–58 (2012).
29. Miyawaki, Y. *et al.* Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* **60**, 915–929 (2008).
30. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* **8**, 679–685 (2005).
31. Kriegeskorte, N., Cusack, R. & Bandettini, P. How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *Neuroimage* **49**, 1965–1976 (2010).
32. Poldrack, R.A. & Gorgolewski, K.J. Making big data open: data sharing in neuroimaging. *Nat. Neurosci.* **17**, 1510–1517 (2014).
33. Abi-Dargham, A. & Horga, G. The search for imaging biomarkers in psychiatric disorders. *Nat. Med.* **22**, 1248–1255 (2016).
34. Hastie, T., Tibshirani, R. & Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn (Springer, 2009).
35. Mohri, M., Rostamizadeh, A. & Talwalkar, A. *Foundations of Machine Learning* (MIT Press, 2012).
36. de Leon, M.J. *et al.* Positron emission tomographic studies of aging and Alzheimer disease. *AJNR Am. J. Neuroradiol.* **4**, 568–571 (1983).
37. Kippenhan, J.S., Barker, W.W., Pascal, S., Nagel, J. & Duara, R. Evaluation of a neural-network classifier for PET scans of normal and Alzheimer's disease subjects. *J. Nucl. Med.* **33**, 1459–1467 (1992).
38. Doyle, O.M. *et al.* Predicting progression of Alzheimer's disease using ordinal regression. *PLoS One* **9**, e105542 (2014).
39. Singh, G. & Samavedham, L. Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: A case study on early-stage diagnosis of Parkinson disease. *J. Neurosci. Methods* **256**, 30–40 (2015).
40. 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).
41. Sørensen, L. *et al.* Early detection of Alzheimer's disease using MRI hippocampal texture. *Hum. Brain Mapp.* **37**, 1148–1161 (2016).
42. Moradi, E., Pepe, A., Gaser, C., Huttunen, H. & Tohka, J. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* **104**, 398–412 (2015).
43. Beardslee, W.R. *et al.* Prevention of depression in at-risk adolescents: longer-term effects. *JAMA Psychiatry* **70**, 1161–1170 (2013).
44. Addington, J. & Heinssen, R. Prediction and prevention of psychosis in youth at clinical high risk. *Annu. Rev. Clin. Psychol.* **8**, 269–289 (2012).
45. Davatzikos, C., Xu, F., An, Y., Fan, Y. & Resnick, S.M. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain* **132**, 2026–2035 (2009).
46. Fan, Y., Batmanghelich, N., Clark, C.M., Davatzikos, C. & Alzheimer's Disease Neuroimaging Initiative\par Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline. *Neuroimage* **39**, 1731–1743 (2008).
47. Misra, C., Fan, Y. & Davatzikos, C. Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. *Neuroimage* **44**, 1415–1422 (2009).
48. Tang, C.C. *et al.* Differential diagnosis of Parkinsonism: a metabolic imaging study using pattern analysis. *Lancet Neurol.* **9**, 149–158 (2010).
49. Pantazatos, S.P., Talati, A., Schneier, F.R. & Hirsch, J. Reduced anterior temporal and hippocampal functional connectivity during face processing discriminates individuals with social anxiety disorder from healthy controls and panic disorder, and increases following treatment. *Neuropsychopharmacology* **39**, 425–434 (2014).
50. Anticevic, A. *et al.* Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. *Cereb. Cortex* **24**, 3116–3130 (2014).
51. Calhoun, V.D., Maciejewski, P.K., Pearlson, G.D. & Kiehl, K.A. Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Hum. Brain Mapp.* **29**, 1265–1275 (2008).
52. Insel, T.R. & Cuthbert, B.N. Medicine. Brain disorders? Precisely. *Science* **348**, 499–500 (2015).
53. Clementz, B.A. *et al.* Identification of distinct psychosis biotypes using brain-based biomarkers. *Am. J. Psychiatry* **173**, 373–384 (2016).
54. Price, R.B. *et al.* Parsing heterogeneity in the brain connectivity of depressed and healthy adults during positive mood. *Biol. Psychiatry* S0006-3223(16)32540-9 (2016).
55. Drysdale, A.T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* (2016).
56. Weinstein, J.N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
57. Roychowdhury, S. & Chinnaiyan, A.M. Translating genomics for precision cancer medicine. *Annu. Rev. Genomics Hum. Genet.* **15**, 395–415 (2014).
58. Hahn, T. *et al.* Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry* **72**, 68–74 (2015).
59. Doehrmann, O. *et al.* Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry* **70**, 87–97 (2013).
60. Whitfield-Gabrieli, S. *et al.* Brain connectomics predict response to treatment in social anxiety disorder. *Mol. Psychiatry* **21**, 680–685 (2016).
61. van Waarde, J.A. *et al.* A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. *Mol. Psychiatry* **20**, 609–614 (2015).
62. Widge, A.S., Avery, D.H. & Zarkowski, P. Baseline and treatment-emergent EEG biomarkers of antidepressant medication response do not predict response to repetitive transcranial magnetic stimulation. *Brain Stimul.* **6**, 929–931 (2013).
63. Sarpal, D.K. *et al.* Baseline striatal functional connectivity as a predictor of response to antipsychotic drug treatment. *Am. J. Psychiatry* **173**, 69–77 (2016).
64. Ye, Z. *et al.* Predicting beneficial effects of atomoxetine and citalopram on response inhibition in Parkinson's disease with clinical and neuroimaging measures. *Hum. Brain Mapp.* **37**, 1026–1037 (2016).
65. Woo, C.W. & Wager, T.D. Neuroimaging-based biomarker discovery and validation. *Pain* **156**, 1379–1381 (2015).
66. Robinson, M., Boissoneault, J., Sevel, L., Letzen, J. & Staud, R. The effect of base rate on the predictive value of brain biomarkers. *J. Pain* **17**, 637–641 (2016).
67. Cronbach, L.J. & Meehl, P.E. Construct validity in psychological tests. *Psychol. Bull.* **52**, 281–302 (1955).
68. Freedman, R. *et al.* The initial field trials of DSM-5: new blooms and old thorns. *Am. J. Psychiatry* **170**, 1–5 (2013).
69. Iidaka, T. Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex* **63**, 55–67 (2015).
70. Duffy, F.H. & Als, H. A stable pattern of EEG spectral coherence distinguishes children with autism from neuro-typical controls - a large case control study. *BMC Med.* **10**, 64 (2012).
71. Deshpande, G., Wang, P., Rangaprakash, D. & Wilamowski, B. Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Trans. Cybern.* **45**, 2668–2679 (2015).
72. Whelan, R. & Garavan, H. When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol. Psychiatry* **75**, 746–748 (2014).
73. Zaki, J., Wager, T.D., Singer, T., Keysers, C. & Gazzola, V. The anatomy of suffering: understanding the relationship between nociceptive and empathic pain. *Trends Cogn. Sci.* **20**, 249–259 (2016).
74. Olivetti, E., Sona, D. & Veeramachaneni, S. Gaussian process regression and recurrent neural networks for fmri image classification. in *Proc. 12th Meeting Org. for Human Brain Mapping, Florence, Italy* (2006).
75. Ribeiro, M.T., Singh, S. & Guestrin, C. "Why should I trust you?": Explaining the predictions of any classifier. Preprint at *arXiv* https://arxiv.org/abs/1602.04938 (2016).
76. HD-200 Consortium. The ADHD-200 Consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Front. Syst. Neurosci.* **6**, 62 (2012).
77. Eloyan, A. *et al.* Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Front. Syst. Neurosci.* **6**, 61 (2012).
78. Eldridge, J., Lane, A.E., Belkin, M. & Dennis, S. Robust features for the automatic identification of autism spectrum disorder in children. *J. Neurodev. Disord.* **6**, 12 (2014).

79. Geurts, J.J., Calabrese, M., Fisher, E. & Rudick, R.A. Measurement and clinical effect of grey matter pathology in multiple sclerosis. *Lancet Neurol.* **11**, 1082–1092 (2012).
80. van den Heuvel, M.P. *et al.* Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry* **70**, 783–792 (2013).
81. Yahata, N. *et al.* A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. Commun.* **7**, 11254 (2016).
82. Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E. & Gallant, J.L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
83. Vemuri, P. *et al.* Antemortem MRI based STructural Abnormality iNDex (STAND)-scores correlate with postmortem Braak neurofibrillary tangle stage. *Neuroimage* **42**, 559–567 (2008).
84. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C. & Wager, T.D. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
85. Yeo, B.T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
86. Glasser, M.F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
87. Bota, M., Dong, H.W. & Swanson, L.W. Brain architecture management system. *Neuroinformatics* **3**, 15–48 (2005).
88. Stephan, K.E. The history of CoCoMac. *Neuroimage* **80**, 46–52 (2013).
89. Power, J.D., Schlaggar, B.L. & Petersen, S.E. Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* **105**, 536–551 (2015).
90. Gorgolewski, K.J. & Poldrack, R.A. A practical guide for improving transparency and reproducibility in neuroimaging research. *PLoS Biol.* **14**, e1002506 (2016).
91. Davatzikos, C., Bhatt, P., Shaw, L.M., Batmanghelich, K.N. & Trojanowski, J.Q. Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol. Aging* **32**, 2322.e19–2322.e27 (2011).
92. Weintraub, D. *et al.* Alzheimer's disease pattern of brain atrophy predicts cognitive decline in Parkinson's disease. *Brain* **135**, 170–180 (2012).
93. Toledo, J.B. *et al.* Memory, executive, and multidomain subtle cognitive impairment: clinical and biomarker findings. *Neurology* **85**, 144–153 (2015).
94. Habes, M. *et al.* White matter hyperintensities and imaging patterns of brain ageing in the general population. *Brain* **139**, 1164–1179 (2016).
95. Asanuma, K. *et al.* Network modulation in the treatment of Parkinson's disease. *Brain* **129**, 2667–2678 (2006).
96. Eidelberg, D. Metabolic brain networks in neurodegenerative disorders: a functional imaging approach. *Trends Neurosci.* **32**, 548–557 (2009).
97. Wu, P. *et al.* Metabolic brain network in the Chinese patients with Parkinson's disease based on 18F-FDG PET imaging. *Parkinsonism Relat. Disord.* **19**, 622–627 (2013).
98. Teune, L.K. *et al.* Validation of parkinsonian disease-related metabolic brain patterns. *Mov. Disord.* **28**, 547–551 (2013).
99. Westfall, J., Judd, C.M. & Kenny, D.A. Replicating studies in which samples of participants respond to samples of stimuli. *Perspect. Psychol. Sci.* **10**, 390–399 (2015).
100. Hashmi, J.A. *et al.* Shape shifting pain: chronification of back pain shifts brain representation from nociceptive to emotional circuits. *Brain* **136**, 2751–2768 (2013).
101. Petersen, S.E. *et al.* Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *J. Cardiovasc. Magn. Reson.* **15**, 46 (2013).
102. Weiner, M.W. *et al.* Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimers Dement.* **11**, 865–884 (2015).
103. Tagliazucchi, E. & Laufs, H. Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron* **82**, 695–708 (2014).
104. Buckner, R.L., Krienen, F.M. & Yeo, B.T.T. Opportunities and limitations of intrinsic functional connectivity MRI. *Nat. Neurosci.* **16**, 832–837 (2013).
105. Glover, G.H. *et al.* Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *J. Magn. Reson. Imaging* **36**, 39–54 (2012).
106. Landis, J.R. *et al.* The MAPP research network: design, patient characterization and operations. *BMC Urol.* **14**, 58 (2014).
107. Thompson, P.M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
108. Borsook, D., Becerra, L. & Hargreaves, R. Biomarkers for chronic pain and analgesia. Part 1: the need, reality, challenges, and solutions. *Discov. Med.* **11**, 197–207 (2011).
109. Hargreaves, R.J. *et al.* Optimizing central nervous system drug development using molecular imaging. *Clin. Pharmacol. Ther.* **98**, 47–60 (2015).
110. López-Solà, M. *et al.* Towards a neurophysiological signature for fibromyalgia. *Pain* (2016).
111. Lombardo, M.V. *et al.* Different functional neural substrates for good and poor language outcome in autism. *Neuron* **86**, 567–577 (2015).
112. Woolf, C.J. & Salter, M.W. Neuronal plasticity: increasing the gain in pain. *Science* **288**, 1765–1769 (2000).
113. Diatchenko, L., Nackley, A.G., Slade, G.D., Fillingim, R.B. & Maixner, W. Idiopathic pain disorders--pathways of vulnerability. *Pain* **123**, 226–230 (2006).
114. Adler, G. & Gattaz, W.F. Pain perception threshold in major depression. *Biol. Psychiatry* **34**, 687–689 (1993).
115. Krishnan, A. *et al.* Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* **5**, e15166 (2016).
116. Woo, C.W., Roy, M., Buhle, J.T. & Wager, T.D. Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. *PLoS Biol.* **13**, e1002036 (2015).
117. Ma, Y. *et al.* Serotonin transporter polymorphism alters citalopram effects on human pain responses to physical pain. *Neuroimage* **135**, 186–196 (2016).
118. Bräscher, A.K., Becker, S., Hoeppli, M.E. & Schweinhardt, P. Different brain circuitries mediating controllable and uncontrollable pain. *J. Neurosci.* **36**, 5013–5025 (2016).
119. Wiecki, T.V., Poland, J. & Frank, M.J. Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. *Clin. Psychol. Sci.* **3**, 378–399 (2015).
120. 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).
121. Brodersen, K.H. *et al.* Generative embedding for model-based classification of fMRI data. *PLOS Comput. Biol.* **7**, e1002079 (2011).
122. Friston, K.J., Harrison, L. & Penny, W. Dynamic causal modelling. *Neuroimage* **19**, 1273–1302 (2003).
123. Hein, G., Morishima, Y., Leiberg, S., Sul, S. & Fehr, E. The brain's functional network architecture reveals human motives. *Science* **351**, 1074–1078 (2016).
124. Fan, Y., Resnick, S.M., Wu, X. & Davatzikos, C. Structural and functional biomarkers of prodromal Alzheimer's disease: a high-dimensional pattern classification study. *Neuroimage* **41**, 277–285 (2008).
125. Casanova, R. *et al.* Alzheimer's disease risk assessment using large-scale machine learning methods. *PLoS One* **8**, e77949 (2013).
126. Tosun, D., Joshi, S. & Weiner, M.W. Neuroimaging predictors of brain amyloidosis in mild cognitive impairment. *Ann. Neurol.* **74**, 188–198 (2013).
127. Vemuri, P. *et al.* Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *Neuroimage* **39**, 1186–1197 (2008).
128. Huang, C. *et al.* Metabolic brain networks associated with cognitive function in Parkinson's disease. *Neuroimage* **34**, 714–723 (2007).
129. Mure, H. *et al.* Parkinson's disease tremor-related metabolic network: characterization, progression, and treatment effects. *Neuroimage* **54**, 1244–1253 (2011).
130. Eckert, T. *et al.* Abnormal metabolic networks in atypical parkinsonism. *Mov. Disord.* **23**, 727–733 (2008).
131. Niethammer, M. *et al.* A disease-specific metabolic brain network associated with corticobasal degeneration. *Brain* **137**, 3036–3046 (2014).
132. Geman, S., Bienenstock, E. & Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **4**, 1–58 (1992).
133. Haxby, J.V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425–2430 (2001).
134. Kriegeskorte, N., Goebel, R. & Bandettini, P. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. USA* **103**, 3863–3868 (2006).
135. Sato, J.R. *et al.* Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression. *Psychiatry Res.* **233**, 289–291 (2015).
136. Wager, T.D., Atlas, L.Y., Leotti, L.A. & Rilling, J.K. Predicting individual differences in placebo analgesia: contributions of brain activity during anticipation and pain experience. *J. Neurosci.* **31**, 439–452 (2011).
137. Dukart, J., Schroeter, M.L. & Mueller, K. Age correction in dementia--matching to a healthy brain. *PLoS One* **6**, e22193 (2011).
138. Naselaris, T., Kay, K.N., Nishimoto, S. & Gallant, J.L. Encoding and decoding in fMRI. *Neuroimage* **56**, 400–410 (2011).
139. Mitchell, T.M. *et al.* Predicting human brain activity associated with the meanings of nouns. *Science* **320**, 1191–1195 (2008).
140. Krishnan, A., Williams, L.J., McIntosh, A.R. & Abdi, H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475 (2011).
141. Nishimoto, S. *et al.* Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* **21**, 1641–1646 (2011).
142. Kay, K.N., Naselaris, T., Prenger, R.J. & Gallant, J.L. Identifying natural images from human brain activity. *Nature* **452**, 352–355 (2008).
143. Ketz, N., O'Reilly, R.C. & Curran, T. Classification aided analysis of oscillatory signatures in controlled retrieval. *Neuroimage* **85**, 749–760 (2014).
144. Kim, J., Calhoun, V.D., Shim, E. & Lee, J.H. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage* **124** Pt A: 127–146 (2016).
145. Kriegeskorte, N. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science* **1**, 417–446 (2015).
146. O'Reilly, R.C. Biologically based computational models of high-level cognition. *Science* **314**, 91–94 (2006).
147. Poldrack, R.A., Halchenko, Y.O. & Hanson, S.J. Decoding the large-scale structure of brain function by classifying mental States across individuals. *Psychol. Sci.* **20**, 1364–1372 (2009).
148. Todd, M.T., Nystrom, L.E. & Cohen, J.D. Confounds in multivariate pattern analysis: Theory and rule representation case study. *Neuroimage* **77**, 157–165 (2013).
149. Etzel, J.A., Zacks, J.M. & Braver, T.S. Searchlight analysis: promise, pitfalls, and potential. *Neuroimage* **78**, 261–269 (2013).
150. Haxby, J.V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).