The functional neuroanatomy of bipolar disorder: a consensus model

Stephen M Strakowski\textsuperscript{a}, Caleb M Adler\textsuperscript{a}, Jorge Almeida\textsuperscript{b}, Lori L Alshuler\textsuperscript{c,d}, Hilary P Blumberg\textsuperscript{e}, Kiki D Chang\textsuperscript{f}, Melissa P DelBello\textsuperscript{a}, Sophia Frangou\textsuperscript{g}, Andrew McIntosh\textsuperscript{h}, Mary L Phillips\textsuperscript{b,l}, Jessika E Sussman\textsuperscript{h}, and Jennifer D Townsend\textsuperscript{c}

\textsuperscript{a}Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, Cincinnati, OH
\textsuperscript{b}Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh PA
\textsuperscript{c}Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, The David Geffen School of Medicine, University of California at Los Angeles
\textsuperscript{d}Department of Psychiatry, VA Greater Los Angeles Healthcare System, Los Angeles, CA
\textsuperscript{e}Department of Psychiatry, Yale University School of Medicine, New Haven, CT
\textsuperscript{f}Pediatric Bipolar Disorders Research Program, Division of Child and Adolescent Psychiatry, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA
\textsuperscript{g}Section of Neurobiology of Psychosis, Department of Psychosis Studies, Institute of Psychiatry, King’s College, London
\textsuperscript{h}Division of Psychiatry, School of Molecular and Clinical Medicine, University of Edinburgh, Edinburgh
\textsuperscript{i}Department of Psychological Medicine, Cardiff, UK

Abstract

Objectives—Functional neuroimaging methods have proliferated in recent years, such that functional magnetic resonance imaging, in particular, is now widely used to study bipolar disorder. However, discrepant findings are common. A workgroup was organized by the Department of Psychiatry, University of Cincinnati (Cincinnati, OH, USA) to develop a consensus functional neuroanatomic model of bipolar I disorder based upon the participants’ work as well as that of others.

Methods—Representatives from several leading bipolar disorder neuroimaging groups were organized to present an overview of their areas of expertise as well as focused reviews of existing data. The workgroup then developed a consensus model of the functional neuroanatomy of bipolar disorder based upon these data.

Results—Among the participants, a general consensus emerged that bipolar I disorder arises from abnormalities in the structure and function of key emotional control networks in the human...
brain. Namely, disruption in early development (e.g., white matter connectivity, prefrontal pruning) within brain networks that modulate emotional behavior leads to decreased connectivity among ventral prefrontal networks and limbic brain regions, especially amygdala. This developmental failure to establish healthy ventral prefrontal–limbic modulation underlies the onset of mania and ultimately, with progressive changes throughout these networks over time and with affective episodes, a bipolar course of illness.

Conclusions—This model provides a potential substrate to guide future investigations and areas needing additional focus are identified.

Keywords
amygdala; bipolar disorder; connectivity; fMRI; neuroimaging; prefrontal cortex

Introduction
Steady advances in neuroimaging techniques have produced a proliferation of neuroimaging research in bipolar disorder during the past decade. Although discrepancies exist among these research reports, a common theme is emerging: namely, that bipolar disorder arises from abnormalities within brain systems that modulate emotional behavior. However, more comprehensive functional neuroanatomic models of bipolar disorder, integrating these various neuroimaging findings, have been relatively uncommon. With these considerations in mind, the University of Cincinnati Department of Psychiatry and Behavioral Neuroscience organized a workgroup of several leading bipolar disorder neuroimaging programs in the US and UK to address these issues that arose from informal discussions among the participants. This workgroup met in Miami, FL, USA in December 2010. The goal of this meeting was to discuss the groups’ past and new work in order to develop a consensus regarding the functional neuroanatomy for bipolar I disorder. The group chose to limit itself to discussion of bipolar I disorder, given that the data in bipolar II disorder and related conditions are relatively sparse. The group also primarily focused on functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) studies consistent with the stated goals. All manuscripts arising from this workgroup were subjected to a full peer-review process and the accepted papers comprise this special issue of Bipolar Disorders (1–9). The International Society for Bipolar Disorders (ISBD) requested this group serve informally as an ISBD task force, to which group members agreed. This article represents a summary of and conclusions from those discussions. Consequently, this article is not intended to be viewed as an exhaustive review; instead, it represents the synthesis of these investigators in order to provide suggestions and a consensus model to guide future research in this important area.

Bipolar disorder and emotional networks
As the workgroup approached this topic, we started with the assumption that the brain systems most likely to underlie bipolar disorder involve those that modulate emotional control. This assumption was based on the long clinical history in which bipolar disorder is considered a primary disorder of mood. Indeed, bipolar disorder is defined by the occurrence of mania and characterized by recurring affective episodes. Nonetheless, these affective symptoms are accompanied by changes in cognition and neurovegetative symptoms, so that bipolar disorder could be considered a disorder of energy, sleep, or cognitive processes. As discussed throughout this special issue, neuroimaging studies have repeatedly identified abnormalities within brain emotional systems in bipolar disorder, and there is no specific evidence suggesting that approaching bipolar disorder as a primary disturbance of mood is flawed. However, as the field advances, other systems (e.g., those involved with sleep
regulation) may be relevant, perhaps in some subgroups, so considerations toward this end are warranted. Nonetheless, given the best available evidence, it is parsimonious to follow the assumption that bipolar disorder arises from dysfunction of brain networks that modulate emotional behavior.

Although the specific control of emotional function in humans is not completely defined, two ventral prefrontal networks appear to modulate emotional behavior (10–13). These networks and related brain regions are extensively reviewed by Townsend and Altshuler (1) and Blond et al. (2) in this issue, but will be briefly summarized here. Namely, both of these networks are similarly organized in that specific ventral prefrontal regions map to specific striatal, pallidal and thalamic brain areas to form iterative feedback loops that process information and modulate amygdala and other limbic brain areas (1, 10–13). One network originates in the ventrolateral prefrontal cortex and is thought to modulate external emotional cues, such as with affective face tasks (1, 14). The other network originates in the ventromedial (orbitofrontal) cortex and is thought to modulate internal emotional stimuli, namely stimuli that arise from internal feeling states, such as paradigms that involve inducing an emotional response to a specific cue (e.g., sadness in response to personalized events) (10, 15). Additionally, different voluntary and automatic (implicit) emotion regulation sub-processes have been identified, centered on ventrolateral and ventromedial prefrontal cortices, respectively (16). These networks serve as likely substrates for the functional neuroanatomy of bipolar disorder (Fig. 1), and they served as the primary focus for the workgroup and consequently for this summary (1).

**Discussion of functional neuroimaging considerations**

The emotional networks illustrated in Figure 1 are iterative feedback and feedforward systems that do not have a ‘starting point’ *per se*; however, since the amygdala is central to these networks, we will begin discussion there. The amygdala is an evolutionarily ancient structure that is responsible for generating flight/flight responses to threats. In humans, it is heavily innervated by an over-developed (relative to other animal species) and evolutionarily recent prefrontal cortex that likely serves to nuance the fight/flight response into the complex emotional behavior that defines human interactions. Along with other nearby medial temporal structures (including paralimbic cortex and hippocampus), the amygdala is responsible for emotion perception and regulation [as reviewed by Blond et al. in this issue (2); also see Chen et al. (14)]. In fMRI studies of bipolar disorder, amygdala dysfunction is commonly observed.

Indeed, one of the most consistent functional neuroimaging findings in bipolar disorder is excessive amygdala activation in response to affective faces during mania, compared with healthy subjects (17). For example, Altshuler et al. (18) used a facial affect matching task, against a baseline geometric form matching task (19), to study amygdala activation in bipolar manic and healthy subjects. The investigators observed increased left amygdala activation in the manic group during facial affect matching. Several other groups reported similar findings in bipolar disorder individuals during depression and remission and with other tasks (14, 20–24). For example, Strakowski et al. (25) observed increased bilateral amygdala activation in medication-free, early-course euthymic bipolar disorder patients while they performed a simple non-emotional continuous performance task (CPT). Patients were asymptomatic for at least one month prior to and one month after the fMRI scan and had also not been receiving medications for at least one month prior to the scan, eliminating these clinical confounds. Consistent with this finding, in their meta-analysis, Chen et al. (14) found increased amygdala activation in euthymia when region-of-interest (ROI) studies were examined. Increased amygdala activation during bipolar depression has also been reported (23), although Almeida et al. (22) found excessive amygdala activation only in
response to specific affective expressions (mildly sad). This latter study suggested that amygdala over-activation in bipolar disorder may occur only in response to specific affective cues, i.e., that it is sensitive to the salience and valence of the affective stimuli. Indeed, other investigators have not observed increased amygdala activation during depression (26, 27) and suggested that amygdala over-activation may be state dependent, as reviewed in this issue (1). This view is not consistent, however, with a recent large meta-analysis of fMRI studies in bipolar disorder (14).

Even during mania, not all studies observed increased amygdala activation in bipolar versus healthy subjects. Strakowski et al. (28) used the continuous performance task with emotional and neutral distracters (CPT-END) (10) to study healthy and bipolar manic subjects. The distracters were neutral and emotionally negative images taken from the International Affective Pictures System (IAPS). In this study, the manic subjects exhibited relatively blunted amygdala response to distracters as compared with healthy subjects. This observation suggested that, in the context of a predominantly attentional task, amygdala was dysregulated, but differently than in facial affects tasks, despite the emotional distracters. However, this finding might nonetheless reflect an underlying over-activation of amygdala during mania (i.e., at baseline), which limited the additional amygdala response that could occur, leading to an apparent blunted activation. The authors suggested that restricted amygdala response flexibility might contribute to the apparent discrepancies in study findings as well as the emotional dyscontrol of bipolar illness (17, 28). Because typical fMRI studies are limited in that absolute activation cannot be assessed, but instead activation always represents a relative (subtraction) measure between two study conditions, interpreting these types of discrepancies among studies can be difficult. Perhaps the more parsimonious view of fMRI findings in bipolar disorder, then, is that, independent of the direction of abnormalities, across bipolar mood states amygdala activation/function is abnormal in response to a variety of stimuli, and is sensitive to both mood state and emotional valence of the task (14, 17, 28). This possibility suggests that future studies of amygdala function in bipolar disorder need to examine the impact of mood interacting with the emotional valence of stimuli in order to clarify the abnormal reactivity of this structure in bipolar disorder. Nonetheless, even with inconsistencies across studies, because many investigators found that amygdala activation is different in subjects with bipolar disorder than in healthy subjects, even in unaffected relatives of bipolar disorder patients (29), amygdala dysfunction appears to be a feature of this condition.

As noted previously, amygdala function in the human brain is modulated by the prefrontal cortex, in particular, ventral prefrontal regions (i.e., orbitofrontal and ventrolateral prefrontal cortex). Consequently, consistent with evidence of dysfunctional amygdala activity in bipolar disorder, abnormalities in ventral prefrontal cortex are also commonly observed. Specifically, decreased fMRI activation in both lateral and medial (orbitofrontal) ventral prefrontal brain regions has been reported across a variety of cognitive tasks and mood states (e.g., 18, 28, 30–33) and also with positron emission tomography (PET) (34). Indeed, ventrolateral prefrontal cortex (specifically inferior frontal gyrus) abnormalities were observed in a recent meta-analysis of fMRI data (14). For example, as reported in this issue, Townsend et al. (7) studied 32 euthymic bipolar disorder subjects and 30 healthy subjects during a Go/NoGo response inhibition task using fMRI. The euthymic bipolar disorder group demonstrated significantly less inferior frontal cortical activation. Similarly in this issue, Liu et al. (6) studied 76 bipolar disorder subjects across a variety of mood states and 58 healthy subjects while performing a facial affect task. Ventral cortical activation was decreased in the bipolar disorder group independent of mood state, although laterality differences were observed; i.e., patients in elevated mood states exhibited lower activation in the right ventral prefrontal cortex, whereas, depressed patients demonstrated decreased activation in the left. Other studies of bipolar disorder patients who are asymptomatic
suggest that some prefrontal recovery may occur during euthymia to compensate for amygdala over-activation in order to maintain attentional and memory performance (25, 35). Together these studies suggest that amygdala dysregulation coupled with ventral prefrontal under-activation may indicate failure of healthy ventral prefrontal network modulation of the limbic brain in bipolar disorder, potentially providing a functional neuroanatomic basis for affective symptoms. Effectively, with diminished prefrontal modulation, the limbic brain is hypothesized to be dysregulated, leading to the emotional extremes of mania, depression, and mixed states. This hypothesis is developed in more detail in this issue by Blond et al. (2).

Consistent with this hypothesis, studies have reported differences in functional connectivity between ventral prefrontal regions and amygdala in bipolar compared with healthy subjects. Functional connectivity can be measured in several ways, but in general it provides an assessment of how brain activation correlates among different regions over time (17). Higher temporal correlations between two or more brain areas suggest that they are more strongly linked, i.e., functionally connected (17). Folan et al. (30) found decreased functional connectivity between amygdala and lateral ventral prefrontal cortex in manic compared with healthy subjects, a finding replicated by Chepenik et al. (36). Abnormalities in functional connectivity between ventral prefrontal cortex and amygdala may also occur during depression. Since the prefrontal cortex remains physically connected with the limbic brain in bipolar individuals, these findings tell us that the synchronization between prefrontal cortex and amygdala is disrupted in depression and mania. For example, work by Almeida et al. (37) measuring effective connectivity between amygdala and ventromedial prefrontal cortex, suggested a pattern of abnormal bilateral disconnectivity between ventromedial prefrontal cortex and amygdala in bipolar depression in response to positive emotional (happy) faces, a pattern that did not occur in individuals with unipolar depression. However, this same disconnectivity was not seen with other mood states, suggesting that the loss of prefrontal–amygdala synchronization was sensitive to the valence of the stimulus. As another example, in this issue Mourão-Miranda et al. (8) applied a novel discrimination paradigm to use fMRI to differentiate between bipolar and unipolar depression. Specifically, differential fMRI activation responses between happy and neutral faces seemed to distinguish depressed bipolar disorder subjects from individuals with unipolar depression and from healthy subjects. Although these findings suggested that as fMRI and functional connectivity techniques progress they may develop diagnostic utility, namely, distinguishing bipolar from unipolar depression (38), understanding how brain regions interact and networks function remains rudimentary. Also as reviewed elsewhere in this issue (2), white matter abnormalities, observed with diffusion tensor imaging, in particular, suggest that impairments in functional connectivity observed in bipolar disorder may have a specific neuroanatomic basis, namely loss of white matter connections among emotional brain regions that may predate the onset of illness and progress with recurrent affective episodes (39–43). Again, these failures of development may put individuals at risk for inadequate prefrontal modulation of limbic brain activity, leading to dysregulation of mood and the development of extreme mood states.

As illustrated in Figure 1, ventral prefrontal cortical areas that modulate human emotional function do so within a structure of relatively independent prefrontal-striatal-pallidal-thalamic iterative circuits (17). Consistent with the previously discussed findings in ventral prefrontal cortex and amygdala, investigators have commonly observed abnormal activation (often excessive) throughout subcortical structures in these networks, particularly during mania and euthymia [for reviews see Marchand and Yurgelun-Todd (44) and Blond et al. in this issue (2)]. However, a recent meta-analysis found both putamen under-activation as well as basal ganglia overactivation, complicating this interpretation (14). Consistent with the former, Liu et al. (6) (in this issue) observed decreased ventral striatal responses to happy
and neutral faces across mood states in a large group of bipolar disorder subjects. Similarly in this issue, Townsend et al. (7) observed decreased activation during a response inhibition task in bipolar disorder subjects compared with healthy subjects in several subcortical brain regions including left caudate, bilateral globus pallidus, lateral putamen, and right thalamus. Discrepancies among studies in relative activation differences between bipolar disorder and healthy subjects may, as with amygdala, support a model of variable dysregulation depending on characteristics of the sample, the task (i.e., valence, salience), or other clinical or demographic factors. For example, in this issue, Jogia et al. (9) observed a significant sex-by-diagnosis interaction within the basal ganglia during incentive decision making in samples of bipolar disorder and healthy men and women. Regardless, striatum and thalamus are critical integrative structures within prefrontal iterative circuits and provide ‘cross-talk’ among otherwise relatively independent prefrontal networks (17). Consequently, disruption of this integrative function may contribute to dysregulation of emotional processes in bipolar disorder.

Another integrative structure that may contribute to the functional neuroanatomy of bipolar disorder is the anterior cingulate cortex. The anterior cingulate cortex sits at the intersection of dorsal (cognitive) and ventral (emotional) prefrontal functions, thereby integrating information processing (10, 28). Specifically, the ventral anterior cingulate, including subgenual regions, is responsive to emotional stimuli and the dorsal cingulate more responsive to cognitive stimuli; the integration may therefore occur along a gradient along this structure. Previous work suggested that emotional and cognitive processing are principally modulated by ventral and dorsal prefrontal regions, respectively, and interact in such a manner to function reciprocally; this reciprocal interaction may be modulated by anterior cingulate (10, 45, 46). In imaging studies of bipolar disorder, the anterior cingulate frequently activated differently in bipolar disorder than healthy subjects (e.g., 6, 18, 28, 46). Often, but not in every study, anterior cingulate activation was increased in bipolar disorder subjects during mania compared with healthy subjects, whereas the converse occurred during bipolar depression (18, 47–49). During mania, the anterior cingulate also demonstrated abnormal functional connectivity with the amygdala (50). During euthymia, the emotional aspects of the anterior cingulate may be over-activated (51), whereas the more cognitive aspects are under-activated (52), indicating a possible dissociation within the anterior cingulate to different types of stimuli in different bipolar mood states (17).

Taken together, fMRI studies suggest dysfunction throughout the putative emotional brain networks illustrated in Figure 1. This observation is important as it does not support single-region models of bipolar disorder, but instead suggests a more general dysregulation of these networks. Unfortunately, fMRI results are not always consistent and several factors, including the type of cognitive task used as a probe and the mood state of the patients, appear to impact differences between bipolar disorder and healthy subjects. As reviewed elsewhere, there are relatively few studies using fMRI to compare subjects with bipolar disorder to other psychiatric groups, such as schizophrenia (5) or unipolar depression (38). Nonetheless, the existing studies suggest that disruption in these ventral prefrontal–striatal–pallidal–thalamic iterative networks as they modulate amygdala may be relatively unique to bipolar I disorder. Clearly, more studies are needed across diagnostic groups to substantiate this assumption. Moreover, as noted previously, the use of pattern recognition techniques with wholebrain functional neuroimaging in response to emotional stimuli shows promise as a new method to differentiate bipolar from unipolar depressed individuals case by case (8).

**Bipolar disorder, imaging, and developmental considerations**

Bipolar disorder most commonly begins in late adolescence and may include pre-syndromal affective, attentional, and behavioral disruptions prior to the first manic or depressive
Moreover, during the first few episodes, bipolar disorder is progressive, in that euthymic periods between affective exacerbations steadily shorten (17). Together these clinical observations suggest that the neurophysiology of bipolar disorder is progressive, particularly early in the illness; consequently, studies of early course and at-risk subjects may be particularly informative toward clarifying this neurophysiology prior to the confounding effects of medication exposure [see in this issue Hafeman et al. (4)] and illness course. Schneider et al. (3) discuss in detail in this issue these study considerations.

Neuroanatomic studies of young people at-risk for bipolar disorder (i.e., with bipolar disorder parents) have had mixed results, but often show few differences with youth of healthy parents (55). However, two studies found increased brain volumes in prefrontal (56) and parahippocampal cortices (57). These authors suggest that increased cortical volume might represent failed pruning or other developmental processes prior to illness onset (55). Consistent with this suggestion, several studies reported white matter abnormalities prior to illness onset in at-risk groups. These findings include increased white matter hyperintensities (58, 59), and evidence of decreased callosal myelination using the proxy of T1 signal intensity in at-risk youth compared with youth of healthy parents (60). Using DTI to examine white matter structure in more detail, several investigators observed abnormalities in at-risk youth, particularly in white matter tracts involved in the connections among regions of the emotional networks of Figure 1. Versace et al. (41) used DTI to study healthy youth with bipolar disorder parents and healthy youth with healthy parents. They found that, even prior to any symptoms or behavioral disturbances, the at-risk group exhibited white matter abnormalities in the corpus callosum and temporal white matter tracts. This finding was also reported in a large cohort by Sprooten et al. (61). Moreover, Frazier et al. (40) observed decreased fractional anisotropy (FA) in the superior longitudinal fasciculus in children at-risk for bipolar disorder compared with healthy subjects; decreased FA may represent loss of bundle coherence within these tracts. This study also examined youth who had developed bipolar disorder; the affected youth demonstrated significantly more widespread FA reductions throughout prefrontal and frontal areas. Consistent with that observation, Adler et al. (39) found decreased FA, suggesting disrupted organization, in prefrontal white matter tracts in adolescent first-episode manic patients, similar to findings by Kafantaris et al. (62) and Barnea-Goraly et al. (63). Together, these studies suggest that white matter abnormalities (i.e., connections among brain regions) may exist well before illness onset, underlying and potentially contributing to additional developmental abnormalities during adolescence, including the loss of prefrontal functional connections with amygdala. These alterations in neurodevelopment may lead to the onset of affective dysregulation, and the eventual onset of the first manic episode. The paucity of longitudinal studies, however, makes this suggestion difficult to directly evaluate.

Reduced amygdala volume has been consistently reported in bipolar disorder compared with healthy adolescents [for a review see Schneider et al. (3) in this issues, as well as Pfeifer et al. (64)]. Recently, Bitter et al. (65) found that, during the year after a first manic episode, bipolar disorder patients failed to demonstrate normal amygdala growth seen in both healthy and attention-deficit hyperactivity disorder (ADHD)-affected youth. This early course effect may underlie the initial development and progression of illness. However, later in life, amygdala volumes have been reported to be larger in bipolar disorder than in healthy subjects (17, 64); these findings suggest dysregulated amygdala growth might be a feature of bipolar disorder throughout the early course, although medication effects may contribute to later-course findings (66). In contrast, amygdala volumes do not appear to be abnormal in at-risk youth prior to illness onset (55).

Perhaps not surprising, then, ventral prefrontal reductions have also been found in bipolar disorder youth that are inversely correlated with age, i.e., that appear to progress during this
critical developmental time (67, 68). Similar relationships between age and regional brain structural volumes have been observed for the striatum (69). Consistent with progressive changes as the first manic episode approaches, Gogtay et al. (70) found decreases in subgenual cingulate cortex volumes after, but not before, illness onset in at-risk youth, suggesting this change may be associated with the onset of mania. Functional imaging studies in bipolar disorder youth typically report abnormalities throughout the emotional networks of Figure 1, similar to adults, although prefrontal cortical function may be relatively spared in the early course of illness (55). Together, these findings suggest that abnormalities in the networks of Figure 1 in bipolar disorder progress during the early course of illness, consistent with the corresponding decrease in euthymic intervals between affective episodes (17). These clinical and neuroimaging features support a model of abnormal anatomic connections between prefrontal cortex and limbic brain structures that evolve into the functional disconnection of mania during adolescence, that progresses over time to produce the recurrent, chronic illness of bipolar disorder. Many of these considerations have been reviewed in detail in this issue by Schneider et al. (3). As noted in that review, conclusions about the specific neurodevelopmental trajectories that lead to bipolar disorder remain inadequately studied and poorly understood. In particular, there is a distinct need for prospective longitudinal studies of individuals at risk for bipolar disorder to identify deviations from healthy neurodevelopment and the associated behavioral manifestations of these changes.

Summary and future directions

Discussion in this workgroup led to the hypothesis that disruption in early development (e.g., white matter connectivity, prefrontal pruning) within brain networks that modulate emotional behavior (Fig. 1) leads to decreased connectivity between ventral prefrontal networks and limbic brain structures, including (perhaps especially) amygdala. This loss of connectivity is associated with abnormal functional responses of emotional networks to various cognitive and emotional tasks in imaging studies, as well as abnormal development of the component brain regions (e.g., failure of amygdala to mature normally and disruption of prefrontal modulation of limbic structures). Dysregulation of limbic brain then leads to loss of emotional homeostasis resulting in mood instability. In the absence of a healthy prefrontal-striatal-pallidal-thalamic-limbic brain networks that can restore this homeostasis, bipolar disorder individuals are at risk for developing extreme mood states and switching among mood states, as well as developing mixed states as different unregulated systems oscillate in the absence of homeostatic control. During euthymia, recovery of prefrontal function, along with compensation from other brain regions as observed by Strakowski et al. (14), temporarily restores homeostasis; nonetheless, the underlying functional neuroanatomic abnormalities leave the bipolar disorder individual at risk for disruption of this fragile homeostasis under even minor stress in the face of ‘stably unstable’ prefrontal-striatal-thalamic-amygdala mood networks (17). In short, we then hypothesized that developmental failure to establish healthy ventral prefrontal–amygdala networks underlies the onset of mania and ultimately, with progressive changes throughout these networks over time, a bipolar course of illness (17). This model provides a potential substrate to guide future investigations using fMRI as well as other imaging modalities.

Functional MRI is data rich; a single study will produce thousands of measurements from voxels throughout the brain. Consequently, investigators are faced with the challenge of determining which measurements signify a meaningful event. Typical approaches (e.g., statistical parametric mapping) involve combinations of data clustering and controlling for multiple comparisons to identify voxels that differ from other voxels and that indicate a response to a task or difference among subjects or groups. However, most approaches are agnostic and make no assumptions that activation in any particular voxel is any more
meaningful than in any other voxel. Consequently, by its very nature, these approaches are exploratory and better designed for hypothesis generating than hypothesis testing. Moreover, these approaches often provide difficult to interpret findings in which activation or activation differences occur in brain regions distinct from those that other neuroscience investigations suggest are prominently involved in a given cognitive task or for a specific pathology (e.g., bipolar disorder). These exploratory voxelwise approaches dominate the bipolar neuroimaging literature, so that it is almost certain that these limitations contribute to the variability seen in fMRI studies of bipolar disorder.

ROI approaches address many of these shortcomings and can be used to directly test specific functional neuroanatomic hypotheses, like those presented here. However, these approaches are conservative by definition, limiting discovery of novel findings that might better explain the functional neuroanatomy of bipolar disorder than our current models. Indeed, the neural substrates underlying behaviors impacted by bipolar disorder are incompletely understood. Perhaps the solution to these limitations is to combine ROI and voxelwise techniques, thereby ensuring specific testing of hypotheses (ROI approach) to support or eliminate functional neuroanatomic models, while continuing to explore alternative models (voxelwise techniques) through novel hypothesis-generating findings, while carefully defining the role of each approach. Regardless, as suggested by Logothetis (71), fMRI findings must be interpreted within the context of known existing networks and network functions; hence our reliance in this discussion on the networks illustrated in Figure 1.

It is almost certain that these issues, coupled with the heterogeneity of bipolar disorder patients and, at times, insufficient clinical rigor when defining phenotypes, contributes to disparity among studies. Differences in data analytic techniques and among cognitive paradigms further contribute variability across studies. Medication exposure may complicate interpretation, as reviewed in detail in this issue by Hafeman et al. (4).

With these considerations in mind, we have several suggestions for future research aimed at defining the functional neuroanatomy of bipolar disorder:

1. **Study narrowly defined groups.** A significant complication across and within studies is the inclusion of heterogeneous bipolar and affectively ill patient samples. Combining unipolar and bipolar disorder, and then various types of bipolar disorder (e.g., type I, type II, not otherwise specified, etc.) may help to increase statistical power by virtue of ease of recruitment and increasing sample sizes, but almost certainly introduces considerable noise. Unlike genetic studies in which a broader inclusion of subjects might help to identify a phenotypically variably expressing gene, there are few advantages in neuroimaging to studying heterogeneous rather than narrowly defined samples. Neuroimaging research benefits from well-designed studies of similar subjects to minimize confounds. Consequently, for the field to move forward, attempts to define more homogeneous clinical samples and then compare across these samples are needed to better understand bipolar neurophysiology. For example, there is a lack of well-powered studies comparing bipolar I and II disorders across different types of cognitive probes and mood states. There are few studies examining the impact of comorbid drug and alcohol abuse on neuroimaging parameters, even though these types of bipolar disorder patients may have differing etiologies (72). Studies of carefully defined subtypes and subgroups are needed to allow investigators to better understand the impact of variable clinical confounds and symptomatic expression in order to interpret differences between groups and across studies. Additionally, studies of carefully defined mood states to create symptomatically homogeneous samples, to again control potential confounds of symptomatic epiphenomena, are needed. Very large samples will be needed for this type of work, but a number of groups have been
scanning subjects for years and are likely developing such samples. Alternatively, working across institutions and geography by collaborating with common imaging approaches and cognitive paradigms might allow these samples to be collected relatively quickly. The Research Domain approach suggested by the National Institute of Mental Health (NIMH) provides one guideline toward this type of homogeneity. These types of studies are unquestionably harder than recruiting ‘all comers’, but the field has moved to the point where the easy studies have been done and advances will only occur with extra effort.

2. **Longitudinal studies across mood states.** The model proposed here suggests that lack of emotional homeostatic mechanisms related to prefrontal modulation of limbic brain underlies mood instability, mood switching, and developing extreme mood states. Specific studies of patients as they experience different mood states and symptom domains will aid our understanding of how these affective states are represented in activation maps of the bipolar brain. There have now been a number of studies that include patients in several mood states within a single sample [e.g., Liu et al. (6) in this issue], which is one approach to identify potential state and trait related activation effects. However, these comparisons are limited by the inevitable differences among the subjects in the different mood states. To advance this work, longitudinal studies are needed within individual subjects as they recover from mania or depression, switch into an alternate affective episode, develop mixed states, or achieve and maintain euthymia. Although medication effects cannot be ignored, they have likely been overstated [see Hafeman et al. (4)], so that studies can be designed around this potential confound. Again, these types of studies are difficult, but are a critical step to understand how brain activation patterns changes across the dynamic course of bipolar illness.

3. **Longitudinal studies of treatment response.** In this issue, Hafeman et al. (4) suggest that medication effects minimally impact fMRI brain measurement differences between bipolar disorder and healthy subjects; however, they also additionally suggest that medication effects may be normalizing. Studying medication affects as they relate to symptomatic improvement provides opportunities to understand how the functional neuroanatomy of bipolar disorder changes in response to treatment and symptom resolution. A few studies to this end have been performed (4), but the processes leading to symptom resolution remain poorly understood. Moreover, lithium is a unique treatment in psychiatry in that it is relatively specific for a specific condition—namely, bipolar mania. Lithium, then, could serve as a pharmacological probe in bipolar disorder that may have unique effects in this population to help define the neurophysiology of illness. Indeed, recruiting a lithium responsive sample would be one way to narrow the phenotype. Larger studies of the impact of different treatment interventions on brain function across time and among medications are needed to define how medications improve symptoms in bipolar disorder. Moreover, such studies offer the promise of identifying treatment response markers and predictors to eventually guide treatment. Since most psychotropic medications have delayed responses in bipolar mood states, studies with scans repeated frequently in early treatment (e.g., every two to three days during the first two weeks) might identify initial changes in function neuroanatomy (e.g., restoration of ventrolateral prefrontal connectivity to amygdala) that predict treatment response. These studies may also clarify how our treatments succeed (or fail).

4. **Studying early course progression.** It is well-established that the early course of bipolar disorder is progressive (17). These clinical observations suggest a progression in the neurophysiology of the illness as it evolves from a single event
(manic episode) to a recurring condition that includes periods of mania, depression, mixed states, and remission. Based on the considerations raised in this article, we suggest that during the early course of bipolar illness ventral prefrontal cortical regions fail to form typical connectivity with and modulation of limbic brain areas to form healthy iterative modulatory prefrontal-striatal-pallidal-thalamic networks. As episodes occur, ventral prefrontal modulation and emotional homeostasis further deteriorates, leading to the illness course progression observed. Whether this progression represents direct effects of affective episodes or instead failure of healthy development (or the intersection of these two) is not known. Longitudinal studies within individual subjects during early illness progression to identify if and how these disruptions in development occur are critically needed. These studies, again difficult, have the potential to provide significant gains by controlling for many other confounding features (e.g., long-term medication exposure, accumulated psychosocial impairment, substance abuse). Specific studies are in need of development in the networks–shown in Figure 1—in both healthy and at-risk youth and recent-onset bipolar disorder subjects to address these considerations. Moreover, these latter two populations are ideal to investigate the intersection of genetic risks and structural and functional brain changes on illness expression and course.

5. **Testing specific hypotheses.** Neuroimaging studies using fMRI in bipolar disorder have been predominantly designed using voxel-wise comparison techniques. As noted, by definition, these are exploratory and hypothesis generating analytic methods. To move forward, the field now needs to advance toward examining how specific networks function in the various phases and across the course of bipolar disorder using more focused analytic models. Functional and effective connectivity models as reviewed by Blond et al. (2), or pattern analyses approaches as described in the original paper by Mourão-Miranda et al. (8) provide suggestions toward this end.

6. **Advancing cognitive probes.** As noted, part of the variability across studies likely reflects differences in how emotional networks respond to probes of varying emotional salience and valence. This issue becomes even more complex across different mood states that might impact these responses and relationships. Moreover, probes have not been consistently employed because they have been shown to specifically activate the regions or networks of interest, but rather have often been incorporated simply because they have commonly been used. There is no defined cognitive probe that identifies a specific cognitive impairment or functional activation abnormality for bipolar disorder in any phase; consequently, future research efforts need to be developed for refining and narrowing the effects of various probes, first, in healthy populations to ensure that the probe activates the networks of interest, and then, across the phases of narrowly defined bipolar disorder groups to understand the interactions of stimuli salience and valence. With this progress, comparisons across fMRI studies will become more meaningful.

7. **Integrating other imaging modalities.** Blood-oxygen-level-dependent (BOLD) fMRI, by itself, provides important functional neuroanatomic information about bipolar disorder, but is limited in its ability to understand mechanisms underlying observed differences from healthy subjects. It is now well established that groups of bipolar disorder subjects have differences in these BOLD fMRI patterns than healthy subjects, but what these differences mean mechanistically is not known. To advance this work from where in the brain problems exist to what those problems are, integration with other imaging (and genetic) techniques are the necessary next step. Multi-modal imaging integrating fMRI with MRS, connectivity measures,
DTI, other structural imaging and techniques, particularly within the longitudinal and narrowed designs we have suggested, offer significant promise to really advance our understanding of the pathophysiology of bipolar disorder.

8. Comparing with other conditions. Finally, most neuroimaging studies have compared bipolar disorder samples to healthy subjects. Although these studies have provided the bulk of the information leading to the discussions in this article, in the end, these studies tell us nothing about the specificity of findings to bipolar disorder versus general effects of mental illness. Indeed, abnormalities within the networks of Figure 1 do not appear to be unique to bipolar disorder. Many more carefully designed studies comparing bipolar disorder to related groups are needed, as reviewed in this issue by Whalley et al. (5) (schizophrenia) and as exemplified by Mourão-Miranda et al. (8). Again, these studies are often more difficult as individual research groups tend to, necessarily, maintain a narrow diagnostic focus, so that cross-group collaborations (e.g., between bipolar disorder and schizophrenia research programs) are required. These studies are, however, the only way that abnormalities specific to bipolar disorder can be defined.

In summary, we suggest that to advance our understanding of bipolar disorder using neuroimaging, studies will need to move toward testing specific abnormalities within recognized brain networks, such as the emotional networks of Figure 1, ideally across mood states and over time within the same subjects. Patient samples must be more carefully evaluated and narrowly defined clinically, and contrasted with similarly narrowly defined alternative patient samples. Longitudinal approaches will help to better define what brain changes are directly associated with specific course of illness features, which is particularly relevant for understanding a dynamic and progressive condition like bipolar disorder. Similarly, longitudinal studies of treatment response can clarify the impact that medications have on brain imaging measures in this patient population as well as potentially clarify mechanisms underlying treatment response. Studies differentiating patients who do or do not respond to a particular treatment may also identify treatment response markers or predictors, as well as define what is unique to bipolar disorder with certain relatively specific treatments (e.g., lithium or lamotrigine). Studies of at-risk youth and early course patients provide the opportunity to define the initial neurophysiological progression that leads to the onset of illness, and that may, therefore, represent the functional neuroanatomic etiology of bipolar disorder. By framing investigations within the context of developmental functional neuroanatomic models, studies may be able to increase comparability. As you read through this special issue, please keep these considerations in mind and be challenged to develop better approaches to advance our understanding of the complexities that underlie the neurophysiology of this uniquely human condition. As we imply here, many of the easy imaging studies have already been done, so in order for the field to properly advance more sophisticated study designs—that may also be more expensive—will be needed. We believe the field is growing, is strong, and is up to this challenge.

Acknowledgments

This workgroup was supported in part by the Department of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, OH, USA. This manuscript was supported in part by NIH grants P50 MH077138 (SMS); R01 MH076971 and R01 MH088371 (MLP); 2R01MH070902, 1RC1MH088366, and R01MH069747 (HPB); 5R21MH075944, 1K24MH001848, and 1R01MH084955 (LLA); R01 MH077047 (KDC); R01 MH080973 (MPD); and R01 MH078043 (CMA).

References


Fig. 1.
Schematic of the proposed ventrolateral and ventromedial prefrontal networks underlying human emotional control [adapted with permission from Oxford University Press (17)]. G. = globus; PFC = prefrontal cortex; OFC = orbitofrontal cortex; BA = Brodmann’s area.