bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Review Article

On the transdiagnostic nature of peripheral biomarkers in major psychiatric disorders: a systematic review

Jairo V. Pinto[1], Thiago C. Moulin[2], Olavo B. Amaral[2]

[1]*Laboratory of Molecular Psychiatry, Department of Psychiatry, Federal University of Rio Grande do Sul. Porto Alegre, RS, Brazil.*

[2] *Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro. Rio de Janeiro, RJ, Brazil.*

Corresponding author:

Olavo B. Amaral
Instituto de Bioquímica Médica
Universidade Federal do Rio de Janeiro
Cidade Universitária, 21941-590
Rio de Janeiro, RJ, Brazil
Phone: (21) 2562-6789
Fax: (21) 2270-8647
E-mail: olavo@bioqmed.ufrj.br

Funding: CNPq, FAPERJ

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ABSTRACT

The search for biomarkers has been one of the leading endeavors in biological psychiatry; nevertheless, in spite of hundreds of publications, it has yet to make an impact in clinical practice. To study how biomarker research has progressed over the years, we performed a systematic review of the literature to evaluate (a) the most studied peripheral molecular markers in major psychiatric disorders, (b) the main features of studies in which they are proposed as biomarkers and (c) whether their patterns of variation are similar across disorders. Out of the six molecules most commonly present as keywords in articles studying plasmatic markers of schizophrenia, major depressive disorder or bipolar disorder, five (BDNF, TNF-alpha, IL-6, C-reactive protein and cortisol) were the same across the three diagnoses. An analysis of the literature on these molecules showed that, while 65% of studies were cross-sectional and 66% compared biomarker levels between patients and controls in specific disorders, only 10% presented an objective measure of diagnostic or prognostic efficacy. Meta-analyses showed that variation in the levels of these molecules was robust across studies, but also similar among disorders, suggesting them to reflect transdiagnostic systemic consequences of psychiatric illness rather than diagnostic markers. Based on this, we discuss how current publication practices have led to research fragmentation across diagnoses, and what steps can be taken in order to increase clinical translation in the field.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

INTRODUCTION

With the growth of biological psychiatry and the widespread adoption of diagnostic classifications, the concept of diagnostic biomarkers has loomed promisingly in the horizon as a major goal of biological psychiatry (Kupfer *et al.* 2002; Insel & Quirion 2005). Although the idea of objective diagnostic tests in psychiatry is not new, and goes back to early promises such as the dexamethasone suppression test (Arana 1985), interest in biomarkers has grown exponentially over the last two decades, as shown by the steep rise in articles including the words "biomarker" and "psychiatry" (Fig. 1A). Not only part of the scientific literature, but also news pieces in scientific journals (Haag 2007) and in the lay media (Nutt 2016) have spread the idea that objective tests could eventually trump the centuries-old method of using clusters of symptoms to diagnose mental illness.

Nevertheless, the application of biomarkers in the psychiatric clinic is still very limited. As recently discussed by various authors (Kapur *et al.* 2012; Lancet Psychiatry 2016), problems for translating research findings into clinical practice include the biological heterogeneity of diagnostic constructs (Heinrichs 2004; Hyman 2010; Pavão *et al.* 2015), the emphasis on statistical rather than clinical significance (Sterne & Davey Smith 2001; Perlis 2011), the use of extreme comparisons between prototypical patients and healthy controls (Kapur *et al.* 2012; Lancet Psychiatry 2016) and issues such as low statistical power, publication bias and lack of replication (Ioannidis 2005; Kapur *et al.* 2012). Moreover, the controversy over the soundness of the DSM as a framework for biological psychiatry (Cuthbert & Insel 2013), as well as evidence showing that major psychiatric disorders are promiscuous in terms of genetic loci (Smoller 2013), risk factors (Caspi *et al.* 2014) anatomical

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

substrates (Goodkind *et al.* 2015; Sprooten *et al.* 2016) and treatment (Chouinard 2006)**,** have led some to argue that the greatest promise in biomarker development might lie not in diagnosis, but in the prediction of prognosis and/or treatment response (Davis *et al.* 2014).

Although such advice seems generally sound, it is uncertain whether such concerns have made a significant impact on biomarker research. Moreover, due to the high degree of separation of psychiatric research into different disorders, the question of whether the promiscuity among disorders observed for genetics, anatomy and risk factors also applies to peripheral biomarkers has only been studied in isolated cases (Fernandes *et al.* 2013; Goldsmith *et al.* 2016). With this in mind, we decided to (a) systematically investigate what are the most studied plasmatic markers for major psychiatric disorders, (b) perform a systematic review of the experimental design features of articles evaluating them as biomarkers and (c) on the basis of available meta-analytical evidence, investigate whether variation in their levels occurs similarly across different diagnoses.

METHODS

Search strategy and biomarker selection

We initially searched the PubMed and Scopus databases on June 5th, 2015 for articles containing the terms “biomarker” OR “biomarkers” AND (“serum” OR “blood” OR “plasma” OR “plasmatic”) AND one of six psychiatric disorders: (a) “bipolar disorder”, (b) “major depression” OR “major depressive disorder”, (c) “schizophrenia”, (d) “post-traumatic stress disorder” OR “PTSD”, (e) “attention-deficit-hyperactivity disorder” OR “ADHD”, (f) “autism” OR “ASD”. Although this search was non-exhaustive, as articles might study peripheral markers without using

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the term "biomarker", our objective was to specifically select papers on this topic for an automated keyword search – thus, the main goal was to minimize the presence of literature not relating to the theme.

Based on search results, we used a MATLAB script (available upon request) to count the numbers of articles containing each individual term in the author and index keywords of the Scopus database (the current PubMed XML field for keywords, "Other Terms", was not present before 2013 and thus could not be used). We then ranked words in order of frequency and manually scanned the tables to build a list of the molecules most frequently included as keywords for each individual disorder (Table 1), taking care to aggregate counts of terms relating to the same molecule (e.g. "BDNF" and "brain-derived neurotrophic factor").

Selection criteria and analysis of original article features

Using the three disorders with the largest number of articles in our search (major depressive disorder, schizophrenia and bipolar disorder), we chose five molecules appearing among the top six keywords for these disorders – brain-derived neurotrophic factor (BDNF), tumor necrosis factor alpha (TNF-α), interleukin-6 (IL-6), C-reactive protein (CRP) and hydrocortisone (cortisol) for further study. We then performed searches in PubMed and Scopus on August 31st, 2015 using the following terms as keywords (as well as MeSH terms for all of them in PubMed):

(a) one of the three disorders ("schizophrenia" / "bipolar disorder" / "major depression" OR "major depressive disorder")

(b) one of the biomarkers above, with inflammatory markers being pooled into a single search due to the high crossover of articles among them ("BDNF" /

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

"interleukin-6" OR "c-reactive protein" OR "tumor necrosis factor alpha" / "hydrocortisone").

(c) "biomarker" OR "biomarkers"

For each biomarker/disorder combination, we screened abstracts (or full text when necessary) for the following inclusion criteria: (a) original articles, (b) in English, (c) including human patients (d) with the disorder of interest, according to DSM or ICD criteria and (e) performing peripheral measurements of the molecule in question (for a summary of the search strategy, see Fig. 2A). We focused on peripheral levels of the molecules, and did not include articles evaluating genetic polymorphisms and/or epigenetic changes in their corresponding genes. We then obtained the full text of every article (except for five articles for which it could not be obtained) and extracted the following information on experimental design features (presented in Table 2, Fig. 3, Supplementary Tables 1-3 and Supplementary Fig. 1):

(a) the specific molecule measured (e.g. BDNF protein, proBDNF protein, BDNF mRNA), site of measurement (e.g. blood, saliva, sweat, urine, hair) and any specific conditions of measurement (e.g. after dexamethasone challenge).

(b) the group comparisons performed in the article (e.g. patients vs. controls; different states of the disorder; patients with the disorder vs. those with another disorder; before vs. after treatment).

(c) the correlations performed within the group of patients (e.g. with symptoms, with illness progression, with treatment response or prognosis, with other markers).

(d) whether the study was cross-sectional or longitudinal.

(e) for longitudinal studies, whether it belonged to a predictive study in a population without the disorder.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(f) whether it offered an objective measure of diagnostic/prognostic efficacy (e.g. receiver operating characteristic (ROC) curves, odds ratio, classifier accuracy) or only presented a statistical comparison between groups.

(g) the total number of markers studied in the article.

(h) the total sample size of the article.

(i) the journal in which it was published.

Information was extracted by two of the authors (O.B.A. and J.V.P.) after extensive discussion of criteria, and 20% of the articles were cross-checked by both investigators, yielding a kappa coefficient of .948 for categorical variables. Controversies were solved with the participation of a third investigator (T.C.M.).

Percentages of articles with or without each experimental design feature were initially calculated for each disorder/marker combination. We then calculated aggregate percentages for each disorder or marker and for the whole sample of articles; in these cases, articles appearing in more than one search were counted only once. Articles in which a particular feature was present for one marker/disorder but not for other(s) were considered to include that feature when calculating aggregate percentages. The complete database of articles retrieved by the search, as well as their categorization for each experimental design feature, is available in tables presented as Supplementary Data.

Meta-analysis search and biomarker evaluation across disorders

In order to study whether variation in biomarker levels is similar or distinct among the three analyzed disorders, we searched PubMed on December 1st, 2016 for articles containing the terms ("meta-analysis"[Publication Type] OR "meta-analysis as topic" [MeSH Terms] OR "meta-analysis"[All Fields]) AND (schizophrenia OR

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

bipolar disorder OR major depression OR major depressive disorder) AND (brain-derived neurotrophic factor OR bdnf OR interleukin-6 OR il-6 OR c-reactive protein OR tumor necrosis factor-alpha OR tnf-alpha OR cortisol OR hydrocortisone), as well as corresponding MeSH terms.

We included meta-analyses studying levels of at least one of the five biomarkers (i.e. BDNF, IL-6, TNF-alpha, C-reactive protein, cortisol) in at least one of the three studied disorders (i.e. MDD, bipolar disorder, schizophrenia) in which patient vs. control comparisons were performed (including comparisons for specific disease states). When more than one meta-analysis was available for the same biomarker, disorder and state, we selected the one with the largest sample size for inclusion in Table 3. Titles and abstracts were screened independently by two of the authors (J.V.P. and T.C.M.), and in case of disagreement consensus was reached using a third investigator (O.B.A). The search process for meta-analyses is described on Fig. 2B.

For each meta-analysis, we extracted sample size, number of studies, effect size with 95% confidence intervals, p value and the I^2 metric as a measure of heterogeneity. We also extracted *p* values for publication bias tests, the number of missing studies when funnel plot/trim-and-fill analyses were performed, and the corrected effect size after accounting for these studies when these were available. These data are included in Table 3 as well as in Supplementary Tables 4 to 8**.**

Statistical analysis

For comparisons of categorical experimental design features between articles on different markers or different disorders, we used chi-square tests for omnibus comparisons followed by Fisher's exact test between specific pairs of

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

markers/disorders. For comparisons of quantitative features (i.e. sample size and number of markers) between articles on different markers or disorders, we used Kruskal-Wallis tests with Dunn's test as a post-hoc. For correlations of experimental design features of articles with the year of publication, we used rank-biserial correlations for categorical variables and Spearman's nonparametric correlations for quantitative variables (as both year and sample size/number of markers presented heavily skewed distributions).

RESULTS

Our automated keyword search (see flowchart in Fig. 2) revealed that, out of the six molecules most commonly present as keywords in articles retrieved using "biomarker" and schizophrenia, major depressive disorder or bipolar disorder, five (BDNF, TNF-alpha, IL-6, C-reactive protein and cortisol) were the same across these disorders (Table 1). The temporal distribution of these articles (as well as of those fulfilling inclusion criteria) is displayed on Fig. 1 (B and C), revealing that articles on BDNF and inflammatory markers increased sharply after around 2005, whereas the number of articles on cortisol has remained relatively stable over two decades. Most articles on BDNF and inflammatory markers studied serum levels of these proteins, while cortisol articles usually studied serum or salivary cortisol levels, with frequent use of pharmacological challenges such as the dexamethasone suppression test as well (Supplementary Table 1).

An analysis of experimental design features of the literature on these molecules (Table 2 and Supplementary Table 2) showed that, whilst 66% of articles performed comparisons between patients and healthy controls, only 35% were

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

longitudinal studies. Moreover, only 10% presented an objective measure of diagnostic or prognostic efficacy, such as a receiver operating characteristic (ROC curve) or an odds ratio for values above a certain threshold (as detailed in Supplementary Table 3). Most of these numbers did not vary significantly across markers or disorders, but some differences were observed: (a) state comparisons and comparisons with other disorders were more frequent for bipolar disorder, (b) pre vs. post-treatment comparisons and correlations of levels with symptoms were more common for BDNF and less frequent for cortisol and (c) cortisol studies presented patient vs. control comparisons less frequently, but included measures of diagnostic efficacy much more commonly than those on other markers (especially inflammation). Median sample size was 71.5 (interquartile range, 45.2-140.8) and larger for inflammatory markers than for BDNF or cortisol ($p=0.004$, Kruskal-Wallis test). The median number of markers addressed in each study was 3 (interquartile range, 1-5), and also larger for studies on inflammation ($p<10^{-4}$, Kruskal-Wallis test).

Temporal trends for various types of comparison, as well as for median sample size and number of markers, are shown on Fig. 3 and Supplementary Fig. 1. No major differences were observed for the frequency of specific comparisons or correlations over time, and rank-biserial correlations testing for a time-related trend did not yield p values under 0.05 except for a slight decrease in the frequency of between-disorder comparisons ($\rho=-0.13$, $p=0.025$). Nevertheless, increases in the total number of articles meant that absolute numbers of articles tended to increase for most categories. Sample size also tended to increase over time ($\rho=0.21$, $p=4\times10^{-4}$), although its distribution varied widely in all periods.

We then moved on to analyze the variation of the five markers in each of the three disorders. Table 3 shows the results of the largest meta-analysis of patient vs.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

control comparisons for each marker/disorder combination (in various states of the disorder when available) (Stetler & Miller 2011; Munkholm *et al.* 2013, 2015; Modabbernia *et al.* 2013; Fernandes *et al.* 2014, 2015, 2016a, 2016b; Girshkin *et al.* 2014; Dargél *et al.* 2015; Polyakova *et al.* 2015; Haapakoski *et al.* 2015; Belvederi Murri *et al.* 2016; Goldsmith *et al.* 2016). Statistically significant variation in the level of all five molecules was found to occur in at least some states of all three disorders. Interestingly, the direction of variation was invariably the same across disorders for all markers (i.e. reductions in BDNF, increases in inflammatory markers and cortisol).

In mood disorders, differences between patients and controls were more marked among acutely ill patients (especially for mania in bipolar disorder) in the case of BDNF, IL-6 and TNF-α, while alterations in schizophrenia were observed both in acute and chronic illness for all markers. Cortisol and C-reactive protein levels appeared to be more fixed across disease states in mood disorders, with variations observed in euthymia as well. Effect sizes were generally comparable across disorders; however, evidence was more robust in some cases. Meta-analyses of BDNF, cortisol and C-reactive protein tended to yield more reliable differences than those studying IL-6 and TNF-α for mood disorders, while in schizophrenia differences were robust for all molecules. Heterogeneity between studies was almost invariably large, and evidence for publication bias was observed in a reasonable number of cases, although corrected effect sizes were infrequently provided.

For analyses other than patient vs. control comparisons (e.g. correlations with symptoms, illness progression, prediction of treatment response, etc.), meta-analytic results were not always available; moreover, results of individual studies were frequently contradictory. This might be related to differences in the clinical samples

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

studied, but also to the low reliability of these secondary analyses in most articles (see Discussion). We summarized results for these additional analyses in Supplementary Tables 4-8, using meta-analyses whenever possible, or individual studies from our sample when meta-analyses were not available. One should note, however, that our search strategy for original studies, unlike the one used for meta-analyses, was not planned to be exhaustive (due to use of the term "biomarker"); thus, these tables should not be taken to include the full range of available literature on the subject.

DISCUSSION

Although the promiscuity of peripheral biomarkers across diagnoses has received relatively little systematic attention in the literature, an automated keyword search revealed that the most frequently studied peripheral biomarkers are generally the same across major psychiatric disorders. Besides the biomarkers we chose to focus on (BDNF, IL-6, TNF-α, CRP and cortisol), other molecules were also studied in many disorders, such as oxidative stress markers (glutathione), other cytokines (IL-10, IL-1β) and the astrocytic protein S100B (Table 1). This is not in itself surprising, as one might expect that literature trends will promote interest in similar molecules across disorders. However, the fact that variation patterns were also similar across diagnoses suggests that there are real biological commonalities among psychiatric disorders in terms of their peripheral manifestations.

The reason for this similarity is worthy of discussion. On one hand, it can be thought of as a sign of overlap between the pathophysiology of major psychiatric disorders, a fact suggested by genetic (Smoller 2013), risk factor (Caspi *et al.* 2014)

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and neuroanatomical (Goodkind *et al.* 2015; Sprooten *et al.* 2016) similarities among diagnoses. After all, symptom-defined diagnostic boundaries in psychiatry should not be expected to "carve nature at its joints" all the way to the molecular level (Kendler 2006; Pavão *et al.* 2015). An additional interesting finding, though, is that many of the markers in question, such as cytokines and cortisol, are increased by various forms of chronic stress in both animal models and humans (McEwen 2007). This suggests that they might be related to the general "wear and tear" or allostatic load associated with mental or clinical illness (Kapczinski *et al.* 2008), rather than to the specific pathophysiology of any given disorder.

Also in favor of this view is the fact that alterations in cytokines and cortisol can be observed in acute stress models even in normal volunteers (Dickerson & Kemeny 2004; Steptoe *et al.* 2007; Slavich *et al.* 2010). And although a BDNF response to acute stress in humans has not been shown, there is evidence that its levels may be altered by traumatic life experiences (Kauer-Sant'Anna *et al.* 2007) as well as by acute and chronic stress in animals (Duman & Monteggia 2006). The fact that these molecules are altered as a consequence of stress does not preclude, of course, that they may also play a causal role in the development and/or progression of mental disorders (Kapczinski *et al.* 2008); however, it does suggest that they are likely to be nonspecific as diagnostic biomarkers.

Interestingly, alterations in both cytokines and cortisol levels have also been found in non-psychiatric medical conditions, such as coronary heart disease (Mendall *et al.* 1997; Danesh *et al.* 2004; Smith *et al.* 2005) and cancer (Martín *et al.* 1999; Rich *et al.* 2005). It is possible, thus, that low-grade inflammation related to psychiatric illness could be linked to the increased clinical morbidity and mortality observed in patients with mood disorders and schizophrenia, particularly due to

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

cardiovascular causes (Olfson *et al.* 2015; Ösby *et al.* 2016). Importantly, systemic inflammation can also have repercussions in the brain, and could be part of a feedback loop that leads to persistence of symptoms in disorders such as major depression (Gold 2014; Wohleb *et al.* 2016).

Thus, although the word "biomarker" is frequently used in a diagnostic sense, most of the literature using the term seems to focus on nonspecific markers of general psychopathology. This is not to say that these markers are devoid of clinical application, as there are multiple potential applications for biomarkers besides diagnosis (Davis *et al.* 2014) – the case has been made, for example, to use BDNF and cytokine levels as markers of illness activity (Fernandes *et al.* 2013) or progression (Kapczinski *et al.* 2009). However, the fact that the literature on biomarkers is still very fragmented across DSM-defined disorders (only 11% of articles in our sample compared markers across more than one disorder, and this percentage has actually decreased over time) probably limits the understanding of these molecules in this sense, and seems to argue in favor of transdiagnostic approaches for their study (Cuthbert & Insel 2013; Lancet Psychiatry 2016).

Evaluation of the clinical utility of biomarkers is also limited by the types of articles that have been found to be most prevalent in the literature – namely, cross-sectional studies focusing on comparisons between patients and healthy controls. Although these studies are obviously necessary as starting points for further research, and have built a solid case for the robustness of biomarker alterations in psychiatric disorders, they are probably not representative of the typical situation in which a marker would be clinically useful (Perlis 2011; Kapur *et al.* 2012). Adding to this, measures of diagnostic efficacy that quantify how much a biomarker adds to clinical reasoning, such as ROC curves or odds ratios, were very infrequent in our sample –

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in fact, they have been *less* reported for recent biomarkers than they were for the evaluation of the dexamethasone suppression test three decades ago (Arana 1985). Taking all of this into account, the available literature on peripheral markers, although extensive and convergent in many points, is still insufficient to adequately assess their clinical potential.

In a more qualitative note, studies varied widely in terms of methodology, sample size and quality of reporting. Particularly notable were discrepancies between the numbers of measured variables and/or proposed analyses in the methods sections and those reported as results. It was not infrequent for articles to examine a large number of biomarkers, correlate them with many clinical variables, and report only significant associations, in a practice best described as selective outcome reporting bias (Dwan *et al.* 2008). The combination of a large number of outcomes with selective reporting will inevitably increase the possibility that reported findings are false positives (Ioannidis 2005), especially in the absence of multiple comparison corrections, which were infrequently performed. Importantly, it also leads to non-reporting of negative findings, and thus bias attempts to systematically review the existing literature.

Meta-analysis of available data can solve some of these issues, and meta-analytic comparisons have shown variations in the assessed biomarkers to be generally robust in common psychiatric disorders (Table 3). However, meta-analyses are intrinsically limited by the quality of the data; thus, their results can be influenced by publication and outcome reporting biases, as recently studied for markers of bipolar disorder (Carvalho *et al.* 2016). Moreover, they have been most successful for comparisons that are prevalent among studies – i.e. those between patients and controls, and eventually between disease states or pre- and post-treatment levels.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Attempts to correlate biomarkers with symptom severity, illness progression or response to treatment have yielded less consistent findings in meta-analyses (with a few exceptions), probably because these correlations were not only less frequent, but also usually presented as secondary analyses, and thus more susceptible to bias.

Such a general overview of the literature naturally has its limitations. First and foremost, the use of the term "biomarker" in our searches was a deliberate choice for specificity over sensitivity, and probably led many studies referring to peripheral markers by other terms to be missed. Moreover, in some articles the markers we assessed were only one among many studied, and not necessarily the main focus of the work – thus, the prevalence of some analyses might have been larger if we had limited ourselves to studies focused on those markers. Finally, our choice to provide a bird's eye view of the literature has led us to focus on the rule rather than on the exceptions – thus, one should keep in mind that there are several articles in our sample that do go beyond patient vs. control comparisons. Moreover, although in terms of prevalence they constitute a minority, the absolute number of these articles has increased as the biomarker literature grows as a whole.

In spite of these limitations, we believe that our general conclusion – namely, that the most frequently studied peripheral markers in psychiatry are nonspecific markers of psychopathology across multiple disorders – is probably correct. In this sense, it is useful to discuss ways in which the literature can change in order to better assess their clinical utility. First and foremost, it is important to define what one means by "biomarker" on a case-by-case basis, as there are many ways in which a molecule might behave as a marker (Davis *et al.* 2014). Moreover, we must move from studies chasing statistical differences between groups to those that evaluate whether a biomarker can make a difference in clinical decision processes (Perlis

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2011; Kapur *et al.* 2012; Lancet Psychiatry 2016). Finally, these studies will likely require more rigid design and larger statistical power – in this sense, development of reporting guidelines for biomarker studies (Gnanapavan *et al.* 2014) pre-registration of study protocols (Kivimaki *et al.* 2013) and formation of multicenter consortia (Sullivan 2010) are foreseeable ways to improve reproducibility. Such changes might ultimately depend on reforming incentive systems for science, which currently favor translational promise over assessment of clinical utility (Joyner *et al.* 2016) and stimulate small-scale studies that maximize the possibility of significant results (Higginson & Munafò 2016; Smaldino & McElreath 2016).

If these changes indeed take place, peripheral biomarker research could follow the path of genetic epidemiology, in which underpowered studies with limited reproducibility (Siontis *et al.* 2010; Bosker *et al.* 2011) have gradually been replaced by much larger studies from multicenter consortia (Sullivan 2010) that can control for the number of comparisons performed and yield more reproducible data (Collins *et al.* 2013). Our data shows that this has not yet happened in the peripheral marker field: aside from weak trends for an increase in sample size and a decrease in the frequency of cross-disorder comparisons, hardly any of the literature patterns we studied seems to have changed significantly over a 20-year period. This picture could change in the near future, however, as the results of ongoing large-scale initiatives in biomarker research (e.g. Kennedy *et al.* 2012; Kessing 2016) become available.

An open question that arises from our data is whether the transdiagnostic nature of the markers we have studied is actually a general rule for peripherally measured molecules. It might be expected, after all, that serum markers will be limited in their correlations with symptoms produced by the highly specialized anatomy of brain circuits. Moreover, the fact that most single molecules have very

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

modest impacts in the risk for psychiatric illness has been made clear by genomic studies (Kendler 2013), and is likely to hold true at the peripheral marker level. Thus, it is possible that serum markers in general will tend to be transdiagnostic, and that most molecules might prove to be no more specific than the ones we focused on. This could feasibly be improved by 'omics' approaches evaluating multiple markers at a time; however, at least in our sample, these articles were relatively uncommon, and although some studies have suggested that combinations of markers might hold greater diagnostic specificity (Domenici *et al.* 2010; Schwarz *et al.* 2012; Frye *et al.* 2015), such findings await independent replication.

Even if most peripheral markers prove to be nonspecific, however, biomarker research can still hold its value. Alterations in the markers we have studied seem to be generally reproducible across psychiatric disorders, suggesting that the consequences of chronic stress might represent a common pathway in the evolution of mental illness. Thus, research in this field might shed light on the connections between psychiatric and medical illness, yield insights on pathophysiology, and perhaps provide ways to assess risk, disease progression or severity, especially if various markers are used in concert (Kapczinski *et al.* 2009, 2010). It can also increase our knowledge of the consequences of chronic stress on the brain and body, and help in the creation of transdiagnostic approaches to bridge the gaps between psychiatric research and neuroscience, or between psychiatry and other fields of medicine (Cuthbert & Insel 2013). For this to happen, however, the field needs to taper down its initial optimism, acknowledge that most peripheral markers are not likely to be diagnostic or specific, increase the rigor of its approaches and focus on questions that can drive research from statistical to clinical significance.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

AUTHOR CONTRIBUTIONS

O.B.A. and J.V.P. designed the study. J.V.P. performed literature searches. T.C.M. programmed the text-mining algorithm. J.V.P., T.C.M. and O.B.A. extracted article data and cross-checked data for consistency. T.C.M., O.B.A. and J.V.P. designed figures and tables. O.B.A. wrote the initial version on the manuscript. All authors provided critical input in revisions of the manuscript.

ACKNOWLEDGEMENTS

The authors are deeply thankful to Marcia Kauer-Sant'Anna and Flávio Kapczinski for thoughtful discussions and constructive criticism on previous versions of the manuscript.

FINANCIAL SUPPORT

This work was supported by FAPERJ grant E-26/201.544/2014 to O.B.A. and by a CNPq scholarship to T.C.M.

CONFLICT OF INTEREST

The authors declare no conflict of interest pertinent to the current work.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

REFERENCES

Arana GW (1985). The dexamethasone suppression test for diagnosis and prognosis in psychiatry. *Archives of General Psychiatry* **42**, 1193-1204.

Belvederi Murri M, Prestia D, Mondelli V, Pariante C, Patti S, Olivieri B, Arzani C, Masotti M, Respino M, Antonioli M, Vassallo L, Serafini G, Perna G, Pompili M, & Amore M (2016). The HPA axis in bipolar disorder: Systematic review and meta-analysis. *Psychoneuroendocrinology* **63**, 327–342.

Bosker FJ, Hartman CA, Nolte IM, Prins BP, Terpstra P, Posthuma D, van Veen T, Willemsen G, DeRijk RH, de Geus EJ, Hoogendijk WJ, Sullivan PF, Penninx BW, Boomsma DI, Snieder H, & Nolen WA (2011). Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Molecular Psychiatry* **16**, 516–532.

Carvalho AF, Köhler CA, Fernandes BS, Quevedo J, Miskowiak KW, Brunoni AR, Machado-Vieira R, Maes M, Vieta E, & Berk M (2016). Bias in emerging biomarkers for bipolar disorder. *Psychological Medicine* **46**, 2287-2297.

Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, Meier MH, Ramrakha S, Shalev I, Poulton R, & Moffitt TE (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science* **2**, 119–137.

Chouinard G (2006). The search for new off-label indications for antidepressant, antianxiety, antipsychotic and anticonvulsant drugs. *Journal of Psychiatry & Neuroscience* **31**, 168–176.

Collins AL, Sullivan PF, Corvin A, Craddock N, Sullivan P, Sullivan P, Geus E de, Willemsen G, James M, Smit J, Zandbelt T, Sullivan P, Daly M, O'Donovan M, Collins A, Kim Y, Sklar P, O'Donovan M, Sullivan P, Visscher P, Brown M, McCarthy M, Yang J, & Sullivan P (2013). Genome-wide association studies in psychiatry: what have we learned? *British Journal of Psychiatry* **202**, 1–4.

Cuthbert BN, & Insel TR (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine* **11**, 126.

Danesh J, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GDO, Pepys MB, & Gudnason V (2004). C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *New England*

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Journal of Medicine **350**, 1387–1397.

Dargél AA, Godin O, Kapczinski F, Kupfer DJ, & Leboyer M (2015). C-reactive protein alterations in bipolar disorder: A meta-analysis. *Journal of Clinical Psychiatry* **76**, 142–150.

Davis J, Maes M, Andreazza a, McGrath JJ, Tye SJ, & Berk M (2014). Towards a classification of biomarkers of neuropsychiatric disease: from encompass to compass. *Molecular Psychiatry* **20**, 152–153.

Dickerson SS, & Kemeny ME (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin* **130**, 355–391.

Domenici E, Willé DR, Tozzi F, Prokopenko I, Miller S, McKeown A, Brittain C, Rujescu D, Giegling I, Turck CW, Holsboer F, Bullmore ET, Middleton L, Merlo-Pich E, Alexander RC, & Muglia P (2010). Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. *PLoS One* **5**, e9166.

Duman RS, & Monteggia LM (2006). A neurotrophic model for stress-related mood disorders. *Biological Psychiatry* **59**, 1116–1127.

Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghersi D, Ioannidis JPA, Simes J, & Williamson PR (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* **3**, e3081.

Fernandes BS, Berk M, Turck CW, Steiner J, & Gonçalves C (2013). Decreased peripheral brain-derived neurotrophic factor levels are a biomarker of disease activity in major psychiatric disorders: a comparative meta-analysis. *Molecular Psychiatry* **19**, 749–751.

Fernandes BS, Molendijk ML, Köhler CA, Soares JC, Leite CMGS, Machado-Vieira R, Ribeiro TL, Silva JC, Sales PMG, Quevedo J, Oertel-Knöchel V, Vieta E, González-Pinto A, Berk M, & Carvalho AF (2015). Peripheral brain-derived neurotrophic factor (BDNF) as a biomarker in bipolar disorder: a meta-analysis of 52 studies. *BMC Medicine* **13**, 289.

Fernandes BS, Steiner J, Berk M, Molendijk ML, Gonzalez-Pinto a, Turck CW, Nardin P, & Gonçalves C (2014). Peripheral brain-derived neurotrophic factor in schizophrenia and the role of antipsychotics: meta-analysis and implications. *Molecular Psychiatry* **20**, 1108–1119.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fernandes BS, Steiner J, Bernstein H-G, Dodd S, Pasco JA, Dean OM, Nardin P, Gonçalves C-A, & Berk M (2016a). C-reactive protein is increased in schizophrenia but is not altered by antipsychotics: meta-analysis and implications. *Molecular Psychiatry* **21**, 554–564.

Fernandes BS, Steiner J, Molendijk ML, Dodd S, Nardin P, Gonçalves C-A, Jacka F, Köhler CA, Karmakar C, Carvalho AF, & Berk M (2016b). C-reactive protein concentrations across the mood spectrum in bipolar disorder: a systematic review and meta-analysis. *Lancet Psychiatry* **3**, 1147–1156.

Frye MA, Nassan M, Jenkins GD, Kung S, Veldic M, Palmer BA, Feeder SE, Tye SJ, Choi DS, & Biernacka JM (2015). Feasibility of investigating differential proteomic expression in depression: implications for biomarker development in mood disorders. *Translational Psychiatry* **5**, e689.

Girshkin L, Matheson SL, Shepherd AM, & Green MJ (2014). Morning cortisol levels in schizophrenia and bipolar disorder: A meta-analysis. *Psychoneuroendocrinology* **49**, 187–206.

Gnanapavan S, Hegen H, Khalil M, Hemmer B, Franciotta D, Hughes S, Hintzen R, Jeromin A, Havrdova E, Tumani H, Bertolotto A, Comabella M, Frederiksen J, Álvarez-Cermeño JC, Villar L, Galimberti D, Myhr K-M, Dujmovic I, Fazekas F, Ionete C, Menge T, Kuhle J, Keir G, Deisenhammer F, Teunissen C, & Giovannoni G (2014). Guidelines for uniform reporting of body fluid biomarker studies in neurologic disorders. *Neurology* **83**, 1210–1216.

Gold PW (2014). The organization of the stress system and its dysregulation in depressive illness. *Molecular Psychiatry* **20**, 32-47.

Goldsmith DR, Rapaport MH, & Miller BJ (2016). A meta-analysis of blood cytokine network alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder and depression. *Molecular Psychiatry* **21**, 1696-1709.

Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, Ortega BN, Zaiko Y V, Roach EL, Korgaonkar MS, Grieve SM, Galatzer-Levy I, Fox PT, & Etkin A (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305–315.

Haag AL (2007). Biomarkers trump behavior in mental illness diagnosis. *Nature Medicine* **13**, 3.

Haapakoski R, Mathieu J, Ebmeier KP, Alenius H, & Kivimäki M (2015). Cumulative meta-analysis of interleukins 6 and 1β, tumour necrosis factor α and C-

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

reactive protein in patients with major depressive disorder. *Brain, Behavior, and Immunity* **49**, 206–215.

Heinrichs RW (2004). Meta-analysis and the science of schizophrenia: variant evidence or evidence of variants? *Neuroscience and Biobehavioral Reviews* **28**, 379–394.

Higginson AD, & Munafò MR (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLOS Biology* **14**, e2000995.

Hyman SE (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology* **6**, 155–179.

Insel TR, & Quirion R (2005). Psychiatry as a clinical neuroscience discipline. *Journal of the American Medical Association* **294**, 2221–2224.

Ioannidis JPA (2005). Why most published research findings are false. *PLoS Medicine* **2**, e124.

Joyner MJ, Paneth N, & Ioannidis JPA (2016). What happens when underperforming big ideas in research become entrenched? *Journal of the American Medical Association* **316**, 1355-1356.

Kapczinski F, Dal-Pizzol F, Teixeira a L, Magalhaes PVS, Kauer-Sant'Anna M, Klamt F, Pasquali M a DB, Quevedo J, Gama CS, & Post R (2010). A systemic toxicity index developed to assess peripheral changes in mood episodes. *Molecular Psychiatry* **15**, 784–786.

Kapczinski F, Dias VV, Kauer-Sant'Anna M, Brietzke E, Vázquez GH, Vieta E, & Berk M (2009). The potential use of biomarkers as an adjunctive tool for staging bipolar disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry* **33**, 1366–71.

Kapczinski F, Vieta E, Andreazza AC, Frey BN, Gomes FA, Tramontina J, Kauer-Sant'Anna M, Grassi-Oliveira R, & Post RM (2008). Allostatic load in bipolar disorder: Implications for pathophysiology and treatment. *Neuroscience and Biobehavioral Reviews* **32**, 675–692.

Kapur S, Phillips a G, & Insel TR (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry* **17**, 1174–1179.

Kauer-Sant'Anna M, Tramontina J, Andreazza AC, Cereser K, da Costa S, Santin A, Yatham LN, & Kapczinski F (2007). Traumatic life events in bipolar disorder: Impact on BDNF levels and psychopathology. *Bipolar Disorders* **9 (Suppl.**

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1), 128–135.

Kendler KS (2006). Reflections on the relationship between psychiatric genetics and psychiatric nosology. *American Journal of Psychiatry* **163**, 1138–1146.

Kendler KS (2013). What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Molecular Psychiatry* **18**, 1058–1066.

Kennedy SH, Downar J, Evans KR, Feilotter H, Lam RW, MacQueen GM, Milev R, Parikh S V., Rotzinger S, & Soares C (2012). The Canadian Biomarker Integration Network in Depression (CAN-BIND): advances in response prediction. *Current Pharmaceutical Design* **18**, 5976–5989.

Kessing LV (2016). The Bipolar Illness Onset Study (the BIO Study) - ClinicalTrials.gov Identifier: NCT02888262. (https://clinicaltrials.gov/ct2/show/NCT02888262). Accessed 21 February 2017.

Kivimaki M, Singh-Manoux A, Ferrie JE, & Batty GD (2013). Post hoc decision-making in observational epidemiology--is there need for better research standards? *International Journal of Epidemiology* **42**, 367–370.

Kupfer D, First MB, & Regier DA (2002). A Research Agenda for DSM-V. Washington, DC: American Psychiatric Association.

Lancet Psychiatry (2016). Blood biomarkers in psychiatry. *Lancet Psychiatry* **3**, 693.

Martín F, Santolaria F, Batista N, Milena A, González-Reimers E, Brito MJ, & Oramas J (1999). Cytokine levels (IL-6 and IFNγ), acute phase response and nutritional status as prognostic factors in lung cancer. *Cytokine* **11**, 80–86.

McEwen BS (2007). Physiology and neurobiology of stress and adaptation: central role of the brain. *Physiological Reviews* **87**, 873–904.

Mendall MA, Patel P, Asante M, Ballam L, Morris J, Strachan DP, Camm AJ, & Northfield TC (1997). Relation of serum cytokine concentrations to cardiovascular risk factors and coronary heart disease. *Heart* **78**, 273–277.

Modabbernia A, Taslimi S, Brietzke E, & Ashrafi M (2013). Cytokine alterations in bipolar disorder: A meta-analysis of 30 studies. *Biological Psychiatry* **74**, 15–25.

Munkholm K, Vinberg M, & Kessing L V (2015). Peripheral blood brain-derived neurotrophic factor in bipolar disorder: a comprehensive systematic review and meta-analysis. *Molecular Psychiatry* **21**, 216–228.

Munkholm K, Vinberg M, & Vedel Kessing L (2013). Cytokines in bipolar disorder: A systematic review and meta-analysis. *Journal of Affective Disorders* **144**,

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

16–27.

Nutt AE (2016). Brain hacking: The mind's biology. The Washington Post. (http://www.washingtonpost.com/sf/national/2016/02/19/brain-hacking-the-minds-biology). Accessed 21 February 2017.

Olfson M, Gerhard T, Huang C, Crystal S, & Stroup TS (2015). Premature mortality among adults with schizophrenia in the United States. *JAMA Psychiatry* **72**, 1172–1181.

Ösby U, Westman J, Hällgren J, & Gissler M (2016). Mortality trends in cardiovascular causes in schizophrenia, bipolar and unipolar mood disorder in Sweden 1987-2010. *European Journal of Public Health* **26**, 867-871.

Pavão R, Tort ABL, & Amaral OB (2015). Multifactoriality in psychiatric disorders: a computational study of schizophrenia. *Schizophrenia Bulletin* **41**, 980–988.

Perlis RH (2011). Translating biomarkers to clinical practice. *Molecular Psychiatry* **16**, 1076–1087.

Polyakova M, Stuke K, Schuemberg K, Mueller K, Schoenknecht P, & Schroeter ML (2015). BDNF as a biomarker for successful treatment of mood disorders: A systematic & quantitative meta-analysis. *Journal of Affective Disorders* **174**, 432–440.

Rich T, Innominato PF, Boerner J, Mormont MC, Iacobelli S, Baron B, Jasmin C, & Lévi F (2005). Elevated serum cytokines correlated with altered behavior, serum cortisol rhythm, and dampened 24-hour rest-activity patterns in patients with metastatic colorectal cancer. *Clinical Cancer Research* **11**, 1757–1764.

Schwarz E, Guest PC, Rahmoune H, Martins-de-Souza D, Niebuhr DW, Weber NS, Cowan DN, Yolken RH, Spain M, Barnes A, & Bahn S (2012). Identification of a blood-based biological signature in subjects with psychiatric disorders prior to clinical manifestation. *World Journal of Biological Psychiatry* **13**, 627–632.

Siontis KCM, Patsopoulos N a, & Ioannidis JP a (2010). Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *European Journal of Human Genetics* **18**, 832–837.

Slavich GM, Way BM, Eisenberger NI, & Taylor SE (2010). Neural sensitivity to social rejection is associated with inflammatory responses to social stress. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 14817–14822.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Smaldino PE, & McElreath R (2016). The natural selection of bad science. *Royal Society Open Science* **3**, 160384.

Smith GD, Ben-Shlomo Y, Beswick A, Yarnell J, Lightman S, & Elwood P (2005). Cortisol, testosterone, and coronary heart disease: prospective evidence from the Caerphilly study. *Circulation* **112**, 332–340.

Smoller JW (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **381**, 1371–1379.

Sprooten E, Rasgon A, Goodman M, Carlin A, Leibu E, Lee WH, & Frangou S (2016). Addressing reverse inference in psychiatric neuroimaging: Meta-analyses of task-related brain activation in common mental disorders. *Human Brain Mapping*, doi:10.1002/hbm.23486.

Steptoe A, Hamer M, & Chida Y (2007). The effects of acute psychological stress on circulating inflammatory factors in humans: a review and meta-analysis. *Brain, Behavior, and Immunity* **21**, 901–912.

Sterne J, & Davey Smith G (2001). Sifting the evidence – what's wrong with significance tests? *British Medical Journal* **322**, 226–231.

Stetler C, & Miller GE (2011). Depression and hypothalamic-pituitary-adrenal activation: a quantitative summary of four decades of research. *Psychosomatic Medicine* **73**, 114–126.

Sullivan PF (2010). The Psychiatric GWAS Consortium: Big science comes to psychiatry. *Neuron* **68**, 182–186.

Wohleb ES, Franklin T, Iwata M, & Duman RS (2016). Integrating neuroimmune systems in the neurobiology of depression. *Nature Reviews Neuroscience* **17**, 497–511.

Table 1.

MDD		BD		Schizophrenia		PTSD		Autism		ADHD	
Molecule	#	Molecule	#	Molecule	#	Molecule	#	Molecule	#	Molecule	#
BDNF	**86**	**BDNF**	**55**	**BDNF**	**48**	**Cortisol**	**19**	Glutathione	20	*Dopamine*	*7*
Cortisol	**61**	**TNF-α**	**21**	*Prolactin*	*31*	**CRP**	**12**	Serotonin	20	Norepinephrine	5
IL-6	**60**	**IL-6**	**18**	**TNF-α**	**22**	**IL-6**	**11**	**BDNF**	**12**	**BDNF**	**4**
CRP	**53**	**Cortisol**	**13**	**IL-6**	**19**	*CRH*	*7*	Homocysteine	12	Adiponectin	4
TNF-α	**46**	**CRP**	**12**	Glutathione	17	**BDNF**	**6**	**TNF-α**	**10**	**Cortisol**	**3**
Fibrinogen	23	S100B	11	*Insulin*	*17*	ACTH	6	**IL-6**	**9**	*Oxyhemoglobin*	*3*
IL-1 beta	21	SOD	10	*Leptin*	*15*	IL-8	5	Cysteine	8	*Serotonin*	*3*
Serotonin	*18*	*Glutamic acid*	*10*	*Glucose*	*14*	Neuropeptide Y	5	Interferon-γ	8	*Cholesterol*	*2*
IL-10	16	IL-10	8	**Cortisol**	**13**	IL-1 beta	4	Oxytocin	8	**IL-6**	**2**
S100B	16	Neurotrophin-3	8	*Dopamine*	13	IL-4	4	Lactic acid	7	**CRP**	**2**
Glutamic acid	*13*	NGF	7	**CRP**	**12**	Interferon-γ	4	Methionine	7	MHPG	2
Prolactin	*12*	*GABA*	*7*	*GABA*	*12*	**TNF-α**	**4**	Melatonin	7	Ferritin	2
Insulin	*11*	IL-1 beta	6	ACTH	12	Annexin a2	4	*Glutamic acid*	7	Homovanillic acid	2
GABA	*11*	IL-2	6	Interferon-γ	11	Serotonin	4	*GABA*	6	Neuropeptide Y	2
Interferon-γ	10	Nitric oxide	6	Homocysteine	11	S100B	4	IL-1 beta	5	DOPAC	2
Kynurenine	10	*Dopamine*	*6*	IL-10	11	IL-2	3	EGF	5	*Iron blood level*	2
Leptin	*10*	*Insulin*	*6*	*Glutamic acid*	*11*	*Leptin*	*3*	*Various*	-	*Various*	-

Table 1. Most frequent molecules among Scopus keywords in peripheral biomarker studies of different psychiatric disorders. Table shows the top-ranked endogenous molecules among Scopus keywords in a database search for "biomarker" AND "serum OR blood OR plasma OR plasmatic" AND individual disorder (see methods). Note that not all molecules included as keywords necessarily represent candidate peripheral biomarkers. The 5 top hits for mood disorders are in bold, and at least 3 of them appear in every disorder. Italic indicates molecules used as markers of treatment effects (e.g. prolactin, insulin) and molecules not used as peripheral markers, which appeared in the text for other reasons (e.g. glutamic acid, serotonin). ACTH: adenocorticotropic hormone; BDNF: brain-derived neurotrophic factor; CRH: corticotrophin releasing hormone; CRP: C-reactive protein; DOPAC: 3, 4 dihydroxyphenylacetic acid; EGF: epidermal growth factor; GABA: gamma-aminobutyric acid: IL-1 beta: interleukin 1 beta; IL-2: interleukin 2; IL-4: interleukin 4; IL-6: interleukin 6; IL-8: interleukin 8; IL-10: interleukin 10; MHPG: 4-hydroxy-3 methoxyphenylethylene glycol; NGF: nerve growth factor; S100B: S100 calcium-binding protein beta; SOD: superoxide dismutase; TNF-α: tumor necrosis factor alpha.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2.

	MDD				BD				Schizophrenia				p	All
	BDNF	**Inflam**	**Cort**	**Total**	**BDNF**	**Inflam**	**Cort**	**Total**	**BDNF**	**Inflam**	**Cort**	**Total**		
Included/total articles	52/149	82/217	56/152	**182/518**	25/65	27/63	10/34	**56/162**	19/68	32/84	14/34	**60/186**	-	**280/866**
Group comparison														
vs. control group	31 (60%)	60 (73%)	31 (55%)	**117 (64%)**	19 (76%)	19 (70%)	5 (50%)	**38 (68%)**	15 (79%)	23 (72%)	6 (43%)	**42 (70%)**	.687	**184 (66%)**
between disease states	2 (4%)	4 (5%)	0 (0%)	**6 (3%)[2]**	8 (32%)	7 (26%)	3 (30%)	**17 (30%)[1,3]**	2 (11%)	1 (3%)	0 (0%)	**3 (5%)[2]**	$<10^{-4}$	**26 (9%)**
vs. other disorder	9 (17%)	7 (9%)	7 (13%)	**21 (12%)[2]**	7 (28%)	6 (22%)	4 (40%)	**15 (27%)[1]**	3 (16%)	5 (16%)	2 (14%)	**8 (13%)**	.018	**28 (10%)**
before vs. after treatment	24 (46%)	28 (34%)	7 (13%)	**56 (31%)**	7 (28%)	2 (7%)	1 (10%)	**10 (18%)**	2 (11%)	7 (22%)	4 (29%)	**12 (20%)**	.075	**77 (28%)**
Correlation														
w/ current symptoms	24 (46%)	31 (38%)	15 (27%)	**66 (36%)**	12 (48%)	5 (19%)	1 (20%)	**17 (30%)**	10 (53%)	12 (37%)	5 (36%)	**24 (40%)**	.550	**103 (37%)**
w/ progression	5 (10%)	5 (6%)	2 (4%)	**12 (7%)**	3 (12%)	5 (19%)	2 (20%)	**9 (16%)**	2 (11%)	6 (19%)	0 (0%)	**8 (13%)**	.064	**29 (10%)**
w/ treatment response	14 (27%)	12 (15%)	5 (9%)	**29 (16%)**	3 (12%)	3 (11%)	0 (0%)	**6 (11%)**	1 (5%)	4 (12%)	3 (21%)	**8 (13%)**	.601	**42 (15%)**
Other														
Risk prediction	3 (6%)	2 (2%)	4 (7%)	**9 (5%)**	1 (4%)	1 (4%)	1 (10%)	**3 (5%)**	0 (0%)	1 (3%)	0 (0%)	**1 (2%)**	.515	**12 (4%)**
Study features														
Longitudinal	27 (51%)	29 (35%)	14 (25%)	**67 (37%)**	10 (40%)	6 (21%)	4 (40%)	**20 (6%)**	2 (11%)	8 (25%)	5 (36%)	**14 (23%)**	.152	**98 (35%)**
Diagnostic efficacy	8 (15%)	4 (5%)	15 (27%)	**23 (13%)**	3 (12%)	1 (4%)	2 (20%)	**5 (9%)**	2 (11%)	4 (12%)	1 (7%)	**5 (8%)**	.557	**29 (10%)**

Table 2. Experimental design features of retrieved articles for each biomarker/disorder combination. Articles measuring more than one marker for a given disorder or the same marker for more than one disorder were counted only once to calculate totals. BDNF, brain-derived neurotrophic factor; MDD, major depressive disorder; BD: bipolar disorder; SCZ: schizophrenia; p values refer to a χ^2 test comparing aggregate values for all markers between the 3 disorders [1]$p<0.05$ vs. MDD, [2]$p<0.05$ vs. BD, [3] $p<0.05$ vs. SCZ, Fisher's exact test comparing pairs of disorders. A similar table with aggregated totals for individual markers instead of disorders is presented as Supplementary Table 2.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 3.

BDNF

Disorder	State	Levels	Sample size	# studies	Effect size	95% CI	*p*	I^2	Egger's *p*	# missing studies	Corrected effect size	Reference
MDD	Euthymia	=	872	11	-0.14	-0.38 to 0.07	0.235	67.9%	---	0	---	Polyakova *et al.* 2015
	Depression	↓	2447	38	-0.80	-1.05 to -0.54	1.1×10^{-9}	91.2%	---	9	-0.48 [-0.72, -0.22]	Polyakova *et al.* 2015
BD	Overall	↓	3538	34	-0.28	-0.51 to -0.04	0.022	92.3%	0.08	6	-0.02 [-0.27, 0.24]	Munkholm *et al.* 2016
	Mania	↓	1397	19	-0.57	-0.99 to -0.14	0.01	92.1%	0.038	0	---	Fernandes *et al.* 2015
	Depression	↓	1074	15	-0.93	-1.37 to -0.50	0.001	87.9%	0.002	0	---	Fernandes *et al.* 2015
	Euthymia	=	3057	24	0.05	-013 to 0.24	0.569	81.2%	0.632	0	---	Fernandes *et al.* 2015
	Mixed	=	213	3	0.09	-0.57 to 0.7	0.787	69.1%	---	---	---	Fernandes *et al.* 2015
SCZ	Overall	↓	5247	39	-0.71	-0.96 to -0.46	0.0001	93.6%	0.26	0	---	Fernandes *et al.* 2014
	1st episode	↓	1560	17	-0.87	-1.23 to -0.51	0.0001	89.6%	---	---	---	Fernandes *et al.* 2014
	Drug naïve	↓	1881	21	-0.84	-1.17 to -0.51	0.0001	90.1%	---	---	---	Fernandes *et al.* 2014
	Chronic	↓	3687	22	-0.56	-0.90 to -0.20	0.0002	95.2%	---	---	---	Fernandes *et al.* 2014

IL-6

Disorder	State	Levels	Sample size	# studies	Effect size	95% CI	*p*	I^2	Egger's *p*	# missing studies	Corrected effect size	Reference
MDD	Overall	↑	2022	31	0.54	0.40 to 0.69	$<1 \times 10^{-4}$	56.2%	0.3	---	---	Haapakoski *et al.* 2015
	Acute	↑	522	10	0.76	0.56 to 0.95	<0.01	90.6%	0.118	---	---	Goldsmith *et al.* 2016
	Chronic	↑	416	7	0.39	0.20 to 0.59	<0.01	77.7%	0.037	---	---	Goldsmith *et al.* 2016
BD	Overall	=	1275	13	0.238	-0.02 to 0.50	0.073*	87.7%	0.312	---	---	Modabbernia *et al.* 2013
	Mania	↑	1024	2	0.59	0.25 to 0.93	<0.01	71.4%	---	---	---	Goldsmith *et al.* 2016
	Depression	=	446	3	0.00	-0.23 to 0.23	0.98	71.2%	---	---	---	Goldsmith *et al.* 2016
	Euthymia	↑	753	7	0.30	0.13 to 0.47	<0.01	78.2%	0.398	---	---	Goldsmith *et al.* 2016
SCZ	1st episode	↑	1083	11	1.16	1.03 to 1.30	<0.01	92.6%	0.624	---	---	Goldsmith *et al.* 2016
	Acute	↑	746	9	0.73	0.56 to 0.90	<0.01	94.0%	0.752	---	---	Goldsmith *et al.* 2016
	Chronic	↑	1385	12	0.27	0.17 to 0.38	<0.01	88.2%	0.069	---	---	Goldsmith *et al.* 2016

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

TNF-α

Disorder	State	Levels	Sample size	# studies	Effect size	95% CI	*p*	I^2	Egger's *p*	# missing studies	Corrected effect size	Reference
MDD	Overall	↑	2476	31	0.40	0.15 to 0.65	<0.05	88.8%	0.12	---	---	Haapakoski *et al.* 2015
	Acute	↑	577	8	0.35	0.17 to 0.53	<0.01	94.1%	0.16	---	---	Goldsmith *et al.* 2016
	Chronic	=	697	7	0.05	-0.10 to 0.19	0.52	86..7%	0.037	---	---	Goldsmith *et al.* 2016
BD	Overall	↑	994	14	0.599	0.15 to 1.05	0.1*	91.6%	0.11	---	---	Modabbernia *et al.* 2013
	Mania	↑	248	4	3.68	1.35 to 6.01	0.002	97.0%	---	---	---	Munkholm *et al.* 2013
	Depression	=	149	2	-0.16	-0.53 to 0.21	0.40	48.3%	---	---	---	Goldsmith *et al.* 2016
	Euthymia	=	532	7	0.04	-0.14 to 0.22	0.66	51.1%	0.398	---	---	Goldsmith *et al.* 2016
SCZ	1st episode	↑	1429	9	0.31	0.22 to 0.39	<0.01	96.6%	0.049	---	---	Goldsmith *et al.* 2016
	Acute	↑	718	7	0.22	0.05 to 0.39	0.01	95.7%	0.993	---	---	Goldsmith *et al.* 2016
	Chronic	↑	1017	9	0.30	0.18 to 0.43	<0.01	80.8%	0.069	---	---	Goldsmith *et al.* 2016

CRP

Disorder	State	Levels	Sample size	# studies	Effect size	95% CI	*p*	I^2	Egger's *p*	# missing studies	Corrected effect size	Reference
MDD	Overall	↑	1425	14	0.47	0.28 to 0.65	$<1 \times 10^{-4}$	62.0%	0.47	---	---	Haapakoski *et al.* 2015
BD	Overall	↑	1618	11	0.39	0.24 to 0.55	$<1 \times 10^{-4}$	47.6%	0.006	---	0.28 [0.11, 0.44]	Dargél *et al.* 2015
	Mania	↑	1669	14	0.87	0.58 to 1.15	$<1 \times 10^{-4}$	79.8%	NS	0	---	Fernandes *et al.* 2016b
	Depression	↑	1363	11	0.67	0.23 to 1.11	0.003	90.9%	NS	0	---	Fernandes *et al.* 2016b
	Euthymia	↑	80637	17	0.65	0.40 to 0.90	$<1 \times 10^{-4}$	85.2%	NS	0	---	Fernandes *et al.* 2016b
SCZ	Overall	↑	82962	24	0.6	0.43 to 0.88	0.001	91.1%	0.11	---	---	Fernandes *et al.* 2016a
	1st episode	↑	708	6	0.63	0.03 to 1.22	0.038	86.9%	---	---	---	Fernandes *et al.* 2016a
	Chronic	↑	82254	18	0.76	0.43 to 1.08	$<1 \times 10^{-4}$	94.9%	---	---	---	Fernandes *et al.* 2016a

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

							Cortisol					
Disorder	**State**	**Levels**	**Sample size**	**# studies**	**Effect size**	**95% CI**	***p***	***I²***	**Egger's *p***	**# missing studies**	**Corrected effect size**	**Reference**
MDD	Overall	↑	18374	354	0.60	0.54 to 0.66	<0.001	---	---	---	---	Stetler and Miller 2011
BD	Overall	↑	2933	37	0.34	0.25 to 0.43	$<1 \times 10^{-4}$	47.8%	0.74	---	0.28 [0.11, 0.44]	Belvederi-Murri *et al*. 2016
	Mania	↑	---	12	0.66	0.42 to 0.89	<0.05	68.0%	---	---	---	Belvederi-Murri *et al*. 2016
	Depression	=	---	6	0.09	-0.15 to 0.34	NS	0.0%	---	---	---	Belvederi-Murri *et al*. 2016
	Euthymia	↑	---	9	0.41	0.16 to 0.65	<0.05	0.0%	---	---	---	Belvederi-Murri *et al*. 2016
SCZ	Overall	↑	2613	44	0.387	0.15 to 0.62	0.001	86.9%	0.753	---	---	Girshkin *et al*. 2014
	1st episode	↑	421	10	-0.096	-0.50 to 0.31	0.001	83.2%	---	---	---	Girshkin *et al*. 2014
	Chronic	↑	1672	33	0.355	0.19 to 0.52	<0.001	63.5%	---	---	---	Girshkin *et al*. 2014

Table 3. **Meta-analytic estimates of variation in serum levels of different markers across disorders.** The table describes the variation in serum/plasma levels of each marker for patient vs. control comparisons in each disorder, as well as in different states of the disorder when available. For each comparison, we used the available meta-analysis with the largest sample size in our search (see Methods). Columns show the direction of variation, total sample size and number of studies included in the meta-analysis, effect size with 95% confidence interval, *p* value for the comparison, heterogeneity measured by I^2, *p* value for publication bias analyses (usually performed by Egger's test), number of missing studies estimated through funnel plot/trim-and-fill analysis, corrected effect size after the imputation of missing studies and reference. State comparisons were defined according to the original meta-analysis. Effect sizes are given as provided in the original meta-analyses (Cohen's *d*, Hedges' *g* or standardised mean deviation). In cortisol meta-analyses, the overall comparison for depression used serum, salivary and urinary cortisol measured either at basal levels or after pharmacological challenges; the overall comparison for bipolar disorder used basal serum or salivary levels measured at any time; state comparisons in bipolar disorder and all comparisons in schizophrenia used morning serum or salivary cortisol levels only. --- stands for absence of information in the included article. * indicates p values corrected for multiple comparisons in the original meta-analysis. MDD; major depressive disorder; BD, bipolar disorder; SCZ, Schizophrenia; BDNF, brain-derived neurotrophic factor; IL-6, interleukin 6; TNF-alpha, tumor necrosis factor alpha.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1.

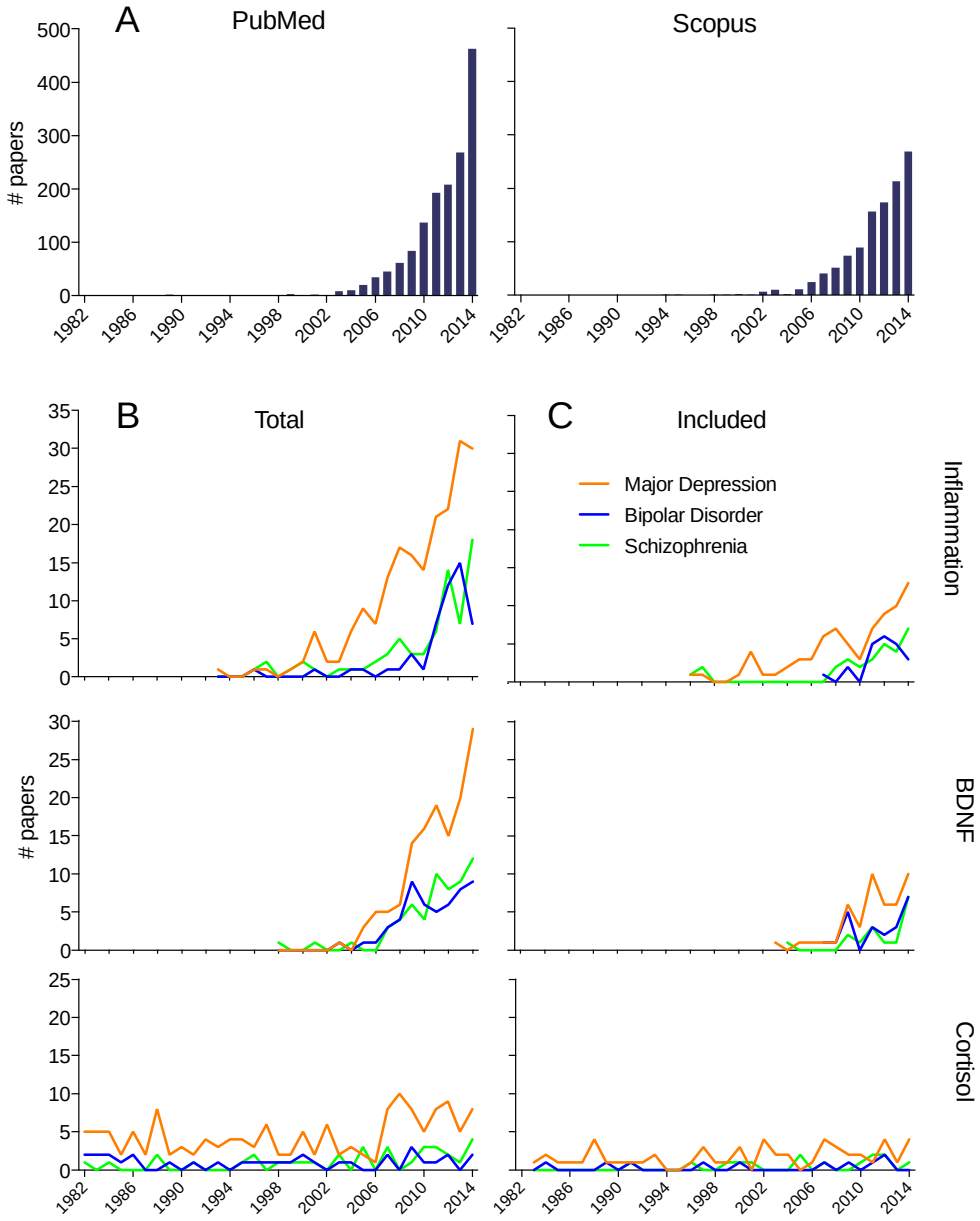

Fig. 1. Growth of biomarker research in psychiatry (A) Number of PubMed (left) and Scopus (right) hits in searches for "biomarker" AND "psychiatry" or "psychiatric" for each year between 1982 and 2014 **(B)** Total number of articles over the same period retrieved for searches in PubMed and Scopus for "biomarker" AND either (a) "tumor necrosis factor alpha OR interleukin-6 OR C-reactive protein" (top) or (b) "BDNF" (middle) or (c) "hydrocortisone" (bottom) AND either (a) "major depression OR depressive disorder" (orange) (b) "bipolar disorder" (blue) or (c) "schizophrenia" (using Pubmed MeSH terms); **(C)** original articles fulfilling criteria for inclusion in our systematic review of experimental design features for each year.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2.

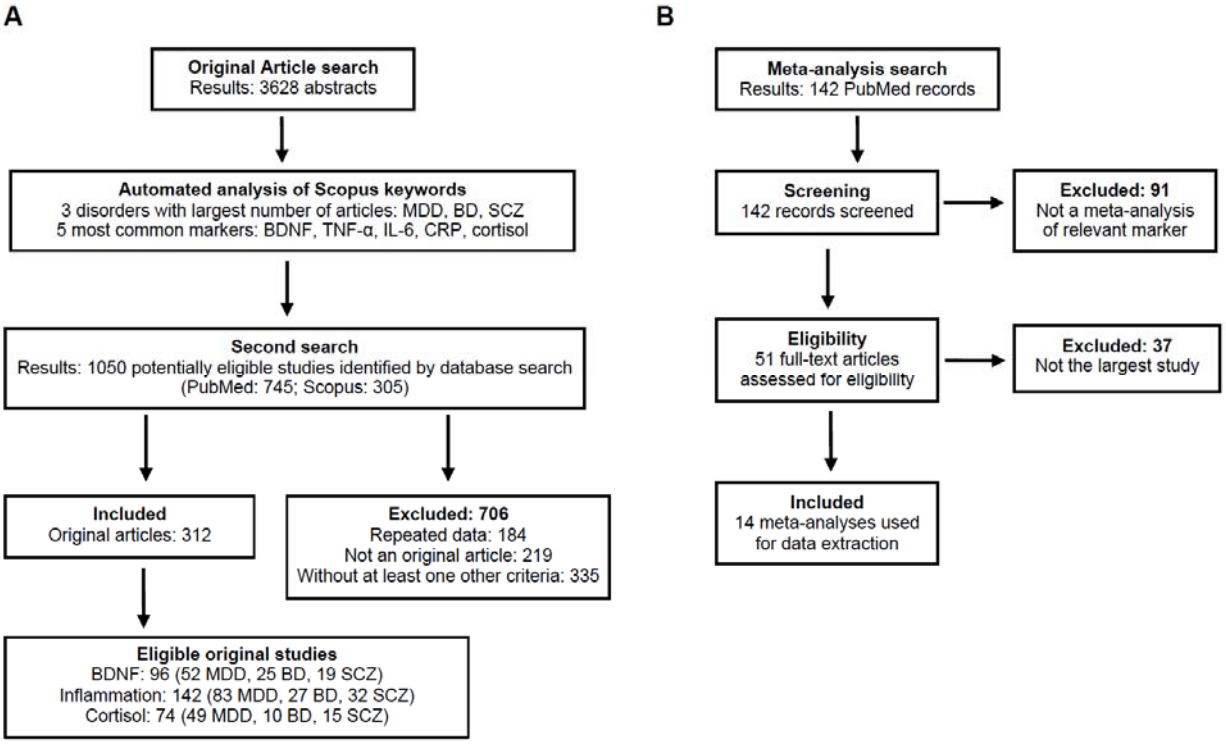

Fig. 2. Flowchart depicting the selection of studies. Flowchart detailing the various stages of study selection in the studies. **(A)** Search strategy for original articles. Eligible studies for each biomarker include articles appearing in more than one search – thus, the sum of articles for individual disorders is greater than the total number of articles for each marker. **(B)** Search strategy for meta-analyses. A more detailed account of the search procedures can be found in the methods section and in the Supplementary Appendix. MDD, major depressive disorder; BD, bipolar disorder; SCZ, schizophrenia; BDNF, brain-derived neurotrophic factor; TNF-α, tumor necrosis factor alpha; IL-6, interleukin-6, CRP, C-reactive protein.

bioRxiv preprint doi: https://doi.org/10.1101/086124; this version posted February 25, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3.

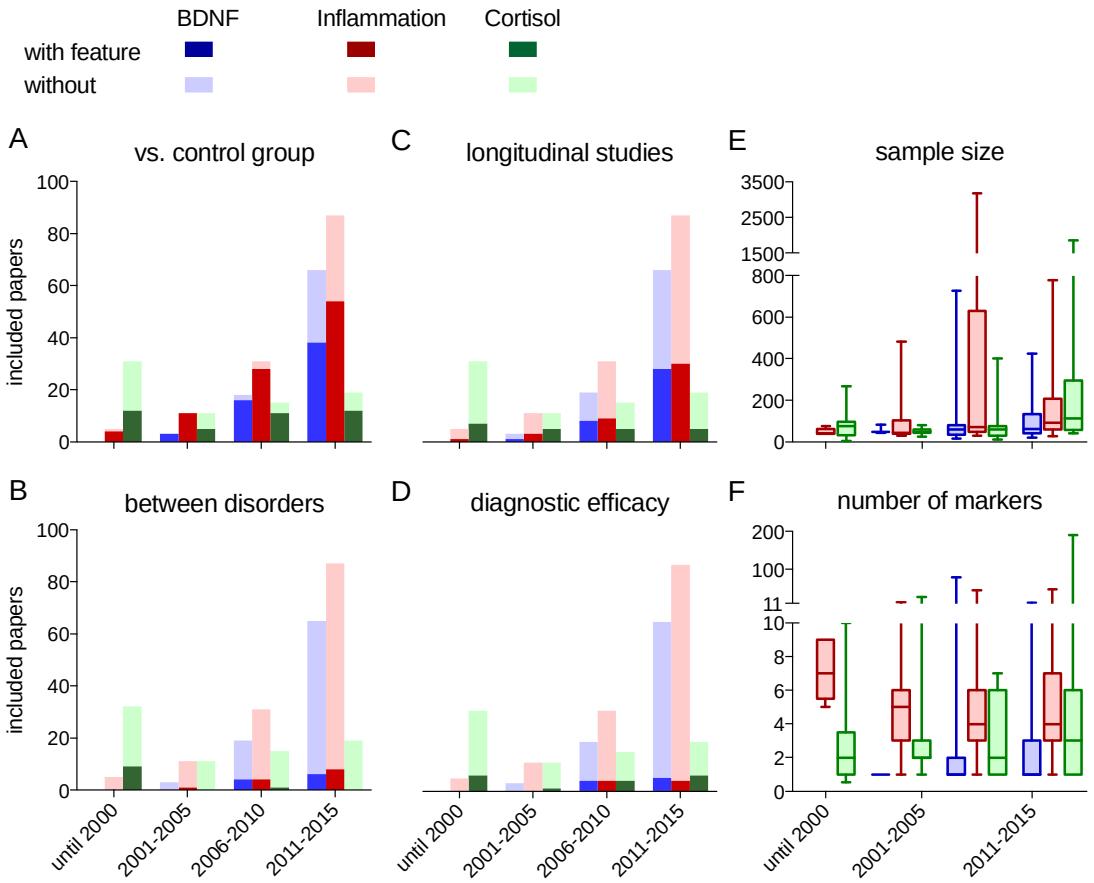

Fig. 3. Variations of experimental design features of articles over different periods. Figure shows the frequency of **(A)** patient vs. control comparisons, **(B)** comparisons between disorders, **(C)** comparisons including a measure of diagnostic efficacy and **(D)** longitudinal studies, as well as the distribution of **(E)** sample size and **(F)** number of peripheral biomarkers studied among analyzed articles measuring BDNF (blue), TNF-α/IL-6/CRP (red) and cortisol (green) over 4 epochs (until 2000, 2001-2005, 2006-2010, 2011-2015). For dichotomous variables (A-D), bars represent the total number of articles, with dark shading representing articles with each experimental design feature and light shading indicating those without them. Distributions of quantitative variables (E-F) are expressed as box-whisker plots (center line, median; box, interquartile range, whiskers, $5^{th}/95^{th}$ percentiles). Spearman's ρ coefficients for correlations between year of publication and each feature are (A) ρ=-0.01, p=0.80, (B) ρ=-0.13, p=0.025, (C) ρ=-0.04, p=0.47, (D) ρ=0.08, p=0.20, (E) ρ=0.21, $p=4x10^{-4}$, (F) ρ=-0.003, p=0.95. Variations of additional features can be visualized in Supplementary Fig. 1.