**Rapid Publication**

**PhenoChipping of Psychotic Disorders: A Novel Approach for Deconstructing and Quantitating Psychiatric Phenotypes**

Alexander B. Niculescu,1* Len L. Lulow,2 Corey A. Ogden,2 Helen Le-Niculescu,1 Daniel R. Salomon,3 Nicholas J. Schork,2 Michael P. Caligiuri,2 and James B. Lohr2++

1Laboratory of Neurophenomics, Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, Indiana
2Department of Psychiatry, UC San Diego, La Jolla, California
3Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California

Psychiatric phenotypes as currently defined are primarily the result of clinical consensus criteria rather than empirical research. We propose, and present initial proof of principle for, a novel approach to characterizing psychiatric phenotypes. We have termed our approach PhenoChipping, by analogy with, and borrowing paradigms and tools from, gene expression microarray studies (GeneChipping). A massive parallel profiling of cognitive and affective state is done with a PhenoChip composed of a battery of existing and new quantitative psychiatric rating scales, as well as hand neuromotor measures. We present preliminary data from 104 subjects, 72 with psychotic disorders (bipolar disorder—41, schizophrenia—17, schizoaffective disorder—14), and 32 normal controls. Microarray data analysis software and visualization tools were used to investigate: 1. relationships between phenotypic items ("phenes"), including with objective motor measures, and 2. relationships between subjects. Our analyses revealed phenotypic overlap among, as well as phenotypic heterogeneity within, the three major psychotic disorders studied. This approach may be useful in helping us move beyond current diagnostic classifications, and suggests a combinatorial building-block (Lego-like) structure underlies psychiatric syndromes. The adaptation of microarray informatic tools for phenotypic analysis readily facilitates direct integration with gene expression profiling of lymphocytes in the same individuals, a strategy for molecular biomarker identification. Empirically derived clusterings of (endo)phenotypes and of patients will better serve genetic, pharmacological, and imaging research, as well as clinical practice. © 2006 Wiley-Liss, Inc.

KEY WORDS: psychiatric profiling; endophenotypes; bipolar; schizophrenia; microarrays; biomarkers


**INTRODUCTION**

Psychiatric phenotypes are currently characterized by consensus criteria derived primarily from clinical experience, as embodied in DSM-IVTR [American Psychiatric Association, Task Force on DSM-IV, 2000] and ICD-10 [Janca et al., 1996]. While this is an advance over the fairly recent past, when categorization relied on somewhat vague clinical descriptions, there is significant room for improvement. The current criteria have, as a strength, provided a common language for psychiatrists across different sites and different countries, with good inter-rater reproducibility [Spitzer and Williams, 1994; Spitzer, 2001; Spitzer and First, 2005]. The major weaknesses are that they are categorical rather than dimensional, and not empirically derived on a consistent basis. As such, they may not entirely and accurately reflect the phenomenological reality, or have a direct correspondence with the underlying biology. There is a need for more quantitative, empirical approaches to psychiatric phenotyping, for both research and clinical purposes. The broad nature of current psychiatric phenotypic constructs is a rate-limiting step for precise and reproducible genetic research, clinical trials, and clinical practice.

The clinical overlap of phenotypes associated with major psychotic disorders such as bipolar disorder, schizoaffective disorder, and schizophrenia, on the one hand, along with the complexity of these psychiatric disorders on the other hand, points to overlapping (shared) mechanisms between disease classes, as well as heterogeneous mechanisms within a disease class. Besides overlap in clinical symptomatology and genetic studies [Berrettini, 2000; Craddock et al., 2006], another body of evidence that supports the existence of shared mechanisms is that various pharmacological treatments are often successful in relieving symptoms across disorders. Conversely, certain pharmacological treatments may only be efficacious for a subgroup of people within a disease class, consistent with the existence of heterogeneity within these disorders [Harrison and Weinberger, 2005; Tamminga and Holcomb, 2005]. Kraepelin and Bleuler, the forefathers of modern psychiatry, had anticipated some of the issues we are facing in terms of overlap among, and heterogeneity within, psychotic disorders.
Nevertheless, it is expected that the next version of the Diagnostic and Statistical Manual for Mental Disorders (DSM), DSM-V, will be an incremental improvement along current categorizations and criteria [Kupfer, 2005], rather than a complete overhaul to include more insights derived from genetics, imaging, and neurophysiological studies. Concerted attempts to be more empirical about the phenomenology of psychotic disorders, particularly bipolar disorders, have been undertaken in the past [Cradock et al., 2004], albeit not yet as part of a concerted integration with genetic work. Personality and temperament measures developed by Cloninger and Akiskal have pointed to dimensional aspects of psychopathology and the existence of a continuum between normality and psychopathology [Maremmani et al., 2005]. There is also the growing realization that endophenotypes [Gottesman and Gould, 2003] are shared in a modular fashion among various psychiatric disorders [Niculescu and Akiskal, 2001; Lenox et al., 2002; Hyman and Fenton, 2003], and an increasing appreciation of the need for approaches that would better define the phenotypic structure of psychiatric disorders [Kelsoe and Niculescu, 2002; Hyman, 2003; Hasler et al., 2004; Krishnan, 2005].

Classifying psychiatric phenotypes based on empirical data analysis may help clarify and quantify the issues of overlap and heterogeneity, and thus place the field on a more biologically relevant footing. Admittedly, this is a large undertaking, with multiple caveats. Nevertheless, if new subtypes can be reliably identified from empirical data analysis of patients profiled on a variety of phenotypic and genetic measures, that may be a starting point towards unraveling their different neurobiological etiologies.

Gene expression profiling with microarrays (GeneChipping) is an empirical, discovery-based approach that has generated new insights in multiple fields, as well as new methodological paradigms. A microarray generally consists of thousands of nucleic acid probes attached to a glass slide. Labeled messenger RNAs, the product of gene transcription (gene expression) from a tissue that is being interrogated, are hybridized with the microarray, and the type and numbers of transcripts that stick to the chip are quantified using a specialized scanner. The readout from the scanner gives a quantitative profile of gene expression in the tissue sample analyzed. A widely used pattern recognition method in microarray analysis is unsupervised hierarchical clustering, in which the similarity between genes determined by expression profiles across multiple conditions is measured. This approach has led to notable successes in cancer biology in terms of improved classification of tumor types, subtypes, and staging, compared to classic histopathological methodologies [Bittner et al., 2000; Lu et al., 2005]. Early attempts at using hierarchical clustering in psychiatry [Mezich, 1978] have not been consistently pursued over the ensuing years.

We reasoned that a similar conceptualization and approach to the one used in cancer biology could be used in psychiatry for empirically studying phenotypes in a massively parallel, quantitative, fashion (“phene” expression). Continuing the analogy with cancer, phene expression may provide advantages compared to classical psychopathologic approaches, similar to those gene expression has provided for tumor classifications compared to classical histopathologic approaches. We have termed the approach PhenoChipping, by analogy to GeneChipping.

It is of interest to have as part of a comprehensive phenotypic profiling (PhenoChipping) approach both subjective measures (quantitative answers to questions about mood, anxiety, cognition) and objective measures (neurophysiology, imaging, gene expression, biochemical assays). New correlations and biomarkers may be revealed by data mining of integrated datasets. Objective phenotypic measurements frequently used include neurophysiology (EEG, neuromotor measures) and brain imaging (fMRI, PET). Hand neuromotor measures, in particular, are easy to administer and deploy, which makes them attractive for large-scale field studies. They engage fundamental fronto-striatal circuits regulating limbic and neuromotor behavior, which may have been recruited also for higher mental functions by evolution. Correlations between motor measures and clinical parameters have been reported in both bipolar disorders and schizophrenia, including in never medicated schizophrenia; moreover, looking at right hand versus left-hand measures may provide a window into brain hemispheric lateralization of pathology [Cortese et al., 2005].

The relationship between cognitive impairment and motor abnormalities remains an important area for further research. Moreover quantitative hand neuromotor measures have been shown, in affective disorders, to be predictive of antidepressant non-response [Caligiuri et al., 2003].

The PhenoChip used in this report consists of a battery of existing psychiatric rating scales (for psychosis, well-being and mood) and one newly developed affective scale, together with right and left hand neuromotor measures, all quantitative in nature. We were particularly interested how responses to questionnaires that reflect an internal subjective experience might correlate with objective neuromotor measures.

Affective abnormalities are an integral part of major psychotic disorders, yet they are often overlooked and not tested for in patients with psychosis, as opposed to patients with mood disorders. We have developed a simple-minded, quantitative, visual analog scale to assess affective state (Total Affective State Scale—TASS), based on combining and placing on a continuum the DSM-IV criteria for depression, mania, and anxiety. This scale has already demonstrated its usefulness in imaging studies of bipolar disorder, as recently reported by us [Caligiuri et al., 2006]. In our current report, we were particularly interested to see if we could detect with it novel differences in individual measures of affect in different psychotic disorders diagnostic categories. From a pragmatic standpoint, we reasoned that we have a higher likelihood of uncovering new phenomenology in an area that has been less explored (mood in psychosis). More generally, we wanted to look at the interdependence of cognition and mood.

With that in mind, we placed on our PhenoChip each of the 11 individual items in TASS, along with the Total Mood subscale, Total Anxiety subscale, and overall Total Affect composite scale (for a total of 14 probes), and placed on it only the composite scales for the 11 other rating instruments (including 4 neuromotor measures, 2 for each hand). Thus, our prototype PhenoChip had 25 probes in total. While being comprehensive, it is purposefully biased towards the finer grained detection of affective phenomenology. Moreover, while we chose to have on the chip the current standard rating scales for psychosis and mood (such as PANSS, HAM-D, YMRS—see below), a variety of other rating scales could have been chosen or added. Thus, we are very cognizant that our results, as with custom microarrays in the gene expression field, may reflect not only the limited number of subjects we sampled, but also the particular choice of probes included on the PhenoChip.

MATERIALS AND METHODS

Demographics and Subject Enrollment

We have collected a sample of 104 subjects, consisting of 41 subjects with bipolar disorder, 17 with schizophrenia, 14 with schizoaffective disorder, and 32 without significant psychiatric illness (normal controls), determined by the Structured Clinical Interview for the DSM-IV Axis I Disorders, Clinician Version (SCID-I).

Subjects consisted of men and women over 18 years of age. A demographic breakdown is shown in Table I. Subjects were
TABLE I. Demographic Data

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>Bipolar</th>
<th>Schizophrenia</th>
<th>Schizoaffective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>32</td>
<td>41</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>Gender</td>
<td>Male: 24</td>
<td>24: 8</td>
<td>14: 3</td>
<td>8: 6</td>
</tr>
<tr>
<td>Age</td>
<td>Mean years (SD)</td>
<td>48.4 (8.5) 32–64</td>
<td>44.3 (10.1) 21–65</td>
<td>47.3 (7.6) 27–59</td>
</tr>
<tr>
<td>Illness duration</td>
<td>Mean years (SD)</td>
<td>18.3 (11.1) 1–47</td>
<td>22.8 (11.3) 3–39</td>
<td>14.9 (8.7) 5–34</td>
</tr>
</tbody>
</table>

recruited from the general population, the patient population at the Veterans Affairs San Diego Healthcare System, and the University of California at San Diego, as well as various facilities that serve people with mental illnesses in San Diego County. The subjects were recruited largely through referrals from care providers, through the use of brochures left in plain sight in public places and mental health clinics, and through word of mouth. Subjects were excluded if they had significant medical, neurological illness or had evidence of active substance abuse or dependence. All subjects understood and signed informed consent forms before assessments began.

### Administration of the PhenoChip

Subjects completed diagnostic assessments (SCID), and then were PhenoChipped. The PhenoChip used consisted of a battery of: (1) existing psychiatric rating scales: Positive and Negative Symptoms Scale (PANSS) (with a positive symptom subscale—PANSSPOS, a negative symptom subscale—PANSSNEG, and a disorganization subscale—PANSSGEN) [Kay et al., 1987], Hamilton Rating Scale for Depression (HAM-D 17 and HAM-D 28) [Hamilton, 1960, 1980], Young Mania Rating Scale (YMRS) [Young et al., 1978], Medical Outcomes Study Short Form-36 (SF-36) [Ware et al., 1996]; (2) a new visual-analog scale: TASS [Caligiuri et al., 2006], as well as (3) hand neuromotor measures: VS-velocity scaling, FI-force instability [Caligiuri et al., 1998].

The battery was administered in one of three predetermined counterbalanced orders. Subjects were paid for their participation. Testers were not blind to the subject’s diagnosis, but were not aware of the study hypotheses or the approach that would be used for empirical data analysis.

### Visual Analog Scale—Total Affective State Subscale (TASS)

The newly developed visual analog scale, the TASS, quantifies mood and anxiety symptoms at the time of administration (Fig. 1). It has a mood subscale and an anxiety subscale. The seven-item mood subscale (Simplified Mood State Subscale—SMS) is based on: (a) combining the DSM-IV criteria for depression and mania, and (b) placing the items on a continuum. The four-item anxiety subscale (Simplified Anxiety State Subscale—SAS), quantifies feelings of uncertainty, fear, and anger. The advantages of TASS, and the reasons for using it, are that: (1) it quantifies state, (2) it measures phenotypes on a continuum, from normal to pathology, (3) it is self-rated, which facilitates administration and ease of use.

We have enrolled more bipolar patients than schizophrenia and schizoaffective patients (41 vs. 17 vs. 14) (Table I) specifically to have a larger sample of patients with known affective symptomatology for the purpose of validating our scale. Besides the face validity of using DSM-IV items for its creation, TASS has internal consistency demonstrated by a high degree of correlation between items, as well as external consistency, demonstrated by the high degree of inverse correlation with HAM-D28, a scale measuring depression (Fig. 1c). Moreover, for the purposes of our studies, we are more interested in the scores of the individual items in TASS, in a modular endophenotypic fashion or as probes on the PhenoChip, rather than how TASS fits as a classic diagnostic measurement scale.

### Data Analysis

To determine which phenes had significantly different scores between each disease group and normal controls, a student’s t-test for independent samples was used. This analysis was performed using Statistica (version 6.1). The average value of the raw scores for each phene is used in our t-test calculation. A P-value <0.05 is considered significant (Fig. 1b and Table II). If, however, we apply a conservative Bonferroni correction for multiple testing, as there are 25 probes on our PhenoChip and three diagnostic groups, the threshold for significance would change to 0.00466.

To analyze the relationships between phenes, a standardization of the data is necessary because of the varying dynamic ranges in which the various psychiatric rating scales measures and neurophysiological hand motor functions are quantified. For example, the HAM-D28 has a score range of 0–82, while the TASS has a range of 0–1,100. The Cohen’s d effect size [Cohen, 1988] was used as our method of standardizing scores for the diagnostic groups, in which Cohen’s d effect size = M1 – M2/σ pooled, where M1 is the average score of the disease group for the phene of interest, and M2 is the average score of the control group for that same phene. σ pooled is the standard deviation of all of the scores that went into calculating both M1 and M2.

To keep the calculations consistent, we used a modified Z score (an individual “effect size”) to calculate the scores for individual subjects, in which Z score = X – M2/σ pooled where X is the individual score for the phene of interest and M2, the average score of the control group for that same phene. σ pooled is the standard deviation of all of the scores that went into calculating both M1 and M2.

### Clustering Analysis Using GeneSpring

We have adapted GeneSpring (Agilent Technologies, Palo Alto, CA) the most widely used, commercially available, microarray gene expression analysis software, for the novel use of analyzing and visualizing phenotypic data. We have inputted the scores on phenotypic items numbers in lieu of the usual use of gene expression intensity numbers. All the subsequent analyses were carried out using the same tools as for gene expression datasets, per the manufacturer’s instructions (www.chem.agilent.com). A “genome” (phenome) was created in the program, consisting of the 25 items on the PhenoChip—each item acting as an individual “gene” (phene). The Z scores for each phene in all samples were imported into GeneSpring. No further normalization was applied to the data.
a  Example of TASS item - Thinking Activity

How high is the amount of mental energy and thinking activity going on in your mind right now? Compare to the most slowed down you ever remember your thinking being, and compared to the most alert and fast you ever remember your thinking being.

\[
\begin{array}{llll}
\text{lowest} & \cdots & \text{highest} \\
0 & \cdots & 100 \text{ mm}
\end{array}
\]

b  Results of measurements using TASS

<table>
<thead>
<tr>
<th>Total Affective State Scale (TASS)</th>
<th>Distance in Spectrum (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal Controls (N=32)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Simplified Mood Scale (SMS)</td>
<td></td>
</tr>
<tr>
<td>Mood</td>
<td>69.25</td>
</tr>
<tr>
<td>Motivation to do things (Motivdo)</td>
<td>70.16</td>
</tr>
<tr>
<td>Movement activity (Mvmntctv)</td>
<td>65.19</td>
</tr>
<tr>
<td>Thinking activity (Thnkactv)</td>
<td>69.13</td>
</tr>
<tr>
<td>Self-esteem (Selffestm)</td>
<td>69.34</td>
</tr>
<tr>
<td>Interest in pleasurable activities (Interest)</td>
<td>69.88</td>
</tr>
<tr>
<td>Appetite</td>
<td>70.06</td>
</tr>
<tr>
<td>Total Mood (TotMood)</td>
<td>483</td>
</tr>
<tr>
<td>Simplified Anxiety State Scale (SAS)</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>42.59</td>
</tr>
<tr>
<td>Uncertainty (Uncertnt)</td>
<td>39.78</td>
</tr>
<tr>
<td>Fear</td>
<td>25.25</td>
</tr>
<tr>
<td>Anger</td>
<td>21.47</td>
</tr>
<tr>
<td>Total Anxiety (TotAnxty)</td>
<td>129.1</td>
</tr>
<tr>
<td>Total Affect (Totalffect)</td>
<td>612.1</td>
</tr>
</tbody>
</table>

c  Total Affective State Scale – correlation with HAM-D28

\[
\begin{align*}
\text{Total Mood} & : r = -0.49, p = 0.000 \\
\text{Total Anxiety} & : r = 0.22, p = 0.025 \\
\text{Total Affect} & : r = -0.39, p = 0.000
\end{align*}
\]

Fig. 1. TASS. a: Example of TASS item - Thinking Activity. b: Results of measurements using TASS. c: TASS-correlation with HAM-D28.
### RESULTS

Phenotypes that were significantly different between each disease group and normal controls are shown in Table II, both with effect size data and t-test data. An effect size of greater than 0.50 is considered medium to high, and significant. We identified 22 phenes in bipolar subjects (11 increased and 11 decreased), 16 phenes in schizophrenic subjects (10 increased and 6 decreased), and 13 phenes in schizoaffective subjects (8 increased and 5 decreased) that were significantly changed ($P < 0.05$).

#### Venn Diagram Analysis

Venn diagrams based on the differentially changed phenes in bipolar disorder, schizophrenia, and schizoaffective disorder, compared with controls, are shown in Figure 2. Figure 2a represents the phenes that were significantly increased, and Figure 2b represents the phenes that were significantly decreased. Several of the differentially expressed phenes were shared between the three psychotic disorders. These shared phenes included four that were increased, and Figure 2b represents the phenes that were significantly decreased. Several of the differentially expressed phenes were shared between the three psychotic disorders.

The effect sizes and the independent t-test P-values for each phenotype in a comparison between disease groups and the normal controls are shown. Numbers in bold text represent phenes that are significantly increased compared to normal controls and numbers in italic text represent phenes that are significantly decreased compared to normal controls (Student’s t-test, $P < 0.05$). All values that have a Cohen’s d effect size greater than 0.50 are bold-italic.
TotAffect. They have to do with well being and mood. These results suggest that the three groups of patients, at the time of PhenoChipping, were overall in a more depressed, psychotic, low well being state compared to normal controls. Furthermore, the results suggest that the areas of endophenotypic and neurobiological overlap common to all three psychotic disorders have to do with both cognition and mood.

The Category II phene increased in common in bipolar disorder and schizophrenia is Uncertnt (Uncertainty), and decreased in common in these two disorders is Thnkactv (Thinking Activity). These results suggest that these two groups of patients, at the time of PhenoChipping, were overall in a state characterized by slow thinking, perhaps in part as a paralyzing consequence of high uncertainty. Furthermore, they suggest that an area of endophenotypic and neurobiological specificity for bipolar disorders compared to schizophrenia spectrum disorders has to do with anxiety and irritability. Last time of the PhenoChipping, were experiencing more negative symptoms than normal controls, and that negative symptoms may be a core endophenotypic and neurobiological feature of schizophrenia spectrum disorders, or a medication side-effect of typical antipsychotics, which are used preponderantly in these two groups of psychotic disorders, compared to bipolar disorder.

The Category III phenes increased only in bipolar disorder patients were Fear, Anger, TotAnxty, MRS. They have to do with anxiety, irritability, and activation. The Category III phenes decreased only in bipolar disorder patients were LVS and Movement Activity (Mvmtactv). They have to do with right hemisphere activity, and overall energy to move. These results suggest that the bipolar patients, at the time of the PhenoChipping, were in an irritable, psychomotorly retarded state, having to do preferentially with their right hemisphere. Furthermore, they suggest that an area of endophenotypic and neurobiological specificity for bipolar disorders compared to schizophrenia spectrum disorders has to do with anxiety and irritability. Last
but not least, an objective neuromotor measure, LVS, having to
do with right hemisphere activity, could potentially be used as
a behavioral biomarker for bipolarity and to monitor treatment
response. This is consistent with recent results in the field
using fMRI [Caligiuri et al., 2004].

The Category III phene decreased only in schizophrenia
patients is Right hand Velocity Scaling (RVS). It has to do with
left hemisphere activity. This result suggests that the schizo-
phrenic patients, at the time of the PhenoChipping, were in a
psychomotorly retarded state, having to do preferentially with
their left hemisphere. Furthermore, they suggest that an area
of endophenotypic and neurobiological specificity for schizo-
phrenia, compared to psychotic disorders with a major affective
component, has to do with left hemisphere function. Last but
not least, an objective neuromotor measure, RVS, having to do
with left hemisphere activity, could potentially be used as a
behavioral biomarker for schizophrenia and to monitor treat-
ment response.

**Clustering of Phenes**

Two-way unsupervised hierarchical clustering of the three
diagnostic groups was first applied, based on the average effect
size for all phenes across the three groups. Results are
displayed in a color-coded "heat map" (Fig. 3), where diagnostic
groups are ordered on the horizontal axis and phenes on the
vertical axis on the basis of similarity of their effect sizes. Of
interest, expression patterns are fairly similar across the three
diagnostic groups, with schizophrenia and schizoaffective
more similar to each other than to bipolar disorder.

The phenes grouped into two main clusters: phenes that
increased in expression compared to normal controls (FIL,
PANSSNEG, FIL, PANSSPOS, PANSSGEN, HAM-D17,
HAM-D28, Uncertnt, Fear, TotAnxty, YMRS, Anxiety, Anger)
and phenes that decreased in expression compared to normal
controls (SF-36, Motivo, TotAffect, Interest, Mood, TotMood,
Mvmtactv, Selfestm, Appetite, RVS, Thnkactv, LVS). Notably,
all of the well-being and mood measures, with the exception of
YMRS, were found to be decreased across all three disorders.
However, HAMD, Fear, and Anger were increased. Taken
together, this suggests that at the time of PhenoChipping,
the subjects were overall in a state of irritable dysphoria. The
score on YMRS may be measuring the activation aspect of this
state rather than true (hypo) mania.

Examples of phenes that clustered together most closely
across all three psychotic disorders groups, in our preliminary
Our non-hypothesis driven, discovery-based approach thus uncovers new empirical relationships between phenotypic items, which are of high neurobiological interest. One such result is the relationship between motivation to do things and affective state. Another on is the relationship between self-esteem and appetite, with clinical implications for abnormal weight changes in these and related disorders.

Our approach also uncovers relationships between objective phenes (hand neuromotor measures) and subjective phenes. One such result is the relationship between RVS and Thinking Activity, suggesting a possible left hemispheric dominance of the neurobiological correlate of this measured phenotype. This may have clinical implications for using left-hemisphere stimulation, through methods such as Transcranial Magnetic Stimulation (TMS), for patients with sluggish thinking, as seen in depression or negative symptoms schizophrenia.

### Clustering of Subjects

Next, unsupervised two-way hierarchical clustering was applied to all of the 104 subjects, based on the Z scores for all the phenes across all of the subjects. Results are displayed in Figure 4 as a color-coded heat map, where the subjects are ordered on the horizontal axis and phenes are ordered on the vertical axis on the basis of similarity of their individual effect sizes. Interestingly, the four major diagnostic groups as established by SCID (normal controls, bipolar, schizophrenia, schizoaffective) fail to cluster together in four distinct groups. The fact that subjects from different diagnostic groups are interspersed speaks to the overlap among current diagnostic classifications (including normal controls), as well as to their internal heterogeneity. More profoundly, it points to the fact that the somewhat artificial boundaries between control and
affected subjects may become blurred when dimensional rather than categorical approaches are used, and should provide a rationale and impetus for population QTL studies in psychiatric genetics. The clustering of this set of individual subjects leads to pairs of highly similar subjects (pseudo-twins) from different diagnostic groups that share more characteristics with each other than with subjects in their own diagnostic group. It has not escaped our attention that this methodology may prove useful in pairing subjects for genetic, pharmacological, and imaging studies. Moreover, clinically, it identifies subjects that may respond similarly to treatments, and should be treated psychiatrically in the same way.

**DISCUSSION**

We have developed, and present initial proof of principle for, an empirical approach to characterizing psychiatric phenotypes, termed PhenoChipping. The approach consists of a massive parallel sampling of cognitive and affective state, employing paradigms and analysis tools from the microarray gene expression field. Our preliminary results revealed overlap among, as well as heterogeneity within, the three major psychotic disorders studied: bipolar disorder, schizophrenia, and schizoaffective disorder. Moreover, the use of hand neuromotor measures has provided preliminary evidence supportive of hemispheric lateralization of cognition and mood, as well as leads for objective behavioral biomarker development.

**Limitations and Confounds**

First, our sample, while well characterized and providing some interesting insights, is limited in size. It is likely that a larger sample size would increase the statistical power of our results, and the discriminatory power of the PhenoChipping approach. Second, our preliminary data is based on cross-sectional analysis at a single timepoint, which may reflect predominately state rather than trait factors. Multiple PhenoChipping measurements, at different timepoints, would likely permit addressing state versus trait issues, by looking at how phenes change over time. Third, as all subjects with major psychotic disorders that were enrolled were on a psychopharmacological regimen, our results may reflect, at least in part, a combined influence of medication (side) effects and underlying disease phenomenology. This may be especially true for hand motor measures in patients on antipsychotic medication (for the Velocity Scaling measure), or mood stabilizing medications (for the Force Instability measure). It remains for future work to address the issue of factoring out possible medication effects by an intra-subject analysis of repeated measurements, while the subjects are on the same pharmacology. Moreover, PhenoChipping of first degree relatives who do not have overt clinical abnormalities, is an important area of future research. Fourth, in the studies described in this report, we did not collect blood for genetic and genomic studies. This should be an integral part of future PhenoChipping work, in view of building a blood sample bank with comprehensive phenotypic information attached to it, for studies integrating phenotype with genotype and gene expression. DNA extracted from whole blood can be used for studies of polymorphisms in candidate genes of interest related to bipolar disorder and schizophrenia [Niculescu et al., 2000; Chowdari et al., 2002; Barrett et al., 2003; Hattori et al., 2003; Geller et al., 2004; Kohn et al., 2004; Ogden et al., 2004; Green et al., 2006]. Whole blood (predominantly lymphocyte) RNA can be extracted for microarray gene expression studies. Lymphocyte gene expression profiling has emerged as a particularly interesting area of research in the quest for peripheral molecular biomarkers of mental illness [Vawter et al., 2004; Glatt et al., 2005; Middleton et al., 2005; Segman et al., 2005; Tsuang et al., 2005]. Fresh blood, with phenotypic state information gathered at time of harvesting, may be more informative than immortalized lymphocytes, and avoid some of the caveats of Epstein–Barr virus (EBV) immortalization and cell culture passaging.

**Conclusions and Future Directions**

Our objective was to develop and provide proof of principle for a novel approach, PhenoChipping, as a way to better understand the phenotypic structure of major psychotic disorders. Our preliminary data documents both overlap among, and heterogeneity within, the three major psychotic disorder studied, and suggests a combinatorial building-block (Lego-like) structure underlying these psychiatric syndromes. It is hoped that our approach, boot-strapped and integrated with genetics, genomics, and imaging studies, will help move the field of psychiatry forward, beyond current categorical diagnostic classifications. The integration of phenotypic data with genotype and gene expression data will be particularly facilitated by the importation of normalized data inside the same analysis software, GeneSpring or another similar program. This was one of the key reasons, we adapted its use for phenotypic data analysis. Applying a methodology developed in one discipline (genomics) can facilitate understanding of another discipline (phenomics) [Kelsoe and Niculescu, 2002], especially if one is trying to connect the two together.

An immediate practical application for our integrative strategy would be in pharmacogenomics; a second would be the identification of peripheral behavioral and molecular biomarkers of illness. A better understanding of major psychotic disorders such as bipolar disorder, schizophrenia, and schizoaffective disorder, will lead to more targeted treatments, with improved efficacy and decreased side-effects. This will have an impact on patient health, well-being, quality of life, and independent functioning. Moreover, early diagnosis and intervention may prevent the full-blown development of illness in genetically susceptible individuals. Ultimately, we propose that our work is an initial step in the direction of developing more individualized diagnosis and treatments for psychiatric patients-personalized psychiatry as a component of personalized medicine [Gould and Manji, 2004], where patients' individual profiles, rather than broad categorical diagnostic constructs, are the targets of therapeutic intervention.

**ACKNOWLEDGMENTS**

This work was supported by an APA/AstraZeneca Young Minds in Psychiatry award to ABN, as well as VISN-22 MIRECC funds and a VA Merit Award to JBL. We thank Dr. John I. Nurnberger, Jr. (Indiana University School of Medicine) for useful discussions and comments.

**REFERENCES**


