The *P*-Value Illusion: How to Improve (Psychiatric) Genetic Studies

Alexander B. Niculescu III[1,2]*,[†] and Helen Le-Niculescu[1]

[1]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana

[2]Indianapolis VA Medical Center, Indianapolis, Indiana

There is an emerging appreciation that genome-wide association studies (GWAS) have failed to live up to expectations and deliver major advances to date. A "surge" strategy, of pooling resources and increasing number of subjects tested, is underway. We argue that, while useful, it will not be enough by itself. Complementary approaches are needed to mine these large datasets. We describe a series of problems, opportunities, and offer a potential comprehensive solution.   © 2010 Wiley-Liss, Inc.

How to Cite this Article:
Niculescu AB, Le-Niculescu H. 2010.
The *P*-Value Illusion: How to Improve (Psychiatric) Genetic Studies.

Am J Med Genet Part B 9999:1–3.

COMPLEXITY

The complexity of most medical and psychiatric disorders as currently defined is such that a large repertoire of genes is likely involved [Le-Niculescu, 2009a; Manolio, 2009]. One way to make things more tractable is to use more narrow phenotypes (sub-phenotypes) for genetic and biomarker studies. Additionally, using a case–case design, comparing different sub-phenotypes or extremes in quantitative sub-phenotypes, may reduce noise and increase power [Kurian, 2009; Le-Niculescu, 2009b], compared to case–control approaches.

HETEROGENEITY

By and large, a small number of housekeeping genes have been identified to date by genome-wide association studies (GWAS) [Anon, 2007; Baum, 2008], as opposed to the more numerous, interesting and biologically relevant genes implicated by other approaches. It may be that the GWAS are under-powered, and can only detect genes in which there is a high degree of conservation (low degree of heterogeneity) in the population, such as housekeeping genes [She, 2009]. Genes that are involved in more discrete and evolved functions, such as specific tissue-enriched genes (e.g., brain-enriched genes like BDNF [Liu, 2009; Petryshen, 2009]), may have a higher degree of heterogeneity due to the evolutionary pressures of adapting to the environment. A mutation in a highly conserved gene will stand out, have a higher impact, and

be detectable by GWAS. A mutation in a highly heterogeneous gene will not stand out across the diverse repertoire of common variants, and will be harder to detect with GWAS. Such interesting and relevant genes, however, may be identified by gene expression studies and other biological experiments, in human tissue samples and animal models.

CNVs may arise all the time, as a paternal (or maternal) age effect of cumulative damage to germline DNA. These random transgenic effects (over-expression, knock-outs) are inadvertent experiments that nature did for us. They are an additional and previously unsuspected source of diversity and heterogeneity in the population, resulting in illness as well as possibly in supra-normative individuals. A note of caution is that studies looking at CNVs need to be carried out on DNA obtained from non-immortalized cell lines, to prevent cell culture and viral (e.g., Ebstein–Barr virus in lymphoblastoid cell lines) induced artifacts. More generally, the tissue source of DNA is relevant, and generalizing from a single one is dangerous, as there is an emerging appreciation of mozaicism and DNA heterogeneity among neural and non-neural cell types [Coufal, 2009].

AJMB-09-0300.R1(31076)

Next generation sequencing, where the whole genome will be sequenced in increasing number of individuals, will reveal rather than solve the problem of heterogeneity. In lieu of just awaiting the hypothetical development of new bioinformatic approaches to deal with that expected torrent of data, sensible strategies feasible with current knowledge and means should be pursued, on the existing and emerging data.

REPRODUCIBILITY

We would like to suggest that *P*-values at best permit the precise ranking of findings *within* a study. At worst, they are over-interpreted and a source of frustration in terms of reproducibility of findings *across* studies. This "*P*-value illusion" in genetics, that a significant *P*-value in one study should necessarily reproduce in another independent study, is based on an under-appreciation of two factors. The first one is the fit-to-cohort effect of classic statistical analyses of genetic studies, the second one is the above discussed complexity and heterogeneity of most disorders. In essence, you are identifying the genes that are a best fit in the cohort you derive your data from. This guarantees that your top findings will not be at the top of the list in a study carried in a different cohort, since complexity and heterogeneity will ensure that no two cohorts are alike. In addition to population-specific factors, which we think are the primary reason, there may be experimental stochastic factors involved as well. This phenomenon is sometimes describes as the "winner's curse," a strong initial finding not being as strong in subsequent independent cohorts.

A solution is to use a fit-to-disease approach, like Convergent Functional Genomics (CFG) [Niculescu, 2000; Ogden, 2004; Le-Niculescu, 2007, 2009a,b; Rodd, 2007; Kurian, 2009]. Such an approach, in addition to *P*-values, uses multiple independent lines of evidence related to illness, including gene expression studies, as a way of prioritizing findings within a cohort, similar to a Google PageRank algorithm. Genes prioritized in such a way may not have the highest *P*-values, but will generalize and reproduce well in independent cohorts. Since this approach is based on a gene level integration of data rather than a SNP level analysis, it reduces heterogeneity. It can and has been used profitably to mine GWAS [Le-Niculescu, 2009a] and biomarker [Kurian, 2009; Le-Niculescu, 2009b] datasets, and to extract panels of top genes or biomarkers that reproduce well in independent cohorts. Studies that by themselves are relatively under-powered can be mined and made to yield results using CFG, by bringing to bear other large datasets and databases relevant to that disease, resulting in essence in a field-wide collaboration. Cross-validating signals from other sources, with different noise factors, can increase the signal–noise ratio and decrease the size of cohort (n) you need.

HERITABILITY

Heritability estimates based on small twin studies may be overly optimistic [Kaminsky, 2009]. Additionally, the cumulative combinatorics of common gene variants and environment (CCxCGV XE) may play a more prominent role than anticipated in complex disorders [Manolio, 2009], such as psychiatric disorders [Le-Niculescu, 2009a].

CONCLUSIONS

We suggest that the large and continuously growing databases from GWAS are a very useful resource for breaking the genetic code of complex disorders, such as psychiatric disorders. Combining them with complementary approaches that (1) look at sub-phenotypes, (2) use a case–case design, (3) integrate and prioritize findings at a gene level using multiple lines of evidence, including gene expression, has a better chance of yielding findings that are disease relevant, as well as reproducible and predictive in independent cohorts. That is the key litmus test, in our opinion, for any biomarker (or genetic) study. For a complete understanding of the illness, the analyses then need to be pursued at a biological pathway and mechanistic level, integrating environmental effects as key modulators of gene expression and phenotype manifestation. A (r)evolution in medical nosology in general, and psychiatric nosology in particular [Kendler et al., 2009; Niculescu et al., 2009], may follow.

ACKNOWLEDGMENTS

REFERENCES

Anon. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678.

Baum AE, et al. 2008. A genome-wide association study implicates diac-ylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 13:197–207.

Coufal NG, et al. 2009. L1 retrotransposition in human neural progenitor cells. Nature 460:1127–1131.

Kaminsky ZA, et al. 2009. DNA methylation profiles in monozygotic and dizygotic twins. Nat Genet 41:240–245.

Kendler KS, Munoz RA, Murphy G. 2009. The development of the Feighner criteria: A historical perspective. Am[Q2] J Psychiatry.

Kurian SM, et al. 2009. Identification of blood biomarkers for psychosis using convergent functional genomics. Mol[Q3] Psychiatry.

Le-Niculescu H, et al. 2007. Towards understanding the schizophrenia code: An expanded convergent functional genomics approach. Am J Med Genet Part B 144B:129–158.

Le-Niculescu H, et al. 2009a. Convergent functional genomics of genome-wide association data for bipolar disorder: Comprehensive identification of candidate genes, pathways and mechanisms. Am J Med Genet Part B 150B:155–181.

Le-Niculescu H, et al. 2009b. Identifying blood biomarkers for mood disorders using convergent functional genomics. Mol Psychiatry 14:156–174.

Liu X, et al. 2009. Family-based association study between brain-derived neurotrophic factor gene and major depressive disorder of Chinese descent. Psychiatry Res 169:169–172.

Manolio TA, et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747–753.

Niculescu AB III, et al. 2000. Identifying a series of candidate genes for mania and psychosis: A convergent functional genomics approach. Physiol Genomics 4:83–91.

Niculescu AB III, Schork NJ, Salomon DR. 2009. Mindscape: A convergent perspective on life, mind, consciousness and happiness. J Affect[Q4] Disord.

Ogden CA, et al. Candidate genes, pathways and mechanisms for bipolar (manic-depressive) and related disorders: An expanded convergent functional genomics approach. 2004. Mol Psychiatry 9:1007–1029.

Petryshen TL, et al. 2009. Population genetic study of the brain-derived neurotrophic factor (BDNF) gene. Mol[Q5] Psychiatry.

Rodd ZA, et al. 2007. Candidate genes, pathways and mechanisms for alcoholism: An expanded convergent functional genomics approach. Pharmacogenomics J 7:222–256.

She X, et al. 2009. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. BMC Genomics 10(269[Q6]).

Q1: Please check the corresponding authors address.
Q2: Please provide the volume number and page range.
Q3: Please provide the volume number and page range.
Q4: Please provide the volume number and page range.
Q5: Please provide the volume number and page range.
Q6: Please provide the page range.

COPYRIGHT TRANSFER AGREEMENT

WILEY-BLACKWELL

Date: _____ Contributor name: _____

Contributor address: _____

Manuscript number (Editorial office only): _____

Re: Manuscript entitled _____

_____ (the "Contribution")

for publication in _____ (the "Journal")

published by _____ ("Wiley-Blackwell").

Dear Contributor(s):

Thank you for submitting your Contribution for publication. In order to expedite the editing and publishing process and enable Wiley-Blackwell to disseminate your Contribution to the fullest extent, we need to have this Copyright Transfer Agreement signed and returned as directed in the Journal's instructions for authors as soon as possible. If the Contribution is not accepted for publication, or if the Contribution is subsequently rejected, this Agreement shall be null and void. **Publication cannot proceed without a signed copy of this Agreement.**

A. COPYRIGHT

1. The Contributor assigns to Wiley-Blackwell, during the full term of copyright and any extensions or renewals, all copyright in and to the Contribution, and all rights therein, including but not limited to the right to publish, republish, transmit, sell, distribute and otherwise use the Contribution in whole or in part in electronic and print editions of the Journal and in derivative works throughout the world, in all languages and in all media of expression now known or later developed, and to license or permit others to do so.

2. Reproduction, posting, transmission or other distribution or use of the final Contribution in whole or in part in any medium by the Contributor as permitted by this Agreement requires a citation to the Journal and an appropriate credit to Wiley-Blackwell as Publisher, and/or the Society if applicable, suitable in form and content as follows: (Title of Article, Author, Journal Title and Volume/Issue, Copyright © [year], copyright owner as specified in the Journal). Links to the final article on Wiley-Blackwell's website are encouraged where appropriate.

B. RETAINED RIGHTS

Notwithstanding the above, the Contributor or, if applicable, the Contributor's Employer, retains all proprietary rights other than copyright, such as patent rights, in any process, procedure or article of manufacture described in the Contribution.

C. PERMITTED USES BY CONTRIBUTOR

1. Submitted Version. Wiley-Blackwell licenses back the following rights to the Contributor in the version of the Contribution as originally submitted for publication:

a. After publication of the final article, the right to self-archive on the Contributor's personal website or in the Contributor's institution's/employer's institutional repository or archive. This right extends to both intranets and the Internet. The Contributor may not update the submission version or replace it with the published Contribution. The version posted must contain a legend as follows: This is the pre-peer reviewed version of the following article: FULL CITE, which has been published in final form at [Link to final article].

b. The right to transmit, print and share copies with colleagues.

2. Accepted Version. Re-use of the accepted and peer-reviewed (but not final) version of the Contribution shall be by separate agreement with Wiley-Blackwell. Wiley-Blackwell has agreements with certain funding agencies governing reuse of this version. The details of those relationships, and other offerings allowing open web use, are set forth at the following website: http://www.wiley.com/go/funderstatement. NIH grantees should check the box at the bottom of this document.

3. Final Published Version. Wiley-Blackwell hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution:

a. Copies for colleagues. The personal right of the Contributor only to send or transmit individual copies of the final published version in any format to colleagues upon their specific request provided no fee is charged, and further-provided that there is no systematic distribution of the Contribution, e.g. posting on a listserve, website or automated delivery.

b. Re-use in other publications. The right to re-use the final Contribution or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications should be accurately noted.

c. Teaching duties. The right to include the Contribution in teaching or training duties at the Contributor's institution/place of employment including in course packs, e-reserves, presentation at professional conferences, in-house training, or distance learning. The Contribution may not be used in seminars outside of normal teaching obligations (e.g. commercial seminars). Electronic posting of the final published version in connection with teaching/training at the Contributor's institution/place of employment is permitted subject to the implementation of reasonable access control mechanisms, such as user name and password. Posting the final published version on the open Internet is not permitted.

d. Oral presentations. The right to make oral presentations based on the Contribution.

4. Article Abstracts, Figures, Tables, Data Sets, Artwork and Selected Text (up to 250 words).

a. Contributors may re-use unmodified abstracts for any non-commercial purpose. For on-line uses of the abstracts, Wiley-Blackwell encourages but does not require linking back to the final published versions.

b. Contributors may re-use figures, tables, data sets, artwork, and selected text up to 250 words from their Contributions, provided the following conditions are met:

(i) Full and accurate credit must be given to the Contribution.
(ii) Modifications to the figures, tables and data must be noted. Otherwise, no changes may be made.
(iii) The reuse may not be made for direct commercial purposes, or for financial consideration to the Contributor.
(iv) Nothing herein shall permit dual publication in violation of journal ethical practices.

CTA-A

D. CONTRIBUTIONS OWNED BY EMPLOYER

1. If the Contribution was written by the Contributor in the course of the Contributor's employment (as a "work-made-for-hire" in the course of employment), the Contribution is owned by the company/employer which must sign this Agreement (in addition to the Contributor's signature) in the space provided below. In such case, the company/employer hereby assigns to Wiley-Blackwell, during the full term of copyright, all copyright in and to the Contribution for the full term of copyright throughout the world as specified in paragraph A above.

2. In addition to the rights specified as retained in paragraph B above and the rights granted back to the Contributor pursuant to paragraph C above, Wiley-Blackwell hereby grants back, without charge, to such company/employer, its subsidiaries and divisions, the right to make copies of and distribute the final published Contribution internally in print format or electronically on the Company's internal network. Copies so used may not be resold or distributed externally. However the company/employer may include information and text from the Contribution as part of an information package included with software or other products offered for sale or license or included in patent applications. Posting of the final published Contribution by the institution on a public access website may only be done with Wiley-Blackwell's written permission, and payment of any applicable fee(s). Also, upon payment of Wiley-Blackwell's reprint fee, the institution may distribute print copies of the published Contribution externally.

E. GOVERNMENT CONTRACTS

In the case of a Contribution prepared under U.S. Government contract or grant, the U.S. Government may reproduce, without charge, all or portions of the Contribution and may authorize others to do so, for official U.S. Govern-ment purposes only, if the U.S. Government contract or grant so requires. (U.S. Government, U.K. Government, and other government employees: see notes at end)

F. COPYRIGHT NOTICE

The Contributor and the company/employer agree that any and all copies of the final published version of the Contribution or any part thereof distributed or posted by them in print or electronic format as permitted herein will include the notice of copyright as stipulated in the Journal and a full citation to the Journal as published by Wiley-Blackwell.

G. CONTRIBUTOR'S REPRESENTATIONS

The Contributor represents that the Contribution is the Contributor's original work, all individuals identified as Contributors actually contributed to the Con-tribution, and all individuals who contributed are included. If the Contribution was prepared jointly, the Contributor agrees to inform the co-Contributors of the terms of this Agreement and to obtain their signature to this Agreement or their written permission to sign on their behalf. The Contribution is submitted only to this Journal and has not been published before. (If excerpts from copy-righted works owned by third parties are included, the Contributor will obtain written permission from the copyright owners for all uses as set forth in Wiley-Blackwell's permissions form or in the Journal's Instructions for Contributors, and show credit to the sources in the Contribution.) The Contributor also warrants that the Contribution contains no libelous or unlawful statements, does not infringe upon the rights (including without limitation the copyright, patent or trademark rights) or the privacy of others, or contain material or instructions that might cause harm or injury.

CHECK ONE BOX:

☐ Contributor-owned work

ATTACH ADDITIONAL SIGNATURE PAGES AS NECESSARY

Contributor's signature _____ Date _____

Type or print name and title _____

Co-contributor's signature _____ Date _____

Type or print name and title _____

☐ Company/Institution-owned work
(made-for-hire in the course of employment)

Company or Institution (Employer-for-Hire) _____ Date _____

Authorized signature of Employer _____ Date _____

☐ U.S. Government work

Note to U.S. Government Employees
A contribution prepared by a U.S. federal government employee as part of the employee's official duties, or which is an official U.S. Government publication, is called a "U.S. Government work," and is in the public domain in the United States. In such case, the employee may cross out Paragraph A.1 but must sign (in the Contributor's signature line) and return this Agreement. If the Contribution was not prepared as part of the employee's duties or is not an official U.S. Government publication, it is not a U.S. Government work.

☐ U.K. Government work
(Crown Copyright)

Note to U.K. Government Employees
The rights in a Contribution prepared by an employee of a U.K. government department, agency or other Crown body as part of his/her official duties, or which is an official government publication, belong to the Crown. U.K. government authors should submit a signed declaration form together with this Agreement. The form can be obtained via http://www.opsi.gov.uk/advice/crown-copyright/copyright-guidance/publication-of-articles-written-by-ministers-and-civil-servants.htm

☐ Other Government work

Note to Non-U.S., Non-U.K. Government Employees
If your status as a government employee legally prevents you from signing this Agreement, please contact the editorial office.

☐ NIH Grantees

Note to NIH Grantees
Pursuant to NIH mandate, Wiley-Blackwell will post the accepted version of Contributions authored by NIH grant-holders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

WILEY
Publishers Since 1807

COLOR REPRODUCTION IN YOUR ARTICLE

These proofs have been typeset using figure files transmitted to production when this article was accepted for publication. Please review all figures and note your approval with your submitted proof corrections. You may contact the journal production team at **AJMBprod@wiley.com** if you wish to discuss specific concerns.

Because of the high cost of color printing, we can only print figures in color if authors cover the expense. If you have submitted color figures, please indicate your consent to cover the cost on the table listed below by marking the box corresponding to the approved cost on the table.

Please note, all color images will be reproduced online in Wiley *InterScience* at no charge, whether or not you opt for color printing.

You will be invoiced for color charges once the article has been published in print.

Failure to return this form with your article proofs may delay the publication of your article.

JOURNAL AMERICAN JOURNAL OF MEDICAL GENETICS PART B _____ MS. NO. _____ NO. COLOR PAGES _____

MANUSCRIPT TITLE _____

AUTHOR(S) _____

No. Color Pages	Color Charge	No. Color Pages	Color Charge	No. Color Pages	Color Charge
☐ 1	$950	☐ 5	$3400	☐ 9	$5850
☐ 2	$1450	☐ 6	$3900	☐ 10	$6350
☐ 3	$1950	☐ 7	$4400	☐ 11	$6850
☐ 4	$2450	☐ 8	$4900	☐ 12	$7350
Contact AJMBprod@wiley.com for a quote if you have more than 12 pages of color					

☐ Please print my figures color   ☐ Please print my figures in black and white

☐ Please print the following figures in color   _____

and convert these figures to black and white   _____

Approved by   _____

Billing Address   _____   E-mail _____

_____   Telephone _____

_____   Fax _____

John Wiley & Sons, Inc.
Publishers Since 1807

PREPUBLICATION REPRINT ORDER FORM

Please complete this form even if you are not ordering reprints. Please be sure to include your article number in the appropriate place to avoid delays and/or errors with your order. This form **MUST** be returned with your corrections. Your reprints will be shipped approximately 4 weeks after print publication. Reprints ordered after printing are substantially more expensive.

JOURNAL: *AMERICAN JOURNAL OF MEDICAL GENETICS PART B: NEUROPSYCHIATRIC GENETICS*

TITLE OF MANUSCRIPT: _____

ARTICLE NO.: _____ NO. OF PAGES: _____ AUTHOR(S): _____

REPRINTS 8 1/4 X 11

No. of Pages	100 Reprints	200 Reprints	300 Reprints	400 Reprints	500 Reprints
	$	$	$	$	$
1-4	336	501	694	890	1,052
5-8	469	703	987	1,251	1,477
9-12	594	923	1,234	1,565	1,850
13-16	714	1,156	1,527	1,901	2,273
17-20	794	1,340	1,775	2,212	2,648
21-24	911	1,529	2,031	2,536	3,037
25-28	1,004	1,707	2,267	2,828	3,388
29-32	1,108	1,894	2,515	3,135	3,755
33-36	1,219	2,092	2,773	3,456	4,143
37-40	1,329	2,290	3,033	3,776	4,528

** REPRINTS ARE ONLY AVAILABLE IN LOTS OF 100. IF YOU WISH TO ORDER MORE THAN 500 REPRINTS, PLEASE CONTACT OUR REPRINTS DEPARTMENT AT (201) 748-8789 FOR A PRICE QUOTE.

COVERS

100 Covers - $90	200 Covers - $145	300 Covers - $200
400 Covers - $255	500 Covers - $325	Additional 100s - $65

**International orders must be paid in U.S. currency and drawn on a U.S. bank

Please send me _____ reprints of the above article at...................... $_____

Please send me _____ generic covers of the above journal at...................... $_____

Please add appropriate State and Local Tax {Tax Exempt No._____} $_____

Please add 5% Postage and Handling... $_____

TOTAL AMOUNT OF ORDER** .... $_____

Please circle one: Check enclosed   Bill me   Credit card

If credit card order, charge to: American Express   Visa   MasterCard   Discover

Credit Card No. _____ Signature _____ Exp. Date _____

Bill To: Ship To:

Name: _____

Address/Institution: _____

Purchase Order No.: _____ Phone: _____ Fax: _____

E-mail: _____