

1 Integrating Multiple Correlated Phenotypes for Genetic Association
2 Analysis by Maximizing Heritability
3

Jin J. Zhou^{1,2}, Michael H. Cho^{3,4,5}, Christoph Lange², Sharon Lutz⁶,
Edwin K. Silverman^{3,4,5}, and Nan M. Laird²

4 January 4, 2015

5 ¹ Division of Epidemiology and Biostatistics, College of Public Health, University of Arizona,
6 Tucson, AZ 85724

7 ²Department of Biostatistics, Harvard University

8 ³Channing Division of Network Medicine, Department of Medicine, Brigham and Womens Hospital

9 ⁴Harvard Medical School

10 ⁵Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and
11 Womens Hospital

12 Boston, MA 02115

13 ⁶Department of Biostatistics and Informatics, University of Colorado, Anschutz Medical Campus,

14 Aurora, CO 80045

15 Running Title:

16 Integrating Correlated Phenotypes

17 Corresponding author:

18 Jin Zhou

19 Division of Epidemiology and Biostatistics, College of Public Health

20 University of Arizona, Tucson, AZ 85724

21 Phone: (520) 626-1393

22 Email: jzhou@email.arizona.edu

23 Supplementary Material

24 How to generate MaxH phenotype

25 We start with a set of phenotypes, i.e., a n by T phenotypic matrix \mathbf{Y} and sample GWAS genotypes.

- 26 1. Generate the standardized residuals of the T phenotypes by regressing out the covariates and
27 standardizing the residuals. At this step no SNPs are include in the covariates.
- 28 2. Estimate the genetic relationship matrix \mathbf{G} from the GWAS genotype data using Genome-wide
29 Complex Trait Analysis (GCTA) (Yang et al., 2011).
- 30 3. Estimate \mathbf{V}_p and \mathbf{V}_g using \mathbf{G} matrix, standardd residuals of phenotypes and either ML (e.g.
31 Mendel (Lange et al., 2013)), REML (e.g. GCTA (Yang et al., 2011)) or MOM methods.
- 32 4. Eigen-decompose \mathbf{V}_g with eigenvectors $\mathbf{\Gamma}$ and set of eigenvalues $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_k\}$, such
33 that,

$$\mathbf{V}_g^{\frac{1}{2}} = \mathbf{\Gamma} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{\Gamma},$$

34 where $\mathbf{\Lambda}^{\frac{1}{2}} = \text{diag}\{\lambda_1^{\frac{1}{2}}, \dots, \lambda_k^{\frac{1}{2}}\}$.

- 35 5. Eigen decompose $\mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_p^{-1} \mathbf{V}_g^{\frac{1}{2}}$ with eigenvectors \mathbf{W} .
- 36 6. Take the eigenvector with the largest eigenvalue, say \mathbf{w} and transform it as $\mathbf{l} = \mathbf{V}_g^{-\frac{1}{2}} \mathbf{w}$. \mathbf{Yl}
37 is the MaxH phenotype.
- 38 7. The first eigenvalue of $\mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_p^{-1} \mathbf{V}_g^{\frac{1}{2}}$ is the maximized heritability.

39 Except steps 2 and 3, the statistical software package R (<http://www.r-project.org/>) is used
40 throughout. Note that besides using the first eigenvector of the matrix $\mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_p^{-1} \mathbf{V}_g^{\frac{1}{2}}$ to generate
41 MaxH phenotype, one can use all the eigenvector of the matrix $\mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_p^{-1} \mathbf{V}_g^{\frac{1}{2}}$ to perform a multivariate
42 regression analysis.

43 Solve for the linear coefficients

44 To optimize the objective function,

$$h_l^2 = \frac{\mathbf{l}' \mathbf{V}_g \mathbf{l}}{\mathbf{l}' \mathbf{V}_p \mathbf{l}},$$

45 with respect to vector \mathbf{l} . One takes the first derivative of h_l^2 with respect to \mathbf{l} , set the formula equal
46 to zero and solve for \mathbf{l} . Since $\mathbf{l}' \mathbf{V}_p \mathbf{l} \neq 0$, it is straightforward to verify that the vector \mathbf{l} needs to

47 satisfy,

$$\mathbf{V}_g \mathbf{l} = \lambda \mathbf{V}_p \mathbf{l} \quad \text{or} \quad \mathbf{V}_p^{-1} \mathbf{V}_g \mathbf{l} = \lambda \mathbf{l}. \quad (1)$$

48 where $\lambda = h_l^2$. This is a generalized eigen-problem if the matrix $\mathbf{V}_p^{-1} \mathbf{V}_g$ is not symmetric. Since \mathbf{V}_g
 49 is positive definite it can be written as $\mathbf{V}_g = \mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_g^{\frac{1}{2}}$, where $\mathbf{V}_g^{\frac{1}{2}}$ is constructed from its eigenvalue
 50 decomposition. Defining the transformation $\mathbf{w} = \mathbf{V}_g^{\frac{1}{2}} \mathbf{l}$, we rewrite equation (1) as

$$\mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_p^{-1} \mathbf{V}_g^{\frac{1}{2}} \mathbf{w} = \lambda \mathbf{w}, \quad (2)$$

51 where $\mathbf{V}_g^{\frac{1}{2}} \mathbf{V}_p^{-1} \mathbf{V}_g^{\frac{1}{2}}$ is a symmetric matrix. We now turn a generalized eigen-problem (1) into a
 52 regular eigenvalue problem (2) and find its eigenvalues λ and eigenvectors \mathbf{w} . Finally the linear
 53 coefficients are defined as, $\mathbf{l} = \mathbf{V}_g^{-\frac{1}{2}} \mathbf{w}$ and the largest eigenvalue is the maximized the heritability
 54 h_l^2 .

55 Supplementary Figures

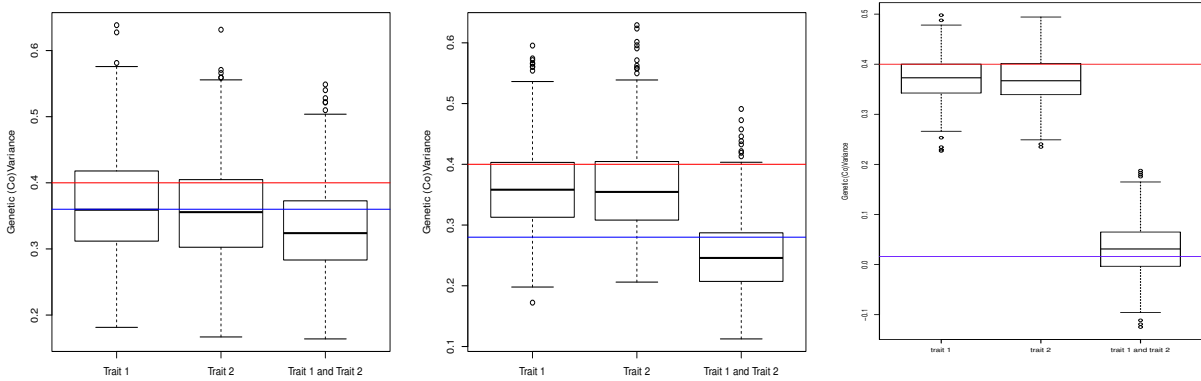


Figure 1: Parameter estimation when combining two phenotypes. Both phenotypes have heritability 0.4. Red line represents true heritability and blue line represents true genetic covariance, i.e., 0.36 (left), 0.28 (middle), and 0.04(right)

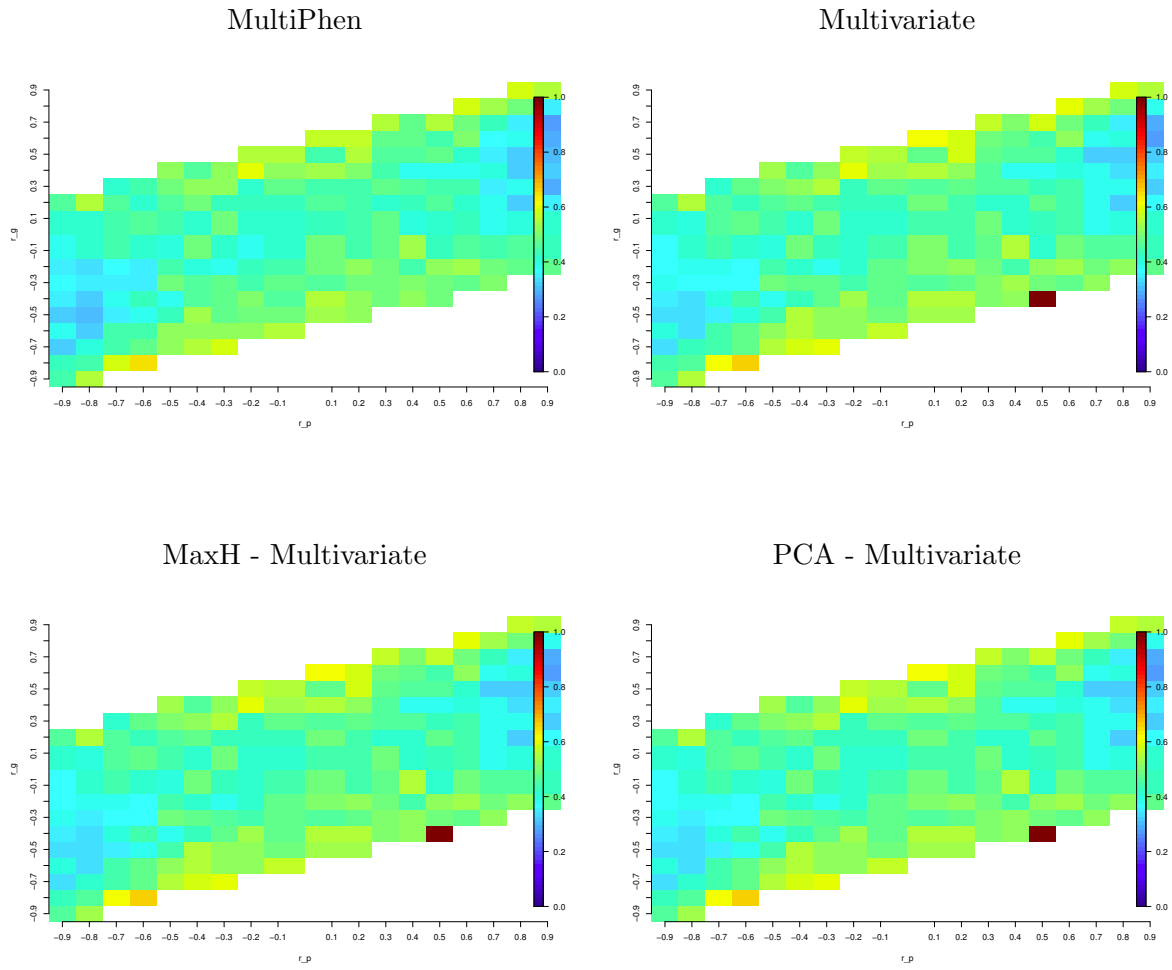
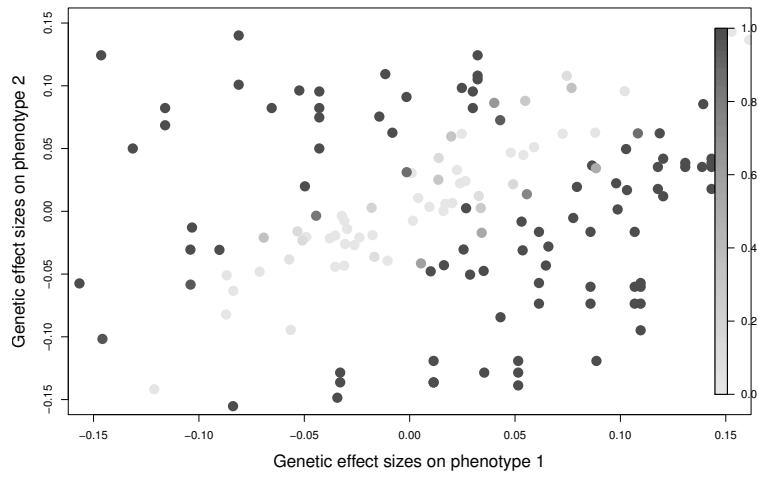


Figure 2: Proportion of 100 SNPs with empirical power greater than 0.8 as a function of r_g and r_p using multivariate associate tests.

MaxH



PCA

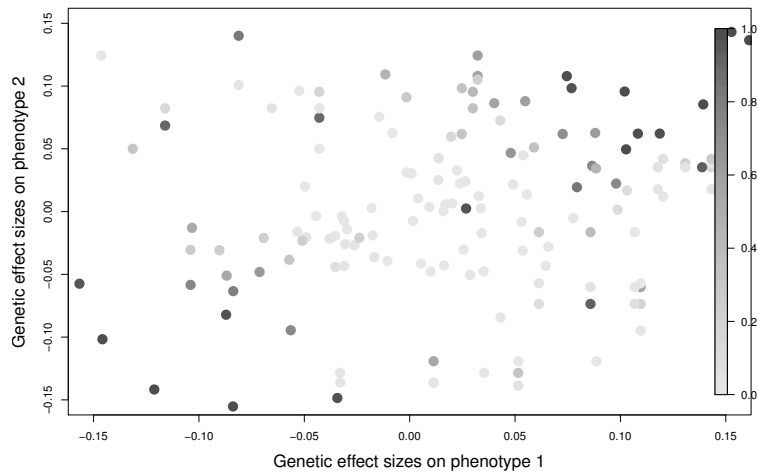


Figure 3: 150 SNPs' empirical power as a function of effects sizes to both traits, when $r_p = 0.8$ and $r_g = 0.7$. 150 SNPs's effect sizes were chosen arbitrarily from uniform distribution in region $[-0.15, 0.15] \times [-0.15, 0.15]$. Color bar represents the scale of power.

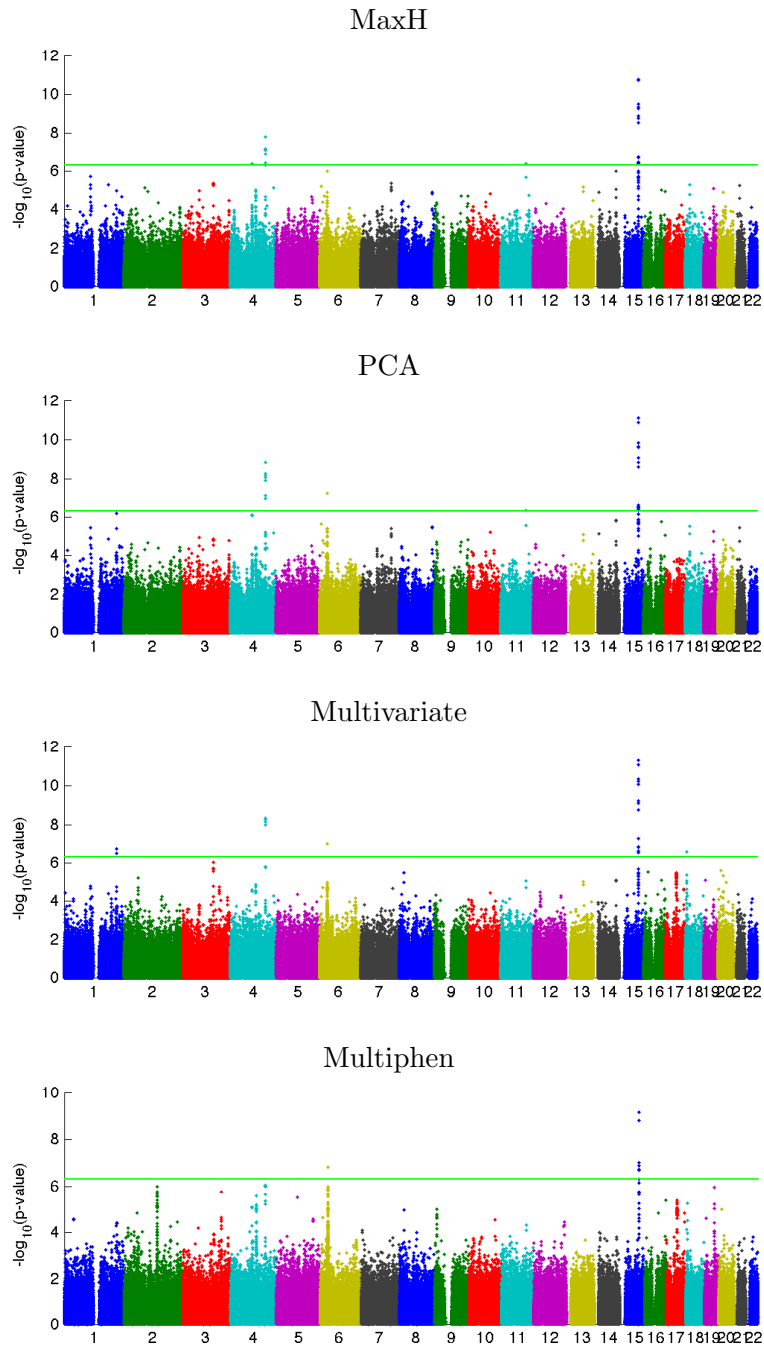


Figure 4: Manhattan plots of results from methods, MaxH, PCA, multivariate regression, and Multiphen when using phenotypes, FEV₁, FEV₁/FVC, log(pctEmph) and COPDGene sample.

56 **Acknowledgments**

57 Appendix 3: COPDGene Investigators - Core Units

58 Administrative Core: James Crapo, MD (PI), Edwin Silverman, MD, PhD (PI), Barry Make, MD,
59 Elizabeth Regan, MD, PhD, Rochelle Lantz, Lori Stepp, Sandra Melanson

60 Genetic Analysis Core: Terri Beaty, PhD, Barbara Klanderma, PhD, Nan Laird, PhD, Christoph
61 Lange, PhD, Michael Cho, MD, Stephanie Santorico, PhD, John Hokanson, MPH, PhD, Dawn
62 DeMeo, MD, MPH, Nadia Hansel, MD, MPH, Craig Hersh, MD, MPH, Peter Castaldi, MD, MSc,
63 Merry-Lynn McDonald, PhD, Jing Zhou, MD, PhD, Manuel Mattheissen, MD, PhD, Emily Wan,
64 MD, Megan Hardin, MD, Jacqueline Hetmanski, MS, Margaret Parker, MS, Tanda Murray, MS

65 Imaging Core: David Lynch, MB, Joyce Schroeder, MD, John Newell, Jr., MD, John Reilly, MD,
66 Harvey Coxson, PhD, Philip Judy, PhD, Eric Hoffman, PhD, George Washko, MD, Raul San Jose
67 Estepar, PhD, James Ross, MSc, Mustafa Al Qaisi, MD, Jordan Zach, Alex Kluiber, Jered Sieren,
68 Tanya Mann, Deanna Richert, Alexander McKenzie, Jaleh Akhavan, Douglas Stinson

69 PFT QA Core, LDS Hospital, Salt Lake City, UT: Robert Jensen, PhD

70 Biological Repository, Johns Hopkins University, Baltimore, MD: Homayoon Farzadegan, PhD,
71 Stacey Meyerer, Shivam Chandan, Samantha Bragan

72 Data Coordinating Center and Biostatistics, National Jewish Health, Denver, CO: Douglas Everett,
73 PhD, Andre Williams, PhD, Carla Wilson, MS, Anna Forssen, MS, Amber Powell, Joe Piccoli

74 Epidemiology Core, University of Colorado School of Public Health, Denver, CO: John Hokanson,
75 MPH, PhD, Marci Sontag, PhD, Jennifer Black-Shinn, MPH, Gregory Kinney, MPH, PhDc, Sharon
76 Lutz, MPH, PhD

77 Appendix 4: COPDGene Investigators - Clinical Centers

78 Ann Arbor VA: Jeffrey Curtis, MD, Ella Kazerooni, MD

79 Baylor College of Medicine, Houston, TX: Nicola Hanania, MD, MS, Philip Alapat, MD, Venkata
80 Bandi, MD, Kalpalatha Guntupalli, MD, Elizabeth Guy, MD, Antara Mallampalli, MD, Charles
81 Trinh, MD, Mustafa Atik, MD, Hasan Al-Azzawi, MD, Marc Willis, DO, Susan Pinero, MD, Linda
82 Fahr, MD, Arun Nachiappan, MD, Collin Bray, MD, L. Alexander Frigini, MD, Carlos Farinas,
83 MD, David Katz, MD, Jose Freytes, MD, Anne Marie Marciel, MD

84 Brigham and Womens Hospital, Boston, MA: Dawn DeMeo, MD, MPH, Craig Hersh, MD, MPH,
85 George Washko, MD, Francine Jacobson, MD, MPH, Hiroto Hatabu, MD, PhD, Peter Clarke, MD,
86 Ritu Gill, MD, Andetta Hunsaker, MD, Beatrice Trotman-Dickenson, MBBS, Rachna Madan, MD

87 Columbia University, New York, NY: R. Graham Barr, MD, DrPH, Byron Thomashow, MD, John
88 Austin, MD, Belinda DSouza, MD
89 Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr., MD, Lacey Washington, MD,
90 H Page McAdams, MD
91 Fallon Clinic, Worcester, MA: Richard Rosiello, MD, Timothy Bresnahan, MD, Joseph Bradley,
92 MD, Sharon Kuong, MD, Steven Meller, MD, Suzanne Roland, MD
93 Health Partners Research Foundation, Minneapolis, MN: Charlene McEvoy, MD, MPH, Joseph
94 Tashjian, MD
95 Johns Hopkins University, Baltimore, MD: Robert Wise, MD, Nadia Hansel, MD, MPH, Robert
96 Brown, MD, Gregory Diette, MD, Karen Horton, MD
97 Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Los Angeles, CA:
98 Richard Casaburi, MD, Janos Porszasz, MD, PhD, Hans Fischer, MD, PhD, Matt Budoff, MD,
99 Mehdi Rambod, MD
100 Michael E. DeBakey VAMC, Houston, TX: Amir Sharafkhaneh, MD, Charles Trinh, MD, Hirani
101 Kamal, MD, Roham Darvishi, MD, Marc Willis, DO, Susan Pinero, MD, Linda Fahr, MD, Arun
102 Nachiappan, MD, Collin Bray, MD, L. Alexander Frigini, MD, Carlos Farinas, MD, David Katz,
103 MD, Jose Freytes, MD, Anne Marie Marciel, MD
104 Minneapolis VA: Dennis Niewoehner, MD, Quentin Anderson, MD, Kathryn Rice, MD, Audrey
105 Caine, MD
106 Morehouse School of Medicine, Atlanta, GA: Marilyn Foreman, MD, MS, Gloria Westney, MD, MS,
107 Eugene Berkowitz, MD, PhD
108 National Jewish Health, Denver, CO: Russell Bowler, MD, PhD, David Lynch, MB, Joyce Schroeder,
109 MD, Valerie Hale, MD, John Armstrong, II, MD, Debra Dyer, MD, Jonathan Chung, MD, Christian
110 Cox, MD
111 Temple University, Philadelphia, PA: Gerard Criner, MD, Victor Kim, MD, Nathaniel Marchetti,
112 DO, Aditi Satti, MD, A. James Mamary, MD, Robert Steiner, MD, Chandra Dass, MD, Libby
113 Cone, MD
114 University of Alabama, Birmingham, AL: William Bailey, MD, Mark Dransfield, MD, Michael Wells,
115 MD, Surya Bhatt, MD, Hrudaya Nath, MD, Satinder Singh, MD
116 University of California, San Diego, CA: Joe Ramsdell, MD, Paul Friedman, MD
117 University of Iowa, Iowa City, IA: Alejandro Cornellias, MD, John Newell, Jr., MD, Edwin JR van
118 Beek, MD, PhD

119 University of Michigan, Ann Arbor, MI: Fernando Martinez, MD, MeiLan Han, MD, Ella Kazerooni,
120 MD
121 University of Minnesota, Minneapolis, MN: Christine Wendt, MD, Tadashi Allen, MD
122 University of Pittsburgh, Pittsburgh, PA: Frank Sciorba, MD, Joel Weissfeld, MD, MPH, Carl
123 Fuhrman, MD, Jessica Bon, MD, Danielle Hooper, MD
124 University of Texas Health Science Center at San Antonio, San Antonio, TX: Antonio Anzueto, MD,
125 Sandra Adams, MD, Carlos Orozco, MD, Mario Ruiz, MD, Amy Mumbower, MD, Ariel Kruger,
126 MD, Carlos Restrepo, MD, Michael Lane, MD

127 **References**

128 Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, et al. (2013) Mendel: the Swiss army knife
129 of genetic analysis programs. *Bioinformatics* 29: 1568–70.

130 Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait
131 analysis. *Am J Hum Genet* 88: 76–82.