

The Open University's repository of research publications  
and other research outputs

## Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits - the Hispanic/Latino Anthropometry Consortium

### Journal Item

#### How to cite:

Fernández-Rhodes, Lindsay; Graff, Mariaelisa; Buchanan, Victoria L.; Justice, Anne E.; Highland, Heather M.; Guo, Xiuqing; Zhu, Wanying; Chen, Hung-Hsin; Young, Kristin L.; Adhikari, Kaustubh; Allred, Nicholette (Palmer); Below, Jennifer E.; Bradfield, Jonathan; Pereira, Alexandre C.; Glover, LáShauntá; Kim, Daeun; Lilly, Adam G.; Shrestha, Poojan; Thomas, Alvin G.; Zhang, Xinruo; Chen, Minhui; Chiang, Charleston W.K.; Pulit, Sara; Horimoto, Andrea; Krieger, Jose E.; Guindo-Martinez, Marta; Preuss, Michael; Schumann, Claudia; Smit, Roelof A.J.; Torres-Mejía, Gabriela; Acuña-Alonzo, Victor; Bedoya, Gabriel; Bortolini, Maria-Cátira; Canizales-Quinteros, Samuel; Gallo, Carla; González-José, Rolando; Poletti, Giovanni; Rothhammer, Francisco; Hakonarson, Hakon; Igo, Robert; Adler, Sharon G.; Iyengar, Sudha K.; Nicholas, Susanne B.; Gogarten, Stephanie M.; Isasi, Carmen R.; Papnicolaou, George; Stimp, Adrienne M.; Qi, Qibin; Kho, Minjung; Smith, Jennifer A.; Langfeld, Carl; Wagenknecht, Lynne; Mckean-Cowdin, Roberta; Gao, Xiaoyi Raymond; Noursome, Darryl; Conti, David V.; Feng, Ye; Allison, Matthew A.; Arzumanyan, Zorayr; Buchanan, Thomas A.; Ida Chen, Yii-Der; Genter, Pauline M.; Goodarzi, Mark O.; Hai, Yang; Hsueh, Willa; Ipp, Eli; Kandeel, Fouad R.; Lam, Kelvin; Li, Xiaohui; Nadler, Jerry L.; Raffel, Leslie J.; Roll, Kaye; Sandow, Kevin; Tan, Jingyi; Taylor, Kent D.; Xiang, Anny H.; Yao, Jie; Audirac-Chalifour, Astride; de Jesus Peralta Romero, Jose; Hartwig, Fernando; Horta, Bernardo; Blangero, John; Curran, Joanne E.; Duggirala, Ravindranath; Lehman, Donna E.; Puppala, Sobha; Fejerman, Laura; John, Esther; Aguilar-Salinas, Carlos; Burt, Noël P.; Florez, Jose C.; García-Ortíz, Humberto; González-Villalpando, Clicerio; Mercader, Josep; Orozco, Lorena; Tusié, Teresa; Blanco, Estela; Gahagan, Sheila; Cox, Nancy J.; Hani, Craig; Butte, Nancy F.; Cole, Shelley A.; Commuzie, Anthony G.; Voruganti, V. Saroja; Rohde, Rebecca; Wang, Yujie; Sofer, Tamar; Ziv, Elad; Grant, Struan F.A.; Ruiz-Linares, Andres; Rotter, Jerome I.; Haiman, Christopher A.; Parra, Esteban J.; Cruz, Miguel; Loos, Ruth J.F. and North., Kari E. (2022). Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits - the Hispanic/Latino Anthropometry Consortium. *Human Genetics and Genomics Advances (Early Access)*.

For guidance on citations see [FAQs](#).

© 2022 The Authors



<https://creativecommons.org/licenses/by/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1016/j.xhgg.2022.100099>

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# Journal Pre-proof

Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits - the Hispanic/Latino Anthropometry Consortium

Lindsay Fernández-Rhodes, Mariaelisa Graff, Victoria L. Buchanan, Anne E. Justice, Heather M. Highland, Xiuqing Guo, Wanying Zhu, Hung-Hsin Chen, Kristin L. Young, Kaustubh Adhikari, Nicholette (Palmer) Allred, Jennifer E. Below, Jonathan Bradfield, Alexandre C. Pereira, LáShauntá Glover, Daeun Kim, Adam G. Lilly, Poojan Shrestha, Alvin G. Thomas, Xinruo Zhang, Minhui Chen, Charleston W.K. Chiang, Sara Pulit, Andrea Horimoto, Jose E. Krieger, Marta Guindo-Martinez, Michael Preuss, Claudia Schumann, Roelof A.J. Smit, Gabriela Torres-Mejía, Victor Acuña-Alonzo, Gabriel Bedoya, Maria-Cátira Bortolini, Samuel Canizales-Quinteros, Carla Gallo, Rolando González-José, Giovanni Poletti, Francisco Rothhammer, Hakon Hakonarson, Robert Igo, Sharon G. Adler, Sudha K. Iyengar, Susanne B. Nicholas, Stephanie M. Gogarten, Carmen R. Isasi, George Papnicolaou, Adrienne M. Stilp, Qibin Qi, Minjung Kho, Jennifer A. Smith, Carl Langfeld, Lynne Wagenknecht, Roberta Mckean-Cowdin, Xiaoyi Raymond Gao, Darryl Noursome, David V. Conti, Ye Feng, Matthew A. Allison, Zorayr Arzumanyan, Thomas A. Buchanan, Yii-Der Ida Chen, Pauline M. Genter, Mark O. Goodarzi, Yang Hai, Willa Hsueh, Eli Ipp, Fouad R. Kandeel, Kelvin Lam, Xiaohui Li, Jerry L. Nadler, Leslie J. Raffel, Kaye Roll, Kevin Sandow, Jingyi Tan, Kent D. Taylor, Anny H. Xiang, Jie Yao, Astride Audirac-Chalifour, Jose de Jesus Peralta Romero, Fernando Hartwig, Bernardo Horta, John Blangero, Joanne E. Curran, Ravindranath Duggirala, Donna E. Lehman, Sobha Puppala, Laura Fejerman, Esther John, Carlos Aguilar-Salinas, Noël P. Burt, Jose C. Florez, Humberto García-Ortíz, Clicerio González-Villalpando, Josep Mercader, Lorena Orozco, Teresa Tusié, Estela Blanco, Sheila Gahagan, Nancy J. Cox, Craig Hanis, Nancy F. Butte, Shelley A. Cole, Anthony G. Commuzzie, V. Saroja Voruganti, Rebecca Rohde, Yujie Wang, Tamar Sofer, Elad Ziv, Struan F.A. Grant, Andres Ruiz-Linares, Jerome I. Rotter, Christopher A. Haiman, Esteban J. Parra, Miguel Cruz, Ruth J.F. Loos, Kari E. North.

PII: S2666-2477(22)00015-X

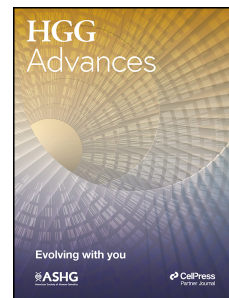
DOI: <https://doi.org/10.1016/j.xhgg.2022.100099>

Reference: XHGG 100099

To appear in: *Human Genetics and Genomics Advances*

Received Date: 11 October 2021

Accepted Date: 6 March 2022



Please cite this article as: Fernández-Rhodes L, Graff M, Buchanan VL, Justice AE, Highland HM, Guo X, Zhu W, Chen H-H, Young KL, Adhikari K, Allred N(P), Below JE, Bradfield J, Pereira AC, Glover L, Kim D, Lilly AG, Shrestha P, Thomas AG, Zhang X, Chen M, Chiang CWK, Pulit S, Horimoto A, Krieger JE, Guindo-Martinez M, Preuss M, Schumann C, Smit RAJ, Torres-Mejía G, Acuña-Alonzo V, Bedoya G, Bortolini M-C, Canizales-Quinteros S, Gallo C, González-José R, Poletti G, Rothhammer F, Hakonarson H, Igo R, Adler SG, Iyengar SK, Nicholas SB, Gogarten SM, Isasi CR, Papnicolaou G, Stilp AM, Qi Q, Kho M, Smith JA, Langfeld C, Wagenknecht L, Mckean-Cowdin R, Gao XR, Noursome D, Conti DV, Feng Y, Allison MA, Arzumanyan Z, Buchanan TA, Ida Chen Y-D, Genter PM, Goodarzi MO, Hai Y, Hsueh W, Ipp E, Kandeel FR, Lam K, Li X, Nadler JL, Raffel LJ, Roll K, Sandow K, Tan J, Taylor KD, Xiang AH, Yao J, Audirac-Chalifour A, de Jesus Peralta Romero J, Hartwig F, Horta B, Blangero J, Curran JE, Duggirala R, Lehman DE, Puppala S, Fejerman L, John E, Aguilar-Salinas C, Burtt NP, Florez JC, García-Ortiz H, González-Villalpando C, Mercader J, Orozco L, Tusié T, Blanco E, Gahagan S, Cox NJ, Hanis C, Butte NF, Cole SA, Commuzzie AG, Voruganti VS, Rohde R, Wang Y, Sofer T, Ziv E, Grant SFA, Ruiz-Linares A, Rotter JI, Haiman CA, Parra EJ, Cruz M, Loos RJF, North KE, Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits - the Hispanic/Latino Anthropometry Consortium, *Human Genetics and Genomics Advances* (2022), doi: <https://doi.org/10.1016/j.xhgg.2022.100099>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 The Author(s).

**Title****Ancestral diversity improves discovery and fine-mapping of genetic loci for anthropometric traits - the Hispanic/Latino Anthropometry Consortium****Authors and Affiliations**

Lindsay Fernández-Rhodes,<sup>1,2\*</sup> Mariaelisa Graff,<sup>2\*</sup> Victoria L. Buchanan<sup>2</sup>, Anne E. Justice,<sup>2,3</sup> Heather M. Highland,<sup>2</sup> Xiuqing Guo,<sup>4</sup> Wanying Zhu,<sup>5</sup> Hung-Hsin Chen,<sup>5</sup> Kristin L. Young,<sup>2</sup> Kaustubh Adhikari,<sup>6</sup> Nicholette (Palmer) Allred,<sup>7</sup> Jennifer E. Below,<sup>5</sup> Jonathan Bradfield,<sup>8</sup> Alexandre C. Pereira,<sup>9</sup> LáShaunta Glover,<sup>2</sup> Daeun Kim,<sup>2</sup> Adam G. Lilly,<sup>10,11</sup> Poojan Shrestha,<sup>2,12</sup> Alvin G. Thomas,<sup>2</sup> Xinruo Zhang,<sup>2</sup> Minhui Chen,<sup>13</sup> Charleston W. K. Chiang,<sup>13,14</sup> Sara Pulit,<sup>15</sup> Andrea Horimoto,<sup>9</sup> Jose E. Krieger,<sup>9</sup> Marta Guindo-Martinez,<sup>16,17</sup> Michael Preuss,<sup>16</sup> Claudia Schumann,<sup>18</sup> Roelof A.J. Smit,<sup>16</sup> Gabriela Torres-Mejía,<sup>19</sup> Victor Acuña-Alonzo,<sup>20</sup> Gabriel Bedoya,<sup>21</sup> Maria-Cátira Bortolini,<sup>22</sup> Samuel Canizales-Quinteros,<sup>23</sup> Carla Gallo,<sup>24</sup> Rolando González-José,<sup>25</sup> Giovanni Poletti,<sup>24</sup> Francisco Rothhammer,<sup>26</sup> Hakon Hakonarson,<sup>8</sup> Robert Igo,<sup>27</sup> Sharon G Adler,<sup>28</sup> Sudha K. Iyengar,<sup>27</sup> Susanne B. Nicholas,<sup>29</sup> Stephanie M. Gogarten,<sup>30</sup> Carmen R. Isasi,<sup>31</sup> George Papnicolaou,<sup>32</sup> Adrienne M. Stilp,<sup>30</sup> Qibin Qi,<sup>31</sup> Minjung Kho,<sup>33</sup> Jennifer A. Smith,<sup>33</sup> Carl Langfeld,<sup>34</sup> Lynne Wagenknecht,<sup>35</sup> Roberta Mckean-Cowdin,<sup>36</sup> Xiaoyi Raymond Gao,<sup>37</sup> Darryl Noursome,<sup>36</sup> David V. Conti,<sup>13</sup> Ye Feng,<sup>36</sup> Matthew A. Allison,<sup>38</sup> Zorayr Arzumanyan,<sup>4</sup> Thomas A. Buchanan,<sup>29,39</sup> Yii-Der Ida Chen,<sup>4</sup> Pauline M. Genter,<sup>40</sup> Mark O. Goodarzi,<sup>41</sup> Yang Hai,<sup>4</sup> Willa Hsueh,<sup>42</sup> Eli Ipp,<sup>29,40</sup> Fouad R. Kandeel,<sup>43</sup> Kelvin Lam,<sup>4</sup> Xiaohui Li,<sup>4</sup> Jerry L. Nadler,<sup>44</sup> Leslie J. Raffel,<sup>45</sup> Kaye Roll,<sup>4</sup> Kevin Sandow,<sup>4</sup> Jingyi Tan,<sup>4</sup> Kent D. Taylor,<sup>4</sup> Anny H. Xiang,<sup>46</sup> Jie Yao,<sup>4</sup> Astride Audirac-Chalifour,<sup>47</sup> Jose de Jesus Peralta Romero,<sup>47</sup> Fernando Hartwig,<sup>48</sup> Bernardo Horta,<sup>48</sup> John Blangero,<sup>49</sup> Joanne E. Curran,<sup>49</sup> Ravindranath Duggirala,<sup>49</sup> Donna E. Lehman,<sup>50</sup> Sobha Puppala,<sup>51</sup> Laura Fejerman,<sup>52</sup> Esther John,<sup>53</sup> Carlos Aguilar-Salinas,<sup>54</sup> Noël P. Burt,<sup>55</sup> Jose C. Florez,<sup>55-57</sup> Humberto García-Ortíz,<sup>58</sup> Clicerio González-Villalpando,<sup>59</sup> Josep Mercader,<sup>55-57</sup> Lorena Orozco,<sup>58</sup> Teresa Tusié,<sup>60</sup> Estela Blanco,<sup>61</sup> Sheila Gahagan,<sup>61</sup> Nancy J. Cox,<sup>5</sup> Craig Hanis,<sup>62</sup> Nancy F. Butte,<sup>63</sup> Shelley A. Cole,<sup>64</sup> Anthony G. Commuzie,<sup>65</sup> V. Saroja Voruganti,<sup>66</sup> Rebecca Rohde,<sup>2</sup> Yujie Wang,<sup>2</sup> Tamar Sofer,<sup>57,67</sup> Elad Ziv,<sup>68</sup> Struan F.A. Grant,<sup>8</sup> Andres Ruiz-Linares,<sup>69-71</sup> Jerome I. Rotter,<sup>4</sup> Christopher A. Haiman,<sup>13</sup> Esteban J. Parra,<sup>72</sup> Miguel Cruz,<sup>47</sup> Ruth J.F. Loos,<sup>16</sup> Kari E. North.<sup>2,73</sup>

<sup>1</sup>Department of Biobehavioral Health, Pennsylvania State University, University Park, PA, USA 16802

<sup>2</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27599

<sup>3</sup>Department of Biomedical and Translational Informatics, Geisinger Health System, Danville, PA, USA 17822

<sup>4</sup>The Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-University of California Los Angeles Medical Center, Torrance, CA, USA 90502

<sup>5</sup>Vanderbilt Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA 37232

<sup>6</sup>School of Mathematics and Statistics, Faculty of Science, Technology, Engineering and Mathematics, The Open University, Milton Keynes, United Kingdom MK7 6AA

<sup>7</sup>Center for Diabetes Research, Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA 27101

<sup>8</sup>Center for Applied Genomics, Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA 19104

<sup>9</sup>Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of São Paulo, Brazil 05508-220

- <sup>10</sup>Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27599
- <sup>11</sup>Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27599
- <sup>12</sup>Division of Pediatric and Public Health, Adams School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27599
- <sup>13</sup>Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA 90033
- <sup>14</sup>Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA 90007
- <sup>15</sup>Vertex Pharmaceuticals, Oxford, United Kingdom W2 6BD
- <sup>16</sup>The Charles Bronfman Institutes for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA 10029
- <sup>17</sup>The Novo Nordisk Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark 2200
- <sup>18</sup>Hasso Plattner Institute, University of Potsdam, Digital Health Center, Potsdam, Germany 14482
- <sup>19</sup>Department of Research in Cardiovascular Diseases, Diabetes Mellitus, and Cancer, Population Health Research Center, National Institute of Public Health, Cuernavaca, Morelos, Mexico 62100
- <sup>20</sup>National Institute of Anthropology and History, Mexico City, Mexico 06600
- <sup>21</sup>Molecular Genetics Investigation Group, University of Antioquia, Medellín, Colombia 1226
- <sup>22</sup>Department of Genetics, Federal University of Rio Grande do Sul, Porto Alegre, Brazil 90040-060
- <sup>23</sup>Population Genomics Applied to Health Unit, the National Institute of Genomic Medicine and the Faculty of Chemistry at the National Autonomous University of Mexico, Mexico City, Mexico 04510
- <sup>24</sup>Research and Development Laboratories, Faculty of the Sciences and Philosophy, Peruvian University Cayetano Heredia, Lima, Peru 15102
- <sup>25</sup>Patagonian Institute of the Social and Human Sciences, Patagonian National Center, Puerto Madryn, Argentina U9120
- <sup>26</sup>Institute of High Studies, University of Tarapacá, Arica, Chile 1000000
- <sup>27</sup>Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA 44106
- <sup>28</sup>Division of Nephrology and Hypertension, Harbor-University of California Los Angeles Medical Center, Torrance, CA, USA 90502
- <sup>29</sup>Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, CA, USA 90095
- <sup>30</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA 98195
- <sup>31</sup>Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA 10461
- <sup>32</sup>National Heart, Lung and Blood Institute, Bethesda, MD, USA 20892
- <sup>33</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA 48109
- <sup>34</sup>Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC USA 27101
- <sup>35</sup>Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC USA 27101
- <sup>36</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA 90032

- <sup>37</sup>Department of Ophthalmology and Visual Sciences, Department of Biomedical Informatics, Division of Human Genetics, The Ohio State University, Columbus, OH, USA 43210
- <sup>38</sup>Department of Family Medicine and Public Health, University of California, San Diego, CA, USA 92161
- <sup>39</sup>Department of Physiology and Biophysics, Keck School of Medicine of USC, Los Angeles, CA USA 90033
- <sup>40</sup>Department of Medicine, Division of Endocrinology, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA 90502
- <sup>41</sup>Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA 90048
- <sup>42</sup>Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, USA 43210
- <sup>43</sup>Department of Translational Research & Cellular Therapeutics, Beckman Research Institute of City of Hope, Duarte, CA USA 91010
- <sup>44</sup>New York Medical College, School of Medicine, Valhalla, NY, USA 10595
- <sup>45</sup>Division of Genetic and Genomic Medicine, Department of Pediatrics, University of California, Irvine, CA, USA 92697
- <sup>46</sup>Research and Evaluation Branch, Kaiser Permanente of Southern California, Pasadena, CA, USA 91101
- <sup>47</sup>Medical Research Unit in Biochemistry, Specialty Hospital, National Medical Center of the Twenty-First Century, Mexican Institute of Social Security, Mexico City, Mexico 06725
- <sup>48</sup>Postgraduate Program in Epidemiology, Federal University of Pelotas, Pelotas, Brazil 96010-610
- <sup>49</sup>Department of Human Genetics and South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville and Edinburg, TX, USA 78520 and 78539
- <sup>50</sup>Department of Medicine, School of Medicine, University of Texas Health San Antonio, San Antonio, TX, USA 78229
- <sup>51</sup>Department of Internal Medicine, Wake Forest University, Winston-Salem, NC, USA 27109
- <sup>52</sup>Department of Public Health Sciences, School of Medicine, and the Comprehensive Cancer Center, University of California Davis, Davis, CA, USA 95616
- <sup>53</sup>Departments of Epidemiology & Population Health and Medicine-Oncology, Stanford University School of Medicine, Stanford, CA, USA 94305
- <sup>54</sup>Division of Nutrition, Salvador Zubirán National Institute of Health Sciences and Nutrition, Mexico City, Mexico 14080
- <sup>55</sup>Programs in Metabolism and Medical and Population Genetics, Broad Institute of the Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA 02142
- <sup>56</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA 02115
- <sup>57</sup>Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA 02114
- <sup>58</sup>Laboratory of Immunogenomics and Metabolic Diseases, National Institute of Genomic Medicine, Mexico City, Mexico 14610
- <sup>59</sup>Center for Diabetes Studies, Research Unit for Diabetes and Cardiovascular Risk, Center for Population Health Studies, National Institute of Public Health, Mexico City, Mexico 14080
- <sup>60</sup>Molecular Biology and Medical Genomics Unity, Institute of Biomedical Research, the National Autonomous University of Mexico and the Salvador Zubirán National Institute of Health Sciences and Nutrition, Mexico City, Mexico 14080<sup>61</sup>Center for Community Health, Division of Academic General Pediatrics, University of California at San Diego, San Diego, CA, USA 92093
- <sup>62</sup>University of Texas Health Science Center at Houston, Houston, TX, USA 77030

<sup>63</sup>United States Department of Agriculture, Agricultural Research Service, The Children's Nutrition Research Center, and the Department Pediatrics, Baylor College of Medicine, Houston, TX, USA 77030

<sup>64</sup>Population Health Program, Texas Biomedical Research Institute, San Antonio, TX, USA 78227

<sup>65</sup>The Obesity Society, Silver Spring, MD, USA 20910

<sup>66</sup>Department of Nutrition, University of North Carolina Nutrition Research Institute, University of North Carolina, Kannapolis, NC, USA 28081

<sup>67</sup>Division of Sleep and Circadian Disorders, Brigham and Women's hospital, Boston MA, USA 02115

<sup>68</sup>Division of General Internal Medicine, Department of Medicine, Helen Diller Family Comprehensive Cancer Center, Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA 94115

<sup>69</sup>Ministry of Education Key Laboratory of Contemporary Anthropology and Collaborative Innovation Center of Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China 200438

<sup>70</sup>Department of Genetics, Evolution and Environment, and Genetics Institute of the University College London, London, UK WC1E 6BT

<sup>71</sup>Laboratory of Biocultural Anthropology, Law, Ethics, and Health, Aix-Marseille University, Marseille, France 13385

<sup>72</sup>Department of Anthropology, University of Toronto- Mississauga, Mississauga, ON, Canada L5L 1C6

<sup>73</sup>Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA 27514

\*Co-first authors.

Present Address for Correspondence:

Lindsay Fernandez-Rhodes, 219 Biobehavioral Health Building, University Park, PA 16802, [fernandez-rhodes@psu.edu](mailto:fernandez-rhodes@psu.edu)

Mariaelisa Graff, 123 West Franklin Street, Chapel Hill, NC 27516  
[migraff@email.unc.edu](mailto:migraff@email.unc.edu)



**ABSTRACT**

Hispanic/Latinos have been underrepresented in genome-wide association studies (GWAS) for anthropometric traits despite their notable anthropometric variability, ancestry proportions, and high burden of growth stunting and overweight/obesity. To address this knowledge gap, we analyzed densely-imputed genetic data in a sample of Hispanic/Latino adults to identify and fine-map genetic variants associated with body mass index (BMI), height, and BMI-adjusted waist-to-hip ratio (WHRadjBMI). We conducted a GWAS of 18 studies/consortia as part of the Hispanic/Latino Anthropometry (HISLA) Consortium (Stage 1, n=59,771) and generalized our findings in 9 additional studies (HISLA Stage 2, n=10,538). We conducted a trans-ancestral GWAS with summary statistics from HISLA Stage 1 and existing consortia of European and African ancestries. In our HISLA Stage 1+2 analyses, we discovered one BMI locus, as well as two BMI signals and another height signal each within established anthropometric loci. In our trans-ancestral meta-analysis, we discovered three BMI loci, one height locus, and one WHRadjBMI locus. We also identified three secondary signals for BMI, 28 for height, and two for WHRadjBMI in established loci. We show that 336 known BMI, 1,177 known height, and 143 known WHRadjBMI (combined) SNPs demonstrated suggestive transferability (nominal significance and effect estimate directional consistency) in Hispanic/Latino adults. Of these, 36 BMI, 124 height, and 11 WHRadjBMI SNPs were significant after trait-specific Bonferroni correction. Trans-ancestral meta-analysis of the three ancestries showed a small-to-moderate impact of uncorrected population stratification on the resulting effect size estimates. Our findings demonstrate that future studies may also benefit from leveraging diverse ancestries and differences in linkage disequilibrium patterns to discover novel loci and additional signals with less residual population stratification.

## INTRODUCTION

A complex interplay between political, social, and economic factors has led to an increasing obesogenic global environment in which many low- to middle- income nations have experienced a rapid transition from under-nutrition and growth stunting to over-nutrition and obesity.<sup>1</sup> In Latin America, by 2016, 35% of the total population was overweight [body mass index (BMI) 25 to <30 kg/m<sup>2</sup>] and another 23% was living with obesity (BMI≥30 kg/m<sup>2</sup>).<sup>2</sup> In Mexico, it is projected that by 2050 only 12% of men and 9% of women will have a healthy weight (BMI <25 kg/m<sup>2</sup>).<sup>3</sup> In South America in 2010-11, the prevalence of obesity was 36%, but abdominal obesity (based on waist circumference) was even more common (53%).<sup>4</sup>

Ancestry may also play a role in anthropometric-related health disparities in Hispanic/Latino populations. Previous studies have described the historical contexts leading to admixture in Latin American populations<sup>5; 6</sup> as characterized by highly diverse (variable) ancestral proportions<sup>7-9</sup> from any of the following regions: the Americas, Europe, Africa and East Asia.<sup>10-15</sup> Proportion of Native American ancestry is associated with obesity-related traits, and even more strongly associated with height.<sup>16; 17</sup> Height is inversely associated with proportion of Native American ancestry, even after taking into account that over time populations globally have become taller due to mainly non-genetic nutritional factors.<sup>16</sup> The ultimate drivers of this association remain unclear; it is possible that genetic factors and/or socio-economic factors strongly associated with Native American ancestry could be responsible. Recent studies are starting to provide relevant insights into this topic, including a recent genome-wide association study (GWAS) in Peru<sup>18</sup> that identified a missense variant in the *FBN1* gene (rs200342067) that has the largest effect size so far described for common height-associated variants in human populations. In the 1000 Genomes Project samples, rs200342067 is only present in two American samples (MXL: 0.78% and PEL: 4.12%), and yet the authors reported that this missense variant shows subtle evidence of positive selection in the Peruvian population.<sup>18</sup>

In the United States (US) like in other high-income nations, both the population size and diversity in national origins (backgrounds) of Hispanic/Latinos have been increasing over several decades,<sup>19</sup> with 24% of the US adult population identifying as Hispanic/Latino by 2065.<sup>19</sup> US Hispanic/Latino adults and children/adolescents face a greater burden of obesity than their non-Hispanic white counterparts.<sup>20-23</sup>

Thus, there is a need to study Hispanic/Latino populations to fully address these disparities.<sup>23; 24</sup> Specifically, we sought to understand the role that Native American or other under-studied components of admixture have on the genetic architecture of anthropometric traits in Hispanic/Latinos, and their relationship with risk of downstream poor health outcomes. To date, no large-scale GWAS of anthropometric traits have been conducted among Hispanic/Latino populations; we therefore performed a genomic study of anthropometric traits, including BMI, height, and waist-to-hip ratio adjusted for BMI (WHRadjBMI), in Hispanic/Latino populations to describe what may be novel loci, or new signals in established loci, for this population.

## **MATERIALS AND METHODS**

### ***Hispanic/Latino Study Samples***

The Hispanic/Latino Anthropometry (HISLA) Consortium is comprised of 27 studies/consortia of adult participants. First, HISLA Stage 1 includes 17 studies and one consortium (Consortium for the Analysis of the Diversity and Evolution of Latin America, CANDELA<sup>17</sup>) collectively representing up to 59,771 adults, depending on the trait, from Brazil, Chile, Colombia, Mexico, Peru, or the US with self-reported heritage from across Spanish-speaking Latin America, or Native American heritage, primarily Pima and Zuni<sup>25</sup> (**Table S1**). HISLA Stage 2 includes nine studies with up to 10,538 adults from across Spanish-speaking Latin America or with related heritage and living in the US (**Table S1**).

This study was approved by the Institutional Review Boards of the University of North Carolina at Chapel Hill, and all contributing studies had received prior Institutional Review Boards approval for each study's activities.

### ***Anthropometric Traits***

BMI is a commonly derived index of obesity risk and is calculated as the ratio of body weight to height squared ( $\text{kg}/\text{m}^2$ ). Adult height was measured or self-reported using either metric or US units and then converted to meters. Waist-to-hip ratio (WHR) is used to capture central fat deposition and is derived from the circumference of the waist at the umbilicus compared to the circumference of the hip at the maximum protrusion of the gluteal muscles.

Residuals were calculated by sex and/or case status, adjusting for age, age<sup>2</sup>, and study-specific covariates (e.g., center; Principal Components, PCs). For WHR, we also adjusted for BMI when creating the residuals to isolate the central deposition of fat from overall body mass. Residuals were then used to create inverse normalizations of BMI and  $\text{WHR}_{\text{adjBMI}}$ , and z-scores of height (=residual/standard deviation for all residuals). In family-based studies, the residuals were calculated in women and men together, adjusting for age, sex, and other study covariates including PCs. Descriptive statistics on the covariates and anthropometric measures are provided for each study's analytic sample in **Table S2**. Only one family-based study in Stage 1 and two non-family-based studies in Stage 2 (Genetics of Latinos Diabetic Retinopathy: 0.3% <18 years, and Mapping the Genes for Hypertension, Insulin Resistance, and Salt Sensitivity Study: 3.9%) included a small subset of adolescents aged 15-17 years, each less than 5% of the total sample. All other study samples included individuals 18-98 years of age.

### ***Childhood/Adolescence Study Samples, Anthropometric Traits, and Obesity***

We assembled an independent sample of children/adolescents with anthropometrics from three studies from the US, Mexico, and Chile (**Table S3**). The distribution of covariates and

anthropometrics of the samples of children/adolescents in each analysis are described in **Table S4**. First, childhood/adolescent obesity was defined as having a  $\geq 95^{\text{th}}$  BMI-for-age percentile versus  $\leq 50^{\text{th}}$  BMI-for-age percentiles, based on the Centers for Disease Control and Prevention growth curves,<sup>26</sup> as done in previous analyses of childhood obesity.<sup>27</sup> We used these two analyses to look up the BMI and height findings from our adult HISLA meta-analysis as well as our trans-ancestral analyses. This resulted in 1,814 children/adolescents aged 2-18 years in a case-control analysis of childhood obesity (**Tables S3-4**). Second, BMI and height-for-age z-scores were calculated in children/adolescents aged 5-18 years from the US and Chile (**Table S4**) based on the more international reference growth curves from the World Health Organization.<sup>28</sup> In Viva la Familia, a family-based study,<sup>29</sup> these residuals were calculated adjusting for sex in the combined sample. The resulting BMI and height-for-age z-scores were available for 1,914 and 1,945 children/adolescents, respectively.

### ***SNP Imputation and Statistical Analyses***

We generated autosomal genome-wide imputed data based on 1000 Genomes Phase 1 and 3 references, except for two studies that contributed Exomechip and MetaboChip (Illumina, Inc.; San Diego, CA) genotypes and one study that blended genotypes from multiple platforms (**Tables S5-6**). Principal component analyses (PCA) were conducted in each study (see select examples provided in **Figures S1-3**) to capture the main components of genetic ancestry from the Americas, Europe, Africa, and Asia. Studies with samples from related individuals accommodated this non-independence by projecting their PCA from the reference to the study sample, and by accounting for relatedness using either generalized estimating equations<sup>30</sup> or mixed linear models.<sup>9;31</sup> Assuming an additive genetic model, we tested for the association of over 20 million autosomal variants on our traits, accounting for all trait or study-specific covariates (e.g., center, PCs).

### **Meta-Analyses of HISLA Stage 1+2**

The studies of the HISLA Consortium were meta-analyzed in two stages, including discovery (Stage 1) and replication (Stage 2). Stage 1 included a total sample of 59,771 individuals with data on BMI, 56,161 with height, and 42,455 with WHRadjBMI. All Stage 1 studies/consortia provided full genome-wide analysis results. All SNPs that met our significance criteria were brought forward for replication in Stage 2, which included 10,538 individuals with data on BMI, 8,110 with height, and 4,393 with WHRadjBMI. All reported association results passed our quality control criteria; i.e., variants with low quality (info score <0.4 or Rsq<0.3), minor allele count (MAC) <5, or sample size <100 were removed. We meta-analyzed effects across all studies using a fixed-effect inverse variance weighted meta-analysis with genomic control in METAL.<sup>32</sup> Given the unique patterns of admixture and ancestry represented by the Brazilian or Native American samples, we conducted sensitivity analyses in Stage 1 studies (i.e., comparing the inclusion and exclusion of the Baependi Heart Study, 1982 Pelotas Birth Cohort Study, and Family Investigation of Nephropathy and Diabetes substudy of individuals of Pima and Zuni heritage) to assess the influence of the three studies on the meta-analysis results. CANDELA was retained in all analyses as <10% of the consortium's samples came from Brazil, primarily originating from the South of Brazil with widespread European heritage with less Native American or African admixture.<sup>17</sup> We provide the Quantile-Quantile plots for all analyses in **Figure S4**.

Regional plots of all GWAS-significant HISLA Stage 1 findings were plotted using LocusZoom (<https://locuszoom.org>). From Stage 1, we selected lead variants that met genome-wide significance ( $P < 5 \times 10^{-8}$ ) that were independent of each other for replication. In cases where Stage 2 studies did not have the lead variant, we selected two proxies per lead variant with a linkage disequilibrium (LD)  $r^2 \geq 0.9$  using 1000 Genomes AMR. Stage 2 studies provided a list of the requested lead variants and/or their proxies from Stage 1 for replication. Stage 2 studies were meta-analyzed and subsequently combined with Stage 1 using METAL<sup>25</sup>. Effect

heterogeneity was assessed through  $I^2$  across all 27 HISLA adult studies/consortia by entering each study separately into the meta-analysis, irrespective of stage. The characteristics of the final SNP array data used in the HISLA adult studies and the children/adolescent Hispanic/Latino studies are summarized separately in **Tables S5-6**.

### ***Meta-Analyses of HISLA Stage 1 with Other Ancestral Consortia***

In addition to a Hispanic/Latino only meta-analysis, we combined the HISLA Stage 1 meta-analysis with data from previous large-scale GWAS meta-analyses from European (the Genetic Investigation of Anthropometric Traits, GIANT, Consortium<sup>33-35</sup>, N ~ 300,000) and/or African (the African Ancestry Anthropometry Genetics Consortium, AAAGC<sup>36:37</sup>, N ~ 50,000) descent populations. We used fixed-effect inverse variance weighted meta-analytic techniques in METAL to generate our trans-ancestral meta-analysis.<sup>32</sup> We then assessed (a) the transferability of the findings from the BMI, height<sup>38</sup> and WHRadjBMI<sup>39</sup> trans-ancestral meta-analyses to an independent sample of Hispanic/Latino children/adolescents or (b) the replication of the signal in the British subsample GWAS of the United Kingdom Biobank (UKBB). Linkage disequilibrium plots and regional plots are shown in the supplement (**Figures S5-53**).

### ***Thresholds for Conditional Signals, Discovery, and Transferability***

We conducted approximate conditional analyses using Genome-wide Complex Trait Analysis (GCTA, version 1.93.1) software. For HISLA analyses, we used our Stage 1 discovery results with the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) as the LD reference dataset. For the approximate conditional trans-ancestral analyses, we used our trans-ancestral results from HISLA Stage 1, AAAGC, or GIANT and a trans-ancestral LD reference dataset of Europeans and African Americans from the Atherosclerosis Risk in Communities (ARIC) cohort, and Hispanic/Latinos from the HCHS/SOL cohort, as a representation of the ancestry distribution of our meta-analysis. In both conditional analyses (HISLA only and trans-

ancestral results), we first identified all independent SNPs using the `--cojo-slc` command. Then, we conditioned each of these independent SNPs on all known SNPs from GWAS (curated in the GWAS Catalog, <https://www.ebi.ac.uk/gwas/>) or reported as part of targeted-array analyses published through December 2019 (BMI<sup>33; 36; 38; 40-55</sup>, Height<sup>34; 38; 47; 50; 55; 56</sup>, WHRadjBMI<sup>35; 36; 39; 44; 46; 54; 55; 57-61</sup>) within 10Mb (+/-5Mb) of the lead SNP. The trans-ancestral meta-analysis results with a  $P\text{-value} < 5 \times 10^{-8}$  after conditioning on known SNPs were taken forward for replication in the British subsample of the UKBB.

In this paper we consider a signal as replicated when the effect of an allele is observed in two independent populations with the same or overlapping ancestral background, whereas generalization (transferability) refers to the observation of the same signal in an independent sample but with a distinct ancestral background, or distinct period of the life course. Furthermore, SNP associations were then defined as either newly-discovered or established, depending on their location. An established locus was defined as a SNP association within  $\pm 500$  kb of at least one previously identified index SNP, otherwise the association was considered a newly-discovered locus.

We designated our Hispanic/Latino SNP-associations within either newly-discovered or established loci as 'novel' if they met the following criteria for replication: 1) were associated at  $P\text{-value} < 5 \times 10^{-8}$  in HISLA Stage 1 and directionally consistent in Stage 2's independent sample, and 2) the addition of Stage 2 samples improved the estimated p-value of the Stage 1+2 meta-analysis. For the trans-ancestral analyses, the designation of a signal as 'novel' was based on SNPs that: 1) were associated at  $P\text{-value} < 5 \times 10^{-8}$  in the combined HISLA, AAAGC and GIANT meta-analysis, and 2) were both directionally consistent and associated at  $P\text{-value} < 5 \times 10^{-2}$  in an independent sample of Hispanic/Latino children/adolescents (generalized across age period) or in the British subsample GWAS from the UKBB (replication).

Hispanic/Latino SNP effects were considered to transfer (or generalize) to Hispanic/Latino children/adolescents or to African or European ancestry adults if they were: 1)



directionally consistent, 2) associated at  $P\text{-value} < 5 \times 10^{-2}$ , and 3) had a heterogeneity of  $I^2 < 75\%$  in either a) the Hispanic/Latino children/adolescent lookups, b) the adult AAAGC, or c) the adult GIANT GWAS lookups. SNP effects of variants previously associated with anthropometric traits in non-Hispanic/Latino populations (i.e., index published SNPs) were considered to be transferable (generalizable) to Hispanic/Latinos only if they were: 1) directionally consistent, 2) displayed a  $P\text{-value} < 5 \times 10^{-2}$ , and 3) had little to moderate effect heterogeneity ( $I^2 < 75\%$ ) in Stage 1.

### ***Fine-Mapping Methods***

We used FINEMAP<sup>62</sup> for analyses of the newly-discovered loci identified as part of the HISLA Stage 1 meta-analysis or trans-ancestral meta-analysis, for both established and novel loci. For the established loci, we included index SNP-associations published as of April 2018 (BMI<sup>33; 36; 40; 42-44; 46; 48; 51-54</sup>, Height<sup>34; 50; 56</sup>, WHRadjBMI<sup>35; 36; 44; 46; 63</sup>) prior to the publications with the UKBB results.<sup>38; 39</sup> We used a 1Mb region subset of the summary statistics from the Stage 1 meta-analyses and HCHS/SOL<sup>9</sup> unrelated sample set (N ~ 7,670) to calculate the LD for each locus. For consistency with the FINEMAP package, we refer to the results using similar language; however, we do not mean to imply that the SNP(s) in the 'causal set' are causative variants.

For trans-ancestral fine-mapping of the novel loci or new signals identified in the trans-ancestral meta-analysis of HISLA, AAAGC, and GIANT, we used a 1Mb region defining each locus using the summary statistics of the given meta-analysis. We calculated the LD for Hispanic/Latino samples using the HCHS/SOL<sup>9</sup> unrelated sample (N ~ 7,670). For African and European ancestry samples, we calculated the LD using the ARIC unrelated sample that included self-reported African ancestry (N ~ 2,800) and European ancestry (N ~ 9,700). We weighted the LD matrices by the GWAS sample sizes for each trait (HISLA range: ~42,400-56,100; AAAGC: 20,300-42,700; GIANT: 210,000-330,000).

All regions allowed up to a maximum of 10 causal variants, as defined by FINEMAP. The cumulative 95<sup>th</sup>% credible set was calculated from the estimated posterior probabilities. Convergence failed for three regions (lead SNPs: rs2902635, rs6900530, and rs4425978, all in known height loci) using the stochastic approach. For these three regions, we used the conditional approach to determine number of causal variants.

### ***Gene Expression & Other Bioinformatic Analyses***

We performed association analyses of measured whole blood gene expression in 606 individuals from the Cameron County Hispanic Cohort.<sup>64</sup> RNA sequencing was conducted using 150bp paired-end reads on the Illumina NovaSeq 6000 by Vanderbilt Technologies for Advanced Genomics. Initial sequencing quality was checked by FastQC.<sup>65</sup> STAR-2.7.8a was applied to align sequencing reads alignment to the human genome reference (UCSC, hg38),<sup>66</sup> and the aligned reads were assigned to genes using featureCounts.<sup>67</sup> We excluded either samples with less than 15M total aligned reads, a rate of successful alignment of less than 20%, or less than 15M total assigned reads. The sequencing library size was normalized using DESeq2<sup>68</sup> and read counts were transformed using variance stabilizing transformations (vst in DESeq2 package). We performed expression quantitative trait loci (eQTL) analysis with our top HISLA SNP findings, by modeling SNP dosages (exposure) in a linear regression of gene expression levels (outcomes), for each gene within the 1 MB interval around each lead SNP. We inverse normalized the gene expression levels and adjusted for age, sex, and three PCs to capture population substructure. Bonferroni correction for each region varied according to the number of SNPs tested.

To gain further insight into the possible functional role of the identified variants and to assess their relevance to other phenotypes, we conducted bioinformatic queries of our potentially novel loci and new signals within known loci in multiple publicly available databases, including PhenoScanner,<sup>69</sup> RegulomeDB,<sup>70</sup> Haploreg,<sup>71</sup> UCSC GenomeBrowser,<sup>72</sup> and GTEx.<sup>73</sup>

### ***Trans-Ancestral Findings to Account for Population Structure in Previous GWAS***

We demonstrated the degree to which the present trans-ancestral meta-analysis could lessen the bias induced by population stratification, using height from HISLA as an example. We first conducted PCA analysis on the four European populations (CEU, GBR, IBS, and TSI) from 1000 Genomes. We excluded the FIN (Finnish in Finland) population because of its known unique demographic history that could drive or dominate the top PCs in a limited sample.<sup>33</sup> We only used biallelic SNPs with minor allele frequency (MAF) > 5% in the four European populations, and then pruned them by both distance and LD using PLINK 1.9.<sup>74</sup> Specifically, we pruned the dataset such that no two SNPs were closer than 2 kb, and then pruned using a 50 SNP LD window (moving in steps of 5 SNPs), such that no SNPs had  $r^2 > 0.2$ . We further removed SNPs in regions of long-range LD.<sup>75</sup> PCA was performed on the remaining SNPs using Eigensoft version 7.2.1 (<https://github.com/DReichLab/EIG/archive/v7.2.1.tar.gz>).

We performed linear regressions of individual PC values on the allelic genotype count for each polymorphic variant in the four European populations from 1000 Genomes and used the resulting regression coefficients as the estimate of the variant's PC loading. For each PC, we then computed Pearson correlation coefficients of PC loadings and effect sizes (of variants with MAF > 1%) from each GWAS summary statistic. We estimated *P-values* based on Jackknife standard errors, by splitting the genome into 1,000 blocks with an equal number of variants. If the GWAS summary statistics are not biased by residual stratification (in this case due to European geographical structure), the correlation coefficients would be expected to be zero. If there was significant correlation in either the GIANT dataset or the HISLA Stage 1, AAAGC and GIANT trans-ancestral meta-analysis, we then further evaluated the improvement of bias due to stratification in trans-ancestral meta-analysis by comparing the correlation coefficients in the trans-ancestral meta-analysis with those in GIANT. Restricting to variants shared between GIANT and the trans-ancestral meta-analysis, we computed their difference in correlation

coefficients of PC loadings and effect sizes, and estimated *P-values* again based on Jackknife standard errors from 1,000 equal sized blocks.

## RESULTS

### ***Discovery of One BMI Locus in Hispanic/Latino Adults***

The first goal of this study was to conduct a genome-wide meta-analysis of anthropometric traits in Hispanic/Latino adults to identify loci in an under-studied population (**Figure 1**). All regional plots of all potentially-novel GWAS-significant HISLA Stage 1 findings are shown in the supplement (**Figures S6-11**).

No novel anthropometric loci were identified in all HISLA samples overall. Yet, when we excluded the samples of exclusively Brazilian or Native American heritage from Stage 1, we discovered one locus for adult BMI at *PAX3* on chromosome 2 in the HISLA Stage 1 sample (**Table S7**) and replicated this locus in HISLA Stage 2 (**Table 1**). The lead SNP at this locus, rs994108, is in moderate LD (rs7559271,  $r^2=0.46$  in 1000 Genomes AMR) (**Figure 2**) and lies on the same haplotype as a SNP reported to influence facial morphology, including position of the nasion (the deepest point on the nasal bridge where the nose meets the forehead) in Europeans<sup>76</sup> and Hispanic/Latino<sup>77</sup> descent individuals. Other *PAX3* variants in lower LD with the lead SNP have also been associated with nasion position,<sup>78</sup> monobrow, and male-pattern baldness.<sup>79; 80</sup> *PAX3* is a well-known transcription factor in normal embryonic neural crest development and differentiation.<sup>81</sup> Neural crest cells can give rise to mesenchymal stem cells,<sup>82</sup> which can in turn give rise to adipocytes;<sup>82-84</sup> thus, the possible role of *PAX3* in adipogenesis may at least partially explain the association signal with BMI near this gene.

Another BMI SNP (rs1505851-T) near *ARRDC3* on chromosome 5 associated at genome-wide significance in HISLA Stage 1 (**Table S7, Figure S6**), but did not replicate in Stage 2 (MAF=71%) or generalize to AAAGC or GIANT (effect allele frequency=31% or 68%;

**Table 1**). However, the association was directionally consistent and showed some signal in AAAGC ( $p=7 \times 10^{-4}$ ). LD patterns at *ARRDC3* appear to be similar across the three ancestries (albeit with a smaller LD block for 1000 Genomes AFR), meaning that lack of generalization may be related to frequency differences, haplotype effects, or a false positive (**Figure S5A**).

We identified two WHRadjBMI loci at *DOCK2* and *TAOK3* at genome-wide significance in HISLA Stage 1 after excluding the Brazilian and Native American samples (**Table S7**, **Figures S7-S8**), yet neither met the p-value threshold for replication in HISLA Stage 2. The *DOCK2* association for WHRadjBMI observed among women only in Stage 1 (**Figure S7**) was directionally consistent in the female Stage 2 sample. There were more SNPs in high LD with rs6879439 in 1000 Genomes AMR ( $0.8 \leq r^2 \leq 1$ ) than for AFR and EUR references (**Figure S5B**), which may explain why this SNP association did not generalize to AAAGC or GIANT (**Table 1**).

The genome-wide significant Stage 1 *TAOK3* association was led by a low frequency variant (rs115981023-A, MAF=0.9%), but the associations at this variant were not directionally consistent across Stages (**Table 1**; **Figure S8**). In fact, patterns of very high LD were seen with rs115981023 in 1000 Genomes AMR and AFR (**Figure S5C**), even though this variant is seen in African ancestry at a higher frequency (e.g. MAF=0.9% in HISLA versus 5% in AAAGC). Similarly, rs115981023 exhibited moderate heterogeneity across Stage 1 samples after excluding Brazilian and Native American samples ( $I^2=45\%$ ); evidence of moderate heterogeneity remained ( $I^2=52\%$ ) in the combined meta-analysis of HISLA Stage 1 and 2 samples (**Table 1**). Lastly, this variant was the least frequent in European ancestry (MAF=0.2% in GIANT), which explained the lack of a proxy for generalization in GIANT (**Table 1**).

No potentially novel loci were identified for height in HISLA Stage 1, and the exclusion of the Brazilian and Native American samples did not reveal additional height or WHRadjBMI loci.

### ***Discovery of Three Signals in Established Loci for BMI and Height in Hispanic/Latino***

#### ***Adults***

At two established loci for BMI, we identified additional signals at *ADCY5* and near *ILRUN* (**Table S7**). These signals were both independent of any previously published anthropometric findings (**Table S8, Figures S9-10**). We replicated these signals in Stage 2 with directional consistency and the combined Stage 1+2 meta-analysis at GWAS significance (**Table 1**). We also identified one additional signal for height in an established height locus, *B4GALNT3*, which was independent of the previously reported SNPs for height (**Tables S7-8, Figure S11**). We replicated this signal in Stage 2 with directional consistency and a Stage 1+2 meta-analysis that was GWAS significant (**Table 1**). In additional gene expression and bioinformatics analyses (**Tables S18-20**), we found that each of the three additional signals in established anthropometric loci is supported by either an eQTL in whole blood in Hispanic/Latino populations (**Table S18**), and/or an eQTL in other relevant tissues, e.g., thyroid, esophagus, artery, using publicly available (non-Hispanic/Latino) datasets (**Tables S19-20**).

### ***Fine-Mapping of Hispanic/Latino Anthropometric Findings***

We fine-mapped the *PAX3* locus for BMI and the three additional signals in known loci (BMI: *ADCY5* and *ILRUN*; height: *B4GALNT3*; **Table S9**). For the three BMI loci, FINEMAP revealed one potential causal set for each locus at *PAX3*, *ADCY5*, and *ILRUN* locus. For the *PAX3* locus, only one causal set was proposed, and the 95<sup>th</sup>% credible set contained only nine plausibly causal SNPs, with the lead SNP rs994108 having a very high posterior probability of being causal (0.89, **Table S21**). However, functional annotation of this SNP was unremarkable (**Tables S22-23**). In contrast, for *ADCY5* and *ILRUN*, FINEMAP revealed one causal configuration for each locus but with much greater uncertainty of the likely functional variant given the size of the credible sets, which contained 14 and 22 SNPs in the credible region for *ADCY5* and *ILRUN*, respectively. The posterior probability of the best lead SNP at these loci had relatively low posterior probabilities of being the causal SNP, with the best posterior probabilities of 0.23 for rs17361324 (*ADCY5*), and 0.11 for rs73420913 (*ILRUN*), respectively.

Interestingly, however, the best candidate for causality at *PAX3* and *ADCY5* loci were the lead SNPs from the HISLA meta-analysis; for *ILRUN*, the FINEMAP and HISLA SNPs were in high LD (rs73420913 had an  $r^2=0.96$  with the lead HISLA SNP rs148899910), providing greater support for the prioritization of these SNPs for functional interrogation. For the *B4GALNT3* height locus, FINEMAP revealed six causal configurations. Four of the variants (rs11063185, rs215230, rs7303572, and rs11063184) with each configuration had a posterior probability  $>0.99$  and contained only the variant itself in the 95<sup>th</sup>% credible set. One variant (rs215223) had a posterior probability of 0.93 and thus included two variants in the 95<sup>th</sup>% credible set. The sixth 95<sup>th</sup>% credible set had a lead variant with a posterior probability of 45% but contained a total of 1621 additional variants, all of which had very small posterior probabilities (i.e.,  $\leq 0.05$ ).

### ***Transferability of Adult Loci/Signals from Hispanic/Latinos to Consortia of Other Ancestral Backgrounds***

To assess how well the effect estimates are transferable to other populations, we looked up the BMI and height findings from Hispanic/Latinos in the AAAGC and GIANT meta-analysis results (**Table 1**). The BMI signal at the *ADCY5* locus (rs17361324) transferred to both AAAGC and GIANT with directional consistency (betas=0.13-0.23) and at nominal significance ( $P$ -values  $< 5 \times 10^{-2}$ ). The lead SNP (rs148899910) representing the BMI signal near *ILRUN* was not available in GIANT; the signal only appeared to be transferable to GIANT (at proxy SNP rs1573905,  $r^2=0.96-1$  in 1000 Genomes AMR and EUR; **Table 1**). The signal for height in *B4GALNT3* (rs215226) was directionally consistent and nominally significant in AAAGC only. In all cases, the effect sizes observed in GIANT and AAAGC were attenuated compared to the effect sizes from HISLA Stage 1.

### ***Relevance of Adult Hispanic/Latino Anthropometric Findings to Childhood/Adolescence***

We looked up our novel HISLA findings in Hispanic/Latino children/adolescents using BMI-for-age and height-for-age z-scores, as well as a case-control study of childhood obesity. Two of the three novel BMI signals were directionally consistent with the anticipated effect on the odds of obesity during childhood/adolescence, one of which was nominally significant (rs17361324 at *ADCY5*;  $P\text{-value}=2.2\times 10^{-2}$ ). None of the HISLA findings generalized at nominal significance with the BMI/height-for-age z-score, but were directionally consistent with the corresponding effect in adulthood (**Table S10**). This may have been due to the small available sample size of Hispanic/Latino children/adolescents.

### ***Transferability of Established Anthropometric Loci to Hispanic/Latino Adults***

We assessed how many established anthropometric loci, described previously in predominantly non-Hispanic/Latino European samples, could be transferred to Hispanic/Latino adults, in light of the available Hispanic/Latino sample size from Stage 1. As shown in **Table S11**, the index SNPs at 336 of 1247 (26.9%) previously reported BMI loci were suggestively transferable at nominal significance to Hispanic/Latinos. Of these BMI loci, 36 SNPs in the HISLA Stage 1 displayed directional consistency with the literature and Bonferroni-significance (**Table S11**). Furthermore, one BMI locus was genome-wide significant at the same published variant and another 12 BMI loci were genome-wide significant at another SNP within 1 MB and in moderate to high LD ( $r^2\geq 0.52$  in AMR) with the reported index SNP (**Table S7**).

**Table S12** shows that a slightly higher percentage of known height loci (1177 of 3806, or 30.9%) were transferable to Hispanic/Latinos. Of these loci, 124 SNPs were directionally consistent and Bonferroni significant (**Table S12**). Ten height loci were genome-wide significant at the same lead variant, and another 39 height loci were associated at genome-wide significance at another SNP within 1 MB ( $0.05\leq r^2\leq 0.98$  in AMR, **Table S7**).

Lastly, **Tables S13-15** show that 143 of 694 (20.6%) known WHRadjBMI in both sexes combined, 133 of 567 (23.5%) in women only, and 28 of 173 (16.2%) in men only loci were



transferable to Hispanic/Latinos at nominal significance. Of these, a total of 15 loci were associated with WHRadjBMI at Bonferroni significance in the combined, women only, or men only analyses (**Tables S13-15**). None of the index SNPs from the previous literature for WHRadjBMI reached genome-wide significance; however, we did observe genome-wide significant evidence for association of a SNP with WHRadjBMI in strong LD with the index variant ( $r^2=0.92$  in AMR) for the *HOXC13* signal (**Table S7**).

### ***Replication of Five Novel Loci and Thirty-Three New Signals in Established Loci for Adult Anthropometric Traits from a Trans-Ancestral Meta-Analysis***

Our secondary goal was to assemble a trans-ancestral meta-analysis of HISLA Stage 1 with the AAAGC and GIANT consortia results to identify additional novel loci and fine-map established loci by leveraging differences in allele frequencies across populations (**Figure 1**). As anticipated, this trans-ancestral meta-analysis of HISLA, AAAGC and GIANT revealed new insights, including eight novel loci and 35 new signals in established loci that were associated at genome-wide significance (**Table S16, Figures S12-S53**), and independent of established SNPs within a 10Mb region (**Table 2**). Of this set, five loci (3 BMI, 1 height, and 1 WHRadjBMI) and 33 signals in established loci (3 BMI, 28 Height, and 2 WHRadjBMI) were generalized using the adult British subsample of the UKBB. In some cases, the significance in the trans-ancestral results were driven more by the AAAGC and/or HISLA consortia, which could explain the lack of association in the UKBB British subsample (**Table S16, Figure S54**).

We looked up the findings from our trans-ancestral meta-analyses in the sample of Hispanic/Latino children/adolescents (**Table S17**). We found that two of the seven BMI and height trans-ancestral loci, and 17 of the 33 trans-ancestral BMI/height signals in established loci, were directionally consistent between their adult directions of association and the BMI/height-for-age z-scores in children/adolescents. However, this directional consistency was not more than what would have been expected by chance alone ( $P\text{-values}_{\text{binomial}} > 0.10$ ). Four

trans-ancestral SNPs were associated at nominal significance in the child/adolescent sample, each having been already replicated in UKBB (**Table S16**). Three of these four loci were directionally consistent in the childhood/adolescence results with the trans-ancestral adult findings (**Table S17**).

### ***Fine-Mapping of Trans-Ancestral Anthropometric Findings***

We also fine-mapped our trans-ancestral findings (**Table S21**) using FINEMAP to pinpoint individual variants and genes within each locus region that have a direct effect on the trait. FINEMAP uses a shotgun stochastic search algorithm<sup>85</sup> that iterates through causal configurations of SNPs by concentrating efforts on the configurations with non-negligible probability. Within a 1MB region, we report a) the causal configuration of SNPs for a given trait that had the highest posterior probability and b) the posterior probability of being causal for each of the SNPs.

For four of the five trans-ancestral loci (three BMI loci and one WHRadjBMI locus), there was one SNP within the configuration with the highest posterior probability. For the height locus near *ANKRD36BP1*, there were two SNPs in the configuration with the highest posterior probability. In all five loci, the SNP with the highest posterior probability from each of these credible sets was either the exact SNP with the strongest GWAS evidence or in high LD ( $r^2$  between 0.70 and 0.99 in each ancestry) with the lead GWAS SNP. Two of these five regions had strong prioritization given high posterior probabilities ( $\geq 0.8$ ) and small 95<sup>th</sup>% credible sets: 1) for BMI, the *CHD1-DT* region had a posterior probability of 0.88 for rs150992 with three SNPs in the credible set, and 2) for height, the *ANKRD36BP1* region had a posterior probability of 0.93 for rs10737541 with five SNPs in the credible set. From the functional annotations (**Tables S22-23**), we find that all three of the BMI loci, the height loci, and WHRadjBMI loci have enhancer marks and eQTLs, most of which are in highly relevant tissues, e.g., adipose, brain, muscle, thyroid.

For the other trans-ancestral loci, the posterior probabilities were lower, between 0.09 and 0.42, yet four loci (rs9860730, rs17375290, rs4324883, and rs9463108) still had relatively few SNPs (<10) in the 95<sup>th</sup>% credible sets, suggesting a narrow window (combination of variants) around the causal variant. For example, functional annotations of rs17375290, the lead GWAS SNP in the NFIA locus associated with height, show it to have promoter markers in muscle, CADD score of 13.29 (CADD > 10 ranks variants among the top 10% potentially deleterious), and an eQTL with *FGGY* in osteoclast tissue (**Tables S22-23**). Three of the other SNPs in the credible set (rs599989, rs1762881, and rs17121184) have nominally significant (p-value 0.01 to 0.005) eQTLs with *FGGY* in osteoclast tissue but are not in high LD with rs17375290 ( $r^2$  range from 0.03 to 0.1). Diseases associated with *FGGY* include autosomal recessive lateral sclerosis and spastic paraplegia 7, which are known to affect height.

Within the 33 trans-ancestral signals in known loci, 31 had configurations with more than one putative causal SNP (e.g. more than one credible set). This made sense given these are loci with multiple independent signals, as described by our earlier conditional analyses. Among the putative causal SNPs within each locus, there were a number of SNPs that represented known signals (either the exact SNP or something in high LD among all ancestries). We found that for many of these, the credible sets contained <10 SNPs. Among the 33 signals in known loci, 26 included a putative causal SNP that is the lead GWAS SNP reported here or a SNP in high LD ( $r^2 > 0.75$ ) with the lead GWAS SNP, suggesting causality for this signal in general, though perhaps maybe not initially discovered at the most-putatively-causal SNP(s). For these 24 putatively-causal SNPs, the posterior probabilities ranged from 0.09 to 1. Twenty-two of these SNPs had 95<sup>th</sup>% credible sets that contained <10 SNPs and 15 also had posterior probability  $\geq 0.8$ .

Many have functional annotations that support the fine-mapping results (**Tables S22-23**). For example, we find eQTLs for the three BMI signals and enhancer marks for rs4807179 in relevant tissues including adipose, brain, muscle, and/or thyroid. The lead SNPs of these

credible sets had posterior probabilities  $>0.75$  and the credible sets included  $<10$  SNPs. Of the 28 identified height signals, we find 13 putatively-causal SNPs that are the lead GWAS SNP or are in high LD ( $r^2 > 0.75$ ) with it, have  $<10$  SNPs in the credible set and have eQTLs in relevant tissues including muscle, thyroid, adipose, lung and osteoclasts. Some also have promoter or enhancer marks in some of the same tissues. For the two WHRadjBMI signals, both have three SNPs in the most probably causal configurations. One of these causal SNPs for each region is either the lead GWAS SNP (rs7975017) or a SNP in high LD (rs17099388 and rs6895040 LD: AFR  $R^2=1.0$ ; AMR  $R^2=1.0$ ; EUR  $R^2=1.0$ ), has a posterior probability  $\geq 0.95$ , and is the only SNP in the credible set. Furthermore, for rs7975017, we find eQTLs in thyroid for multiple genes (*BHLHE41*, *SSPN*, and *AC022509.3 from GTEX*) and enhancer marks in multiple tissues including those related to the WHRadjBMI trait, e.g., thyroid, muscle, fat, bone, and adrenal gland. Overall, across many of the loci and secondary signals, FINEMAP revealed SNPs with somewhat strong prioritization (posterior probability  $\geq 0.8$ ), and at some loci, putatively causal SNPs have small 95<sup>th</sup>% credible sets, thus demonstrating the utility of trans-ancestral approaches to fine mapping GWAS loci.

### ***Trans-Ancestral Findings to Account for Population Structure in Previous GWAS***

Previous height GWAS utilizing only European-ancestry samples are known to exhibit signatures of residual stratification, which manifest in effect size estimates of height-associated SNPs being correlated with geographical structure in Europe.<sup>86-89</sup> In theory, this bias should be lessened with addition of non-European samples in a trans-ancestry GWAS, since geographical structure across different continental ancestries are not expected to be correlated with each other. We demonstrate this hypothesis empirically using the HISLA data. The first two PCs in the PCA of European populations (**Figure S55**) reflect geographical or population structure in Europe, corresponding to the North-South and Southeast-Southwest axes of variation, respectively. We found that the bias in effect size estimates due to stratification is most obvious

for height as this phenotype is known to differ across Europe.<sup>86; 90; 91</sup> Effect sizes on height estimated from the GIANT and our trans-ancestral meta-analysis were both highly correlated with the loadings of the first PCA ( $\rho = 0.125$ ,  $P\text{-value} = 3.2 \times 10^{-94}$  in GIANT;  $\rho = 0.105$ ,  $P\text{-value} = 3.4 \times 10^{-70}$  in meta-analysis). The correlation was much lower in AAAGC and HISLA ( $\rho = 0.012$ ,  $P\text{-value} = 2.17 \times 10^{-4}$  in AAAGC;  $\rho = 0.007$ ,  $P\text{-value} = 9.2 \times 10^{-2}$  in HISLA; **Figure 4A**). Importantly, the magnitude of correlation was lessened in meta-analysis compared with GIANT ( $P\text{-value} = 6.6 \times 10^{-9}$ ), consistent with our hypothesis. Other traits were not *a priori* known to be as differentiated across Europe as height, and thus the degree of correlation between effect sizes and PC loadings are much lower in GIANT (e.g.  $\rho = -0.025$  for BMI; **Figure 4B-E**).

## DISCUSSION

Hispanic/Latinos are a unique population with continental admixture from the Americas, Africa, and Europe,<sup>10-14</sup> and yet are underrepresented in GWAS. Herein, we present results from a large-scale meta-analysis of anthropometric traits on an ancestrally diverse sample of Hispanic/Latino adults (**Figures S1-3**). We have assembled a landmark consortium of Hispanics/Latinos to discover and map a total of six novel loci and 36 novel signals using both Hispanic/Latino population-specific and trans-ancestral discovery efforts (**Figure 1**). Numerous previously-reported anthropometric-SNP associations were suggestively (at nominal significance) or strongly (at Bonferroni significance) transferable to Hispanic/Latino adults. For example, between 16-31% of anthropometric variants transferred to Hispanic/Latino adults, depending on the given trait or sex-specific analyses conducted (**Tables S11-13**). In total, 67 previously reported loci reached genome-wide significance in our Hispanic/Latino adult sample at the same index or another lead SNP, the majority of which were in LD in 1000 Genomes EUR or AMR (**Table S7**). Moreover, we observed that four of seven of our HISLA findings were transferable to other ancestral populations at nominal significance.

We note that even though these findings provide additional evidence for transferability of common loci for anthropometrics,<sup>92</sup> still a number of previously-reported anthropometric loci may not be transferable to this population in part due to variability in allele frequencies, effect sizes across ancestral populations, or our relatively smaller sample compared to European consortia.<sup>55</sup> Thus, absence of generalization does not equate to a lack of relevance to Hispanic/Latino adults or children, especially given that Hispanic/Latinos are under-studied population in genetic research and larger/comparable sample sizes are currently unavailable.

Our conditional and fine-mapping analyses revealed 36 signals in established anthropometric loci, which independently replicated in HISLA Stage 2 or the UKBB British subsample. In addition, our lead SNPs for the BMI signals discovered at *ADCY5* (from the HISLA meta-analysis) and *ADAMTS9-AS2* (from the trans-ancestral meta-analysis) are both nominally associated with childhood obesity status between 2-18 years of age. Three of our trans-ancestral signals in established height loci also displayed association with height-for-age z-scores in children/adolescents between 5-18 years of age. These observations support the premise that diverse and trans-ancestral studies represent a valuable tool for leveraging ancestral differences and similarities both within and across populations to identify multiple signals in established association regions, identify putative variants that may account for some of the missing heritability of complex diseases, or reveal promising genes and SNPs for functional follow-up.

In light of the notable ancestral, geographical or environmental diversity of the studies analyzed in our meta-analyses, we observed evidence of allele frequency differences for many of our Hispanic/Latino (**Figure 3**) and trans-ancestry findings (**Figure S53**). Similar to reports from other diverse genome-wide analyses,<sup>55</sup> this allele frequency heterogeneity may explain heterogeneity in effects seen across consortia in our trans-ancestral HISLA, AAAGC, and GIANT meta-analysis (e.g., *IGF2BP2*  $I^2=78.7$ ; *MYO6* with  $I^2=84.4$ , **Tables 2 and S16**). Our use of fixed-effect meta-analyses may have failed to identify loci with effect heterogeneity unrelated

to allele frequency or LD differences across populations; future studies should address this limitation by considering trans-ancestral random-effects meta-analysis, local ancestry and haplotype analyses as they explore for sources of heterogeneity in large, diverse datasets. These observations reinforce how studies of one predominant ancestry group, such as Europeans, may fail to identify additional loci or, more likely, new signals in known loci that have allele frequency differences across ancestral populations.

Residual uncorrected stratification in GWAS could result in biased estimates of effect sizes.<sup>34</sup> For example, effect sizes on height from GIANT were reported to be significantly correlated with North-South axis of variation in Europe, suggesting residual uncorrected stratification,<sup>86-88</sup> which we also observe here. Note that the residual stratification is subtle, and while the effect sizes may be biased, this does not imply the identified associations are spurious. For example, compared with effect sizes on height from UKBB, which is based on a single homogeneous population and results in better control of population stratification, the genetic correlation between GIANT and UKBB was 0.94.<sup>86</sup>

Of the three traits studied here, height is the most stratified in Europe. The correlation coefficient between effect sizes on height and PC loadings reached 0.125 in the GIANT only for PC1, while it was much smaller for other traits (e.g., the maximum  $|\rho| = 0.042$  in GIANT on WHR using only males on PC1). The decrease in bias in the trans-ancestral meta-analysis was also obvious in height. The correlation with PC1 was non-significant in HISLA ( $\rho = 0.007$ ) and statistically significant but weak in AAAGC ( $\rho = 0.012$ ), consistent with a decreased impact of European population stratification on the estimate of effect size in AAAGC and HISLA. This decreased correlation could be due to large non-European ancestries known in these populations (African and Native American, respectively) that are less affected by population stratification in Europe; it could also be that by using European ancestry-based loadings, we are less likely to detect non-European based population stratification patterns or that smaller sample sizes in these cohorts result in greater noise in effect size estimates. Regardless of the reason, compared to

GIANT alone, trans-ancestral meta-analysis of the three cohorts showed less impact of uncorrected stratification on effect size estimates, even though the sample size in AAAGC and HISLA are comparably small. For other traits, the conclusions are qualitatively similar: that trans-ancestral meta-analysis lessened the bias due to stratification, even though the bias in GIANT was not as strong in the first place.

Gene expression and bioinformatic analyses of our population-specific (**Table S18-20**) and trans-ancestral findings in newly discovered loci (**Tables S22-23**) revealed important insights into the underlying biology of obesity, bone development and growth. For example, the previously reported BMI locus *ILRUN* has also been associated with adult height<sup>93; 94</sup> and height change during puberty.<sup>95</sup> The previously-describe BMI signal at *ILRUN* was at index SNP, rs205262, an eQTL for another gene within the region (*SNRPC*) in European ancestry samples.<sup>33</sup> A second signal (rs75398113) has also been reported at *SNRPC* for extremes of the body mass index distribution.<sup>96</sup> Yet, our signal led by rs148899910 is more than 300kb away and in low LD with these two index SNPs ( $r^2=0.01-0.05$  in 1000 Genomes AMR). More recently, rs148899910 has been associated with height in Korean women.<sup>97</sup> Furthermore, variants in high LD with rs148899910 in 1000 Genomes AMR are associated with Type 2 diabetes in individuals of East Asian ancestry<sup>98</sup> (rs4711389 has  $r^2=0.9$  in 1000 Genomes AMR with rs148899910), and with BMI-adjusted waist circumference in individuals of European ancestry (rs202228093 and rs2780226 each have an  $r^2>0.7$  with rs148899910 in 1000 Genomes AMR).<sup>44;99</sup> Using whole blood gene expression data from 606 participants of the Cameron County Hispanic Cohort, we find evidence that our BMI signal at rs148899910 is an eQTL for increased gene expression of *C6orf1* ( $p\text{-value}=3\times 10^{-7}$ ) and not any other genes in the region (**Table S18**). Taken together, this signal shows associations across a wide array of anthropometric phenotypes.

In general, the lead SNPs from our HISLA only meta-analyses appear relatively benign (not pathogenic) based on CADD and FATHMM-XF scores (**Table S20**). Yet, all SNPs potentially change motifs. Both rs17361324 (*ADCY5*) and rs215226 (*B4GALNT3*) have



enhancer and promoter histone marks and eQTLs in the respective genes in relevant tissues. For BMI, there is an eQTL for rs17361324-*ADCY5* in thyroid, and *ADCY5* has been previously associated with type 2 diabetes,<sup>100</sup> BMI,<sup>101</sup> central obesity traits,<sup>39</sup> height,<sup>47</sup> birth outcomes,<sup>102-104</sup> and a number of other phenotypes. Additionally, rs17361324 is proximate to an *ADCY5* intronic variant (rs1093467,  $r^2=0.3$  in 1000 Genomes AMR) that is highly conserved across species (Haploreg v4.1). For height, there is an eQTL for rs215226-*B4GALNT3* in aortic (coronary), and tibial nerve. The lead SNP for the height signal in *B4GALNT3*, rs215226, has enhancer histone marks in bone and muscle, and promoter marks in muscle tissue. In addition, the variant rs215226 (*B4GALNT3*) has a posterior probability of 1 in FINEMAP analyses (see **Table S9**). Other interesting information about these regions is provided in **Table S19**.

The lead SNPs in our trans-ancestral loci were mainly located in intronic and intergenic regions (**Table S22**) and were benign. One exception was the locus *C11orf63* associated with height led by rs11605693, which showed pathogenic scores for CADD and FATHMM-XF (CADD score=17.1 and FATHMM-XF score=0.87). This lead SNP has an eQTL in *C11orf63* for adipose, tibial nerve, and testis. *C11orf63*, junctional cadherin complex regulator, is responsible for ependymal cells that line the brain and spinal cord.

Among the trans-ancestral findings, a BMI signal in the established locus *RNH1* was led by rs10540 had a posterior probability of 0.82, and is an eQTL for a wide range of tissues and genes (see **Tables S21 and S23**), potentially making it relevant to body mass. Another signal in a known locus for height, led by rs12918773, has a posterior probability of 0.98 and is one of four casual variants suggested from fine-mapping in the locus (**Table S21**), has an eQTL (in lung, thyroid, tibial nerve and artery, breast, testis) with *CDK10*, a gene also associated with growth retardation.<sup>105</sup> In addition, rs1342330 led the newly-discovered signal in a known height locus, has a low regulomeDB score at 2b, and several enhancer and promoter histone marks in relevant tissues (**Tables S22**). As an intronic variant, it is an eQTL in the pancreas with *PHACTR2* (**Table S23**), a gene associated with body dysmorphic disorder.<sup>106</sup> While many of the

our discovered loci/signals appeared to be benign based on CADD and FATHMM-XF scores, they still show enhancer and promoter histone marks in trait relevant tissues such as adipose, bone, and muscle, thymus, brain, and adrenal gland.

As described above, in this study we were able to 1) discover six additional loci with a notably smaller analytic size than other anthropometric consortia like GIANT. We also 2) discovered 36 signals in established loci in HISLA or our trans-ancestral meta-analysis, and 3) generated trans-ancestral effect estimates with better control for population structure. Taken together, these findings indicate the added value of building large, more diverse GWAS in the near future.

Large-scale analyses of diverse populations hold great potential for advancing the field of genetic epidemiology.<sup>55</sup> This study illustrates how studying admixed populations, like Hispanics/Latinos, and highlighting them in trans-ancestral epidemiologic investigations can yield additional insights into the genetic architecture of anthropometric traits. Future discovery efforts in Hispanic/Latino populations and with other ancestrally diverse populations will help address the concerning research gap between who is studied and who is affected by conditions like obesity, to the benefit of both public health and precision medicine.

## Supplemental Data

Supplemental Data include 23 tables and 59 figures.

## Declaration of Interests

SMG and AMS receive funding from Seven Bridges Genomics to develop tools for the NHLBI BioData Catalyst consortium. All other authors declare no competing interests.

## Data and Code Availability Statements

The HISLA meta-analysis results [GWAS Catalog Access IDs: Height- GCST90095033, BMI- GCST90095034, WHRadjBMI (sex combined)- GCST90095035, WHRadjBMI (men)- GCST90095036, WHRadjBMI (women)-GCST90095037] and the trans-ancestral HISLA, AAAGC, and GIANT meta-analysis results [GWAS Catalog Access ID: Height- GCST90095038, BMI- GCST90095039, WHRadjBMI (sex combined)- GCST90095040, WHRadjBMI (men)- GCST90095041, WHRadjBMI (women)- GCST90095042] are available through the NHGRI-EBI Catalog (<https://www.ebi.ac.uk/gwas/>).

## Acknowledgements

AGL was supported by T32 HD091058, P2C HD050924, and P30 AG066615. AGT was supported by T32HL007055. ARL is supported by the Leverhulme Trust (F/07 134/DF), BBSRC (BB/I021213/1), the Excellence Initiative of Aix-Marseille University-A\*MIDEX, the National Natural Science Foundation of China (#31771393), Universidad de Antioquia (CODI sostenibilidad de grupos 2013- 2014 and MASO 2013-2014). KA was supported by 0044 (0)20 3108 4003. LF was supported by R01CA204797. LF-R was supported by an American Heart Association (AHA) predoctoral grant (13PRE16100015). MG, KLY and KEN were supported by AHA (13GRNT16490017, 15GRNT25880008), R01DK089256 and R01DK101855. LF-R, CAH, and RFJL were supported by R01DK101855. XRG was supported by R01EY022651. KEN was

supported by R01HD057194, R01DK122503, R01HG010297, R01HL142302, R01HL143885, and R01HG009974. LG was supported by T32 HL129982. QQ was supported by R01HL060712, R01HL140976, and R01DK119268. SFAG was supported by the Daniel B. Burke Endowed Chair for Diabetes Research and R01 HD056465. XG, MAA, Y-DIC, JY, and JIR were supported by EY14684, HL-0767711, HL-0697974, HL-088457, NEI EY11753, UL1-TR-001881. XL and KR were supported by R01 HL0767711 and DK-079888. ZA, TAB, JT, and AHX were supported by HTN-IR funding (HL-0697974); PMG, YH, EI, and KDT were supported by EY-14684. MOG, WH, KL, and KS were supported by HL-088457. AEJ was supported by K99/R00 HL130580. EMJ was supported by R01 CA063446, R01 CA077305, DOD RP9590546, CBCRP 7PB-0068. JM was supported by the American Diabetes Association 1-19-ICTS-068 and by U01HG011723. Study-specific acknowledgements are available as **Supplemental Data**.

## References

1. World Health Organization. (2020). Obesity and overweight, <https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight>.
2. Organización de las Naciones Unidas para la Alimentación y la Agricultura (2017). América Latina y el Caribe: Panorama de la seguridad alimentaria y nutricional. Sistemas alimentarios sostenibles para poner fin al hambre y la malnutrición, 2016, <http://iris.paho.org/xmlui/handle/123456789/33680>.
3. Mueller, M., Purnell, T.S., Mensah, G.A., and Cooper, L.A. (2015). Reducing racial and ethnic disparities in hypertension prevention and control: what will it take to translate research into practice and policy? *Am J Hypertens* 28, 699-716.
4. Lanas, F., Bazzano, L., Rubinstein, A., Calandrelli, M., Chen, C.S., Elorriaga, N., Gutierrez, L., Manfredi, J.A., Seron, P., Mores, N., et al. (2016). Prevalence, Distributions and Determinants of Obesity and Central Obesity in the Southern Cone of America. *PLoS One* 11, e0163727.
5. Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J.K., Muzzio, M., Rodriguez-Flores, J.L., Kenny, E.E., Gignoux, C.R., Maples, B.K., Guiblet, W., et al. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet* 9, e1004023.
6. Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J.L., Byrnes, J.K., Gignoux, C.R., Ortiz-Tello, P.A., Martinez, R.J., Hedges, D.J., Morris, R.W., et al. (2013). Reconstructing the population genetic history of the Caribbean. *PLoS Genet* 9, e1003925.
7. Gonzalez Burchard, E., Borrell, L.N., Choudhry, S., Naqvi, M., Tsai, H.J., Rodriguez-Santana, J.R., Chapela, R., Rogers, S.D., Mei, R., Rodriguez-Cintron, W., et al. (2005). Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *Am J Public Health* 95, 2161-2168.

8. Moreno-Estrada, A., Gignoux, C.R., Fernandez-Lopez, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuna-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 344, 1280-1285.
9. Conomos, M.P., Laurie, C.A., Stilp, A.M., Gogarten, S.M., McHugh, C.P., Nelson, S.C., Sofer, T., Fernandez-Rhodes, L., Justice, A.E., Graff, M., et al. (2016). Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet* 98, 165-184.
10. Bryc, K., Velez, C., Karafet, T., Moreno-Estrada, A., Reynolds, A., Auton, A., Hammer, M., Bustamante, C.D., and Ostrer, H. (2010). Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A* 107 Suppl 2, 8954-8961.
11. Halder, I., Yang, B.Z., Kranzler, H.R., Stein, M.B., Shriver, M.D., and Gelernter, J. (2009). Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mutat* 30, 1299-1309.
12. Shtir, C.J., Marjoram, P., Azen, S., Conti, D.V., Le Marchand, L., Haiman, C.A., and Varma, R. (2009). Variation in genetic admixture and population structure among Latinos: the Los Angeles Latino eye study (LALES). *BMC Genet* 10, 71.
13. Wang, S., Ray, N., Rojas, W., Parra, M.V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A.M., et al. (2008). Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 4, e1000037.
14. Manichaikul, A., Palmas, W., Rodriguez, C.J., Peralta, C.A., Divers, J., Guo, X., Chen, W.M., Wong, Q., Williams, K., Kerr, K.F., et al. (2012). Population structure of Hispanics in the United States: the multi-ethnic study of atherosclerosis. *PLoS Genet* 8, e1002640.
15. Chiang, C.W., Gajdos, Z.K., Korn, J.M., Kuruvilla, F.G., Butler, J.L., Hackett, R., Guiducci, C., Nguyen, T.T., Wilks, R., Forrester, T., et al. (2010). Rapid assessment of genetic

- ancestry in populations of unknown origin by genome-wide genotyping of pooled samples. *PLoS Genet* 6, e1000866.
16. Spear, M.L., Diaz-Papkovich, A., Ziv, E., Yracheta, J.M., Gravel, S., Torgerson, D.G., and Hernandez, R.D. (2020). Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. *Elife* 9.
  17. Ruiz-Linares, A., Adhikari, K., Acuna-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., et al. (2014). Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet* 10, e1004572.
  18. Asgari, S., Luo, Y., Akbari, A., Belbin, G.M., Li, X., Harris, D.N., Selig, M., Bartell, E., Calderon, R., Slowikowski, K., et al. (2020). A positively selected *FBN1* missense variant reduces height in Peruvian individuals. *Nature* 582, 234-239.
  19. Pew Research Center (2015). Modern Immigration Wave Brings 59 Million to U.S., Driving Population Growth and Change Through 2065, <http://www.pewhispanic.org/2015/09/28/modern-immigration-wave-brings-59-million-to-u-s-driving-population-growth-and-change-through-2065/>.
  20. Ogden, C.L., Carroll, M.D., Kit, B.K., and Flegal, K.M. (2012). Prevalence of Obesity and Trends in Body Mass Index Among US Children and Adolescents, 1999-2010. *JAMA*.
  21. Flegal, K.M., Carroll, M.D., Kit, B.K., and Ogden, C.L. (2012). Prevalence of Obesity and Trends in the Distribution of Body Mass Index Among US Adults, 1999-2010. *JAMA*.
  22. Ogden, C.L., Carroll, M.D., Kit, B.K., and Flegal, K.M. (2014). Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama* 311, 806-814.
  23. Isasi, C.R., Ayala, G.X., Sotres-Alvarez, D., Madanat, H., Penedo, F., Loria, C.M., Elder, J.P., Daviglius, M.L., Barnhart, J., Siega-Riz, A.M., et al. (2015). Is acculturation related to obesity in Hispanic/Latino adults? Results from the Hispanic community health study/study of Latinos. *J Obes* 2015, 186276.

24. NCD Risk Factor Collaboration. (2020). Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: a pooled analysis of 2181 population-based studies with 65 million participants. *Lancet* 396, 1511-1524.
25. Knowler, W.C., Coresh, J., Elston, R.C., Freedman, B.I., Iyengar, S.K., Kimmel, P.L., Olson, J.M., Plaetke, R., Sedor, J.R., Seldin, M.F., et al. (2005). The Family Investigation of Nephropathy and Diabetes (FIND): design and methods. *J Diabetes Complications* 19, 1-9.
26. Centers for Disease Control and Prevention. (2016). CDC Growth Charts. In. (
27. Bradfield, J.P., Voegelezang, S., Felix, J.F., Chesi, A., Helgeland, O., Horikoshi, M., Karhunen, V., Lowry, E., Cousminer, D.L., Ahluwalia, T.S., et al. (2019). A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. *Hum Mol Genet* 28, 3327-3338.
28. World Health Organization. Growth reference data for 5-19 years, [https://www.cdc.gov/growthcharts/cdc\\_charts.htm](https://www.cdc.gov/growthcharts/cdc_charts.htm).
29. Butte, N.F., Cai, G., Cole, S.A., and Comuzzie, A.G. (2006). Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clin Nutr* 84, 646-654; quiz 673-644.
30. Lin, D.Y., Tao, R., Kalsbeek, W.D., Zeng, D.L., Gonzalez, F., Fernandez-Rhodes, L., Graff, M., Koch, G.G., North, K.E., and Heiss, G. (2014). Genetic Association Analysis under Complex Survey Sampling: The Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet* 95, 675-688.
31. Chen, M.H., and Yang, Q. (2010). GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 26, 580-581.
32. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190-2191.



33. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197-206.
34. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 46, 1173-1186.
35. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Magi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 518, 187-196.
36. Ng, M.C.Y., Graff, M., Lu, Y., Justice, A.E., Mudgal, P., Liu, C.T., Young, K., Yanek, L.R., Feitosa, M.F., Wojczynski, M.K., et al. (2017). Discovery and fine-mapping of adiposity loci using high density imputation of genome-wide association studies in individuals of African ancestry: African Ancestry Anthropometry Genetics Consortium. *PLoS Genet* 13, e1006719.
37. Graff, M., Justice, A.E., Young, K.L., Marouli, E., Zhang, X., Fine, R.S., Lim, E., Buchanan, V., Rand, K., Feitosa, M.F., et al. (2021). Discovery and fine-mapping of height loci via high-density imputation of GWASs in individuals of African ancestry. *Am J Hum Genet* 108, 564-582.
38. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., Visscher, P.M., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* 27, 3641-3649.
39. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association

- studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 28, 166-174.
40. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* 49, 1458-1467.
41. Berndt, S.I., Gustafsson, S., Magi, R., Ganna, A., Wheeler, E., Feitosa, M.F., Justice, A.E., Monda, K.L., Croteau-Chonka, D.C., Day, F.R., et al. (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 45, 501-512.
42. Gong, J., Nishimura, K.K., Fernandez-Rhodes, L., Haessler, J., Bien, S., Graff, M., Lim, U., Lu, Y., Gross, M., Fornage, M., et al. (2018). Trans-ethnic analysis of metabochip data identifies two new loci associated with BMI. *Int J Obes (Lond)* 42, 384-390.
43. Gong, J., Schumacher, F., Lim, U., Hindorff, L.A., Haessler, J., Buyske, S., Carlson, C.S., Rosse, S., Buzkova, P., Fornage, M., et al. (2013). Fine Mapping and Identification of BMI Loci in African Americans. *Am J Hum Genet* 93, 661-671.
44. Graff, M., Scott, R.A., Justice, A.E., Young, K.L., Feitosa, M.F., Barata, L., Winkler, T.W., Chu, A.Y., Mahajan, A., Hadley, D., et al. (2017). Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PLoS Genet* 13, e1006528.
45. Hoffmann, T.J., Choquet, H., Yin, J., Banda, Y., Kvale, M.N., Glymour, M., Schaefer, C., Risch, N., and Jorgenson, E. (2018). A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics* 210, 499-515.
46. Justice, A.E., Winkler, T.W., Feitosa, M.F., Graff, M., Fisher, V.A., Young, K., Barata, L., Deng, X., Czajkowski, J., Hadley, D., et al. (2017). Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun* 8, 14977.

47. Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* 104, 65-75.
48. Monda, K.L., Chen, G.K., Taylor, K.C., Palmer, C., Edwards, T.L., Lange, L.A., Ng, M.C., Adeyemo, A.A., Allison, M.A., Bielak, L.F., et al. (2013). A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat Genet* 45, 690-696.
49. Okada, Y., Kubo, M., Ohmiya, H., Takahashi, A., Kumasaka, N., Hosono, N., Maeda, S., Wen, W., Dorajoo, R., Go, M.J., et al. (2012). Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat Genet* 44, 302-306.
50. Tachmazidou, I., Suveges, D., Min, J.L., Ritchie, G.R.S., Steinberg, J., Walter, K., Lotchkova, V., Schwartzentruber, J., Huang, J., Memari, Y., et al. (2017). Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *Am J Hum Genet* 100, 865-884.
51. Turcot, V., Lu, Y., Highland, H.M., Schurmann, C., Justice, A.E., Fine, R.S., Bradfield, J.P., Esko, T., Giri, A., Graff, M., et al. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* 50, 26-41.
52. Wen, W., Cho, Y.S., Zheng, W., Dorajoo, R., Kato, N., Qi, L., Chen, C.H., Delahanty, R.J., Okada, Y., Tabara, Y., et al. (2012). Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* 44, 307-311.
53. Wen, W., Zheng, W., Okada, Y., Takeuchi, F., Tabara, Y., Hwang, J.Y., Dorajoo, R., Li, H., Tsai, F.J., Yang, X., et al. (2014). Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet* 23, 5492-5504.

54. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpelainen, T.O., et al. (2015). The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* 11, e1005378.
55. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570, 514-518.
56. Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al. (2017). Rare and low-frequency coding variants alter human adult height. *Nature* 542, 186-190.
57. Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M., et al. (2009). A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* 41, 527-534.
58. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinhorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Magi, R., et al. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 42, 949-960.
59. Justice, A.E., Karaderi, T., Highland, H.M., Young, K.L., Graff, M., Lu, Y., Turcot, V., Auer, P.L., Fine, R.S., Guo, X., et al. (2019). Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nat Genet* 51, 452-469.
60. Scuteri, A., Sanna, S., Chen, W.M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orru, M., Usala, G., et al. (2007). Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 3, e115.

61. Wen, W., Kato, N., Hwang, J.Y., Guo, X., Tabara, Y., Li, H., Dorajoo, R., Yang, X., Tsai, F.J., Li, S., et al. (2016). Genome-wide association studies in East Asians identify new loci for waist-hip ratio and waist circumference. *Sci Rep* 6, 17958.
62. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32, 1493-1501.
63. Justice, A.E., Karaderi, T., Highland, H.M., Young, K.L., Graff, M., Lu, Y., Turcot, V., Auer, P.L., Fine, R.S., Guo, X., et al. (2019). Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nat Genet* 51, 452-469.
64. Fisher-Hoch, S.P., Rentfro, A.R., Salinas, J.J., Perez, A., Brown, H.S., Reininger, B.M., Restrepo, B.I., Wilson, J.G., Hossain, M.M., Rahbar, M.H., et al. (2010). Socioeconomic status and prevalence of obesity and diabetes in a Mexican American community, Cameron County, Texas, 2004-2007. *Prev Chronic Dis* 7, A53.
65. Babraham Bioinformatics. (2010). FastQC. A quality control tool for high throughput sequence data, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
66. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.
67. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930.
68. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, 550.
69. Staley, J.R., Blackshaw, J., Kamat, M.A., Ellis, S., Surendran, P., Sun, B.B., Paul, D.S., Freitag, D., Burgess, S., Danesh, J., et al. (2016). PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics (Oxford, England)* 32, 3207-3209.

70. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22, 1790-1797.
71. Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 44, D877-881.
72. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res* 12, 996-1006.
73. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature genetics* 45, 580-585.
74. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
75. Price, A.L., Weale, M.E., Patterson, N., Myers, S.R., Need, A.C., Shianna, K.V., Ge, D., Rotter, J.I., Torres, E., Taylor, K.D., et al. (2008). Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* 83, 132-135; author reply 135-139.
76. Paternoster, L., Zhurov, A.I., Toma, A.M., Kemp, J.P., St Pourcain, B., Timpson, N.J., McMahon, G., McArdle, W., Ring, S.M., Smith, G.D., et al. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am J Hum Genet* 90, 478-485.
77. Adhikari, K., Fuentes-Guajardo, M., Quinto-Sanchez, M., Mendoza-Revilla, J., Camilo Chacon-Duque, J., Acuna-Alonzo, V., Jaramillo, C., Arias, W., Lozano, R.B., Perez, G.M., et al. (2016). A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. *Nat Commun* 7, 11616.

78. Xiong, Z., Dankova, G., Howe, L.J., Lee, M.K., Hysi, P.G., de Jong, M.A., Zhu, G., Adhikari, K., Li, D., Li, Y., et al. (2019). Novel genetic loci affecting facial shape variation in humans. *Elife* 8.
79. Yap, C.X., Sidorenko, J., Wu, Y., Kemper, K.E., Yang, J., Wray, N.R., Robinson, M.R., and Visscher, P.M. (2018). Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nat Commun* 9, 5407.
80. Hagenaaars, S.P., Hill, W.D., Harris, S.E., Ritchie, S.J., Davies, G., Liewald, D.C., Gale, C.R., Porteous, D.J., Deary, I.J., and Marioni, R.E. (2017). Genetic prediction of male pattern baldness. *PLoS Genet* 13, e1006594.
81. Nelms, B.L., and Labosky, P.A. (2010). Pax Genes. In *Transcriptional Control of Neural Crest Development*. (San Rafael (CA), Morgan & Claypool Life Sciences.
82. Shakhova O., and L., S. (2010). Neural crest-derived stem cells. In. (Cambridge, MA, Harvard Stem Cell Institute.
83. Lee, G., Kim, H., Elkabetz, Y., Al Shamy, G., Panagiotakos, G., Barberi, T., Tabar, V., and Studer, L. (2007). Isolation and directed differentiation of neural crest stem cells derived from human embryonic stem cells. *Nat Biotechnol* 25, 1468-1475.
84. Billon, N., Iannarelli, P., Monteiro, M.C., Glavieux-Pardanaud, C., Richardson, W.D., Kessar, N., Dani, C., and Dupin, E. (2007). The generation of adipocytes by the neural crest. *Development* 134, 2283-2292.
85. Hans, C., Dobra, A., and West, M. (2007). Shotgun Stochastic Search for “Large p” Regression. *Journal of the American Statistical Association* 102, 507-516.
86. Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W., Hirschhorn, J., Daly, M.J., Patterson, N., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife* 8.

87. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., et al. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* 8.
88. Chen, M., Sidore, C., Akiyama, M., Ishigaki, K., Kamatani, Y., Schlessinger, D., Cucca, F., Okada, Y., and Chiang, C.W.K. (2020). Evidence of Polygenic Adaptation in Sardinia at Height-Associated Loci Ascertained from the Biobank Japan. *Am J Hum Genet* 107, 60-71.
89. Chen, M., and Chiang, C.W.K. (2021). Allele frequency differentiation at height-associated SNPs among continental human populations. *Eur J Hum Genet* 29, 1542-1548.
90. Campbell, C.D., Ogburn, E.L., Lunetta, K.L., Lyon, H.N., Freedman, M.L., Groop, L.C., Altshuler, D., Ardlie, K.G., and Hirschhorn, J.N. (2005). Demonstrating stratification in a European American population. *Nat Genet* 37, 868-872.
91. Robinson, M.R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T., Shakhbazov, K., Powell, J.E., Vinkhuyzen, A., Berndt, S.I., Gustafsson, S., et al. (2015). Population genetic differentiation of height and body mass index across Europe. *Nat Genet* 47, 1357-1362.
92. Carlson, C.S., Matisse, T.C., North, K.E., Haiman, C.A., Fesinmeyer, M.D., Buyske, S., Schumacher, F.R., Peters, U., Franceschini, N., Ritchie, M.D., et al. (2013). Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol* 11, e1001661.
93. Soranzo, N., Rivadeneira, F., Chinappan-Horsley, U., Malkina, I., Richards, J.B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A., et al. (2009). Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet* 5, e1000445.



94. Weedon, M.N., Lango, H., Lindgren, C.M., Wallace, C., Evans, D.M., Mangino, M., Freathy, R.M., Perry, J.R., Stevens, S., Hall, A.S., et al. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 40, 575-583.
95. Cousminer, D.L., Berry, D.J., Timpson, N.J., Ang, W., Thiering, E., Byrne, E.M., Taal, H.R., Huikari, V., Bradfield, J.P., Kerkhof, M., et al. (2013). Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum Mol Genet* 22, 2735-2747.
96. Riveros-McKay, F., Mistry, V., Bounds, R., Hendricks, A., Keogh, J.M., Thomas, H., Henning, E., Corbin, L.J., Understanding Society Scientific, G., O'Rahilly, S., et al. (2019). Genetic architecture of human thinness compared to severe obesity. *PLoS Genet* 15, e1007603.
97. Cho, H.W., Jin, H.S., and Eom, Y.B. (2021). A Genome-Wide Association Study of Novel Genetic Variants Associated With Anthropometric Traits in Koreans. *Front Genet* 12, 669215.
98. Spracklen, C.N., Horikoshi, M., Kim, Y.J., Lin, K., Bragg, F., Moon, S., Suzuki, K., Tam, C.H.T., Tabara, Y., Kwak, S.H., et al. (2020). Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 582, 240-245.
99. Christakoudi, S., Evangelou, E., Riboli, E., and Tsilidis, K.K. (2021). GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. *Sci Rep* 11, 10688.
100. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., et al. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* 52, 680-691.
101. Zhu, Z., Guo, Y., Shi, H., Liu, C.L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links

- between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol* 145, 537-549.
102. Freathy, R.M., Mook-Kanamori, D.O., Sovio, U., Prokopenko, I., Timpson, N.J., Berry, D.J., Warrington, N.M., Widen, E., Hottenga, J.J., Kaakinen, M., et al. (2010). Variants in *ADCY5* and near *CCNL1* are associated with fetal growth and birth weight. *Nat Genet* 42, 430-435.
103. Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L.M., Juodakis, J., Miller, D.E., Litterman, N., Jiang, P.P., Russell, L., et al. (2017). Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *N Engl J Med* 377, 1156-1167.
104. Yang, X.L., Zhang, S.Y., Zhang, H., Wei, X.T., Feng, G.J., Pei, Y.F., and Zhang, L. (2019). Three Novel Loci for Infant Head Circumference Identified by a Joint Association Analysis. *Front Genet* 10, 947.
105. Windpassinger, C., Piard, J., Bonnard, C., Alfadhel, M., Lim, S., Bisteau, X., Blouin, S., Ali, N.B., Ng, A.Y.J., Lu, H., et al. (2017). *CDK10* Mutations in Humans and Mice Cause Severe Growth Retardation, Spine Malformations, and Developmental Delays. *Am J Hum Genet* 101, 391-403.
106. Daelemans, C., Ritchie, M.E., Smits, G., Abu-Amero, S., Sudbery, I.M., Forrest, M.S., Campino, S., Clark, T.G., Stanier, P., Kwiatkowski, D., et al. (2010). High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genet* 11, 25.

## Figure Titles and Legends

**Figure 1. Flowchart of the design and discovery of 6 loci and 36 signals in known loci in the Hispanic/Latino Anthropometry Consortium (HISLA) meta-analysis and the trans-ancestral meta-analysis of HISLA and consortia of other ancestries.**

\*Stage 1 maximum sample sizes varied from 59,771 for BMI, 56,161 for height, to 42,455 for WHRadjBMI (sex combined). \*\*Stage 2 sample sizes varied from 10,538 for BMI, 8,110 for height, to 4,393 for WHRadjBMI (sex combined). Actual sample sizes may vary by SNP. \*\*\*The BMI and height-for-age z-score models were conducted using up to 1,914 and 1,945 of children/adolescents, respectively. In contrast, the obesity case-control study compared up to 1,814 children/adolescents who were  $\geq 95$ th versus  $\leq 50$ th BMI-for-age percentiles.

**Figure 2. Regional plot, unconditioned (A) and conditioned (B) on established variants within +/-500kb of lead variant, at the body mass index (BMI) locus at PAX3 in the Hispanic/Latino Anthropometry Consortium (HISLA, after excluding Brazilian and Native American samples). Linkage disequilibrium patterns are based on rs994108 (shown by the purple diamond) from the Hispanic Communities in Health Study/Study of Latinos.**

**Figure 3. Variability in HISLA Stage 1+2, AAAGC, and GIANT P-values, effect sizes and coded allele frequencies for genome-wide significant anthropometric loci from HISLA Stage 1.**

African American Anthropometry Genetics Consortium (AAAGC); Body mass index (BMI); Coded Allele Frequency (CAF); Genetic Investigation of ANthropometric Traits (GIANT); Hispanic/Latino Anthropometry Consortium (HISLA); WHRadjBMI - waist to hip ratio adjusted for BMI. \*SNPs that remained significant ( $P$ -value  $< 5 \times 10^{-8}$ ) in HISLA Stage 1+2.

**Figure 4. Correlations ( $\rho$ ) between effect estimates and the loadings of the principal components 1-5 in each consortium and the meta-analysis of all 3 consortia by trait. (A) height, (B) BMI, (C) Waist-to-hip ratio adjusted for BMI (WHRadjBMI) for men and women combined, (D) WHRadjBMI for women only, (E) WHRadjBMI for men only.**

Hispanic/Latino Anthropometry Consortium (HISLA); African American Anthropometry Genetics Consortium (AAAGC); Genetic Investigation of ANthropometric Traits (GIANT); WHRadjBMI - waist to hip ratio adjusted for BMI.

## Tables

**Table 1. Potentially novel loci and new signals in known loci from the Stage 1: Adult HISLA Discovery, combined with the Stage 2: Adult HISLA Replication sample,<sup>1</sup> and lookups of results from the AAAGC and GIANT Consortia.**

Trait	Locus Name	SNP rsid	Genomic region <sup>3</sup>	Chr	Position (hg19)	Effect/ Other Alleles	Stage	EAF	Beta	SE	P-value	HetI <sup>2</sup>	N	Novel? Yes/ No <sup>3</sup>
<b>Novel loci<sup>1</sup></b>														
BMI	PAX3 <sup>2</sup>	rs994108	intergenic	2	223057288	C/A	Stage 1: Discovery	0.390	0.041	0.007	<b>1.62E-08</b>	0	43048	Yes
							Stage 2: Validation	0.394	0.038	0.016	5.65E-02	8.0	9336	
							Stage 1 + 2	0.389	0.038	0.006	<b>2.19E-09</b>	0	52384	
							AAAGC	0.526	0.007	0.007	3.26E-01	0	42751	
	GIANT	0.342	<0.001	0.004	9.81E-01	-	233955							
	ARRDC3	rs1505851	intronic	5	90893954	T/C	Stage 1: Discovery	0.741	0.041	0.007	<b>2.287E-08</b>	0	52365	No
							Stage 2: Validation	0.709	0.005	0.017	7.62E-01	33.3	9336	
							Stage 1 + 2	0.735	0.035	0.007	1.16E-07	14.1	61701	
AAAGC							0.307	0.027	0.008	7.00E-04	46.6	42752		
GIANT	0.680	0.001	0.004	7.90E-01	-	233999								
WHRadjBMI - Women only	DOCK2 <sup>2</sup>	rs6879439	intronic	5	169314869	C/T	Stage 1: Discovery	0.520	0.060	0.010	<b>1.02E-08</b>	0	18591	No
							Stage 2: Validation	0.526	0.013	0.028	6.54E-01	28.7	2747	
							Stage 1 + 2	0.515	0.049	0.0093	1.57E-07	1.9	23382	
							AAAGC	0.440	0.012	0.012	3.09E-01	0	15600	
							GIANT	0.610	0.003	0.005	6.30E-01	-	86317	
WHRadjBMI - Sex combined	TAOK3	rs115981023	intronic	12	118751105	A/G	Stage 1: Discovery	0.009	0.328	0.057	<b>1.08E-08</b>	44.8	19640	No
							Stage 2: Validation	0.004	-0.339	0.687	6.22E-01	0	1340	
							Stage 1 + 2	0.009	0.308	0.057	5.18E-08	52.0	20980	
							AAAGC	0.050	0.027	0.027	3.07E-01	0	15601	
							GIANT	0.002	no proxy					
<b>New signals in known loci</b>														
BMI	ADCY5 <sup>2</sup>	rs17361324	intronic	3	123131254	T/C	Stage 1: Discovery	0.280	0.042	0.008	<b>2.60E-08</b>	0	43333	Yes
							Stage 2: Validation	0.269	0.035	0.018	4.70E-02	0	9035	
							Stage 1 + 2	0.278	0.041	0.007	<b>2.84E-09</b>	0	52368	
							AAAGC	0.119	0.023	0.011	3.85E-02	0	42682	
	GIANT	0.253	0.013	0.004	9.90E-04	-	320704							
	ILRUN	rs148899910	intergenic	6	34232259	C/G	Stage 1: Discovery	0.275	0.040	0.007	<b>9.03E-09</b>	0	54105	Yes
							Stage 2: Validation	0.282	0.049	0.017	4.43E-03	0	9035	
							Stage 1 + 2	0.276	0.041	0.006	<b>1.24E-10</b>	0	63140	
AAAGC							0.316	-0.016	0.008	5.02E-02	30.0	42750		
GIANT <sup>4</sup>	0.017	0.036	0.012	3.99E-03	-	216522								
Height	B4GALNT3	rs215226	intronic	12	591300	A/G	Stage 1: Discovery	0.550	-0.032	0.005	<b>5.53E-09</b>	22.1	52156	Yes
							Stage 2: Validation	0.565	-0.020	0.017	2.37E-01	19.3	6906	
							Stage 1 + 2	0.551	-0.031	0.005	<b>1.98E-09</b>	21.8	59062	
							AAAGC	0.772	-0.031	0.009	8.99E-04	24.0	41327	
							GIANT	0.633	0.006	0.004	1.10E-01	-	220370	

Abbreviations: Chr - chromosome; EAF - effect allele frequency; HetI<sup>2</sup> - heterogeneity I-square; N - sample size; WHRadjBMI - waist to hip ratio adjusted for BMI; AAAGC- African American Anthropometry Genetics Consortium; GIANT- Genetic Investigation of ANthropometric Traits Consortium

<sup>1</sup>All studies were meta-analyzed using METAL,<sup>32</sup> with each study entered individually into the Stage 1+2 meta-analysis.

<sup>2</sup>These BMI and WHRadjBMI analyses did not include Brazilian and/or Native American samples.

<sup>3</sup>New loci or signals are those that were replicated by HISLA Stage 2 results that are directionally consistent with Stage 1 and remained genome-wide significant after meta-analysis with Stage 1. <sup>4</sup>Proxy GIANT, rs1573905 ( $r^2= 0.96$  AMR).

<sup>3</sup>Human Genome Organisation approved gene names.

Journal Pre-proof

**Table 2. Novel loci and new signals in established loci by trait from a trans-ancestral meta-analysis of Adult samples from the HISLA, AAAGC, and GIANT Consortia.**

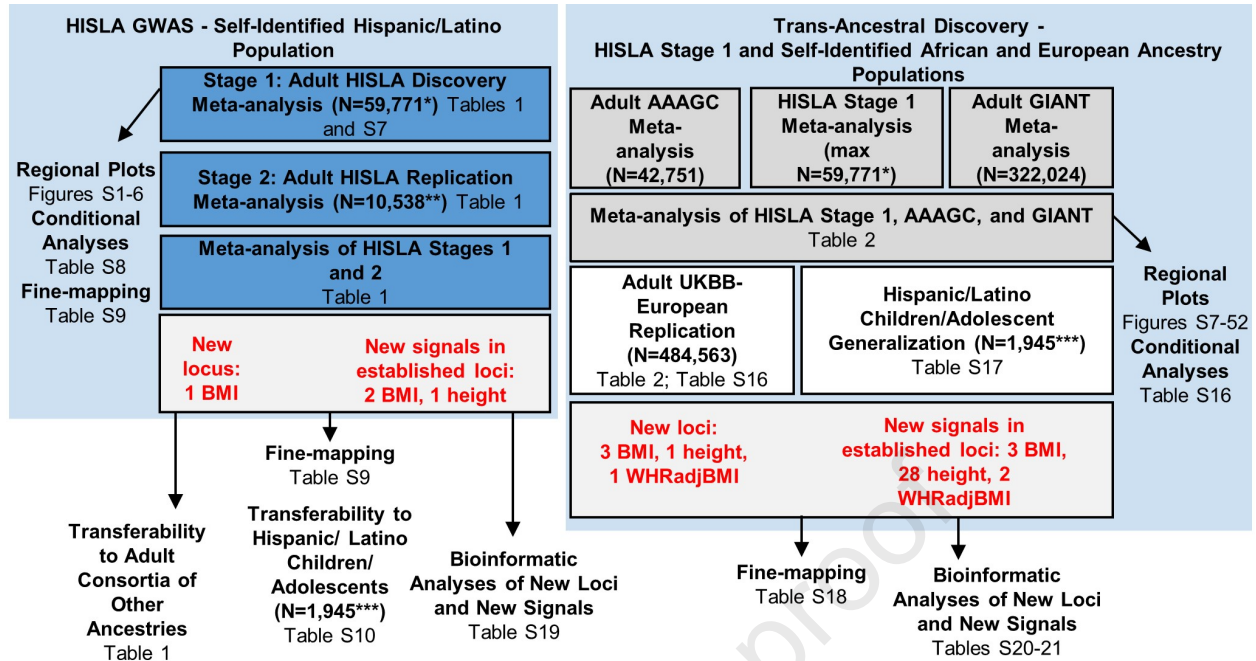
Trait	SNP rsid	Chr	Position (hg19)	Locus Name <sup>3</sup>	Effect/Other Alleles	EAF			N	Unconditioned meta-analysis results				Conditioned on all known SNPs within 10Mb region			UKBB (Validation results)				
						HISLA	AAAGC	GIANT		Beta	SE	P-value	HetI Sq	Beta	SE	P-value	EAF	N	Beta	SE	P-value
<b>Novel loci<sup>1</sup></b>																					
BMI	rs4675117	2	227769794	RHBDD1	T/C	0.421	0.104	0.383	343628	0.017	0.003	8.56E-08	0	0.019	0.003	2.23E-09	0.383	336107	0.006	0.002	1.82E-02
BMI	rs9860730	3	64701146	ADAMTS9-AS2	A/G	0.354	0.222	0.767	428763	-0.016	0.003	1.67E-08	0	-0.015	0.003	3.80E-08	0.712	336107	-0.008	0.003	4.54E-03
BMI	rs150932	5	92273197	CHD1-DT	A/G	0.778	0.585	0.793	438077	0.019	0.003	5.40E-10	0	0.017	0.003	1.02E-08	0.693	336107	0.005	0.003	3.74E-12
Height	rs17375290	1	61334177	NFYA	A/G	0.830	0.645	0.793	364656	0.017	0.003	3.47E-08	0	0.017	0.003	3.00E-08	0.794	336474	0.002	0.002	4.58E-01
Height	rs10737541	1	168214098	ANKRD36BP1	T/C	0.399	0.645	0.196	319809	-0.018	0.003	1.60E-09	0	-0.018	0.003	3.16E-10	0.226	336474	-0.004	0.002	4.37E-02
Height	rs4618485	6	73565917	KCNQ5	A/G	0.750	0.645	0.592	348626	0.014	0.003	4.72E-08	39.2	0.018	0.003	4.34E-12	0.604	336474	0.003	0.002	7.02E-02
Height	rs17493997	8	82044302	PAG1	C/G	0.483	0.645	0.340	325906	-0.015	0.003	3.42E-08	0	-0.017	0.003	2.73E-10	0.299	336474	-0.003	0.002	8.56E-01
WHRadjBMI sex-combined	rs16873543	6	45577134	RUNX2	T/C	0.716	0.645	0.767	209552	-0.018	0.004	3.20E-06	0	-0.022	0.004	9.65E-09	0.724	484563	-0.008	0.002	5.50E-04
<b>New signals in established loci<sup>2</sup></b>																					
BMI	rs10540	11	494662	RNH1	A/G	0.200	0.031	0.092	470714	-0.021	0.004	1.01E-07	0	-0.023	0.004	5.75E-09	0.135	336107	-0.007	0.004	4.16E-02
BMI	rs4807179	19	1956035	CSNK1G2	A/G	0.494	0.218	0.525	309507	0.020	0.003	2.75E-10	0	0.018	0.003	1.52E-08	0.632	336107	0.014	0.002	1.06E-08
BMI	rs4813428	20	21451848	NKX2-2	T/C	0.181	0.093	0.108	321797	0.029	0.005	2.89E-10	0	0.029	0.005	1.46E-10	0.093	336107	0.013	0.004	2.47E-03
Height	rs4912122	1	19876438	NKX2-2	A/G	0.383	0.837	0.383	334951	-0.015	0.003	6.33E-09	0	-0.019	0.003	1.45E-13	0.350	336474	-0.012	0.002	1.58E-11
Height	rs4425978	1	42243878	HIVEP3	T/C	0.389	0.189	0.580	351587	0.014	0.003	2.14E-08	0	0.016	0.003	3.63E-10	0.533	336474	0.008	0.002	5.41E-06
Height	rs618555	1	86481084	COL24A1	T/C	0.241	0.064	0.353	320239	0.019	0.003	5.47E-12	0	0.016	0.003	2.90E-08	0.311	336474	0.008	0.002	1.67E-05
Height	rs6545538	2	56217900	MIR216A	A/G	0.258	0.628	0.258	305704	0.022	0.003	1.23E-13	0	0.019	0.003	2.59E-11	0.286	336474	0.011	0.002	1.25E-08
Height	rs274311	2	233239743	ALP	T/C	0.076	0.018	0.050	463609	0.046	0.005	1.41E-21	49.6	0.030	0.005	4.78E-10	0.080	336474	0.033	0.003	2.39E-24
Height	rs6938964	6	26255451	HIST1H2BH	A/G	0.306	0.103	0.408	345378	0.042	0.003	3.06E-09	46.5	0.018	0.003	3.29E-12	0.425	336474	-0.027	0.002	1.10E-54
Height	rs6900530	6	35280971	DEFB	T/C	0.094	0.385	0.042	123137	-0.057	0.005	3.09E-28	79.0	-0.036	0.005	6.75E-12	0.027	336474	-0.073	0.005	1.03E-42
Height	rs9472006	6	43067487	PTK7	A/G	0.149	0.249	0.059	212931	-0.027	0.005	4.37E-09	0	-0.034	0.005	7.04E-14	0.041	336474	-0.013	0.004	2.86E-03
Height	rs3822957	6	76607280	MYO6	A/G	0.173	0.538	0.158	279818	-0.023	0.003	4.88E-12	84.4	-0.023	0.003	3.44E-12	0.142	336474	-0.015	0.002	1.98E-09
Height	rs1342330	6	144065685	PHACTR2	A/T	0.601	0.738	0.508	353259	0.014	0.003	1.70E-08	0	0.017	0.003	6.95E-12	0.520	336474	0.006	0.002	8.84E-04
Height	rs6936615	6	154355100	OPRM1	A/G	0.887	0.901	0.850	415248	-0.018	0.003	2.47E-08	0	-0.020	0.003	2.33E-09	0.830	336474	-0.003	0.002	1.73E-01
Height	rs991946	6	166329862	RP11-252P19.3 <sup>3</sup>	T/C	0.481	0.486	0.500	379912	-0.019	0.002	1.28E-14	36.7	-0.019	0.002	9.96E-15	0.479	336474	-0.013	0.002	3.92E-13
Height	rs7816300	8	109787856	TMEM74	T/C	0.262	0.116	0.260	397735	-0.015	0.003	2.24E-08	0	-0.016	0.003	5.47E-09	0.299	336474	-0.002	0.002	3.94E-01
Height	rs4520250	9	88924057	TUT7	A/C	0.235	0.069	0.433	295945	0.015	0.003	3.80E-08	55.9	0.016	0.003	1.48E-08	0.339	336474	0.010	0.002	4.36E-08
Height	rs7029157	9	97000863	snoU13 <sup>3</sup>	T/C	0.084	0.325	0.067	262808	0.028	0.004	5.85E-11	16.8	0.026	0.004	1.00E-09	0.088	336474	0.030	0.003	4.55E-22
Height	rs12347744	9	97575273	AOPEP	T/C	0.102	0.015	0.033	454111	-0.032	0.005	6.55E-11	65.8	-0.030	0.005	6.06E-10	0.061	336474	-0.031	0.004	1.04E-17
Height	rs7024254	9	109498129	ZNF462	A/G	0.255	0.721	0.207	324351	0.017	0.003	1.04E-08	23.7	0.036	0.003	6.10E-35	0.204	336474	0.010	0.002	6.50E-06
Height	rs10119624	9	118305438	DELEC1	A/G	0.612	0.437	0.683	353374	0.021	0.003	2.56E-16	31.1	0.022	0.003	8.02E-17	0.671	336474	0.012	0.002	3.20E-10
Height	rs2902635	10	105476045	SH3PX2A	T/G	0.710	0.382	0.821	308762	-0.021	0.003	1.59E-12	0	-0.017	0.003	1.62E-08	0.805	336474	-0.015	0.002	7.65E-12
Height	rs17659078	11	2284590	ASCL2	A/C	0.211	0.149	0.258	354951	0.019	0.003	1.78E-10	9.0	0.016	0.003	3.45E-08	0.273	336474	0.004	0.002	2.24E-02
Height	rs11605993	11	122837037	JHY	T/C	0.591	0.384	0.448	380447	-0.017	0.002	3.44E-12	0	-0.018	0.002	3.07E-14	0.447	336474	-0.013	0.002	4.45E-13
Height	rs621794	11	125848462	CDON	A/G	0.500	0.430	0.475	380049	-0.014	0.002	1.46E-08	0	-0.014	0.002	7.29E-09	0.429	336474	-0.009	0.002	1.76E-07
Height	rs11221442	11	12857624	FLI1	C/G	0.171	0.109	0.217	352360	-0.022	0.003	2.03E-12	78.8	-0.023	0.003	2.03E-13	0.252	336474	-0.008	0.002	2.59E-05
Height	rs12300112	12	103147575	LINC00485	C/G	0.075	0.242	0.008	154014	0.041	0.006	4.31E-13	45.2	0.038	0.006	3.11E-11	0.027	336474	0.038	0.005	4.61E-12
Height	rs11616067	12	116393174	MED13L	A/G	0.841	0.904	0.750	327941	0.021	0.003	6.38E-12	21.0	0.018	0.003	2.80E-09	0.768	336474	0.012	0.002	6.01E-09
Height	rs17197170	14	21977962	MEITL3	A/G	0.931	0.960	0.842	310344	-0.026	0.004	5.40E-12	64.9	-0.025	0.004	3.74E-11	0.828	336474	-0.018	0.002	3.07E-14
Height	rs11076551	16	51109492	RP11-883G14.4 <sup>3</sup>	A/G	0.341	0.343	0.267	434688	0.014	0.003	2.14E-08	21.6	0.015	0.003	5.87E-09	0.376	336474	0.009	0.002	2.51E-07
Height	rs12918773	16	89741403	SPATA33	A/G	0.121	0.036	0.149	279104	-0.024	0.004	7.45E-09	46.1	-0.024	0.004	3.50E-09	0.112	336474	-0.024	0.003	4.71E-19
Height	rs1346490	19	7244233	INSR	A/C	0.546	0.611	0.500	333896	0.015	0.003	3.99E-09	12.3	0.014	0.003	4.86E-08	0.620	336474	0.009	0.002	1.67E-19
Height	rs17457472	19	17493810	PLVAP	A/C	0.037	0.062	0.028	356712	-0.051	0.007	5.70E-14	0	-0.042	0.007	9.89E-10	0.040	336474	-0.019	0.004	1.12E-05
WHRadjBMI sex-combined	rs17099388	5	142095250	FGF1	A/G	0.166	0.256	0.008	105460	0.039	0.007	3.14E-09	0	0.037	0.007	2.47E-08	0.041	484563	0.025	0.005	2.00E-07
WHRadjBMI sex-combined	rs7975017	12	26428793	SSPN	T/C	0.303	0.717	0.167	267044	-0.021	0.004	7.57E-09	5.5	-0.021	0.004	1.05E-08	0.239	484563	-0.014	0.002	2.00E-09

Abbreviations: Chr - chromosome; EAF - effect allele frequency; HetI sq - heterogeneity I-square; N - sample size; WHRadjBMI - waist to hip ratio adjusted for BMI; AAAGC- African American Anthropometry Genetics Consortium; GIANT- Genetic Investigation of ANthropometric Traits Consortium

<sup>1</sup>Each novel locus was defined by the absence of known (previously published) SNPs within 1Mb (+/-500 Kb) of the lead SNP.

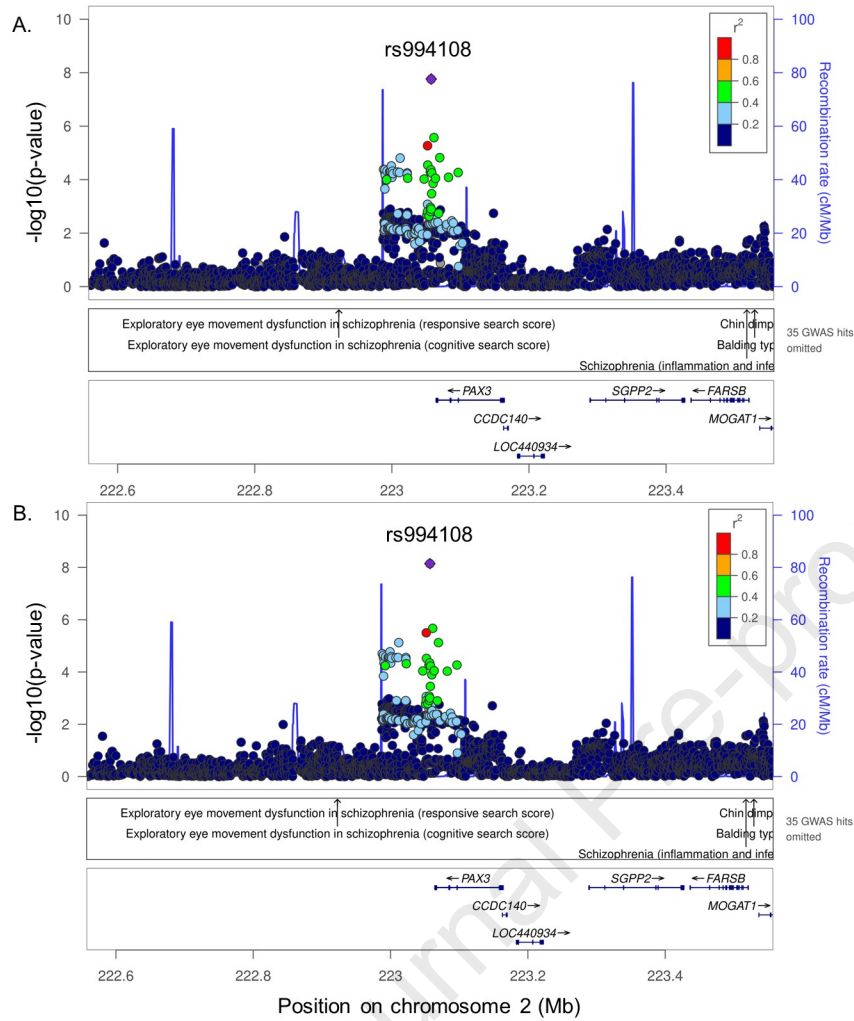
<sup>2</sup>Each known locus was defined by a 1Mb region around previously identified SNP(s) for the indicated trait; the known SNP(s),  $P < 5 \times 10^{-8}$ , at each established locus can be found in Table S16.

<sup>3</sup>Human Genome Organisation approved gene names, unless otherwise indicated.



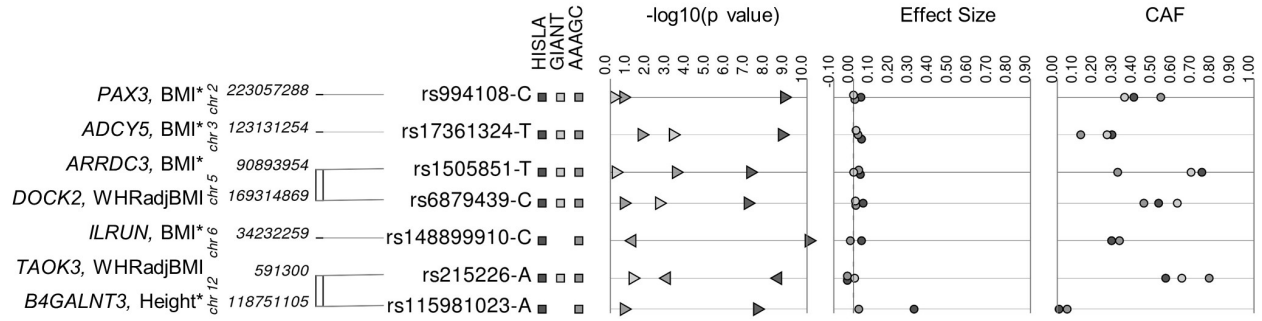
**Figure 1.** Flowchart of the design and discovery of 6 loci and 36 signals in known loci in the Hispanic/Latino Anthropometry Consortium (HISLA) meta-analysis and the trans-ancestral meta-analysis of HISLA and consortia of other ancestries.

\*Stage 1 maximum sample sizes varied from 59,771 for BMI, 56,161 for height, to 42,455 for WHRadjBMI (sex combined). \*\*Stage 2 sample sizes varied from 10,538 for BMI, 8,110 for height, to 4,393 for WHRadjBMI (sex combined). Actual sample sizes may vary by SNP. \*\*\*The BMI and height-for-age z-score models were conducted using up to 1,914 and 1,945 of children/adolescents, respectively. In contrast, the obesity case-control study compared up to 1,814 children/adolescents who were  $\geq 95$ th versus  $\leq 50$ th BMI-for-age percentiles.



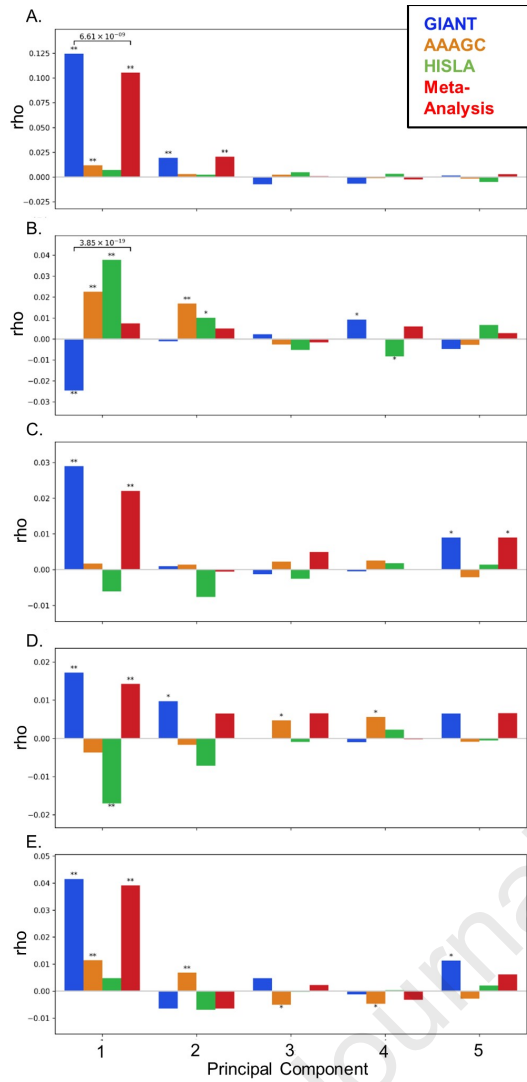
**Figure 2.** Regional plot, unconditioned (A) and conditioned (B) on established variants within  $\pm 500\text{kb}$  of lead variant, at the body mass index (BMI) locus at *PAX3* in the Hispanic/Latino Anthropometry Consortium (HISLA, after excluding Brazilian and Native American samples). Linkage disequilibrium patterns are based on rs994108 (shown by the purple diamond) from the Hispanic Communities in Health Study/Study of Latinos.





**Figure 3.** Variability in HISLA Stage 1+2, AAAGC, and GIANT P-values, effect sizes and coded allele frequencies for genome-wide significant anthropometric loci from HISLA Stage 1.

African American Anthropometry Genetics Consortium (AAAGC); Body mass index (BMI); Coded Allele Frequency (CAF); Genetic Investigation of ANthropometric Traits (GIANT); Hispanic/Latino Anthropometry Consortium (HISLA); WHRadjBMI - waist to hip ratio adjusted for BMI. \*SNPs that remained significant ( $P$ -value  $< 5 \times 10^{-8}$ ) in HISLA Stage 1+2.



**Figure 4.** Correlations ( $\rho$ ) between effect estimates and the loadings of the principal components 1-5 in each consortium and the meta-analysis of all 3 consortia by trait. (A) height, (B) BMI, (C) Waist-to-hip ratio adjusted for BMI (WHRadjBMI) for men and women combined, (D) WHRadjBMI for women only, (E) WHRadjBMI for men only. Hispanic/Latino Anthropometry Consortium (HISLA); African American Anthropometry Genetics Consortium (AAAGC); Genetic Investigation of ANthropometric Traits (GIANT); WHRadjBMI - waist to hip ratio adjusted for BMI.

**Highlights**

- Largest genome-wide association study of anthropometrics in Hispanic/Latinos to date
- We identified 42 novel loci or novel secondary signals in known loci
- Trans-ancestral analyses revealed insights on obesity, bone development and growth
- Future studies should build large, more diverse genome-wide resources

**eTOC Blurp**

Large-scale genetic analyses of diverse populations hold great potential for advancing the field. In our sample of Hispanic/Latino adults, and when combined with publicly available results from other ancestries, we identified, described, and characterized 42 novel genomic findings. We additionally illustrated the novel biologic insights can be garnered from studying ancestrally diverse populations. Overall, our findings indicate the added value of building large, more diverse genome-wide association studies in the future.