ARTICLE

Discovery and fine-mapping of height loci via high-density imputation of GWASs in individuals of African ancestry

Mariaelisa Graff,[1,*] Anne E. Justice,[1,2] Kristin L. Young,[1] Eirini Marouli,[3,4] Xinruo Zhang,[1] Rebecca S. Fine,[5] Elise Lim,[6] Victoria Buchanan,[1] Kristin Rand,[7] Mary F. Feitosa,[8] Mary K. Wojczynski,[8] Lisa R. Yanek,[9] Yaming Shao,[1] Rebecca Rohde,[1] Adebowale A. Adeyemo,[10] Melinda C. Aldrich,[11,12,13] Matthew A. Allison,[14] Christine B. Ambrosone,[15] Stefan Ambs,[16] Christopher Amos,[17] Donna K. Arnett,[18] Larry Atwood,[19] Elisa V. Bandera,[20] Traci Bartz,[21,22] Diane M. Becker,[9] Sonja I. Berndt,[23] Leslie Bernstein,[24] Lawrence F. Bielak,[25] William J. Blot,[11,12,26] Erwin P. Bottinger,[27] Donald W. Bowden,[28,29,30] Jonathan P. Bradfield,[31] Jennifer A. Brody,[21] Ulrich Broeckel,[32] Gregory Burke,[33] Brian E. Cade,[34,35] Qiuyin Cai,[11,12] Neil Caporaso,[23] Chris Carlson,[36] John Carpten,[37] Graham Casey,[7,38] Stephen J. Chanock,[23] Guanjie Chen,[10] Minhui Chen,[39] Yii-Der I. Chen,[40]

(Author list continued on next page)

Summary

Although many loci have been associated with height in European ancestry populations, very few have been identified in African ancestry individuals. Furthermore, many of the known loci have yet to be generalized to and fine-mapped within a large-scale African ancestry sample. We performed sex-combined and sex-stratified meta-analyses in up to 52,764 individuals with height and genome-wide genotyping data from the African Ancestry Anthropometry Genetics Consortium (AAAGC). We additionally combined our African ancestry meta-analysis results with published European genome-wide association study (GWAS) data. In the African ancestry analyses, we identified three novel loci (*SLC4A3*, *NCOA2, ECD/FAM149B1*) in sex-combined results and two loci (*CRB1, KLF6*) in women only. In the African plus European sex-combined GWAS, we identified an additional three novel loci (*RCCD1, G6PC3, CEP95*) which were equally driven by AAAGC and European results. Among 39 genome-wide significant signals at known loci, conditioning index SNPs from European studies identified 20 secondary signals. Two of the 20 new secondary signals and none of the 8 novel loci had minor allele frequencies (MAF) $< 5\%$. Of 802 known European height signals, 643 displayed directionally consistent associations with height, of which 205 were nominally significant ($p < 0.05$) in the African ancestry sex-combined sample. Furthermore, 148 of 241 loci contained ≤ 20 variants in the credible sets that jointly account for 99% of the posterior probability of driving the associations. In summary, trans-ethnic meta-analyses revealed novel signals and further improved fine-mapping of putative causal variants in loci shared between African and European ancestry populations.

Introduction

Human height is a highly heritable, polygenic trait that results from the interplay between many complex growth processes.[1–7] GWASs have identified more than 800, mostly common (MAF $> 5\%$), variants associated with adult height variation, primarily in European populations[1–6] but also to some extent across multiple race/ethnic

[1]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [2]Population Health Services, Geisinger Health, Danville, PA 17822, USA; [3]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK; [4]Centre for Genomic Health, Life Sciences, Queen Mary University of London, London EC1M 6BQ, UK; [5]Vertex Pharmaceuticals Inc., Boston, MA 02210, USA; [6]Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA; [7]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; [8]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108, USA; [9]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [10]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [11]Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; [12]Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; [13]Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; [14]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA 92093, USA; [15]Department of Cancer Prevention and Control, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA; [16]Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA; [17]Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [18]School of Public Health, University of Kentucky, Lexington, KY 40563, USA; [19]Framingham Heart Study, Boston University School of Medicine, Boston, MA 02118, USA; [20]Department of Population Science, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ 08903, USA; [21]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA; [22]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [23]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD 20892, USA; [24]Division of Biomarkers of Early Detection and Prevention, Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA 91010, USA; [25]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA; [26]International

(Affiliations continued on next page)

Wei-Min Chen,[41] Alessandra Chesi,[42] Charleston W.K. Chiang,[39] Lisa Chu,[43] Gerry A. Coetzee,[7,38,44] David V. Conti,[7] Richard S. Cooper,[45] Mary Cushman,[46] Ellen Demerath,[47] Sandra L. Deming,[11,12] Latchezar Dimitrov,[28] Jingzhong Ding,[48] W. Ryan Diver,[49] Qing Duan,[50] Michele K. Evans,[51] Adeyinka G. Falusi,[52] Jessica D. Faul,[53] Myriam Fornage,[54] Caroline Fox,[55,56,57] Barry I. Freedman,[58] Melissa Garcia,[59] Elizabeth M. Gillanders,[60] Phyllis Goodman,[61] Omri Gottesman,[27] Struan F.A. Grant,[31,62,63] Xiuqing Guo,[40] Hakon Hakonarson,[31,42,63] Talin Haritunians,[64] Tamara B. Harris,[65] Curtis C. Harris,[16] Brian E. Henderson,[7,38] Anselm Hennis,[66,67,68] Dena G. Hernandez,[69] Joel N. Hirschhorn,[70,71,72] Lorna Haughton McNeill,[73] Timothy D. Howard,[28] Barbara Howard,[74] Ann W. Hsing,[43,75] Yu-Han H. Hsu,[70,71,72] Jennifer J. Hu,[76,77] Chad D. Huff,[17] Dezheng Huo,[78] Sue A. Ingles,[7,38] Marguerite R. Irvin,[79] Esther M. John,[80] Karen C. Johnson,[81] Joanne M. Jordan,[82] Edmond K. Kabagambe,[79] Sun J. Kang,[83] Sharon L. Kardia,[25] Brendan J. Keating,[84] Rick A. Kittles,[85] Eric A. Klein,[86] Suzanne Kolb,[36] Laurence N. Kolonel,[87] Charles Kooperberg,[36] Lewis Kuller,[88] Abdullah Kutlar,[89] Leslie Lange,[90] Carl D. Langefeld,[91] Loic Le Marchand,[87] Hampton Leonard,[69,92] Guillaume Lettre,[93,94] Albert M. Levin,[95] Yun Li,[50,96,97] Jin Li,[98] Yongmei Liu,[99] Youfang Liu,[82] Simin Liu,[100] Kurt Lohman,[91] Vaneet Lotay,[27] Yingchang Lu,[11,27] William Maixner,[101] JoAnn E. Manson,[102] Barbara McKnight,[21,22] Yan Meng,[70] Keri L. Monda,[1,103] Kris Monroe,[7] Jason H. Moore,[104] Thomas H. Mosley,[105] Poorva Mudgal,[30] Adam B. Murphy,[106] Rajiv Nadukuru,[27] Mike A. Nalls,[69,92] Katherine L. Nathanson,[107] Uma Nayak,[41] Amidou N'Diaye,[93] Barbara Nemesure,[66] Christine Neslund-Dudas,[95] Marian L. Neuhouser,[36] Sarah Nyante,[1,108] Heather Ochs-Balcom,[109]

(Author list continued on next page)

Epidemiology Institute, Rockville, MD 20850, USA; [27]The Charles R. Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [28]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [29]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [30]Center for Diabetes Research, Wake Forest school of Medicine, Winston-Salem, NC 27157, USA; [31]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [32]Department of Pediatrics, Section of Genomic Pediatrics, Medical College of Wisconsin, Milwaukee, WI 53226, USA; [33]Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [34]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA 02115, USA; [35]Harvard Medical School, Boston, MA 02115, USA; [36]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [37]Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; [38]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90033, USA; [39]Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; [40]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA; [41]Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA; [42]Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [43]Cancer Prevention Institute of California, Fremont, CA 94538, USA; [44]Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, LA 90033, USA; [45]Department of Public Health Sciences, Stritch School of Medicine, Loyola University Chicago Stritch School of Medicine, Maywood, IL 60153, USA; [46]Department of Medicine, University of Vermont College of Medicine, Burlington, VT 05405, USA; [47]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN 55455, USA; [48]Section on Gerontology and Geriatric Medicine, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [49]Epidemiology Research Program, American Cancer Society, Atlanta, GA 30303, USA; [50]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [51]Health Disparities Research Section, Clinical Research Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA; [52]Institute for Medical Research and Training, University of Ibadan, Ibadan, Nigeria; [53]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA; [54]Center for Human Genetics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [55]Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA; [56]Framingham Heart Study, Framingham, MA 01702, USA; [57]Division of Endocrinology and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [58]Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [59]National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA; [60]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD 20892, USA; [61]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [62]Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia Research Institute, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [63]Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; [64]Medical Genetics Institute, Cedars Sinai Medical Center, Los Angeles, CA 90048, USA; [65]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA; [66]Department of Preventive Medicine, Stony Brook University, Stony Brook, NY 11794, USA; [67]Chronic Disease Research Centre and Faculty of Medical Sciences, University of West Indies, Bridgetown, Barbados; [68]Ministry of Health, Bridgetown, Barbados; [69]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20814, USA; [70]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [71]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA 02115, USA; [72]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA; [73]Department of Health Disparities Research, Division of OVP, Cancer Prevention and Population Sciences, and Center for Community Implementation and Dissemination Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [74]MedStar Research Institute, Hyattsville, MD 20782, USA; [75]Department of Medicine, Stanford Prevention Research Center and Cancer Institute, Stanford, CA 94305, USA; [76]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL 33136, USA; [77]Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL 33136, USA; [78]Department of Public Health Sciences, University of Chicago, Chicago, IL 60637, USA; [79]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [80]Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA 94304, USA; [81]Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN 38163, USA; [82]Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [83]Genetic Epidemiology Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892, USA; [84]Institute for Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104, USA; [85]Division of Health Equities, Department of Population Sciences, City of Hope Medical Center, Duarte, CA 91010, USA;

(Affiliations continued on next page)

Temidayo O. Ogundiran,[110] Adesola Ogunniyi,[111] Oladosu Ojengbede,[112] Hayrettin Okut,[30] Olufunmilayo I. Olopade,[113] Andrew Olshan,[1,108] Badri Padhukasahasram,[114] Julie Palmer,[115] Cameron D. Palmer,[70,71] Nicholette D. Palmer,[116] George Papanicolaou,[117] Sanjay R. Patel,[118] Curtis A. Pettaway,[119] Patricia A. Peyser,[25] Michael F. Press,[38] D.C. Rao,[120] Laura J. Rasmussen-Torvik,[121] Susan Redline,[102] Alex P. Reiner,[36] Suhn K. Rhie,[7,38] Jorge L. Rodriguez-Gil,[76,77] Charles N. Rotimi,[10] Jerome I. Rotter,[40] Edward A. Ruiz-Narvaez,[122] Benjamin A. Rybicki,[95] Babatunde Salako,[112] Michele M. Sale,[41] Maureen Sanderson,[123] Eric Schadt,[124] Pamela J. Schreiner,[47] Claudia Schurmann,[27] Ann G. Schwartz,[125,126] Daniel A. Shriner,[10] Lisa B. Signorello,[11,12,26] Andrew B. Singleton,[69] David S. Siscovick,[127] Jennifer A. Smith,[25,53] Shad Smith,[101] Elizabeth Speliotes,[128,129] Margaret Spitz,[17] Janet L. Stanford,[36,130] Victoria L. Stevens,[49] Alex Stram,[7] Sara S. Strom,[17] Lara Sucheston,[15] Yan V. Sun,[131] Salman M. Tajuddin,[59] Herman Taylor,[105] Kira Taylor,[132] Bamidele O. Tayo,[45] Michael J. Thun,[49] Margaret A. Tucker,[23] Dhananjay Vaidya,[9,133] David J. Van Den Berg,[7,38] Sailaja Vedantam,[70,71] Mara Vitolins,[33] Zhaoming Wang,[134] Erin B. Ware,[25,53] Sylvia Wassertheil-Smoller,[135] David R. Weir,[53] John K. Wiencke,[136] Scott M. Williams,[137] L. Keoki Williams,[114,138] James G. Wilson,[139] John S. Witte,[140,141] Margaret Wrensch,[136,140] Xifeng Wu,[17] Jie Yao,[40] Neil Zakai,[46] Krista Zanetti,[60] Babette S. Zemel,[63,142] Wei Zhao,[25] Jing Hua Zhao,[143] Wei Zheng,[11,12] Degui Zhi,[144] Jie Zhou,[10] Xiaofeng Zhu,[137] Regina G. Ziegler,[23] Joe Zmuda,[88] Alan B. Zonderman,[65] Bruce M. Psaty,[21] Ingrid B. Borecki,[8,145] L. Adrienne Cupples,[6,19] Ching-Ti Liu,[6] Christopher A. Haiman,[7,38] Ruth Loos,[27,146,147] Maggie C.Y. Ng,[28,30] and Kari E. North[1]

[86]Department of Urology, Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH 44195, USA; [87]Epidemiology Program, Cancer Research Center, University of Hawaii Cancer Center, Honolulu, HI 96813, USA; [88]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; [89]Sickle Cell Center, Medical College of Georgia, Augusta University, Augusta, GA 30912, USA; [90]Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA; [91]Department of Biostatistics and Data Science, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA; [92]Data Tecnica Int'l, LLC, Glen Echo, MD 20812, USA; [93]Montreal Heart Institute, Montréal, QC H1T 1C8, Canada; [94]Department of Medicine, Université de Montréal, Montréal, QC H1T 1C8, Canada; [95]Department of Public Health Sciences, Henry Ford Health System, Detroit, MI 48202, USA; [96]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [97]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [98]Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Palo Alto, CA 94304, USA; [99]Department of Medicine, Division of Cardiology, Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC 27701, USA; [100]Department of Epidemiology, Brown University, Providence, RI 02912, USA; [101]Center for Translational Pain Medicine, Department of Anesthesiology, Duke University Medical Center, Durham, NC 27710, USA; [102]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [103]The Center for Observational Research, Amgen, Inc., Thousand Oaks, CA 91320, USA; [104]Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA; [105]Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA; [106]Department of Urology, Northwestern University, Chicago, IL 60611, USA; [107]Department of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [108]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; [109]Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, The State University of New York, Buffalo, NY 14214, USA; [110]Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria; [111]Department of Medicine, College of Medicine, University of Ibadan, Ibadan, Nigeria; [112]Centre for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria; [113]Center for Clinical Cancer Genetics and Global Health, University of Chicago Medical Center, Chicago, IL 60637, USA; [114]Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, MI 48202, USA; [115]Slone Epidemiology Center, Boston University School of Medicine, Boston, MA 02118, USA; [116]Department of Biochemistry, School of Medicine, Wake Forest University, Winston-Salem, NC 27157, USA; [117]Division of Cardiovascular Sciences, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA; [118]Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA; [119]Department of Urology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA; [120]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO 63110, USA; [121]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA; [122]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI 48109, USA; [123]Department of Family and Community Medicine, Meharry Medical College, Nashville, TN 37208, USA; [124]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [125]Department of Oncology, Wayne State University School of Medicine, Detroit, MI 48201, USA; [126]Karmanos Cancer Institute, Detroit, MI 48201, USA; [127]The New York Academy of Medicine, New York, NY 10029, USA; [128]Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, MI 48109, USA; [129]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA; [130]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA 98195, USA; [131]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA; [132]Department of Epidemiology and Population Health, School of Public Health and Information Sciences, University of Louisville, Louisville, KY 40202, USA; [133]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA; [134]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN 38105, USA; [135]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA; [136]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94158, USA; [137]Departments of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106, USA; [138]Department of Internal Medicine, Henry Ford Health System, Detroit, MI 48202, USA; [139]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA; [140]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158, USA; [141]Department of Urology, University of California, San Francisco, San Francisco, CA 94158, USA; [142]Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA 19146, USA; [143]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK; [144]School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [145]BioData Catalyst Program, National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA; [146]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [147]The Mindich Child Health Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

*Correspondence: migraff@email.unc.edu
https://doi.org/10.1016/j.ajhg.2021.02.011.

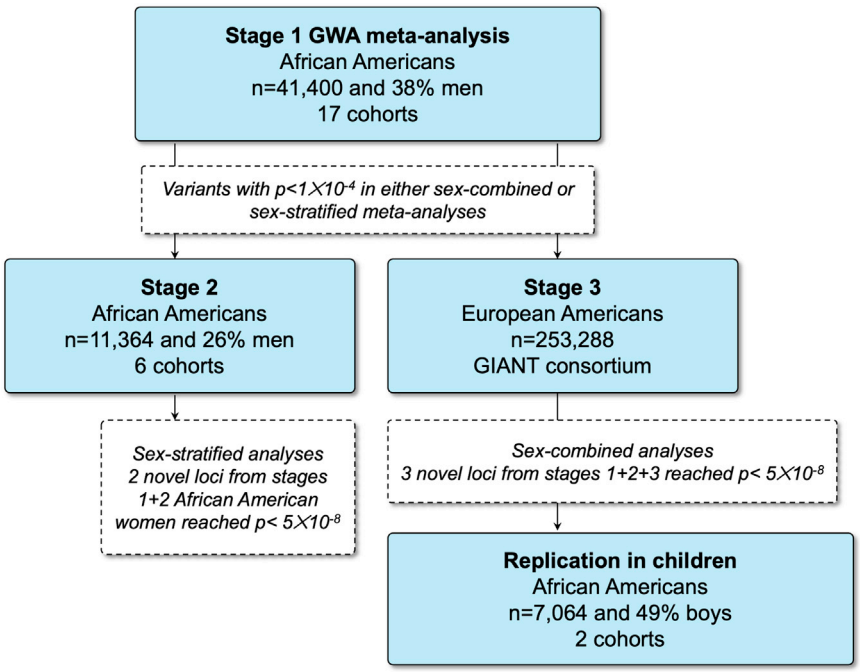

Figure 1. Three-stage design to evaluate genetic association with height in sex-combined and sex-stratified samples

In stage 1, genome-wide association results from 17 studies including 41,400 individuals (16,032 men and 25,368 women) of African ancestry (AA) were meta-analyzed. For variants with p < 1E−4 in either the sex-combined or the sex-stratified meta-analyses, stage 2 replication was performed in an additional 11,364 individuals (2,915 men and 8,449 women) of AA from AAAGC. In stage 3 we completed a meta-analysis of stage 1 and stage 2 results of AA individuals and 253,288 individuals of European ancestry (EA) from the GIANT consortium. Variants that reached genome-wide significance (p < 5E−8) in stage 2 and stage 3 were assessed for associations in two AA children's cohorts (N = 7,064).

groups.[6,7] A recent analysis using a multiethnic sample and an exome array in >700,000 individuals identified height associations with 32 rare and 51 low-frequency coding variants[6] that were not well-captured in previous GWASs imputed to the HapMap reference panel.[8,9] However, with the decreasing cost of whole-genome sequencing, higher density reference panels with larger numbers of haplotypes have become available, such as the 1000 Genomes Project (38M variants in 1,092 individuals from phase 1, 84M variants in 2,504 individuals from phase 3)[10,11] and the Haplotype Reference Consortium (39M variants in 32,611 primarily European individuals; see web resources). These reference panels have substantially improved imputation quality, particularly for low frequency and rare variants down to MAFs of 0.1%–0.5%.[12,13]

Herein, we report the results from our African Ancestry Anthropometry Genetics Consortium (AAAGC) meta-analysis of height associated with variants imputed to the 1000 Genomes reference panel in up to 52,764 individuals of African ancestry. We aimed to (1) discover novel variants, (2) fine-map established loci, and (3) evaluate the coverage and contribution of variants in genetic associations to height in populations of African ancestry.

Subjects and methods

Study design

We used a three-stage design to evaluate genetic association with height in sex-combined and sex-stratified samples (Figure 1). In stage 1, GWAS results from 17 studies including 41,400 individuals (16,032 men and 25,368 women) of African ancestry (AA), most of whom were African American, were meta-analyzed. In stage 2, we took forward variants with p < 1E−4 in either the sex-combined or the sex-stratified meta-analyses from stage 1, for a meta analysis with 11,364 (2,915 men and 8,449 women) additional AA individuals. In stage 3, we meta-analyzed stage 1 and stage 2 results with 253,288 individuals of European ancestry (EA) from the Genetic Investigation of Anthropometric Traits (GIANT) consortium. Variants that reached genome-wide significance (p < 5E−8) in either stage 1 and stage 2 or stages 1, 2, and 3 were assessed for associations in two AA pediatric cohorts (N = 7,064). All AA participants in these studies provided written informed consent for the research, and approval for the study was obtained from the ethics review boards at all participating institutions. Detailed descriptions of each participating study and measurement and collection of height and age are provided in Tables S1, S2, and S16.

Genotyping, imputation, and quality control

Genotyping in each study was performed with Illumina or Affymetrix genome-wide SNP arrays. Pre-phasing and imputation of missing genotypes in each study was performed using MaCH/minimac[14] or SHAPEIT2/IMPUTEv2[12,15] using the 1000 Genomes Project cosmopolitan reference panel (Phase I Integrated Release v.3, March 2012).[10] The details of the array, genotyping, and imputation quality-control procedures and sample exclusions for each study are listed in Table S3. Samples reflecting duplicates, low call rates, gender mismatch, or population outliers were excluded. Variants were excluded by the following criteria: call rate < 0.95, minor allele count (MAC) ≤ 6, Hardy-Weinberg Equilibrium (HWE) p < 1E−4, imputation quality score < 0.3 for minimac or < 0.4 for IMPUTE, or absolute allele frequency difference > 0.3 compared with expected allele frequency (calculated as 1000 Genomes frequency of AFR × 0.8 + EUR × 0.2).

We note as a limitation that genotyping arrays used in the studies for these analyses were designed primarily for European ancestry sample and do not comparably tag variation in individuals of African ancestry.

Study-level association analyses

At all stages, GWASs were performed by each of the participating studies. Height was regressed on age, age squared, principal components (PCs), and study site (if needed) to obtain residuals, separately by sex and case-control status, if needed. PCs were included to adjust for admixture proportion and population structure within each study. Residuals were inverse-normally transformed to obtain a standard normal distribution with a mean of zero and standard deviation of one. For studies with unrelated subjects, each variant was tested assuming an additive genetic model with each trait by regressing the transformed residuals on the number of copies of the variant effect allele. For studies that included related individuals, association tests were conducted that took into consideration the genetic relationships among the individuals by a linear mixed model with genetic relationship matrix as random effect which controls for population structure and cryptic relatedness (see Table S3). Sex-stratified, case/control-stratified, and combined analyses were performed. Association results with extreme values (absolute beta coefficient or standard error $\geq$ 10), primarily due to small sample sizes and/or low minor allele count, were excluded from meta-analysis. EasyQC (see web resources) was used to perform quality control on all study-specific results.

Imputation of European GWAS summary statistics to 1000 Genomes

The latest summary statistics of sex-combined meta-analyses of height imputed to the HapMap reference panel in EA from the Genetic Investigation of Anthropometric Traits (GIANT) consortium were obtained (see web resources). These association summary statistics were used to impute z-scores of unobserved variants at the 1000 Genomes Project EUR reference panel (Phase I Integrated Release v.3) using the ImpG program. In brief, palindromic variants (AT/CG) and variants with allele mismatch with the reference were removed from the data. Using the ImpG-Summary method, the z-score of an unobserved variant was calculated as a linear combination of observed z-scores weighted by the variance-covariance matrix between variants induced by LD within a 1 Mb window from the reference haplotypes. The sample size of each unobserved variant was also interpolated from the sample sizes of observed variants using the same weighting method for z-score as $N_i = \sum_{t=1}^{t=T} \frac{|w_{i,t}|}{\sum |w_{i,t}|} N_t$ Here, t = 1,2,….,T, where T is the number of observed variants and $w_{i,t}$ is the element of the covariance matrix $S_{i,t}$ for the unobserved variant i and the observed variant t within window. The performance of imputation was assessed by r^2pred, with similar characteristics as the standard imputation accuracy metric r^2hat. Results of variants with $r^2pred \geq 0.6$ were used in subsequent analyses.

Meta-analysis

In the discovery stage 1, association results were combined across studies in sex-stratified and sex-combined samples using inverse-variance weighted fixed-effect meta-analysis implemented in the program METAL. The study-specific l values of association ranged from 0.97 to 1.11 for height (Table S3). Genomic control correction was applied to each study before meta-analysis, and to the overall results after meta-analysis (l = 1.00 for height, Figures S1 and S2). Variants with results generated from <50% of the total sample size for each trait were excluded. After filtering, the numbers of variants reported in the meta-analyses were 17,972,087. EasyStrata (see web resources) was used to perform quality control on the meta-anlaysis results and create manhattan and QQ plots.

Variants with p < 1E−4 in stage 1 sex-stratified or sex-combined meta-analyses were carried forward for replication in additional AA individuals (stage 2) and EA individuals (stage 3). For each of the replication AA studies, association analyses with height (inverse-normally transformed residuals of height) were performed as in stage 1 and results were meta-analyzed using the inverse-variance method in METAL. For the replication study in EA, HapMap-imputed summary statistics from the GIANT consortium were used to impute z-scores of unobserved variants at the 1000 Genomes.

Variants taken forward from stage 1 were meta analyzed with the samples from stage 2, using the inverse-variance weighted method. In stage 3, meta-analysis results were expressed as signed z-scores using the fixed effect sample size weighted method in METAL, due to the lack of beta and standard error estimates from the ImpG program. Evidence of heterogeneity of allelic effects between males and females, within and across stages were assessed by the I^2 statistic in METAL. Variants that reached genome-wide significance (p < 5E−8) in either the sex-stratified and sex-combined meta-analysis including AA and/or combined AA and EA individuals were considered our main study results. For comparison purposes between the lead EA results with the AA stage 1 and 2 results, we calculate z-scores from effect size and standard error, $Z = \beta/SE$. For lead variants, differences in the magnitude of effects between men and women were assessed using Cochran's Q test and with a p value [HetPval] < 0.001 was declared significant based on Bonferroni correction. As a sensitivity analysis for any heterogeneity between AA studies in stages 1 and 2 results, we also ran a meta-analysis of all lead variants by entering all studies separately. We defined evidence of moderate to high heterogeneity from Cochran's Q test with a p value [HetPval] < 0.001 or the I-square (HetISq) statistic > 50%.[16]

A lead variant in a locus was defined as the most significant variant within a 1 Mb region. A novel locus was defined as a lead variant with distance > 500 kb from any established lead variant reported in previous studies. By convention, a locus was named by the closest gene(s) to the lead variant.

Variance explained

For lead genome-wide significant variants within a locus, we calculated the variance explained for stage 1 and stage 2 meta-analysis results using the equation $\beta^2(1-f)2f$ where β is the effect size and f is the effect allele frequency. Allele frequencies were based on the combined frequency for stage 1 + stage 2. The effect sizes used came from stage 2 in one calculation (Table 1) and using the implementation of winner's curse correction as described in Zhong and Prentice[17] and Palmer and Pe'er[18] (Table S4b).

Meta-analysis of lead variants in pediatric cohorts

Two pediatric cohorts, the Children's Hospital of Philadelphia's Center for Applied Genomics (CHOP/CAG) and Bone Mineral Density in Childhood Study (BMDCS), provided results for variants reaching p < 1E−4 in stage 1 of height-for-age z scores. Results from these two studies were meta-analyzed together using the inverse-variance method in METAL. For the lead variants by locus that reached genome-wide significance in stage 1 + stage 2, we ran analyses by pubertal status in the CHOP/CAG pediatric cohort, the larger of the two pediatric cohorts. Pre-pubertal was defined as

Table 1. Lead variant in novel and previously identified height loci at p < 5E−8 in African ancestry stage 1 and stage 2 samples, and European ancestry samples

							AA Stage 1 + Stage 2				EA			Stages 1+2+3: AA + EA[d]		Stage 2	
Lead SNP	Chr	Position (b37/ hg19)	Known locus (if yes, lead published variant)	Known signal in known locus[a]	Locus[b]	Effect/ other alleles	EAF	Effect (SE)	Z-score[c]	p	n	Z-score[c]	p	n	p	Total sample size	Effect	Variance explained (%)[e]
rs3176468	1	51,438,881	yes rs12855	no	*CDKN2C*	C/T	0.203	0.058 (0.008)	7.25	1.39E−12	52,763	7.10	8.79E−13	246,918	2.839E−21	299,681	0.026	0.022
rs3770820	2	36,763,620	yes rs711245	yes	*CRIM1*	T/C	0.231	0.044 (0.008)	5.50	3.99E−08	52,764	4.50	5.63E−06	238,723	1.157E−10	291,487	0.043	0.065
rs58680090	2	56,080,379	yes rs1367226	yes	*EFEMP1*	T/A	0.931	0.081 (0.013)	6.23	1.56E−09	52,764	15.10	2.44E−51	247,170	3.917E−59	299,934	0.105	0.131
rs10180829	2	128,931,951	yes rs744265	no	*UGGT1*	T/C	0.429	0.032 (0.007)	4.57	1.68E−06	52,764	5.20	2.60E−07	243,301	2.358E−11	296,065	0.037	0.067
rs2553033	2	218,100,996	yes rs17181956	no	*DIRC3*	G/A	0.171	0.061 (0.009)	6.78	8.48E−12	52,765	1.60	1.13E−01	73,838	1.792E−08	126,603	0.078	0.159
rs11677783	2	220,706,985	no N/A	no	*SLC4A3/ MIR4268*	T/A	0.289	0.051 (0.008)	6.38	1.64E−10	49,507	−0.50	6.25E−01	239,545	0.0275	289,052	0.044	0.078
rs6431539	2	238,360,708	yes rs6719451	yes	*COL6A3/ MLPH*	G/T	0.191	0.029 (0.009)	3.22	5.19E−04	52,763	4.60	4.32E−06	236,261	1.896E−08	289,024	−0.010	0.003
rs4681725	3	56,692,321	yes rs9835332	yes	*FAM208A*	T/G	0.224	0.046 (0.008)	5.75	3.80E−09	52,694	9.00	2.26E−19	253,165	2.005E−26	305,859	0.032	0.034
rs200396883	3	58,031,200	yes rs1658351	no	*FLNB*	i/d	0.830	0.056 (0.01)	5.60	3.88E−08	44,714	N/A	N/A	N/A	N/A	N/A	0.049	0.068
rs6785012	3	141,109,348	yes rs724016	yes	*ZBTB38*	T/C	0.668	0.062 (0.009)	6.89	6.93E−13	51,945	26.10	1.91E−150	237,535	2.98E−157	289,480	0.050	0.069
rs7652177	3	171,969,077	yes rs7652177	yes	*FNDC3B*	G/C	0.856	0.054 (0.01)	5.40	2.27E−08	51,944	12.80	2.94E−37	224,593	5.636E−44	276,537	0.033	0.024
rs925098	4	17,919,811	yes rs7692995	yes	*LCORL*	G/A	0.350	0.05 (0.007)	7.14	2.99E−13	52,675	14.50	6.24E−48	252,926	1.867E−59	305,601	0.035	0.056
rs1662837	4	82,168,889	yes rs17556750	yes	*PRKG2*	C/T	0.731	0.047 (0.008)	5.88	2.49E−10	52,763	14.10	6.46E−45	253,076	1.035E−53	305,839	0.049	0.091
rs112226333	5	31,525,207	yes rs17410035	yes	*DROSHA*	T/G	0.038	0.108 (0.02)	5.40	3.25E−08	50,697	4.10	4.30E−05	235,546	1.75E−09	286,243	0.083	0.037
rs10071837	5	33,381,581	yes rs11745439	yes	*TARS*	C/T	0.575	0.041 (0.007)	5.86	1.12E−09	52,695	6.40	1.67E−10	252,200	7.583E−17	304,895	0.050	0.123
rs1150781	6	34,214,322	yes rs12214804	no	*C6orf1*	C/G	0.436	0.052 (0.007)	7.43	2.02E−13	50,788	13.90	3.71E−44	229,929	9.237E−56	280,717	0.041	0.083
rs12332985	6	35,278,924	yes rs6899744	no	*DEF6*	C/A	0.825	0.071 (0.009)	7.89	1.66E−15	51,752	N/A	N/A	N/A	N/A	N/A	0.045	0.060
rs148342137	6	36,010,674	yes rs4713902	no	*MAPK14*	I/D	0.753	0.051 (0.009)	5.67	3.72E−09	43,791	N/A	N/A	N/A	N/A	N/A	0.000	0.000
rs7742789	6	43,345,803	yes rs2242416	no	*ZNF318*	C/T	0.329	0.028 (0.007)	4.00	7.10E−05	52,697	5.30	1.16E−07	230,890	8.447E−11	283,587	0.009	0.004
rs2071454	6	152,126,824	yes rs6902771	no	*ESR1*	G/T	0.398	0.044 (0.007)	6.29	1.56E−10	52,763	6.10	1.43E−09	236,456	2.017E−16	289,219	0.014	0.009
rs6463331	7	46,532,407	yes rs6949739	no	*IGFBP3/ TNS3*	C/T	0.797	0.049 (0.008)	6.13	4.01E−09	52,764	2.50	1.24E−02	242,574	1.906E−06	295,338	0.021	0.015
rs2926701	8	71,170,604	no N/A	no	*NCOA2*	C/T	0.366	0.04 (0.007)	5.71	9.41E−09	52,764	−0.40	7.26E−01	252,246	0.03846	305,010	0.028	0.037
rs7905296	10	74,918,196	no N/A	no	*ECD*	C/A	0.175	0.057 (0.009)	6.33	8.23E−11	51,741	N/A	N/A	N/A	N/A	N/A	0.040	0.049
rs941873	10	81,139,462	yes rs1923367	yes	*ZCCHC24*	G/A	0.589	0.044 (0.007)	6.29	6.91E−10	52,753	10.00	1.52E−23	251,172	2.126E−31	303,925	0.030	0.044
rs634552	11	75,282,052	yes rs606452	yes	*SERPINH1*	T/G	0.363	0.055 (0.007)	7.86	3.61E−15	52,764	9.80	1.40E−22	236,321	3.162E−34	289,085	0.034	0.054

(*Continued on next page*)

Table 1. ***Continued***

Lead SNP	Chr	Position (b37/ hg19)	Known locus (if yes, lead published variant)	Known signal in known locus[a]	Locus[b]	Effect/ other alleles	AA Stage 1 + Stage 2					EA			Stages 1+2+3: AA + EA[d]		Stage 2	
							EAF	Effect (SE)	Z-score[c]	p	n	Z-score[c]	p	n	p	Total sample size	Effect	Variance explained (%)[e]
rs79241096	12	14,503,656	yes rs12228415	yes	*ATF7IP*	T/C	0.680	0.041 (0.007)	5.86	2.20E−08	52,763	2.10	3.86E−02	235,100	1.823E−05	287,863	0.029	0.037
rs12307687	12	47,175,866	yes rs10880969	no	*SLC38A4*	T/A	0.245	0.045 (0.008)	5.63	1.00E−08	51,752	N/A	N/A	N/A	N/A	N/A	0.045	0.078
rs2070808	12	66,217,872	yes rs8756	yes	*RPSAP52*	T/A	0.675	0.053 (0.008)	6.63	2.03E−12	46,247	1.70	9.49E−02	234,123	1.117E−05	280,370	0.038	0.065
rs11107175	12	94,161,719	yes rs10859567	no	*CRADD*	C/T	0.891	0.062 (0.011)	5.64	1.47E−08	51,753	N/A	N/A	N/A	N/A	N/A	0.077	0.112
rs75823898	13	50,669,173	yes rs2687950	no	*DLEU1/ DLEU2*	A/C	0.027	0.203 (0.022)	9.23	4.70E−21	51,753	N/A	N/A	N/A	N/A	N/A	0.183	0.185
rs4899520	14	74,987,572	yes rs862034	yes	*LTBP2*	A/G	0.583	0.044 (0.007)	6.29	1.06E−10	52,764	8.90	7.98E−19	241,593	4.88E−27	294,357	0.023	0.026
rs3917155	14	76,444,685	yes rs2303345	no	*TGFB3*	G/C	0.949	0.103 (0.016)	6.44	1.65E−10	51,753	N/A	N/A	N/A	N/A	N/A	0.137	0.188
rs28566535	15	51,601,141	yes rs16964211	yes	*CYP19A1*	A/C	0.516	0.041 (0.007)	5.86	6.84E−10	52,765	6.80	1.00E−11	244,124	1.972E−18	296,889	0.063	0.198
rs12904319	15	75,816,649	yes rs4886707	yes	*PTPN9*	C/A	0.066	0.073 (0.015)	4.87	2.04E−06	49,704	4.40	1.34E−05	245,002	2.702E−09	294,706	−0.028	0.009
rs1600640	15	84,603,034	yes rs7162542	yes	*ADAMTSL3*	G/T	0.822	0.052 (0.009)	5.78	2.58E−09	52,763	6.40	1.97E−10	252,729	1.301E−16	305,492	0.047	0.064
rs146576224	15	89,387,846	yes rs16942341	yes	*ACAN*	C/G	0.882	0.078 (0.011)	7.09	1.91E−13	51,752	N/A	N/A	N/A	N/A	N/A	0.052	0.059
rs10852140	15	91,500,296	no N/A	no	*RCCD1*	T/C	0.174	0.047 (0.009)	5.22	2.31E−07	52,763	4.10	3.67E−05	234,866	2.669E−09	287,629	0.066	0.114
rs2871865	15	99,194,896	yes rs2871865	yes	*IGF1R*	C/G	0.576	0.047 (0.007)	6.71	1.62E−10	51,029	11.30	1.66E−29	238,470	2.686E−38	289,499	0.037	0.067
rs228758	17	42,148,205	no N/A	no	*G6PC3*	C/T	0.877	0.049 (0.011)	4.45	3.61E−06	52,764	4.00	6.33E−05	253,102	2.709E−08	305,866	0.042	0.031
rs113229779	17	45,398,018	yes rs80267077	yes	*ITGB3/ EFCAB13*	T/C	0.953	0.095 (0.023)	4.13	3.44E−05	40,759	6.00	2.27E−09	243,679	1.237E−12	284,438	0.091	0.056
rs113121081	17	59,575,304	yes rs2378870	yes	*TBX4/ NACA2*	A/G	0.194	0.064 (0.009)	7.11	1.93E−13	52,763	3.30	1.15E−03	229,134	8.509E−10	281,897	0.049	0.078
rs2955250	17	61,959,740	yes rs2854207	yes	*GH2*	T/C	0.704	0.059 (0.007)	8.43	1.17E−15	52,758	13.10	3.67E−39	243,859	1.369E−52	296,617	0.053	0.113
rs8082122	17	62,534,459	no N/A	no	*CEP95*	C/T	0.695	0.033 (0.007)	4.71	1.07E−05	52,764	4.00	6.54E−05	236,486	3.849E−08	289,250	0.030	0.036
rs357900	18	46,585,235	yes rs12458127	no	*DYM*	A/T	0.366	0.04 (0.007)	5.71	4.02E−09	52,764	8.10	6.02E−16	240,293	8.968E−23	293,057	0.005	0.001
rs224333	20	34,023,962	yes rs143384	yes	*GDF5*	A/G	0.858	0.055 (0.01)	5.50	3.42E−08	52,763	22.50	4.15E−112	250,545	2.31E−114	303,308	0.035	0.026

AA, African ancestry; EA, European ancestry; EAF, effect allele frequency; HetISq, heterogeneity measured by I-square; SE, standard error.
[a]The results of the conditional analysis for the tested variants on published variants and other variants in LD with published variants in known loci are shown in Table S8.
[b]Locus is the nearest gene or previous reported locus.
[c]Z-scores in AA Stage 1 + Stage 2 are calculated for each variant as Z-score = Effect/SE to use as a comparison with the EA z-scores. Z-scores from the EA as a linear combination of observed z-scores weighted by the variance-covariance matrix between variants induced by LD within a 1 Mb window from the reference haplotypes. (based on the ImpG-Summary Method).
[d]Previously identified loci with $p < 5E{-}8$ in the combined African and European ancestry samples were not shown.
[e]The variance explained for each variant is calculated from the variant effect size (b) and effect allele frequency (f) as follows: b2(1 - f)2f. We used the effect sizes and the effect allele frequency from AA stage 2.

<12 years in boys and <11 years in girls, while post-pubertal was defined as 12–18 years in boys and 11–18 years in girls. By meta-analyzing respective combinations of these strata using METAL and calculating Cochran's Q-test for heterogeneity and I-square, we looked for "considerable" heterogeneity (as defined by Deeks et al.[16]) between pre- and post-pubertal status and between girls and boys defined as I-Square [HetISq] > 75% and p [HetPVal] < 0.05. We also estimate the regression slope from variant effects in children and the variant effects of stage 1 + stage 2 after correcting for winner's curse.

Conditional and joint analyses of summary statistics

For the genome-wide significant loci identified in the sex-combined meta-analyses in AA (stages 1+2), we used GCTA[19,20] to select the top independent associated variants. This method uses the LD correlations between variants estimated from a reference sample to perform an approximate conditional association analysis. We used 8,054 unrelated individuals of African ancestry from the WHI cohort with ~15.7M variants available as the reference sample for LD estimation. To select the top independent variants in the discovery and replication meta-analysis results, we first selected all variants that had p < 5E−8 and conducted analysis conditioning on these selected variants to search for additional variants iteratively via a stepwise model. This was serially conducted until no variant had a conditional p value that passed the significance level p < 5E−8. We used default settings in GCTA for the following: (1) allowable differences in alleles frequencies up to 0.2 between the meta-analysis and the LD reference, (2) the distance of 10 Mb for which LD is considered, and (3) a collinearity cut-off of 0.9 between variants tested.

We also tested whether the genome-wide significant variants identified from sex-combined GWASs in AA and the locus-wide significant variants identified from sex-combined locus transferability studies in AA were independent from nearby established loci identified from EA studies.[1,6,21] First, the published lead variants from EA studies were used to search for all surrogate variants that were in high LD ($r^2 > 0.8$ in 1000 Genomes Project EUR population). Second, these variants were pruned to select only variants in low LD in AA ($r^2 < 0.3$ in the 1000 Genomes Project AFR population) to avoid collinearity in conditional analysis. Third, association analysis was conducted on the AA significant variants conditioned on the selected EA lead and surrogate variants, using the program GCTA and estimated LD correlation from the WHI cohort. For genome-wide significant loci, an AA derived association signal is considered as independent from the established EA signals when the difference in $-\log p < 2.5$ and difference in effect size < 1 standard error after conditional analysis. For locus-wide significant loci, given the lower level of significance, independence is only considered as difference in effect size < 1 standard error after conditional analysis.

SNP and locus transferability analyses

We investigated the transferability of EA height-associated variants and loci in AA individuals using the stage 1 sex-combined meta-analyses. First, we tested for replication of lead variants previously reported to be associated with height (802 lead signals from 627 loci) at genome-wide significance in sex-combined analyses from the GIANT consortium studies. We defined SNP transferability as an EA lead variant sharing the same trait-raising allele and p < 0.05 in AA individuals. To account for differences in local LD structure across populations, we also interrogated the flanking 0.1 M regions of the lead variants to search for the best variants with the smallest association p in AA individuals. Locus-wide significance was declared as $p_{locus} < 0.05$ by Bonferroni correction for the effective number of tests within a locus, estimated using the Li and Ji approach.[22]

Fine-mapping analyses

We compared the credible set intervals of established loci that showed locus-wide significance ($p_{locus} < 0.05$) from this study in summary statistics datasets including the 1000 Genomes imputed results from GIANT, AAAGC, and meta-analysis of GIANT and AAAGC. In each dataset, a candidate region is defined as the flanking 0.1 M region of the lead variant reported by the GIANT consortium. Under the assumption of one causal variant in a region of M variants, the posterior probability of a variant j with association statistics Z driving the association, $P(C_j|Z)$, was calculated using the formula $P(C_j|Z) = \frac{\exp\left(\frac{1}{2}z_j^2\right)}{\sum_{i=1}^{M}\exp\left(\frac{1}{2}z_j^2\right)}$. A 99% credible set was constructed by ranking all variants by their posterior probability, followed by adding variants until the credible set has a cumulative posterior probability > 0.99.[23] The posterior probability of a variant depends on the relative z-score of this variant against all other variants. Variants in high LD will have similar z-scores and similar posterior probability. A locus with a causal variant that is not well tagged will have higher posterior probability than a locus with a causal variant that is tagged by many nearby variants.

Bioinformatics

Functional annotation of novel variants

To determine whether any of our GWAS lead variants in new loci or new signals in known loci identified in the sex-specific and sex-combined analyses might be tagging potentially functional variants, we identified all variants within 1 Mb and in LD ($r^2 > 0.7$, 1000 Genomes AFR) with our lead variants. As such, we identified variants and annotated each of them using ANNOVAR[24] and Haploreg, v.4.[25] The predicted functional impact for coding variants were assessed via the Exome Variant Server (see web resources) for PhastCons,[26] GERP,[27] and PolyPhen,[28] as well as SIFT.[29]

We further characterized the variants that were in LD with the novel variants using the web-based tool RegulomeDB.[30] The variants that were likely to affect binding and linked to expression of a gene target (scores 1a-1f) based on "eQTL, transcription factor (TF) binding, matched TF motif, matched DNase footprint, and DNase peak" or were only likely to affect binding (scores 2a-2c) based on "TF binding, matched TF motif, matched DNase footprint, and DNase peak" were selected. For these variants, the sequence conservation (GERP and SiPhy), the epigenomic data from the Roadmap Epigenomic project (ChromHMM states corresponding to enhancer or promoter elements, histone modification ChIP-seq peaks, and DNase hypersensitivity data peaks), the transcription factor binding and motif data from the ENCODE project and the eQTLs from Genotype-Tissue Expression (GTEx v6) project were extracted from web-based HaploReg v.4 and listed in Table S12. For variants within the tractable credible sets in the fine mapping analyses, similar analyses were also conducted.

Cross-trait associations

To assess whether the novel loci identified in the sex-specific and sex-combined analyses were associated with any related cardiometabolic and anthropometric traits, or may be in high LD with

known eQTLs, we examined the NHGRI-EBI GWAS Catalog and the GRASP (Genome-Wide Repository of Associations Between SNPs and Phenotypes) catalog for reported variant-trait associations near our lead variants. We supplemented the catalogs with additional genome-wide significant associations of interest from the literature PMID. We used PLINK to identify variants within 1 Mb of lead variants. All variants within the specified regions with $r^2 > 0.7$ (1000 Genomes AFR) were retained from the catalogs for further evaluation.

Power analysis

Given our sample sizes in stage 1 and stage 2 of our AA populations, we estimated >80% power to detect variants explaining 0.08% variance for height, which corresponds to effect sizes of 0.09 and 0.20 SD units for MAF of 0.05 and 0.01, respectively. Effect sizes > 0.1 SD units are less likely suggesting that we are not well powered to detect variants below MAF of 0.05. Power analyses are shown in Figure S3 based on a sample size of 50,000 (the sample size of stage 1 + stage 2 results) at genome-wide significance, or to validate in the sample of children with a sample size of 7,000 at nominal significance ($p < 0.05$) at varying minor allele frequencies.

Pathway enrichment analyses: DEPICT

DEPICT is a gene set enrichment analysis method for GWAS data, originally designed for analysis of European-ancestry samples.[31] Its primary innovation is the use of "reconstituted" gene sets, where many different types of gene sets (e.g., canonical pathways, protein-protein interaction networks, and mouse phenotypes) were extended through the use of large-scale microarray data (see Pers et al.[31] for details). We adapted DEPICT for use with African ancestry results by using 1000 Genomes phase 3 samples of west African ancestry, including ESN (Esan in Nigeria), GWD (Gambian in Western Divisions in the Gambia), MSL (Mende in Sierra Leone), and YRI (Yoruban in Nigeria) (total N = 405). We used these samples as a reference panel for (1) clumping the input GWAS loci and (2) defining locus boundaries to include all SNPs with an $r^2 > .5$ to each index SNP. Genes within or overlapping these boundaries were included in the analysis. For p value calculation, we generated 500 new "null" GWASs based on 2,098 unrelated African-Americans from the Atherosclerosis Risk in Communities (ARIC) cohort[32] (using 500 sets of normally distributed "null" phenotypes). We note that the AA GWAS data analyzed here contains some European admixture, which our DEPICT reference data does not; therefore, while our approach is a substantial improvement over using the European default reference data, it will not have perfect accuracy and should be considered only as an approximation. We also defined "meta-gene sets" by using affinity propagation clustering[33] to group the most similar reconstituted gene sets and choose one representative gene set for each one, which are reported in Table S15 (for more details, see Marouli et al.[6]).

Trans-ethnic findings to account for population structure in previous GWASs

We first conducted principal component 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 demography history.[34] We only used bi-allelic SNPs with MAF > 0.05 in the four European populations, and then pruned them by both distance and linkage disequilibrium (LD) using plink 1.9.[35] Specifically, we pruned the dataset such that no two SNPs were closer than 2 kb, and then pruned in LD in windows of 50 SNPs, moving in steps of 5 variants, such that no two SNPs had $r^2 > 0.2$. We further removed SNPs in regions of long-range LD.[36] Principal components analysis was performed on the remaining SNPs using Eigensoft v.7.2.1 (see web resources).

To measure the impact of uncorrected stratification on estimated effect sizes, we computed the correlation between principal component (PC) loadings and beta effects estimated from each GWAS, i.e., GIANT, AAAGC (stage 1), and GIANT+AAAGC trans-ethnic meta-analysis. We performed linear regressions of individual PC value 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 > 0.01) from each GWAS panel. We estimated p values based on Jackknife standard errors by splitting the genome into 1,000 blocks with an equal number of variants. If there is significant correlation in either the GIANT or the AAAGC meta-analysis, we then further evaluated the improvement of bias due to stratification in the trans-ethnic meta-analysis (GIANT+ AAAGC) by comparing the correlation coefficients in the trans-ethnic meta-analysis with those in GIANT. Restricting to variants shared between GIANT and stage 1 AAAGC 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

Study overview

We conducted sex-combined and sex-stratified meta-analyses of GWAS summary statistics for height across 17 studies of 41,401 individuals (16,032 men and 25,368 women) in AA individuals in stage 1 discovery (Tables S1 and S2). Among all variants with MAF ≥ 0.1% in the largest study (Women's Health Initiative [WHI]), the average info score was 0.81, and 90.5% had imputation info score ≥ 0.3.[15] Genomic control corrections were applied to each study and after meta-analysis ($\lambda = 1.09$) (Table S3). Association results for ~18M variants were subsequently interrogated further.

From stage 1 meta-analyses, variants at $p < 1E{-}4$ (9,872 in all, 3,018 in men, 5,725 in women) were carried forward for replication in AA (stage 2) and EA individuals (stage 3). Stage 2 included 11,364 AA (2,915 men and 8,449 women). Stage 3 included 253,288 EA individuals by imputing HapMap summary statistics results by Wood et al.[1] to 1000 Genomes.[16] Meta-analyses were performed to combine results from AA individuals (stage 1 + stage 2, N ≤ 59,475 in sex-combined analyses) and both AA and EA individuals (stage 1 + stage 2 + stage 3, N ≤ 312,204 in sex-combined analyses). Variants that reached genome-wide statistical significance ($p < 5E{-}8$) were assessed for generalization of associations with height to children in two additional AA cohorts (N = 7,064).

Genome-wide significant loci in meta-analyses

Sex-combined analyses

In the sex-combined meta-analysis of height in AA individuals (stage 1 + stage 2), 39 previously established

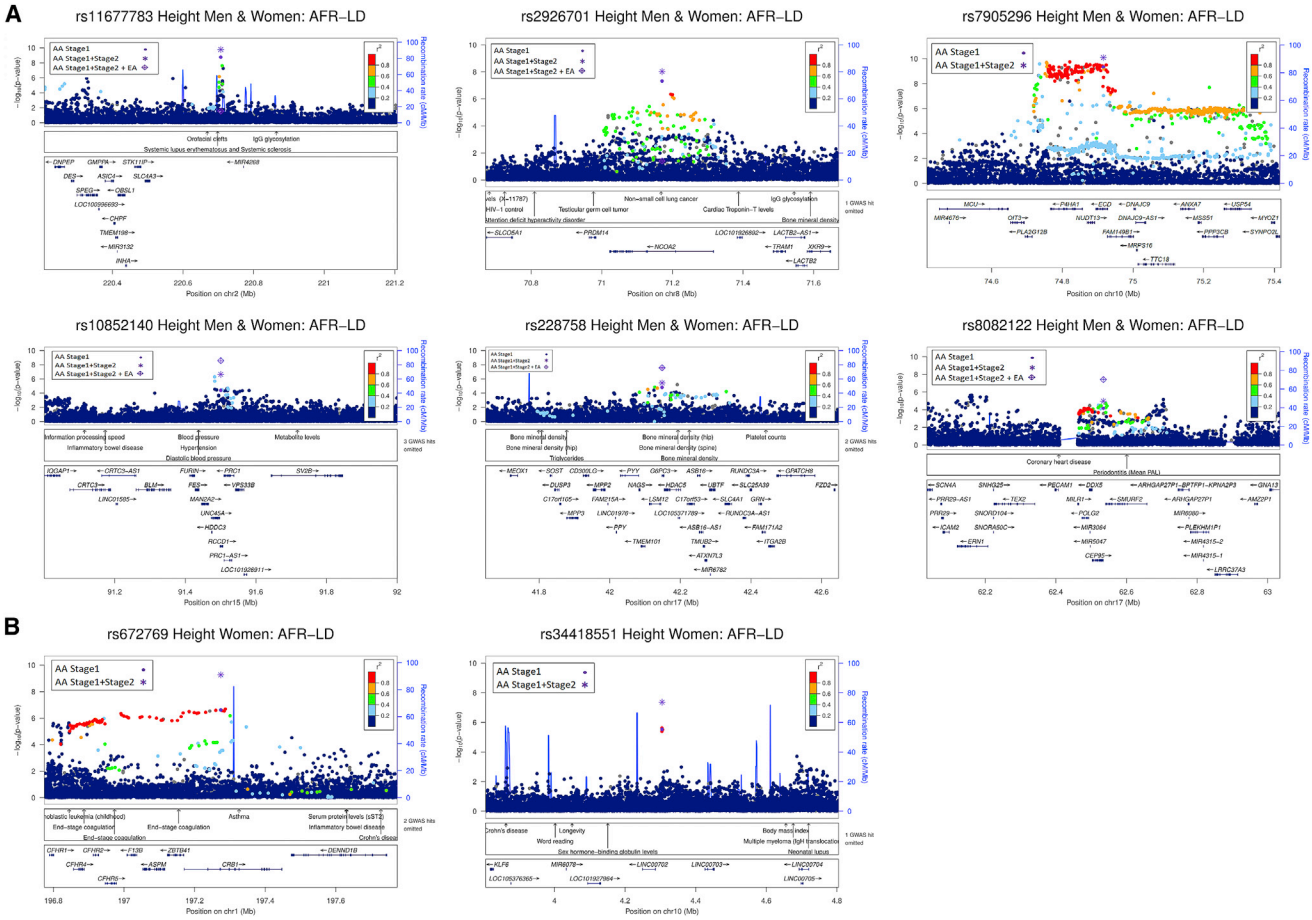

Figure 2. Locuszoom plots of six novel height loci
SLC4A3/MIR4268, NCOA2, ECD, RCCD1, G6PC3, and *CEP95* in men and women combined (A) and *CRB1* and *KLF6/LINC00704* in women only (B). All plots use AFR LD from the 1000 Genomes phase 1 reference panel. In each plot, the most significant variant within a 1 Mb regional locus is highlighted. p values for all variants including the most significant variant are based on the African American discovery phase only (AA Stage1). In addition, for the most significant variant, p values are annotated and illustrated from the African American discovery and replication phases (AA Stage1+Stage2). For loci *SLC4A3/MIR4268, NCOA2, RCCD1, G6PC3,* and *CEP95*, the lead SNP are also shown for the European ancestry from the GIANT consortium effort[1] combined with the African American discovery and replication phases (AA Stage1+Stage2 + EA).

European-derived loci reached genome-wide significance ($p < 5E{-}8$) (Tables 1 and S4, Figure S4). Three novel loci not previously identified in Europeans were found near *SLC4A3/MIR4268* (lead variant rs11677783 at chr 2:220,706,985), *NCOA2* (lead variant rs2926701 at chr 8:71,170,604), and *ECD/FAM149B1* (lead variant rs7905296 at chr 10:74,918,196) (Figure 2). In the trans-ethnic meta-analyses (stage 3), three new loci were identified including *RCCD1* (lead variant rs10852140 at chr 15:91,500,296), *G6PC3* (lead variant rs228758 at chr 17:42,148,205), and *CEP95* (lead variant rs8082122 at chr 17:62,534,459). The 6 novel loci explained ~0.2 to 0.3% of the variance for height among AA individuals, and the 39 known height loci explained ~2.5% of the variance for height.

Using the AA only analyses (stage 1 + stage 2), we used conditional and joint association analyses to examine the genome-wide significant loci for secondary signals. We identified multiple secondary signals in five known loci: *TARS/NPR3, RPSAP52/MHGA2, DLEU1/DLEU2, ACAN,* and *IGF1R/ADAMTS17* (Table 3).

Sex-stratified analyses

In the sex-stratified meta-analysis in AA individuals (stage 1 + stage 2), two novel loci were observed for women only in *CRB1* and *KLF6* (Table 2, Figures 2 and S6), both of which were significantly different ($p < 0.001$) between men and women (Table S4). We also tested the lead variants of novel and previously known height loci that reached genome-wide significance in the sex-combined analyses for differences between men and women in the magnitude of effects. No differences between men and women for these lead variants reached Bonferroni-corrected significance. However, there were three loci, one that was novel in the sex-combined EA and AA analyses (*CEP95*) and two that were known (*FAM208A* and *MAPK14*), that displayed nominally significant differences between men and women in AA ($p_{het} < 0.05$) (Table S4). There were no novel loci found for men only (Table 2, Figure S5).

Table 2. Lead variant for additional novel height loci at p < 5E−8 in analyses of African ancestry stage 1 and stage 2 women only

rsid	Chr	Position (b37/ hg19)	Known signal in known locus[a]	Locus[b]	Effect/ other alleles	EAF	Stage	Effect (SE)	p	HetISq	n	Variance explained (%)[c]
rs672769	1	197,274,118	no	*CRB1*	T/C	0.978	stage 1	0.176 (0.033)	6.75E−08	18.4	25,368	0.216
							stage 2	0.224 (0.064)	4.34E−04	0	7,043	
							stage 1 + stage 2	0.187 (0.03)	5.38E−10	N/A	32,411	
rs34418551	10	4,304,458	no	*KLF6/ LINC00704*	C/G	0.902	stage 1	0.083 (0.017)	1.02E−06	0	25,368	0.128
							stage 2	0.085 (0.03)	4.05E−03	38.2	7,748	
							stage 1 + stage 2	0.084 (0.015)	4.27E−08	N/A	33,116	

AA, African ancestry; EAF, effect allele frequency; HetISq, heterogeneity measured by I-square; SE, standard error.
[a]Results of conditional analysis on published variants and other variants in LD with published variants in known loci are shown in Table S8.
[b]Locus is the nearest gene or previous reported locus.
[c]The variance explained for each variant is calculated from the variant effect size (b) and effect allele frequency (f) as follows: b2(1 - f)2f. We used the effect sizes and the effect allele frequency from AA stage 2.

Additional QC of meta-analysis results

Two studies, GeneSTAR and HyperGEN, had slightly elevated lambdas, 1.11 and 1.13, respectively (Table S3). In addition, four of the lead variants (three variants of sex-combined analyses in Tables 1 and S4a and one variant of women-only analyses Tables 2 and S4a) had effect sizes > 0.1 SD. Therefore, we looked for evidence of heterogeneity for the lead variants in Table 1 by running a meta-analysis of all stage 1 and 2 studies added individually in METAL (rather than meta-analyzing stage 1 results with stage 2 results). Based on a Bonferroni corrected p (HetPval) of < 0.001 or I-square (HetISq) > 50%, none of the variants showed any evidence of heterogeneity. For variants with effect sizes > 0.1 SD, we also looked at forest plots of study and meta-analysis effects to see if it appeared that any of the smaller studies were driving the associations (Figure S7). From observation it seemed that the meta results tended to be driven by studies with the larger sample sizes N > 1,500.

Replication in children

We evaluated the 45 sex-combined genome-wide significant height loci for associations in 7,064 AA children (3,494 boys and 3,570 girls). Thirty-four of 45 lead variants displayed directional consistency, and five of these, including *FLNB, PRKG2, DROSHA, MAPK14,* and *CEP95* (the latter is a novel locus), showed nominally significant associations (Table S5a), suggesting some support for a role of these loci in influencing height in AA children. Results for association analyses and tests for heterogeneity by pubertal status and sex from the CHOP/CAG pediatric cohort are provided in Table S5b. In some of the lead variants, we find evidence of high heterogeneity (defined by a heterogeneity I-square [HetIsq] > 75 and heterogeneity p value [HetPVal] < 0.05) between boys and girls and by pubertal status. The lead variant at the *COL6A3/MLPH* locus was different between pre-pubertal and post-pubertal children. Lead variants in the *DROSHA* and *TGFB3* loci displayed heterogeneity by sex (HetIsq = 83.1 and 86.5 and HetPVal = 0.015 and 0.0065, respectively). For the lead variant in the *TGFB3* locus, effects were restricted to prepubertal girls versus boys (HetIsq = 86 and HetPVal = 0.0074), while for the lead variant in the *DROSHA* locus, the lead variant was restricted to post-pubertal girls (HetIsq = 90.7 and HetPVal = 0.001). We found heterogeneity between pre- and post-pubertal girls for *ATF7IP* (HetIsq = 82.3 and HetPVal = 0.0176). Finally, the lead variant at locus RPSAP52 was heterogeneous by pubertal status in both girls (HetIsq = 78.3 and HetPVal = 0.0317) and boys (HetIsq = 75.4 and HetPVal = 0.0438).

The estimated regression slope from variant effects in children and the variant effects of stage 1 + stage 2, after correcting for winner's curse, showed no correlation with $R^2 = 0.035$ for pre-pubertal versus stages 1 + stage 2 and $R^2 = 0.0074$ for post-pubertal versus stage 1 + stage 2 (Figure S8).

Functional characterization of novel loci

We used multiple complementary approaches to elucidate the putative causal genes and/or variants associated with the eight novel height loci from the sex-combined and sex-stratified analyses, including annotating nearby coding variants, *cis*-expression quantitative trait loci (*cis*-eQTL) analyses, and functional regulatory genomic element analyses. We identified six putative coding variants in high LD ($r^2 > 0.7$) with three of the lead variants within the flanking 1 Mb-regions (*CRB1*/rs672769, *ECD/ FAM149B1*/rs7905296, and *CEP95*/rs8082122) (Table S7). Two of the variants, rs112230218 and rs113611857 (both in perfect LD, $r^2 = 1$, with rs672769), had PolyPhen2 and SIFT scores suggesting possible damaging impact. In addition, four variants in LD ($r^2 > 0.7$) in the same locus (rs113054309, rs78537329, rs58690198, rs78306439) are *cis*-eQTLs for *MRPS16, TTC18,* and *NUDT13* in several tissues (Table S12). Both *NUDT13* and *MRPS16* are involved in energy metabolism. Six variants in LD ($r^2 > 0.7$) with

rs8082122 are also *cis*-eQTLs for *MILR1* in blood (Table S12) including rs3744409, a missense variant. For the new signal in the known locus *MAPK14*, three variants in high LD with the lead variant, rs148342137, are *cis*-eQTLs for *SLC26A8* in several tissues. *SLC26A8* is an anion transporter, transferring a variety of monovalent and divalent anions between cells, including chloride, bicarbonate, sulfate, and oxalate. For another new signal in the known locus *ZNF318*, the variant rs1214759, which is in high LD with the lead variant rs7742789 ($r^2 = 0.99$), is a *cis*-eQTL for *ZNF318* in osteoclasts.

Cross-trait associations of novel loci

We searched the NHGRI-EBI GWAS[37] and Genome-Wide Repository of Associations Between SNPs and Phenotypes (GRASP)[38] catalogs to assess whether any of the eight novel lead variants were in high LD ($r^2 > 0.7$) with variants that were genome-wide significantly ($p < 5E{-}8$) or nominally ($p < 0.05$) associated with related anthropometric and cardiometabolic traits or gene expression in prior studies. We did not find results for the lead variants in the GWAS catalog, but we did find some results in GRASP and have listed all results down to $p < 0.05$ (Table S14). There is possibly some shared biology but also definitely some overlap by chance particularly because of the large number of height loci across the genome. We noted several variants in high LD with rs8082122 that are associated with height in women of African ancestry,[39] just slightly below the genome-wide significance threshold (at $p < 5E{-}7$). To get a sense of independence of the eight novel height SNPs with the SNPs in LD at $r^2 < 0.7$ from Table S14, we performed a conditional analysis of the lead SNPs in each locus with the other SNPs in LD in that locus. p values before and after conditioning (columns L and M, respectively) show dependence between SNPs within each locus. We also used the Open Targets Genetics resource to further interrogate the evidence for co-localization in the UKBB resource (see web resources). In contrast to our in-house conditional analyses, no evidence for co-localization of the 8 novel height loci (looking up the 8 novel lead SNPs and the SNPs in LD at $r^2 > 0.7$) was observed.

Evaluation of established European loci in African ancestry populations

Conditional analysis in GWAS loci

Among the 39 height loci that achieved genome-wide significance in AA that were previously reported in EA,[1,6,21] we tested whether the African-derived lead variants were independent of the reported European signals by conditioning on the European lead variants or their surrogates (Table S8). Five of the 39 known loci had more than one signal, while the other 34 loci had just one signal per locus. Among the five loci with multiple signals, we identified a total of 14 independent signals that reached genome-wide significance. Of these 14 signals, 7 were signals that have previously been reported and 7 were independent of published signals (Tables 3 and S8). Of the remaining 34 known loci that only had one signal per locus that reached genome-wide significance, 13 were independent of the previously published signals (Tables 1 and S8).

SNP transferability

We further examined all height loci identified from previous EA studies[1,6,21] in our AA data. Among 802 EA lead signals from 627 height loci, 643 variants displayed directionally consistent associations in our data, and 205 (~25%) of these were nominally significant at $p < 0.05$ ($p_{binomial} = 3.02 \times 10^{-84}$ among 802 variants) (Table S9). Among the 205 lead variants that were nominally significant and directionally consistent in AA, 58% and 59% of the effect sizes and allele frequencies, respectively, were larger in the EA than the AA populations. The correlations of both effect sizes and allele frequency of the transferable variants were high for allele frequencies but only moderate for effect sizes, 0.71 and 0.45, respectively (Figure S9). Only 25% of lead variants were transferable from EA to AA individuals, suggesting either that many loci are not implicated in AA populations or that population differences in LD mask the detection of associated variants in AA individuals. Those variants that were transferable explain relatively similar levels of variances in both populations.

Locus transferability

We further investigated locus transferability in EA loci derived from the sex-combined analyses by considering varying LD between EA and AA populations. Using our AA results, we conditioned each of 796 lead EA signals that could be tested (in 627 loci)[1,6,21] on the most significant variant within 0.1 cM from our AA sex-stratified and sex-combined data (Table S10). We found that 289 (36%) of the lead regional variants across 201 loci (these loci were further fine-mapped below) remained significant ($p_{locus} < 0.05$) after adjustment for the number of independent variants tested at each locus. Yet, only 46 (16%) and 81 (28%) of these 289 lead regional variants were in LD ($r^2 > 0.2$) with the EA height lead variants based on 1000 Genomes AFR and CEU LD, respectively. Using the conditional analyses of variants meeting genome-wide significance, we found that 19 of these 46 variants had <1 standard error decrease in effect sizes after conditional analyses, representing distinct association signals in AA populations (Table S10).

Fine mapping of novel AA loci and known EA loci that were generalizable to AA

We performed fine mapping to localize putative causal variants. We constructed 99% credible sets containing variants that jointly accounted for 99% posterior probability of driving the association in a locus using the sex-combined meta-analysis results from AA, EA, and combined ancestry (Table S11). A smaller number of variants in a credible set represents a higher resolution of fine mapping, and we considered a credible set containing ≤20 variants as "tractable" for follow up. We tested the 201 locus-wide significant established loci mentioned above (which included a total of 235 tractable sets; some loci had overlapping sets or more than one credible set) and 6 novel

Table 3. Height loci with multiple distinct association signals at $p_{cond} < 5E{-}8$ after conditional analysis in African ancestry stage 1 and stage 2 samples

Signal	SNP	Chr	Position (b37/ hg19)	Known locus (if yes, lead published variant)	Known signal in known locus[a]	Locus[b]	Function	Effect/ other alleles	EAF	n	Unconditioned AA stage 1+ 2		Conditioned AA stage 1 + 2[c]		Stage 2	Variance explained (%)[d]
											Effect (SE)	p	Effect (SE)	p	Effect	
primary signal	rs10071837	5	33,381,581	yes, rs11745439	yes	*TARS*	intergenic	C/T	0.578	52,695	0.041 (0.007)	1.12E−09	0.04 (0.007)	2.26E−09	0.050	0.121
2nd signal	rs3811968	5	32,765,489		no	*NPR3*	intronic	A/C	0.436	51,627	0.041 (0.007)	1.31E−09	0.053 (0.007)	2.73E−14	0.043	0.090
2nd signal	rs7727858	5	32,924,873		no	*LOC340113*	intergenic	A/G	0.606	52,692	0.03 (0.007)	6.26E−06	0.041 (0.007)	4.05E−09	0.004	0.001
primary signal	rs2070808	12	66,217,872	yes, rs8756	yes	*RPSAP52*	intronic	T/A	0.680	46,247	0.053 (0.008)	2.03E−12	0.062 (0.008)	9.57E−16	0.038	0.064
2nd signal	rs8756	12	66,359,752		yes	*HMGA2*	3′-UTR	C/A	0.409	47,066	0.039 (0.007)	4.05E−08	0.049 (0.007)	1.28E−11	0.024	0.027
primary signal	rs75823898	13	50,669,173	yes, rs2687950	no	*DLEU1/ DLEU2*	intronic	A/C	0.026	51,753	0.203 (0.022)	4.70E−21	0.208 (0.022)	6.63E−22	0.183	0.172
2nd signal	rs114656078	13	50,714,388		yes	*DLEU2*	intergenic	G/A	0.042	51,752	0.086 (0.017)	3.90E−07	0.093 (0.017)	4.73E−08	0.091	0.066
primary signal	rs146576224	15	89,387,846	yes, rs16942341	no	*ACAN*	intronic	C/G	0.884	51,752	0.078 (0.011)	1.91E−13	0.097 (0.011)	1.92E−19	0.052	0.055
2nd signal	rs4932426	15	89,349,539		yes		intronic	A/G	0.489	52,764	0.036 (0.007)	1.07E−07	0.037 (0.007)	4.44E−08	0.049	0.118
2nd signal	rs111680044	15	89,394,117		yes		intronic	G/A	0.897	51,752	0.048 (0.011)	1.73E−05	0.062 (0.011)	2.54E−08	0.033	0.020
2nd signal	rs80095362	15	89,397,640		no		intronic	G/A	0.934	51,752	0.079 (0.014)	1.95E−08	0.103 (0.014)	3.29E−13	0.085	0.090
2nd signal	rs34543273	15	89,402,227		no		synonymous	C/T	0.971	52,764	0.152 (0.021)	3.21E−13	0.158 (0.021)	4.51E−14	0.142	0.115
primary signal	rs2871865	15	99,194,896	yes, rs2871865	yes	*IGF1R*	intronic	C/G	0.580	51,029	0.047 (0.007)	1.62E−10	0.046 (0.007)	3.76E−10	0.037	0.067
2nd signal	rs2573652	15	100,514,614		no	*ADAMTS17*	missense	C/T	0.805	52,693	0.04 (0.008)	1.34E−06	0.047 (0.008)	1.81E−08	0.039	0.048

Chr, chromosome; EAF, effect allele frequency; n, sample size; SE, standard error

[a]Results of conditional analysis on published variants and other variants in LD with published variants in known loci are shown in Table S8.

[b]Locus is the nearest gene or previous reported locus.

[c]The SNPs were selected by an approximate conditional and joint multiple-SNP analysis (GCTA-COJO) of the summary statistics from the meta-analysis. The primary signal represents the most significant SNP within 1 Mb region, others are defined as secondary.

[d]The variance explained for each variant is calculated from the variant effect size (b) and effect allele frequency (f) as follows: b2(1 - f)2f. We used the effect sizes and the effect allele frequency from AA stage 2.

loci. The credible sets in the EA analyses were generally smaller than those in the AA analyses given their larger sample size. As compared to the EA analyses, the number of tractable loci in the meta-analyses of AA and EA individuals increased from 104 loci (including 125 sets) to 128 loci (including 148 sets). Of these 148 sets, 106 (in 99 loci) also contained fewer SNPs than in the EA credible set.

Among the 148 tractable lead sets, the lead variants in the combined ancestry analyses had posterior probability ≥ 0.95 in 23 height loci, including 28 total credible sets (*ACBD4, AXIN2, DNM3, EFEMP1, ENPP2, FBXW11, FGFR4, FKBP5, FNDC3B, GDF5, HHIP, HLA-C, HMGA2, IGF1R, LIN28B, LTBP1, MC4R, PML, PTCH1, PTPRG, TET2, ZBTB4, ZFAT*) (Table S11). We functionally characterized the variants within the tractable credible sets (Table S13) and report some of the more interesting findings here. For locus *EFEMP1*, the intronic variant, rs3791675 (posterior probability = 0.97) is a *cis*-eQTL for E*FEMP1* in the thyroid. The *PTBRG* locus included a nonsynonymous variant (rs7652177, *T179S*) with a high posterior probability of 0.99; this variant showed enhancer-like histone marks in mesenchymal cells but was not an annotated *cis*-eQTL. The *FBXW11* locus contained two non-coding variants with a posterior probability of 0.98 (rs153753 and rs4868126) and may influence enhancers in fat, muscle, bone, skin, and the stomach based on data from the Roadmap and Encode projects.[25] The *ZFAT* locus contained an intronic variant rs2277138 with posterior probability of 0.98 that also showed enhancer histone marks in adrenal, brain, and thymus, and was a *cis*-eQTL for *ZFAT* in lymphocytes. Two rare missense (MAF < 1%) variants in ZFAT, rs112892337 and rs75596750, for height were reported by Marouli et al.[6] The *PML* locus included a nonsynonymous variant (rs5742915, F645L) with posterior probability of 1.0 that also has enhancer histone marks in several tissues including adipose, muscle, gastro-intestinal, lung, heart, and others, and the variant is also a *cis*-eQTL for *PML* in lung. At the *IGF1R* locus, the intronic variant rs2871865 (posterior probability = 1.0) has both enhancer and promoter histone marks in almost all tissues. The *ZBTB4* locus included the variant rs9217 (posterior probability of 0.98), which lies in the 3-prime UTR region of the gene and is a *cis*-eQTL for *CHRNB1* in several tissues, including lung, blood, gastro-intestinal, adipose, muscle, and skin. The *ABCD4* locus included the intronic variant rs11657325 (posterior probability = 0.97), which has enhancer histone marks in several tissues, including adipose, muscle, gastro-intestinal, lung, and heart. It is also a *cis*-eQTL for (1) *DCAKD* in gastro-intestinal, muscle, skin, and thyroid, and (2) *ABCD4* in thyroid. The locus *AXIN2* included the intergenic variant, rs757558 (posterior probability = 0.99), which has enhancer marks in several tissues, including muscle, adipose, lung, and heart. It is also a *cis*-eQTL for *AXIN2* in blood and lymphocytes. The *MC4R* locus included the variant rs6567160 (posterior probability = 0.99), which has enhancer marks in several tissues including muscle, adipose, lung, and heart. The locus *GDF5* included the variant intergenic variant, rs143384 (posterior probability = 1.0), which has enhancer marks in several tissues including muscle, adipose, lung, and heart. It is also a *cis*-eQTL with *UQCC1* in adipose, muscle, lung, esophagus, and blood.

Gene set and pathway enrichment analysis

To determine whether the significant variants from African ancestry height results highlight novel biological pathways and/or provide additional support for previously identified biological pathways, we applied a pathway analysis method using DEPICT (Data-driven Expression Prioritized Integration for Complex Traits).[31,33] We examined all variants with suggestive significance ($p < 1E{-}4$) from the stage 1 analyses. We used 1000 Genomes Phase 3 genotype files based on western African ancestry samples (specific populations ESN, GWD, MSL, YRI) rather than EUR genotypes to clump the input data based on LD, which produced 551 loci. We observed 449 significant gene sets (Table S15). The top 10 gene sets included "SMAD2 PPI subnetwork," "chordate embryonic development," "embryo development ending in birth or egg hatching," "absent stapes," "rib fusion," "protein localization to nucleus," "skeletal system development," "pathways in cancer," "Wnt signaling pathway," and vertebral transformation. In general, the biology defined by the gene sets were similar to those reported in Europeans ($R^2 = 0.617$, $p < 1E{-}300$ with Wood et al.[1]).

Trans-ethnic findings to account for population structure in previous GWASs

The first two PCs in PCA (Figure S10) reflected geographical or population structure in Europe, corresponding to the North-South and Southeast-Southwest axes of variation, respectively. Consistent with subtle but persistent uncorrected bias in effect sizes due to stratification, we found that effect sizes estimated from GIANT and the AAAGC+GIANT trans-ethnic meta-analysis were both highly correlated with the loadings of the first principal component of population structure (rho = 0.125, p = 3.24E−94 in GIANT; rho = 0.110, p = 1.64E−82 in the trans-ethnic meta-analysis). The correlation is much lower in AAAGC (rho = 0.012, p = 2.17E−4; Figure 3). Importantly, the magnitude of correlation was lessened in *trans*-ethnic meta-analysis compared with GIANT (p = 3.84E−5).

Discussion

We undertook a large-scale GWAS meta-analysis of height in African ancestry individuals imputed to the 1000 Genomes reference panel, complemented by a meta-analysis with a European GWAS, with both sex-stratified and sex-combined analyses considered. In total, our results among African Ancestry individuals revealed 42 genome-wide significant loci associated with height, 39 known and 3 novel

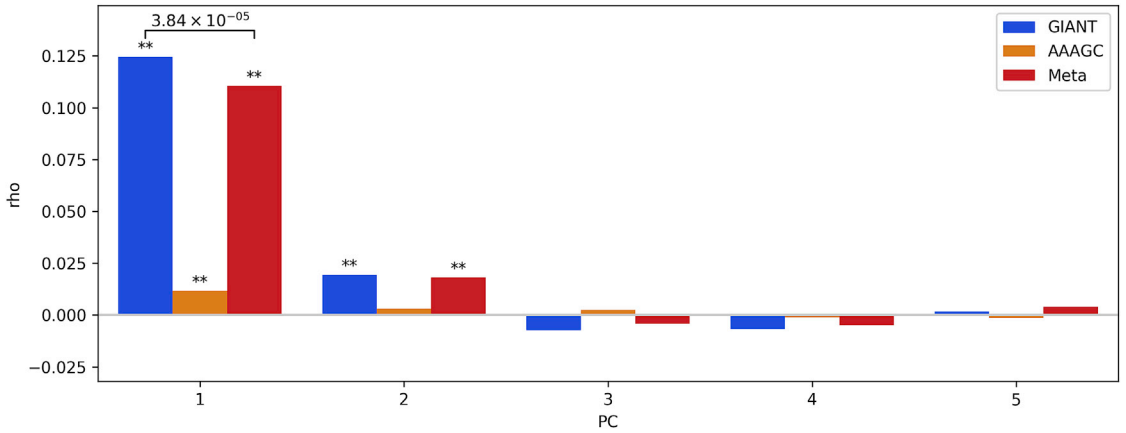

Figure 3. Correlations (rho) between effect estimates and the loadings of the principal components 1–5 in each consortium
GIANT (Genetic Investigation of ANthropometric Traits) and AAAGC (African American Anthropometry Genetics Consortium) and the meta-analysis of both (Meta).

loci. Two more novel loci were identified from the sex-stratified analyses. After we combined with European ancestry results, three more novel associations were identified. Among the 39 known loci, we identified a total of 20 new independent signals that reached genome-wide significance. In total, eight of the identified SNPs (3 sex-combined AA, 2 sex-stratified AA, 3 when combined with EA) were in novel regions based on height publications up to January 2018, in or near *SLC4A3/MIR4268, NCOA2, ECD/FAM149B1, RCCD1, G6PC3, CEP95, CRB1*, and *KLF6*. While 2 out of 39 known loci had MAF < 5%, none of the 8 novel loci did. After accounting for winner's curse, the variance explained by the eight newly identified variants from the sex-combined analyses was ~0.3%, bringing the total variance explained for height to ~28%, when considering all the 627 known loci plus the 8 new loci. We may have been overly conservative in our approach to controlling type 1 errors, by implementing a double GC correction. Thus, we formally assessed, post hoc, the evidence for over-correction using the intercepts from LD score regression of the double GC corrected genome-wide (stage 1) meta-analyses. We estimated a deflation (i.e., 1/intercept) of 0.857, 0.907, and 0.933 for the sex-combined, women-only, and men-only strata, respectively. Such findings indeed support an over-correction of our study results and this is an inherent limitation to this study, in that we have likely missed additional real signals that influence height. Given that much large meta-analyses of height, including these same study populations, are now ongoing, we have chosen to simply acknowledge this limitation and additionally provide single GC corrected meta-analysis results to the NHGRI-EBI GWAS catalog upon publication of this study. In addition, we are aware of the Yengo et al.[5] publication and analyses that included a larger set of European descent individuals (i.e., the GIANT study[1] plus UK Biobank data) published in 2018. We looked up the lead SNPs or something in high LD ($r^2 \geq 0.7$ in AFR or EUR) for each of the 8 novel loci in the Yengo et al.[5] publicly available results. None met genome-wide significance in the larger European ancestry study. Two (*CRB1* and *KLF6*) were monomorphic or very rare (MAF < 0.001) in Europeans and one (rs228758 in/near *G6PC3*) was almost genome-wide significant with a p value = 5.2E−8. Thus, our results highlight new genes with evidence of involvement in skeletal development and disease that advance current knowledge of height genetics and biology.

Our analyses revealed the contribution of non-coding variants in several genes, some of them related to skeletal growth and bone development. Nuclear Receptor Coactivator 2 (*NCOA2*) has been found to be involved in translocations that result in fusions with other genes in various cancers, including mesenchymal chondrosarcoma, which is a rare cancer type usually beginning from the bones. The nuclear receptor coactivator protein acts as a transcriptional coactivator for nuclear hormone receptors including vitamin D receptors (see web resources).

Among the 8 novel height loci, 2 were genome-wide significant only in the sex-specific analyses and thus appear to be driven by females. The *CRB1* gene provides instructions for making a protein that plays an essential role in normal vision.[40] Gene Ontology annotations related to this gene include "calcium ion binding." Calcium ions have a crucial role for skeletal muscle function, plasticity, and disease.[41] However, the functional characterization at the *CRB1* locus does not point to this gene, but to the *CFHR4* and *ASPM* genes described in more detail below, illustrating the limitation of naming loci after the closest gene. Why the effect sizes for this locus is much larger in women than men (~0.19 in women and −0.003 in men), is not clear. We cannot exclude the possibility that the larger sample size in women had an undue influence on our findings. The second locus driven by women is near *KLF6* which is involved in the TGF-beta Signaling Pathway,[42] which plays a fundamental role in both embryonic skeletal development and postnatal bone homeostasis.[43] We note that this locus was barely significant and should be considered with caution.

Among the eight novel height loci, we identified six putative coding variants in high LD with three of the leading variants. The first lead SNP rs672769 was in high LD with nonsynonymous SNPs in two genes, *CFHR4* and *ASPM*, and not in *CRB1* the closest gene to the lead index SNP. Mutations in *ASPM* are the most common cause of autosomal-recessive primary microcephaly (MCPH), a condition where the size of the cerebral cortex is significantly reduced.[44] *ASPM* is necessary for normal mitotic spindle function in embryonic neuroblasts.[44] Gene Ontology annotations related to this gene also include "calcium ion or calmodulin binding." The second lead SNP, rs7905296, was also in high LD with two coding variants. One of such variants was located in *P4HA1*, which encodes proteins involved in the synthesis of collagen, an important component of the extracellular matrix. Bi-allelic mutations in *P4HA1* were reported in a family with congenital disorder of connective tissue.[45] The other variant was in *ECD*, which is involved in cell cycle arrest and apoptosis (see web resources). Two of the novel loci were nominally significantly different ($p < 0.05$) in effect sizes between AA men and women. The third lead SNP, rs8082112 was in high LD with a coding variant (rs1427463) which is located in *POLG2*. *POLG2* is required for mitochondrial DNA replication, and mutations in *POLG2* have been liked to a variety of diseases, including progressive external ophthalmoplegia (PEO).[46] *PEO* is characterized by symptoms including progressive weakening of the external eye muscle (ophthalmoparesis).[46] In addition, mitochondrial disease is linked to short stature.[47] The rs8082122 signal is mainly driven by women.

We also identified additional signals in known loci. For instance, two variants, rs1150781 and rs2070808, are located close to *HMGA1* and *HMGA2,* respectively. These two genes are important genetic determinants of human adult height.[5,7] At many of the identified signals, fine-mapping resolution provided further specification of plausible causal variants. We highlight 23 height loci as human-validated targets based on causal variant effects. We also provide insights into the potential biological mechanisms implicated by several of the fine-mapped signals. Interestingly, the identified variant in locus *EFEMP1* is a *cis*-eQTLs with *EFEMP1* in the thyroid tissue, and epidemiological studies have reported that prolonged hypothyroidism may result in compromised height.[48] In contrast, hyperthyroidism has been reported to accelerate growth in children and individuals with Turner syndrome.[49]

In the follow-up analyses in children for the lead variants from stage 1 and stage 2 results, we find some support of that these loci influence height in all children or in children by pubertal status. We find heterogeneity for some loci by sex and/or pubertal status, possibly indicating distinct genetic effects across the life course. There is a lack of correlation between effect sizes in all children and the adult stage 1 + stage 2 effect sizes, which could be due to a true lack of correlation, differences in growth by pubertal status, or low power as the sample size of children is small.

As the vast majority of GWASs are performed in Europeans, transferability to other populations is dependent on several parameters, including genetic architecture, allele frequency differences, and population differences in LD. In the SNP and locus transferability analyses, 80% of EA variants displayed directionally consistent associations in our AA samples, and a quarter were nominally significant. More than 50% of the variants that demonstrated directional consistency and were nominally significant in AA analyses had larger effect sizes and allele frequencies in EA compared to AA populations. Many of these variants did not reach genome-wide significance in the AA meta-analysis likely due to a smaller sample, although it is also possible that some of the signals may not represent true signals in AA. This could be another indication of a non-extensive transferability across populations. It is also reflected in the low correlation of the effect sizes between the two populations as well as in the low transferability of lead variants from EA to AA populations. Even though the EA and AA lead SNPs are uncorrelated in some cases, they still could be tagging the same causal variant given the differences in LD between the two ancestries.

Residual uncorrected stratification in GWASs could result in biased estimates of effect sizes.[1] 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.[50–52] The high biological plausibility of the top pathways also emphasizes the point that the subtle inflation across the genome does not alter the relevance of the top signals and pathways. Note that the residual stratification effect 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 UK Biobank, which is based on a single homogeneous population and results in better control of population stratification, the genetic correlation between GIANT and UKB was 0.94.[52]

Meta-analysis using GWAS summary statistics from GIANT and an ancestrally diverse population is expected to alleviate concern of uncorrected stratification because any biases in the non-European population should be independent of the structure in Europe. This is indeed what we observed (Figure 2). First, we found that in AAAGC the effect size correlation with PC1, although significant, was much less than what we observed in GIANT, suggesting the effect sizes are less biased by European population stratification. This could be due to a large proportion of African ancestry in the AAAGC cohort; it could also be due to a smaller sample size in the AAAGC resulting in lower precision in effect size estimates. Either way, the magnitude of correlation was lessened in the meta-analysis of GIANT and AAAGC consortia, as we expected, despite the significantly smaller sample size in AAAGC. As non-European cohorts increase in sample size, we would expect the bias in effect size estimates from meta-analysis to continue to decrease.

Overall, our results provide evidence for an ancestry-specific genetic influence on height in AA populations that had not been captured by large-scale meta-analyses in Europeans, and we report eight novel loci. These findings have important implications on transferability of genetic variability across populations and generally for prediction of complex phenotypes and diseases. They also give us additional signals to follow up on in wet-lab functional studies. Focusing on the identification of population-specific genetic variants will pave the way to more accurate prediction tools, which will have significant impact in the era of customized care and precision medicine. As medical genomics studies are extensively large and diverse, shedding light toward the direction of transferability of the identified genetic component of complex traits is critical.

Data and code availability

Meta-analysis results can be accessed through the NHGRI-EBI GWAS Catalog under the following accession numbers: AAAGC_Height_All, GCST90013466; AAAGC_Height_Men, GCST90013467; AAAGC_Height_Women, GCST90013468.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.02.011.

Acknowledgments

See full acknowledgments across all studies in Table S16.

Declaration of interests

The authors have nothing to declare.

Received: June 5, 2020
Accepted: February 9, 2021
Published: March 12, 2021

Web resources

DEPICT, https://data.broadinstitute.org/mpg/depict/index.html
EasyQC R package, https://homepages.uni-regensburg.de/~wit59712/easyqc/EasyQC_9.0_Commands_140918_2.pdf
EasyStrata R package, https://homepages.uni-regensburg.de/~wit59712/easystrata/EasyStrata_8.6_Commands_140615.pdf
Ecdysoneless cell cycle regulator (ECD), https://omim.org/entry/616464
Eigensoft version 7.2.1, https://github.com/DReichLab/EIG/archive/v7.2.1.tar.gz
GCTA software with the cojo-slct and cojo-cond methods, https://cnsgenomics.com/software/gcta/#COJO
GeneCards, *NOCA2* Gene, https://www.genecards.org/cgi-bin/carddisp.pl?gene=NCOA2
GIANT consortium data files, https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
Haplotype Reference Consortium, http://www.haplotype-reference-consortium.org/home
METAL, https://genome.sph.umich.edu/wiki/METAL_Documentation
NHGRI-EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/downloads/summary-statistics
NHLBI Exome Sequencing Project Exome Variant Server, https://evs.gs.washington.edu/EVS/
Open Targets Genetics, https://genetics.opentargets.org/

References

1. 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.; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; and LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.
2. 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.; Diabetes Genetics Initiative; Wellcome Trust Case Control Consortium; and Cambridge GEM Consortium (2008). Genome-wide association analysis identifies 20 loci that influence adult height. Nat. Genet. *40*, 575–583.
3. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.
4. Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., Eyheramendy, S., Voight, B.F., Butler, J.L., Guiducci, C., et al.; Diabetes Genetics Initiative; FUSION; KORA; Prostate, Lung Colorectal and Ovarian Cancer Screening Trial; Nurses' Health Study; and SardiNIA (2008). Identification of ten loci associated with height highlights new biological pathways in human growth. Nat. Genet. *40*, 584–591.
5. 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.; and GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum. Mol. Genet. *27*, 3641–3649.
6. 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.; EPIC-InterAct Consortium; CHD Exome+ Consortium; ExomeBP Consortium; T2D-Genes Consortium; GoT2D Genes Consortium; Global Lipids Genetics Consortium; ReproGen Consortium; and MAGIC Investigators (2017). Rare and low-frequency coding variants alter human adult height. Nature *542*, 186–190.
7. N'Diaye, A., Chen, G.K., Palmer, C.D., Ge, B., Tayo, B., Mathias, R.A., Ding, J., Nalls, M.A., Adeyemo, A., Adoue, V., et al. (2011). Identification, replication, and fine-mapping of Loci associated with adult height in individuals of african ancestry. PLoS Genet. *7*, e1002298.
8. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al.; International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
9. Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu,

F., Peltonen, L., et al.; International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. Nature *467*, 52–58.
10. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., McVean, G.A.; and 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
11. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., Abecasis, G.R.; and 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
12. Delaneau, O., Marchini, J.; and 1000 Genomes Project Consortium (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat. Commun. *5*, 3934.
13. Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L., Danecek, P., Malerba, G., Trabetti, E., Zheng, H.F., et al.; UK10K Consortium (2015). Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. Nat. Commun. *6*, 8111.
14. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. *34*, 816–834.
15. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
16. Deeks, J.J., Higgins, J.P., and Altman, D.G. (2008). Analysing Data and Undertaking Meta-analyses. In Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series, J.P.T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M.J. Page, and V.A. Welch, eds..
17. Zhong, H., and Prentice, R.L. (2008). Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. Biostatistics *9*, 621–634.
18. Palmer, C., and Pe'er, I. (2017). Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. PLoS Genet. *13*, e1006916.
19. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Weedon, M.N., Loos, R.J., et al.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; and DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. *44*, 369–375, S1–S3.
20. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
21. Tachmazidou, I., Süveges, D., Min, J.L., Ritchie, G.R.S., Steinberg, J., Walter, K., Iotchkova, V., Schwartzentruber, J., Huang, J., Memari, Y., et al.; SpiroMeta Consortium; GoT2D Consortium; arcOGEN Consortium; Understanding Society Scientific Group; and UK10K Consortium (2017). Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. Am. J. Hum. Genet. *100*, 865–884.
22. Li, J., and Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity *95*, 221–227.
23. Kichaev, G., and Pasaniuc, B. (2015). Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. Am. J. Hum. Genet. *97*, 260–271.
24. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.
25. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. *40*, D930–D934.
26. Siepel, A., and Haussler, D. Phylogenetic Hidden Markov Models. In Statistical Methods in Molecular Evolution, R. Nielsen, ed. (New York, NY: Springer).
27. Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. *6*, e1001025.
28. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Chapter 7, Unit7 20.
29. Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31*, 3812–3814.
30. 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.
31. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. *6*, 5890.
32. (1989). The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am. J. Epidemiol. 129, 687–702.
33. Frey, B.J., and Dueck, D. (2007). Clustering by passing messages between data points. Science *315*, 972–976.
34. Locke, A.E., Steinberg, K.M., Chiang, C.W.K., Service, S.K., Havulinna, A.S., Stell, L., Pirinen, M., Abel, H.J., Chiang, C.C., Fulton, R.S., et al.; FinnGen Project (2019). Exome sequencing of Finnish isolates enhances rare-variant association power. Nature *572*, 323–328.
35. 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.
36. 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.
37. Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. *42*, D1001–D1006.
38. Leslie, R., O'Donnell, C.J., and Johnson, A.D. (2014). GRASP: analysis of genotype-phenotype results from 1390 genome-wide association studies and corresponding open access database. Bioinformatics *30*, i185–i194.

39. Carty, C.L., Johnson, N.A., Hutter, C.M., Reiner, A.P., Peters, U., Tang, H., and Kooperberg, C. (2012). Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe). Hum. Mol. Genet. *21*, 711–720.
40. Jacobson, S.G., Cideciyan, A.V., Aleman, T.S., Pianta, M.J., Sumaroka, A., Schwartz, S.B., Smilko, E.E., Milam, A.H., Sheffield, V.C., and Stone, E.M. (2003). Crumbs homolog 1 (CRB1) mutations result in a thick human retina with abnormal lamination. Hum. Mol. Genet. *12*, 1073–1078.
41. Berchtold, M.W., Brinkmeier, H., and Müntener, M. (2000). Calcium ion in skeletal muscle: its crucial role for muscle function, plasticity, and disease. Physiol. Rev. *80*, 1215–1265.
42. Dionyssiou, M.G., Salma, J., Bevzyuk, M., Wales, S., Zakharyan, L., and McDermott, J.C. (2013). Krüppel-like factor 6 (KLF6) promotes cell proliferation in skeletal myoblasts in response to TGFβ/Smad3 signaling. Skelet. Muscle *3*, 7.
43. Wu, M., Chen, G., and Li, Y.P. (2016). TGF-β and BMP signaling in osteoblast, skeletal development, and bone formation, homeostasis and disease. Bone Res. *4*, 16009.
44. Bond, J., Roberts, E., Mochida, G.H., Hampshire, D.J., Scott, S., Askham, J.M., Springell, K., Mahadevan, M., Crow, Y.J., Markham, A.F., et al. (2002). ASPM is a major determinant of cerebral cortical size. Nat. Genet. *32*, 316–320.
45. Zou, Y., Donkervoort, S., Salo, A.M., Foley, A.R., Barnes, A.M., Hu, Y., Makareeva, E., Leach, M.E., Mohassel, P., Dastgir, J., et al. (2017). P4HA1 mutations cause a unique congenital disorder of connective tissue involving tendon, bone, muscle and the eye. Hum. Mol. Genet. *26*, 2207–2217.
46. Young, M.J., and Copeland, W.C. (2016). Human mitochondrial DNA replication machinery and disease. Curr. Opin. Genet. Dev. *38*, 52–62.
47. Boal, R.L., Ng, Y.S., Pickett, S.J., Schaefer, A.M., Feeney, C., Bright, A., Taylor, R.W., Turnbull, D.M., Gorman, G.S., Cheetham, T., and McFarland, R. (2019). Height as a Clinical Biomarker of Disease Burden in Adult Mitochondrial Disease. J. Clin. Endocrinol. Metab. *104*, 2057–2066.
48. Kandemir, N., and Yordam, N. (2001). Height prognosis in children with late-diagnosed congenital hypothyroidism. Turk. J. Pediatr. *43*, 303–306.
49. Massa, G., de Zegher, F., Dooms, L., and Vanderschueren-Lodeweyckx, M. (1992). Hyperthyroidism accelerates growth in Turner's syndrome. Acta Paediatr. *81*, 362–364.
50. 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., and Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. eLife *8*, 8.
51. Chen, M., Sidore, C., Akiyama, M., Ishigaki, K., Kamatani, Y., Schlessinger, D., Cucca, F., Yukinori, O., and Chiang, C.W.K. (2019). Evidence of polygenic adaptation at height-associated loci in mainland Europeans and Sardinians. Amer. J. Hum. Genet. *107*, 60–71.
52. 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*, 8.