Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction

David V. Conti[1†¥], Burcu F. Darst[1†], Lilit Moss[1], Edward J. Saunders[2], Xin Sheng[1], Alisha Chou[1], Fredrick R. Schumacher[3,4], Ali Amin Al Olama[5,6], Sara Benlloch[5], Tokhir Dadaev[2], Mark N. Brook[2], Ali Sahimi[1], Thomas J. Hoffmann[7,8], Atushi Takahashi[9,10], Koichi Matsuda[11,12], Yukihide Momozawa[13], Masashi Fujita[14], Kenneth Muir[15,16], Artitaya Lophatananon[15], Peggy Wan[1], Loic Le Marchand[17], Lynne R. Wilkens[17], Victoria L. Stevens[18], Susan M. Gapstur[18], Brian D. Carter[18], Johanna Schleutker[19,20], Teuvo L. J. Tammela[21], Csilla Sipeky[19], Anssi Auvinen[22], Graham G. Giles[23,24,25], Melissa C. Southey[25], Robert J. MacInnis[23,24], Cezary Cybulski[26], Dominika Wokołorczyk[26], Jan Lubiński[26], David E. Neal[27,28,29], Jenny L. Donovan[30], Freddie C. Hamdy[31,32], Richard M. Martin[30,33,34], Børge G. Nordestgaard[35,36], Sune F. Nielsen[35,36], Maren Weischer[36], Stig E. Bojesen[35,36], Martin Andreas Røder[37], Peter Iversen[37], Jyotsna Batra[38,39], Suzanne Chambers[40], Leire Moya[38,39], Lisa Horvath[41,42], Judith A. Clements[38,39], Wayne Tilley[43], Gail P. Risbridger[44,45], Henrik Gronberg[46], Markus Aly[46,47,48], Robert Szulkin[46,49], Martin Eklund[46], Tobias Nordström[46,50], Nora Pashayan[51,52,53], Alison M. Dunning[52], Maya Ghoussaini[54], Ruth C. Travis[55], Tim J. Key[55], Elio Riboli[56], Jong Y. Park[57], Thomas A. Sellers[57], Hui-Yi Lin[58], Demetrius Albanes[59], Stephanie Weinstein[59], Lorelei A. Mucci[60], Edward Giovannucci[60], Sara Lindstrom[61], Peter Kraft[62], David J. Hunter [63], Kathryn L. Penney[64], Constance Turman[62], Catherine M. Tangen[65], Phyllis J. Goodman[65], Ian M. Thompson Jr. [66], Robert J. Hamilton[67,68], Neil E. Fleshner[67], Antonio Finelli[69], Marie-Élise Parent[70,71], Janet L. Stanford[72,73], Elaine A. Ostrander[74], Milan S. Geybels[72], Stella Koutros[59], Laura E. Beane Freeman[59], Meir Stampfer[64], Alicja

Wolk[75,76], Niclas Håkansson[75], Gerald L. Andriole[77], Robert N. Hoover[59], Mitchell J. Machiela[59], Karina Dalsgaard Sørensen[78,79], Michael Borre[79,80], William J. Blot[81,82], Wei Zheng[81], Edward D. Yeboah[83,84], James E. Mensah[83,84], Yong-Jie Lu[85], Hong-Wei Zhang[86], Ninghan Feng[87], Xueying Mao[85], Yudong Wu[88], Shan-Chao Zhao[89], Zan Sun[90], Stephen N. Thibodeau[91], Shannon K. McDonnell[92], Daniel J. Schaid[92], Catharine M. L. West[93], Neil Burnet[94], Gill Barnett[95], Christiane Maier[96], Thomas Schnoeller[97], Manuel Luedeke[98], Adam S. Kibel[99], Bettina F. Drake[77], Olivier Cussenot[100], Géraldine Cancel-Tassin[101,100], Florence Menegaux[102], Thérèse Truong[102], Yves Akoli Koudou[103], Esther M. John[104], Eli Marie Grindedal[105], Lovise Maehle[105], Kay-Tee Khaw[106], Sue A. Ingles[107], Mariana C. Stern[107], Ana Vega[108,109,110], Antonio Gómez-Caamaño[111], Laura Fachal[5,108,109,110], Barry S. Rosenstein[112,113], Sarah L. Kerns[114], Harry Ostrer[115], Manuel R. Teixeira[116,117], Paula Paulo[116,118], Andreia Brandão[116,118], Stephen Watya[119], Alexander Lubwama[119], Jeannette T. Bensen[120,121], Elizabeth T.H. Fontham[59], James Mohler[121,122], Jack A. Taylor[123,124], Manolis Kogevinas[125,126,127,128], Javier Llorca[129,128], Gemma Castaño-Vinyals[125,126,127,128], Lisa Cannon-Albright[130,131], Craig C. Teerlink[130,131], Chad D. Huff[132], Sara S. Strom[132], Luc Multigner[133], Pascal Blanchet[134], Laurent Brureau[134], Radka Kaneva[135], Chavdar Slavov[136], Vanio Mitev[135], Robin J. Leach[137], Brandi Weaver[137], Hermann Brenner[138,139,140], Katarina Cuk[138], Bernd Holleczek[141], Kai-Uwe Saum[138], Eric A. Klein[142,143], Ann W. Hsing[144], Rick A. Kittles[145], Adam B. Murphy[146], Christopher J. Logothetis[147], Jeri Kim[147], Susan L. Neuhausen[148], Linda Steele[148], Yuan Chun Ding[148], William B. Isaacs[149], Barbara Nemesure[150], Anselm J. M. Hennis[150,151], John Carpten[152], Hardev Pandha[153], Agnieszka Michael[153], Kim De Ruyck[154], Gert De Meerleer[155], Piet Ost[155], Jianfeng Xu[156], Azad Razack[157], Jasmine Lim[157], Soo-Hwang Teo[158], Lisa F. Newcomb[73,159], Daniel W. Lin[73,159], Jay H.

Fowke[160], Christine Neslund-Dudas[161], Benjamin A. Rybicki[161], Marija Gamulin[162], Davor Lessel[163], Tomislav Kulis[164], Nawaid Usmani[165,166], Sandeep Singhal[165], Matthew Parliament[165,166], Frank Claessens[167], Steven Joniau[168], Thomas Van den Broeck[167,168], Manuela Gago-Dominguez[169,170], Jose Esteban Castelao[171], Maria Elena Martinez[172], Samantha Larkin[173], Paul A. Townsend[153,174], Claire Aukim-Hastie[153], William S. Bush[175], Melinda C. Aldrich[176], Dana C. Crawford[175], Shiv Srivastava[177], Jennifer C. Cullen[177], Gyorgy Petrovics[177], Graham Casey[178], Monique J. Roobol[179], Guido Jenster[179], Ron H.N. van Schaik[180], Jennifer J. Hu[181], Maureen Sanderson[182], Rohit Varma[183], Roberta McKean-Cowdin[1], Mina Torres[183], Nicholas Mancuso[1], Sonja I. Berndt[59], Stephen K. Van Den Eeden[184,185], Douglas F Easton[5], Stephen J. Chanock[59], Michael B. Cook[59], Fredrik Wiklund[46], Hidewaki Nakagawa[14], John S. Witte[7,8,185], Rosalind A. Eeles[2,186¥], Zsofia Kote-Jarai[2¥], Christopher A. Haiman[1¥*]

[1]Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA, [2]The Institute of Cancer Research, London, UK, [3]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA, [4]Seidman Cancer Center, University Hospitals, Cleveland, Ohio, USA, [5]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK, [6]University of Cambridge, Department of Clinical Neurosciences, Stroke Research Group, Cambridge Biomedical Campus, Cambridge, UK, [7]Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA, [8]Institute for

Human Genetics, University of California, San Francisco, California, USA, [9]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [10]Department of Genomic Medicine, National Cerebral and Cardiovascular Center Research Institute, Suita, Japan, [11]Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan, [12]Biobank Japan, [13]Laboratory for Genotyping Development, RIKEN Center of Integrative Medical Sciences, Yokohama, Japan, [14]Laboratory for Cancer Genomics, RIKEN Center of Integrative Medical Sciences, Yokohama, Japan, [15]Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK, [16]Warwick Medical School, University of Warwick, Coventry, UK, [17]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA, [18]Behavioral and Epidemiology Research Group, Research Program, American Cancer Society, Atlanta, Georgia, USA, [19]Institute of Biomedicine, University of Turku, Turku, Finland, [20]Department of Medical Genetics, Genomics, Laboratory Division, Turku University Hospital, Turku, Finland, [21]Department of Urology, Tampere University Hospital and Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland, [22]Unit of Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland, [23]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia, [24]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia, [25]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia, [26]International Hereditary Cancer Center, Department of Genetics and Pathology,

Pomeranian Medical University, Szczecin, Poland, [27]Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Headington, Oxford, UK, [28]University of Cambridge, Department of Oncology, Addenbrooke's Hospital, Cambridge, UK, [29]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK, [30]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK, [31]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, [32]Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK, [33]National Institute for Health Research (NIHR) Biomedical Research Centre, University of Bristol, Bristol, UK, [34]Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK, [35]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [36]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Copenhagen, Denmark, [37]Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Herlev, Copenhagen, Denmark, [38]Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia, [39]Translational Research Institute, Brisbane, Queensland, Australia, [40]University of Technology, Sydney, Australia, [41]Chris O'Brien Lifehouse (COBLH), Camperdown, Sydney NSW, Australia, [42]Garvan Institute of Medical Research, Sydney NSW, Australia, [43]Dame Roma Mitchell Cancer Research Laboratories, Adelaide Medical School, University of Adelaide, Adelaide, South Australia, Australia, [44]Department of Anatomy and Developmental Biology, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia, [45]Prostate Cancer Translational Research Program, Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne,

Victoria, Australia, [46]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden, [47]Department of Molecular Medicine and Surgery, Karolinska Institutet, and Department of Urology, Karolinska University Hospital, Solna, Stockholm, [48]Department of Urology, Karolinska University Hospital, Stockholm, Sweden, [49]SDS Life Science, Danderyd, Sweden, [50]Department of Clinical Sciences at Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden, [51]University College London, Department of Applied Health Research, London, UK, [52]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Cambridge, UK, [53]Department of Applied Health Research, University College London, London, UK, [54]Open Targets, Wellcome Sanger Institute, Hinxton, Saffron Walden, UK, [55]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK, [56]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK, [57]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, Florida, USA, [58]School of Public Health, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA, [59]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, USA, [60]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA, [61]Department of Epidemiology, University of Washington, Seattle, Washington, USA, [62]Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA, [63]Nuffield Department of Population Health, University of Oxford, Oxford, UK, [64]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, Massachusetts, USA, [65]SWOG Statistical Center, Fred Hutchinson Cancer

Research Center, Seattle, Washington, USA, [66]CHRISTUS Santa Rosa Hospital – Medical Center, San Antonio, Texas, USA, [67]Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Canada, [68]Dept. of Surgery (Urology), University of Toronto, Toronto, Canada, [69]Division of Urology, Princess Margaret Cancer Centre, Toronto, Canada, [70]Epidemiology and Biostatistics Unit, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique, Laval, Quebec, Canada, [71]Department of Social and Preventive Medicine, School of Public Health, University of Montreal, Montreal, Quebec, Canada, [72]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA, [73]Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA, [74]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA, [75]Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, [76]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden, [77]Washington University School of Medicine, St. Louis, Missouri, USA, [78]Department of Molecular Medicine, Aarhus University Hospital, Aarhus N, Denmark, [79]Department of Clinical Medicine, Aarhus University, Aarhus N, Denmark, [80]Department of Urology, Aarhus University Hospital, Aarhus N, Denmark, [81]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [82]International Epidemiology Institute, Rockville, Maryland, USA, [83]University of Ghana Medical School, Accra, Ghana, [84]Korle Bu Teaching Hospital, Accra, Ghana, [85]Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London, UK, [86]Second Military Medical University, Shanghai, China, [87]Wuxi Second Hospital, Nanjing Medical University, Wuxi, Jiangzhu Province, China, [88]Department of

Urology, First Affiliated Hospital, The Academy of Medical Sciences, Zhengzhou University, Zhengzhou, China, [89]Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou, China, [90]The People's Hospital of Liaoning Proviouce, The People's Hospital of China Medical University, Shenyang, China, [91]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA, [92]Division of Biomedical Statistics & Informatics, Mayo Clinic, Rochester, Minnesota, USA, [93]Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester, UK, [94]Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science Centre, and The Christie NHS Foundation Trust, Manchester, UK, [95]University of Cambridge Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK, [96]Humangenetik Tuebingen, Tuebingen, Germany, [97]Department of Urology, University Hospital Ulm, Germany, [98]genetikum, Neu-Ulm, Germany, [99]Division of Urologic Surgery, Brigham and Womens Hospital, Boston, Massachusetts, USA, [100]Sorbonne Universite, Tenon Hospital, Paris, France, [101]CeRePP, Tenon Hospital, Paris, France., [102]"Exposome and Heredity", CESP (UMR 1018), Faculté de Médecine, Université Paris-Saclay, Inserm, Gustave Roussy, Villejuif, France, [103]CESP (UMR 1018), Faculté de Médecine, Université Paris-Saclay, Inserm, Villejuif, France, [104]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA, [105]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway, [106]Clinical Gerontology Unit, University of Cambridge, Cambridge, UK, [107]Department of Preventive Medicine, Keck School of Medicine,

University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA, [108]Fundación Pública Galega Medicina Xenómica, Santiago De Compostela, Spain, [109]Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago De Compostela, Spain, [110]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Spain, [111]Department of Radiation Oncology, Complexo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain, [112]Department of Radiation Oncology and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [113]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA, [114]Department of Radiation Oncology, University of Rochester Medical Center, Rochester, New York, USA, [115]Department of Pathology, Albert Einstein College of Medicine, Bronx, New York, USA, [116]Department of Genetics, Portuguese Oncology Institute of Porto (IPO-Porto), Porto, Portugal, [117]Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal, [118]Cancer Genetics Group, IPO-Porto Research Center (CI-IPOP), Portuguese Oncology Institute of Porto (IPO-Porto), Porto, Portugal, [119]Uro Care, Kampala, Uganda, [120]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, [121]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, [122]Department of Urology, Roswell Park Cancer Institute, Buffalo, New York, USA, [123]Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA, [124]Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, [125]ISGlobal, Barcelona, Spain, [126]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain, [127]Universitat Pompeu Fabra

(UPF), Barcelona, Spain, [128]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain, [129]University of Cantabria-IDIVAL, Santander, Spain, [130]Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA, [131]George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah, USA, [132]Department of Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas, USA, [133]Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), Rennes, France, [134]CHU de Pointe-à-Pitre, Univ Antilles, Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), Pointe-à-Pitre, France, [135]Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, Bulgaria, [136]Department of Urology and Alexandrovska University Hospital, Medical University of Sofia, Sofia, Bulgaria, [137]Department of Urology, Cancer Therapy and Research Center, University of Texas Health Science Center at San Antonio, San Antonio Texas, USA, [138]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, [139]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany, [140]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany, [141]Saarland Cancer Registry, Saarbrücken, Germany, [142]Cleveland Clinic Lerner Research Institute, Cleveland, Ohio, USA, [143]Glickman Urological & Kidney Institute, Cleveland Clinic, Cleveland, Ohio, USA, [144]Department of Medicine and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA, [145]Division of Health Equities, Department of Population Sciences, City of Hope, Duarte, California, USA, [146]Department of Urology, Northwestern University,

Chicago, Illinois, USA, [147]The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston, Texas, USA, [148]Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, California, USA, [149]James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institution, Baltimore, Maryland, USA, [150]Department of Family, Population and Preventive Medicine, Stony Brook University, Stony Brook, New York, USA, [151]Chronic Disease Research Centre and Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados, [152]Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, California, USA, [153]Faculty of Health and Medical Sciences, The University of Surrey, Guildford, Surrey, UK, [154]Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Gent, Belgium, [155]Ghent University Hospital, Department of Radiotherapy, Gent, Belgium, [156]Program for Personalized Cancer Care and Department of Surgery, NorthShore University HealthSystem, Evanston, Illinois, USA, [157]Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, [158]Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, Subang Jaya, Selangor, Malaysia, [159]Department of Urology, University of Washington, Seattle, Washington, USA, [160]Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, USA, [161]Department of Public Health Sciences, Henry Ford Hospital, Detroit, Michigan, USA, [162]Department of Oncology, University Hospital Centre Zagreb, University of Zagreb School of Medicine, Zagreb, Croatia, [163]Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [164]Department of Urology, University Hospital Center Zagreb, University of Zagreb School of Medicine , Zagreb, Croatia, [165]Department of

Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta, Canada, [166]Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta, Canada, [167]Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, Leuven, Belgium, [168]Department of Urology, University Hospitals Leuven, Leuven, Belgium, [169]Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago de Compostela, Spain, [170]University of California San Diego, Moores Cancer Center, La Jolla, California, USA, [171]Genetic Oncology Unit, CHUVI Hospital, Complexo Hospitalario Universitario de Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), Vigo (Pontevedra), Spain, [172]University of California San Diego, Moores Cancer Center, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA, [173]The University of Southampton, Southampton General Hospital, Tremona Road, Southampton, UK, [174]Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, University of Manchester, Manchester, UK, [175]Case Western Reserve University, Department of Population and Quantitative Health Sciences, Cleveland Institute for Computational Biology, Cleveland, Ohio, USA, [176]Vanderbilt University Medical Center, Department of Thoracic Surgery, Nashville, Tennessee, USA, [177]Center for Prostate Disease Research, Uniformed Services University, Bethesda, Maryland, USA, [178]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA, [179]Department of Urology, Erasmus University Medical Center, Rotterdam, The

Netherlands, [180]Department of Clinical Chemistry, Erasmus University Medical Center, Rotterdam, The Netherlands, [181]The University of Miami School of Medicine, Sylvester Comprehensive Cancer Center, Miami, Florida, USA, [182]Department of Family and Community Medicine, Meharry Medical College, Nashville, Tennessee, USA, [183]Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles, California, USA, [184]Division of Research, Kaiser Permanente, Northern California, Oakland, California, USA, [185]Department of Urology, University of California San Francisco, San Francisco, California, USA, [186]Royal Marsden NHS Foundation Trust, London, UK

Equal contributions

† Drs. Conti and Darst are co-first authors

¥ Drs. Conti, Eeles, Kote-Jarai and Haiman contributed equally to this article.

*Corresponding Author: Christopher A. Haiman, Harlyne J. Norris Cancer Research Tower, USC Norris Comprehensive Cancer Center, 1450 Biggy Street, Rm 1504, Los Angeles, CA 90033 or haiman@usc.edu

Abstract

Prostate cancer is a highly heritable disease with large disparities in incidence rates across ethnic populations. We conducted a multiethnic meta-analysis of prostate cancer genome-wide association studies (107,247 cases and 127,006 controls) and identified 269 genetic risk variants independently associated with prostate cancer risk, of which 86 were novel. The top genetic risk score (GRS) decile was associated with odds ratios that ranged from 5.06 [95% confidence interval (CI) 4.84-5.29] for men of European ancestry to 3.74 [95% CI 3.36-4.17] for men of African ancestry. Men of African ancestry were estimated to have a mean GRS that was 2.18-times higher [95% CI 2.14-2.22], and men of East Asian ancestry 0.73-times lower [95% CI 0.71-0.76], than men of European ancestry. These findings support the role of germline variation contributing to population differences in prostate cancer risk, with the GRS offering an approach for personalized risk prediction.

Prostate cancer incidence varies across ethnic groups and is approximately 75% higher in African Americans and 45% lower in Asians, compared with non-Hispanic Whites.[1] Age, family history of prostate cancer and germline variation are the most established risk factors for prostate cancer, with as much as 57% of the variability in prostate cancer risk estimated to be due to genetic factors.[2] Accordingly, it is hypothesized that genetic factors are likely to contribute, in part, to ethnic disparities in prostate cancer incidence.[3] Genome-wide association and fine-mapping studies of prostate cancer have been conducted mainly in populations of European ancestry and have discovered ~180 germline risk variants for prostate cancer, with some more frequent in specific populations.[4-14] Genetic risk scores (GRS) comprised of these variants have been demonstrated to identify men at higher risk of prostate cancer; however, they have been developed and optimized for populations of European ancestry.[12]

In this study, we combined data from genome-wide association studies (GWAS) for 107,247 prostate cancer cases and 127,006 controls, including men from European, African, East Asian and Hispanic populations, to identify common genetic variants associated with disease risk across populations. We also developed a GRS for prostate cancer to evaluate risk stratification due to genetic factors across racial and ethnic groups, with GRS validation conducted in two independent studies. Based on the GRS, we estimated relative risks for ethnic differences in prostate cancer risk as well as lifetime and age-specific absolute risks of prostate cancer due to genetic factors.

Results

Multiethnic GWAS meta-analysis. The multiethnic meta-analysis was based on summary statistics from 85,554 prostate cancer cases and 91,972 controls of European ancestry, 10,368 cases and 10,986 controls of African ancestry, 8,611 cases and 18,809 controls of East Asian ancestry and 2,714 cases and 5,239 controls of Hispanic ethnicity that are part of the Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome and Collaborative Oncological Gene-Environment Study Consortium (PRACTICAL iCOGS), the Elucidating Loci Involved in Prostate Cancer Susceptibility OncoArray Consortium (ELLIPSE OncoArray), the United Kingdom GWAS (UK GWAS1 and UK GWAS2), Cancer of the Prostate in Sweden (CAPS1 and CAP2), the National Cancer Institute (NCI) Prostate cancer Genome-wide Association Study of Uncommon Susceptibility loci study (PEGASUS), the NCI Breast and Prostate Cancer Cohort Consortium (BPC3), the ProHealth GWAS Study within the Research Program on Genes, Environment and Health Kaiser Permanente cohort (ProHealth Kaiser GWAS), the African Ancestry Prostate Cancer Consortium (AAPC GWAS), BioBank Japan (RIKEN GWAS1 and GWAS2), GWAS of prostate cancer in Latinos (LAPC GWAS) and Japanese (JAPC GWAS) in the Multiethnic Cohort Study (MEC) and the Ghana Prostate Study (GPS) (**Online Methods**, **Table 1** and **Supplementary Table 1**). Ethnicity was self-reported, with the additional exclusion of men whose genetic ancestry was inconsistent with a self-report of either African, Asian, or European ancestry (**Online Methods**). Imputation in each study was performed using the October 2014 (Phase 3) release of the 1000 Genomes Project[15] data as the reference panel. Across the studies, 5.8-16.8M genotyped and imputed SNPs as well as insertion/deletion variants with ≥ 1% frequency were examined in association with prostate cancer risk (**Supplementary Table 2**). We performed a fixed-effects

meta-analysis within populations and overall, and inflation statistics ranged from 1.03 (Hispanic) to 1.25 (East Asian), with the corresponding λ_{1000} (i.e. an inflation statistic scaled to a sample size of 1,000 cases and 1,000 controls) ranging from 1.002 to 1.022. The overall multiethnic meta-analysis GWAS had a λ of 1.13 and λ_{1000} of 1.001 (**Supplementary Table 3** and **Supplementary Fig. 1**).

In combining summary statistics of single variant tests from analyses of 107,247 prostate cancer cases and 127,006 controls (**Table 1**), we identified 269 independent genetic loci associated with prostate cancer risk at the genome-wide significance threshold of P-value < 5.0×10^{-8}, including 86 novel loci, defined as newly reported loci that were not correlated with known prostate cancer risk variants (**Supplementary Fig. 2** and **Supplementary Table 4**). Of the 86 novel associations, 36 were genome-wide significant for at least one ancestry group (32 for men of European ancestry, 1 for men of African ancestry and 5 for men of East Asian ancestry). Thirty-three of the novel risk variants were located within 1 megabase of a previously reported risk variant and were independently associated with risk in analyses conditioning on previously discovered risk variants in the region (**Online Methods**). Of the 183 previously reported prostate cancer risk variants, 121 variants or close proxies ($r^2 > 0.9$ in men of European ancestry) were observed to remain the lead signal in these regions, while stronger markers of risk were discovered for 62 variants (**Supplementary Table 4**). Of the 269 risk variants, eight were poorly imputed and replaced with suitable surrogate variants with imputation scores > 0.8 across studies and populations (**Supplementary Table 5**).

In multiethnic case-only analyses, the 269 risk variants were generally equally associated with risk of aggressive disease (i.e. high-risk), defined as tumor stage T3/T4, regional lymph node involvement, metastatic disease, Gleason Score ≥ 8, a

prostate-specific antigen (PSA) level ≥ 20 ng/mL, or prostate cancer as the underlying cause of death, and non-aggressive disease (i.e. intermediate and low-risk), defined as Gleason ≤ 7, PSA < 20 and stage ≤ T2 (**Supplementary Table 6**). Exceptions were nominally significant (P < 0.05) inverse associations (OR < 0.9) observed with variants at the *KLK3* locus on chromosome 19 (rs76765083, OR = 0.71, P = 1.54×10^{-39} and rs61752561, OR = 0.89, P=1.43×10^{-4}) and positive associations (OR > 1.1) observed with variant rs183373024 at 8q24 (OR = 1.14, P = 0.0047) and non-synonymous variant rs138708 (NP_001186508.1:p.Arg369Cys) in the *SUN2* gene on chromosome 22 (OR = 1.12, P = 0.01) (**Supplementary Table 6**).

In multiethnic case-only analyses, 105 of the 269 risk variants were nominally associated (P < 0.05) with age at prostate cancer diagnosis (only three were nominally associated with older age at prostate cancer diagnosis), with 15 associated at P-value threshold < 5×10^{-8}, including rs76765083 in *KLK3* (0.78 years younger at diagnosis per allele, multiethnic P-value = 4.1×10^{-20}), rs10993994 upstream of *MSMB* (0.33, multiethnic P-value = 1.2×10^{-18}), rs72725854 at 8q24 (1.46, African P-value = 7.1×10^{-15}), rs183373024 at 8q24 (1.19, multiethnic P-value = 1.5×10^{-15}) and *HOXB13* variant rs138213197 (1.55, European P-value = 1.2×10^{-10}) (**Supplementary Table 7**). In age-stratified case-control analyses, 188 of the 269 variants (69.9%) had larger effects in younger (≤55 years) compared to older (>55 years) men, 31 of which differed with a nominal P-value < 0.05 (**Supplementary Table 8** and **Extended Data 1**).

European versus African ancestry effect estimates (odds ratios) of the 269 risk variants were correlated with an r = 0.45, while European versus East Asian ancestry estimates were correlated at r = 0.37 and estimates for men of European

ancestry versus Hispanic men were correlated at r = 0.51 (**Extended Data 2**). In comparing risk allele frequencies of the 269 risk variants across populations, average frequencies were similar between men of European ancestry (0.490), African ancestry (0.494) and Hispanic men (0.492) and were lowest in men of East Asian ancestry (0.479). However, variants with multiethnic odds ratios > 1.10 (71 variants, 26.4%) were on average more common in men of African ancestry (average risk allele frequency: 0.509 for men of African ancestry, 0.482 for men of European ancestry, 0.472 for men of East Asian ancestry and 0.476 for Hispanic men; **Supplementary Table 9**).

Based on a familial risk estimate of 2.5 for prostate cancer[16], the 269 risk variants were estimated to capture 33.6% of familial relative risk (FRR) in men of East Asian ancestry, 39.3% in Hispanic men, 42.6% in men of European ancestry, and 43.2% in men of African ancestry (**Supplementary Table 10**). The 86 newly identified prostate cancer risk variants alone capture 5.4% of the FRR in men of European ancestry, 5.7% in both Hispanic men and men of East Asian ancestry, and 6.5% in men of African ancestry, which corresponds to 12.8-17.1% of the total FRR represented by the 269 risk variants.

Risk variant annotation. *In silico* annotation of the 269 lead variants re-affirmed known prostate cancer susceptibility genes and identified a number of new strong candidate genes that may be involved in prostate tumorigenesis. (**Supplementary Table 11**). Fourteen of the lead variants are non-synonymous in 12 unique genes, two are situated in the 5'UTR and five in the 3'UTR of a gene, including a novel variant within the 3'UTR of the tumor suppressor *TP53*, for which a role in tumorigenesis is well established.[17] We have also established the cancer-related

458 1100delC frameshift deletion in *CHEK2* (NP_009125.1:p.Thr367fs)[18] as a genome-
459 wide significance risk variant for prostate cancer. A number of other lead variants
460 demonstrate high or moderate evidence for regulatory potential, intersecting putative
461 enhancer, repressor or promoter sites (**Supplementary Table 11**). For example,
462 rs111595856 is located upstream of *INHBB* and is an expression quantitative trait
463 loci (eQTL) for Inhibin subunit Beta B, a member of the transforming growth factor-
464 beta superfamily involved in pituitary and gonadal hormone secretion and endocrine-
465 related cancers, including prostate cancer.[19] We observed overlap with a significant
466 eQTL signal for 133 of the 269 lead variants (49.5%) in one or more prostate tissue
467 datasets (**Online Methods**), including 36 of the 86 novel risk variants (41.9%), with
468 265 unique eGenes (genes for which expression is significantly associated with an
469 eQTL) represented by the 133 lead variants (**Supplementary Table 12**). It is notable
470 that of the 269 lead variants, 54 are situated within or adjacent to, or are associated
471 with expression of, a transcription factor[20], of which seven are enriched in prostate
472 tissue in the Human Protein Atlas.[21,22] An example includes *SOX14* on chromosome
473 3, where the novel risk variant also intersects binding sites for regulatory factors *AR*,
474 *FOXA1* and *HOXB13* involved in prostate cancer.
475
476 **Developing genetic risk scores for prostate cancer.** To understand the aggregate
477 effect of the 269 variants on prostate cancer risk, we constructed a genetic risk score
478 (GRS) using the multiethnic weights of the risk variants associated with disease
479 (**Online Methods**). Compared with men at average genetic risk in the 40-60% GRS
480 category, the estimated odds ratio for men in the top 10% of the GRS (90-100%
481 GRS category) was 5.06 [95% CI 4.84-5.29] for men of European ancestry, 3.74
482 [95% CI 3.36-4.17] for men of African ancestry, 4.47 [95% CI 3.52-5.68] for men of

East Asian ancestry and 4.15 [95% CI 3.33-5.17] for Hispanic men (**Table 2**). Men in the top 1% of the GRS distribution (99-100%) had higher odds of disease, ranging from 11.65 [95% CI 10.56-12.85] for men of European ancestry to 5.68 [95% CI 4.44-7.28] for men of African ancestry. Category specific GRS risk estimates were very similar using weights from bias corrected estimates (**Online Methods**, **Supplementary Table 13**). GRS differences by population were comparable when using weights based on similar sample sizes of each population and equal weights for the 269 variants (**Online Methods** and **Supplementary Table 14**).

We examined GRS replication in two independent studies in men of European ancestry from the UK Biobank and in men of African ancestry from the California and Uganda (CA UG) study, neither of which were included in the multiethnic GWAS meta-analyses; additional studies in Asian and Hispanic men are currently not available for GRS replication in these groups. The GRS associations with prostate cancer risk replicated in both men of European and African ancestry (**Table 2**). For men of European ancestry, the odds ratio was 4.17 [95% CI 3.85-4.51] for those in the top 10% of the GRS and 9.03 [95% CI 7.87-10.35] for those in the top 1%. For men of African ancestry, the odds ratio was 3.53 [95% CI 2.66-4.69] for those in the top 10% of the GRS and 7.05 [95% CI 3.66-13.56] for those in the top 1%.

The discriminative improvement of the GRS was evaluated in the UK Biobank using area under the curve (AUC). Compared to a model of age and family history (AUC = 0.784, 95% CI 0.779-0.789), incorporating the GRS into the model resulted in improved discrimination (AUC = 0.836, 95% CI 0.832-0.840, Δ = +0.052). Comparatively, a model of age and GRS (AUC = 0.833, 95% CI 0.828-0.837) was minimally improved upon incorporating family history (AUC = 0.836, 95% CI 0.832-0.840, Δ = +0.003; **Online Methods** and **Supplementary Table 15**). In the UK

Biobank, relative to a model of age and family history, the addition of the GRS to the risk model also resulted in a 59.5% (95% CI 57.1-62.1%) net reclassification improvement (NRI), with similar improvement observed in both cases (29.4%, 95% CI 27.6-31.1%) and controls (30.2%, 95% CI 29.1-31.4%; **Online Methods** and **Supplementary Table 15**).

We also derived a genome-wide GRS that included the 269 genome-wide significant risk variants and additional variants independently associated ($r^2 < 0.10$ and > 800 kb from the 269 variants) with prostate cancer with a P-value < $1.0x10^{-5}$ from the multiethnic meta-analysis (605 total variants) (**Online Methods**). While effect sizes were typically larger for the genome-wide GRS than the 269-variant GRS in the discovery sample, associations with the genome-wide GRS and 269-GRS were similar in the replication studies of men of European ancestry from the UK Biobank and men of African ancestry from the CA UG study (**Supplementary Table 15 and 16** and **Extended Data 3**). A genome-wide GRS was similarly constructed based on the African ancestry meta-analysis (917 total variants) (**Online Methods**); however, performance was poorer for men of both European and African ancestry (**Supplementary Table 17** and **Extended Data 4**).

The relationship between GRS, age at diagnosis, family history and prostate cancer risk. We found the GRS to be significantly associated with younger age at diagnosis in each population. Men with prostate cancer in the top 10% of the GRS distribution were diagnosed 2.84 years younger (95% CI -3.24, -2.44, P-value = $4.1x10^{-44}$) on average, while men in the top 1% were diagnosed 3.88 years younger (95% CI -4.31, -3.44) on average than men in the bottom 10% across populations (**Extended Data 5** and **Supplementary Table 18**). Men of both European and

533 African ancestry with prostate cancer in the top 10% of the GRS were also 2.0-fold
534 (95% CI 1.78-2.64, P = $1.4x10^{-14}$) more likely to have a first-degree family history of
535 prostate cancer compared to men in the bottom 10% (**Extended Data 6** and
536 **Supplementary Table 19**).

537 We also found age to modify the GRS association with prostate cancer risk for
538 men in higher GRS categories (**Supplementary Table 20**). In men of European
539 ancestry included in the GWAS meta-analysis (**Fig. 1A**), the top decile GRS
540 category was associated with an odds ratio of 6.71 [95 % CI 5.99-7.52] for men ages
541 55 years or younger and 4.39 [95% CI 4.19-4.60] for men older than 55 years (P-
542 heterogeneity for age = $1.5x10^{-11}$). Effect modification of the GRS by age was
543 similarly observed in men of African ancestry (P-heterogeneity = 0.02) and men in
544 the UK Biobank (P-heterogeneity = 0.004) (**Fig. 1A**, **Fig. 1B** and **Supplementary**
545 **Table 20**). Odds ratios were even greater for the top 1% of the GRS (99-100%
546 category) for younger men of European and African ancestry ages 55 years or
547 younger (**Fig. 1A** and **1B**). We did not observe evidence of effect modification of the
548 top GRS decile by family history of prostate cancer in men of European or African
549 ancestry (P-heterogeneity = 0.29 and 0.34, respectively; **Supplementary Table 21**).
550

551 **The relationship between GRS and disease aggressiveness.** We observed no
552 evidence of the GRS differentiating risk of aggressive versus non-aggressive
553 prostate cancer (i.e. case-only odds ratios in each decile were ~1 and case-control
554 odds ratios were similar for cases with non-aggressive and aggressive phenotypes
555 versus controls in stratified analyses; **Supplementary Table 22** and **23**). However,
556 45-51% of all men with prostate cancer in these populations have a GRS in the top
557 20% (**Extended Data 7** and **8**). Thus, while the GRS does not predict who is more

likely to develop aggressive disease (vs. non-aggressive disease), it can define a subset of men (i.e. 20% of the population) in which a substantial fraction of aggressive cases will develop.

Comparing GRS distributions across populations. In comparing the GRS across populations, we found that the GRS distribution in controls was higher for men of African ancestry and lower for men of East Asian ancestry compared with men of European ancestry (**Fig. 2**). Relative to the mean prostate cancer GRS for men of European ancestry, 20% of men of European ancestry, 54% of men of African ancestry, 9% of men of East Asian ancestry and 18% of Hispanic men had a relative risk for the GRS greater than 2.0. Using the GRS distribution in controls, compared to the mean prostate cancer GRS in men of European ancestry, men of African ancestry had a mean prostate cancer GRS that was associated with a relative risk of 2.18 [95% CI 2.14-2.22], while Hispanic men and men of East Asian ancestry had relative risks of 0.97 [95% CI, 0.94-1.00] and 0.73 [95% CI 0.71-0.76], respectively. Within the admixed African and Hispanic populations, associations were similar in GRS analyses stratified by global European ancestry (**Supplementary Table 24**). All tests of heterogeneity had a P-value > 0.40 (**Online Methods**).

Estimating absolute risk of prostate cancer by GRS. Lifetime absolute risks of prostate cancer by GRS category and ethnic group are shown in **Fig. 3** (**Supplementary Table 25**). The absolute risk for men in the top decile of the GRS reached 38% for both men of African [95% CI 36%-41%] and European [95% CI 37%-39%] ancestry, 31% [95% CI 27%-36%] for Hispanics and 26% [95% CI 22%-30%] for East Asians. Absolute risk estimates were only slightly reduced when using

583 GRS estimates from men of European and African ancestry in the UK Biobank and
584 CA UG replication studies, respectively (**Extended Data 9** and **Supplementary**
585 **Table 25**). Men with a first-degree family history of prostate cancer had increased
586 absolute risks for each GRS category, with 67% [95% CI 59%-76%] and 56% [95%
587 CI 52%-60%] lifetime absolute risks estimated for men in the top 10% for African and
588 European ancestry men, respectively (**Supplementary Table 26** and **Extended**
589 **Data 10**).
590
591 **Discussion**
592 Through this large multiethnic GWAS meta-analysis, we identified 86 novel risk
593 variants that influence prostate cancer susceptibility and point to a number of novel
594 candidate genes potentially involved in prostate cancer development. We integrated
595 these discoveries with known risk loci for prostate cancer to derive a GRS based on
596 269 risk variants for prostate cancer that could effectively stratify prostate cancer risk
597 across populations, with GRS associations replicating in two independent studies in
598 men of European and African ancestry.
599 The inclusion of non-European ancestry samples, especially those of African
600 ancestry, allows for better refinement of signal(s) within regions.[23] However, the
601 discovery of novel variants and lead variants in known regions was largely
602 determined by the size of the European ancestry sample, which represented 79.8%
603 of the cases included in the GWAS. The smaller sample size of the African, Hispanic
604 and East Asian studies resulted in an imbalance in the discovery of risk variants and
605 in the precision of risk estimation in these groups. Because of this, for each variant,
606 we used the multiethnic weight in the GRS estimation, as the effect is likely to more
607 closely reflect that of the underlying causal allele, assuming little or no effect

heterogeneity by population. While inflation of the GRS associations could result from using the same sample for risk variant discovery as GRS testing, the GRS predictive ability was comparable in the independent UK Biobank and CA UG studies, and sensitivity analyses incorporating weights with a bias correction had little impact on GRS associations.

Despite population sample size differences, the magnitudes of GRS associations were similar across populations, except for men of African ancestry, in which the odds ratio in the top GRS decile was attenuated by ~20% for men of African ancestry compared to men of European ancestry. This consistency of GRS performance across ancestral populations has not generally been observed for GRS derived for cancers or many other diseases or traits[24] and is likely the result of prostate cancer having a strong genetic component, the multiethnic approach we employed, which allowed for the discovery of novel pan-ethnic variants and the refinement of lead variants in known risk regions, and the use of multiethnic weights in the GRS. However, GRS distributions were observed to vary widely across populations, signifying the importance of incorporating an individual's ancestry before GRS-associated risk can be assigned to an individual, particularly for admixed populations.

While larger GRS effect sizes were observed in men of European ancestry, the greater disease incidence for men of African ancestry resulted in our reporting comparable lifetime risk estimates for GRS deciles. Ethnic-specific GRS cutoffs were used to determine the 10% of men in each population at highest risk, who had estimated lifetime risks of developing prostate cancer that ranged from 38% for African and European ancestry men, 31% for Hispanic men and 26% for East Asian men. Estimated lifetime risks for men in the top GRS decile were > 50% for African

633 and European ancestry men who also had a family history of prostate cancer.

634 We found little evidence that a genome-wide GRS improved risk prediction
635 beyond the 269-variant GRS. Of the 269 variants, those with odds ratios > 1.10,
636 which have a larger contribution to the GRS than variants with weaker effects (odds
637 ratios ≤ 1.10), were more common in men of African ancestry, resulting in a greater
638 contribution of the GRS to the overall risk of prostate cancer for this ancestry group.
639 Based on our observed 2-fold difference in the mean GRS distribution in controls
640 between men of European and African ancestry, in aggregate, the known risk
641 variants are estimated to account for a substantial fraction of the ~70% greater
642 prostate cancer incidence observed in men of African ancestry. However, it will be
643 important to incorporate the biologically functional variants and local ancestry
644 differences in order to better understand how GRS distributions relate to population
645 differences in prostate cancer incidence.

646 For men between 55 and 69 years of age, the U.S. Preventive Task Force
647 recommends that the decision to undergo PSA screening should be an individual
648 one, following consultation with a physician and considering information about family
649 history of prostate cancer and African ancestry.[25] Currently, genetic information is
650 not incorporated into the decision-making process for PSA screening. However, men
651 with a high GRS may benefit from earlier and more frequent screening, while
652 knowledge of a low GRS may help to reduce unnecessary biopsies for men with
653 borderline screening PSA levels. While the lifetime risk of developing prostate cancer
654 is heavily dependent on age, the odds ratio associated with the top GRS decile was
655 greater for younger compare to older men. For cancer, younger age at diagnosis
656 typically indicates a genetic influence on disease onset, which is supported by our
657 findings of common genetic variants having a greater impact on prostate cancer risk

for earlier versus later onset disease. As such, regular PSA screening may be beneficial even earlier than age 55 for a subset of men at high genetic risk.

Consistent with previous findings, we found that common variants are equally associated with risk of aggressive and non-aggressive prostate cancer. Although we found little evidence that the GRS can differentiate risk of aggressive versus non-aggressive disease, the GRS could define ~20% of men in each ancestral and ethnic population at high risk, which includes one-half of the men who will be diagnosed with aggressive disease. While the benefit/harm tradeoffs of including GRS in future risk-tailored screening programs need to be evaluated, these data suggest that GRS greatly improves upon discriminative models based on age and family history and that a substantial fraction of men who will develop aggressive tumors may be identified earlier through risk-based screening.

In summary, we have applied a multiethnic approach to discover novel risk variants for prostate cancer, refine lead variants in known risk regions and develop a GRS for prostate cancer that is effective in stratifying prostate cancer across populations. These findings also provide further support for a contribution of germline variation to ethnic differences in prostate cancer incidence. The clinical benefit of GRS profiling for targeted screening and early diagnosis needs to be examined, and larger prostate cancer consortia in men of non-European ancestry, particularly in men of African ancestry, will be required to identify additional risk variants, improve precision of risk estimation and enhance the predictive ability of the GRS across populations.

Acknowledgements

This project was support by the US National Institutes of Health (NIH) grants U19CA148537 (CAH), U01CA194393 (SL), and K99CA246063 (BFD). We acknowledge the ARCS Foundation, Inc., Los Angeles Chapter, for their generous support of L. Moss through the Margaret Kirsten Ponty Fellowship and B. Darst through the John and Edith Leonis Family Foundation. This research has been conducted using the UK Biobank Resource under application number 42195. A full description of funding and acknowledgements for each of the contributing studies can be found in the Supplementary Note.

Author Contributions

DVC, RAE, ZK-J and CAH contributed to study conception, and DVC, BFD, RAE, ZK-J and CAH contributed to interpretation and wrote the manuscript. EJS performed a literature search. MB, TD, SB and XS provided data management and bioinformatics support. DVC, LM, BD, EJS, TD, ZK-J and CAH contributed to data analysis and interpretation. All authors contributed data to the study, revised, critically reviewed and approved the final version of the manuscript: DVC, BDF, LM, EJS, XS, AC, FRS, AAAO, SB, TD, MNB, AS, TJH, AT, KM, YM, MF, KM, AL, PW, LLM, LRW, VLS, SMG, BDC, JS, TLJT, CS, AA, GGG, MCS, RJM, CC, DW, JL, DEN, JLD, FCH, RMM, BGN, SFN, MW, SEB, MAR, PI, JB, SC, LM, LH, JAC, WT, GPR, HG, MA, RS, ME, TN, NP, AMD, MG, RCT, TJK, ER, JYP, TAS, H-YL, DA, SW, LAM, EG, SL, PK, DJH, KLP, CT, CMT, PJG, IAM, RJH, NEF, AF, M-ÉP, JLS, EAO, MSG, SK, LEBF, MS, AW, NH, GLA, RNH, MJM, KDS, MB, WJB, WZ, EDY, JEM, Y-JL, H-WZ, NF, XM, YW, S-CZ, ZS, SNT, SKM, DJS, CMLW, NB, GB, CM,

TS, ML, ASK, BFD, OC, GC-T, FM, TT, YAK, EMJ, EMG, LM, K-TK, SAI, MCS, AV, AG-C, LF, BSR, SLK, HO, MRT, PP, AB, SW, AL, JTB, ETHF, JM, JAT, MK, JL, GC-V, LC-A, CCT, CDH, SSS, LM, PB, LB, RK, CS, VM, RJL, BW, HB, KC, BH, K-US, EAK, AWH, RAK, ABM, CJL, JK, SLN, LS, YCD, WBI, BN, AJMH, JC, HP, AM, KDR, GDM, PO, JX, AR, JL, S-HT, LFN, DWL, JHF, CN-D, BAR, MG, DL, TK, NU, SS, MP, FC, SJ, TVB, MG-D, JEC, MEM, SL, PAT, CA-H, WSB, MCA, DCC, SS, JCC, GP, GC, MJR, GJ, RHNS, JJH, MS, RV, RM-C, MT, NM, SIB, SKVDE, DFE, SJC, MBC, FW, HN, JSW, RAE, ZK-J, CAH. CAH and RAE had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing Interests Statement

RAE reports the following disclosures: 1) GU-ASCO meeting in San Francisco (Jan 2016) – Received $500 honorarium as speaker; 2) RMH FR meeting (Nov 2017) – received support from Janssen and £1100 honorarium as speaker; 3) University of Chicago invited talk (May 2018) – received $1000 honorarium as speaker; 4) EUR 200 education honorarium paid by Bayer & Ipsen to attend GU Connect “Treatment sequencing for mCRPC patients within the changing landscape of mHSPC” at a venue at ESMO, Barcelona (Sept 2019); 5) Prostate Dx Advisory Panel – Member of external Expert Committee (June 2020) / 3 hours / £900.

References

1. Groups., U.S.C.S.W. U.S. Cancer Statistics Data Visualizations Tool, based November 2018 submission data (1999-2016). *U.S. Department of Health and Human Services, Centers for Disease Control and Preventions and National Cancer Institute* **www.cdc.gov/cancer/dataviz,** (June 2019).
2. Mucci, L.A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68-76 (2016).
3. Freedman, M.L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* **103**, 14068-73 (2006).
4. Al Olama, A.A. *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* **46**, 1103-9 (2014).
5. Amundadottir, L.T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat Genet* **38**, 652-8 (2006).
6. Conti, D.V. *et al.* Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry. *J Natl Cancer Inst* **109**(2017).
7. Dadaev, T. *et al.* Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nat Commun* **9**, 2256 (2018).
8. Eeles, R.A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* **45**, 385-91, 391e1-2 (2013).
9. Gudmundsson, J. *et al.* A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nat Genet* **44**, 1326-9 (2012).

10. Gudmundsson, J. *et al.* Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* **40**, 281-3 (2008).
11. Hoffmann, T.J. *et al.* A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov* **5**, 878-91 (2015).
12. Schumacher, F.R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* (2018).
13. Takata, R. *et al.* Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* **42**, 751-4 (2010).
14. Wang, M. *et al.* Large-scale association analysis in Asians identifies new susceptibility loci for prostate cancer. *Nat Commun* **6**, 8469 (2015).
15. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
16. Kicinski, M., Vangronsveld, J. & Nawrot, T.S. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. *PLoS One* **6**, e27130 (2011).
17. Bode, A.M. & Dong, Z. Post-translational modification of p53 in tumorigenesis. *Nat Rev Cancer* **4**, 793-805 (2004).
18. Dong, X. *et al.* Mutations in CHEK2 associated with prostate cancer risk. *Am J Hum Genet* **72**, 270-80 (2003).
19. Dowling, C.R. & Risbridger, G.P. The role of inhibins and activins in prostate cancer pathogenesis. *Endocr Relat Cancer* **7**, 243-56 (2000).
20. Lambert, S.A. *et al.* The Human Transcription Factors. *Cell* **172**, 650-665 (2018).

21. O'Hurley, G. *et al.* Analysis of the Human Prostate-Specific Proteome Defined by Transcriptomics and Antibody-Based Profiling Identifies TMEM79 and ACOXL as Two Putative, Diagnostic Markers in Prostate Cancer. *PLoS One* **10**, e0133449 (2015).
22. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol* **28**, 1248-50 (2010).
23. Zaitlen, N., Pasaniuc, B., Gur, T., Ziv, E. & Halperin, E. Leveraging genetic variability across populations for the identification of causal variants. *Am J Hum Genet* **86**, 23-33 (2010).
24. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* **10**, 3328 (2019).
25. Moyer, V.A. & Force, U.S.P.S.T. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* **157**, 120-34 (2012).

Figure Legends

Figure 1: Odds ratio for prostate cancer by GRS category stratified by age. Results are shown for A. Men of European ancestry (N=124,101 from the GWAS and 199,969 from independent replication) and B. Men of African ancestry (N=17,828 from the GWAS and 2,633 from independent replication). The x-axis indicates the GRS category [0-10% (low-risk), 40-60% (average risk), 60-70%, 80-90%, 90-100% (high-risk) and 99-100% (high-risk)]. The y-axis indicates odds ratios with error bars representing 95% CIs for each GRS category compared to the 40-60% GRS as the reference. Odds ratios and 95% CIs for each decile and strata are provided in **Supplementary Table 20**.

Figure 2. Comparison of prostate cancer GRS distributions for controls. A. Men of European ancestry versus men of African ancestry; B. Men of European ancestry versus men of East Asian ancestry; and C. Men of European ancestry versus Hispanic men. The x-axis indicates the relative risk calculated by exponentiation of the difference in the mean GRS in controls for men of European ancestry and the mean GRS in controls for each of the other populations. The y-axis indicates the GRS density. Solid areas and corresponding percentages indicate the proportion of a given population with a relative risk greater than or equal to 2.0 in comparison to the mean GRS for men of European ancestry.

Figure 3. Absolute risks of prostate cancer by GRS category. A. European ancestry; B. African ancestry; C. East Asian ancestry; and D. Hispanic. SEER data is used for mortality and incidence rates corresponding to non-Hispanic White, Black, Asian, and Hispanic men. The x-axis indicates the age of an individual and the y-axis is the

818 absolute risk by a given age.

Tables

Table 1. Baseline Characteristics of the Participants.

	Multiethnic GWAS Sample Population Group										Replication Sample Population Group			
	Total		European		African		East Asian		Hispanic		European (UK Biobank)		African (AFR CA UG)	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
No. of participants	107,247	127,006	85,554	91,972	10,368	10,986	8,611	18,809	2,714	5,239	6,852	193,117	1,586	1,047
No. with individual level data [a]	84,574	65,134	71,570	52,531	9,126	8,702	1,652	1,803	2,226	2,098	6,852	193,117	1,586	1,047
No. ≤ 55 years of age	8,959	13,562	7,099	11,471	1,628	1848	47	81	185	162	481	79,347	354	277
No. with aggressive disease [b]	26,374	-	21,917	-	2,934	-	753	-	770	-	-	-	-	-

[a] These participants are also included in GRS and stratified analyses.

[b] Aggressive disease defined as stage T3/T4, regional lymph node involvement (N1), metastatic disease (M1), a tumor with a Gleason Score ≥ 8, or a prostate-specific antigen (PSA) level ≥ 20 ng/mL, or, prostate cancer as the underlying cause of death.

Table 2. Genetic Risk Score (GRS) by Population.

	Multiethnic GWAS Sample Population Group								Replication Sample Population Group			
GRS Category	European 71,570 cases, 52,531 controls		African 9,126 cases, 8,702 controls		East Asian 1,652 cases, 1,803 controls		Hispanic 2,226 cases, 2,098 controls		European (UK Biobank) 6,852 cases, 193,117 controls		African (CA UG) 1,586 cases, 1,047 controls	
	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
0 - 10%	0.24	0.23 - 0.26	0.30	0.26 - 0.36	0.37	0.26 - 0.55	0.39	0.28 - 0.54	0.28	0.24 - 0.34	0.31	0.21 - 0.47
10 - 20%	0.42	0.40 - 0.45	0.52	0.45 - 0.60	0.48	0.34 - 0.68	0.59	0.44 - 0.79	0.40	0.35 - 0.47	0.49	0.34 - 0.71
20 - 30%	0.57	0.54 - 0.60	0.61	0.53 - 0.70	0.75	0.55 - 1.02	0.69	0.52 - 0.91	0.62	0.55 - 0.71	0.61	0.43 - 0.86
30 - 40%	0.73	0.69 - 0.77	0.77	0.67 - 0.87	0.76	0.56 - 1.03	0.80	0.61 - 1.05	0.79	0.70 - 0.89	0.72	0.52 - 1.01
40 - 60%	1.00	ref.	1.00	ref.	1.00	ref.	1.00	ref.	1.00	ref.	1.00	ref.
60 - 70%	1.36	1.29 - 1.42	1.43	1.27 - 1.60	1.25	0.95 - 1.65	1.46	1.15 - 1.87	1.29	1.17 - 1.43	1.45	1.07 - 1.97
70 - 80%	1.73	1.65 - 1.82	1.63	1.45 - 1.83	1.8	1.42 - 2.39	1.77	1.40 - 2.25	1.62	1.47 - 1.78	1.66	1.23 - 2.23
80 - 90%	2.45	2.34 - 2.56	2.37	2.12 - 2.65	2.37	1.84 - 3.06	2.47	1.97 - 3.11	2.43	2.23 - 2.65	1.78	1.32 - 2.40
90 - 100%	5.06	4.84 - 5.29	3.74[a]	3.36 - 4.17	4.47	3.52 - 5.68	4.15	3.33 - 5.17	4.17	3.85 - 4.51	3.53	2.66 - 4.69
99 - 100%	11.65	10.56 - 12.85	5.68[a]	4.44 - 7.28	9.41	5.60 - 15.82	6.85	4.20 - 11.18	9.03	7.87 - 10.35	7.05	3.66 - 13.56

[a] P-value < 0.001 for heterogeneity testing for each GRS category versus men of European ancestry.

Online Methods

Study Subjects in the Multiethnic GWAS. This investigation includes the Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome and Collaborative Oncological Gene-Environment Study Consortium (PRACTICAL iCOGS), the Elucidating Loci Involved in Prostate Cancer Susceptibility OncoArray Consortium (ELLIPSE OncoArray), the United Kingdom GWAS (UK GWAS1 and UK GWAS2), Cancer of the Prostate in Sweden (CAPS1 and CAP2), the National Cancer Institute (NCI) Prostate cancer Genome-wide Association Study of Uncommon Susceptibility loci study (PEGASUS), the NCI Breast and Prostate Cancer Cohort Consortium (BPC3), the ProHealth GWAS Study within the Research Program on Genes, Environment and Health Kaiser Permanente cohort (ProHealth Kaiser GWAS), the African Ancestry Prostate Cancer Consortium (AAPC GWAS), BioBank Japan (RIKEN GWAS1 and GWAS2), GWAS of prostate cancer in Latinos (LAPC GWAS) and Japanese (JAPC GWAS) in the Multiethnic Cohort Study (MEC) and the Ghana Prostate Study (GPS). In total, 136 studies contributed samples and/or summary statistics to the analysis. An overview of each study is provided in **Supplementary Table 1**. Informed consent was obtained from all participants and study protocols were approved by respective Institutional Review Boards.

Genotyping and Imputation in the Multiethnic GWAS. The genotyping array, sample and variant quality control, imputation and the basic statistical software used for each study or consortium are summarized in **Supplementary Table 2**. Details for each individual study or consortium have been described elsewhere (see references in **Supplementary Table 1**). In general, samples and variants were excluded with a corresponding study-specific sample or genotyping call rate < 95%. Most studies limited variants analyzed to those with a MAF ≥ 1%, although there were exceptions, including

852 the ELLIPSE OncoArray Consortium that included all variants. Most studies screened
853 variants with a test of Hardy-Weinberg equilibrium (with varying significance thresholds),
854 but a few studies did not implement such a screen. Imputation used either MACH[26],
855 Minimac3/Minimac4[27] or IMPUTE2[28] using Phase 3 of the 1000 Genomes Project[15] as
856 the reference panel. Post-imputation variant inclusion criteria included MAF ≥ 1% and an
857 imputation INFO/r^2 ≥ 0.3.
858
859 **Study Subjects Included in GRS Replication**. We used GWAS data for 199,969 men
860 of European ancestry from the UK Biobank (https://www.ukbiobank.ac.uk), which
861 included 6,852 cases and 193,117 controls (**Supplementary Table 1** and **2**). Genotype
862 data was generated in the UK Biobank using the Affymetrix UK Biobank Axiom Array and
863 the Affymetrix UK BiLEVE Axiom Array and imputation was performed using the
864 Haplotype Reference Consortium (HRC), UK10K and 1000 Genomes Project
865 panels.[29] All samples had GWAS data, were genetically identified as male, did not have
866 high heterozygosity or missingness prior to imputation, and were unrelated (2nd degree
867 or higher relationships with a kinship > 0.0884 were excluded).
868 For men of African ancestry, GRS replication was conducted among 1,586 cases and
869 1,086 controls from California and Uganda (CA UG Study) genotyped with the Illumina H3
870 Africa array and imputed using Phase 3 of the 1000 Genomes Project[15] as the reference
871 panel and Minimac4 on the Michigan Imputation Server[27] (**Supplementary Tables 1** and **2**).
872 All samples were genetically identified as male, had a genotyping call rate ≥ 95%, and were
873 unrelated to men in our multiethnic GWAS meta-analysis.
874
875 **Statistical Analysis for GWAS.** Genetic ancestry was estimated using a principal
876 component analysis performed in each study based on uncorrelated single nucleotide

polymorphisms (SNPs). Ancestry/ethnicity was based on self-report with extremely admixed individuals (e.g. ± 4SD outside of ancestry-specific clusters defined with principal components) removed for non-Hispanic population-specific analyses. In total, 29,235,255 variants (SNPs and indels) on autosomal chromosomes 1-22 and the X chromosome were examined for association with prostate cancer risk using logistic regression adjusting for age, sub-study (described in **Supplementary Table 1**) and principal components with PLINK[30], SNPtest[31], or R. Per-allele odds ratios and standard errors from individual studies were combined by a fixed-effects inverse-variance weighted meta-analysis using METAL[32] in ancestry-specific analyses and across all four populations to obtain multiethnic estimates. All statistical tests conducted were two-sided. A marginal P-value less than 5.0×10^{-8} in either the population-specific or multiethnic analysis was used to define statistically significant genetic associations, with regions bounded within +/- 800 kb from the most significant variant. To determine if multiple independent associations exist within each region, we implemented a forward stepwise selection starting with the inclusion of the lowest multiethnic marginal P-value into a multivariate logistic regression model. We used Joint Analysis of Marginal summary statistics (JAM)[33] to obtain population-specific conditional summary statistics from multivariate models. Conditional statistics were combined with an inverse-variance weighted fixed effects meta-analysis to obtain multiethnic conditional summary statistics (**Supplementary Table 4**). Variants with a conditional multiethnic P-value < 5.0×10^{-8} were retained in the model. We excluded variants with a marginal multiethnic P-value > 5.0×10^{-4}, MAF < 1% in all four populations, and correlation $r^2 \geq 0.2$ to any variants included in the current model at each step. Poorly imputed selected variants (n=8) were replaced with suitable surrogate variants with imputation scores > 0.8 across studies and populations (**Supplementary Table 5**).

We conducted stratified case-control and case-case analyses to evaluate the impact of the novel variants on disease aggressiveness (**Supplementary Table 6**). As previously defined[4], aggressive prostate cancer (i.e. high-risk) was defined as tumor stage T3/T4, regional lymph node involvement, metastatic disease, Gleason Score ≥ 8, prostate-specific antigen (PSA) level ≥ 20 ng/mL or prostate cancer as the underlying cause of death and non-aggressive disease (i.e. intermediate and low-risk) was defined as Gleason ≤ 7, PSA < 20 and stage ≤ T2. Studies missing these clinical features were excluded (**Table 1**).

Genetic Risk Score (GRS) Construction. Genetic risk scores (GRS) were constructed using all studies with individual-level data (**Supplementary Table 1**) by summing variant-specific weighted allelic dosages. The initial GRS included the 269 variants independently associated with risk at a genome-wide significance threshold, including established rare (<1% frequency) moderate penetrance risk variants at 8q24 (rs183373024)[9], *HOXB13* (rs138213197, NP_006352.2:p.Gly84Glu)[34] and *CHEK2* (c.1100delC, rs555607708, NP_009125.1:p.Thr367fs)[35] (**Supplementary Table 4**). Specifically, for individual *i*, $GRS_i = \sum_{m=1}^{M} w_m g_{im}$, where g_{im} is the genotype dosage for individual *i* for variant *m* and $w_m$is a variant-specific weight (on the log odds ratio scale) calculated by meta-analyzing the ethnic-specific conditional effects from the JAM analysis using an inverse Z-score weighted fixed effects meta-analysis. An inverse Z-score weight was used rather than an inverse variance weight to up-weight noteworthy population-specific variants that may not have evidence in other populations. *M* is the total number of variants included.

The risk of the GRS on prostate cancer was estimated using indicator variables for the percentile categories of the GRS distribution: [0-10%], (10-20%], (20-30%], (30-

40%], (40-60%], (60-70%], (70-80%], (80-90%], and (90-100%]. An additional analysis was also performed by splitting the top decile into two categories to obtain the GRS risk for the top 1%: (90-99%], (99-100%]. GRS thresholds were determined using the observed distribution among controls for the corresponding ancestry group. Logistic regression was used to estimate odds ratios corresponding to each GRS category, adjusting for principal components, age and sub-study, using the (40-60%] category as the reference. To obtain ethnic-specific GRS estimates, an inverse-variance weighted fixed effects meta-analysis was performed within each population. Multiethnic estimates were obtained via an inverse-variance fixed effects meta-analysis using the ethnic-specific results.

GRS Replication Analysis. We examined the GRS in men of European ancestry in the UK Biobank and African ancestry in the CA UG study; additional studies in Asian and Hispanic men are currently unavailable. Of the 269 variants identified in the multiethnic meta-analysis, 267 were present in the UK Biobank sample, all of which had an imputation info score > 0.50 (median info score=0.99), and 266 were present in the CA UG Study and had an imputation info score > 0.36 (median info score=0.98). The GRS used the multiethnic conditional weights from the previous GRS analysis. Odds ratios were estimated within populations comparing each GRS decile to the 40–60% category using logistic regression models adjusted for age, ten principal components and sub-study (African American vs. Uganda in the CA UG study). GRS models were further evaluated in analyses stratified by age, as described below.

Bias Correction and Sensitivity Analysis for GRS. Since a subset of the data used in the overall multiethnic meta-analysis was initially used to evaluate the GRS, there is the

potential for bias to exist in GRS estimates from these data (note that this does not apply to replication analyses, which were performed in independent samples). As shown in Zhong and Prentice,[12,13] this bias becomes very small as the sample size increases. Given the overall sample size contributing to the multiethnic GWAS, bias potential exists only for very small true variant effects. To correct for this potential bias, the variant-specific weights used in our primary GRS analysis (i.e. the weights from the multiethnic meta-analysis of ethnic-specific conditional JAM effects) were corrected using the approach outlined Zhong and Prentice[12] and used to construct a second GRS to investigate this potential bias (**Supplementary Table 13**).

To investigate the influence of the large sample of European ancestry men on GRS weights, we recalculated weights for the 269 variants limiting the number of European ancestry men to 10,000 cases and 10,000 controls (roughly the same size as the African ancestry sample). Resulting weights were highly correlated with original weights (r^2=95.1%). These weights were used to calculate a GRS, and the association between this GRS and prostate cancer was evaluated. We also developed an equally weighted GRS using the average conditional effect of the 269 variants and evaluated the association between this GRS and prostate cancer.

Discriminative Improvement of GRS. The discriminative improvement of the GRS was evaluated in men of European ancestry from the UK Biobank using area under the curve (AUC) and net reclassification improvement (NRI). AUCs were calculated using four separate logistic regression models of prostate cancer, which included the following variables: 1) age, 2) age and family history of prostate cancer, 3) age and GRS and 4) age, family history and GRS. Each model was additionally adjusted for ten principal components of ancestry. NRI indicates the amount of reclassification improvement of

977 cases and controls resulting from the addition of a variable to a model.[36] NRI was
978 calculated comparing model 2 (age and family history) and model 4 (age, family history
979 and GRS), both of which additionally included ten principal components. These
980 calculations were based on the continuous NRI model, suggested by Pencina et al.[36] to
981 be the most versatile measure of improvement in risk prediction and appropriate for
982 case-control data. The 95% confidence intervals for NRI estimates were calculated using
983 1,000 bootstrap replications.
984
985 **Expanded Genome-Wide GRS.** A genome-wide GRS was developed using 605
986 variants independently associated ($r^2 < 0.10$) with prostate cancer risk at a multiethnic P-
987 value < $1.0x10^{-5}$, which included the 269 variants associated with prostate cancer risk at
988 the genome-wide significance threshold, while excluding variants within 800 kb of these
989 269 variants. Independence was determined using PriorityPruner
990 (prioritypruner.sourceforge.net) and the 1000 Genomes Project[15] reference populations,
991 first identifying independent variants within the AFR, followed by EUR, EAS and AMR
992 populations. Variants with an imputation info score < 0.30 were excluded, as were
993 variants with a MAF < 1% in all four discovery populations. The GRS was constructed
994 using the same individual-level data used in the genome-wide significant GRS, summing
995 allelic dosages weighted by variant-specific marginal multiethnic weights. Odds ratios
996 were estimated for each GRS decile relative to the average 40-60% category, adjusting
997 for principal components, age and sub-study. Ethnic-specific GRS estimates were
998 obtained using an inverse-variance weighted fixed effects meta-analysis performed
999 within each population, and multiethnic estimates were obtained using an inverse-
1000 variance fixed effects meta-analysis performed across the ethnic-specific results. For
1001 comparison, we also calculated the genome-wide GRS using subsets of these variants

with a multiethnic GWAS meta-analysis P-value < $1.0x10^{-6}$ and P-value < $1.0x10^{-7}$, retaining the 269 variants in each. We also calculated the AUC and odds ratio for the 90-100% versus 40-60% GRS categories upon iteratively adding each variant to the GRS, first adding the most significant variants within the list of 269 followed by our identified genome-wide variants, sorted by their multiethnic GWAS meta-analysis P-values.

This process was repeated to develop and test an African ancestry-based genome-wide GRS using 917 variants independently associated (r^2 < 0.10) with prostate cancer risk at an African ancestry P-value < $1.0x10^{-4}$ (this larger P-value was used to identify a comparable number of variants), also including the 269 variants. African ancestry variant-specific weights were used in the African ancestry genome-wide GRS.

Stratification of Risk Estimation for GRS. We investigated the GRS effect stratified by age and first-degree family history of prostate cancer and its association with aggressive disease phenotypes, including Gleason Score and metastatic disease (**Supplementary Tables 20-23**). For age and family history, cases and controls were stratified into age groups (age ≤ 55 vs. age > 55) or family history positive vs. negative. For aggressive disease strata, cases were stratified by disease aggressiveness and corresponding stratified analyses used all controls. Stratified analyses were also performed comparing aggressive cases to non-aggressive cases. Logistic regression was performed with prostate cancer status (either case vs. control or aggressive vs. non-aggressive) as the outcome and GRS categories as the independent predictors, adjusting for principal components, age and sub-study. Ancestry-specific GRS estimates were obtained via an inverse-variance weighted fixed effects meta-analysis performed within each population. Overall multiethnic estimates were obtained via an inverse-variance fixed effects meta-analysis using ancestry-specific results (European and African only). The sample sizes of

the other populations (East Asian and Hispanic) were too small for stratified analyses. Heterogeneity was assessed via a Q-statistic between effect estimates with corresponding tests of significance.

We also estimated the GRS effect stratified by global ancestry in African and Hispanic populations, given the high admixture of these populations, using logistic regression models adjusted for age, sub-study and principal components (**Supplementary Table 24**). Global ancestry estimates were calculated as previously described[6] using RFMix[37] and the 1000 Genomes data.[15] African and Hispanic populations were stratified by their median percentages of global European ancestry (15% and 58%, respectively). Analyses were also performed stratifying Hispanic men by their median percentage of global Amerindian ancestry (37%). Heterogeneity was assessed to determine whether effects differed between those with more versus less European or Amerindian ancestry by adding to logistic regression models an interaction term between the continuous GRS and dichotomized ancestry indicator.

Estimation of Relative Risk for Ancestry/Ethnicity. To estimate the relative risk between ethnic groups due to the GRS, we used the distributions of the GRS in controls across the four populations. As the GRS is calculated on the log odds scale, we can estimate the relative risk between any two populations as the exponential of the difference between the corresponding mean GRS distributions in controls. Specifically, the relative risk comparing population *a* vs. population *b* is given by: $RR_{\text{a vs. b}} = \exp\left[\log\left(\frac{a}{b}\right)\right] = \exp[\log(a) - \log(b)] = \exp\left[\mu_{GRS}^{a} - \mu_{GRS}^{b}\right]$, where μ_{GRS}^{a} is the mean GRS in population *a*. As the difference in means can be viewed as a two-sample test, corresponding standard errors and confidence intervals were calculated in a similar

fashion as a two sample t-test with unequal variance using the observed population means, μ^a_{GRS}, standard deviations, σ^a_{GRS}, and corresponding sample sizes for controls.

Age-Specific Absolute Risk Estimation. As an alternative way to investigate the impact of the GRS, we calculated the absolute risk for a given age for each GRS category and each ethnicity.[38-41] The approach constrains the GRS-specific absolute risks for a given age to be equivalent to the age-specific incidences for the entire population. In other words, age-specific incidence rates are calculated to increase or decrease based on the GRS category estimated risk and the proportion of the population within the GRS category. The calculation accounts for competing causes of death.

Specifically, for a given ethnic group and a given GRS risk category *k* (e.g. 80-90%, 90-100%), the absolute risk by age *t* is computed as: $AR_k(t) = \sum_0^t P_{ND}(t)\, S_k(t) I_k(t)$. This calculation consists of three components:

(1) $P_{ND}(t)$ is the probability of not dying from another cause of death by age *t* using age-specific mortality rates, $\mu_D(t)$: $P_{ND}(t) = \exp[-\sum_0^t \mu_D(t-1)]$. Age-specific mortality rates are provided from a reference cohort.

(2) $S_k(t)$ is the probability of surviving prostate cancer by age *t* in the GRS category *k* and uses the prostate cancer incidence by age *t* for category *k*: $S_k(t) = \exp[-\sum_0^t I_k(t-1)]$.

(3) The prostate cancer incidence by age *t* for GRS category *k* is $I_k(t)$ and is calculated by multiplying the population prostate cancer incidence for the reference category, $I_0(t)$ and the corresponding risk ratio for GRS category *k,* as estimated from the odds ratio obtained from the population-specific individual-level GRS analysis as described above: $I_k(t) = I_0(t)\exp(\beta_k)$.

Prostate cancer incidence for age *t* for the reference category, $I_0(t)$, is obtained by constraining the weighted average of the population cancer incidences for the GRS categories to the population age-specific prostate cancer incidence, $\mu(t)$. $I_0(t) = \mu(t)\frac{\sum_K f_k S_k(t-1)}{\sum_K f_k S_k(t-1)\exp(\beta_k)}$, where f_k is the frequency of the GRS category *k* with $f_k = 0.1$ for all non-reference categories in our primary GRS analysis by deciles (e.g. [0-10%], (10-20%], (20-30%], etc.).

By leveraging the definition that $S_k(t=0) = 1$, for all *k*, the absolute risks were calculated iteratively by first getting $I_0(t=1)$, then $I_k(t=1)$, then $S_k(t=1)$ and finally $AR_k(t=1)$. Subsequent values were then calculated recursively for all *t*. Confidence intervals for absolute risk estimates were obtained via a parametric bootstrap repeating the above calculations for 1,000 bootstraps with the β_k's sampled from their corresponding estimated distributions using the standard error of the estimate.

For each ethnic group, absolute risks by age *t* were calculated using age-specific prostate cancer incidence, $\mu(t)$, from the Surveillance, Epidemiology, and End Results (SEER) Program (1999-2013)[10] and age-specific mortality rates, $\mu_D(t)$, from the National Center for Health Statistics, CDC (1999-2013).[11] Using the same analytic framework, absolute risks were also calculated using the family history stratified estimates for the GRS combined with mortality and incidence rates estimated from men from the Multiethnic Cohort (MEC) with a positive family history of prostate cancer. Rates were based on 35,711 White and African American men and 4,060 incidence cases identified over a 20-year period (1993-2013). For absolute risks in those with a positive family history, the log odds ratio estimates, β_k, were obtained from the corresponding stratified analysis.

Proportion of familial risk explained. The contribution of the 269 variants to the familial risk (i.e. sibling recurrence risk) of prostate cancer was computed using the formula: $\frac{\sum_k(log\lambda_k)}{(log\lambda_0)}$, where λ_0 is the observed familial risk to first degree relatives of prostate cancer cases, assumed to be 2.5[16], and λ_k is the familial relative risk (FRR) due to locus *k*, given by: $\lambda = \frac{p_k r_k^2 + q_k}{(p_k r_{k+ q_k})^2}$, where p_k is the frequency of the risk allele for locus *k*, $q_k = 1 - p_k$ and r_k is the estimated per-allele odds ratio.[42,43]

In Silico Annotation. The 269 variants selected in the multiethnic conditional analysis were annotated for putative evidence of biological functionality (**Supplementary Table 11**) using publicly available datasets according to the framework described by Dadaev et al.[7]

Variants were annotated for genomic context and proximity to genes (ENSEMBL/Gencode definitions) using wANNOVAR[44], with additional manual review of exonic variants. Annotation of variants against intersection with chromatin marks indicative of regulatory DNA regions were performed relative to peak data from publicly available datasets conducted in the prostate derived cell-lines LNCaP, PC3, PrEC and VCaP. Peak data were analyzed according to a standardized pipeline and QC procedures were downloaded from the Cistrome Data Browser[45] (http://cistrome.org/db/) and converted from GRCh38 to GRCh37/hg19 reference assembly co-ordinates in R using rtracklayer v1.42.2 liftOver.[46] Variants were assessed for intersection within DNaseI hypersensitivity site peaks in three datasets (GSM1024742, GSM736565 and GSM822387) and ATAC-seq peaks in three datasets (GSM2186481, GSM3075372 and GSM3075374). Histone modification site data was obtained for H3K27Ac (GSM1249447, GSM1249448 and ENCSR826UTD_1), H3K9Ac (GSM2527582 and GSM2527583),

H3K4me1 (GSM1145323 and GSM2187238), H3K4me2 (GSM353635 and GSM1891829) and H3K4me3 (GSM1383874 and GSM945240). Transcription factor-binding site ChIP-seq peak data were obtained for the Androgen Receptor (GSM1274871, GSM1576447 and GSM1527834), CTCF (GSM1006874 and GSM2825574), ERG (GSM1193657 and GSM1328978), FOXA1 (GSM1274873, GSM1691142 and GSM2219863), GABPA (GSM1193660), GATA2 (GSM941195 and GSM1600544), HOXB13 (GSM1716764 and GSM2537218), NKX3.1 (GSM989640) and POLR2A (GSM353623, GSM969566, GSM1059393 and GSM1059394).

eQTL Analyses. To determine the possible target genes through which the risk signals identified may operate, we assessed the 269 risk variants against expression quantitative-trait loci (eQTL) data in three prostate tissue cohorts. Normal prostate tissue significant variant-gene pair data were downloaded for GTEx[48] v8 from the GTEx portal (n=221; https://gtexportal.org/home/datasets) and converted to GRCh37/hg19 reference assembly co-ordinates in R using rtracklayer v1.42.2 liftOver.[46] Normalized prostate expression levels, genotypes and relevant covariates were obtained for the Thibodeau et al.[47] tumor-adjacent normal prostate dataset from dbGaP (n=471; accession phs000985.v1.p1). Prostate adenocarcinoma data was obtained from TCGA (n=359; https://gdc-portal.nci.nih.gov), QC filtered and rank-normalized as described previously.[7] For the phs000985.v1.p1 and TCGA data, genotype array data was imputed using the 1000 Genomes Project[15] European panel from the Michigan Imputation Server.[27] A *cis*-eQTL scan was performed using FastQTL[49] separately for each study using a 1Mb window up- and down-stream of each gene's transcription start site and adaptive permutations between 1,000 and 10,000. Beta distribution-adjusted P-values were used to calculate Q-values, and a false discovery rate (FDR) threshold of ≤ 0.05 was applied

to identify significant variant-gene pairs. Identified eGenes are shown in **Supplementary Table 12**. For lead variants correlated with multiple eGenes within the same cohort or between cohorts, we report all significantly associated genes.

Data Availability

The full summary statistics resulting from this investigation are available through dbGaP under accession code phs001120.v2.p1. The genotype data and relevant covariate information (ancestry, country, principal components, etc.) used in this study are deposited in dbGaP under accession codes phs001391.v1.p1, phs000306.v4.p1, phs001120.v1.p1, phs001221.v1.p1, phs000812.v1.p1, and phs000838.v1.p1. Publicly available data described in this manuscript can be found from the following websites: 1000 Genomes Project (https://www.internationalgenome.org/); SEER (https://seer.cancer.gov/); National Center for Health Statistics, CDC (https://www.cdc.gov/nchs/index.htm); Cistrome Data Browser (http://cistrome.org/db/); GTEx (https://gtexportal.org/home/datasets); and TGCA (https://gdc-portal.nci.nih.gov).

Code Availability

Imputation was performed using IMPUTE2, MACH 1.0, Minimac3, and Minimac4. Association testing was performed using PLINK 1.07, SNPtest v2.5.2, and R v3.5. Meta-analyses were conducted using METAL v2011-03-25 and fine-mapping with JAM. Other analyses were performed with PriorityPruner v0.1.4, RFMix v1.0.2, and wANNOVAR (accessed 04/21/2020). Custom code modifying the JAM approach was developed for these analyses and is available on GitHub (https://github.com/USCmec/Conti_NatGen_2020). Code for analyses using other indicated software is readily available from the websites of the corresponding software.

Methods-only References

26. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**, 816-34 (2010).

27. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).

28. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).

29. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

30. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

31. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).

32. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).

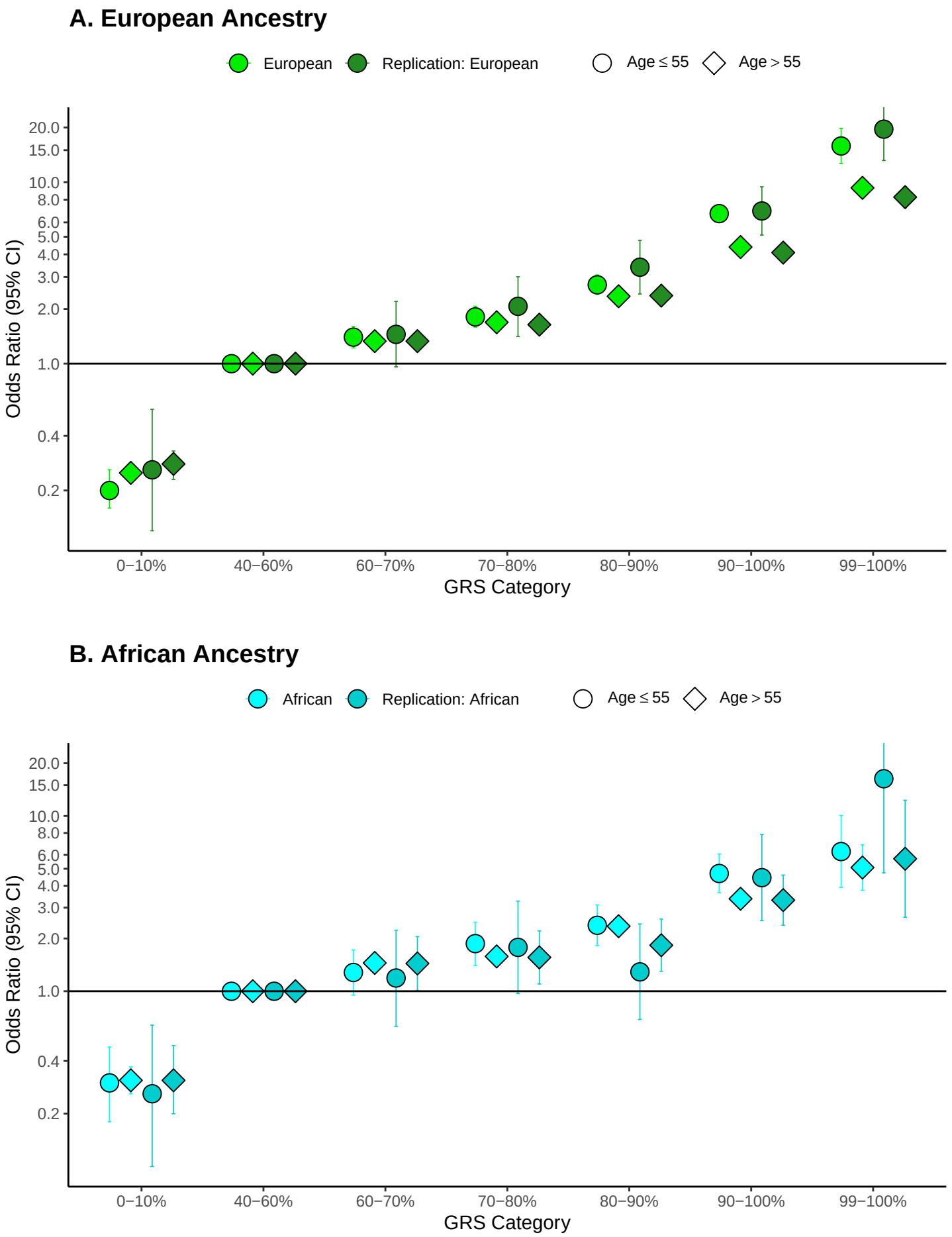
33. Newcombe, P.J., Conti, D.V. & Richardson, S. JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. *Genet Epidemiol* **40**, 188-201 (2016).

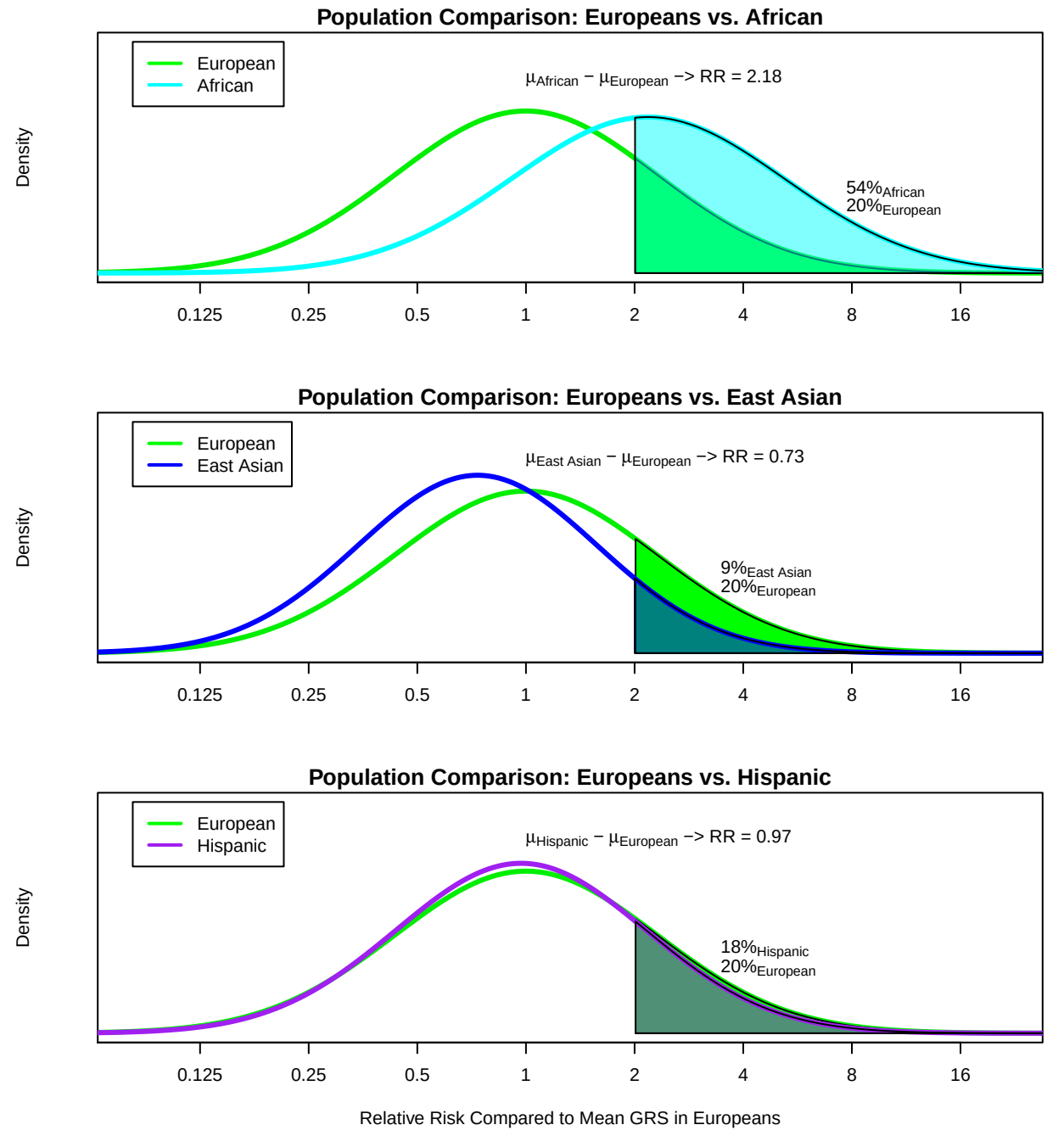
34. Ewing, C.M. *et al.* Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* **366**, 141-9 (2012).

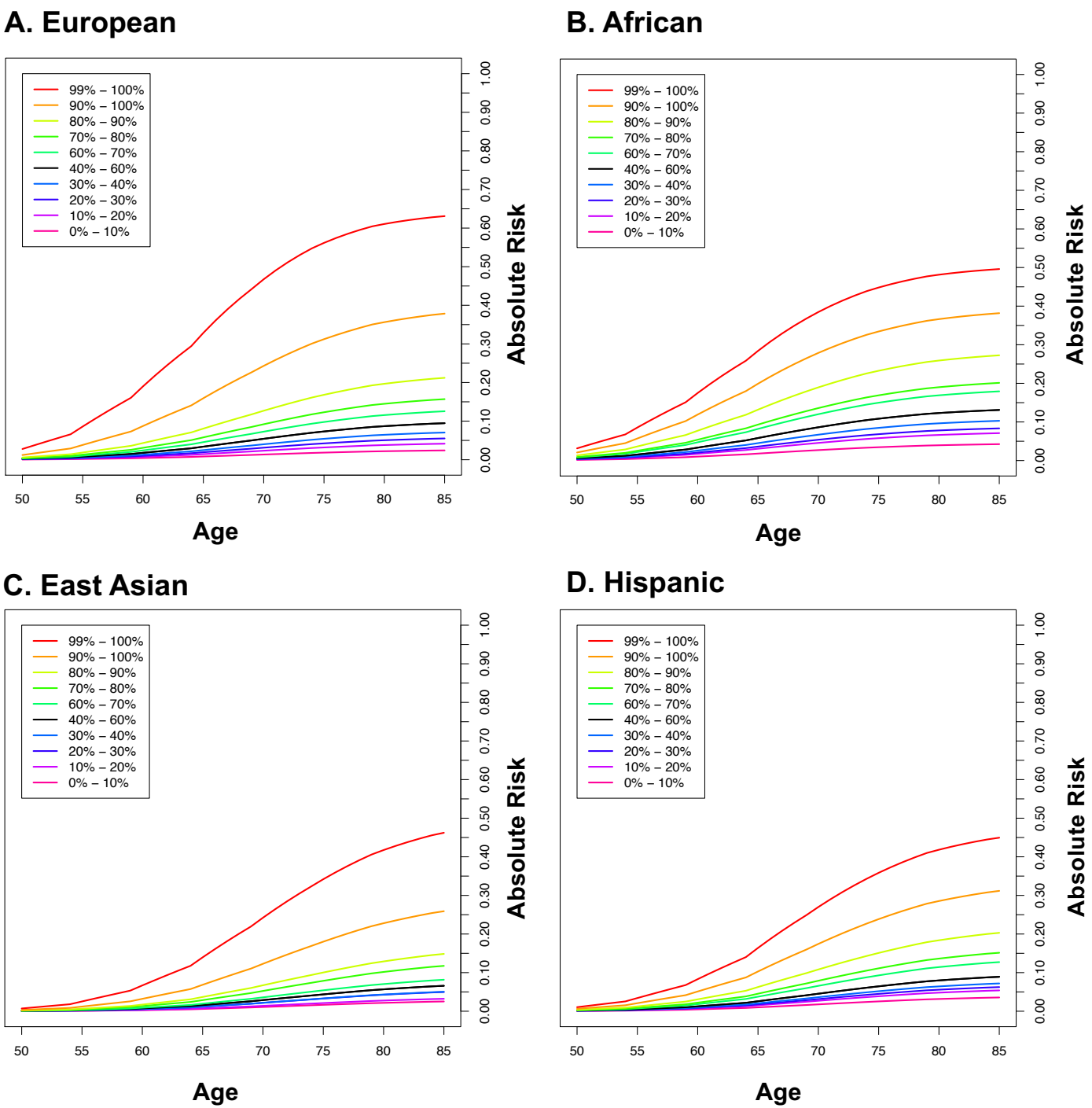
35. Seppala, E.H. *et al.* CHEK2 variants associate with hereditary prostate cancer. *Br J Cancer* **89**, 1966-70 (2003).

36. Pencina, M.J., D'Agostino, R.B., Sr. & Steyerberg, E.W. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* **30**, 11-21 (2011).
37. Maples, B.K., Gravel, S., Kenny, E.E. & Bustamante, C.D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* **93**, 278-88 (2013).
38. Amin Al Olama, A. *et al.* Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **24**, 1121-1129 (2015).
39. Antoniou, A.C. *et al.* Common Breast Cancer Susceptibility Alleles and the Risk of Breast Cancer for BRCA1 and BRCA2 Mutation Carriers: Implications for Risk Prediction. *Cancer research* **70**, 9742-9754 (2010).
40. Antoniou, A.C. *et al.* Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic epidemiology* **21**, 1-18 (2001).
41. Kuchenbaecker, K.B. *et al.* Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *Journal of the National Cancer Institute* **109**(2017).
42. Wang, K. *et al.* Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet* **86**, 730-42 (2010).
43. Witte, J.S., Visscher, P.M. & Wray, N.R. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet* **15**, 765-76 (2014).

44. Chang, X. & Wang, K. wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet* **49**, 433-6 (2012).
45. Mei, S. *et al.* Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res* **45**, D658-D662 (2017).
46. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841-2 (2009).
47. Thibodeau, S.N. *et al.* Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. *Nat Commun* **6**, 8653 (2015).
48. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
49. Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479-85 (2016).


A. European Ancestry
European
Replication: European
Age ≤ 55
Age > 55
Odds Ratio (95% CI)
20.0
15.0
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.0
0.4
0.2
0–10%
40–60%
60–70%
70–80%
80–90%
90–100%
99–100%
GRS Category
B. African Ancestry
African
Replication: African
Age ≤ 55
Age > 55
Odds Ratio (95% CI)
20.0
15.0
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.0
0.4
0.2
0–10%
40–60%
60–70%
70–80%
80–90%
90–100%
99–100%
GRS Category


Population Comparison: Europeans vs. African
European
African
μAfrican − μEuropean −> RR = 2.18
54%African
20%European
Density
0.125
0.25
0.5
1
2
4
8
16
Population Comparison: Europeans vs. East Asian
European
East Asian
μEast Asian − μEuropean −> RR = 0.73
9%East Asian
20%European
Density
0.125
0.25
0.5
1
2
4
8
16
Population Comparison: Europeans vs. Hispanic
European
Hispanic
μHispanic − μEuropean −> RR = 0.97
18%Hispanic
20%European
Density
0.125
0.25
0.5
1
2
4
8
16
Relative Risk Compared to Mean GRS in Europeans


A. European
B. African
C. East Asian
D. Hispanic
99% – 100%
90% – 100%
80% – 90%
70% – 80%
60% – 70%
40% – 60%
30% – 40%
20% – 30%
10% – 20%
0% – 10%
Absolute Risk
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Age
50
55
60
65
70
75
80
85