CANCER DIAGNOSIS & PROGNOSIS *5*: 146-152 (2025)
doi: 10.21873/cdp.10424

The Role of HMGA1 and Height in Breast Cancer Risk and Prognosis: Insights from UK Biobank Data

STEVEN LEHRER[1] and PETER H. RHEINSTEIN[2]

[1]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York City, NY, U.S.A.;
[2]Severn Health Solutions, Severna Park, MD, U.S.A.

Abstract

Background/Aim: Tall women are more likely to develop breast cancer (BC). High Mobility Group AT-Hook 1 (HMGA1), an oncofetal protein, plays a role in BC progression. Variants near HMGA1 have been associated with increased height. This study examines the relationship between HMGA1, height, and BC risk and prognosis using UK Biobank data.

Patients and Methods: Data from 10,527 women with invasive BC were analyzed. Subjects were grouped by height: short (<155 cm), medium (155-175 cm), and tall (>175 cm). HMGA1 SNP rs41269028, a single nucleotide intron variant, was evaluated for its influence on height, BC risk, and survival. Statistical analysis included Fisher's exact test, regression models, and survival analysis using the log-rank test.

Results: HMGA1 SNP rs41269028 carriers (CT+TT) were taller (162.88 cm) compared to homozygotes for the major allele (162.29 cm, p=0.005). Tall women with BC showed poorer survival than short women (p=0.032). However, HMGA1 genotype did not significantly affect BC risk (p=0.602) or survival (p=0.439). Multivariate analysis confirmed an independent effect of age and HMGA1 genotype on height.

Conclusion: While HMGA1 influences height, no direct association with increased BC risk or poor prognosis in tall women was demonstrated. Nevertheless, tall women with BC had worse survival, suggesting height might be considered in treatment decisions. Future studies should explore mechanisms linking height to BC outcomes.

Keywords: Breast cancer, height, genetics, risk, survival.

Introduction

Tall women are more likely to develop breast cancer (BC) (1). Women who are 176 cm or taller have a 20%-30% higher risk of breast cancer than women who are roughly 155 cm or lower, according to a pooled analysis that included data from 20 prospective cohort studies (2). The growth spurts tall women experienced as children have

Steven Lehrer, Box 1236, Radiation Oncology, Mount Sinai Medical Center, 1 Gustave L. Levy Place, New York City, NY 10029, U.S.A. Tel: +1 2127657132, e-mail: steven.lehrer@mssm.edu

Received December 19, 2024 | Revised January 8, 2025 | Accepted January 10, 2025

been postulated to elevate the risk of breast cancer associated with height. Increased hormone levels like IGF-1 or other growth factors can trigger growth spurts (3). Higher hormone levels and rapid cell proliferation during a growth spurt are thought to influence risk of breast cancer in later life.

High Mobility Group AT-Hook 1 (HMGA1), an oncofetal protein, plays a role in the progression of breast cancer (4). HMGA1 establishes an autocrine loop in invasive triple-negative breast cancer (TNBC) cells, which mediates the migration, invasion, and metastasis of TNBC cells and predicts the onset of metastasis in these patients. Hawkes *et al.* performed a whole genome sequencing association analysis for height using 333,100 individuals from three datasets: UK Biobank, TOPMed and All of Us. They identified non-coding sequences proximal to HMGA1 containing variants associated with a 4.83 cm taller height (5). In the current study, we used UK Biobank data to examine the relationship of HMGA1 to height, risk, and prognosis of women with breast cancer.

Patients and Methods

Patients. The UK Biobank is a large prospective observational study of men and women with no link to MedWatch. Participants were recruited from across 22 centers located throughout England, Wales, and Scotland between 2006 and 2010 and continue to be longitudinally followed for capture of subsequent health events (6). This methodology is like that of the ongoing Framingham Heart Study (7), with the exception that the UKB program collects postmortem samples, which Framingham did not.

UK Biobank has approval from the Northwest Multi-center Research Ethics Committee (MREC) to obtain and disseminate data and samples from the participants, and these ethical regulations cover the work in this study. Written informed consent was obtained from all participants. Details can be found at www.ukbiobank.ac.uk/ethics.

Our UK Biobank application was approved as UKB project 57245 (S.L., P.H.R.). Our analysis included all subjects with invasive BC that occurred either before or after participant enrollment and was recorded in the UK Biobank database using self-reported data and the International Classification of Diseases (ICD10, ICD9).

Methods. We divided the subjects into three previously described height groups (1): Short (<155 cm), Medium (155 cm to 175 cm), Tall (>175 cm). We analyzed the HMGA1 SNP rs41269028, a single nucleotide intron variant, C>T, minor allele frequency 0.044. SNP rs41269028 was previously evaluated in subjects with diabetes (8, 9).

Statistical analysis. Descriptive statistics. Mean and standard deviations of height were calculated for different genotype groups (homozygous major allele CC *vs.* minor allele carriers CT+TT). A two-tailed *t*-test was used to assess the significance of height differences between genotype groups.

Logistic regression analysis. To evaluate the association between BC risk and independent variables (age, menopause status, height group, and HMGA1 genotype), logistic regression models were applied. Odds ratios (ORs) with 95% confidence intervals were calculated for each variable to determine its contribution to BC risk.

Survival analysis. Survival outcomes for BC patients were assessed using Kaplan-Meier curves. Survival differences between height groups (short, medium, tall) and HMGA1 genotypes (CC *vs.* CT or TT) were evaluated with the log-rank test. Statistical significance for survival outcomes was set at $p<0.05$.

Multivariate linear regression. To correct for potential confounding effects of age on height (as aging is associated with height loss), multivariate linear regression was conducted. The dependent variable was height, while independent variables included HMGA1 genotype and age. Regression coefficients (ß) and significance values were reported to determine the independent effect of genotype and age on height.

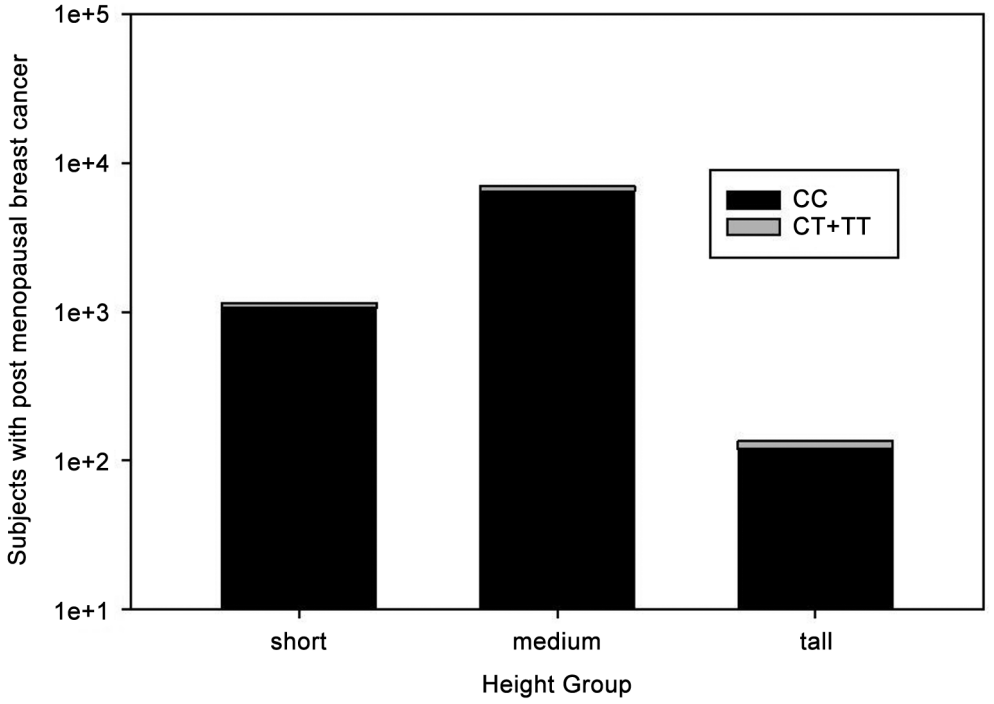

Figure 1. *8,327 post-menopausal breast cancer cases stratified by height group and HMGA1 SNP rs41269028 genotype (CC versus CT or TT). Note that tall women have the greatest proportion of carriers or homozygotes (CT+TT) for the minor allele T (p=0.02, two tail Fisher exact test).*

Software used. Statistical analysis was performed using SPSS (version 26, IBM, Armonk, NY, USA), and figures illustrating survival and genotype distributions were generated to visualize the results (10).

Results

Data from 273,378 women, of which 10,527 were invasive breast cancer cases, was analyzed. Breast cancer patients were aged 60±7 [mean±standard deviation (SD)]. 95% of subjects were white British. Table I shows the HMGA1 SNP rs41269028 genotype *versus* height group in 8,327 post-menopausal breast cancer cases. A greater proportion of tall women with breast cancer (11.9%) than short women (6.8%) were carriers or homozygotes (CT+TT) of the minor allele T (p=0.02, two tail Fisher exact test). No significant effect was present in pre-menopausal women (p=0.441).

The height of 9583 women with BC homozygous for the HMGA1 SNP rs41269028 major allele (CC) was 162.29 cm±6.18. The height of 944 women with BC who were carriers or homozygotes (CT+TT) of the minor allele T was 162.88 cm±6.001. This difference was significant (p=0.005). Figure 1 illustrates 8,327 post-menopausal breast cancer

Table I. *HMGA1 SNP rs41269028 genotype versus height in 8,327 post-menopausal breast cancer cases.*

		Genotype		
Height		CC	CT or TT	Total
Short	Count	1,068	78	1,146
	% within height group	93.2%	6.8%	100%
Medium	Count	6,414	632	7,046
	% within height group	91.0%	9.0%	100%
Tall	Count	119	16	135
	% within height group	88.1%	11.9%	100%
Total	Count	7,601	726	8,327
	% within height group	91.3%	8.7%	100%

A greater proportion of tall women with breast cancer (11.9%) than short women (6.8%) were carriers or homozygotes (CT+TT) of the minor allele T (p=0.02, two-tailed Fisher exact test).

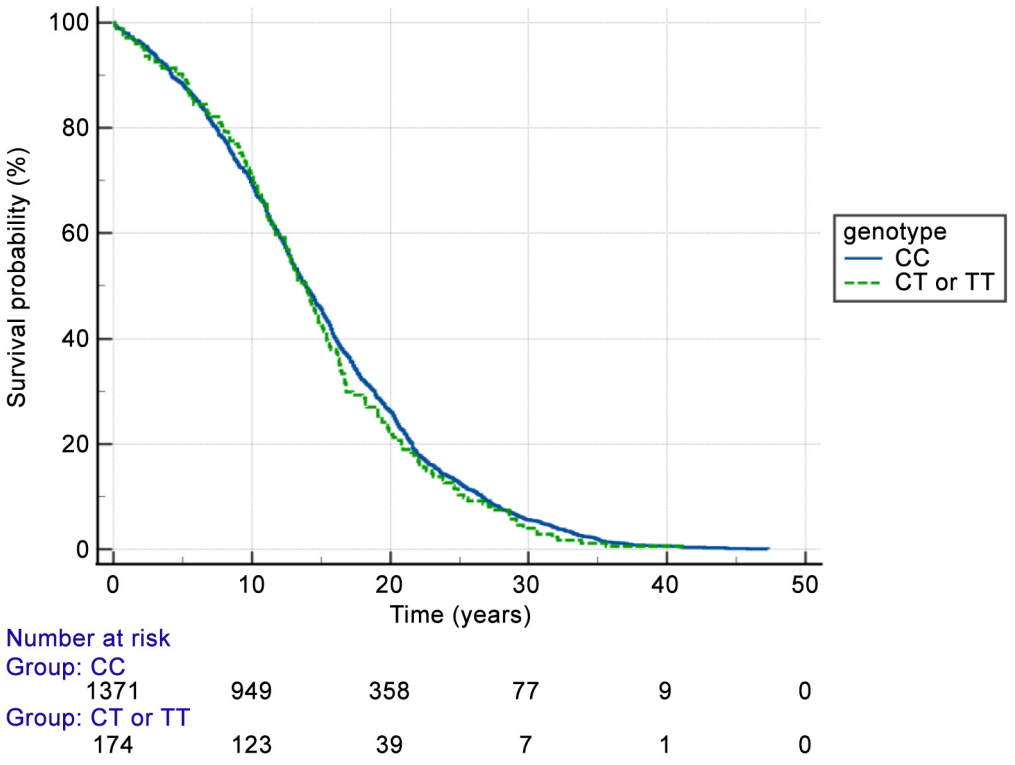

Figure 2. *Survival of breast cancer subjects stratified by HMGA1 SNP rs41269028 genotype. The effect of genotype was insignificant (p=0.439, log rank test).*

patients stratified by height group and the HMGA1 SNP rs41269028 genotype (CC *versus* CT or TT).

Aging is linked to a reduction in height over time (11). To account for this effect, multivariate linear regression was conducted using breast cancer cases. In this analysis, height group was the dependent variable, while HMGA1 SNP rs41269028 genotype and age were included as independent variables. The effect of genotype on height groups was significant (β=0.040, p=0.002) and independent of the effect of age (β=-0.005, p<0.001). In other words, carriers or homozygotes of the minor allele T were taller than homozygotes for the major allele (CC); while older women were shorter than younger women.

Figure 2 illustrates the survival of breast cancer subjects stratified by HMGA1 SNP rs41269028 genotype. The effect of genotype was insignificant (p=0.439, log rank test). Figure 3 illustrates the survival outcomes of breast cancer patients categorized by height group. Height group had a significant impact on survival (p=0.032, log-rank test), with tall women exhibiting the poorest survival rates. Table II presents the results of the logistic regression analysis. The dependent variable is breast cancer status (yes or no), with data from 228,611 women. The independent variables include age, menopausal status, and height group. The risk of breast cancer was increased in post-menopausal women (OR=4.315, p<0.001). The risk increased with each year of age (OR=1.034, p<0.001). Short women were at decreased risk compared to tall women (OR=0.819, p=0.026). Women of medium height were at decreased risk that was not significant compared to tall women (OR=0.929, p=0.391). HMGA1 SNP rs41269028 had no significant relationship to BC risk (p=0.602).

Discussion

Most of the genetic variation associated with complex traits, such as height, is located in non-coding regions of

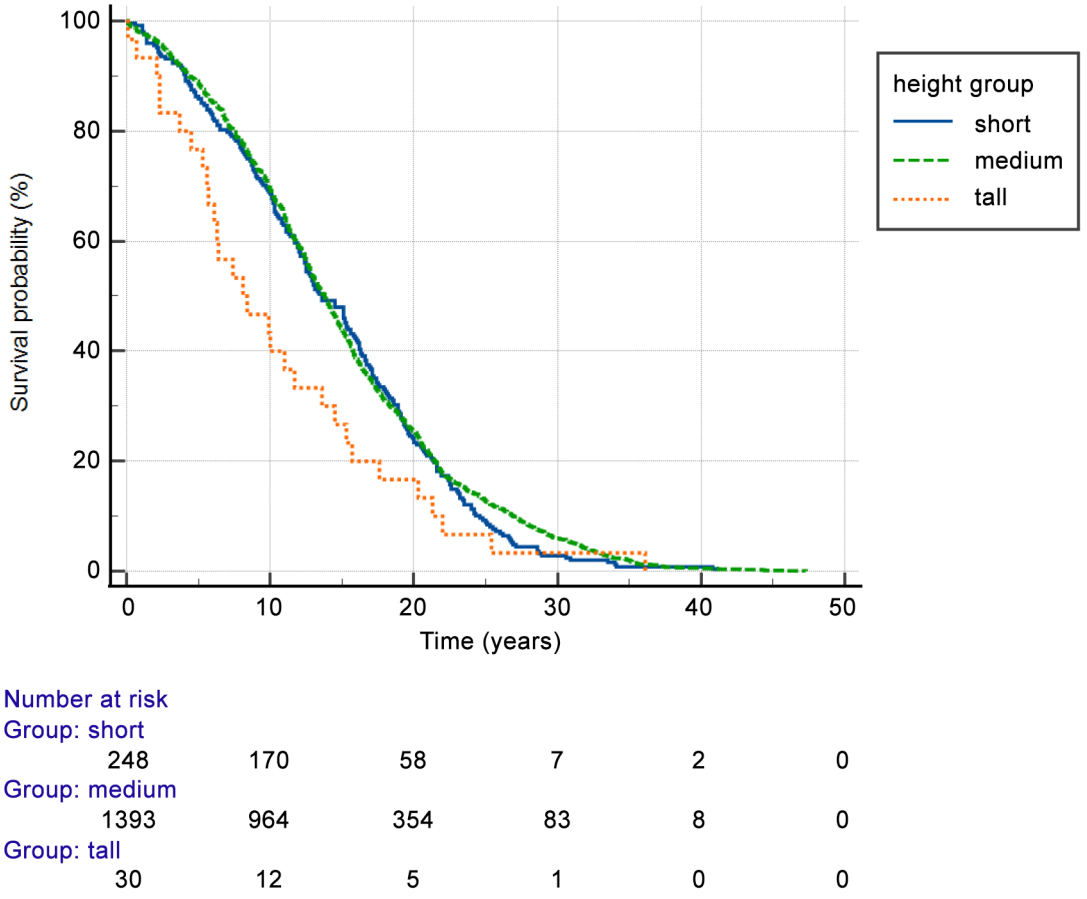

Figure 3. *Survival of breast cancer subjects stratified by height. The effect of height was significant (p=0.032, log rank test). Tall women had the poorest survival.*

the genome. Since 99% of the human genome is non-coding, most inherited genetic variations are rare and occur within these regions. Identifying these rare non-coding variations associated with common traits and diseases can reveal new regulatory gene pathways and significantly enhance our understanding of human biology and disease mechanisms (5).

HMGA1 SNP rs41269028 is an intron variant in a non-coding region of the genome, chromosome 6. Rare variants such as those of HMGA1 are said to confer most of the heredity for height, about 79% (12). In other words, in a large group of people, 79% of height differences are genetic (13). HMGA1 is a protein that has been found to play a role in the progression of breast

Table II. *Logistic regression results for breast cancer risk by menopause status, age, and height group.*

Variable	95% LB	OR	95% UB	*p*-Value
Menopause	3.866	4.315	4.815	<0.001
Age	1.030	1.034	1.038	<0.001
Short	0.689	0.819	0.973	0.026
Medium	0.788	0.929	1.095	0.391
HMGA1	0.908	0.980	1.058	0.602

The risk of breast cancer was increased in post-menopausal women (OR=4.315, $p<0.001$). The risk increased with each year of age (OR=1.034, $p<0.001$). Short women were at decreased risk compared to tall women (OR=0.819, $p=0.026$). Women of medium height were at decreased risk that was not significant compared to tall women (OR=0.929, $p=0.391$). HMGA1 SNP rs41269028 had no significant relationship to BC risk ($p=0.602$). LB, Lower bound; UB, upper bound; OR, odds ratio.

cancer. One study suggests that HMGA1 establishes an autocrine loop in invasive triple-negative breast cancer (TNBC) cells, which mediates the migration, invasion, and metastasis of TNBC cells and predicts the onset of metastasis in these patients (14).

HMGA1 has been reported to promote breast cancer angiogenesis by supporting the stability, nuclear localization, and transcriptional activity of FOXM1 (15). FOXM1 is an oncogenic transcription factor that is greatly upregulated in breast cancer and many other cancers where it promotes tumorigenesis, cancer growth and progression (16). It is expressed in all subtypes of breast cancer and is the factor most associated with risk of poor patient survival, especially in TNBC.

Study limitations. We did not have any data on tumor size, histology, grade, or hormone receptor status. We did not have the recently released UKBB whole genome sequence data for HMGA1 that Hawkes *et al.* used (5, 17). Instead, we evaluated imputed genotypes from the UKB data field 22828 (18). We found that the HMGA1 SNP rs41269028 minor allele T carriers (CT) and homozygotes (TT) were significantly taller, but the effect size was small (0.59 cm) compared to the Hawkes *et al.* report (4.83 cm) (5).

Conclusion

We conclude that HMGA1 influences height, but we were unable to demonstrate that HMGA1 is related to increased incidence or poor prognosis of tall women with breast cancer. Our finding that tall women have a worse prognosis is important because it could help the oncologist decide, along with other prognostic factors, whether adjuvant therapy is warranted.

Data Availability

Data sources described in the article are publicly available or can be accessed after an approved application to the UK Biobank.

Conflicts of Interest

None.

Authors' Contributions

SL and PHR contributed equally to the conception, writing, and data analysis of this study.

Acknowledgements

This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences.

References

1 Gremke N, Griewing S, Kalder M, Kostev K: Positive association between body height and breast cancer prevalence: a retrospective study with 135,741 women in Germany. Breast Cancer Res Treat 196(2): 349-354, 2022. DOI: 10.1007/s10549-022-06730-0

2 van den Brandt PA, Ziegler RG, Wang M, Hou T, Li R, Adami HO, Agnoli C, Bernstein L, Buring JE, Chen Y, Connor AE, Eliassen AH, Genkinger JM, Gierach G, Giles GG, Goodman GG, Håkansson N, Krogh V, Le Marchand L, Lee IM, Liao LM, Martinez ME, Miller AB, Milne RL, Neuhouser ML, Patel AV, Prizment A, Robien K, Rohan TE, Sawada N, Schouten LJ, Sinha R, Stolzenberg-Solomon RZ, Teras LR, Tsugane S, Visvanathan K, Weiderpass E, White KK, Willett WC, Wolk A, Zeleniuch-Jacquotte A, Smith-Warner SA: Body size and weight change over adulthood and risk of breast cancer by menopausal and hormone receptor status: a pooled analysis of 20 prospective cohort studies. Eur J Epidemiol 36(1): 37-55, 2021. DOI: 10.1007/s10654-020-00688-3

3 Cole TJ, Ahmed ML, Preece MA, Hindmarsh P, Dunger DB: The relationship between Insulin-like Growth Factor 1, sex steroids and timing of the pubertal growth spurt. Clin Endocrinol (Oxf) 82(6): 862-869, 2015. DOI: 10.1111/cen.12682

4 Unachukwu U, Chada K, D'Armiento J: High mobility group AT-Hook 2 (HMGA2) oncogenicity in mesenchymal and epithelial neoplasia. Int J Mol Sci 21(9): 3151, 2020. DOI: 10.3390/ijms21093151

5 Hawkes G, Beaumont RN, Li Z, Mandla R, Li X, Albert CM, Arnett DK, Ashley-Koch AE, Ashrani AA, Barnes KC,

Boerwinkle E, Brody JA, Carson AP, Chami N, Chen YI, Chung MK, Curran JE, Darbar D, Ellinor PT, Fornage M, Gordeuk VR, Guo X, He J, Hwu C, Kalyani RR, Kaplan R, Kardia SL, Kooperberg C, Loos RJ, Lubitz SA, Minster RL, Mitchell BD, Murabito JM, Palmer ND, Psaty BM, Redline S, Benjamin Shoemaker M, Silverman EK, Telen MJ, Weiss ST, Yanek LR, Zhou H, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Liu C, North KE, Justice AE, Locke J, Owens N, Murray A, Patel K, Frayling TM, Wright CF, Wood AR, Lin X, Manning A, Weedon MN: Whole genome association testing in 333,100 individuals across three biobanks identifies rare non-coding single variant and genomic aggregate associations with height. biorxiv, 2023. DOI: 10.1101/2023.11.19.566520

6 Arthur RS, Wang T, Xue X, Kamensky V, Rohan TE: Genetic factors, adherence to healthy lifestyle behavior, and risk of invasive breast cancer among women in the UK biobank. J Natl Cancer Inst 112(9): 893-901, 2020. DOI: 10.1093/jnci/djz241

7 Mahmood SS, Levy D, Vasan RS, Wang TJ: The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 383(9921): 999-1008, 2014. DOI: 10.1016/S0140-6736(13)61752-3

8 Marquez M, Huyvaert M, Perry JR, Pearson RD, Falchi M, Morris AP, Vivequin S, Lobbens S, Yengo L, Gaget S, Pattou F, Poulain-Godefroy O, Charpentier G, Carlsson LM, Jacobson P, Sjöström L, Lantieri O, Heude B, Walley A, Balkau B, Marre M, Froguel P, Cauchi S, DIAGRAM Consortium: Low-frequency variants in HMGA1 are not associated with type 2 diabetes risk. Diabetes 61(2): 524-530, 2012. DOI: 10.2337/db11-0728

9 Chiefari E, Tanyolaç S, Paonessa F, Pullinger CR, Capula C, Iiritano S, Mazza T, Forlin M, Fusco A, Durlach V, Durlach A, Malloy MJ, Kane JP, Heiner SW, Filocamo M, Foti DP, Goldfine ID, Brunetti A: Functional variants of the HMGA1 gene and type 2 diabetes mellitus. JAMA 305(9): 903-12, 2011. DOI: 10.1001/jama.2011.207

10 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4: 7, 2015. DOI: 10.1186/s13742-015-0047-8

11 Sagiv M, Vogelaere PP, Soudry M, Ehrsam R: Role of physical activity training in attenuation of height loss through aging. Gerontology 46(5): 266-270, 2000. DOI: 10.1159/000022170

12 Wainschtein P, Jain D, Zheng Z, TOPMed Anthropometry Working Group, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Cupples LA, Shadyab AH, McKnight B, Shoemaker BM, Mitchell BD, Psaty BM, Kooperberg C, Liu CT, Albert CM, Roden D, Chasman DI, Darbar D, Lloyd-Jones DM, Arnett DK, Regan EA, Boerwinkle E, Rotter JI, O'Connell JR, Yanek LR, de Andrade M, Allison MA, McDonald MN, Chung MK, Fornage M, Chami N, Smith NL, Ellinor PT, Vasan RS, Mathias RA, Loos RJF, Rich SS, Lubitz SA, Heckbert SR, Redline S, Guo X, Chen Y-I, Laurie CA, Hernandez RD, McGarvey ST, Goddard ME, Laurie CC, North KE, Lange LA, Weir BS, Yengo L, Yang J, Visscher PM: Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat Genet 54(3): 263-273, 2022. DOI: 10.1038/s41588-021-00997-7

13 Geddes L: Genetic study homes in on height's heritability mystery. Nature 568(7753): 444-445, 2019. DOI: 10.1038/d41586-019-01157-y

14 Méndez O, Pérez J, Soberino J, Racca F, Cortés J, Villanueva J: Clinical implications of extracellular HMGA1 in breast cancer. Int J Mol Sci 20(23): 5950, 2019. DOI: 10.3390/ijms20235950

15 Zanin R, Pegoraro S, Ros G, Ciani Y, Piazza S, Bossi F, Bulla R, Zennaro C, Tonon F, Lazarevic D, Stupka E, Sgarra R, Manfioletti G: HMGA1 promotes breast cancer angiogenesis supporting the stability, nuclear localization and transcriptional activity of FOXM1. J Exp Clin Cancer Res 38(1): 313, 2019. DOI: 10.1186/s13046-019-1307-8

16 Katzenellenbogen BS, Guillen VS, Katzenellenbogen JA: Targeting the oncogenic transcription factor FOXM1 to improve outcomes in all subtypes of breast cancer. Breast Cancer Res 25(1): 76, 2023. DOI: 10.1186/s13058-023-01675-8

17 Callaway E: World's biggest set of human genome sequences opens to scientists. Nature 624(7990): 16-17, 2023. DOI: 10.1038/d41586-023-03763-3

18 Lehrer S, Rheinstein PH: Association of Kallikrein Related Peptidase 3 (KLK3) gene with dermatophytosis in the UK biobank cohort. Mycoses 66(12): 1050-1055, 2023. DOI: 10.1111/myc.13649