

## RESEARCH ARTICLE

# What is creating the height premium? New evidence from a Mendelian randomization analysis in China

Jun Wang<sup>1,2</sup>, Qihui Chen<sup>3</sup>, Gang Chen<sup>4</sup>, Yingxiang Li<sup>4</sup>, Guoshu Kong<sup>1,2</sup>, Chen Zhu<sup>3\*</sup>

**1** School of Public Administration and Policy, Renmin University of China, Beijing, China, **2** Center for Health Policy Research and Evaluation, Renmin University of China, Beijing, China, **3** College of Economics and Management, China Agricultural University, Beijing, China, **4** WeGene, Shenzhen, China

\* [zhuchen@cau.edu.cn](mailto:zhuchen@cau.edu.cn)



## OPEN ACCESS

**Citation:** Wang J, Chen Q, Chen G, Li Y, Kong G, Zhu C (2020) What is creating the height premium? New evidence from a Mendelian randomization analysis in China. *PLoS ONE* 15(4): e0230555. <https://doi.org/10.1371/journal.pone.0230555>

**Editor:** Petri Böckerman, University of Jyväskylä, FINLAND

**Received:** November 18, 2019

**Accepted:** March 2, 2020

**Published:** April 10, 2020

**Copyright:** © 2020 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data cannot be shared publicly because of the Regulation on the Administration of Human Genetic Resources (effective since July 1, 2019) promulgated by China's State Council. Data are available from the WeGene Research Team (contact via Dr. Lizhong Wang <[wanglz@wegene.com](mailto:wanglz@wegene.com)>) for researchers who meet the criteria for access to confidential data.

**Funding:** The authors received no specific funding for this work. WeGene provided data support of

## Abstract

This study uses a Mendelian randomization approach to resolve the difficulties of identifying the causal relationship between height and earnings by using a unique sample of 3,427 respondents from mainland China with sociodemographic information linked to individual genotyping data. Exploiting genetic variations to create instrumental variables for observed height, we find that while OLS regressions yield that an additional centimeter in height is associated with a 10–13% increase in one's annual earnings, IV estimates reveal only an insubstantial causal effect of height. Further analyses suggest that the observed height premium is likely to pick up the impacts of several cognitive/noncognitive skills on earnings confounded in previous studies, such as mental health, risk preference, and personality factors. Our study is the first empirical study that employs genetic IVs in developing countries, and our results contribute to the recent debate on the mechanism of height premium.

## 1. Introduction

Height is widely believed to be a key determinant of professional success. Indeed, taller people are usually found to earn more in both developed countries [1–5] and developing countries [6–10]. Yet why is a height-wage premium observed? More intriguingly, why is it observed even in labor markets where height is unlikely to be a crucial productive factor, say, sedentary white-collar jobs in industrialized countries? This question has led many researchers to interpret the observed height premium as evidence of labor-market discrimination by employers who stigmatize shorter people [11–12]. However, this interpretation has recently been challenged on the ground that the observed link between height and better labor-market performance may reflect the influence of unobserved factors (that affect one's labor-market performance). For example, height has been found to be correlated with both cognitive skills [1–2,13–14] and non-cognitive skills [14–15], skills that contribute greatly to one's labor-market success [16].

These differing interpretations point into vastly different directions of policy-making. If labor-market discrimination is at work, then policies aiming to prevent discrimination may be

collecting and analyzing genotyping data, but did not have any additional role in financially supporting the current study.

**Competing interests:** Gang Chen and Yingxiang Li are employees of WeGene. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

needed to protect employee welfare. In contrast, if the observed height premium is, in fact, capturing the impact of some hard-to-observe factors, such as emotional stability and perseverance, then using height as a marker to gauge returns to these factors is justifiable (and may even be desirable). Therefore, to better inform policy in response to inequality, it is crucial to investigate the origins and driving channels of the observed height premium. Such an investigation is of particular interest in transitory economies, where the role of height as a productive factor is rapidly changing, as a result of structural transformations undergoing in the labor market (e.g., changes in the relative shares of blue-collar and white-collar jobs) and in the economy as a whole (e.g., changes in the relative shares of different sectors).

China provides an interesting case to study. Partly due to China's market-oriented reform, employment opportunities and wage determination rely more on the market rather than institutional and political factors [17–18], which leads to complex implications for the role height plays in wage determination. On the one hand, as the Chinese economy advances and its industrial structure upgrades, demand for unskilled labor has been falling since the early 1980s [19], which downplays the earning potential of physical strength. On the other hand, the rapid development of the service sector in the recent decade, exemplified by the booming express delivery services for on-line purchased daily and food products, creates new room for physical strength and endurance to play a role. Echoing such complexity, empirical studies conducted in China usually find that while the observed height premium does pick up some impacts of factors such as political capital, it does not entirely vanish after these factors have been controlled for in estimation [20–21]. Thus, it is natural to ask: to what extent is height itself rewarded in China's labor market? And to what extent does the height premium capture the impact of other factors?

Needless to say, estimating the causal effect of height on wages using observational data is challenging, as it is difficult to control for all potential confounders in estimation. A standard solution is to find instrumental variables (IVs) for height, i.e., exogenous variables that affect the wage only through their impacts on height, to achieve identification. Some researchers exploited twinning experiments to identify the impact of height [22]. Since identical twins share the same genes and similar family environment, it is possible to eliminate shared environmental factors, such as the family background, neighborhood and peer effects, and genetic factors, using twin data. However, the external validity of twin studies is usually quite limited, since twins account for only 2% in any human population. Previous studies conducted in developing countries routinely used shocks to childhood nutrition, such as occurrences of droughts or disease outbreaks, relatives' height or community-level average height, to identify the impact of height. Yet the effect so identified mainly captures the impact of (the part of height that is determined by) childhood environmental factors, which may be correlated with other unobserved productive factors whose impacts we would like to purge out in estimation.

Taking another path, this paper adopts an IV strategy based on genetic information to achieve identification. Similar to those recently adopted by von Hinke et al. [23], Tyrrell et al. [24], and Böckerman et al. [25], we exploit variations in individuals' genetic markers to create an IV for height within the framework of Mendelian randomization (MR) [26]. To the extent that genetic markers strongly predict one's adult height but do not directly affect one's labor-market performance, they provide a source of exogenous variations in height needed for identification. More specifically, we instrument height using the polygenic score with gene variations that have been found to be significantly associated with height in genome-wide association studies (GWAS) using extensive population samples [27]. Polygenic scores (PGSs; also called "polygenic risk scores," "genetic risk scores," or "genome-wide scores") are aggregated effects of hundreds and thousands of trait-associated DNA variants identified in GWAS studies and can be used to predict propensities toward certain traits and outcomes [28–29].

While GWAS were mostly based on samples of European ancestry, recent studies indicate that the results can well apply to East Asian (e.g., Chinese) populations. For example, Duncan et al. [30] demonstrate that the polygenic score performance is reasonably reliable in East Asian samples (95%) relative to European samples (100%). However, the predictive performance of PGSs is much lower in African (42%) and South Asian (60%) samples.

By analyzing a unique dataset recently collected in China, we find that the statistically and economically significant height premium estimated by OLS regressions, i.e., an additional centimeter in height is associated with a 10–13% increase in one's annual earnings, largely reflects impacts of other factors. While our OLS estimates are similar to their counterparts in most previous studies conducted in China [20–21], our IV estimates reveal only a negligible causal impact of height. Further analysis reveals that the observed height premium is likely picking up the impacts of several cognitive/non-cognitive skills on earnings (e.g., mental health, risk preference).

Our analysis makes three contributions to the literature. First, our study is the first height-premium study that employs genetic IVs in China, if not the first in developing countries. Not only can our findings provide a deeper understanding of the working of China's labor market, but they are also complementary to findings in other, more developed countries [23–25]. Second, our findings are complementary to the recent discovery that the observed height premium largely captures the impact of other factors, rather than a reflection of labor-market discrimination believed by many previous researchers. Third, we extensively test the relevance, independence, and exclusion restrictions of the genetic instrument of height. The demonstration of the instrument validity not only strengthens our own findings, but also lends support to previous findings that are based on such genetic IV.

The remainder of this paper is structured as follows. The next section briefly reviews the relevant literature, paying particular attention to the channels through which height may affect one's wage income. Section 3 describes the data. Section 4 details our empirical methods. Section 5 presents our main empirical results. The final section draws conclusions and points out a number of directions for future research.

## 2. Relevant literature

### 2.1. Why may height affect wages?

To better understand what genetic IVs can identify, it is helpful to consider first the potential channels through which height may affect wages. Thus far, four channels have been put forward in the literature. First of all, height itself may be a productive factor, especially in developing countries where many jobs value physical strength and endurance of physical labor [31–33]. Yet these physical features are less critical for sedentary, white-collar jobs in more developed settings.

Secondly, height may be the *generator* of some other productive factors. For example, taller people may enjoy more social dominance [34–35] and have higher self-esteem [36], all of which may improve their competitiveness and labor-market performance [37]. The height premium may also be generated through *perception* and *expectation*. For example, taller people may be *perceived* as more productive by employers [38]. The perceived productivity of taller people may, in turn, influence the expectation of their employers, inducing them to assign more challenging and high-return tasks to taller employees [39]. Thus taller employees may have more opportunities to outperform their shorter counterparts in the labor market through a self-fulfilling prophecy effect. To the extent that these channels also reflect the productive effects of height, their existence should not be considered as evidence of labor-market discrimination.

Thirdly, height may be correlated with productive factors that are not generated by itself. The most well-known examples include cognitive skills [1] and non-cognitive skills [15]. While taller children tend to develop more cognitive [21,40] and non-cognitive skills [14], these skills need not be products of height *per se* to act as confounders in a conventional height-wage regression. As illustrated in Case and Paxson [1], for example, both height and these skills could be outcomes of some underlying *endowment*, say, insulin-like growth factors that stimulate simultaneous neural and physical growth [41–42], which is sufficient to generate a significant correlation between height and cognitive/non-cognitive skills in adulthood. Childhood environmental factors, such as food abundance in childhood, parental education, and household income, can also lead to correlations between height and cognitive developments [14]. Height may even be correlated with other types of human capital, such as political capital. Yamamura et al. [21], for example, show that taller youth in China are more likely to join the Chinese Communist Party (CCP) and CCP membership is rewarded in the labor market.

Finally, the observed height premium may simply reflect pure discrimination against shorter employees in the labor market [11–12,43]. Under pure discrimination, taller employees are paid more for reasons that are not related to their productivity, whether actual or perceived.

## 2.2. Recent methodological development

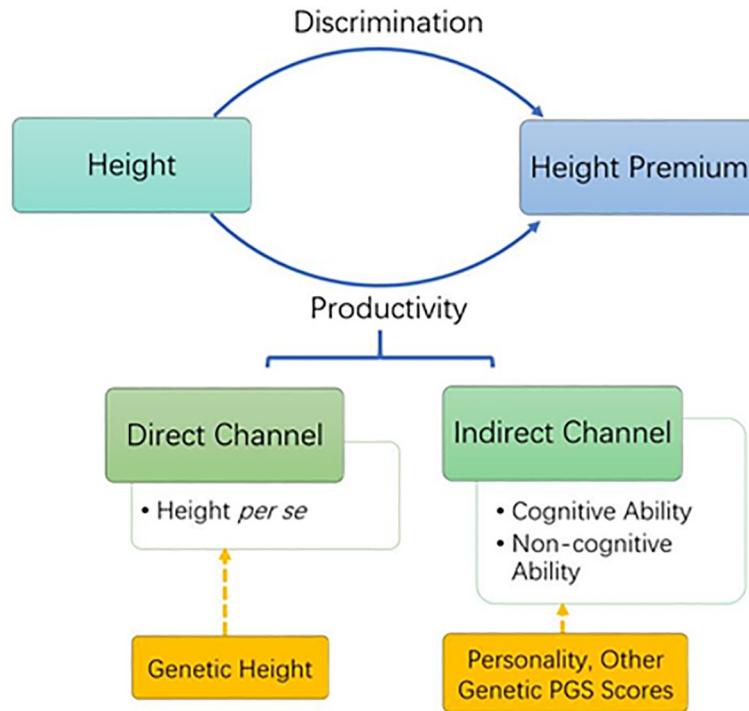
While earlier studies tend to examine what the height premium captures in an indirect manner, estimating OLS models that control for variables capturing impacts through the aforementioned channels, more-recent studies tend to estimate the impact of genetic height on wages more directly. The recent decade witnessed a new strand of studies that employ variations in one's genetic markups induced by Mendelian randomization to identify the impact of height [23–25], which is the impact of height working through the first two channels discussed above (Fig 1). More specifically, these studies instrument height using the polygenic score with gene variations that have been found to be significantly associated with height in genome-wide association studies based on extensive population samples [27]. An empirical regularity emerging from these studies is that the impacts of height identified by genetic variations are routinely much smaller than their OLS counterparts.

The present paper also exploits genetic variations to identify the causal impact of height in the context of China for the first time. We control for several personality characteristics as well as additional genetic factors, which allows us to see more clearly the role of height in wage determination.

## 3. Data

### 3.1. Database and survey

Our empirical analysis is based on a unique dataset collected in collaboration with WeGene, a leading private genetic testing company founded in 2014 in Shenzhen, China. The study has been approved by the Institutional Review Board of College of Economics and Management at China Agricultural University (CAU-CEM-IRB; approval number: HC0028401A). All participants of this study were drawn from the customer database of WeGene, all of whom provided (signed) informed consent and participated in the online geno-economic survey sometime between April 2018 and January 2019. Approximately 3,600 participants completed the full survey. The online survey collected information on each participant's demographic characteristics such as height, year and month of birth, their socioeconomic characteristics such as



**Fig 1. Channels identified using genetic instruments.**

<https://doi.org/10.1371/journal.pone.0230555.g001>

**Table 1. Summary statistics of the analytical sample (N = 3,427).**

Variable	Mean/Percentage	Std.Dev.
<i>Socio-demographic Characteristics:</i>		
Annual Earnings (in CNY)	138,927	179,554
Natural Logarithm of Annual Earnings	11.570	0.893
Age	30.245	6.667
Years of Schooling	15.604	2.237
Self-reported Height (in centimeter)	169.088	8.331
Male	56.1%	-
<i>Genetic Instrumental Variable:</i>		
Polygenic Score of Height [47]	-0.598	0.446
<i>Economic Preferences:</i>		
Risk Loving	5.613	2.136
Altruism	4.862	2.490
Trust	5.084	2.719
<i>Other Polygenic Score Controls:</i>		
Polygenic Score of Cognitive Ability [48]	-0.010	0.171
Polygenic Score of Depression [49]	32.628	4.233
Polygenic Score of Delay Discounting [50]	0.074	0.087
Polygenic Score of Reproduction Preference [51]	0.029	0.166

Source: Data are drawn from the consumer information base of WeGene. Summary statistics are calculated by the author.

<https://doi.org/10.1371/journal.pone.0230555.t001>

annual income and educational attainment, as well as their preferences such as risk attitude, altruism and trust [44–45].

Important to our study, all sample subjects were genotyped on a WeGene custom genotyping array platform (AffyMetrix). Imputation and quality control were performed by WeGene using PLINK (1.90 Beta), SHAPEIT (v2.17) and IMPUTE2 (v2.3.1). A total of 10,670,107 Single Nucleotide Polymorphisms (SNPs) were identified for each participant, which were then used to construct polygenic scores and genetic instrumental variables. Respondents' economic preferences (risk-loving, altruism and trust) were assessed using the Global Preferences Survey on a 10-point scale [45].

### 3.2. Sample characteristics

Excluding individuals who were younger than the age of 16, older than 60, or still a student at the time of the survey from the original dataset yields an analytical sample with 3,427 observations (1,922 males and 1,505 females). Table 1 provides summary statistics of the major characteristics of respondents in the analytical sample. Consistent with previous findings that Direct-to-Consumer genetic testing customers are generally well-educated middle-class professionals [46], an average respondent in the sample earned CNY138,927 (1 US Dollar  $\approx$  CNY 6.8) in 2017 and completed 15.6 years of education, both being significantly higher than the state-wide averages. The average annual wage of an urban employee is CNY 74,318 in 2017 (Source: <http://www.chinadaily.com.cn/a/201805/21/WS5b02d4d6a3103f6866ee9b15.html>). The average educational attainment of an employee is 10.2 years in 2015 (Source: [http://www.gov.cn/xinwen/2017-07/25/content\\_5213292.htm](http://www.gov.cn/xinwen/2017-07/25/content_5213292.htm)). The spatial distribution of respondents' home provinces, shown in Fig 2, also mirrors the above pattern: the top three provinces are Jiangsu, Shandong and Zhejiang, which are all located the most developed eastern coastal areas, whereas provinces in the less-developed western China (i.e., Ningxia and Qinghai) have the least respondents.

## 4. Empirical methods

Our empirical analysis mainly consists of two related parts. The first part measures the potential height premium in China's urban labor market. The second part then seeks to explain what the observed height premium really captures by (a) sequentially adding a set of covariates in the model, and (b) instrumenting height using the genetic IV.

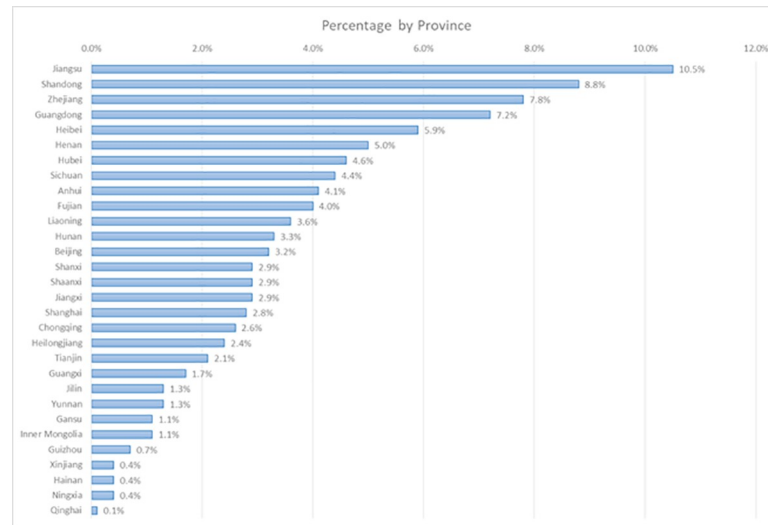
As a starting point, consider a statistical model that links one's wage income and adult height:

$$\log(Y) = \beta_0 + \beta_1 H + \mathbf{X}\beta_2 + u, \quad (1)$$

where  $Y$  is one's annual earnings (in log),  $H$  is his or her adult height measured in centimeters (cm),  $\mathbf{X}$  is a set of personal characteristics to be detailed below, and  $u$  is an error term capturing the influence of all unobserved factors. If Eq (1) is correctly specified, the coefficient on height,  $\beta_1$ , captures the causal effect of being one cm taller on one's earnings, which can be estimated by ordinary least-squares (OLS) techniques. However, as discussed above, in many situations, OLS estimates of  $\beta_1$  may capture something other than the causal effect of height.

### 4.1. Measuring the height premium

Assuming away identification problems with OLS for the moment, one subtle issue in estimating the height premium is: What covariates should be included in Eq (1)? A number of studies conducted in China adopted an extended version of Mincer's earnings equation to capture the height premium [52], which includes years of schooling, post-school experience, experience



**Fig 2. Spatial distributions of observations.**

<https://doi.org/10.1371/journal.pone.0230555.g002>

squared, marital status, self-reported health, household registration (*Hukou*) status, political affiliation, as well as dummies for occupation, industry, and firm ownership [20–21], as covariates. However, as argued by Neal and Johnson [53], Heckman [54], and Persico et al. [15], many of these covariates, such as years of schooling and occupation, are *choice* variables that can be influenced by height or are jointly determined with height, and thus the inclusion of these variables may yield misleading results. Therefore, we adopt a specification similar to that used by Persico et al. [15] as the benchmark specification, which includes only gender and age as covariates. But for comparison and exploratory purposes, we expand the set of covariates sequentially to include proxies for skills such as years of schooling, measures of personal preferences such as risk preference, altruism and trust, as well as other genetic controls—i.e., polygenic scores of cognitive ability [48], depression [49], delay discounting [50], and reproduction preference [51]. Expanding the set of covariates sequentially provides an opportunity for us to see what is being captured by the height premium estimated in the benchmark model.

#### 4.2. Genetic instrument: The polygenic score of height

Another issue is raised by the existence of unobserved confounders that are correlated with both one’s height and labor market performance. A standard solution is to find IVs for height, i.e., variables that are highly correlated with height (conditional on the covariates  $X$ ) but are orthogonal to potential confounders, to eliminate the influence of confounders in estimating Equ (1). We construct an IV exploiting one’s genetic information.

The genetic IV is constructed based on the notion of “Mendelian randomization”, which refers to the random assignment of an individual’s genotype at conception [55–56]. Since adult height is known as one of the highest heritable traits in human beings, determined approximately 80% by genetic factors and 20% by environmental factors [57–58], genetic variations related to height can naturally serve as unconfounded proxies for observed adult height [23–26].

It is worth noting that there is no *single* gene for height in humans. Instead, there are hundreds of genes and DNA segments that are determining the adult height in human genome. In other words, the genetic basis of human height is quantitative, rather than qualitative. More

specifically, this IV is constructed as the weighted polygenic score of height (*PGS\_Height*) based on 697 SNPs previously identified to be associated with adult height in human beings [47]. Mathematically, *PGS\_Height<sub>i</sub>* is calculated as the sum of genotyped or imputed allele *j* dosages carried by a respondent *i* (*SNP<sub>j,i</sub>*) multiplied by the estimated effect size ( $\beta_j$ ) reported by Wood et al. [47] (We use the Affymetrix WeGene V1 Arrays to assay common SNP variation in the genome of all participants, and then imputed additional SNPs):

$$PGScore\_Height_i = \sum_{j=1}^{J=697} \beta_j SNP_{j,i} \quad (2)$$

Note that because each person can have 0, 1, or 2 risky (effect) allele(s) for each of the 697 SNPs, *SNP<sub>j,i</sub>* can be 0, 1, or 2. Fig 3 presents a plot of the distribution (with kernel-smoothed density) of the constructed polygenic score of height in our sample. As demonstrated in the plot, the genetic spectrum of *PGS\_Height* could range from low to high, and the higher score that an individual has, the more likely that he/she would be taller.

Note that whereas Mendelian randomization is orthogonal to confounders that are correlated with but are not determined by height, it may not be orthogonal to those that are (at least partly) determined or triggered by height, such as social dominance and self-esteem discussed in Section 2.1. In this sense, the ‘causal’ effect of height identified using *PGS\_Height* captures not only the impact of genetic height, but also the impact of social reactions that are *triggered* by genetic height—in other words, it serves as an upper bound of the economic returns to genetic height.

With the genetic IV discussed above, we estimate the following first-stage equation,

$$H = \alpha_0 + \mathbf{G}\alpha_1 + \mathbf{X}\alpha_2 + v, \quad (3)$$

where  $\mathbf{G} = PGS\_Height_i$  is the genetic IV, along with Eq (1) in a two-stage least-squares (2SLS) framework.

Note finally that the above framework (Eqs 1–3) implicitly invokes the common-return assumption that the economic return to height is the same for all individuals in the underlying population—in other words,  $\beta_1$  does not have a subscript for individual *i*. von Hinke et al. [14,23] formally extended the genetic-IV framework to allow for heterogeneous return to height in the spirit of Imbens and Angrist [59] and Angrist et al. [60] and provided assumptions needed to identify these effects. Under the heterogeneous-return framework,  $\beta_1$  captures the local-average treatment-effect (LATE) for those individuals who are taller *because* they have the particular genes that made them tall. In contrast, the impact of height identified in many previous studies in developing countries using shocks to childhood nutrition (e.g., occurrences of droughts or disease outbreaks) captures the LATE for those individuals who are shorter *because* they encountered these shocks.

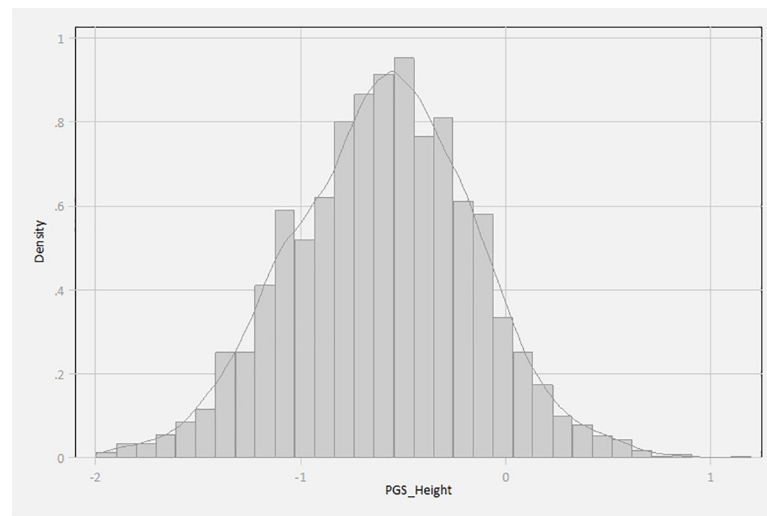
### 4.3. Instrument validity

The validity and credibility of *PGS\_Height* rely on several critical criteria that we discuss in detail below [25,61–63].

**4.3.1. Relevance.** The relevance criterion requires that the genetic IV of *PGS\_Height* must be correlated with the endogenous variable of observed height. In our data, the explanatory power of *PGS\_Height* is 0.0459, comparable to the 0.048 that Böckerman and colleagues found using a Finnish sample [25]. We further test and confirm the association between *PGS\_height* and observed height in the first stage regressions of 2SLS (reported in Table 7).

**4.3.2. Independence.** The independence criterion requires that the genetic IV of *PGS\_Height* must not be correlated with unmeasured confounders through population stratification. We first divide our analytical sample into two groups: the higher *PGS\_Height*





**Fig 3. Distribution of the genetic instrumental variable.**

<https://doi.org/10.1371/journal.pone.0230555.g003>

subsample with an above-average value and the lower PGS\_Height subsample with a below-average value of the polygenic score. We then assess whether the two groups differ in major observable characteristics significantly. As reported in Table 2, the subsample averages are in general similar between the two groups, except for the endogenous variable of observed height, suggesting little correlation between the genetic IV and other observed variables.

As a second strategy to address the potential population stratification problem, we include a total of 42 individual ancestral composition variables based on each respondent's genetic data as additional control variables (The 42 ancestries are, from high to low: Northern Han, Southern Han, Mongolian, Naxi/Yi, Japanese, Gaoshan, Korean, Dai, She, Kinh, Tibetan, Tungus, Ashkenazi, Balkan, Bantusa, Bengali, Cambodian, Egyptian, English, Eskimo, Finnish/Russian, French, Hungarian, Iranian, Kyrgyz, Lahu, Mala, Mayan, Mbuti, Miao/Yao, Papuan, Pima, Sardinian, Saudi, Sindhi, Somali, Spanish, Thai, Uygur, Uzbek, Yakut, and Yoruba.). In our data, individual ancestry composition is estimated by using the ADMIXTURE program, developed by the Department of Human Genetics, University of California Los Angeles [64–65]. Table 3 reports the descriptive information of the top ten ancestries estimated in our sample. In all subsequent analyses, we include a full set of 42 ancestral controls unless otherwise noted.

**4.3.3. Exclusion.** The exclusion criterion requires that the genetic IV of PGS\_Height have no direct effect on income through pleiotropy. In our design, if some of the SNPs used to construct PGS\_Height can directly affect the outcome variable of interest (i.e., income), then the exclusion condition of the genetic IV would be violated [23]. Although it is empirically impossible to prove the null hypothesis that PGS\_Height is uncorrelated with the error term in 2SLS regressions by using a single instrumental variable, we perform a series of tests assessing the validity of the exclusion restriction assumption [62–63].

First, we follow VanderWeele and colleagues [62] and estimate a reduced-form model, that of log earnings on the genetic score of height (*PGS\_Height*), while excluding the endogenous explanatory variable, *Height*, from the model. If height does not have a causal effect on (log) earnings, then one would expect the genetic score of height to have little impact on (log) earnings. The results, reported in Table 4 verify this expectation. The estimated coefficients on *PGS\_Height* in all three specifications are small and statistically insignificant, suggesting little predictive power of the genetic IV for the outcome variable.

Second, we compare the magnitude of the coefficients of *Height* in the OLS and the IV models. If our genetic IV satisfies the exclusion restriction, one would expect to see  $\hat{\beta}_{IV} < \hat{\beta}_{OLS}$  under the assumption that there is positive unmeasured confounding of height and income [63]. Our results reported in Table 5 (column 2,  $\hat{\beta}_{OLS} = 0.0103$ ) and Table 7 (column 1,  $\hat{\beta}_{IV} = 0.0056$ ) are consistent with this expectation (i.e.,  $0.0056 < 0.0103$ ). This result indicates no obvious violation of the exclusion restriction assumption.

It is worth noting that Böckerman et al. and Palmer et al. point out that the use of polygenic scores can significantly reduce the risk of pleiotropy compared with a single gene variant or SNP [25,66]. Böckerman et al. also explicitly demonstrate the robustness of the causal effect of height on earnings after adjusting for potential pleiotropy [25], which lends further support to the validity of using the polygenic score of height as an IV for the observed height in an MR setting.

#### 4.4. Personality traits and other genetic factors

In **X**, we include personality traits (i.e., risk attitude, altruism and trust) and additional genetic controls (i.e., cognitive ability, depression, delay discounting, and reproduction preference) for two reasons. First, personality characteristics (also known as noncognitive or soft skills) have been found to play crucial roles in determining one's labor market success [25,67–68]. Second, it is well-known that cognitive ability is closely related to labor market success [16]. Although the phenotypic variable of cognitive ability is not available in the current study, the inclusion of its proxy—the polygenic scores of cognitive ability, can help mitigate the potential omitted variable bias [48]. Similarly, if any of the behavioral traits of depression, delay discounting, or reproduction preference contributes to one's labor market success, then although their phenotypic variables are not available, it is intuitive to also include their genetic proxies (i.e., polygenic scores of depression, delay discounting, and reproduction preference) to avoid the omitted variable bias, and further strength our empirical results.

### 5. Results

Turing to our empirical results, this section first documents the existence of a height premium using OLS regressions, with various specifications that have been adopted in previous studies. By sequentially adding more explanatory variables in the model, these OLS estimates help shed light on what is being captured in the estimated height premium. We then use the genetic IV to estimate the causal effect of height, to see how genetic height affects one's earnings while netting out the influence of confounding factors.

#### 5.1. OLS results: Is there a height premium?

Table 5 reports main results of estimating Eq (1) by OLS. Overall, these results are consistent with those of previous observational studies conducted in China [20–21], which suggest a statistically significant and positive association between one's adult height and wage income.

Model 1 and 2 presents results for the full sample. As a benchmark, Model 1 controls for gender, age, years of schooling, province fixed effects, and individual ancestral composition. It estimates that one additional cm in height is associated with a 1.30% increase in annual salary income, which is similar to the findings of Gao and Smyth [20] and Yamamura et al. [21] using Chinese data.

Model 2 further includes a set of personality traits (namely, risk-loving, altruism and trust) and additional genetic markers (including polygenic scores for cognitive ability, depression, delay discounting, and reproduction preference) that can potentially affect one's labor-market

Table 2. Comparison of observable characteristics by *PGS\_Height* (n = 3,427).

Observable characteristics	Higher <i>PGS_Height</i>	Lower <i>PGS_Height</i>	Difference	t-statistics	p-value
Natural Logarithm of Annual Earnings	11.585 (0.026)	11.553 (0.029)	0.032	0.835	0.404
Height (in centimeter)	171.342 (0.225)	168.663 (0.242)	2.679	8.109***	0.000
Age	30.150 (0.186)	30.353 (0.191)	-0.202	-0.756	0.450
Years of Schooling	16.429 (0.060)	16.445 (0.062)	-0.016	-0.186	0.852
Risk loving	5.604 (0.060)	5.623 (0.062)	-0.019	-0.221	0.825
Altruism	4.814 (0.067)	4.916 (0.075)	-0.102	-1.017	0.309
Trust	5.096 (0.075)	5.070 (0.080)	0.026	0.234	0.815
PGS of Cognitive Ability	-0.009 (0.005)	-0.011 (0.005)	0.002	0.302	0.763
PGS of Depression	32.591 (0.131)	32.671 (0.112)	-0.080	0.466	0.641
PGS of Delay Discounting	0.074 (0.002)	0.074 (0.003)	0.000	-0.010	0.992
PGS of Reproduction Preference	0.028 (0.005)	0.031 (0.005)	-0.004	-0.520	0.604

<https://doi.org/10.1371/journal.pone.0230555.t002>

performance in the regression. The estimated height premium declines—a 1-cm increase in height is associated with a 1.03% increase in annual income. It is worth noting that higher risk-loving preference and lower polygenic score of delay discounting (i.e., genetically less hyperbolic or temporal discounting, and preferring future return or gratification rather than immediate rewards) are positively associated with annual income. Results in Table 6, panel A also suggest that height is significantly correlated with risk-loving and altruism (columns 3–6), even after controlling for the genetic determinant of height (i.e., *PGS\_Height*). Table 6, panel B further explores the co-occurrence of height and those polygenic scores, and interestingly it appears that higher genetic cognitive ability (column 2) and lower genetic risk of depression

Table 3. Descriptive statistics of top 10 ancestries.

Ancestry/Population	Mean	Std.Dev.	Min	Max
Northern Han	0.5530	0.2963	0.0000	0.9996
Southern Han	0.2579	0.2645	0.0000	0.9996
Mongolian	0.0604	0.1124	0.0000	0.7774
Naxi/Yi	0.0292	0.0610	0.0000	0.9996
Japanese	0.0202	0.0369	0.0000	0.2173
Gaoshan	0.0084	0.0171	0.0000	0.1163
Korean	0.0075	0.0105	0.0000	0.0575
Dai	0.0070	0.0203	0.0000	0.1696
She	0.0056	0.0096	0.0000	0.0540
Kinh	0.0054	0.0147	0.0000	0.1260

Source: author's calculation.

<https://doi.org/10.1371/journal.pone.0230555.t003>

**Table 4. Additional tests of the exclusion restriction assumption.**

	(1)	(2)	(3)
	ln(income)	ln(income)	ln(income)
<i>PGS_Height</i>	0.0047 (0.0436)	0.0057 (0.0403)	0.0046 (0.0435)
Male		0.027 (0.0364)	0.0277 (0.0392)
Age		0.2042*** (0.0188)	0.2031*** (0.0202)
Age^2		-0.0023*** (0.0003)	-0.0023*** (0.0003)
Years of Schooling		0.0498*** (0.0090)	0.0525*** (0.0096)
Constant	11.5921*** (0.0328)	6.8570*** (0.3264)	6,236.7441** (2,836.0718)
Additional Controls of Personality and Other Polygenic Scores	Yes	Yes	Yes
Province FE	No	No	Yes
Ancestral Controls	No	No	Yes
Observations	3,427	3,427	3,427
R-squared	0.0006	0.1468	0.1845

\*\*\*

\*\*, and

\* indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

<https://doi.org/10.1371/journal.pone.0230555.t004>

(column 4) are associated with higher genetic height (but not higher observed height), which is consistent with the simultaneous neural and height growth discussed by Tanner [41] and Thompson and Potter [42]. These results imply that one should control for those polygenic scores in the regression to avoid and/or mitigate the omitted-variable bias and potential pleiotropy whenever possible. In the following models, we include these personality traits and additional genetic markers unless otherwise noted.

Model 3 and 4 of Table 4 report OLS estimates of male and female respondents, respectively. Similar to the results of Gao and Smyth [20], the estimated return to height is more pronounced among males than among females. Model 5 analyzes the subsample of prime working-age individuals (age 30–50), in which a 1-cm increase in height is associated with a 1.28% increase in annual income, higher than that of the full sample.

Taken together, the OLS results reported in Table 5 indicate that height has a statistically and economically significant effect on (the log of) annual salary income, even after controlling for a large set of covariates that are likely to be correlated with height. Since it is difficult to control for all potential confounding factors, it remains unclear whether the estimated height premium will further decline when more covariates are included in the OLS model. Assessing this issue from another angle, the next subsection aims to estimate the economic return to genetic height, using genetic markers as an IV for observed height.

## 5.2. 2SLS results: Is the height premium causal?

Table 7 presents the main results of two-stage least squares (2SLS) regressions, with odd-numbered columns reporting results of second-stage regressions and even-numbered ones

Table 5. OLS results.

Variables	(1)	(2)	(3)	(4)	(5)
	Pooled	Pooled	Male	Female	Age 30–50
	ln(income)	ln(income)	ln(income)	ln(income)	ln(income)
Height	0.0130*** (0.0021)	0.0103*** (0.0023)	0.0138*** (0.0030)	0.0084** (0.0037)	0.0128*** (0.0035)
Male	-0.0105 (0.0355)	0.0167 (0.0382)	- -	- -	0.0036 (0.0566)
Age	0.1994*** (0.0181)	0.1995*** (0.0194)	0.2347*** (0.0269)	0.1732*** (0.0290)	0.1888** (0.0767)
Age^2	-0.0023*** (0.0003)	-0.0023*** (0.0003)	-0.0028*** (0.0004)	-0.0018*** (0.0004)	-0.0021** (0.0010)
Years of schooling	0.0525*** (0.0088)	0.0555*** (0.0093)	0.0249** (0.0124)	0.0880*** (0.0148)	0.0522*** (0.0134)
<i>Personality traits:</i>					
Risk Loving	- -	0.0542*** (0.0092)	0.0415*** (0.0122)	0.0734*** (0.0147)	0.0374*** (0.0139)
Altruism	- -	0.0011 (0.0081)	0.0033 (0.0111)	0.0025 (0.0124)	0.0098 (0.0120)
Trust	- -	0.0006 (0.0073)	-0.0094 (0.0097)	0.0090 (0.0116)	-0.0001 (0.0113)
<i>Additional genetic factors:</i>					
Cognitive ability	- -	-0.1138 (0.1128)	-0.1152 (0.1519)	-0.0079 (0.1751)	-0.0325 (0.1699)
Depression	- -	-0.0050 (0.0043)	-0.0071 (0.0057)	0.0002 (0.0068)	-0.0144** (0.0066)
Delay discounting	- -	-0.5264** (0.2218)	-0.6943** (0.2980)	-0.1819 (0.3415)	-0.3947 (0.3257)
Reproduction preference	- -	0.0092 (0.1144)	-0.0318 (0.1562)	-0.0686 (0.1745)	0.0583 (0.1678)
Constant	4.3258*** (0.4841)	4,431.2922 (2,756.9803)	3,905.0754 (3,648.9270)	4,307.0775 (4,355.0797)	6,813.2636* (4,118.1286)
Province FE	Yes	Yes	Yes	Yes	Yes
Ancestral controls	Yes	Yes	Yes	Yes	Yes
Observations	3,427	3,427	1,922	1,505	1,843
R-squared	0.1631	0.2187	0.2344	0.2829	0.1835

\*\*\*

\*\*, and

\* indicate statistical significance at the 1%, 5%, and 10% levels, respectively. In all models, we control for province fixed effects and 42 individual ancestry composition variables.

<https://doi.org/10.1371/journal.pone.0230555.t005>

reporting first-stage regression results. In all models, the polygenic score of height (*PGS\_Height*) is used as the sole instrument for observed height. An important finding emerging from Table 7 is that the 2SLS estimates of the impact of height become much smaller and are no longer statistically significant at any conventional level, implying little causal link between genetic height and earnings. This finding, in turn, suggests that labor-market discrimination is unlikely to be the key factor driving the observed height premium documented by OLS regressions in Table 5.

**Table 6. Relationship between height and additional control variables.**

A. Correlations between height and personality traits								
	-1	-2	-3	-4	-5	-6	-7	-8
	Years of Schooling	Years of Schooling	Risk Loving	Risk Loving	Altruism	Altruism	Trust	Trust
Height	-0.0018	-0.0008	0.0294***	0.0298***	0.0247***	0.0241***	-0.0124	-0.0115
<i>PGS_Height</i>	-0.0062	-0.0065	-0.0065	-0.0067	-0.0076	-0.0079	-0.0083	-0.0086
	-	-0.1119	-	-0.0986	-	-0.0694	-	0.0508
	-	-0.1162	-	-0.1208	-	-0.1416	-	-0.1547
Constant	16.7378***	16.4893***	0.6234	0.5003	0.7132	0.7736	7.2942***	7.1546***
	-1.0582	-1.114	-1.1026	-1.157	-1.2967	-1.3559	-1.4131	-1.4808
Observations	3,427	3,427	3,427	3,427	3,427	3,427	3,427	3,427
R-squared	0.0001	0.0007	0.013	0.0128	0.0067	0.0062	0.0014	0.0012

B. Correlations between height and other PGS								
Variables	-1	-2	-3	-4	-5	-6	-7	-8
	PGS: Cognitive Ability	PGS: Cognitive Ability	PGS: Depression	PGS: Depression	PGS: Delay Discounting	PGS: Delay Discounting	PGS: Reproduction Preference	PGS: Reproduction Preference
Height	0.0003	0	-0.0164	-0.0183	-0.0002	-0.0002	-0.0004	-0.0005
	-0.0005	-0.0005	-0.0182	-0.0121	-0.0003	-0.0003	-0.0005	-0.0005
<i>PGS_Height</i>	-	0.0206**	-	-0.2373**	-	0.0028	-	0.0059
	-	-0.0096	-	-0.1118	-	-0.005	-	-0.0096
Constant	-0.0578	-0.0026	34.7815***	35.6082***	0.1066**	0.1135**	0.0965	0.1112
	-0.0863	-0.0915	-3.0966	-2.0798	-0.0462	-0.0478	-0.0887	-0.0918
Observations	3,427	3,427	3,427	3,427	3,427	3,427	3,427	3,427
R-squared	0.0002	0.0032	0.0005	0.0028	0.0003	0.0005	0.0004	0.0007

\*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

<https://doi.org/10.1371/journal.pone.0230555.t006>

More specifically, the first model of full sample (column 1) suggests that the impact of height on annual income identified by Mendelian randomization of height-related genes upon conception is very small compared to its OLS counterpart (Table 5, column 2)—other things being equal, an additional cm in (genetic) height increases one’s annual salary by less than 0.6%. Such an impact is not statistically significant at any conventional level, echoing the finding of Böckerman et al. using Finnish data [25]. The corresponding first-stage regression (column 2) reveals no sign of the weak-IV problem: the IV used, *PGS\_Height*, has a very strong predictive power for height, with an associated F-statistic of 85.1053, far exceeding the rule-of-thumb value of 10 [69]. Models 2–4 further perform 2SLS estimations by subsamples of males, females, and prime working-age participants, respectively, but neither model yields a statistically or economically significant height premium.

As additional checks, we perform 2SLS by including the quadratic form of *PGS\_Height* (i.e., *PGS\_Height*<sup>2</sup>) as an additional genetic IV to the above models again. The key results remain robust (reported in S1 Appendix in Table A1), whereas first-stage regression results show little predictive power of *PGS\_Height*<sup>2</sup> (i.e., insignificant in all models), implying a linear effect of *PGS\_Height* on observed height. It is worth pointing out that the use of two genetic IVs facilitates an overidentification test for the validity of the exclusion restriction. For example, in

Table 7. 2SLS results.

Outcome variables	Model 1: Pooled		Model 2: Male		Model 3: Female		Model 4: Age 30–50	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2SLS	First Stage	2SLS	First Stage	2SLS	First Stage	2SLS	First Stage
	Ln(income)	Height	Ln(income)	Height	Ln(income)	Height	Ln(income)	Height
Height	0.0056 (0.0107)	- -	0.0063 (0.0172)	- -	0.0029 (0.0137)	- -	0.0075 (0.0152)	- -
Male	0.0217 (0.0377)	10.1720*** (0.3530)	- -	- -	- -	- -	0.0031 (0.0554)	10.5547*** (0.5591)
Age	0.2106*** (0.0195)	-0.1478 (0.1608)	0.2482*** (0.0264)	-0.2206 (0.2207)	0.1771*** (0.0290)	-0.1977 (0.2440)	0.2033*** (0.0765)	0.0973 (0.7687)
Age^2	-0.0024*** (0.0003)	0.0004 (0.0024)	-0.0030*** (0.0004)	0.0013 (0.0032)	-0.0019*** (0.0004)	0.0014 (0.0036)	-0.0023** (0.0010)	-0.0022 (0.0101)
Years of schooling	0.0495*** (0.0092)	0.0419 (0.0848)	0.0212* (0.0121)	0.0464 (0.1146)	0.0755*** (0.0142)	-0.0428 (0.1318)	0.0509*** (0.0132)	0.0165 (0.1326)
Risk loving	0.0537*** (0.0100)	0.4047*** (0.0833)	0.0434*** (0.0135)	0.3397*** (0.1128)	0.0693*** (0.0153)	0.5015*** (0.1290)	0.0340** (0.0145)	0.4152*** (0.1331)
Altruism	0.0012 (0.0080)	0.1350* (0.0737)	0.0036 (0.0107)	0.0677 (0.1019)	0.0017 (0.0123)	0.1835 (0.1116)	0.0093 (0.0115)	0.1253 (0.1161)
Trust	0.0010 (0.0076)	-0.2059*** (0.0669)	-0.0100 (0.0096)	-0.1447 (0.0902)	0.0115 (0.0120)	-0.2547** (0.1044)	0.0013 (0.0109)	-0.0918 (0.1089)
Cognitive ability	-0.1092 (0.1104)	-0.1979 (1.0313)	-0.1027 (0.1470)	-0.1497 (1.4084)	-0.0250 (0.1673)	-0.1688 (1.5711)	-0.0316 (0.1634)	-1.8792 (1.6424)
Depression	-0.0052 (0.0042)	0.0323 (0.0407)	-0.0070 (0.0055)	0.0091 (0.0547)	-0.0002 (0.0065)	0.0410 (0.0631)	-0.0145** (0.0063)	0.0043 (0.0647)
Delay discounting	-0.5255** (0.2182)	-2.3509 (2.0091)	-0.7054** (0.2869)	-0.8149 (2.7791)	-0.1790 (0.3279)	-3.1801 (2.9905)	-0.3644 (0.3129)	-1.8665 (3.1492)
Reproduction preference	0.0148 (0.1125)	-1.3113 (1.0587)	-0.0257 (0.1519)	0.8948 (1.4567)	-0.0201 (0.1768)	-4.3051*** (1.5927)	0.0949 (0.1636)	-2.2624 (1.6250)
Constant	109.9583 (80.6906)	96.1696 (200.3519)	12.0476 (146.3101)	86.0966 (204.1028)	-0.4013 (12.8593)	39.8194 (210.7858)	3.8526 (40.5838)	15.5517 (174.1510)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ancestral controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Instrumental Variable:</i>								
<i>PGS_Height</i>	- -	3.9284*** (0.3947)	- -	3.2287*** (0.5302)	- -	4.6578*** (0.6137)	- -	3.6183*** (0.6200)
First-stage F statistic	85.1053		29.9235		52.1446		34.3092	
p-value	0.0000		0.0000		0.0000		0.0000	
Observations	3,427		1,922		1,505		1,843	

\*\*\*, \*\*, and \* indicate statistical significance at the 1%, 5%, and 10% levels, respectively. In all models, we control for province fixed effects and 42 individual ancestry composition variables.

<https://doi.org/10.1371/journal.pone.0230555.t007>

Model 1 of the full sample (S1 Appendix in Table A1, column 1–2), results of the standard overidentification test (Sargan statistic = 2.2653,  $p = 0.1323$ ) reveal no sign of violation of the exclusion restriction, lending further support to the IV validity of *PGS\_Height*.

## 6. Concluding remarks

The income disparities associated with one's physical appearance (such as height, stature, BMI, facial appearance) have motivated a large body of studies that seek to identify the underlying causal mechanisms [33]. This paper picks up this agenda and uses very recent data from China to show that without correcting for endogeneity, body height is indeed associated with higher salary income. However, once we have instrumented the observed height with its genetic determinants (i.e., the polygenic score of height), such height premium vanishes. Taking advantage of the unique prosperous individual genotyping data available in our sample, we then explore and illustrate that genetic cognitive ability, genetic risk of depression, various personality traits, as well as remaining confounders that are correlated with height are likely to be those truly captured in the observed height premium.

While we find an economically insubstantial premium of height *per se*, which suggests that discrimination from the workplace may not be a crucial component to wage disparity with respect to height, we interpret these results as empirical support of channels through which both cognitive and non-cognitive skills are correlated with height during individual development. In other words, the observed height may act as a signal of beneficial circumstances for developing higher cognitive/non-cognitive skills during childhood or early-life [25]. This naturally suggests a role for redistributive policies to support children and youths from lower SES households. If individuals from households with limited resources do not reach their height and/or human capital potentials due to insufficient human capital investment in childhood, then simply implementing anti-discrimination laws in the labor market may not substantially reduce income inequality.

Before closing, there are several limitations of the present study. First, the current study is based on a sample of 3,427 observations, which may lack statistical power due to small sample size [70]. Second, as mentioned earlier, the WeGene sample we use in this study is not nationally representative with respect to the total population in China. More studies may be needed to confirm the generality of our findings. Third, the polygenic score of height in this paper is constructed as an aggregate measure of 697 SNPs, in which most of these SNPs are involved in skeletal growth and bone development from pathway analyses [47]. Although this strongly suggests that the LATE presented in the current study can be mainly interpreted as resulting from the main biological pathways of skeletal growth/bone development, we lack a comprehensive understanding of the potential different LATEs through other minor biological mechanisms due to data limitations, and it can be an intriguing direction for future research. Fourth, although we find that height and cognitive/non-cognitive abilities are correlated at the genetic level, the underlying biological and/or environmental mechanisms are still unclear. One possible explanation of that may be *positive selection* from the realm of evolutionary biology, in which beneficial traits (i.e. those that can make carriers survive and reproduce more successfully) tend to be selected by exogenous environments and become more frequent in populations over time [71]. Future research may look for supporting evidence or explore additional explanation to gain a deeper understanding of the co-occurrence of height and cognitive/non-cognitive abilities in modern humans.

## Supporting information

**S1 Appendix.**  
(DOCX)

## Author Contributions

**Conceptualization:** Jun Wang, Chen Zhu.



**Data curation:** Gang Chen, Yingxiang Li, Guoshu Kong.

**Formal analysis:** Chen Zhu.

**Methodology:** Qihui Chen, Gang Chen, Yingxiang Li.

**Resources:** Jun Wang.

**Validation:** Jun Wang.

**Writing – original draft:** Qihui Chen, Chen Zhu.

**Writing – review & editing:** Qihui Chen, Chen Zhu.

## References

1. Case A., Paxson C., 2008a. Stature and status: height, ability: and labor market outcomes. *Journal of Political Economy* 116, 499–532.
2. Case A., Paxson C., 2008b. Height, health and cognitive function at older ages. *American Economic Review*, 98(2), 463–467.
3. Gladwell M. 2005. *Blink: The Power of Thinking Without Thinking*. New York: Little Brown & Company.
4. Heineck G. 2005. Up in the skies? The relationship between height and earnings in Germany. *Labour*, 19(3), 469–489.
5. Kortt M., Leigh A. 2010. Does size matter in Australia? *Economic Record* 86(272), 71–83.
6. Dinda S., Gangopadhyay P.K., Chattopadhyay B.P., Saiyed H.N., Pal M. and Bharati P. 2006. Height, weight and earnings among coalminers in India. *Economics and Human Biology*, 4(3), 342–350. <https://doi.org/10.1016/j.ehb.2005.10.003> PMID: 16377265
7. LaFave D., Thomas D. 2017. Height and Cognition at Work: Labor Market Productivity in a Low Income Setting, *Economics and Human Biology* 25, 52–64. <https://doi.org/10.1016/j.ehb.2016.10.008> PMID: 27843117
8. Schultz T.P. 2002. Wage gains associated with height as a form of health human capital. *American Economic Review* 92(3), 349–453.
9. Sohn K. 2015. The height premium in Indonesia. *Economics and Human Biology* 16, 1–15. <https://doi.org/10.1016/j.ehb.2013.12.011> PMID: 24480546
10. Vogl TS. Height, skills, and labor market outcomes in Mexico. *Journal of Development Economics*. 2014 Mar 1; 107:84–96.
11. Cinnirella, F., Winter, J., 2009. Size matters! Body height and labor market discrimination: A cross-European analysis. CESifo Working Paper No. 2733.
12. Hamstra M.R.W. 2014. 'Big' men: male leaders' height positively relates to followers' perception of charisma. *Pers. Individ. Diff.* 56, 190–192.
13. Heineck G. 2009. Too tall to be smart: The relationship between height and cognitive abilities. *Economics Letters* 105(1), 78–80.
14. von Hinke Kessler Scholder S., Smith G.D., Lawlor D.A., et al., 2013. Child height, health and human capital: evidence using genetic markers. *European Economic Review* 57, 1–22. <https://doi.org/10.1016/j.eurocorev.2012.09.009> PMID: 25673883
15. Persico N, Postlewaite A, Silverman D. The effect of adolescent experience on labor market outcomes: The case of height. *Journal of Political Economy*. 2004 Oct; 112(5):1019–53.
16. Heckman J.J., Stixrud J., Urzua S. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3), 411–482.
17. Walder A. 1995. Career mobility and the Communist political order. *American Sociological Review* 60 (3), 309–328.
18. Groves T., Hong Y., McMillan J. Naughton B. 1995. China's evolving managerial labor market. *Journal of Political Economy* 103(4), 873–892.
19. Lardy NR. Markets over Mao: The rise of private business in China. Columbia University Press; 2014 Sep 10.
20. Gao W., Smyth R. 2009. Health human capital, Height and Wages in China. *Journal of Development Studies* 46(3), 466–484.

21. Yamamura E., Smyth R., Zhang Y., 2015. Decomposing the effect of height on income in China: the role of market and political channels. *Economics and Human Biology* 19, 62–74. <https://doi.org/10.1016/j.ehb.2015.08.003> PMID: 26344779
22. Böckerman P., Vainiomäki J., 2013. Stature and life-time labor market outcomes: accounting for unobserved differences. *Labour Economics* 24, 86–96.
23. von Hinke S., Smith G.D., Lawlor D.A., Propper C, Windmeijer F. 2016. Genetic markers as instrumental variables. *Journal of Health Economics* 45, 131–148. <https://doi.org/10.1016/j.jhealeco.2015.10.007> PMID: 26614692
24. Tyrrell J., Jones S.E., Beaumont R., Astley C.M., Lovell R., Yaghoobkar, et al. 2016. Height, body mass index, and socioeconomic status: mendelian randomisation study in UK Biobank. *BMJ*, 352, p.i582. <https://doi.org/10.1136/bmj.i582> PMID: 26956984
25. Böckerman P., Viinikainen J., Vainiomäki J., Hintsanen M., Pitkänen N., Lehtimäki, et al. 2017. Stature and long-term labor market outcomes: Evidence using Mendelian randomization. *Economics and Human Biology* 24, 18–29. <https://doi.org/10.1016/j.ehb.2016.10.009> PMID: 27846416
26. Hingorani A., Humphries S. 2005. Nature's randomised trials. *Lancet* 366, 1906–1908. [https://doi.org/10.1016/S0140-6736\(05\)67767-7](https://doi.org/10.1016/S0140-6736(05)67767-7) PMID: 16325682
27. Allen H., Estrada K., Lettre G., Berndt S., Weedon M., et al. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467, 832–838. <https://doi.org/10.1038/nature09410> PMID: 20881960
28. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome medicine*. 2017 Dec; 9(1):96. <https://doi.org/10.1186/s13073-017-0489-y> PMID: 29132412
29. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. *Molecular psychiatry*. 2018 May; 23(5):1368. <https://doi.org/10.1038/mp.2017.163> PMID: 28785111
30. Duncan L., Shen H., Gelaye B. et al. 2019. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, 10, 3328, <https://doi.org/10.1038/s41467-019-11112-0> PMID: 31346163
31. Haddad L.J., Bouis H.E. 1991. The impact of nutritional status on agricultural productivity: Wage evidence from the Philippines. *Oxford Bulletin of Economics and Statistics*, 53(1), 45–68.
32. Steckel R.H. 1995. Height and the standard of living. *Journal of Economic Literature*, 33(4), 1903–1940.
33. Strauss J., Thomas D. 1998. Health, Nutrition, and Economic Development. *Journal of Economic Literature* 36(2), 766–817.
34. Hensley W.E. 1993. Height as a measure of success in Academe. *Psychology: Journal of Human Behavior* 30(1), 40–46.
35. Martel LF, Biller HB. Stature and Stigma: The biopsychosocial development of short males. Lexington Books/DC Heath and Com; 1987.
36. Judge T., Cable D.M. 2004. The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology* 89(3), 428–411. <https://doi.org/10.1037/0021-9010.89.3.428> PMID: 15161403
37. Fessler, D., Gneezy, U., List, J., Hoffman, M. 2010. Height and Competitiveness. Rady University of California San Diego Working Paper.
38. Macintyre S., West P. 1991. Lack of class variation in health in adolescence: an artefact of an occupational measure of social class. *Social Science and Medicine*, 32, 395–402. [https://doi.org/10.1016/0277-9536\(91\)90340-i](https://doi.org/10.1016/0277-9536(91)90340-i) PMID: 2024154
39. Dipboye R.L., Fromkin H.L., Wiback K. 1975. Relative importance of applicant sex, attractiveness and scholastic standing in evaluation of job applicant resumes. *Journal of Applied Psychology* 60(1), 39–43.
40. Magnusson P., Rasmussen F., Gyllenstein U. 2006. Height at age 18 is a strong predictor of attained education later in life: cohort study of over 950000 Swedish men. *International Journal of Epidemiology* 35, 658–663. <https://doi.org/10.1093/ije/dyl011> PMID: 16446353
41. Tanner J. 1978. *Fetus into Man: Physical Growth from Conception to Maturity*. Cambridge: Harvard University Press.
42. Thompson C., Potter G.B. 2000. Thyroid Hormone Action in Neural Development. *Cerebral Cortex* 10 (10), 939–945. <https://doi.org/10.1093/cercor/10.10.939> PMID: 11007544
43. Loh E.S. 1993. The economic effects of physical appearance. *Social Science Quarterly*, 74(2), 420–434.

44. Dohmen T., Falk A., Huffman D. and Sunde U., 2010. Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3), 1238–60.
45. Falk A., Becker A., Dohmen T., Enke B., Huffman D., Sunde U. (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics*, Volume 133, Issue 4, November 2018, Pages 1645–1692, <https://doi.org/10.1093/qje/qjy013>.
46. Roberts J.S. and Ostergren J., 2013. Direct-to-consumer genetic testing and personal genomics services: a review of recent empirical studies. *Current genetic medicine reports*, 1(3), pp.182–200. <https://doi.org/10.1007/s40142-013-0018-2> PMID: 24058877
47. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*. 2014 Nov; 46(11):1173. <https://doi.org/10.1038/ng.3097> PMID: 25282103
48. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, De Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature genetics*. 2018 Jul; 50(7):912. <https://doi.org/10.1038/s41588-018-0152-6> PMID: 29942086
49. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics*. 2018 May; 50(5):668. <https://doi.org/10.1038/s41588-018-0090-3> PMID: 29700475
50. Sanchez-Roige S, Fontanillas P, Elson SL, Pandit A, Schmidt EM, Foerster JR, et al. Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nature neuroscience*. 2018 Jan; 21(1):16. <https://doi.org/10.1038/s41593-017-0032-x> PMID: 29230059
51. Barban N, Jansen R, De Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature genetics*. 2016 Dec; 48(12):1462. <https://doi.org/10.1038/ng.3698> PMID: 27798627
52. Mincer JA. Schooling, Experience, and Earnings. NBER Books. 1974.
53. Neal D.A., Johnson W.R. 1996. The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy* 104, 869–95.
54. Heckman J. J. 1998. Detecting Discrimination. *Journal of Economic Perspectives* 12, 101–16.
55. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease?. *International journal of epidemiology*. 2003 Feb 1; 32(1):1–22. <https://doi.org/10.1093/ije/dyg070> PMID: 12689998
56. Smith G.D. and Ebrahim S., 2008. Mendelian randomization: genetic variants as instruments for strengthening causal inference in observational studies. In Biosocial Surveys. National Academies Press (US).
57. Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, 42(7), p.565. <https://doi.org/10.1038/ng.608> PMID: 20562875
58. Jelenkovic A., Sund R., Hur Y.M., Yokoyama Y., Hjelmborg J.V.B., Möller S., et al. 2016. Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts. *Scientific reports*, 6, p.28496. <https://doi.org/10.1038/srep28496> PMID: 27333805
59. Imbens GW, Angrist JD. Identification and Estimation of Local Average Treatment Effects. *Econometrica*. 1994 Mar; 62(2):467–75.
60. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*. 1996 Jun 1; 91(434):444–55.
61. Willage B. (2018). The effect of weight on mental health: New evidence using genetic IVs. *Journal of Health Economics*, 57, 113–130. <https://doi.org/10.1016/j.jhealeco.2017.11.003> PMID: 29289809
62. VanderWeele T. J., Tchetgen Tchetgen E. J. T., Cornelis M., & Kraft P. (2014). Methodological challenges in mendelian randomization. *Epidemiology* (Cambridge, Mass.), 25(3), 427.
63. Glymour M. M., Tchetgen Tchetgen E. J., & Robins J. M. (2012). Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *American Journal of Epidemiology*, 175(4), 332–339. <https://doi.org/10.1093/aje/kwr323> PMID: 22247045
64. Alexander D.H., Novembre J. and Lange K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*.
65. Alexander D.H. and Lange K., 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC bioinformatics*, 12(1), p.246.
66. Palmer T. M., Lawlor D. A., Harbord R. M., Sheehan N. A., Tobias J. H., Timpson N. J., et al. (2012). Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*, 21(3), 223–242. <https://doi.org/10.1177/0962280210394459> PMID: 21216802

67. Heckman J. J., & Rubinstein Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, 91(2), 145–149.
68. Kautz T., Heckman J. J., Diris R., Ter Weel B., & Borghans L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success (No. w20749). *National Bureau of Economic Research*.
69. Bound J., Jaeger D.A., Baker R. 1995. Problems with instrumental variables estimation when the correlations between the instruments and the endogenous variable is weak. *Journal of the American Statistical Association* 90, 443–450.
70. Freeman G., Cowling B. J., & Schooling C. M. (2013). Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *International Journal of Epidemiology*, 42(4), 1157–1163. <https://doi.org/10.1093/ije/dyt110> PMID: 23934314
71. Sabeti P.C., Schaffner S.F., Fry B., Lohmueller J., Vargill P., Shamovsky O., et al. 2006. Positive natural selection in the human lineage. *Science*, 312(5780), pp.1614–1620. <https://doi.org/10.1126/science.1124309> PMID: 16778047