Assortative Mating on Blood Type

Yao Hou[1], Ke Tang[2], Jingyuan Wang[3], Danxia Xie[4] and Hanzhe Zhang[5]

June 8, 2022

Abstract: Blood type is one of the most fundamental phenotypes in biological, medical, and psychological studies. Using unique data from 1 million Chinese couples, we find strong and robust evidence from a group of statistical tests for assortative mating on blood type, especially for type-O blood. An evolutionary matching model is developed to rationalize these findings. The paper shows that the biological and psychological advantages of sharing blood types are manifested in mate choice behavior.

Keywords: Assortative mating, blood type

[1] houy20@mails.tsinghua.edu.cn; Institute of Economics, School of Social Sciences, Tsinghua University, Beijing, China.
[2] ketang@tsinghua.edu.cn; Institute of Economics, School of Social Sciences, Tsinghua University, Beijing, China.
[3] jywang@buaa.edu.cn; New Main Building G810, Beihang Unversity, Beijing, China, 100191.
[4] xiedanxia@tsinghua.edu.cn; Institute of Economics, School of Social Sciences, Tsinghua University, Beijing, China.
[5] hanzhe@msu.edu; Department of Economics, Michigan State University, East Lansing, Michigan, United States.

Section 1. Introduction

It has been found that spousal pairs share a wide array of biological, socioeconomic, and psychological traits, showing the highly nonrandom nature of human mate choice[1-11]. Assortative mating may bring about important genetic consequences by increasing trait variance in a population, intensifying trait divergence, and providing life-saving benefits, for example, during organ exchanges[12-14]. Motivated by significant multi-disciplinary implications, studies reveal that many phenotypes, including body mass index (BMI), body height, and intelligence quotient (IQ)[10,15-20], are highly correlated within spousal pairs. However, blood type, one of the most fundamental phenotypes that was first discovered over one century ago, remains untapped in studies on assortative mating[21]. It is therefore of significance to investigate whether there is assortative mating on blood type in the population.

In this study, we aim to investigate whether there is mating is assortative on blood type and distinguish it from other possible sources of spousal concordance. First, we make use of the full sample of 967,329 couples obtained from 2014-2015 Chinese prepregnancy checkup data and perform a chi-square test on blood-type associations within spousal pairs to evaluate the degree of assortative mating on blood type. Because prepregnancy checkup is mandatory in China, our dataset represents an unbiased—in fact, comprehensive—sample of the population. We then take a series of alternative measures to check whether the obtained results are robust. Third, we perform a meta-analysis using subsamples divided according to regions where couples receive prepregnancy checkups to control the effect of population stratification on the estimation. Next, we perform logistic regression analysis and linear regression analysis to isolate our estimates from population stratification or the relationship maintenance effect. Specifically, we regress one's blood type on spousal blood type and incorporate the share of blood type in the local population, the share of the individual's blood type in her ethnicity, and length of marriage into the regression. Finally, we develop a simple evolutionary model of matching to explain the findings from the data.

Section 2. Results

We first use Pearson's chi-square test on a contingency table for spousal pairs' blood types to evaluate whether blood type influences human mate choice. Assortative mating typically refers to a mating choice pattern in which individuals with similar phenotypes mate with each other more frequently than the theoretical prediction under a random mating pattern. This definition implies that the degree of assortative mating on a certain phenotype can be measured by a chi-square test that compares the contingency table for pairs' blood types to the contingency table generated under a random matching pattern. To perform this test, we use the full sample of Chinese prepregnancy checkup data and aggregate information of each spousal pair's blood type by a contingency table. The raw dataset consists of 1,137,010 couples from 31 provinces of China. For convenience for our later analysis, we remove observations with incomplete information related to the couple's blood types, living areas, birthplaces, ethnicity, and marital status. We obtain a sample of 967,329 couples with complete personal information to serve as our full sample. As the contingency table reports, diagonal elements have a higher frequency than expected by random mating, showing that spousal pairs with the same blood type are more likely to marry each other (Table 1). Pearson's chi-square test (chi-square statistic: 4159.012, p value: 0.000, degree of freedom: 9, Cramer's V: 0.038) further validates nonrandom mating

in the blood type observed in the full sample.

Table 1 Observed mating pairs and ratios over expected mating pairs

		Female			
		A Group	B Group	AB Group	O Group
Male	A Group	91,537 +5.1%	80,187 +0.2%	25,219 -0.0%	89,652 -4.9%
	B Group	82,603 +2.2%	78,706 +6.0%	23,616 +0.9%	81,002 -7.4%
	AB Group	25,119 -1.9%	24,852 +5.6%	8,780 +18.4%	25,491 -8.0%
	O Group	94,640 -5.8%	86,431 -6.4%	27,528 -5.4%	121,966 +12.2%

Note: The percentages of observed mating pairs above the expected mating pairs are underlined.

We subsequently adopt alternative measures to evaluate the degree of assortative mating to investigate whether our observation obtained from the chi-square test is robust. The chi-square test shows whether we can reject the null hypothesis of random mating. However, it cannot tell us the specific pattern that the mating process follows if the null hypothesis is rejected. To learn whether mating choice is assortative, we are particularly interested in the diagonal elements of the contingency table. Specifically, we want to know whether the numbers of matching in these four specific cells are significantly different than expected. To do this, we compute adjusted Pearson residuals that indicate whether the number of mating pairs in a specific cell is significantly different than expected and perform statistical tests on the statistics of diagonal elements at the $\alpha = 5\%$ level. It shows that the four pairs of blood types, i.e., the diagonal elements of country-level contingency table (A, A), (B, B), (AB, AB), and (O, O), have adjusted Pearson residuals of 21.604, 22.498, 17.375, and 60.498, respectively, all of which are significantly higher than the two-sided critical value after Bonferroni correction $\pm$ 2.498. The results suggest that individuals of all blood types—and especially those with type-O blood—tend to marry who share the same blood type.

We also use the Altham index, which is widely used in testing the association of unordered rows and columns of a $r \times s$ contingency table, to alternatively measure the overall extent of assortative mating[33-35]. The Altham index uses the odds ratio of the likelihood of matching within different blood type pairs to capture the distance between the row-column associations in the observed contingency table and those generated by random matching patterns (see Methods). Its value is equal to zero if mating choice is random and increases with the degree of nonrandom row-column association. The estimate of the Altham index with the full sample is 2.067, showing a significant correlation in spousal blood types.

We then restrict our analysis to local subsamples (couples who were born in the same region and receive a prepregnancy checkup in this region) and perform a meta-analysis on assortative mating to investigate whether the assortative mating pattern still exists after controlling for possible subpopulation structure and whether this pattern is universal among different areas in China[14]. We identify locally matched observations by birthplace information as well as living area information provided by Chinese prepregnancy checkup data. Meta-analysis is carried out at the city level (that is, we evaluate assortative mating within a city). We report the statistical test results of cities whose sample sizes are larger than 10,000 (Table

2). In most cities, Pearson's chi-square statistics indicate that mating choice is highly nonrandom. Adjusted Pearson residuals further suggest that assortative mating of blood type is common among different areas even after the subpopulation structure is controlled for. Blood type pairs (O, O) typically have higher adjusted Pearson residuals than other blood type pairs, showing a higher degree of assortative mating. It is also noticed that the values of adjusted Pearson residuals we obtain from locally matched samples in each city are generally lower than those we obtain from the country-level full sample. This result implies that compounded by population stratification, biased estimates of assortative mating are obtained in the analysis of the full sample.

Meta-analysis helps to eliminate subpopulation structure. However, another possible mechanism, relationship maintenance, may also confound the estimation of assortative mating. The similarity of spousal pairs' blood types can be explained by the concordance of living with a partner with the same blood type. Those couples with different blood types might feel it hard to get along with each other and then fail to be enrolled in the prepregnancy checkup dataset, as their relationship has a higher probability of breaking up before they get married. Therefore, we perform regression analysis to isolate the effect of assortative mating on spousal concordance on blood type from that of relationship maintenance as well as that of subpopulation structure. The regression can also help us quantitatively evaluate how much variation in the blood type of one individual's partner assortative mating as well as the other two alternative mechanisms, population stratification and relationship maintenance, can explain. Specifically, we regress one's blood type on the individual's partner's blood type and incorporate a group of control variables into the regression, which includes the share that the individual's blood type takes up in the population living in her birthplace, the share that this type takes up in her ethnicity, length of marriage, and its interaction term with the partner's blood type (see Methods for details). For comparison, we also run regressions without controlling for other factors.

The results are reported in Table 3. In the odd-numbered columns, positive coefficients of indicator variables showing whether the individual and her partner have the same blood type offer evidence for assortative matching on blood type. After incorporating control variables, as reported in even-numbered columns, the magnitudes of indicator variables' coefficients decline considerably, and most control variables show statistical significance, which validates that estimates of assortative mating can be biased by confounding factors such as population stratification and relationship maintenance. Although the estimates of indicator variables' coefficients decrease by approximately 30% to 40% after incorporation of control variables, they are still statistically significant at the 1% level, indicating a highly nonrandom matching on blood type.

The coefficients of variables with statistical significance represent the extent to which the corresponding mechanism can explain the individual's blood type. One's blood type is mostly explained by the distribution of blood types in the population of her birthplace. A 1% rise in the share increases the odds ratio of having the corresponding blood type by approximately 4% to 5% for individuals with type A, B or O blood and by approximately 12% for those with AB-type blood. In addition to local population structure, the individual's ethnicity explains a considerable fraction of her blood type. A 1% increase in the proportion that her blood type takes up in her ethnicity increases the odds ratio of having the corresponding blood type by

approximately 1% to 2% for those with type A, B, or O blood and by approximately 5% for those with AB-type blood. The blood type of the individual's partner—our primary purpose—is also an effective predictor of her blood type, which provides evidence for assortative mating. If her partner is of a specific blood type, the odds ratio of having the same blood type will increase by approximately 8% for those with A-type blood, 7% for those with B-type blood, 15% for those with AB-type blood, and 20% for those with O-type blood (Table 3). More assortative matching is observed within spousal pairs with AB-type blood or O-type blood. The almost equal estimates in Panel A and Panel B in Table 3 suggest that the degrees of assortative mating are similar between females and males. We do not find strong evidence supporting the argument that the similarity of spousal pairs' blood types is associated with the length of marriage, which is used as a proxy for their relationship maintenance. Most of the coefficients of the interaction terms of length of marriage and the indicator variables of partner's blood type are insignificant or of small magnitudes lower than 0.01.

The odds ratio increase estimated by logistic regression model roughly indicates the degree of assortative mating on blood type among individuals. To estimate the increase in probability of matching among individuals attributed to blood type assortative mating in a more explicit way, we repeat the above regression analysis with linear regression model, and report the results in Table 4. Similar to the results obtained by logistic regression analysis, it can be seen that individuals within the same blood type have a significantly higher probability of matching with each other than those with different blood types. The results are robust after controlling for a group of control variables. If one's partner is of a specific blood type, the probability of having the same blood type will increase by approximately 1.6% for those with A-type blood, 1.3% for those with B-type blood, 1.2% for those with AB-type blood, and 4.4% for those with O-type blood (Table 4). Again, linear regression analysis provides clear evidence supporting assortative mating on blood type.

Table 2 Statistical tests on locally matched samples at the city level

City	Sample Size	Pearson's Chi-square statistic	Cramer's V	Adjusted Pearson Residuals			
				A, A	B, B	AB, AB	O, O
Yueyang, Hunan	34,327	753.586**	0.086	16.310**	8.813**	5.730**	25.151**
Nanyang, Henan	30,736	232.789**	0.050	1.721	-5.217**	-0.538	10.468**
Shaoyang, Hunan	23,342	133.264**	0.044	0.473	-4.324**	1.998	6.444**
Jingzhou, Hubei	19,017	286.661**	0.071	7.909**	7.206**	0.107	10.350**
Shangqiu, Henan	18,536	14.527	0.016	1.395	1.177	0.154	3.665**
Chenzhou, Hunan	17,355	138.999**	0.052	6.162**	2.783**	4.599**	8.680**
Yiyang, Hunan	17,164	8.005	0.012	1.154	0.726	-0.364	-0.431
Luoyang, Henan	16,941	60.553**	0.035	2.317	2.909**	1.551	7.105**
Baoding, Hebei	16,420	27.294**	0.024	3.251**	0.939	1.428	3.232**
Changde, Hunan	16,237	231.617**	0.069	-9.406**	-2.523**	0.987	6.269**
Hengyang, Hunan	15,661	147.266**	0.056	-3.139**	-1.543	6.404**	-0.189
Loudi, Hunan	15,194	62.474**	0.037	-2.298	-3.152**	2.571**	3.272**
Zhoukou, Henan	15,134	12.157	0.016	1.620	0.63	1.710	1.401
Yongzhou, Hunan	14,578	52.825**	0.035	1.702	3.947**	2.867**	5.553**
Zhumadian, Henan	13,080	41.007**	0.032	4.499**	0.788	2.459	4.438**
Xingtai, Hebei	11,869	268.322**	0.087	-5.061**	-5.720**	1.166	10.944**
Maoming, Guangdong	10,458	202.239**	0.080	6.774**	3.222**	4.778**	12.680**
Xuchang, Henan	10,014	41.188**	0.037	-0.187	3.825**	2.796**	3.977**

Note: ** Significant at the 5% level. The 5%-level critical value for Pearson's Chi-square test is 19.023 with a degree of freedom of 9. The 5%-level critical value for adjusted Pearson residuals is 2.498. It is corrected by Bonferroni correction (see Methods).

Table 3 Logistic regression analysis on the full sample

Panel A. Wife's mate choice.

Variable	I^{A}_{wife}		I^{B}_{wife}		I^{AB}_{wife}		I^{O}_{wife}	
$I^{A}_{Husband}$	0.104***	0.076***						
	(0.005)	(0.006)						
$I^{B}_{Husband}$			0.113***	0.066***				
			(0.005)	(0.006)				
$I^{AB}_{Husband}$					0.206***	0.147***		
					(0.012)	(0.014)		
$I^{O}_{Husband}$							0.273***	0.197***
							(0.005)	(0.005)
BS^{A}_{Wife}		4.559***						
		(0.058)						
BS^{B}_{Wife}				4.791***				
				(0.049)				
BS^{AB}_{Wife}						12.295***		
						(0.201)		
BS^{O}_{Wife}								4.222***
								(0.035)
EGS^{A}_{Wife}		1.870***						
		(0.175)						
EGS^{B}_{Wife}				0.715**				
				(0.315)				
EGS^{AB}_{Wife}						4.650***		
						(0.599)		
EGS^{O}_{Wife}								0.569***
								(0.096)
Length of Marriage		0.001		0.003***		0.001		-0.004***
		(0.001)		(0.001)		(0.001)		(0.001)
$I^{A}_{Husband}$ ×Length of Marriage		-0.001						
		(0.002)						
$I^{B}_{Husband}$ ×Length of Marriage				-0.002				
				(0.002)				
$I^{AB}_{Husband}$ ×Length of Marriage						0.008*		
						(0.004)		
$I^{O}_{Husband}$ ×Length of Marriage								0.004**
								(0.002)
Observation	967,329	966,764	967,329	966,764	967,329	966,764	967,329	966,764

Panel B. Husband's mate choice.

Variable	$I^{A}_{Husband}$		$I^{B}_{Husband}$		$I^{AB}_{Husband}$		$I^{O}_{Husband}$	
I^{A}_{Wife}	0.104***	0.076***						
	(0.005)	(0.006)						
I^{B}_{Wife}			0.113***	0.065***				
			(0.005)	(0.006)				
I^{AB}_{Wife}					0.206***	0.148***		
					(0.012)	(0.014)		
I^{O}_{Wife}							0.273***	0.195***
							(0.005)	(0.005)
$BS^{A}_{Husband}$		4.636***						
		(0.056)						
$BS^{B}_{Husband}$				4.882***				
				(0.047)				
$BS^{AB}_{Husband}$						11.743***		
						(0.170)		
$BS^{O}_{Husband}$								4.190***
								(0.035)
$EGS^{A}_{Husband}$		1.414***						
		(0.182)						
$EGS^{B}_{Husband}$				0.761***				
				(0.270)				
$EGS^{AB}_{Husband}$						4.728***		
						(0.634)		
$EGS^{O}_{Husband}$								0.474***
								(0.093)
Length of		0.001		-0.001		-0.002*		0.000
Marriage		(0.001)		(0.001)		(0.001)		(0.001)
I^{A}_{Wife} ×Length		-0.001						
of Marriage		(0.002)						
I^{B}_{Wife} ×Length				-0.001				
of Marriage				(0.002)				
I^{AB}_{Wife} ×Length						0.008**		
of Marriage						(0.004)		
I^{O}_{Wife} ×Length								0.005***
of Marriage								(0.002)
Observation	967,329	966,764	967,329	966,764	967,329	966,764	967,329	966,764

Note: Robust standard errors are in parentheses. * Significant at 10% level, ** 5% level, *** 1% level. $I^{g}_{Husband,i}$ and $I^{g}_{Wife,i}$ indicate whether the wife or the husband of couple i has type-g blood. $BS^{g}_{Wife,i}$ and $BS^{g}_{Husband,i}$ suggest the share that individuals with type-g blood in the population of the birthplace of the wife or the husband in couple i. $EGS^{g}_{Wife,i}$ and $EGS^{g}_{Husband,i}$ represent the share of individuals with type-g blood in the population of the ethnicity of the wife or the husband in couple i. $\text{Length of Marriage}_i$ indicates the time gap between the time when couple i gets married and the time when they receive a prepregnancy checkup. The same applies to tables below.

Table 4 Linear regression analysis on the full sample

Panel A. Wife's mate choice.

Variable	I^{A}_{wife}		I^{B}_{wife}		I^{AB}_{wife}		I^{O}_{wife}	
$I^{A}_{Husband}$	0.022***	0.016***						
	(0.001)	(0.001)						
$I^{B}_{Husband}$			0.023***	0.013***				
			(0.001)	(0.001)				
$I^{AB}_{Husband}$					0.018***	0.012***		
					(0.001)	(0.001)		
$I^{O}_{Husband}$							0.061***	0.044***
							(0.001)	(0.001)
BS^{A}_{Wife}		0.959***						
		(0.012)						
BS^{B}_{Wife}				0.983***				
				(0.010)				
BS^{AB}_{Wife}						0.961***		
						(0.016)		
BS^{O}_{Wife}								0.940***
								(0.008)
EGS^{A}_{Wife}		0.294***						
		(0.031)						
EGS^{B}_{Wife}				0.140**				
				(0.064)				
EGS^{AB}_{Wife}						0.238***		
						(0.037)		
EGS^{O}_{Wife}								0.164***
								(0.023)
Length of Marriage		0.000		0.001***		0.000		-0.001***
		(0.000)		(0.000)		(0.000)		(0.000)
$I^{A}_{Husband}$ ×Length of Marriage		-0.000						
		(0.000)						
$I^{B}_{Husband}$ ×Length of Marriage				-0.000				
				(0.000)				
$I^{AB}_{Husband}$ ×Length of Marriage						0.001**		
						(0.000)		
$I^{O}_{Husband}$ ×Length of Marriage								0.001*
								(0.000)
Observation	967,329	966,764	967,329	966,764	967,329	966,764	967,329	966,764

Panel B. Husband's mate choice.

Variable	$I^{A}_{Husband}$		$I^{B}_{Husband}$		$I^{AB}_{Husband}$		$I^{O}_{Husband}$	
I^{A}_{Wife}	0.022***	0.016***						
	(0.001)	(0.001)						
I^{B}_{Wife}			0.023***	0.013***				
			(0.001)	(0.001)				
I^{AB}_{Wife}					0.018***	0.012***		
					(0.001)	(0.001)		
I^{O}_{Wife}							0.062***	0.044***
							(0.001)	(0.001)
$BS^{A}_{Husband}$		0.970***						
		(0.012)						
$BS^{B}_{Husband}$				0.987***				
				(0.010)				
$BS^{AB}_{Husband}$						0.977***		
						(0.015)		
$BS^{O}_{Husband}$								0.945***
								(0.008)
$EGS^{A}_{Husband}$		0.207***						
		(0.032)						
$EGS^{B}_{Husband}$				0.139**				
				(0.054)				
$EGS^{AB}_{Husband}$						0.163***		
						(0.036)		
$EGS^{O}_{Husband}$								0.143***
								(0.022)
Length of Marriage		0.000		-0.000		-0.000**		0.000
		(0.000)		(0.000)		(0.000)		(0.000)
I^{A}_{Wife} ×Length of Marriage		-0.000						
		(0.000)						
I^{B}_{Wife} ×Length of Marriage				-0.000				
				(0.000)				
I^{AB}_{Wife} ×Length of Marriage						0.001**		
						(0.000)		
I^{O}_{Wife} ×Length of Marriage								0.001**
								(0.000)
Observation	967,329	966,764	967,329	966,764	967,329	966,764	967,329	966,764

Section 3. An evolutionary matching model

We can explain our observations of assortative matching using a simple evolutionary matching model.

Blood types are inherited from both parents. The ABO blood type is controlled by a single gene (the ABO gene) with three types of alleles inferred from classical genetics: i, iA, and iB. The iA allele gives type A, iB gives type B, and i gives type O. As both iA and iB are dominant over i, only ii people have type O blood. Individuals with iAiA or iAi have type A blood, and individuals with iBiB or iBi have type B blood. iAiB people have both phenotypes because A and B express a special dominance relationship: codominance, which means that type A and B parents can have an AB child. A couple with type A and type B can also have a type O child if they are both heterozygous (iBi, iAi).

Suppose a person would like to have a child with the same blood type as herself and will gain 1 unit of utility (or evolutionary gain) if she has a child sharing the same blood type with her. Then, we have the payoff matrix of matches (Table 5, see Methods for details of how these numbers are arrived at).

If the utilities of spousal pairs are additive, then the payoff matrix is further simplified as shown in Table 6.

When AA and AO (and BB and BO) appear with the same frequency, we reduce the payoffs of matches to the matrices shown in Table 7.

We see that the diagonal terms produce the highest payoffs. This is a simple theory showing how evolutionary motives result in positive assortative matching. In particular, this theory helps rationalize our observations: (i) Assortative matching is more likely; (ii) (O, O) assortative matching is the most likely.

Table 5 Payoff matrix of matches on pair of genes

	AA	AO	BB	BO	AB	OO
AA	1, 1	1, 1	0.5, 0.5	0.5, 0	0.5, 0.5	1, 0
AO	1, 1	0.75, 0.75	0, 0.5	0.25, 0.25	0.5, 0.25	0.5, 0.5
BB	0.5, 0.5	0.5, 0	1, 1	1, 1	0.5, 0.5	1, 0
BO	0, 0.5	0.25, 0.25	1, 1	0.75, 0.75	0.5, 0.25	0.5, 0.5
AB	0.5, 0.5	0.25, 0.5	0.5, 0.5	0.25, 0.5	0.5, 0.5	0, 0
OO	0, 1	0.5, 0.5	0, 1	0.5, 1	0, 0	1, 1

Table 6 Payoff matrix of matches on pairs of genes when utilities are additive

	AA	AO	BB	BO	AB	OO
AA	2	2	1	0.5	1	1
AO	2	1.5	0.5	0.5	0.75	1
BB	1	0.5	2	2	1	1
BO	0.5	0.5	2	1.5	0.75	1
AB	1	0.75	1	0.75	0.75	0
OO	1	1	1	1.5	0	2

Table 7 Payoff matrix of matches on blood type

Panel A. Nontransferable utilities

	A	B	AB	O
A	0.9375, 0.9375	0.3125, 0.3125	0.5, 0.375	0.75, 0.25
B	0.3125, 0.3125	0.9375, 0.9375	0.5, 0.375	0.75, 0.25
AB	0.375, 0.5	0.375, 0.5	0.5, 0.5	0, 0
O	0.5, 0.5	0.5, 0.5	0, 0	1, 1

Panel B. Transferable utilities

	A	B	AB	O
A	**1.875**	0.625	0.875	1
B	0.625	**1.875**	0.875	1
AB	0.875	0.875	**1**	0
O	1	1	0	**2**

Section 4. Conclusion

In summary, we provide evidence of assortative mating on blood type. The degree of assortative mating varies among individuals with different blood types. The findings are robust after we control for the effect of other possible mechanisms, such as environmental confounding and relationship maintenance. To explain these observations, we develop a simple evolutionary model. The theoretical prediction of this model fits the findings we gain from the data.

Our study makes two contributions. First, our empirical and theoretical results shed light on nonrandom matching on blood type, one of the most well-known human phenotypes, which has not been fully investigated to the best of our knowledge. Assortative mating on blood type will have important genetic consequences by influencing the direction of evolution of blood type distribution in the population, making it interesting to investigate this issue.

Second, we improve the causal inference of assortative mating by using a group of approaches. To mitigate the estimation bias caused by population stratification, we restrict our analysis to locally matched subsamples to perform meta-analysis. We further address this concern by running regressions with control variables. Other possible mechanisms, such as relationship maintenance, are also controlled for in our analysis. The robust results show causal evidence for assortative mating on blood type.

We acknowledge the limitations of our study. First, although overall evidence for assortative mating was found and a possible theoretical explanation for the findings was developed, we did not investigate the specific mechanisms through which blood types influence human mate choice from an empirical perspective due to data availability. It is of interest to understand the underlying mechanisms in future research. Second, we should be cautious when extrapolating the findings in the Chinese population to other populations. Further evidence is needed to investigate whether the nonrandom matching pattern of blood types is robust among other populations.

It is difficult to infer a causal relationship between blood type similarity and mate choice, as spousal concordance is affected by many confounding factors[6,20,22-31]. In addition to assortative mate choice, mate similarity may be induced by alternative mechanisms, including social factors such as population stratification and relationship maintenance after spousal pairing. When latent subpopulation structure underlies the observed sample, the positive association of blood type within spousal pairs can be attributed to systematic differences among subpopulations arising from the ancestry differences they inherit. Under this circumstance, assortative mating will thus be overestimated if homogeneous sampling is assumed. An alternative explanation for the similarity of blood types within spousal pairs is that blood type concordance helps spousal pairs maintain their relationship. Although empirical evidence is still unclear, some research argues that there is a certain association between blood type and personality[32]. Partners with the same blood type might share a similar personality and thus find it easier to get along together and maintain their relationship, which increases their probability of being enrolled in the study and leads to biased estimation. All three possible mechanisms may result in spousal concordance on blood type (Figure 1). To study assortative mating on blood type, we need to carefully distinguish this mechanism from the other two possible explanations.

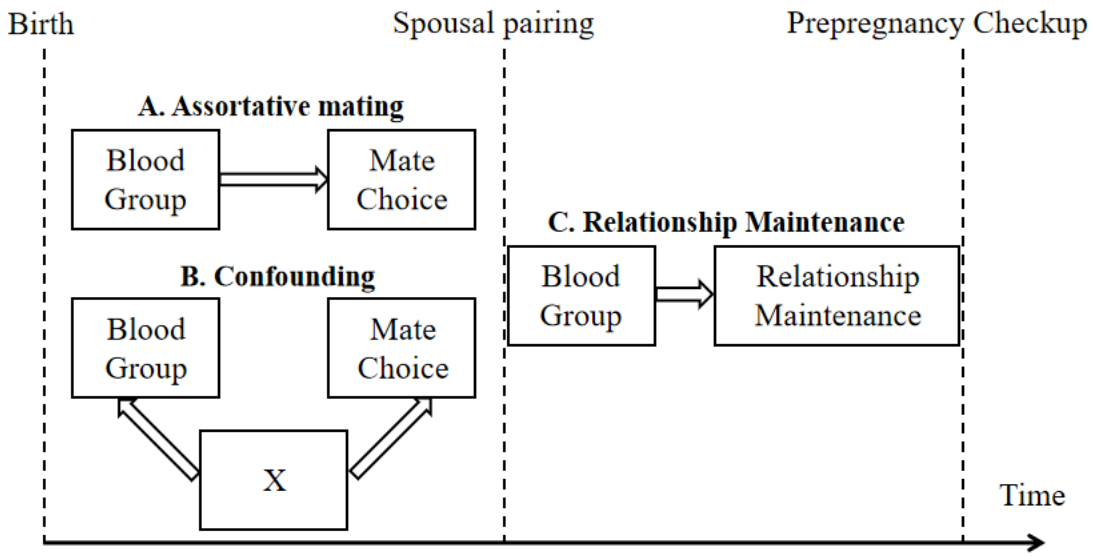

Fig. 1 Possible mechanisms of spousal concordance on blood type. Mechanism A (Assortative mating) refers to the process in which partners choose each other based on their similarity in blood type. Mechanism B (Confounding) indicates confounding factors that may influence both blood type and mate choice. Population stratification is the most common confounding factor. Mechanism C (Relationship Maintenance) shows the possibility that spousal pairs with the same blood type are more likely to maintain their long-term relationship and be enrolled in the observed sample.

Section 5. Methods

Section 5A. Data

Chinese prepregnancy checkup data.

In this study, we used 2014-2015 Chinese prepregnancy checkup data to perform statistical analysis. This dataset is collected from the National Free Preconception Health Examination Project (NFPHEP), a program launched by the National Health Commission and the Ministry of Finance of the People's Republic of China offering free prepregnancy examinations for married couples. Blood samples are collected during the examination and are sent to local laboratories for blood type testing. In addition to the results of physical examinations, the NFPHEP also collected all participants' basic personal information, including sex, address of residence, ethnicity, birthplace, and marriage, via a standardized questionnaire as well as participants' identification cards[36-38]. Informed consent was signed by all project participants before enrollment. A total of 1,137,010 couples from 31 provinces of China are covered in the dataset. This study was approved by Institute of Science and Technology, the National Health Commission of the People's Republic of China. Written informed consent forms were obtained from all participants before enrolment.

Subsample with complete personal information.

We remove observations with incomplete information related to the couple's blood types, living areas, birthplaces, ethnicity, and marriage, which is utilized in our statistical analysis, and obtain a sample of 967,329 couples as the full sample of our study. The raw dataset of Chinese prepregnancy checkup data was filtered in Stata.

Section 5B. Statistical analysis

Contingency table.
We broke down the numbers of matches among individuals with different blood types by sex to produce a 4×4 contingency table, with the number in grid (i, j) $N_{i,j}$ representing the frequencies of matching between males with type-i blood and females with type-j blood. The ratios of observed frequencies over expected frequencies are also reported, with gird (i, j) computed as $N_{i,j}/E_{i,j} = N_{i,j}/(R_i \times C_j/N)$, where R_i and C_j denote row and column marginal totals and N represents the grand total. The contingency table was conducted using R.

Pearson's Chi-square test.
We performed a chi-square test on a contingency table of blood types obtained from the full sample to test the association between an individual's blood type and her partner's blood type, with a null hypothesis that matching behavior is independent of spousal pairs' blood types. The 5%-level critical value for Pearson's Chi-square test is 19.023 with a degree of freedom of 9. Cramer's V was also reported to measure the effect size for the chi-square test. The Pearson's chi-square test was performed using R.

Adjusted Pearson residuals.
The adjusted Pearson residual of grid (i, j) is computed as

$$\tilde{r}_{i,j} = \frac{N_{i,j} - E_{i,j}}{\sqrt{E_{i,j}(1 - R_i/N)(1 - C_j/N)}} \quad (1)$$

It follows a standard normal distribution $N(0,1)$ in the null hypothesis of random matching, which enables us to perform statistical tests on nonrandom mating. The 5%-level critical value for adjusted Pearson residuals is ±2.498. It is corrected by Bonferroni correction. To test the 4 diagonal elements in the contingency table, the new alpha level should be set as $\alpha_{Bon} = \frac{5\%}{4} = 1.25\%$, and the corresponding critical value $N(0,1)_{1-\alpha_{Bon}/2} = 2.498$. Computation of adjusted Pearson residuals and related statistical tests were performed using R.

The Altham index.
We use the Altham index as an alternative measure of assortative mating. It can be written as

$$\text{The Altham index} = [\sum_{i \in G}\sum_{j \in G}\sum_{k \in G}\sum_{l \in G} |\log(\frac{P(I^i_{Wife} = 1, I^j_{Husband} = 1)P(I^k_{Wife} = 1, I^l_{Husband} = 1)}{P(I^i_{Wife} = 1, I^l_{Husband} = 1)P(I^k_{Wife} = 1, I^j_{Husband} = 1)})|^2]^{1/2},$$
$$G = \{A, B, AB, O\} \quad (2)$$

where $I^g_{Husband}$ and I^g_{Wife} refer to indicator variables showing whether the wife or the husband has type-g blood, and $P(I^g_{Wife} = 1, I^h_{Husband} = 1)$ represents the probability of matching between females with type-g blood and males with type-h blood, which is estimated

by the share this type of matching takes up in the full sample. The higher the index is, the more nonrandom the mate choice is. It is equal to zero when matching is random. The Altham index was computed using R.

Meta-analysis.
As a sensitivity analysis, we restricted samples for Pearson's chi-square test to locally matched couples to prevent our test on assortative mating from being confounded by population stratification. Specifically, we cover couples who were born in the same birthplace and received prepregnancy checkups in this area and stratify them by their birthplace. Meta-analysis is performed both at the city level, which allows more granular segmentation for subpopulations. Full sample was filtered in Stata to obtain locally matched subsamples. Meta-analysis on locally matched subsamples was performed using R.

Regression Analysis.
We investigated the degree of assortative mating on blood type with logistic regression models. The models are specified as equations (3) to (6), where $p^{g}_{Wife,i}$ and $p^{g}_{Husband,i}$ indicate the probability that the wife or the husband of couple i has type-g blood. $I^{g}_{Husband,i}$ and $I^{g}_{Wife,i}$ refer to indicator variables that show whether the wife or the husband of couple i has type-g blood. $BS^{g}_{Wife,i}$ and $BS^{g}_{Husband,i}$ suggest the share that individuals with type-g blood take up in the population of the birthplace of the wife or the husband in couple i, which is estimated by the population in this city covered in our full sample. $EGS^{g}_{Wife,i}$ and $EGS^{g}_{Husband,i}$ represent the share that individuals with type-g blood take up in the population of the ethnicity of the wife or the husband in couple i, which is estimated by the population in this ethnicity covered in our full sample. Finally, $\text{Length of Marriage}_i$ indicates the time gap between the time when couple i get married and the time when they receive a prepregnancy checkup. $I^{g}_{Wife,i} \times \text{Length of Marriage}_i$ and $I^{g}_{Husband,i} \times \text{Length of Marriage}_i$ are interaction terms of the indicator variables and the length of marriage of this couple. For all models, we adopt robust standard errors in the model estimation to ensure that our estimates are robust when the model specification is incorrect. We also repeated the above steps using linear regression model, to explicitly estimate the increase in probability of matching induced by assortative mating on blood type. Regression analysis was performed using Stata.

$$\ln\left(\frac{p^{g}_{Wife,i}}{1-p^{g}_{Wife,i}}\right) = \beta_0 + \beta_1 I^{g}_{Husband,i}, \qquad g \in \{A, B, AB, O\} \tag{3}$$

$$\ln\left(\frac{p^{g}_{Wife,i}}{1-p^{g}_{Wife,i}}\right) = \beta_0 + \beta_1 I^{g}_{Husband,i} + \beta_2 BS^{g}_{Wife,i} + \beta_3 EGS^{g}_{Wife,i} + \beta_4 \text{Length of Marriage}_i + \beta_5 I^{g}_{Husband,i} \times \text{Length of Marriage}_i, \qquad g \in \{A, B, AB, O\} \tag{4}$$

$$\ln\left(\frac{p^{g}_{Husband,i}}{1-p^{g}_{Husband,i}}\right) = \beta_0 + \beta_1 I^{g}_{Wife,i}, \qquad g \in \{A, B, AB, O\} \tag{5}$$

$$\ln\left(\frac{p^{g}_{Husband,i}}{1-p^{g}_{Husband,i}}\right)$$

$$= \beta_0 + \beta_1 I^{g}_{Wife,i} + \beta_2 BS^{g}_{Husband,i} + \beta_3 EGS^{g}_{Husband,i} + \beta_4 \text{Length of Marriage}_i$$

$$+$$

$$\beta_5 I^{g}_{Wife,i} \times \text{Length of Marriage}_i, \qquad g \in \{A, B, AB, O\} \tag{6}$$

Section 5C. Model

Model specification.

Suppose a person would like to have a child with the same blood type as herself, and she will gain 1 unit of utility from having a child with the same blood type.

Payoffs of matches of pairs of genes.

Given the model specification, we are able to derive the payoffs of matches of pairs of genes by computing the probability of having a child with the same blood for the wife and the husband. For instance, suppose there is (AB, AB) pairing.

	A	B
A	AA	AB
B	BA	BB

Therefore, ½ AB + ¼ AA + ¼ BB; AB parents pass on their type with a probability of ½, both of which therefore gain 0.5 units of utility from the matching.

Payoff matrix of matches.

We obtain the following matrix by computing the payoffs of matching different pairs of genes in turn.

	AA	AO	BB	BO	AB	OO
AA	1, 1	1, 1	0.5, 0.5	0.5, 0	0.5, 0.5	1, 0
AO	1, 1	0.75, 0.75	0, 0.5	0.25, 0.25	0.5, 0.25	0.5, 0.5
BB	0.5, 0.5	0.5, 0	1, 1	1, 1	0.5, 0.5	1, 0
BO	0, 0.5	0.25, 0.25	1, 1	0.75, 0.75	0.5, 0.25	0.5, 0.5
AB	0.5, 0.5	0.25, 0.5	0.5, 0.5	0.25, 0.5	0.5, 0.5	0, 0
OO	0, 1	0.5, 0.5	0, 1	0.5, 1	0, 0	1, 1

By assuming that AA and AO (and BB and BO) appear with the same frequency, we obtain the reduced form of the payoff matrix, which indicates the utilities that spousal pairs with different blood types gain from matching.

	A	B	AB	O
A	0.9375, 0.9375	0.3125, 0.3125	0.5, 0.375	0.75, 0.25

B	0.3125, 0.3125	0.9375, 0.9375	0.5, 0.375	0.75, 0.25
AB	0.375, 0.5	0.375, 0.5	0.5, 0.5	0, 0
O	0.5, 0.5	0.5, 0.5	0, 0	1, 1

These payoff matrices can be further simplified when we assume that the utilities of spousal pairs are additive.

Section 5D. Data availability

We utilize 2014-2015 Chinese prepregnancy checkup data data available from Institute of Science and Technology, the National Health Commission of the People's Republic of China, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Institute of Science and Technology, the National Health Commission of the People's Republic of China.

Section 5E. Code availability

All analyses performed in this study used pre-existing packages mentioned in Section 5B. No custom code was generated.

References

1. Vandenberg, S. G. Assortative mating, or who marries whom? *Behavior Genetics* 2, 127–157 (1972).
2. Eaves, L. J. & Heath, A. C. Detection of the effects of asymmetric assortative mating. *Nature* 289, 205–206 (1981).
3. Buss, D. M. Human mate selection: opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable. *American Scientist* 73, 47–51 (1985).
4. Heath, A. C. et al. No decline in assortative mating for educational level. *Behavior Genetics* 15, 349–369 (1985).
5. Heath, A. C. & Eaves, L. J. Resolving the effects of phenotype and social background on mate selection. *Behavior Genetics* 15, 15–30 (1985).
6. Heath, A. C., Eaves, L. J., Nance, W. E. & Corey, L. A. Social inequality and assortative mating: Cause or consequence? *Behavior Genetics* 17, 9–17 (1987).
7. Mare, R. D. Five decades of educational assortative mating. *American Sociological Review*. 56, 15–32 (1991).
8. Krueger, R. F., Moffitt, T. E., Caspi, A., Bleske, A. & Silva, P. A. Assortative mating for antisocial behavior: developmental and methodological implications. *Behavior Genetics* 28, 173–186 (1998).
9. Eaves, L. et al. Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Research and Human Genetics* 2, 62–80 (1999).
10. Silventoinen, K., Kaprio, J., Lahelma, E., Viken, R. J. & Rose, R. J. Assortative mating by body height and BMI: Finnish twins and their spouses. *American Journal of Human Biology* 15, 620–627 (2003).
11. Watson, D. et al. Match makers and deal breakers: analyses of assortative mating in newlywed couples. *Journal of Personality* 72, 1029–1068 (2004).
12. Kirkpatrick, M. Reinforcement and divergence under assortative mating. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 267, 1649–1655 (2000).
13. Jiang, Y., Bolnick, D. I., & Kirkpatrick, M. Assortative mating in animals. *The American Naturalist* 181, E125–E138 (2013).
14. Stulp, G., Simons, M. J., Grasman, S., & Pollet, T. V. Assortative mating for human height: A meta-analysis. *American Journal of Human Biology* 29, e22917 (2017).
15. Eaves, L. J. Assortative mating and intelligence: An analysis of pedigree data. *Heredity* 30, 199–210 (1973).
16. Eaves, L. J. & Heath, A. C. Detection of the effects of asymmetric assortative mating. *Nature* 289, 205–206 (1981).
17. Maes, H., Neale, M. & Eaves, L. Genetic and environmental factors in relative body weight and human adiposity. *Behavior Genetics* 27, 325–351 (1997).
18. Zietsch, B. P., Verweij, K. J. H., Heath, A. C. & Martin, N. G. Variation in human mate choice: simultaneously investigating heritability, parental influence, sexual imprinting, and assortative mating. *The American Naturalist* 177, 605–616 (2011).

19. Keller, M. C. et al. The genetic correlation between height and IQ: Shared genes or assortative mating? *PLoS Genetics* 9, e1003451 (2013).
20. Robinson, M. R., et al. Genetic evidence of assortative mating in humans. *Nature Human Behaviour* 1, 1–13 (2017).
21. Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten* 27: 357–362 (1900).
22. Falconer, D. S. Introduction to Quantitative Genetics (Ronald, 1960).
23. Nance, W. E., Corey, L. A., Rose, R. J. & Eaves, L. J. Relevance of the marriages of twins to the causal analysis of nonrandom mating. *Progress in Clinical and Biological Research* 69, 61–71 (1981).
24. Abdellaoui, A. et al. Association between autozygosity and major depression: stratification due to religious assortment. *Behavior Genetics* 43, 455–467 (2013).
25. Abdellaoui, A., Verweij, K. J. H. & Zietsch, B. P. No evidence for genetic assortative mating beyond that due to population stratification. *Proceedings of the National Academy of Sciences* 111, E4137 (2014).
26. Abdellaoui, A. et al. Educational attainment influences levels of homozygosity through migration and assortative mating. *PLoS One* 10, e0118935 (2015).
27. Tenesa, A., Rawlik, K., Navarro, P., & Canela-Xandri, O. Genetic determination of height-mediated mate choice. *Genome Biology* 16(1), 1–8 (2015).
28. Howe, L. J. et al. Genetic evidence for assortative mating on alcohol consumption in the UK Biobank. *Nature Communications* 10, 1–10 (2019).
29. Richardson, K., & Jones, M. C. Why genome-wide associations with cognitive ability measures are probably spurious. *New Ideas in Psychology* 55, 35–41 (2019).
30. Morris, T. T., Davies, N. M., Hemani, G., & Smith, G. D. Population phenomena inflate genetic associations of complex social traits. *Science Advances* 6, eaay0328 (2020).
31. Yengo, L., Sidari, M., Verweij, K. J., Visscher, P. M., Keller, M. C., & Zietsch, B. P. No evidence for social genetic effects or genetic similarity among friends beyond that due to population stratification: a reappraisal of Domingue et al (2018). *Behavior Genetics* 50, 67–71 (2020).
32. Tsuchimine, S., Saruwatari, J., Kaneda, A., & Yasui-Furukori, N. ABO blood type and personality traits in healthy Japanese subjects. *PLoS One* 10, e0126983 (2015).
33. Altham, P. M. The Measurement of Association of Rows and Columns for an r× s Contingency Table. *Journal of the Royal Statistical Society: Series B (Methodological)* 32, 63–73 (1970).
34. Long, J., & Ferrie, J. Intergenerational Occupational Mobility in Great Britain and the United States Since 1850. *The American Economic Review* 103, 1109–1137 (2013).
35. Eika, L., Mogstad, M., & Zafar, B. Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy* 127, 2795–2835 (2019).
36. Liu, J. et al. Seroepidemiology of hepatitis B virus infection in 2 million men aged 21–49 years in rural China: a population-based, cross-sectional study. *The Lancet Infectious Diseases* 16, 80-86 (2016).

37. Liu, J. et al. Frequencies and ethnic distribution of ABO and RhD blood types in China: a population-based cross-sectional study. *BMJ Open* 7, e018476 (2017).
38. Liu, J. et al. Maternal pre-pregnancy infection with hepatitis B virus and the risk of preterm birth: a population-based cohort study. *The Lancet Global Health* 5, e624-e632 (2017).

Competing interests. The authors declare no competing interests.