



# A decade of research on the genetics of entrepreneurship: a review and view ahead

Cornelius A. Rietveld · Eric A.W. Slob · A. Roy Thurik

Accepted: 1 April 2020 / Published online: 28 April 2020  
© The Author(s) 2020

**Abstract** Studies analyzing the heritability of entrepreneurship indicate that explanations for why people engage in entrepreneurship that ignore genes are incomplete. However, despite promises that were solidly backed up with ex ante power calculations, attempts to identify specific genetic variants underlying the heritable variation in entrepreneurship have until now been unsuccessful. We describe the methodological issues hampering the identification of associations between genetic variants and entrepreneurship, but we also outline why this search will eventually be successful. Nevertheless, we argue that the benefits of using these individual genetic variants for empirical research in the entrepreneurship domain are likely to be small. Instead, the use of summary indices comprising multiple genetic variants, so-called polygenic risk scores, is advocated. In doing so, we stress the caveats associated with applying population-level results to the individual level. By drawing upon the promises of “genoeconomics,” we

sketch how the use of genetic information may advance the field of entrepreneurship research.

**Keywords** Entrepreneurship · Genetics · Polygenic risk scores

**JEL classification** L26 · D01

## 1 Introduction

In 2000, the field of psychology concluded the nature-nurture debate to be “over” by posing that all human behavioral traits are heritable (Turkheimer 2000). This “first law” of behavior genetics is backed by a vast body of literature comprising thousands of heritability studies (Polderman et al. 2015; Turkheimer 2000). Since 2008, several studies have shown that this law also holds for entrepreneurship (Nicolaou et al. 2008a, b, 2010; Shane and Nicolaou 2015; Van der Loos et al. 2013; Zhang et al. 2009). Inspired by these findings and advances in genetics research, Koellinger et al. (2010) provided a sketchy forecast in this journal of the expected identification of relationships between genetic variants and entrepreneurship. Nevertheless, despite several attempts in the past decade (Nicolaou et al. 2011; Quaye et al. 2012; Van der Loos et al. 2011, 2013; Wernert et al. 2012), no single robust association between a genetic variant and entrepreneurship has been found. Therefore, the first question we address in the present study is “*Why has the identification of robust associations between genetic variants and entrepreneurship been*

---

C. A. Rietveld (✉) · E. A. Slob · A. R. Thurik  
Department of Applied Economics, Erasmus School of  
Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000  
DR Rotterdam, The Netherlands  
e-mail: nrrietveld@ese.eur.nl

C. A. Rietveld · E. A. Slob · A. R. Thurik  
Erasmus University Rotterdam Institute for Behavior and Biology  
(EURIBEB), Erasmus University Rotterdam, Rotterdam,  
The Netherlands

A. R. Thurik  
Montpellier Business School, Montpellier, France

*unsuccessful in the last decade?*” We answer this question from a methodological point of view. In doing so, we also provide a review of the literature in this field of research.

The second question we address is “*Would the identification of associations between genetic variants and entrepreneurship help to advance the field of entrepreneurship research?*” Despite the unsuccessful attempts so far, we provide methodological and empirical reasons for why we may expect the identification of the first robust associations between genetic variants and entrepreneurship in the not too distant future. Entrepreneurship scholars have argued that the prediction of entrepreneurial behavior using genetic data could have practical applications in business and for individual decision-making (Nicolaou et al. 2008a; Nicolaou and Shane 2010; Shane 2010). Moreover, several private companies already offer genetic tests to predict someone’s leadership and managerial qualities.<sup>1</sup> We explain how summary indices of genetic variants (so-called polygenic risk scores) can be used for such prediction analyses, but by drawing on the broader behavior genetics literature, we stress the caveats associated with applying population-level results to the individual level. By relating the promises of “genoeconomics” as outlined by Benjamin et al. (2012a) to entrepreneurship research, we then sketch how we think the use of genetic information may advance the field of entrepreneurship research.

To illustrate the answers to our two research questions, we include an empirical analysis of data from the US Health and Retirement Study. The inclusion of the empirical analyses in this study serves three purposes. First, the results of the analyses show how polygenic risk scores constructed for a range of traits (and not just entrepreneurship) can help to identify regions in the human genome particularly important for entrepreneurial behavior. Second, these analyses illustrate how polygenic risk scores can significantly predict entrepreneurship (even when proxied by the relatively episodic activity of self-employment). Third, we use these analyses to illustrate that the estimated relationships between polygenic risk scores and entrepreneurship at the population level only marginally improve the prediction of entrepreneurial behavior at the individual level.

<sup>1</sup> For example, such tests are provided by Leadership Consultants (<https://leapership.com/shop/karmagene-dna-based-personality-test/>) and Goldman Genetics (<https://goldmen.eu/>).

In the following section, we review the studies providing evidence for the heritability of entrepreneurship. By exploiting family-based relationships rather than molecular genetic information, these studies show that approximately 40% of the differences in entrepreneurial behavior can be explained by genes. In Section 3, we review the molecular genetic analyses of entrepreneurship. We provide a comprehensive overview and discussion of the methodological approaches taken to identify relationships between genetic variants and entrepreneurship. Our empirical analyses are introduced and presented in Section 4. Finally, Section 5 concludes by discussing the added value of genetics for entrepreneurship research.

## 2 The heritability of entrepreneurship<sup>2</sup>

Heritability is a technical term denoting the proportion of observed differences in a trait among individuals from a certain population that is due to the genetic differences among these individuals (Visscher et al. 2008). The main challenge in the estimation of heritability is the statistical separation of the effect of genes from the effect of the family environment on the trait of interest. One way to address this challenge is to compare adoptees with biological children. Using this approach, Lindquist et al. (2015) find that parental entrepreneurship increases the likelihood of children’s entrepreneurship by 60%. In their Swedish sample, they show that post-birth factors (i.e., adoptive parents) are two times more important than pre-birth factors (i.e., biological parents) for explaining entrepreneurial involvement.

Another, more common approach to separating the effect of genes from the effect of the family environment is the comparison of monozygotic and dizygotic twins reared together because the number of available twin samples is much larger than the available samples of adoptees (Knopik et al. 2016). Monozygotic twins are genetically identical; however, dizygotic twins are as genetically similar to each other as regular siblings. Under the assumption that monozygotic and dizygotic twins are influenced by their family environment to the

<sup>2</sup> Nofal et al. (2018) provide a review of the literature about “biology and management.” Studies analyzing entrepreneurship are also included in this overview. All studies related to entrepreneurship in their category “Quantitative genetics” are discussed in this section (besides other studies). All entrepreneurship studies in their category “Molecular genetics” are discussed in Section 3 (again, besides other studies).

same extent, it is possible to decompose the variance in a trait into three components: the additive genetic effect, the common environment (family specific) effect, and the unique (individual specific) environment effect. Nicolaou et al. (2008a, b, 2010), Shane and Nicolaou (2015), Van der Loos et al. (2013), and Zhang et al. (2009) use the classical twin study methodology to estimate the heritability of entrepreneurship in American, British, and Swedish samples. These studies draw on a broad range of empirical measures for entrepreneurship, such as self-employment and the number of start-up efforts, and provide general support for the heritability of entrepreneurship. Overall, the heritability estimates are in the neighborhood of 40%, indicating that almost one-half of the differences in entrepreneurship in these countries can be attributed to genetic differences across population members.<sup>3</sup>

Although adoptee and twin studies can establish that genetic factors account for variation in a trait, they do not identify specific genes or the biological pathways through which genes function, because the genetic component is inferred from family relationships rather than observed in these studies. The completion of the sequencing of the human genome at the beginning of the present century (Venter et al. 2001) enabled the identification and measurement of locations in the human genome that differ among population members and hence led to the search for the specific genes underlying the heritable variation in entrepreneurship.

### 3 The molecular genetic analysis of entrepreneurship

#### 3.1 The human genome

A complete human genome consists of 23 pairs of chromosomes, from which the 23rd pair determines the biological sex of an individual. One of each pair of chromosomes is inherited from the mother, and the other is inherited from the father. A chromosome is composed of two intertwined strands of deoxyribonucleic acid (DNA), each made up of a sequence of nucleotide molecules. There are four different nucleotide molecules in the DNA: adenine, cytosine, thymine, and guanine.

<sup>3</sup> Nicolaou et al. (2009) use an extended version of the classical twin study to show that the genes influencing the tendency to be an entrepreneur and the genes influencing opportunity recognition partially overlap.

Adenine on one strand is always paired with thymine on the other strand, and cytosine is always paired with guanine. These combinations are called base pairs. Every human genome consists of approximately 3 billion base pairs. The stretches of base pairs in the DNA coding of a protein are called genes. There are approximately 20,000 genes in the human genome with varying lengths.

A random pair of individuals shares approximately 99.9% of their DNA (National Human Genome Research Institute 2018b), and most genetic differences across population members can be attributed to single nucleotide polymorphisms (SNPs, pronounced “snips”). Therefore, behavioral genetics researchers focus primarily on SNPs when analyzing heritable genetic variation. A SNP is defined as a location in the DNA strand at which two different nucleotides are present in the population. Each of the two possible nucleotides is called an allele for that SNP. The allele that is least common in the population is called the minor allele; the other allele is called the major allele. For each SNP, an individual’s genotype is coded as 0, 1, or 2, depending on the number of minor alleles present. Individuals who inherited the same allele from each parent are called homozygous for that SNP (and have genotype 0 or 2), while individuals who inherited different alleles are called heterozygous (and have genotype 1). SNPs can be found in every part of the genome, within genes or in regions in between genes, and may influence the production of proteins.

In the human genome, there are approximately 85 million SNPs with a minor allele prevalence of at least 1% (The 1000 Genomes Project Consortium 2015). When relating so many SNPs  $x_{ij}$  (coded as 0, 1, or 2) to a specific outcome  $y_i$  in a regression framework such as

$$y_i = \mu + \sum_{j=1}^J \beta_j x_{ij} + \varepsilon_i,$$

with intercept  $\mu$ , SNP effects  $\beta_j$ , and residual term  $\varepsilon_i$ , it is evident that we have to deal with an overidentified model with fewer individuals  $I$  than SNPs  $J$  (Benjamin et al. 2012a).<sup>4</sup> For this purpose, two basic approaches

<sup>4</sup> Advanced statistical methods, such as GREML (genome-based restricted maximum likelihood), use two-step procedures to jointly estimate the explained variance of all SNPs (Yang et al. 2010). With this method, Van der Loos et al. (2013) show that all SNPs in their sample explain 25% of the variance in entrepreneurship. However, such approaches do not identify which individual SNPs are associated with the outcome variable.

have been developed to deal with the overidentification problem. Hypothesis-driven methods such as the candidate gene approach do not consider all  $J$  SNPs, and hypothesis-free methods such as the genome-wide association study (GWAS) consider all  $J$  SNPs but not in one model. We continue by discussing these two basic approaches from a methodological point of view, and we review how they have been used for unravelling the genetic architecture of entrepreneurship.

### 3.2 Hypothesis-driven approaches

The candidate gene approach consists of testing a subset of genetic variants for association with the outcome of interest. These genetic variants are selected based on what is known or believed about their biological function (Benjamin et al. 2012a, b; Ebstein et al. 2010; Nicolaou and Shane 2009). This approach resembles the classic way of justifying and then testing a hypothesis. A clear advantage of this approach is that the interpretation of revealed significant relationships is relatively straightforward. Adopting this approach, Nicolaou et al. (2011) were the first to report an association between a SNP in the *DRD3* gene (a dopamine receptor gene) and entrepreneurial behavior in a British sample. Their selection of candidate SNPs was based on the observation that dopamine receptor genes have been associated with novelty seeking/sensation seeking and attention deficit hyperactivity disorder (ADHD). These traits were reported to be particularly prevalent among entrepreneurs (Nicolaou et al. 2008b; Antshel 2017). Unfortunately, Van der Loos et al. (2011) failed to replicate this association in a Dutch sample seven times larger than the sample Nicolaou et al. (2011) drew upon.

This non-replication is exemplary for candidate gene studies (Benjamin et al. 2012a, b; Ioannidis 2005; Rietveld et al. 2014a). In principle, a theoretical framework guides empirical research in reducing the number of hypotheses being tested. However, the analytical rigor that a theory-guided approach provides is not helpful in the context of behavioral genetics because it is difficult to reduce the number of plausible hypotheses purely on theoretical grounds. For instance, 70% of all genes (thus approximately 14,000) are expressed in the brain (Ramsköld et al. 2009), and for many of these genes

(and hence the SNPs within these genes), a seemingly plausible relation between genes and behavior—including entrepreneurship—could be hypothesized ex ante. As a matter of fact, in 2012, the editor of the leading field journal *Behavior Genetics* issued an editorial policy on candidate gene studies of behavioral traits that reads “The literature on candidate gene associations is full of reports that have not stood up to rigorous replication” and went on to say “...it now seems likely that many of the published findings of the last decade are wrong or misleading and have not contributed to real advances in knowledge” (Hewitt 2012). This editorial policy outlines strict quality criteria that candidate gene studies must meet to be considered for publication. Most importantly, the editors stressed the importance of sufficient statistical power in genetic discovery studies (Hewitt 2012).

Statistical power refers to the probability of rejecting the null hypothesis when it is not true. Statistical power of 80% or higher is generally considered to be adequate (Ellis 2010). Low statistical power results in a high chance of false negatives, i.e., non-rejections of the null hypothesis when the alternative hypothesis is true. Even more problematic, because of the winner’s curse, low statistical power also results in the overestimation of effect sizes for significant findings (Benjamin et al. 2018; Button et al. 2013; Wacholder et al. 2004). Statistical power is (among other things) a function of the effect size (of the SNP), the size of the analysis sample, and the significance level adopted. Nicolaou et al. (2011) report that their identified SNP explained 0.5% of the likelihood of being an entrepreneur. With their sample of 1335 individuals, they had only 6% power to detect such an effect at  $p < 0.05$ .<sup>5</sup> Hence, it is not surprising that this finding could not be further replicated (Van der Loos et al. 2013).<sup>6</sup>

<sup>5</sup> In their analysis, Nicolaou et al. (2010) adopted a significance level of  $6 \times 10^{-4}$  to account for the correlation between SNPs. As a result, the power of their analysis was almost zero. To be adequately powered (80%), one would have needed a sample of 3643 individuals to find an effect of 0.5% (at  $p = 6 \times 10^{-4}$ ).

<sup>6</sup> The working paper by Wemerfelt et al. (2012) reports an association between a genetic polymorphism and entrepreneurship (proxied by the number of companies founded) in a sample of 135 participants of an executive education course at Harvard Business School. It is evident that in such a sample, the same concerns about statistical power hold.



### 3.3 Hypothesis-free approaches

#### 3.3.1 Genome-wide association studies

GWAS is a hypothesis-free approach to genetic discovery because no prior selection is made on the set of SNPs used in the analysis. To deal with the overidentification problem, a GWAS runs a single regression for every SNP. Hence, millions of regressions are performed in a GWAS. An advantage of the hypothesis-free study design of GWAS is that it makes the need to correct for multiple testing transparent. If the null hypothesis of no association is true for all these millions of SNPs, one still finds a  $p$  value  $< 0.05$  for 5% of the SNPs. Therefore, in a GWAS, the significance threshold is set to  $0.05/1,000,000 = 5 \times 10^{-8}$  (“genome-wide significance”) because of the approximately 1 million independent SNPs in the human genome (adjacent SNPs in the genome are often inherited together). A clear disadvantage of this approach is that GWASs may prioritize SNPs for which the biological function is yet unknown or unclear.<sup>7</sup> Hence, GWAS usually identifies SNPs that need to be subjected to further analyses to understand the pathways between the SNPs and the outcome. Close collaboration with geneticists and biologists in consortia, such as the Entrepreneur Consortium (Van der Loos et al. 2010) and the Social Science Genetic Association Consortium,<sup>8</sup> is therefore a prerequisite for the success of GWAS analysis.

The combination of a very stringent significance level and the small effect sizes of individual SNPs implies that large samples are needed to be adequately powered for gene discovery. The typical dataset has only several thousands of observations, and therefore, datasets need to be combined into mega-analyses or meta-analyses. In a mega-analysis, individual-level genetic data are merged and jointly analyzed. However, legal and privacy issues generally make it impossible to pursue this strategy. In a meta-analysis, the summary results of specific analyses are combined. The GWAS meta-analysis approach has enabled an unprecedented

surge in genetic discoveries that are consistently replicated (Hindorff et al. 2009; Visscher et al. 2017), including the discovery of genetic associations with behavioral outcomes such as educational attainment (Lee et al. 2018; Okbay et al. 2016b; Rietveld et al. 2013), subjective well-being (Okbay et al. 2016a), and more recently preferences such as attitudes toward risk-taking (Linnér et al. 2019). The large sample sizes in these studies ( $N > 1,000,000$  in some of them) could be obtained due to the dramatic decline in the cost of genotyping in the last decade (National Human Genome Research Institute 2018a).

In 2010, Koellinger et al. (2010) calculated that at least 30,000 observations were needed to find a relationship between an individual genetic variant and entrepreneurship at the genome-wide significance level. Quaye et al. (2012) used the GWAS approach in a sample of 3933 British females to assess whether there are associations between specific SNPs and entrepreneurship. Not surprisingly, because of the small sample size, they did not find SNPs that are significant at the genome-wide significance level. Van der Loos et al. (2013) conducted a large-scale GWAS meta-analysis on entrepreneurship in a combined sample of 53,898 individuals from Europe and the USA. Despite the sample size, this study did not find any genome-wide significant SNPs. Moreover, this study found no evidence that any of the genes that were previously suggested in the literature to influence entrepreneurship (Shane 2010) show significant associations with entrepreneurship. From a statistical point of view, this null result could have been driven by the attenuation of the effect sizes through the meta-analysis of samples from different countries and with different birth year profiles. However, GWASs from the past few years on other behavioral outcomes indicate that the effect sizes used in the power calculations by Koellinger et al. (2010) were too high.

The past years of research in behavioral genetics showed that individual SNPs typically explain less than 0.02% of the variance in a behavioral outcome (Chabris et al. 2015; Rietveld et al. 2014a). These findings imply that a sample of at least 197,984 individuals is needed to identify a SNP at the genome-wide significance level with 80% power. Hence, by now, we know that the GWAS meta-analysis of Van der Loos et al. (2013) was underpowered. Although the availability of genetic data is rapidly increasing, genetic data are collected primarily for medical purposes, and measures for

<sup>7</sup> Relatedly, GWAS models usually use a very small number of control variables to capture the full relationship between the SNP and the outcome. For example, Van der Loos et al. (2013) control for only sex, age, and genetic relatedness in their GWAS on self-employment. The use of a small number of control variables causes the interpretation of the estimated effects to be not as straightforward because there may be many pathways through which a SNP influences a behavioral outcome.

<sup>8</sup> <https://www.thessgac.org/>.

entrepreneurship are not always available in medical datasets. There is progress in the collection of genetic data in surveys with an economic focus (such as the US Health and Retirement Study and the English Longitudinal Study of Ageing), but at this moment, a sufficiently large analysis sample for a GWAS on entrepreneurship is not available.

Nevertheless, the heritability estimates for entrepreneurship and the successful discovery of SNPs related to other behavioral outcomes indicate that we can be confident about the eventual success of a GWAS on entrepreneurship. Visscher et al. (2017) showed that the number of identified genetic associations in a GWAS is positively related to the size of the (meta-) analysis sample. For example, whereas the first GWAS meta-analysis on educational attainment ( $N \approx 100,000$ ) found only three genome-wide significant SNPs (Rietveld et al. 2013), the second one (Okbay et al. 2016b) identified 74 SNPs ( $N \approx 300,000$ ), and the third one (Lee et al. 2018) identified 1271 SNPs ( $N \approx 1,100,000$ ). Hence, a GWAS with a sufficiently large sample size—at least four times larger than the sample of  $\sim 50,000$  individuals used by Van der Loos et al. (2013)—will also reveal the SNPs that are associated with entrepreneurship.

### 3.3.2 Genetic discovery using proxy traits

A novel way to boost statistical power in GWASs is the identification of genetic associations using a two-step procedure in the so-called proxy-phenotype method. Rietveld et al. (2014b) introduced this approach to identify genetic associations with cognitive performance. Similar to entrepreneurship, cognitive performance is not often measured in genotyped samples. Therefore, the first step in this method is conducting a large-scale GWAS on a genetically related trait. In the second step, the genetic variants associated with this proxy trait are tested for association with the main trait of interest. In this spirit, Rietveld et al. (2014b) used the results of a GWAS on educational attainment to select 69 independent SNPs, which were then tested for association with cognitive performance. The significance threshold adopted in the second step equals  $\alpha = 0.05/69$  rather than the genome-wide significance threshold of  $\alpha = 5 \times 10^{-8}$ .

Linnér et al. (2019) used this approach in their GWAS on risk tolerance to study the genetic architecture of related traits, such as self-employment. Based on

their main GWAS on risk tolerance, 99 SNPs were selected for further analysis regarding their association with entrepreneurship. In the second stage, the *discovery* GWAS ( $N = 50,627$ ) results of Van der Loos et al. (2013) were used. Using a more lenient threshold for significance, Linnér et al. (2019) found one SNP that was significantly associated with entrepreneurship. The sign of the effect was in the expected direction, meaning that the SNP was related to higher risk tolerance and a higher likelihood of being an entrepreneur. Linnér et al. (2019) claimed in their supplementary materials that “if the association with rs7387531 is robust, this would be the first genetic variant to be found to be significantly associated with self-employment.” However, in the *replication* sample ( $N = 3271$ ) of Van der Loos et al. (2013), the effect of the SNP (rs7387531) was in the opposite direction with  $p > 0.05$ , so it seems that the first robust association between a SNP and entrepreneurship is yet to be identified. Nevertheless, this approach illustrates that the genetic analysis of related traits may help to find genetic variants associated with entrepreneurship.

### 3.4 Polygenic risk scores

Individual SNPs typically explain less than 0.02% of the variance in a behavioral outcome (Chabris et al. 2015), and the GWAS on self-employment by Van der Loos et al. (2013) has shown that the effects of individual SNPs on entrepreneurship are also small (otherwise they would have been found). Hence, individually, genetic variants are practically useless for use in empirical studies. However, the tiny explanatory power of individual genetic variants has encouraged researchers to develop methods that combine individual genetic variants into so-called polygenic risk scores with larger explanatory power. A polygenic risk score is a weighted sum of SNPs and is constructed as follows<sup>9</sup>:

$$PGS_i = \sum_{j=1}^J \beta_j x_{ij},$$

where  $PGS_i$  is the value for the polygenic risk score for individual  $i$ ,  $\beta_j$  is the regression coefficient of SNP  $j$  from the GWAS, and  $x_{ij}$  is the genotype of individual  $i$

<sup>9</sup> More advanced methods for constructing polygenic risk scores exist, for example, methods that better deal with the correlation structure across SNPs within the genome (see, e.g., So and Sham 2017 and Vilhjálmsson et al. 2015). However, the main rationale behind these methods is similar to the basic (still commonly used) approach presented in the main text.

for SNP  $j$  (coded as 0, 1, or 2). This simple approach has been proven to be effective in the out-of-sample prediction of behavioral outcomes. For example, Rietveld et al. (2013) found only three SNPs significantly associated with educational attainment at the genome-wide significance level. Each SNP explained approximately 0.02% of the variance in educational attainment. However, the polygenic risk score based on all SNPs (including the non-significant ones) explained approximately 2.5% of the variance. This percentage increased with the sample size of the GWAS. For example, the most recent polygenic risk score for educational attainment now explains 9.4% (Lee et al. 2018). The prediction attempt of Van der Loos et al. (2013) was unsuccessful in the sense that their polygenic risk score for entrepreneurship captured less than an insignificant 0.2% of the variance. Nevertheless, this percentage will increase if the GWAS for entrepreneurship increases in terms of sample size (Dudbridge 2013).

The weights ( $\beta_j$ ) used in the calculation of the polygenic risk score capture almost the full relationship between the SNP and entrepreneurship: the only control variables used in the GWAS on self-employment by Van der Loos et al. (2013) are sex, age, and variables to account for genetic relatedness between individuals. The relationship between someone's genetic makeup and behavior is assumed to be extremely complex and to run through many (possibly also multiplicative) pathways. Therefore, a "direct" relationship between a SNP and entrepreneurship is unlikely to exist. Many pathways, possibly comprising gene-gene and gene-environment interactions, are likely to explain the relationship between a SNP and behavior. Nevertheless, in a GWAS, these pathways are all included in  $\beta_j$  and therefore also in the polygenic risk score. In the spirit of the proxy-phenotype approach used in GWAS (see Section 3.3.2), we can therefore use the polygenic risk scores of traits that we think are in the pathway between some SNPs and entrepreneurship to foster our understanding about the genetic architecture of entrepreneurship.

One obvious example of such a pathway is risk tolerance. The recent GWAS by Linnér et al. (2019) on risk tolerance shows how the polygenic risk score for risk tolerance does indeed predict entrepreneurship out of sample. Although the explanatory power of this polygenic risk score is relatively small, between 0.57 and 1.36 in terms of (pseudo-)  $R^2$  for different proxies of entrepreneurship, it contributes significantly to the fit

of the model. Moreover, the variance explained is already larger than we may expect it to be for individual SNPs. Risk tolerance may be an obvious trait to investigate when analyzing the pathway between SNPs and entrepreneurship. However, other less obvious traits may also be investigated. For example, earlier research shows that body height is associated with entrepreneurship (Rietveld et al. 2015). The newest polygenic risk score for height explains approximately 34.7% of the variance (Yengo et al. 2018). If the effect of the SNPs explaining entrepreneurship runs through height, we will be able to find an association between the polygenic risk score for body height and entrepreneurship.

Hence, polygenic risk scores constructed for traits other than entrepreneurship may help to identify regions in the human genome that are related to entrepreneurship. Moreover, these genetic summary indices may facilitate the gene-based prediction of entrepreneurship. In the next section, we present empirical analyses that illustrate these two conclusions.

#### 4 Empirical illustration

For our empirical illustration, we draw on data from the US Health and Retirement Study. The HRS is a representative panel of Americans over 50 years old and their spouses. The HRS focuses on a variety of labor markets and health and retirement outcomes. Genetic data were collected from consenting HRS participants between 2006 and 2012 (Health and Retirement Study 2012). We use the RAND HRS Longitudinal File 2014 (V2) for the data on self-employment (Health and Retirement Study 2018a). This longitudinal data file includes the harmonized biennial data of the HRS (1992–2014). Our dependent variable indicating whether an individual is self-employed or not is derived from the question: "Do you work for someone else, are you self-employed, or what?". The respondents could answer "for someone else" or "self-employed." If respondents said they were self-employed, they were coded as 1, and if they replied that they worked for some else, they were coded as 0. Self-employment is the most commonly used measure for entrepreneurship studies drawing on survey data (Parker 2018), although engagement in self-employment can be episodic. We restrict our analyses to those aged between 50 and 65 to exclude individuals active in the labor market after retirement age. Moreover, following the recommendations of the genotyping

center, we restrict the analysis to individuals of recent European descent to preempt bias from unobserved relationships between genetic and environmental factors (Health and Retirement Study 2012).

For the polygenic risk scores, which are the main independent variables in our regressions, we use the HRS Polygenic Scores 2006–2012 Genetic Data - Release 3 (Health and Retirement Study 2018b). In the present illustrative analyses, we use all available polygenic risk scores in this file that relate to mental health.<sup>10</sup> We choose to limit ourselves to the polygenic risk scores of only these traits, as the recent entrepreneurship literature suggests an important link between entrepreneurship and mental health in terms of person-job fit (Benz and Frey 2008; Stephan 2018). In total, we analyze 16 different polygenic risk scores. In our analyses, we control for sex, birth year (dummies for each birth year), and survey waves (dummies for each survey wave). We also control for the first ten principal components of the genetic relationship matrix, as is common in genetic association studies. The latter ten variables control for the genetic aspects of common ancestry that could be spuriously correlated with the polygenic risk scores and the outcome of interest, such as cultural or environmental factors (Rietveld et al. 2014a). To estimate the relationships between self-employment and the polygenic risk scores, we use a linear probability model with random effects (to deal with the time-invariant nature of the polygenic risk scores as well as the longitudinal nature of our data)<sup>11</sup>:

$$SE_{it} = \sum_{k=1}^K \gamma_k PGS_{ik} + \delta Z_{it} + \alpha_i + \varepsilon_{it},$$

where  $SE_{it}$  is the binary variable indicating the self-employment status of individual  $i$  at time  $t$ ,  $\gamma_k$  is the effect of the polygenic risk score  $PGS_{ik}$  for trait  $k$ ,  $\delta$  is a

<sup>10</sup> For some polygenic risk scores, there are multiple versions, reflecting the publication of increasingly large GWAS studies on these traits. In these cases, we use the newest polygenic risk score. For some other traits, there are separate scores for males, females, and the combined sample of males and females. In these cases, we use the combined score.

<sup>11</sup> We present the results of a linear probability model despite the binary nature of our dependent variable because the interpretation of the regression coefficients in a linear probability model with random effects is more straightforward than in a logit model with random effects. However, we note that this choice does not affect our results from a qualitative point of view. In a logit model with random effects, ADHD, autism, bipolar disorder, educational attainment, and cognition are still significant at  $p < 0.05$ . However, the  $p$  value for well-being (0.062) is slightly above the significance threshold.

vector of coefficients for the vector of control variables  $Z_{it}$ ,  $\alpha_i$  is an unobserved random variable for individual  $i$ , and  $\varepsilon_{it}$  is the residual for individual  $i$  at time  $t$ .<sup>12</sup>

Overall, 31,927 (person-year) observations are available from 7948 different individuals. In this sample, 47% of the individuals are male, the average age is 57.4 years (with standard deviation 4.1), and 19.9% of the person-year observations report self-employment. Table 1 displays the estimates of the associations between the different polygenic risk scores and self-employment. We observe that there are six (out of 16) significant associations at the 5% level: the polygenic risk scores for ADHD, autism, bipolar disorder, educational attainment, general cognition, and well-being.<sup>13</sup> For these traits, an increase of one standard deviation leads to an increase or decrease in the likelihood of being self-employed of approximately 1%. These results indicate that polygenic risk scores can significantly predict entrepreneurship (even when proxied by the relatively episodic activity of self-employment) and that genes influencing entrepreneurship are likely to be found in regions in the human genome associated with these six traits.<sup>14</sup>

At the same time, these results illustrate that the predictive power of these polygenic risk scores is small (although larger than the predictive power of individual SNPs). Compared to that of a model without the polygenic risk scores, the explained variance of this model increased by only 0.42%.<sup>15</sup> Table 2 shows that, from a prediction point of view (by taking the percentage of

<sup>12</sup> In the analysis, we estimate the effect of several polygenic risk scores in one single model. As some traits are genetically correlated, such as ADHD and bipolar disorder (Faraone and Larsson 2019), we also analyze models in which we separately include the polygenic risk scores. From a qualitative point of view, the results are similar to the results presented in the main text.

<sup>13</sup> Even with a stringent Bonferroni correction (0.05 divided by the number of polygenic risk scores analyzed), the association with ADHD remains significant.

<sup>14</sup> For illustration purposes, we analyzed all available mental health related polygenic risk scores in the Health and Retirement Study in the present study. The set of polygenic risk scores includes traits for which the link with entrepreneurship is not always evident. Therefore, future studies may use theoretical or other insights for selecting the most promising candidates from the set of available polygenic risk scores rather than using them all. However, the fact that ADHD is found to be the strongest association in our analyses builds confidence in our approach since there are several nongenetic studies showing a similar link (Verheul et al. 2015, 2016; Antshel 2017; Wiklund et al. 2017; Lerner et al. 2019). Nevertheless, future studies need to replicate the current findings in independent datasets to investigate their robustness and generalizability.

<sup>15</sup> Individual SNPs typically explain less than 0.02% of the variance in a behavioral outcome (Chabris et al. 2015; Rietveld et al. 2014a).



person-year observations in our sample in self-employment—19.9%—as the classification threshold), the correct individual-level prediction of self-employment status increases only marginally with the current model (0.14% point increase).

## 5 Conclusion: a second decade?

The “quest for the entrepreneurial gene” (Thurik 2015; Van der Loos et al. 2011) is largely motivated by the struggle of scholars to have a better understanding of entrepreneurs and entrepreneurship: what makes entrepreneurs decide to start a business, what motivates them, what makes them successful or fail, and what makes them different from other people? Various research approaches, as well as tools and theories from economics, psychology, and sociology, have been proposed and applied to these questions. However, the answers to “what makes an entrepreneur” remain uncertain and incomplete (Shane and Venkatamaran 2000; Parker 2018). Empirical evidence that genes may be part of the answer (Nicolaou et al. 2008a, b, 2009, 2011; Shane and Nicolaou 2013; Van der Loos et al. 2011, 2013; Zhang et al. 2009) has been received by scholars and the media with both hopes and enthusiasm, as well as with skepticism and criticism.

Despite several attempts in the past decade, until now, no robust association between genetic variants and entrepreneurship has been discovered. Our overview and discussion of these works gives a clear answer to our first research question, “*Why has the identification of robust associations between genetic variants and entrepreneurship been unsuccessful in the last decade?*” Irrespective of whether a hypothesis-driven or hypothesis-free approach was used, genetic discovery studies on entrepreneurship have until now been underpowered. Nevertheless, based on the results of large-scale genetic discovery studies on other behavioral traits (such as educational attainment), we may expect that robust associations between genetic variants and entrepreneurship will be identified if a sufficiently large sample can be gathered. Datasets that contain both genetic data and entrepreneurship information are relatively scarce (Van der Loos et al. 2013), but the advent of large genotyped biobanks such as the UK Biobank (Bycroft et al. 2018) and the Estonian Biobank (Leitsalu et al. 2015) is currently changing the landscape. Hence, a sufficiently powered GWAS on entrepreneurship may soon become feasible.

Because of data constraints, the latest and largest GWAS on entrepreneurship used self-employment as a proxy for entrepreneurship (Van der Loos et al. 2013). With more data becoming available, future GWASs of entrepreneurship may benefit from the analysis of an entrepreneurship measure less episodic in nature, such as serial or high-performance entrepreneurship. With more precise classification of individuals into occupational groups, the GWAS becomes more powerful and hence the chance to detect associations between individual genetic variants and entrepreneurship becomes larger. Nevertheless, in combination with other GWAS results, the analysis of the relatively heterogeneous self-employment measure may help identify specific underlying types of self-employment. For example, by drawing on GWAS results for schizophrenia and educational attainment, Bansal et al. (2018) reveal that the binary schizophrenia diagnosis aggregates over at least two different subtypes. The first type is associated with high intelligence and bipolar disorder, while the second type is a cognitive disorder that is independent of bipolar disorder. With GWAS results for many publicly available traits,<sup>16</sup> similar analyses may also be interesting to conduct on self-employment to possibly identify unexpected subtypes.

However, rather than directly analyzing entrepreneurship, it is possible to shift attention (at least for the time being) to variables mediating the relationship between genes and entrepreneurship. Examples of such variables that can be measured in large samples include traits such as preferences for risk and uncertainty, confidence, and optimism. In addition to these well-known measures in the world of entrepreneurship research, one may also consider characteristics such as body height, body mass index, and mental disorders (possibly in a hypothesis-free setting). One advantage of this approach is that genetic effects on more proximate outcomes are likely to be stronger and hence easier to detect, for a given sample size, than the genetic effects on distal outcomes, such as entrepreneurship (Rietveld et al. 2014b). By using the proxy-phenotype approach, as discussed in Section 3.3.2, it will be possible to identify associations with entrepreneurship, for example, by using the (publicly available) GWAS results of Van der Loos et al. (2013) in the second step of the analysis.<sup>17</sup> This approach circumvents to some extent the problem

<sup>16</sup> For example, in the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>).

<sup>17</sup> The results of the GWAS on self-employment by Van der Loos et al. (2013) are publicly available via [www.thessgac.org](http://www.thessgac.org).

**Table 1** The association between the polygenic risk scores for traits in the mental health domain and self-employment (random effects regression,  $N_{\text{individual-year}} = 31,927$ ,  $N_{\text{individual}} = 7948$ )

Polygenic risk score	Coefficient	Standard error	<i>p</i> -value
<i>Attention deficit hyperactivity disorder</i>	0.017	0.004	0.000
Anxiety (factor score)	0.001	0.004	0.796
<i>Autism</i>	-0.013	0.006	0.049
<i>Bipolar disorder</i>	0.010	0.005	0.047
Depressive symptoms	0.007	0.005	0.187
<i>Educational attainment</i>	0.013	0.005	0.004
Extraversion	0.007	0.004	0.100
<i>General cognition</i>	-0.012	0.005	0.010
Major depressive disorder	-0.005	0.005	0.367
Mental health (cross disorder)	-0.004	0.007	0.558
Neuroticism	0.008	0.006	0.202
Obsessive compulsive disorder	-0.001	0.004	0.752
Post-traumatic stress disorder	0.001	0.005	0.900
Schizophrenia	0.005	0.008	0.509
<i>Well-being</i>	0.010	0.005	0.032

The regression model includes control variables for sex, age, survey waves, and genetic relatedness. Italicized traits are significant at the 5% level

of the currently insufficient sample size needed for a well-powered GWAS on entrepreneurship.

Although a regular GWAS looks only at the linear association between a genetic variant and entrepreneurship, the genetic architecture of entrepreneurship may comprise interactions between two or more genetic variants. Theoretically, it is possible to include cross-products of SNPs as explanatory variables in a GWAS to advance our understanding of the possibly complex biological mechanisms that are associated with entrepreneurship. However, in a hypothesis-free setting, such

an approach would also require an even more stringent correction of the significance level (as the number of statistical tests increases exponentially with the number of interacting SNPs). Hence, if we assume the size of the interaction effects is not larger than the effects of individual SNPs, this approach is unlikely to be productive in the distant future because of data limitations. The interaction effect may also be identified with (nonlinear) machine learning techniques. Relatively simple machine learning techniques have been proven to have relatively high predictive power for traits such as human height (Pare et al. 2017; Lello et al. 2018). Despite the massive computational burden of these methods, it is promising to analyze to what extent these techniques are also useful for predicting entrepreneurship. Nevertheless, the biological interpretation of the results obtained with machine learning techniques is arguably even more difficult than that of results obtained with a regular GWAS.

To answer our second research question, “*Would the identification of associations between genetic variants and entrepreneurship help to advance the field of entrepreneurship research?*,” we relate the promises of “genoeconomics,” as outlined by Benjamin et al. (2012a), to entrepreneurship research in light of the recent development in behavioral genetics. Benjamin

**Table 2** In-sample prediction results for self-employment (versus wage work) for the models with and without polygenic risk scores; observations in the top 19.9% (percentage of person-year observations reporting self-employment in the sample) of the predicted values in each model are classified as self-employed

Actual occupation	Predicted occupation based on model without polygenic risk scores		Predicted occupation based on model with polygenic risk scores	
	Self-employment	Wage work	Self-employment	Wage work
Self-employment	5.75%	14.11%	5.82%	14.04%
Wage work	14.10%	66.04%	14.03%	66.11%

et al. (2012a) outlined four main reasons why the genetic analysis of behavioral traits is important and relevant. *First*, studies using directly observed genes may reveal the genetic pathways and mechanisms underlying behavior and may lead to a more complete understanding of entrepreneurial behavior. For example, as already discussed above in light of the findings of Bansal et al. (2018), it may be possible to identify to what extent different mechanisms and cognitive processes are involved in the identification and exploitation of business opportunities. *Second*, these studies have the potential to provide measures for constructs that are difficult to measure empirically. Benjamin et al. (2012a) use the example that specific genetic variants can be used as a proxy for the taste for fatty foods. In this spirit, rather than using self-reported measures for entrepreneurial intention, one could draw on the genes related to entrepreneurship. *Third*, based on someone's genetic profile, interventions may be channeled. In this vein, entrepreneurship scholars argue that the prediction of entrepreneurial behavior using genetic data could have practical applications in business and for individual decision-making (Nicolaou et al. 2008a; Nicolaou and Shane 2010; Shane 2010). *Fourth*, genes can be used to enrich otherwise nongenetic models. For example, the inclusion of control variables for genetic endowments may absorb the residual variance in regression models or experimental settings and allow for stronger statistical inference (DiPrete et al. 2018; Rietveld and Webbink 2016). In some instances, it will also be possible to infer causal relationships in observational data by using genes as instrumental variables (Van Kippersluis and Rietveld 2018; Von Hinke et al. 2016). Hence, the use of genes may be instrumental for better understanding the effects of environmental factors.

Regarding the first two promises, we have seen that for behavioral outcomes (such as entrepreneurship), one should not expect values of  $R^2$  in excess of 0.02% for individual SNPs. Hence, it is unlikely that such a SNP will provide much information about the mechanisms underlying entrepreneurship behavior. In contrast to focusing on individual genetic variants, there are good arguments for shifting our attention to polygenic risk scores that summarize the contribution of several genetic variants to a trait. A clear advantage of this approach is that polygenic risk scores can be used as regular variables in empirical research, and expertise for working with raw

genetic data is not necessary, as some polygenic risk scores are already publicly available (such as in the HRS).<sup>18</sup> In the present absence of a polygenic risk score for entrepreneurship with significant explanatory power, we have to shift our focus to the analysis of polygenic risk scores for entrepreneurship-related traits. By doing so, we also come closer to the common practice in entrepreneurship research of testing particular hypotheses (i.e., particular pathways through which genes influence entrepreneurship). For example, we may hypothesize and test whether the genetic variants contributing to the development of ADHD are also related to entrepreneurship. In this spirit, a polygenic risk score can also serve as a proxy for a trait. For example, Patel et al. (2019) use the polygenic risk score for ADHD to study the influence of ADHD on entrepreneurship and entrepreneurial performance in a sample of individuals for which the diagnosis of ADHD was not available.

Regarding the third and fourth promise (the use of genetic information to predict individual behavior and to enrich otherwise nongenetic models), the current state of the behavioral genetics literature as well as the analyses presented in the present study makes clear that the added value of genetics for entrepreneurship scholars should be thought of in terms of enriching population-level models rather than improving individual-level prediction (Morris et al. 2019). Van der Loos et al. (2013) show that all SNPs together may explain up to 25% of differences in entrepreneurial behavior between individuals. Even if we are able to realize this prediction  $R^2$ , the likelihood of misclassification of individual into occupational groups remains great. Hence, early speculations about the use of molecular genetic data for understanding and predicting entrepreneurship (Shane 2010) remain premature, at a minimum. Even though it may be useful to capture some of the (otherwise residual) variance in polygenic risk scores, the gene-based prediction of individual entrepreneurial behavior will remain of limited value

<sup>18</sup> There is currently an important initiative to make a repository of polygenic risk scores for several datasets. However, the exact time window of this initiative is unknown (Okbay et al. 2018). More (future) data sources can be found through portals such as the Database of Genotypes and Phenotypes (dbGaP, Mailman et al. 2007) and the European Genome-phenome Archive (EGA, Lappalainen et al. 2015).

for individuals and entities such as governments and banks.<sup>19</sup>

Nevertheless, capturing residual variance in polygenic scores may improve the understanding of the effects of environmental factors. In so-called gene-by-environment (“GxE”) studies (Keller 2014; Thompson 2014), polygenic risk scores could be used to investigate how entrepreneurship results from the interplay between genetic endowments and environmental factors. For example, a recent study argues that cultural factors (as proxied by the taste for alcoholic drinks) may influence how genes shape different types of entrepreneurship (Acs and Lappi 2019). In general, a good fit between individuals and their occupations has been shown to be important for high levels of productivity (Kristof-Brown et al. 2005). Importantly, the identifiable occurrence of matches and mismatches between an individual and his or her career choices and the possible impact on stress and health was a crucial argument for the medical profession to cooperate with behavioral researchers in the search for the genes associated with entrepreneurship (Koellinger et al. 2010; Van der Loos et al. 2010). Because of the large-scale collections of genetic data and expertise on the biological functioning of genes in the medicine and biology fields, the involvement of researchers in these fields will remain crucial to find associations between genetic variants and entrepreneurship.

In sum, although the attempts to identify specific genetic variants underlying the heritable variation in entrepreneurship have until now been unsuccessful, there is reason to be confident about the eventual success of the “quest for the entrepreneurial gene” (Van der Loos et al. 2011). The benefits of using individual genetic variants for empirical research in the entrepreneurship domain are likely to be small. However, the use of polygenic risk scores may promote the realization of the promises of genoconomics for entrepreneurship research. Although the gene-based prediction of individual entrepreneurial behavior will be of limited value, the use of polygenic risk scores in models may help to increase our understanding of which regions in the genome and which combinations of genetic endowments and environmental circumstances drive entrepreneurship and person-job fit at the population level.

<sup>19</sup> Besides, ethical considerations are needed to determine whether such gene-based prediction of entrepreneurship is actually desirable.

**Acknowledgments** The authors thank André van Stel and Kristel de Groot for providing extensive feedback on earlier drafts of this manuscript.

**Funding information** The HRS (Health and Retirement Study) is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. C.A.R. acknowledges funding from the New Opportunities for Research Funding Agency Cooperation in Europe (NORFACE-DIAL grant 462-16-100). A.R.T. is member of the Entrepreneurship and Innovation Chair, which is part of Labex Entrepreneurship (University of Montpellier, France) and is funded by the French government (Labex Entreprendre, ANR-10-Labex-11-01).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Acs, Z., & Lappi, E. (2019). Entrepreneurship, culture, and the epigenetic revolution: a research note. *Small Business Economics*. <https://doi.org/10.1007/s11187-019-00230-0>.
- Antshel, K. M. (2017). Attention deficit/hyperactivity disorder (ADHD) and entrepreneurship. *Academy of Management Perspectives*, 32(2), 243–265.
- Bansal, V., Mitjans, M., Burik, C. A., Linner, R. K., Okbay, A., Rietveld, C. A., et al. (2018). Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. *Nature Communications*, 9(1), 1–12.
- Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., et al. (2012a). The promises and pitfalls of genoconomics. *Annual Review of Economics*, 4(1), 627–662.
- Benjamin, D. J., Cesarini, D., Van Der Loos, M. J., Dawes, C. T., Koellinger, P. D., Magnusson, P. K., et al. (2012b). The genetic architecture of economic and political preferences. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21), 8026–8031.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10.
- Benz, M., & Frey, B. S. (2008). The value of doing what you like: evidence from the self-employed in 23 countries. *Journal of Economic Behavior & Organization*, 68(3–4), 445–455.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure:



- why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726), 203–209.
- Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, 24(4), 304–312.
- DiPrete, T. A., Burik, C. A. P., & Koellinger, P. D. (2018). Genetic instrumental variable regression: explaining socioeconomic and health outcomes in nonexperimental data. *Proceedings of the National Academy of Sciences*, 115(22), 4970–4979.
- Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genetics*, 9(3), e1003348.
- Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S., & Knafo, A. (2010). Genetics of human social behavior. *Neuron*, 65(6), 831–844.
- Ellis, P. D. (2010). *The essential guide to effect sizes: an introduction to statistical power, meta-analysis and the interpretation of research results*. Cambridge: Cambridge University Press.
- Faraone, S. V., & Larsson, H. (2018). Genetics of attention deficit hyperactivity disorder. *Molecular Psychiatry*, 24, 562–575.
- Health and Retirement Study. (2012). Quality control report for genotypic data [PDF file]. Retrieve from [http://hrsonline.isr.umich.edu/sitedocs/genetics/HRS\\_QC\\_REPORT\\_MAR2012.pdf](http://hrsonline.isr.umich.edu/sitedocs/genetics/HRS_QC_REPORT_MAR2012.pdf). Accessed 23 Jan 2019.
- Health and Retirement Study (2018a). *HRS Polygenic Scores 2006-2012 Genetic Data - Release 3*. Retrieved from: <https://hrs.isr.umich.edu/news/hrs-polygenic-scores-2006-2012-genetic-data-release-3>. Accessed 23 Jan 2019.
- Health and Retirement Study (2018b). *RAND HRS longitudinal file 2014 (V2) documentation* [PDF file]. Retrieved from [https://www.rand.org/content/dam/rand/www/external/labor/aging/dataprod/randhrs1992\\_2014v2.pdf](https://www.rand.org/content/dam/rand/www/external/labor/aging/dataprod/randhrs1992_2014v2.pdf). Accessed on 23 Jan 2019.
- Hindorf, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, 106(23), 9362–9367.
- Hewitt, J. (2012). Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behavior Genetics*, 42(1), 1–2.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), 124.
- Keller, M. C. (2014). Gene×environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biological Psychiatry*, 75(1), 18–24.
- Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioral genetics* (7th ed.). New York: Worth Publishers.
- Koellinger, P. D., van der Loos, M. J., Groenen, P. J., Thurik, A. R., Rivadeneira, F., van Rooij, F. J., et al. (2010). Genome-wide association studies in economics and entrepreneurship research: promises and limitations. *Small Business Economics*, 35(1), 1–18.
- Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: a meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, 58(2), 281–342.
- Lappalainen, I., Almeida-King, J., Kumanduri, V., Senf, A., Spalding, J. D., Saunders, G., et al. (2015). The European Genome-phenome Archive of human data consented for biomedical research. *Nature Genetics*, 47(7), 692–695.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, 50(8), 1112–1121.
- Leitsalu, L., Haller, T., Esko, T., Tammesoo, M. L., Alavere, H., Snieder, H., et al. (2015). Cohort profile: Estonian biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology*, 44(4), 1137–1147.
- Lello, L., Avery, S. G., Tellier, L., Vazquez, A. I., de los Campos, G., & Hsu, S. D. (2018). Accurate genomic prediction of human height. *Genetics*, 210(2), 477–497.
- Lerner, D. A., Verheul, I., & Thurik, A. R. (2019). Entrepreneurship and attention deficit/hyperactivity disorder: a large-scale study involving the clinical condition of ADHD. *Small Business Economics*, 53(2), 381–392.
- Lindquist, M. J., Sol, J., & Van Praag, M. (2015). Why do entrepreneurial parents have entrepreneurial children? *Journal of Labor Economics*, 33(2), 269–296.
- Linnér, R. K., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., et al. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, 51(1), 245–257.
- Mailman, M. D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nature Genetics*, 39(10), 1181.
- Morris, T. T., Davies, N. M., & Smith, G. D. (2019). Can education be personalised using pupils' genetic data?. *BioRxiv*, 645218.
- National Human Genome Research Institute (2018a). DNA sequencing costs: Data. Retrieved from <https://www.genome.gov/27541954/dna-sequencing-costs-data/>. Accessed 29 Jan 2019.
- National Human Genome Research Institute (2018b). *Frequently asked questions about genetic and genomic science*. Retrieved from <https://www.genome.gov/19016904/faq-about-genetic-and-genomicscience/>. Accessed 29 Jan 2019.
- Nicolaou, N., & Shane, S. (2009). Can genetic factors influence the likelihood of engaging in entrepreneurial activity? *Journal of Business Venturing*, 24(1), 1–22.
- Nicolaou, N., & Shane, S. (2010). Entrepreneurship and occupational choice: genetic and environmental influences. *Journal of Economic Behavior and Organization*, 76(1), 3–14.
- Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J., & Spector, T. D. (2008a). Is the tendency to engage in entrepreneurship genetic? *Management Science*, 54(1), 167–179.
- Nicolaou, N., Shane, S., Cherkas, L., & Spector, T. D. (2008b). The influence of sensation seeking in the heritability of entrepreneurship. *Strategic Entrepreneurship Journal*, 2(1), 7–21.
- Nicolaou, N., Shane, S., Cherkas, L., & Spector, T. D. (2009). Opportunity recognition and the tendency to be an entrepreneur: a bivariate genetics perspective. *Organizational Behavior and Human Decision Processes*, 110(2), 108–117.

- Nicolaou, N., Shane, S., Adi, G., Mangino, M., & Harris, J. (2011). A polymorphism associated with entrepreneurship: evidence from dopamine receptor candidate genes. *Small Business Economics*, 36(2), 151–155.
- Nofal, A. M., Nicolaou, N., Symeonidou, N., & Shane, S. (2018). Biology and management: a review, critique, and research agenda. *Journal of Management*, 44(1), 7–31.
- Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., et al. (2016a). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, 48(6), 624–633.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., et al. (2016b). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533(7604), 539–542.
- Okbay, A., Becker, J., Benjamin, D., Burik, C. A. P., Cesarini, D., & Turley, P. (2018). A repository of polygenic scores. *Behavior Genetics*, 49(6), 507.
- Pare, G., Mao, S., & Deng, W. Q. (2017). A machine-learning heuristic to improve gene score prediction of polygenic traits. *Scientific Reports*, 7(1), 1–11.
- Parker, S. C. (2018). *The economics of entrepreneurship* (2nd ed.). Cambridge: Cambridge University Press.
- Patel, P. C., Rietveld, C. A., & Verheul, I. (2019). Attention deficit hyperactivity disorder (ADHD) and earnings in later-life self-employment. *Entrepreneurship Theory and Practice*. <https://doi.org/10.1177/1042258719888641>.
- Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47(7), 702–709.
- Quaye, L., Nicolaou, N., Shane, S., & Mangino, M. (2012). A discovery genome-wide association study of entrepreneurship. *International Journal of Developmental Science*, 6(3–4), 127–135.
- Ramsköld, D., Wang, E. T., Burge, C. B., & Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Computational Biology*, 5(12), 1000598.
- Rietveld, C. A., & Webbink, D. (2016). On the genetic bias of the quarter of birth instrument. *Economics and Human Biology*, 21(1), 137–146.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, 340(6139), 1467–1471.
- Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A., et al. (2014a). Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological Science*, 25(11), 1975–1986.
- Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Benyamin, B., Chabris, C. F., & Koellinger, P. D. (2014b). Common genetic variants associated with cognitive performance identified using proxy-phenotype method. *Proceedings of the National Academy of Sciences*, 111(38), 13790–13794.
- Rietveld, C. A., Hessels, J., & van der Zwan, P. (2015). The stature of the self-employed and its relation with earnings and satisfaction. *Economics and Human Biology*, 17(1), 59–74.
- Shane, S. (2010). *Born entrepreneurs, born leaders: how your genes affect your work life*. New York: Oxford University Press.
- Shane, S., & Nicolaou, N. (2015). Creative personality, opportunity recognition and the tendency to start businesses: a study of their genetic predispositions. *Journal of Business Venturing*, 30(3), 407–419.
- Shane, S., & Venkatamaran, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25(1), 217–226.
- Shane, S., & Nicolaou, N. (2013). The genetics of entrepreneurial performance. *International Small Business Journal*, 31(5), 473–495.
- So, H. C., & Sham, P. C. (2017). Improving polygenic risk prediction from summary statistics by an empirical Bayes approach. *Scientific Reports*, 7(1), 41262.
- Stephan, U. (2018). Entrepreneurs' mental health and well-being: a review and research agenda. *Academy of Management Perspectives*, 32(3), 290–322.
- The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74.
- Thompson, J. N. (2014). *Interaction and coevolution*. University of Chicago Press.
- Thurik, A. R. (2015). Determinants of entrepreneurship: the quest for the entrepreneurial gene. In D. B. Audretsch, Ch. S. Hayter & A. N. Link (Eds.), *Concise guide to entrepreneurship, technology and innovation* (pp. 28–38). Cheltenham: Edward Elgar Publishing Limited.
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9(5), 160–164.
- Van der Loos, M. J., Koellinger, P. D., Groenen, P. J., & Thurik, A. R. (2010). Genome-wide association studies and the genetics of entrepreneurship. *European Journal of Epidemiology*, 25(1), 1–3.
- Van der Loos, M. J., Koellinger, P. D., Groenen, P. J., Rietveld, C. A., Rivadeneira, F., van Rooij, F. J., et al. (2011). Candidate gene studies and the quest for the entrepreneurial gene. *Small Business Economics*, 37(3), 269–275.
- Van der Loos, M. J., Rietveld, C. A., Eklund, N., Koellinger, P. D., Rivadeneira, F., Abecasis, G. R., et al. (2013). The molecular genetic architecture of self-employment. *PLoS One*, 8(4), e60542.
- Van Kippersluis, H., & Rietveld, C. A. (2018). Pleiotropy-robust Mendelian randomization. *International Journal of Epidemiology*, 47(4), 1279–1288.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science*, 291(5507), 1304–1351.
- Verheul, I., Block, J., Burmeister-Lamp, K., Thurik, A. R., Tiemeier, H., & Turtura, R. (2015). ADHD-like behavior and entrepreneurial intentions. *Small Business Economics*, 45(1), 85–101.
- Verheul, I., Rietdijk, W., Block, J., Franken, I., Larsson, H., & Thurik, R. (2016). The association between attention-deficit/hyperactivity (ADHD) symptoms and self-employment. *European Journal of Epidemiology*, 31(8), 793–801.
- Vilhjálmsdóttir, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The American Journal of Human Genetics*, 97(4), 576–592.

- Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nature Reviews Genetics*, 9(4), 255–266.
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *American Journal of Human Genetics*, 101(1), 5–22.
- Von Hinke, S., Smith, G. D., Lawlor, D. A., Propper, C., & Windmeijer, F. (2016). Genetic markers as instrumental variables. *Journal of Health Economics*, 45(1), 131–148.
- Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L., & Rothman, N. (2004). Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *Journal of the National Cancer Institute*, 96(6), 434–442.
- Wernerfelt, N., Rand, D. G., Dreber, A., Montgomery, C., & Malhotra, D. K. (2012). *Arginine vasopressin 1a receptor (AVPR1a) RS3 repeat polymorphism associated with entrepreneurship*. Retrieved from [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2141598](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2141598). Accessed 29 Jan 2019.
- Wiklund, J., Yu, W., Tucker, R., & Marino, L. (2017). ADHD, impulsivity, and entrepreneurship. *Journal of Business Venturing*, 32(6), 627–656.
- 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), 565.
- Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., et al. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics*, 27(20), 3641–3649.
- Zhang, Z., Zyphur, M. J., Narayanan, J., Arvey, R. D., Chaturvedi, S., Avolio, B. J., et al. (2009). The genetic basis of entrepreneurship: effects of gender and personality. *Organizational Behavior and Human Decision Processes*, 110(2), 93–107.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.