Determinants of Entrepreneurship: The quest for the entrepreneurial gene

Roy Thurik

Erasmus School of Economics and Montpellier Business School

Source: The present text is equal to Chapter Six of the *Concise Guide to Entrepreneurship. Technology and Innovation* edited by David B. Audretsch, Christopher S. Hayter and Albert N. Link.

Ackowledgement: The present chapter on entrepreneurship and genes would not have been possible without the following persons: Patrick Groenen, Albert Hofman, Philipp Koellinger, Matthijs van der Loos, Niels Rietveld, Fernando Rivadeneira, Frank van Rooij, André Uitterlinden and many others. I thank them for letting me use jointly written texts. I will refrain from burdening the below text with constantly citing the four publications on which it is based: Koellinger et al. (2010), Van der Loos et al. (2011, 2013) and Rietveld & Van der Loos (2012).

Bibliographical details: Thurik, A.R., (2015), Determinants of entrepreneurship, in D.B. Audretsch, Ch.S. hayter and A.N. Link (eds), *Concise Guide to Entrepreneurship. Technology and Innovation*, Cheltenham, UK: Edward Elgar Publishing Limited, 28-38.

Introduction

Research on the determinants of entrepreneurship has a long history (Parker, 2009). The most recent and promising approaches include the investigation of biological determinants, such as genes, hormones and brain activity measures (electroencephalography or magnetic resonance imaging). The present chapter reviews several recent endeavors to connect genes to proxies of entrepreneurship, in particular self-employment. It should be read as an overview of the shortcomings and potential of two research methods: candidate gene studies and genome-wide association (GWA) studies.

There are two popular views on what makes an entrepreneur. The first is that anyone can learn the necessary skills, provided they dedicate sufficient time and effort. The second view is that people are either born with the right personality and skills or they are not, and developing these traits is impossible. Which of these two views — the nurture or the nature hypothesis — is true, or the interplay between the two, has far-reaching implications for individual behavior and economic policies. Evidence suggests that inherited qualities play a role in occupational choice, with recent scientific advances showing different pathways through which genes can influence entrepreneurial behavior. The current view in this debate concludes that neither nurture nor nature alone are responsible for behavioral outcomes such as entrepreneurial choice. Rather, it is a complex interplay of both.

Self-employed parents may transfer relevant skills and a familiarity with entrepreneurial behavior to their children. Alternatively, inherited characteristics that can affect the tendency to become an entrepreneur may also explain the observed intergenerational effects. Examples of such characteristics include preferences for risk-seeking, altruism in dictator games, job satisfaction, vocational interests, work values, novelty-seeking, gambling, general cognitive ability and intelligence, educational attainment and overconfidence. Moreover, several twin studies suggest a genetic influence on the propensity to become self-employed (Nicolaou et al., 2008a, 2008b; Nicolaou and Shane, 2009; Zhang et al., 2009). In these studies, the heritability of proxies for entrepreneurship is consistently estimated to be in the range of 40 to 60 percent.

Entrepreneurship — which is proxied by self-employment in the present chapter — has been the target of attempts to identify specific genetic polymorphisms underlying its heritable variation. These attempts have been unsuccessful so far because the candidate gene studies were not replicable, while genome-wide association studies did not have sufficient sample sizes for genetic discovery, and the available proxy for entrepreneurship (i.e., self-employment) is too broad. In this chapter, the results of both the candidate gene and the genome-wide association approach are presented. These two approaches make it possible to find individual genetic variants associated with entrepreneurship, and complement other methods that consider (scaled) combinations of genetic variants simultaneously, such as twin studies and genomic-relatedness-matrix restricted maximum likelihood (GREML) studies (Benjamin et al., 2012).

Before presenting the dos and don'ts of the quest for the entrepreneurial gene, it is important to note that, in contrast to popular views, a genetic influence would not imply determinism or the irrelevance of the environment or free will; a genetic influence only implies a shift in an individual's probability of exhibiting a behavior, such as the tendency to become self-employed.

Basic genetic concepts

When a trait is heritable, it is, in principle, possible to locate the sites in the human genome that influence it. The human genome consists of all of the genetic information in human cells and is composed of 23 chromosomal pairs; half of the chromosomes are inherited from the mother and half from the father. These chromosomes 'package' DNA molecules and encode the genetic information along two DNA strands. A DNA strand is a polymer of nucleotides. Each nucleotide is a building block containing a base, which can be adenine (A), cytosine (C), guanine (G) or thymine (T); thus, there are four distinct nucleotides. DNA is structured as a double helix, where two DNA strands are held together by weak hydrogen bonds. Hydrogen bonding occurs between the bases of opposing nucleotides along the two strands: adenine always binds to thymine, and cytosine always binds to guanine. Consequently, two DNA strands of a DNA duplex have complementary sequences, and the sequence of one DNA strand can easily be inferred if the DNA sequence of its complementary strand is already known. DNA sequences are usually described by writing the sequence of the bases for only one strand. For example, an individual may have inherited the AA nucleotides for one particular position on a pair of chromosomes (i.e., a genotype). This inheritance would imply that the individual inherited an A base from the paternal chromosome and an A base from the maternal one. Another individual may have inherited the AG nucleotides at the same position, i.e., a different base from each of the two parents, while a third may have inherited both GG nucleotides from each parent. Alternative bases in a nucleotide at the same physical locus are called alleles. A DNA sequence on one position of the genome that exhibits at least a 1 percent variation between members of a species is called a single nucleotide polymorphism (SNP). The minor allele frequency (MAF) refers to the frequency of the less common allele of a SNP in a population.

Almost all human DNA is identical from person to person. To date, geneticists have identified 27 million SNPs among humans, while the entire human genome consists of some three billion nucleotides. These SNPs and other types of genomic variation are what make humans different from each other. The total number and locations of SNP markers that need to be genotyped to detect an association between common genetic variants and an outcome of interest (a phenotype of an individual) were identified by the HapMap project (International HapMap Consortium, 2005).

Until recently, genotyping was performed with arrays of 550,000 SNPs that, after data cleaning, deliver information about the specific alleles for approximately 500,000 SNPs. Although these arrays already give a high-resolution image of the human genome, the newest generation microarrays allow researchers to array two to 12 million markers per sample, including comprehensive coverage of both common and rare variants.

Candidate gene studies and genome-wide association studies

When the DNA of a sufficient number of individuals has been genotyped, their genotypes for certain SNPs can be associated with an outcome of interest, such as the presence of a disease, an IQ score, or the employment status of an individual. For a binary outcome such as entrepreneurship (with $y = 1$ for the individual being an entrepreneur, and $y = 0$ otherwise), we can test for an association by conducting a logistic regression for each SNP. When 500,000 SNPs are available for statistical analyses, 500,000 logistic regressions must be conducted.

The question is whether it is really necessary to test each SNP for association, or whether the analysis can be limited to a subset of SNPs? This choice is, in essence, the difference between candidate gene studies and genome-wide association (GWA) studies. Candidate gene studies hypothesize the relation between certain genes and the outcome of interest (phenotype). Only the hypothesized SNPs are tested for an association. GWA studies are hypothesis-free: no association between certain genes and a particular phenotype is hypothesized, and therefore, all available SNPs are tested. Hence, the GWA study is an exploratory method that does not rely on prior hypotheses.

As the number of independent statistical tests increases, so does the problem of multiple testing. By pure chance, a number of SNPs will show significant associations, even if there is no actual relationship between the SNPs and the phenotype. For example, assume that we analyze 500,000 SNPs where none are truly associated with the phenotype, i.e., the statistical null hypothesis of no association between the SNP and the outcome is correct. If we adopt a 1 percent significance level for hypothesis testing, performing 500,000 tests should yield 5,000 incorrect rejections of the null hypothesis (i.e., false positives).

Following this reasoning, the number of false positives could be reduced by testing only a small set of SNPs instead of all available genotyped SNPs. The question is, however, whether selecting a limited set of SNPs and, consequently, testing a limited number of hypotheses is appropriate. Current knowledge of DNA does not enable us to predict which genes and how many are associated with entrepreneurship nor how

strong their association is. For example, we know that approximately 70 percent of all genes are expressed in the brain and that brain function influences behavior. Thus, it is very possible to derive a seemingly plausible hypothesis for practically every gene (and therefore every SNP), and each of these hypotheses may sound credible for different reasons.

As a consequence, empirical research that focuses on a small subset of these hypotheses, such as candidate gene studies, is forced to make arbitrary choices regarding these hypotheses. A large number of false positive results can be expected if the statistical confidence intervals are not appropriately adjusted to reflect the total number of plausible hypotheses (Ioannidis, 2005) and if researchers yield to publication bias sentiments. Such adjustments are typically missing in candidate gene studies because researchers only point to their 'theory' as a justification for focusing on a small number of candidates. As a result, most findings of candidate gene studies are not replicable, while replication in an independent study dramatically lowers the chance of a false positive. Examples of replication failure in the social sciences include genetic loci associated with personality traits, behavior in dictator games and harm avoidance. For instance, Israel et al. (2009) report an association between a variant of the OXTR gene and the dictator game, which Apicella et al. (2010) failed to replicate. Vormfelde et al. (2006) report an association between a variant in the serotonin transporter gene and anxiety-related traits such as harm avoidance, which Lang et al. (2004) failed to replicate. Ioannidis (2005) showed that the pre-study probability of a genetic association being true is generally extremely low, and consequently, the post-study probability is also low.

To keep the false positive rate at an acceptably low level, stringent significance tests must be used to compensate for multiple testing. Even if researchers do not test all SNPs for an association, correcting for the existence of these alternative hypotheses is imperative. For individuals of European descent, the consensus is to account for one million independent tests. Based on this number, the often used so-called Bonferroni correction proposes a significance level of 5×10^{-8} to obtain a family-wise significance level of 5 percent (the probability of making one or more type I errors among all hypotheses while performing multiple testing). This significance level is often referred to as 'genome-wide significance,' and only SNPs that pass this threshold are considered to be true positives (Beauchamp et al., 2010). This also makes clear that very large sample sizes are needed in GWA studies to discover true associations.

Non-replication of a candidate gene study

In a recent paper, Nicolaou et al. (2011) reported a significant association between a SNP in the dopamine receptor D3 (*DRD3*) gene and the tendency to be an entrepreneur in a group of 1,335 British subjects. In this candidate gene study, SNPs in a set of nine genes were tested for an association with the tendency to be an entrepreneur, resulting in a single significant association. The set of candidate genes consisted of five dopamine receptor genes associated with novelty- or sensation-seeking and four genes associated with attention deficit hyperactivity disorder (ADHD). These specific genes were selected based on the

observation that sensation-seeking and ADHD are more common among entrepreneurs. The authors claimed that this is the first evidence of an association between variants of a specific gene and entrepreneurship.

As argued above, the appropriate significance threshold in candidate gene studies should be 5×10^{-8}. The reported association by Nicolaou et al. (2011) has a *p*-value of 0.0002, much higher than the genome-wide significance level. To evaluate this result, Van der Loos et al. (2011) tried to replicate their findings by performing an association analysis of the 18 SNPs reported in Nicolaou et al. (2011), including the significant association between a SNP in the *DRD3* gene and entrepreneurship, in three much larger independent groups of Dutch subjects from the Rotterdam Study (Hofman et al., 2009).

The Rotterdam Study (RS) consists of three cohorts: 5,974 participants in RS-I have been successfully genotyped, 2,129 in RS-II and 2,030 in RS-III. Because the type of array differs between the candidate gene study and the replication study, not all 18 reported SNPs were readily available in the Rotterdam Study cohorts. Therefore, these SNPs from the available genotype data were imputed using MACH (Li et al., 2009).

Van der Loos et al. (2011) constructed a binary variable indicating whether a subject (i) had never been self-employed or (ii) had been self-employed at least once during his/her complete working life (RS-I) or in his/her current or last occupation (RS-II and RS-III). For RS-I, the individuals with an incomplete working life history and the individuals who had never had a job were excluded, except those who were self-employed at least once. The rationale for this exclusion is that incomplete work life histories could 'contaminate' the control group with people who were self-employed at least once. Complete SNP and self-employment data were available for 5,374 subjects (531 cases, 4,843 controls) in RS-I, 2,066 subjects (197 cases, 1,869 controls) in RS-II, and 1,925 subjects (209 cases, 1,716 controls) in RS-III. The measure of entrepreneurship in Van der Loos et al. (2011) is equivalent to the definition used by Nicolaou et al. (2011), i.e., 'Have you ever started a business in your working life?' This equivalence was confirmed by a correlation coefficient of 0.87 between the two proxies for self-employment and starting a new business (Nicolaou et al., 2008a).

An association analysis was performed by Van der Loos et al. (2011) for each SNP by logistic regression (Li et al. 2009). For each SNP, two models were estimated: Model 1, which includes the SNP as an independent variable, and Model 2, which controls for sex and possible population stratification by including the first four principal components of the genotypic covariance-variance matrix. For RS-III, a dummy for age (>=50) was included in the latter model. Because Van der Loos et al. (2011) were replicating previously reported associations, it was appropriate to correct only for the number of SNPs that are replicated. The Bonferroni correction results in a significance level of 0.0028 (0.05/18 tests), which corresponds to a significance level of 0.05 for all tests. This level is much higher than the genome-wide significance level of 5×10^{-8}. The full estimation results are given in Van der Loos et al. (2011): none of the SNPs are even remotely significant in both models.

The estimation results for RS-I require additional explanation. Nicolaou et al. (2011) reported a significant association between a certain SNP in the *DRD3* gene and the tendency to be an entrepreneur. This SNP was not significantly associated in RS-I at the chosen significance level of 0.0028. Moreover, the negative coefficient suggests the opposite association, which demonstrates that the original finding was probably a false positive.

Further inspection of the results indicates that three SNPs within the *DRD3* gene survive the Bonferroni-corrected significance level of 0.0028. However, the direction of the associations is opposite to the associations reported in the original candidate gene study. Although the hypothesis that the *DRD3* gene is associated with entrepreneurship cannot be rejected based on these results, they do not support the effect of the G allele of the SNP reported by Nicolaou et al. (2011).

Discussion of non-replication

There are several shortcomings in the candidate gene studies, exemplified in Nicolaou et al. (2011), that lead to the skepticism that a reported association is a false positive and that all of the results in this area so far should also be interpreted with care. These shortcomings are lessons learned from the era of candidate gene studies, usually pursued with ill-defined markers across genes, small samples and/or lacking replication. The fact that the reported associations cannot be replicated underlines several arguments.

First, there is the suspicion that the selection of candidates, although seemingly sound, is largely arbitrary. This selection consists of genes previously thought to be associated with novelty- or sensation-seeking and ADHD, characteristics that are hypothesized to be more common among entrepreneurs. However, there are many other candidate genes, such as the serotonin 2A and 1B transporters (*HTR2A* and *HTR2B*), dopamine and serotonin transporters (*SLC6A3*, *SLC6A4*), dopamine beta-hydroxylase (*DBH*), monoamine oxidase B (*MAOB*), and genes associated with testosterone levels. Furthermore, the majority of genes are related to either brain function or to the expression of proteins in the brain and could therefore be candidates. Hence, there may be hundreds of thousands of potential candidate loci. This large number of potential genes makes the candidate gene approach, at present, infeasible for the study of complex behaviors such as entrepreneurship.

Second, the selection criteria of SNPs within the chosen candidate genes are confined to the coding regions. A complete overview of the selected SNPs is lacking, although Nicolaou et al. (2011) report that the SNPs from the coding regions of the nine candidate genes were selected. SNPs in regulatory non-coding regions are not considered, although these could have substantial effects on a given phenotype. For an overview, see www.genome.gov/gwastudies.

Third, the hypothesis that dopamine receptor genes are associated with novelty- or sensation-seeking is based on mixed evidence from small-scale studies that could not always be replicated. For example, one study reported a significant association between a variant of the *DRD4* gene and novelty-seeking, but this association could not be replicated by a different study. A recent meta-analysis concludes that the *DRD4*

gene may be associated with measures of novelty seeking and impulsivity, but significant evidence of publication bias was found (Munafo et al., 2008).

Therefore, unfortunately, the candidate gene study of Nicolaou et al. (2011) should be interpreted with care because it does not sufficiently adjust for multiple testing. Even the significance level of 0.0028 used in the replication study is potentially too liberal. The reported association from Nicolaou et al. (2011) is likely to be a false positive and, hence, not a serious candidate for replication studies.

A genome-wide association study

In Van der Loos et al. (2013), self-employment (having started, owned and managed a business) is used as a proxy for entrepreneurship. A meta-analysis of GWA studies of self-employment was performed using 16 studies to identify genetic variants that are robustly associated with self-employment. Together, these studies were composed of 50,627 participants of European ancestry who are part of the Gentrepreneur Consortium. This study is the first large-scale effort to identify common genetic variants that are associated with an economic variable. A second study is Rietveld et al. (2013), which analyzed educational attainment.

Theoretical and empirical evidence from entrepreneurship research suggests that there are differences between males and females with respect to the types of businesses they start. These differences also extend to individuals' motivations, goals and resources (Verheul et al., 2012) and exist because women face different — and often more — barriers to entrepreneurship than men (Verheul and Thurik, 2001). Therefore, both pooled and sex-stratified analyses were performed.

The discovery stage of this study did not identify any genome-wide significant associations: there are no common SNPs for self-employment with moderate to large effect sizes. Gene-based tests for approximately 17,700 genes, including several candidate genes for entrepreneurship that have been previously suggested in the literature (Shane, 2010), did not reveal any significant associations. A SNP that is located in the DRD3 gene and was identified by Nicolaou et al. (2011) did not correlate with the tendency to be an entrepreneur. Lastly, 58 SNPs in the discovery stage with a *p*-value below 10^{-5} (a threshold that was predefined in the analysis plan) were tested in a replication sample of 3,271 individuals, but none were replicated.

As the heritability of entrepreneurship is consistently measured between 40 and 60 percent (Nicolaou et al., 2008a; Zhang et al., 2009; Van der Loos et al., 2010), a plausible interpretation of these results is that the molecular genetic architecture of self-employment is highly polygenic: there are hundreds or maybe thousands of genetic variants that individually have a small effect, which together explain a substantial proportion of the heritability. Additionally, a complex interplay between genes and the environment may play a role. Finally, the possibility cannot be ruled out that rare genetic variants or other currently unknown and unmeasured variants that are insufficiently correlated with the SNPs have large effects on an individual's tendency to be self-employed. However, if these genetic variants are rare, they would not

contribute a great deal to the population-based variance in self-employment, and large samples would be required to identify these variants.

The results of Van der Loos et al. (2010) are similar to those that have been reported for biologically more proximate human traits, such as height, and diseases, such as schizophrenia, for which a polygenic molecular genetic architecture has also been suggested. One implication of this similarity is that with sufficiently large sample sizes, SNPs that are associated with self-employment can, in principle, be discovered, as has been the case for height (Wood et al., 2014). A discovery sample of approximately 50,000 individuals is apparently still too small for a meta-analysis of GWA studies on a biologically distal, complex, and relatively rare human behavior such as self-employment. Moreover, self-employment is a fuzzy phenotype, having different meanings in different environments, i.e., cohorts. Lastly, there is the possibility of gene-environment interactions (the interplay between nature and nurture), which would make it even more difficult to identify the effects of individual SNPs in a GWA study that pools results from very different cohorts and environments.

Conclusion

Twin study estimates show that part of the variance in the propensity to engage in entrepreneurship can be explained by genetic variation. Hence, in principle, it should be possible to find the genetic loci that influence this propensity. In this chapter, it is argued that GWA studies are the best scientific approach given our current knowledge of DNA. There are several reasons: first, theories for selecting SNPs for candidate gene approaches are typically weak and arbitrary and, therefore, not convincing. Second, GWA studies make clear the need to correct for multiple testing. The non-replication of the candidate gene study illustrates this argument.

Therefore, large-scale GWA studies are the best method to conduct research in the ‘quest for the entrepreneurial gene.’ However, a discovery sample from Van der Loos et al. (2013) of approximately 50,000 individuals was apparently still too small for a meta-analysis of GWA studies on a biologically distal, complex, and relatively rare human behavior such as self-employment. A potential opportunity for future research includes performing GWA studies on endophenotypes such as risk preferences, confidence, and independence. The effect sizes of individual SNPs on these endophenotypes may be larger because of their greater biological proximity. However, these variables are difficult to measure reliably and are not (yet) available in many genotyped samples. An alternative may be to use less noisy proxies for entrepreneurship than just a measurement of self-employment, such as serial self-employment or successful business ownership. Finally, very large datasets (some say that at least 200,000 individuals are needed) may uncover the molecular architecture of entrepreneurship, even when the measurement is self-employment.

Scholars in the social sciences widely adopted the so-called standard social science model, which assumes that the mind is a cognitive device shaped by culture and socialization only. The model implies that variation in economic outcomes, such as human decisions, is the result of nurture (the environment) rather

than nature or the interplay of nurture and nature. The quest for the entrepreneurial gene is one of the first initiatives to introduce biology into the realm of economic outcomes.

Why is the role of genetics in explaining entrepreneurship interesting? Koellinger et al. (2010) give various reasons. The first is simple curiosity. Genetics can help find the origins of individual differences and how they shape behaviors. Second, genetic differences across populations may be identified that will help explain aggregate economic outcomes, such as the share of nascent entrepreneurship. Third, knowledge of the genetics of economic behavior may improve our understanding of the boundaries of economic policies: a poor fit between genetic predisposition and occupational choice may result in an inferior performance. The rapid progress in the field of genetics, the advent of the so-called bio-banks with their extensive datasets and the limited progress in the traditional approaches of the determinants of entrepreneurship point toward a bright future for the quest for the entrepreneurial gene.

References

Apicella, C., Cesarini, D., Johanneson, M., Dawes, C., Lichtenstein, P., Wallace, B., Beauchamp, J. and Westberg, L. (2010), No association between oxytocin receptor (OXTR) gene polymorphisms and experimentally elicited social preferences, *PLoS ONE*, **5**(6), e11153.

Beauchamp, J.P., Cesarini, D., Johannesson, M., van der Loos, M.J.H.M., Koellinger, P.D., Groenen, P.J.F., Fowler, J.H., Rosenquist, J.N., Thurik, A.R. and Christakis, N.A. (2011), Molecular genetics and economics, *Journal of Economic Perspectives*, **25**, 57–82.

Benjamin, D.J., Cesarini, D., van der Loos, M.J.H.M., Dawes, C.T., Koellinger, P.D., Magnusson, P.K.E., Chabris, C.F., Conley, D., Laibson, D., Johannesson, M. and Visscher, P.M. (2012), 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.

Hofman, A., Breteler, M.M.B., van Duijn, C.M., Janssen, H.L.A., Krestin, G.P., Kuipers, E.J., Stricker, B.H.C., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C.M. (2009), The Rotterdam Study: 2010 objectives and design update, *European Journal of Epidemiology*, **24**, 553–572.

International HapMap Consortium (2005), A haplotype map of the human genome, *Nature* **437**, 1299–1320.

Ioannidis, J.P.A. (2005), Why most published research findings are false, *PLoS Medicine*, **2**(8), e124.

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Reibold, M., Laiba, E., Bachner-Melman, R., Maril, A., Bornstein, G., Knafo, A. and Ebstein, R.P. (2009), The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task, *PLoS ONE*, **4**(5), e5535.

Koellinger, P.D., Van der Loos, M.J.H.M., Groenen, P.J.F., Thurik, A.R., Rivadeneira, F., van Rooij, F.J.A, Uitterlinden, A.G. and Hofman, A. (2010), Genome-wide association studies in economics and entrepreneurship research: promises and limitations, *Small Business Economics*, **35**, 1–18.

Lang, U.E., Bajbouj, M., Wernicke, C., Rommelspacher, H., Danker-Hopfe, H. and Gallinat, J. (2004), No association of a functional polymorphism in the serotonin transporter gene promoter and anxiety-related personality traits, *Neuropsychobiology*, **49**, 182–184.

Li, Y., Willer, C.J., Sanna, S. and Abecasis, G.R. (2009), Genotype imputation, *Annual Review of Genomics and Human Genetics*, **10**, 387–406.

Munafo, M.R., Yalcin, B., Willis-Owen, S.A. and Flint, J. (2008), Association of the dopamine D4 receptor (DRD4) gene and approach-related personality traits: meta-analysis and new data, *Biological Psychiatry*, **63**, 197–206.

Nicolaou, N. and Shane, S. (2009), Can genetic factors influence the likelihood of engaging in entrepreneurial activity? *Journal of Business Venturing*, **24**, 1–22.

Nicolaou, N., Shane, S., Cherkas, L., Hunkin, J. and Spector, T.D. (2008a), Is the tendency to engage in entrepreneurship genetic? *Management Science*, **54**, 167–179.

Nicolaou, N., Shane, S., Cherkas, L. and Spector, T.D. (2008b), The influence of sensation seeking in the heritability of entrepreneurship, *Strategic Entrepreneurship Journal*, **2**, 7–21.

Nicolaou, N., Shane, S., Adi, G., Mangino, M. and Harris, J. (2011), A polymorphism associated with entrepreneurship: evidence from dopamine receptor candidate genes, *Small Business Economics*, **36**, 151–155.

Parker, S.C. (2009), *The Economics of Entrepreneurship*, Cambridge: Cambridge University Press.

Rietveld, C.A.& Van der Loos, M.J.H.M. (2012) The quest for the entrepreneurial gene: the importance of study design. *Medium Econometrische Toepassingen* 19(1), 14-21.

Rietveld, C.A. et al. (2013), GWAS of 126,559 individuals identifies genetic variants associated with educational attainment, *Science*, **340**(6139), 1467–1471.

Shane, S. (2010), *Born Entrepreneurs, Born Leaders: How Your Genes Affect Your Work Life*, New York: Oxford University Press.

Van der Loos, M.J.H.M., Koellinger, P.D., Groenen, P.J.F. and Thurik, A.R. (2010), Genome-wide association studies and the genetics of entrepreneurship, *European Journal of Epidemiology*, **25**, 1–3.

Van der Loos, M.J.H.M., Koellinger, P.D., Groenen, P.J., Rietveld, C.A., Rivadeneira, F., van Rooij, F.J.A, Uitterlinden, A.G., Hofman, A. and Thurik, A.R. (2011), Candidate gene studies and the quest for the entrepreneurial gene, *Small Business Economics*, **37**, 269–275.

Van der Loos, M.L.H.M. et al. (2013), The molecular genetic architecture of self-employment, *PLOS One*, **8**(4).

Verheul, I. and Thurik, A.R. (2001), Start-up capital: does gender matter? *Small Business Economics*, **16**: 329–346.

Verheul, I., Thurik, R., Grilo, I. and van der Zwan, P. (2012), Explaining preferences and actual involvement in self-employment: gender and the entrepreneurial personality, *Journal of Economic Psychology*, **33,** 325–341.

Vormfelde, S.V., Hoell, I., Tzvetkov, M., Jamrozinski, K., Sehrt, D., Brockmöller, J. and Leibing, E. (2006), Anxiety- and novelty seeking-related personality traits and serotonin transporter gene polymorphisms, *Journal of Psychiatric Research*, **40**, 568–576.

Wood, A.R. et al. (2014), Defining the role of common variation in the genomic and biological architecture of adult human height, *Nature Genetics*, **46**, 1173–1186.

Zhang, Z., Zyphur, M.J., Narayanan, J., Arvey, R.D., Chaturvedi, S., Avolio, B.J., Lichtenstein, P. and Larsson, G. (2009), The genetic basis of entrepreneurship: effects of gender and personality, *Organizational Behavior and Human Decision Processes*, **110**, 93–107.