Supplementary Online Content

Belsky DW, Moffitt TE, Baker TB, et al. Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence [published online March 27, 2013]. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2013.736.

This supplementary material has been provided by the authors to give readers additional information about their work.

Supplemental Materials to DW Belsky et al. "Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence."

Supplemental Methods.

To select single-nucleotide polymorphisms (SNPs) for inclusion in our genetic risk score (GRS), we examined the results of 3 meta-analyses of genome-wide association studies (GWASs) of smoking quantity.[1-3] We first selected regions with genome-wide significant associations (*P*<.001) in at least 2 of the 3 meta-analyses. Next, we pruned the set of genome-wide significant SNPs using a linkage disequilibrium (LD) threshold of R^2=0.60. This procedure grouped SNPs into clusters on the basis of LD queried from the Hap Map v22 and 1000 Genomes databases (CEU samples) using the Broad Institute's SNAP tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php). From within each cluster, we selected the SNP with the lowest p-value in the meta-analysis by Thorgeirsson and colleagues[1] (the only meta-analysis to report exact p-values for all SNPs).

We then compared the set of 6 SNPs derived as described above to a set of available genotypes from an Illumina BeadPlex Array previously constructed to genotype SNPs identified in GWAS of obesity, asthma, and smoking. We selected the best available matches among SNPs passing quality control (Hardy Weinberg equilibrium >0.01; call rate of >95% in samples with valid calls for >95% of SNPs).

Supplemental Materials to DW Belsky et al. "Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence."

eTable 1. Replication of Association Between the Genetic Risk Score and Smoking Quantity in the Dunedin Cohort and the ARIC and SAGE Databases[a]

	Ordinary Least Squares		Negative Binomial	
	B	95% CI	IRR	95% CI
Dunedin (n=627)				
Full GRS	0.08	(0.03-0.14)	1.16	(1.05-1.29)
GRS excluding proxy for rs16969968	0.09	(0.03-0.14)	1.18	(1.06-1.31)
GRS excluding proxy for rs6495308	0.09	(0.04-0.15)	1.19	(1.07-1.32)
ARIC (n=4,804)				
Full GRS	0.08	(0.05-0.12)	1.07	(1.05-1.10)
GRS excluding proxy for rs16969968	0.07	(0.03-0.11)	1.06	(1.04-1.09)
GRS excluding proxy for rs6495308	0.07	(0.05-0.10)	1.07	(1.04-1.09)
SAGE (n=1,992)				
Full GRS	0.06	(0.02-0.11)	1.09	(1.04-1.15)
GRS excluding proxy for rs16969968	0.04	(0.01-0.08)	1.07	(1.01-1.13)
GRS excluding proxy for rs6495308	0.05	(0.01-0.10)	1.08	(1.03-1.14)

[a]The dependent variable in the analyses is a 4-category measure of lifetime smoking quantity: 0-10 cigarettes/day; 11-20 cigarettes/day; 21-30 cigarettes/day; 31+ cigarettes/day. The categories match those used by the Fagerstrom Test of Nicotine Dependence.[4] Analyses tested the full genetic risk score (gray highlights) and versions of this score that excluded each of the two best-replicated GWAS hits for smoking phenotypes. Effects are presented for a 1-standard deviation increase in the genetic risk scores (GRSs). Two regression approaches were used to analyze the association between genetic risk and smoking quantity: ordinary least squares (OLS), which treated the 4-category smoking quantity measure as a continuous variable, and negative binomial regression, which treated the 4-category measure as a count variable. Regression coefficients (B) are presented from the OLS models. Incident rate ratios (IRRs) are presented from the negative binomial regression models. Analyses include European-descent individuals who ever smoked. The ARIC database included individuals sampled from 3 US regions (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000090.v1.p1). The SAGE database included individuals drawn from 3 separate studies (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1). Analyses of these databases accounted for their

multi-part structure using fixed and random effects in the case of the OLS models and fixed effects in the case of the negative binomial models.

eTable 2. Single-Nucleotide Polymorphisms (SNPs) Included in the Genome Risk Score[a]

Genes	GWAS SNP	Alleles	Effect Allele	Frequency (HapMap), %	Dunedin SNP	LD With GWAS SNP	Alleles	Effect Allele	Frequency (Dunedin), %
CHRNA5, CHRNA3, CHRNB4	rs16969968	A/G	A	39	rs10519203	0.93	A/G	G	34
	rs6495308	C/T	T	80	rs4887069	1.00	A/G	A	79
ADAMTS7, MORF4L1	rs12595538	A/T	A	62	rs7164529	0.90	A/G	G	61
	rs8032771	A/G	A	52	rs11072810	0.97	C/T	T	50
EGLN2	rs7937	C/T	T	55	rs7937	1.00	C/T	T	57
CYP2A6	rs4105144	A/G	G	74	rs8102683	0.87	C/T	C	73

[a]Effect allele frequencies for the genome-wide association study (GWAS) SNPs are based on the HapMap CEU sample (release 22 for SNPs rs12595538, rs8032771, and rs4105144; version 3 release 2 for SNPs rs16969968 and rs6495308). Linkage disequilibrium (LD in terms of R^2) was obtained from 1000 Genomes project data for all SNPs except rs4105144. LD between this SNP and rs8102683 was obtained using HapMap Release 22 data. All allele frequency and linkage queries were run through the Broad Institute’s SNAP tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php). Effect allele frequencies for the SNPs genotyped in the Dunedin sample are based on n=880 European-descent study members.

eTable 3. Correlations Among Measures of Adult Smoking Problems[a]

		(1)	(2)	(3)
(1)	Lifetime cigarette consumption (pack-years)	1.00		
(2)	Total Fagerstr□m symptoms	0.81	1.00	
(3)	Total No. of cessation failures (relapse)	0.52	0.51	1.00

[a]Pearson correlations (*r*) were calculated among n=627 ever smokers. Total Fagerstr□m Symptoms was the sum of Fagerstr□m Test of Nicotine Dependence (FTND) scores across all measurements. All correlations were statistically significant at *P*<.001.

Supplemental Materials to DW Belsky et al. “Polygenic risk accelerates the developmental progression to heavy, persistent smoking and nicotine dependence.”

eTable 4. Associations Between Genetic Risk and Clinical Phenotypes of Smoking Behavior Are Mediated by Developmental Phenotypes of Rapid Progression From Smoking Initiation to Heavy Smoking[a]

Latent Adult Smoking Problems Factor	Total Effect of Genetic Risk	Direct (Unmediated) Effect of Genetic Risk	Indirect Effect of Genetic Risk Mediated Through Developmental Phenotypes	Proportion of Total Effect Accounted for by the Indirect Effect, %
B	0.15	0.03	0.12	81
95% CI	0.05 to 0.26	-0.04 to 0.10	0.05 to 0.20	
P value	<.001	<.001	<.001	

[a]The dependent variable in the mediation analysis was the latent smoking problems factor. Indirect, direct, and total effects were estimated from the structural equation described by MacKinnon and Dwyer implemented using the methods described by Preacher and colleagues. [5-7] Percentile-based 95% CIs were estimated from 1000 bootstrap repetitions. Developmental phenotypes were early conversion to daily smoking (by age 15 years) and rapid progression to heavy smoking (by age 18 years). Both developmental phenotypes were associated with the latent adult smoking problems factor and with the individual clinical phenotypes (*P*<.001 for all). Collectively, early conversion to daily smoking and rapid progression to heavy smoking explained 23% of the variance in the latent smoking problems factor.

References

1. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet.* 2010;42(5):448-453.
2. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet.* 2010;42(5):436-440.
3. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441-447.
4. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstr□m Test for Nicotine Dependence: a revision of the Fagerstr□m Tolerance Questionnaire. *Br J Addict.* 1991;86(9):1119-1127.
5. Mackinnon DP, Dwyer JH. Estimating mediated effects in prevention studies. *Eval Rev.* 1993;17(2):144-158.
6. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Meth.* 2008;40(3):879-891.
7. Preacher KJ, Kelley K. Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychol Methods.* 2011;16(2):93-115.