GENOMICS OF HUMAN AGGRESSION: Current state of genome-wide studies and an automated systematic review tool

Odintsova V.V., Roetman P.J., Ip H.F., Pool R., Van der Laan C.M., Tona K.D., Vermeiren R.R.J.M., Boomsma D.I.

Supplement S3. Additional materials on automated screening

Automated screening of titles and abstracts was performed with use of Automated Systematic Review Software (ASR) developed by researchers from Utrecht University, the Netherlands (PI A.G.J. van de Schoot) for screening abstracts and titles. The software is hosted at https://github.com (*Automated systematic reviews by using Deep Learning and Active Learning*, 2019). ASR is based on supervised machine learning approach with classification approach (the papers are classified in categories—i.e. 1=included or 0=not-included). The oracle modus is used to perform a systematic review with interaction by the reviewer. During the training phase, the model is created, and in the prediction phase, the model is used to predict the future results of a literature search (see Figure S3.1).

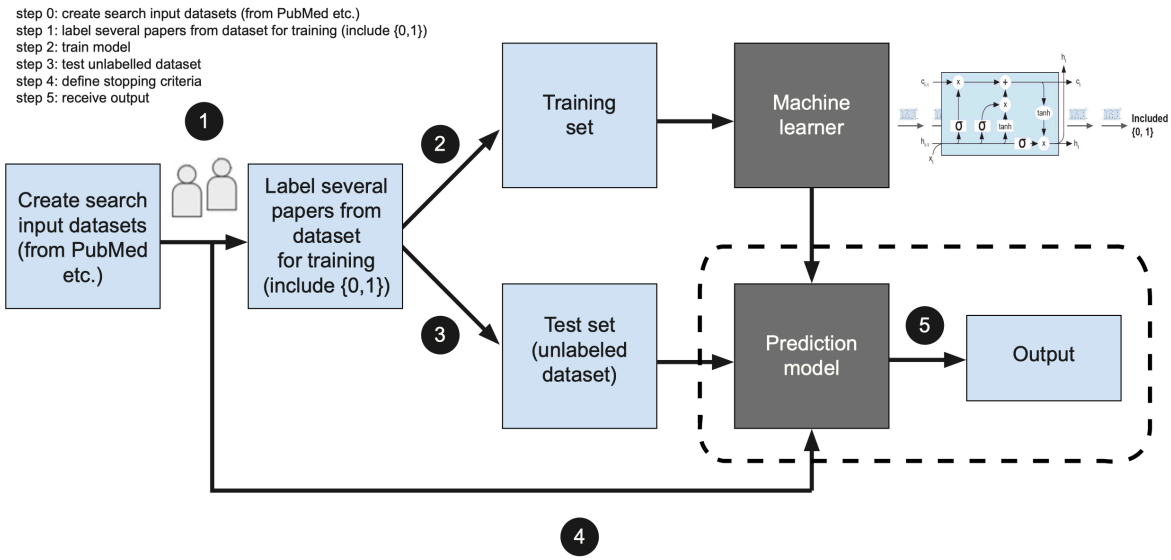

Figure S3.1. Process scheme of training and testing sets using ASR

We had two objectives in applying ASR:

1) To analyze screening parameters of ASR (time of screening, inclusion and exclusion rates, false positive rates (FPR), false negative rates (FNR), true positive rates (TPR), true negative rates (TNR), and receiver operating characteristics (ROC)) and compare it with parameters of manual screening (time of screening, inclusion and exclusion rates as workload characteristics);
2) To contribute to the current systematic review by predicting inclusion/exclusion in a large data set of records based on generated ASR models. To make automated screening of ASR on large dataset of records to make a new contribution to the current systematic review.

The following steps were done in our systematic review:

0. several literature searches were done in PubMed to create a training dataset with key words “human aggression GWAS”, “human aggression genetic association studies”, “human aggression epigenetics” (2,955 records)

1. the training dataset was labelled by reviewers to create training sets (0=not-included, 1=included) and comprised 152 positives and 2803 negatives labels
2. ASR models were trained with training sets from the labelled training dataset (500 records)
3. models with different parameters were used for screening
4. the ROC analyses were performed to define FNR and thresholds of positive and negative results

Receiver operating characteristics (ROC) analyses were performed on the models including different number of records labelled as "included": $N_{label=1}$ = [10, 20, 30, 40, 50, 60, 70] from the randomly selected training set of size $N_{training\ dataset}$ = 500 from the prelabeled list of N = 2,955 records. All models perform considerably better than random, since AUC ∈ [0.79,0.92] (see Figure S3.2). We selected the model where we used $N_{label=1}$ = 50, since it resulted in the minimal FPR=0.39 at FNR ≤ 0.03 with optimal threshold of prediction.

Table S3.1. ROC parameters used for model selection.

$N_{label=1}$	**Minimal false positive rate at FNR≤ 0.03**	**Maximum threshold of prediction at FNR≤ 0.03**
10	0.934363	0.01
20	0.878205	0.03
30	0.604671	0.09
40	0.571186	0.03
50*	**0.386431**	**0.12**
60	0.583788	0.05
70	0.455537	0.06

**The model using $N_{label=1}$ = 50 exhibits the lowest minimal FPR at FNR ≤ 0.03*

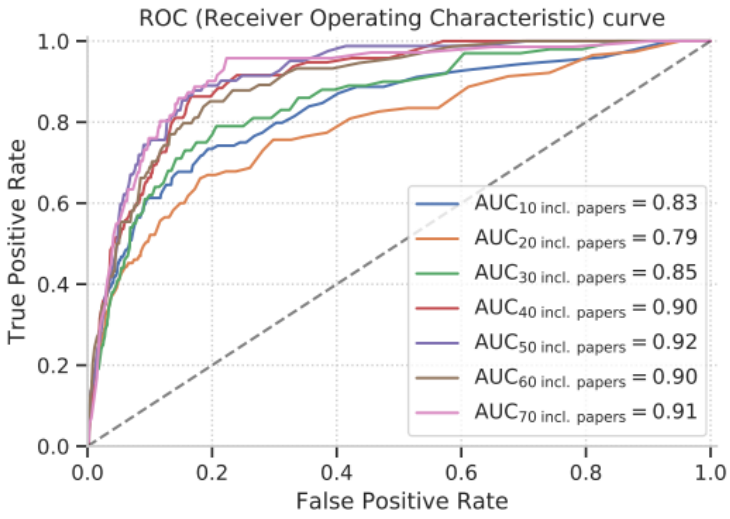

Figure S3.2. ROC curves for the trained models
AUC=area under the curve

Once the optimal model was defined, screenings were repeated on different datasets:

(1) 1,713 records of potential reviews on genetics of human aggression (see Supplement S2);
(2) 356 records of potential GWASs on genetics of human aggression (see Supplement S2);
(3) 2,069 records that join together (1) and (2) datasets;
(4) a new dataset of 14,400 records done with a wide search "humanANDaggressionANDgenes" in the same databases as previous datasets.

Screenings (1)-(3) were used to compare the parameters of automated screening with manual screening (see Table S3.2).

By screening dataset (3) with $N = 2,069$ ASR predicted relevant records and recovered 50 of the 51 expert-labelled true positives, yielding TPR = 0.980. The ASR model mislabeled 1 record as not-relevant from expert labeled true positive, yielding FNR = 0.020. The performance of the model applied to the above search is high. FPR was 0.600, meaning that a reduction in reading time of ~40% is expected.

It is worth noting that model generation and using it for predicting takes ~ 1 hour on a regular computer.

Table S3.2. Comparison of titles and abstracts screening performed manually and automated

Step	Dataset	Screening type	Input Sample	Inclusion*	Inclusion rate	Exclusion	Exclusion rate
Training set	Training dataset	ASR	2,955	152	5,1%	2,803	94,9%
Titles and abstracts screening	Reviews	Manual	1,713	26	1,5%	1,687	98,5%
		ASR	1,713	1,018	59,4%	695	40,6%
	GWASs	Manual	356	25	7,0%	331	93,0%
		ASR	356	243	68,3%	113	31,7%
	"Human aggression genes"	ASR	14,400	7,297	50,7%	7,103	49,3%

Note * The inclusion numbers done on the base of titles and abstracts screening (not the final number of articles included in the review)
ASR=Automated Systematic Review

False-negative result

Sonuga-Barke EJ, Lasky-Su J, Neale BM, Oades R, Chen W, Franke B, et al. Does parental expressed emotion moderate genetic effects in ADHD? An exploration using a genome wide association scan. Am J Med Genet B Neuropsychiatr Genet. 2008;147B(8):1359-68.

Papers selected by researchers from automated selection in addition to traditional selection

Reviews

Baud P. Personality traits as intermediary phenotypes in suicidal behavior: genetic issues. Am J Med Genet C Semin Med Genet. 2005 Feb 15;133C(1):34-42. Review. PubMed PMID: 15648080.

Beaver K.M., Connolly E.J., Nedelec J.L., Schwartz J.A. On the genetic and genomic basis of aggression, violence, and antisocial behavior. Oxford Handbook of Evolution, Biology, and Society. 2018. p.1-18 DOI: 10.1093/oxfordhb/9780190299323.013.15

Davydova J.D., Litvinov S.S., Enikeeva R.F., Malykh S.B., Khusnutdino- va E.K. Recent advances in genetics of aggressive behavior. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2018;22(6):716-725. DOI 10.18699/VJ18.415

Tuvblad C, Beaver KM. Genetic and environmental influences on antisocial behavior. J Crim Justice. 2013;41(5):273–276. doi:10.1016/j.jcrimjus.2013.07.007

Empirical genetic studies

Neumann, A., Pappa, I., Lahey, B. B., Verhulst, F. C., Medina-Gomez, C., Jaddoe, V. W., . . . Tiemeier, H. (2016). Single nucleotide polymorphism heritability of a general psychopathology factor in children. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(12), 1038-1045. e1034.

Criado, J. R., Gizer, I. R., Slutske, W. S., Phillips, E., & Ehlers, C. L. (2012). Event-related oscillations to affective stimuli: heritability, linkage and relationship to externalizing disorders. *J Psychiatr Res, 46*(2), 256-263. doi:10.1016/j.jpsychires.2011.10.017

Dick, D. M., Li, T. K., Edenberg, H. J., Hesselbrock, V., Kramer, J., Kuperman, S., . . . Foroud, T. (2004). A genome-wide screen for genes influencing conduct disorder. *Molecular Psychiatry, 9*(1), 81-86. doi:10.1038/sj.mp.4001368