DOI:
10.1038/mp.2016.85

Document Version
Peer reviewed version

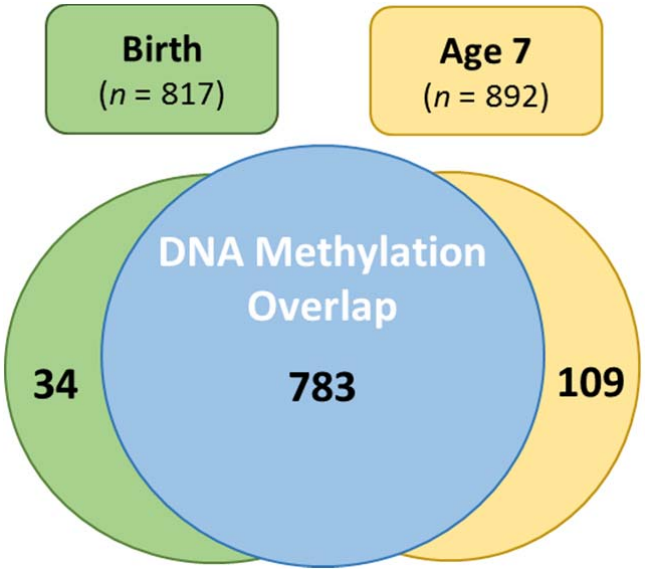
IN PRESS; MOLECULAR PSYCHIATRY

Supplementary Information

Methods

1.1 Sample description and power

For this study, we included youth from the Avon Longitudinal Study of Parents and Children (ALSPAC) – specifically the subsample forming the Accessible Resource for Integrated Epigenomics Studies (ARIES), who had available data on ADHD symptomatology ratings (age 7-15) as well as epigenetic data at birth (n = 817, 49% male) and/or age 7 (n = 892, 50% male). The overlap was n = 783 participants with DNA methylation at birth and age 7 as well as ADHD ratings, while a minority of participants with ADHD symptom ratings had DNA methylation data only at birth (n = 34) or only at age 7 (n = 109; Supplementary Figure 1).



Supplementary Figure 1. DNA methylation samples at birth and age 7 (all participants with ADHD trajectory assignment).

Given the dearth of longitudinal epigenetic research, it is not currently possible to have precise pre-specified effect sizes for power calculation. Therefore, any power calculation is bound to be uncertain, all the more as power calculations in epigenetic studies depend on many factors.[1] However, we note that, based on existing work by Tsai

& Bell,[1] our sample size of 817 appears sufficient to detect 5% methylation differences at an epigenome-wide level, which is within the range of effect sizes reported by other EWAS studies of psychiatric phenotypes, including the smaller, comparable ADHD study by Wilmot et al.[2]

1.2 DAWBA measures

The Development and Well-Being Assessment (DAWBA- www.DAWBA.com; Goodman et al., 2000)[3] is a semi-structured interview administered to parents of children age 4–16, and to children over the age of 11. A shorter version is administered to teachers. The questions for each disorder closely follow the diagnostic criteria in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV)[4] or the International Classification of Diseases (ICD).[5] Each section contains 20-25 questions with skip-rules so that the full set of questions is administered only to children reported to have relevant problems in initial screening questions. When an informant completes a structured section, he/she is invited to answer to open-ended questions. Experienced clinicians can then review all the available information provided by each available informant to propose a diagnosis. This is referred to as "clinician-rated DAWBA diagnoses".[6]

More recently, DAWBA risk bands have been developed.[6] These are based on computer algorithms that use the symptoms and impact information recorded in the structured sections of the DAWBA to generate ordered-categorical measures of the risk of clinician-based disorders. Six levels were created:

- level 0; < 0.1% of children in this band have the disorder in question;

- level 1; ≈ 0.5% of children in this band have the disorder in question;

- level 2; ≈ 3% of children in this band have the disorder in question;

- level 3; ≈15% of children in this band have the disorder in question;

- level 4; ≈ 50% of children in this band have the disorder in question;

- level 5; >70% of children in this band have the disorder in question.

In the present study, we used DSM-IV DAWBA bands derived from parent ratings for the following reasons:

- **Only parent ratings were available across all ages in ALSPAC, with teacher ratings only available at 7 years, and only for a subset of participants.[7] Children were not asked in detail about hyperactivity because of the lower reliability of the information they provide, so that DAWBA bands are not available for children reports.[6] In addition, as reported in the DAWBA bands validation study,[6] prevalence estimates generated from the parent DAWBA bands were generally closer to the clinician-rated prevalence than teacher bands.**
- **The validation study also concluded that information derived from the DAWBA band and clinician ratings correlated similarly to risk factors. The fact that the two approaches generated similar conclusions about the importance of different risk factors for the disorder makes DAWBA bands suitable for large epidemiological studies whose primary interest is to explore the risk factors for a disorder.**
- **Finally, the DAWBA bands also showed a dose-response association with service use and risk factors, including in the low risk bands (0-3), making the DAWBA bands useful to make distinctions in the normal range, which is particularly indicated in general population studies where the prevalence of the disorder is low.**

Supplementary Table 1 provides percentages of children in each band in the British and Norwegian samples used in the validation article, compared to percentages in ALSPAC and ARIES.[6] Results are very similar across the three cohorts except for the band 1, with a higher percentage in ALSPAC/ARIES. However, this band has a risk very close to zero (~0.5%), which makes it hard to distinguish from the band 0 (risk <0.1%), as mentioned in the validation article. When categories 0 and 1 are added, the percentages are very similar in all studies.

Hyperactivity DSM-IV criteria Dawba bands, Parent	B-CAMHS (N = 7,777) %	Norwegian sample (N = 1364) %	ALSPAC (N = 8122) %	AIRIES (N = 817) %
0	80.7	90.8	67.4	69.3
1	7.9	3.7	20.7	20.7
2	4.2	1.9	6.5	5.5

3	**5.0**	**2.8**	**3.9**	**3.6**
4	**1.6**	**0.5**	**0.9**	**0.5**
5	**0.6**	**0.4**	**0.6**	**0.4**

Supplementary Table 1. The B-CAMHS (The British Child and Adolescent Mental Health Surveys) and the Norwegian samples were used in the validation study of the Dawba bands. Mean age for the B-CAMHS was 13.3 years (range 7–19 years). The Norwegian sample was part of the Bergen Child study and included seventh grade (aged 11–13 years) mean age of 12.1 years. For ALSPAC and ARIES, the percentages were averaged over the 4 assessment times from 7 to 15 years.

1.3 DNA methylation preprocessing

The protocol followed manufacturer instructions using the recommended alternative incubation conditions for use with Illumina Infinium arrays. Illumina HumanMethylation450 BeadChips (Illumina, San Diego, USA) were run following the manufacturer's protocol with no modifications and arrays were scanned using an Illumina iScan (software version 3.3.28). Initial quality control of data generated was conducted using GenomeStudio (version 2011.1) to determine the status of staining, extension, hybridization, target removal, bisulfite conversion, specificity, non-polymorphic and negative controls.
Samples were distributed across slides in a semi-random approach to minimise the potential relationship between batch effects and other variables. During the data generation process a wide range of batch variables were recorded in a purpose-built laboratory information management system (LIMS). The LIMS also reported QC metrics from the standard control probes on the 450k BeadChip for each sample back to the laboratory.

1.4 K-means for longitudinal data

K-means for longitudinal data partitions the data into sub-groups following homogeneous developmental trajectories. In this nonparametric procedure, each participant is first assigned arbitrarily to one of two seed trajectories. The center (mean) of each of these two seed trajectories is calculated. Then each participant is reassigned to

the closest of these two seed trajectories. Based on this reassignment, the center is recalculated for each of the two trajectories and participants are reassigned to the closest trajectory. This operation of calculating the center and reassigning participants is repeated until convergence (i.e. until no further change occurs in the trajectories). The process from initial assignment to convergence is then repeated (500 times in the present study) to make sure that the solution is not dependent on the initial assignment (i.e. avoid local maxima). The best solution is determined by the Calinksi-Harabatz criterion, which maximizes a ratio computed by dividing the trace of the between-variance by the trace of the within-variance (i.e. maximizing the differences between trajectories and maximizing the homogeneity within trajectory). Finally, the whole procedure is repeated for a range of solutions, in the present study from 2 to 6 trajectories (as no more progress was achieved by including more trajectories).

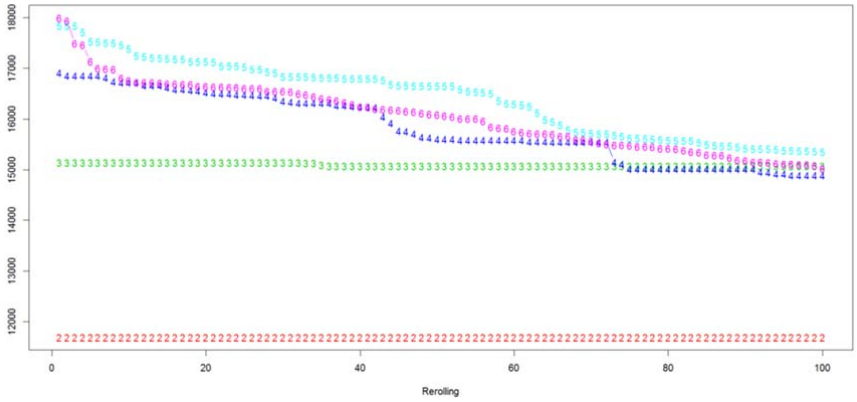
Because this procedure is non-parametric, it does not require any assumption regarding the shape of the trajectory, which can help to fit the data more closely. Indeed, certain shapes cannot be easily approximated by linear, quadratic or even cubic parameters; see ref.[8] for an example where a semi-parametric procedure is unable to recover existing trajectories whilst the non-parametric procedure can. In addition, it does not require any normality or parametric assumptions within clusters.[9] This was important in the present case due to the nature of the variable that was clustered, i.e. DAWBA ordinal 'probability bands'. Various examples of the implementation of this method, including to estimate developmental trajectories of hyperactivity/inattention symptoms, can be found in the following references.[10–13]

1.5 Number of trajectories and binary variable computation

The number of trajectories best representing the data was determined based on two criteria: 1) A fit criterion derived from Calinski-Harabatz and; 2) content of classes, e.g., the number of participants per trajectory. Values of the fit criterion improved steadily from a 2 trajectory solution up to a 4 trajectory solution and improved more modestly after 4 trajectories **(see Supplementary Figure 2 as well as Supplementary Table 2)**. The 4 trajectory solution was not satisfactory because the percentage of participants in the high

trajectory was reduced to 3.8%, which decreases the statistical power to detect associations. **In addition, the global post-probability of accurate classification decreased below 0.90 for the 4 trajectory solution (see Supplementary Table 2 and legend for an explanation). As a result, we selected the 3 trajectory solution for our analyses.**

In addition to the 6-band prediction, the Dawba can be used to predict cases (1 for predicted cases, and 0 for predicted non-cases). At age 7 years, 10 of the 11 predicted cases in the sample were in the high trajectory, with 28% of the participants in this category being predicted cases. Accordingly, participants in the high trajectory were likely either cases or participants with a subthreshold ADHD symptomatology. Conversely, only 1 of the predicted cases was in the low/intermediate trajectories, corresponding to 0.0% of participants in these trajectories being predicted cases. We therefore chose to group these two low trajectories and contrast them with the high trajectory.



Supplementary Figure 2. Trajectories sorted by a fit criterion derived from Calinski-Harabatz. Higher values on the Y axis indicate better fit. There is a clear increase in the criterion values from trajectories 2 to 4, with smaller increase from 4 onwards. Each number corresponds to one trial (from arbitrary assignment to convergence, see supplementary information explanations regarding the procedure). The Figure shows the 100 best solutions (X axis) among the 500 trials.

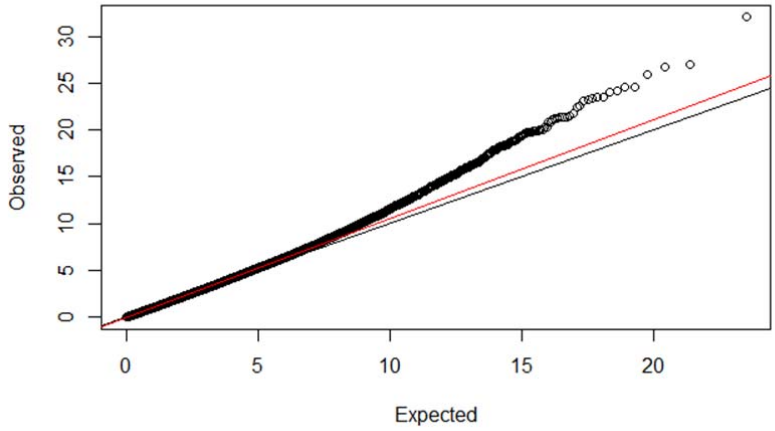
	Number of trajectories				
	2	**3**	**4**	**5**	**6**
Calinksi-Harabatz	**11693.6**	**15143.7**	**16907.6**	**17838.0**	**17989.2**
Post-Proba	**0.98**	**0.93**	**0.89**	**0.87**	**0.85**
Proportion per trajectory (%)					

Trajectory 1	**85.8**	**67.4**	**61.8**	**59.2**	**54.2**
Trajectory 2	**15.2**	**26.5**	**25.0**	**18.0**	**14.1**
Trajectory 3		**6.1**	**9.4**	**10.9**	**11.5**
Trajectory 4			**3.8**	**8.1**	**9.0**
Trajectory 5				**3.8**	**7.6**
Trajectory 6					**3.6**

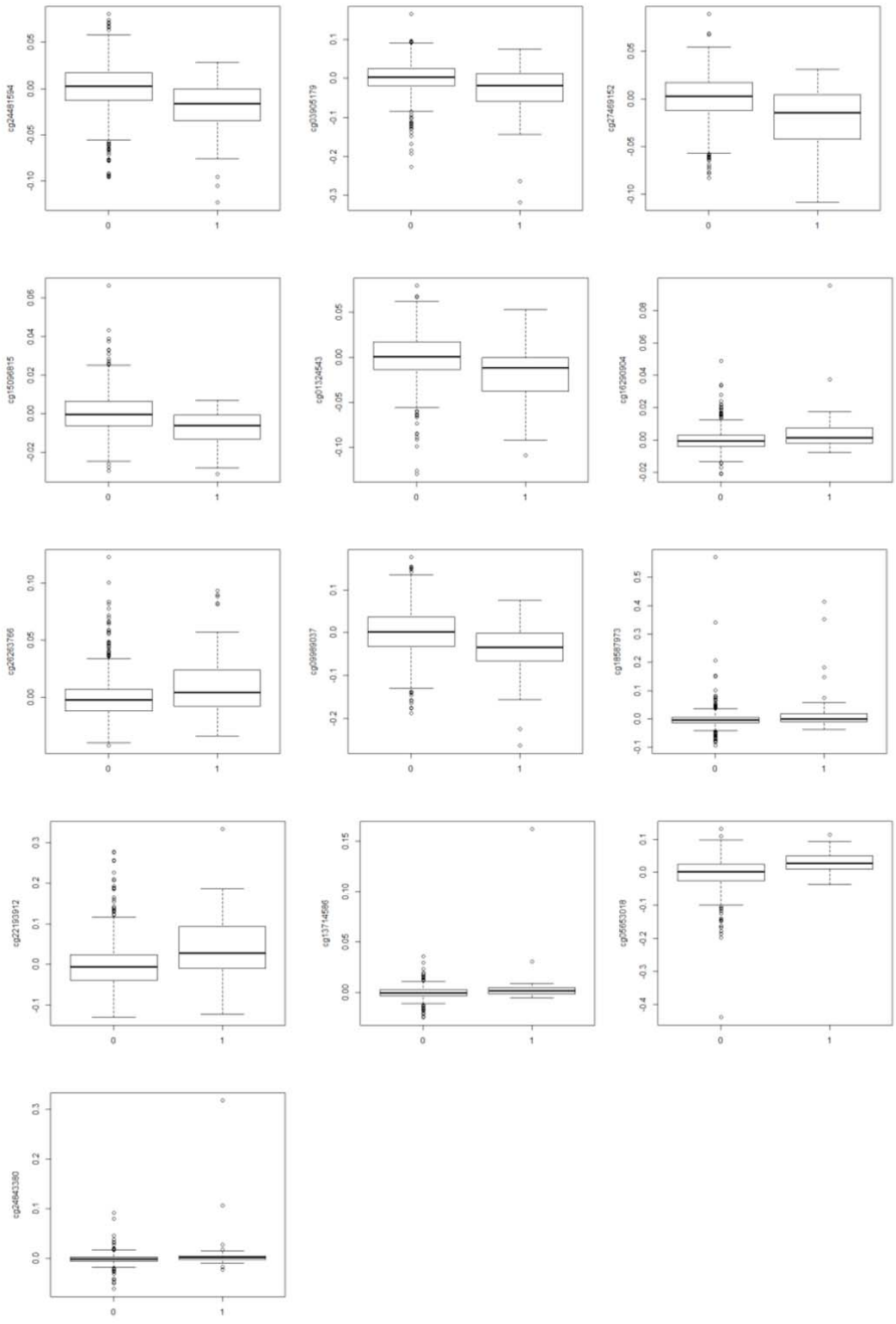
Supplementary Table 2. Fit statistics and proportion of sample per trajectory. The global post-probability (Post-proba) is the global average of post-probabilities and is an index of classification accuracy (similar to entropy in parametric analyses). The post-probability is the average across all participants of each participant's probability of being effectively classified in the trajectory he/she is classified in. For instance, in a two trajectory solution, a participant with very high and consistent ADHD symptoms will have a post-probability close to 100% of being classified in the high trajectory, and a probability close to 0% of belonging to the low trajectory. Conversely, for a participant who is at equal distance from the high and the low trajectory, the corresponding probabilities will be 50/50%. The global post-probability is the average of the highest probability for each participant (here average of 100% and 50%). The global post-probability decreases as the number of trajectories increases. This is because, as the number of trajectories increases, they get closer to each other and the classification becomes ambiguous for more participants.

Results

2.1 QQ-plot ***and boxplots***



Supplementary Figure 3. QQ-plot of observed vs expected p-values indicated no appreciable inflation of test statistics. Lambda = 1.056.


cg24481594
cg03905179
cg27469152
cg15096815
cg01324543
cg16290904
cg26263766
cg09969037
cg18587973
cg21193912
cg13714586
cg25653018
cg24643360

Supplementary Figure 4. Boxplots indicated no strong evidence for a violation of equal variance assumption between trajectories (0 = low vs 1 = high) for any of the 13 probes, which passed FDR-correction.

2.2 Network analysis

As further described in ref,[14,15] GeneMANIA uses a basic weighting method to derive a functional association network based on an input gene list. Weights are chosen automatically using linear regression, to make genes, which are part of the input list, interact as much as possible with each other, and as little as possible with genes not on that list (non-query genes). Then, a variation of Gaussian field label propagation[16,17] is used to score all non-query genes. The score assigned to each gene reflects how often paths that start at a given gene node end up in one of the query nodes and how long and heavily weighted those paths are. Functional enrichment is assessed using Gene Ontology categories and Q-values derived from a FDR corrected hypergeometric test for enrichment using the Benjamini-Hochberg procedure (Supplementary Table 3). Results were not adjusted for a custom-based background.

Function	FDR	Coverage
protein import into peroxisome matrix	8.26E-03	3/12
integral component of peroxisomal membrane	8.26E-03	3/12
intrinsic component of peroxisomal membrane	8.26E-03	3/13
protein localization to peroxisome	1.17E-02	3/18
establishment of protein localization to peroxisome	1.17E-02	3/18
protein targeting to peroxisome	1.17E-02	3/18
peroxisomal transport	1.19E-02	3/19
transcription corepressor activity	1.85E-02	5/162
intracellular protein transmembrane import	2.18E-02	3/25
intracellular protein transmembrane transport	2.79E-02	3/28
peroxisome organization	3.47E-02	3/31
protein transmembrane transport	4.22E-02	3/34
peroxisomal membrane	5.93E-02	3/40
microbody membrane	5.93E-02	3/40
regulation of DNA-templated transcription in response to stress	5.96E-02	3/41
transcription coactivator activity	7.29E-02	5 / 249

Supplementary Table 3. Genetic network analysis by GeneMANIA. "Function" refers to Gene Ontology (GO) terms enriched among the genes in the network displayed by GeneMANIA. Enriched terms are derived based on genes from the input list and non-query genes. Terms related to peroxisomal processes were driven by *PEX2*, which was the only directly measured gene in this pathway. The involvement of *PEX10* and *PEX12* was indirectly inferred based on shared protein domains and physical interactions.

2.3 Sensitivity analyses

Effect sizes at birth were consistent in direction and size in an analyses in a subsample with complete DNA methylation data at both time points (n = 783; Supplementary Table 4)

CpG probe	Gene	Chr	Position	Birth		Age 7	
				StdB	*p*	*StdB*	*p*
cg24481594	*SKI*	1	2190850	-0.172	1.70E-06	0.020	0.587
cg03905179	*MAFK*	7	1582588	-0.177	8.01E-07	0.010	0.779
cg27469152	*EPX*	17	56282313	-0.180	4.43E-07	-0.098	0.007
cg15096815	*JUN*	1	59249838	-0.183	2.54E-07	0.012	0.736
cg01324543	*CCDC30*	1	42999439	-0.149	3.17E-05	0.014	0.706
cg16290904	*PEX2*	8	77912348	0.188	1.31E-07	-0.053	0.138
cg26263766	*ZNF544*	19	58739734	0.159	8.98E-06	0.004	0.911
cg09989037	*ST3GAL3*	1	44300942	-0.145	5.13E-05	-0.064	0.076
cg18587973	*CDADC1*	13	49822535	0.128	3.75E-04	0.005	0.893
cg22193912	*MAFG*	17	79881523	0.165	3.97E-06	0.054	0.134
cg13714586	*FBXW5*	9	139838358	0.166	3.87E-06	-0.033	0.357
cg05653018	*ELF3*	1	201979533	0.168	2.59E-06	0.064	0.075
cg24843380	*ZNF454*	5	178367827	0.170	2.25E-06	0.025	0.492

Supplementary Table 4. Sensitivity analysis in a subsample with complete DNA methylation data at both time points (n = 783)

References

1 Tsai P-C, Bell JT. Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. *Int J Epidemiol* 2015; : dyv041.

2 Wilmot B, Fry R, Smeester L, Musser ED, Mill J, Nigg JT. Methylomic analysis of salivary DNA in childhood ADHD identifies altered DNA methylation in VIPR2. *J Child Psychol Psychiatry* 2015. doi:10.1111/jcpp.12457.

3 Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000; **41**: 645–655.

4 American Psychiactric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR®*. American Psychiatric Pub, 2000.

5 Organization WH, others. *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization, 1992http://apps.who.int/iris/handle/10665/37958 (accessed 18 Feb2016).

6 Goodman A, Heiervang E, Collishaw S, Goodman R. The 'DAWBA bands' as an ordered-categorical measure of child mental health: description and validation in British and Norwegian samples. *Soc Psychiatry Psychiatr Epidemiol* 2011; **46**: 521–532.

7 Scott N, Blair PS, Emond AM, Fleming PJ, Humphreys JS, Henderson J *et al.* Sleep patterns in children with ADHD: a population-based cohort study from birth to 11 years. *J Sleep Res* 2013; **22**: 121–128.

8 Genolini C, Falissard B. KmL: k-means for longitudinal data. *Computational Statistics* 2010; **25**: 317–328.

9 Genolini C, Falissard B. KmL: a package to cluster longitudinal data. *Comput Methods Programs Biomed* 2011; **104**: e112–121.

10 Pingault J-B, Côté S, Galéra C, Genolini C, Falissard B, Vitaro F *et al.* Childhood trajectories of inattention, hyperactivity and oppositional behaviors and prediction of substance abuse/dependence: a 15-year longitudinal population-based study. *Mol Psychiatry* 2013; **18**: 806–812.

11 Genolini C, Pingault JB, Driss T, Côté S, Tremblay RE, Vitaro F *et al.* KmL3D: a non-parametric algorithm for clustering joint trajectories. *Computer methods and programs in biomedicine* 2013; **109**: 104–111.

12 Pingault J-B, Tremblay RE, Vitaro F, Japel C, Boivin M, Côté SM. Early Nonparental Care and Social Behavior in Elementary School: Support for a Social Group Adaptation Hypothesis. *Child Development* 2015; **86**: 1469–1488.

13 Abdennour M, Reggio S, Le Naour G, Liu Y, Poitou C, Aron-Wisnewsky J *et al.* Association of adipose tissue and liver fibrosis with tissue stiffness in morbid obesity: links with diabetes and BMI loss after gastric bypass. *The Journal of Clinical Endocrinology & Metabolism* 2014; **99**: 898–907.

14 Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010; **38**: W214–220.

15 Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 2008; **9 Suppl 1**: S4.

16 Zhou D, Bousquet O, Lal TN, Weston J, Schölkopf B. Learning with local and global consistency. *Advances in neural information processing systems* 2004; **16**: 321–328.

17 Zhu X, Ghahramani Z, Lafferty J, others. Semi-supervised learning using gaussian fields and harmonic functions. In: *ICML*. 2003, pp 912–919.