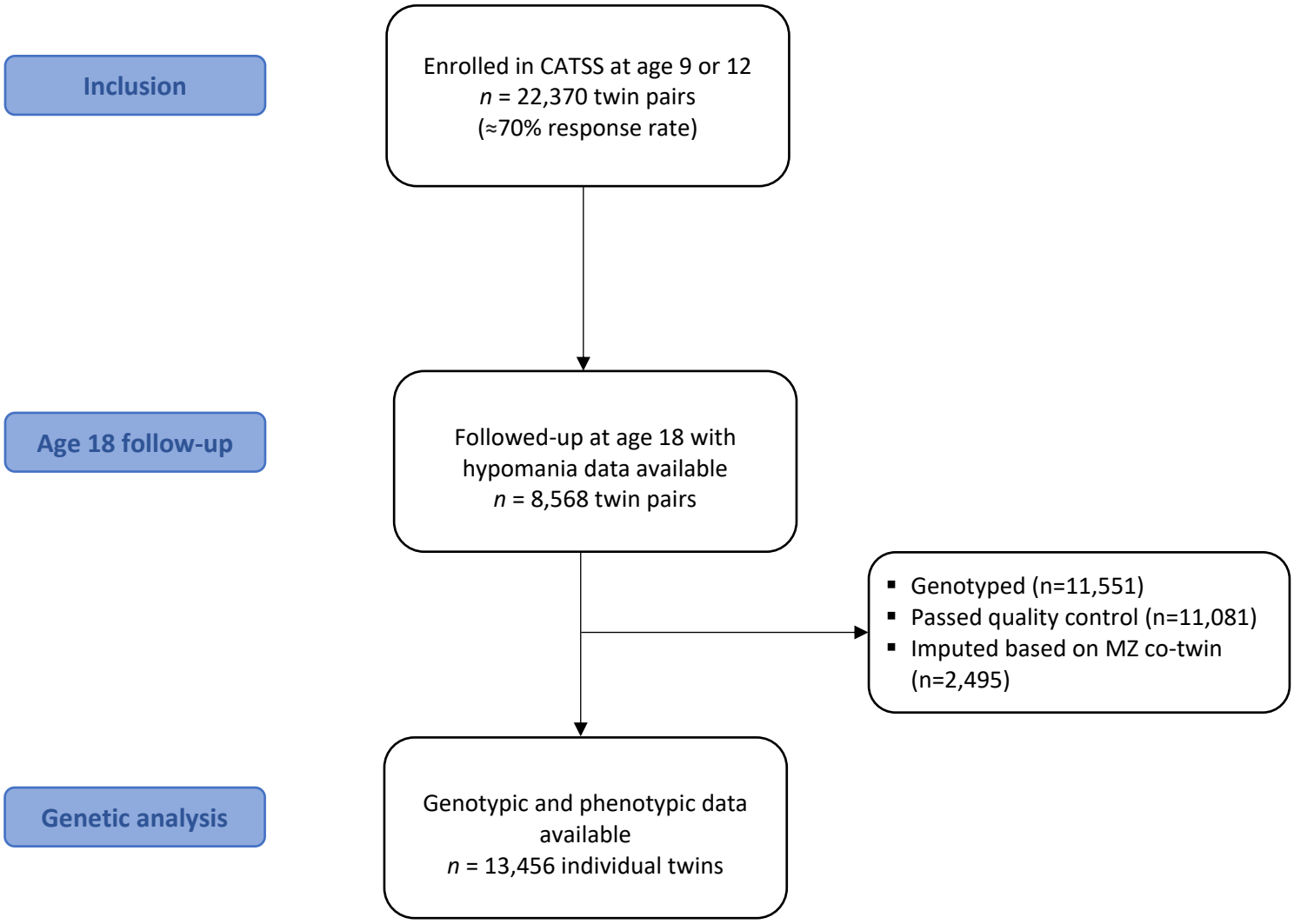
Supplementary Materials

Hosang, GM, Martin, J, Karlsson, R, Ronald, A, Lundström, S, Larsson, H, Lichtenstein, P & Taylor, MJ. Do subsyndromal hypomanic symptoms show etiological similarity to bipolar disorder and other severe mental illnesses? Twin and polygenic risk score analysis

Supplementary Figure 1. Strobe flow diagram of CATSS participation



Polygenic risk score calculation

The CATSS cohort is too young to have participated in discovery samples for schizophrenia, major depressive disorder and bipolar disorder, moreover CATSS data has only recently been genotyped[1] and not contributed to any GWAS consortia. Therefore, the discovery and target data are independent.

Each discovery GWAS results file was filtered to retain high quality common variants, based on available information for minor allele frequency (MAF>0.01) and imputation quality (INFO>0.8). Indels, symmetric/ strand-ambiguous (A/T, T/A, C/G, G/C), multi-allelic and duplicate position SNPs were excluded. SNPs were matched across the target and discovery data based on chromosomal position and alleles. Finally, markers were restricted to those present in the HapMap3 reference sample[2].

Filtered discovery datasets were processed with PRS-CS[3] (version Jun 4, 2021) to generate posterior SNP effects accounting for LD structure and genetic architecture. PRS-CS was run with default values for parameters *a* (1), and *b* (0.5), and automatic estimation of the global shrinkage parameter *phi*. Estimated *phi* values (weighted means of per-chromosome estimates, with number of HapMap3 markers per chromosome as weights) were 1.07e-04 (MDD), 1.76e-04 (SCZ), 1.23e-04 (BIP), 1.20e-04 (BDI), and 9.68e-05 (BDII). The total sample size of each discovery GWAS was used for the sample size parameter. To account for LD structure, we used precomputed LD information provided with PRS-CS, based on the European ancestry subset of the 1000 genomes [4] phase 3 reference sample.

PRS were then calculated for each CATSS participant by scoring the number of effect alleles (weighted by the PRS-CS posterior SNP effect) across each discovery set of SNPs in

PLINK2 v2.00a3LM (28 Mar 2021)[5] (using the --score command). Scores were derived in imputed genotype data (dosage format) after filtering out SNPs with MAF<0.01 and INFO<0. 8 and restricting to HapMap3 SNPs. The number of markers used for scoring (i.e., present and passing quality filters in discovery GWAS, CATSS, and HapMap3) ranged from 949,081 to 963,617. The PRS were standardized using z-score transformations; effect sizes can be interpreted as increase in risk of the outcome, per standard deviation increase in PRS. Nagelkerke R^2 differences between null and full models were calculated to obtain estimates of variance explained.

Principal components analysis

Population covariates were derived from principal components analysis (PCA) in all individuals. First, common autosomal genotyped SNPs passing all quality control were LD-pruned once SNPs located in long-range LD regions were removed. Next, relatives were identified using an identity-by-descent analysis and one of each pair of related individuals (pi-hat>0.2) were temporarily excluded. Allele frequencies were obtained for the set of LD-pruned SNPs for unrelated individuals. PCs were then estimated by calculating variant weights on unrelated individuals and then projecting remaining samples to the PC scales set by these unrelated individuals. All analyses were carried out using PLINK.v.1.9. The first 10 principal components were used as covariates for subsequent analyses.

1. Brikell I, Larsson H, Lu Y, et al. The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. *Mol Psychiatry*. 2020;25(8):1809-1821. doi:10.1038/s41380-018-0109-2

2. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010;467(7311):52-58. doi:10.1038/nature09298

3. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776. doi:10.1038/s41467-019-09718-

4. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

5. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(1). doi:10.1186/s13742-015-0047-8

Supplementary Table 1. Univariate assumptions testing

Model	-2LL	Parameters	df	AIC	Comparison Model	$\Delta\chi^2$	Δdf	p-value
Hypomania								
Fully Saturated	22837.45	25	8276	6285.45	-----	-----	-----	-----
Submodel 1	22838.54	21	8280	6278.54	Fully Saturated	1.10	4	0.895
Submodel 2	22850.45	17	8284	6282.45	Fully Saturated	13.00	8	0.112
Submodel 3	22854.65	15	8286	6282.65	Fully Saturated	17.20	10	0.07
Submodel 4	22884.54	13	8288	6308.54	Fully Saturated	47.09	12	<0.001
Bipolar Disorder								
Fully Saturated	1129.78	6	35612	-70094.22	-----	-----	-----	-----
Submodel 1	1133.48	4	35614	-70094.52	Fully Saturated	3.70	2	0.157
Submodel 2	1133.54	3	35615	-70096.46	Fully Saturated	3.76	3	0.288

All submodels were compared to the fully saturated model using the likelihood-ratio test, with a p-value below 0.05 indicating a statistically significant violation of a given assumption.
For hypomania, the following assumptions were tested: equal means within twin pair (Submodel 1), equal variances within twin pairs (Submodel 2), equal means across zygosity (Submodel 3), and equal variances across zygosity (Submodel 4). For bipolar disorder, the following assumptions were tested: equal thresholds (i.e. prevalence) within twin pairs (Submodel 1) and equal thresholds across zygosity (Submodel 2)

Supplementary Table 2. Twin model fit statistics

Model	-2LL	Parameters	df	AIC	Comparison Model	$\Delta\chi^2$	Δdf	p-value
Hypomania								
Fully Saturated	22837.45	25	8276	6285.45	-----	-----	-----	-----
ACE	22903.06	9	8292	6319.06	Fully Saturated	65.62	16	<0.001
ADE	22908.05	9	8292	6324.05	Fully Saturated	70.60	16	<0.001
Quantitative	22903.06	8	8293	6317.06	ACE	0.00	1	1.00
Homogeneity	22926.09	5	8296	6334.09	Quan	23.03	3	<0.001
AE	22919.63	6	8295	6329.63	Quan	16.56	2	<0.001
CE	23019.37	6	8295	6429.37	Quan	116.31	2	<0.001
E	23547.17	4	8297	6953.17	Quan	644.1	4	<0.001
Bipolar Disorder								
Fully Saturated	1129.78	6	35612	-70094.22	-----	-----	-----	-----
ACE	1133.86	4	35616	-70098.14	Fully Saturated	4.08	4	0.396
AE	1133.86	3	35617	-70100.14	ACE	0.00	1	1.00
CE	1144.97	3	35617	-70089.03	ACE	11.12	1	0.001
E	1184.17	2	35618	-70051.83	ACE	50.31	2	<0.001
Hypomania and bipolar disorder								
Saturated Model	23961.43	24	43895	-63828.57	-----	-----	-----	-----
ACE	23994.87	11	43910	-63825.13	Saturated Model	33.44	15	0.004
AE	23994.87	8	43913	-63831.13	ACE	0	3	1.00
CE	24132.19	8	43913	-63693.81	ACE	137.32	3	<0.001
E	24672.43	5	43916	-63159.57	ACE	677.56	6	<0.001

Supplementary Table 3. ACE joint categorical-continuous bivariate model between hypomania and bipolar disorder

	rPH	rA	rC	rE
Hypomania and bipolar disorder	0.38 (0.29-0.47)	0.40 (0.21-0.73)	-0.95 (-1.00-1.00)	0.41 (0.03-0.75)

Using an ACE joint categorical-continuous bivariate model, the hypomania-BD phenotypic correlation was found to be mainly explained by genetic factors (72%, 95 confidence intervals [CI] 39%-110%) with a smaller contribution from unique environmental factors (28%, 95% CI 2-59%), shared environmental influences contributions were negligible.