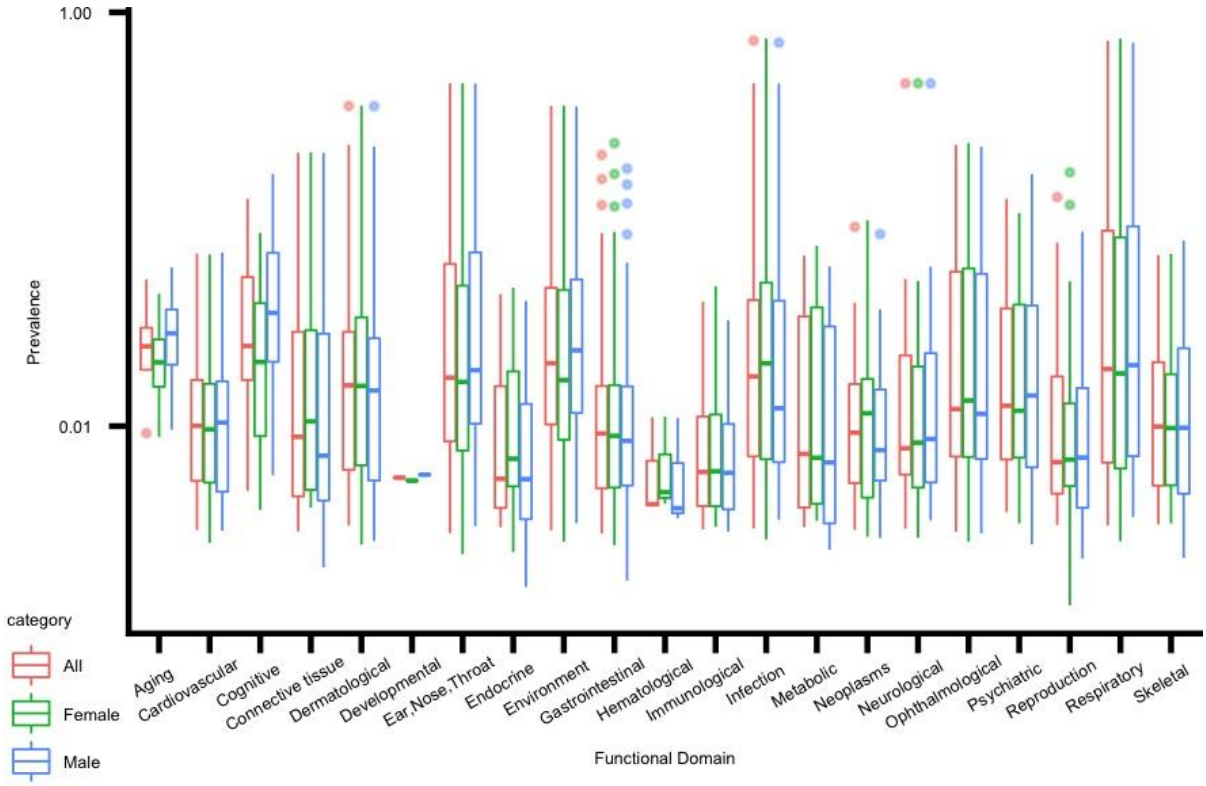
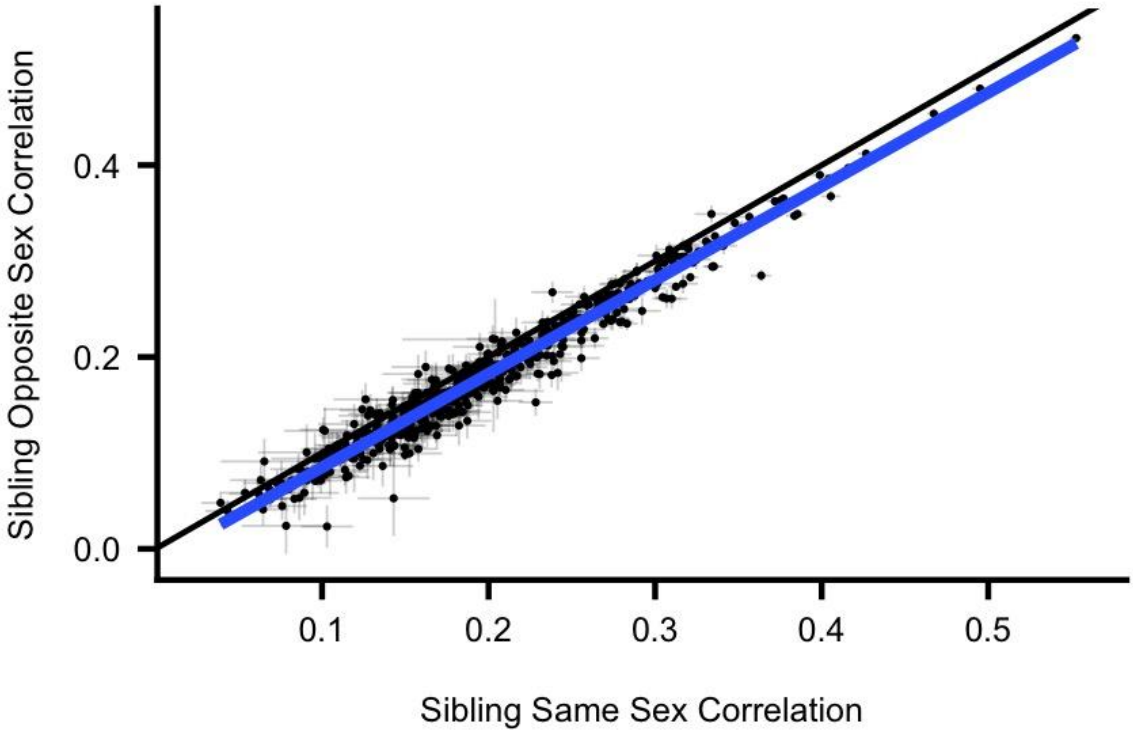
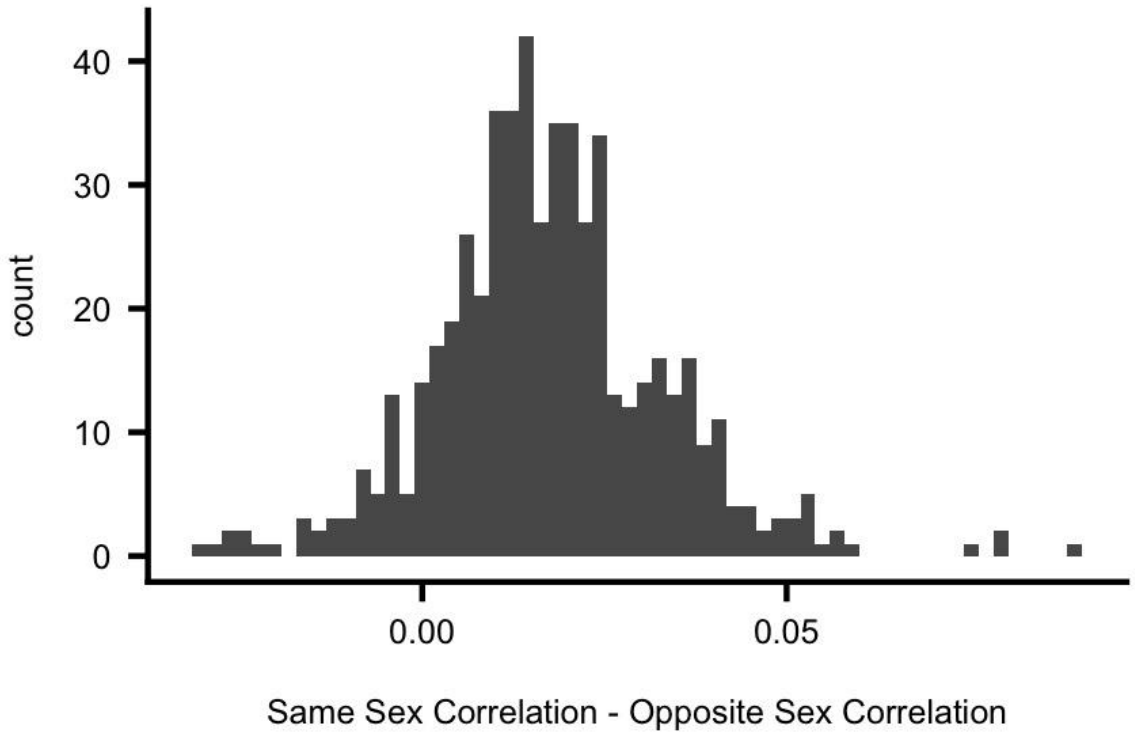
Supplementary Figures



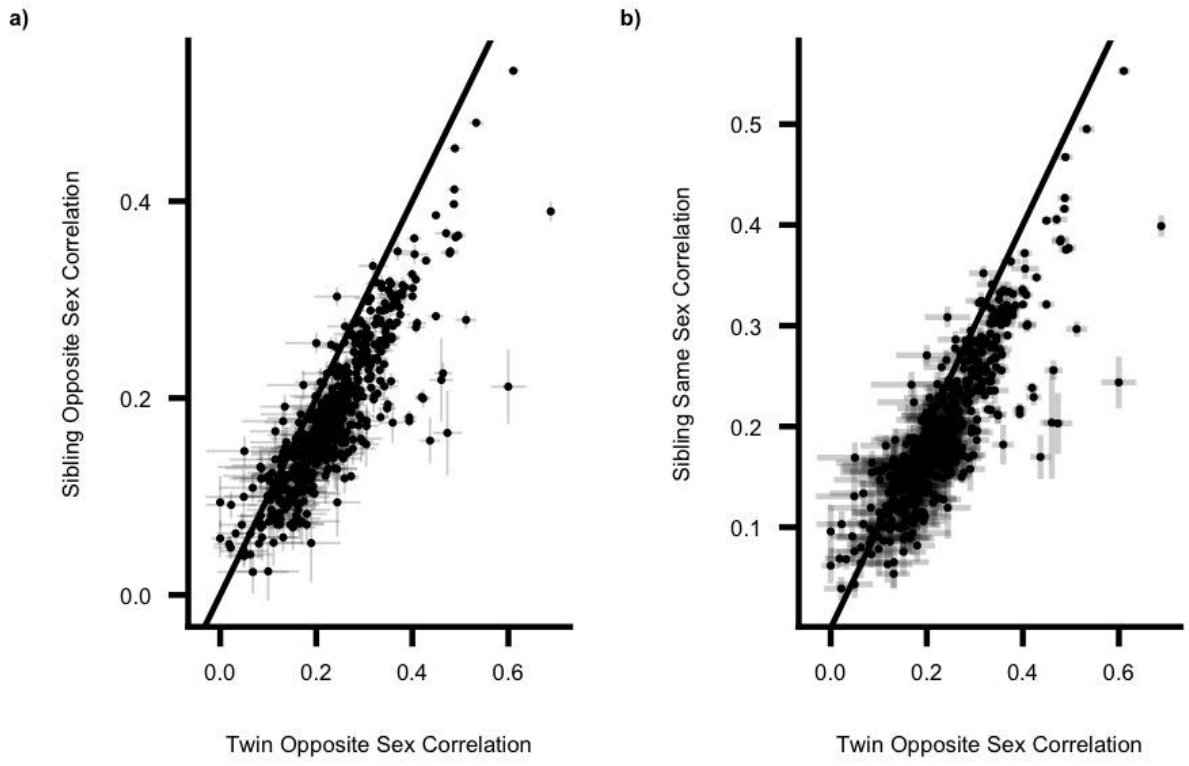
Supplementary Figure 1: Prevalence of phenotypes among different functional domains. Boxplot of prevalence values for 56,396 twin pairs in CaTCH among 560 phenotypes stratified by all genders, just males, and just females. Bottom line, center line, and top line represent 25th percentile, median, and 75th percentile of prevalence values for functional domain and gender strata, respectively. Bottom and top whiskers represent smallest and largest value within 1.58 times IQR range for functional domain and gender strata, respectively, respectively. Dots represents phenotypes where prevalence is outside 1.58 times IQR range for each functional domain and gender strata.



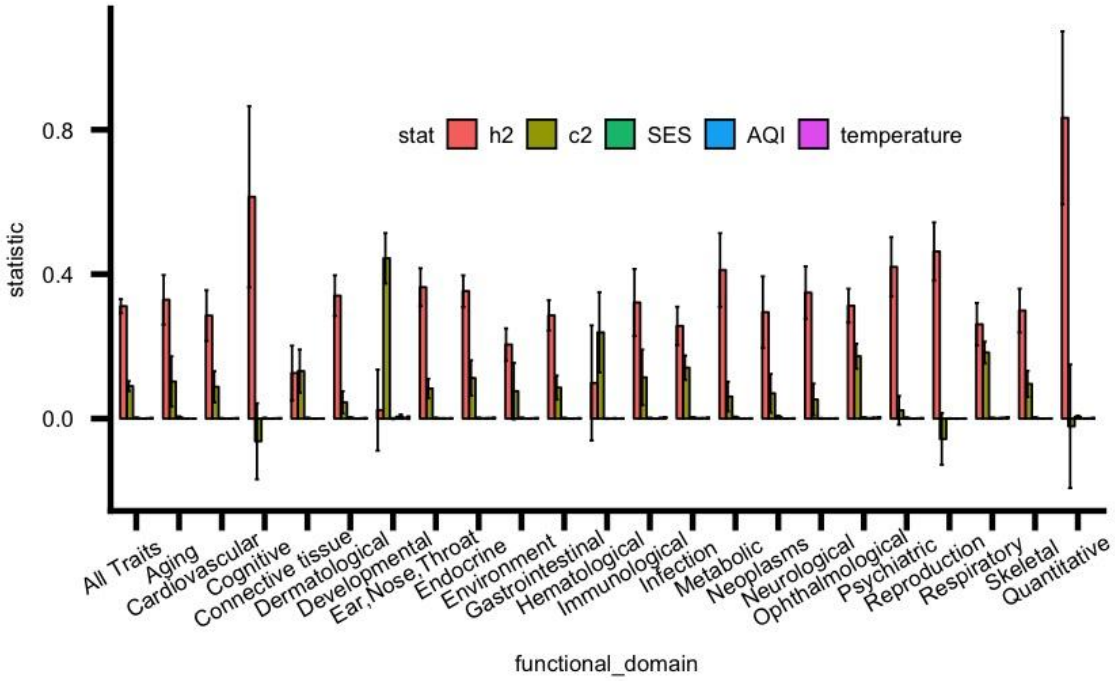
Supplementary Figure 2: Comparison of sibling correlations. Scatterplot comparing estimate of sibling same sex correlation (r_{sibSS}) versus estimate of sibling opposite sex correlation(r_{sibOS}) based on 724,513 sibling pairs for 551 binary phenotypes. Black line represents line with slope 1 and intercept 0, blue line is line of best fit, and horizontal and error bars represent 95% CI for r_{sibSS} and r_{sibOS}, respectively.



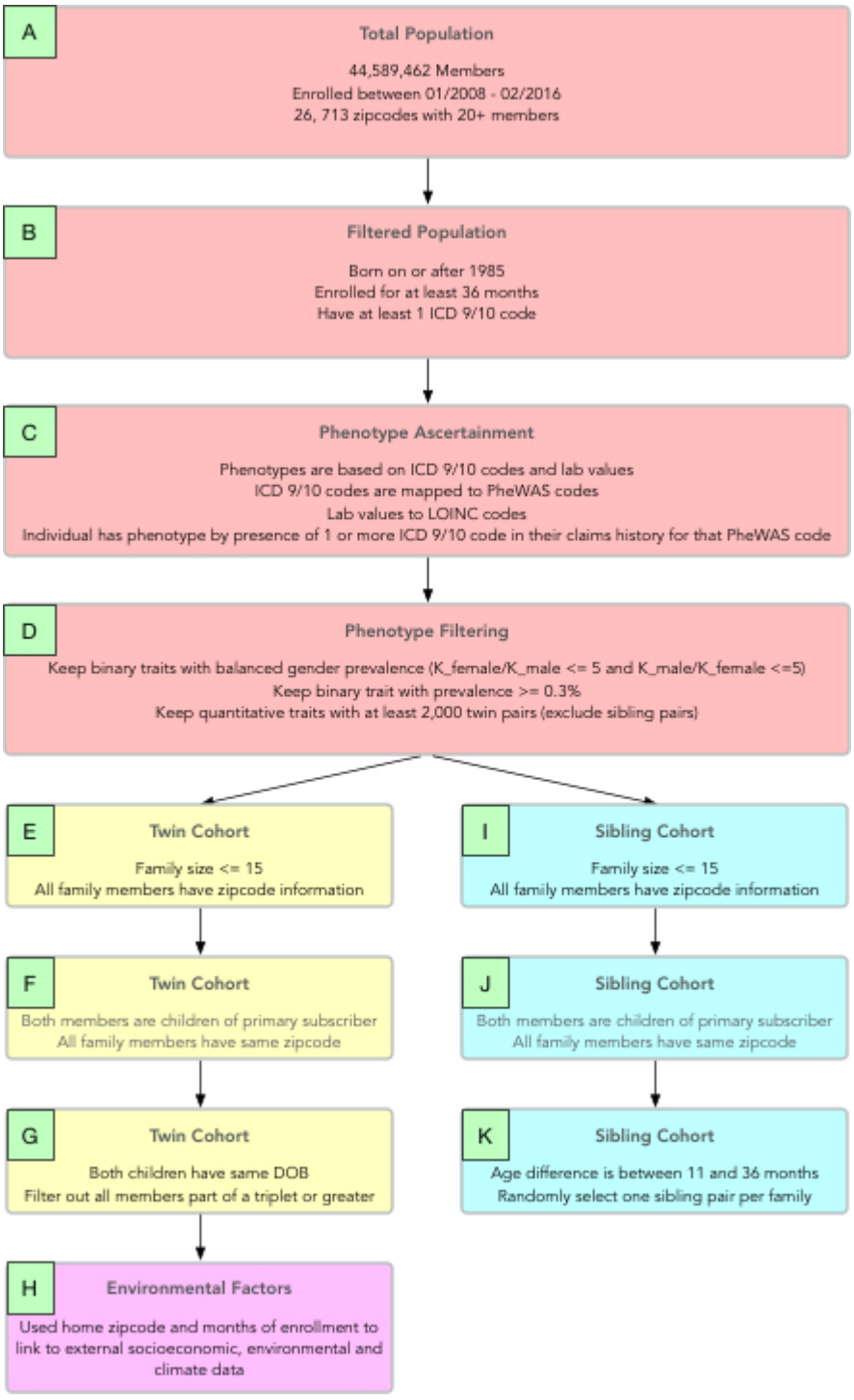
Supplementary Figure 3: Distribution of difference between sibling correlations. Histogram of difference between estimate of sibling same sex correlation (r_{sibSS}) and estimate of sibling opposite sex correlation (r_{sibOS}) using 724,513 sibling pairs for 551 binary phenotypes.



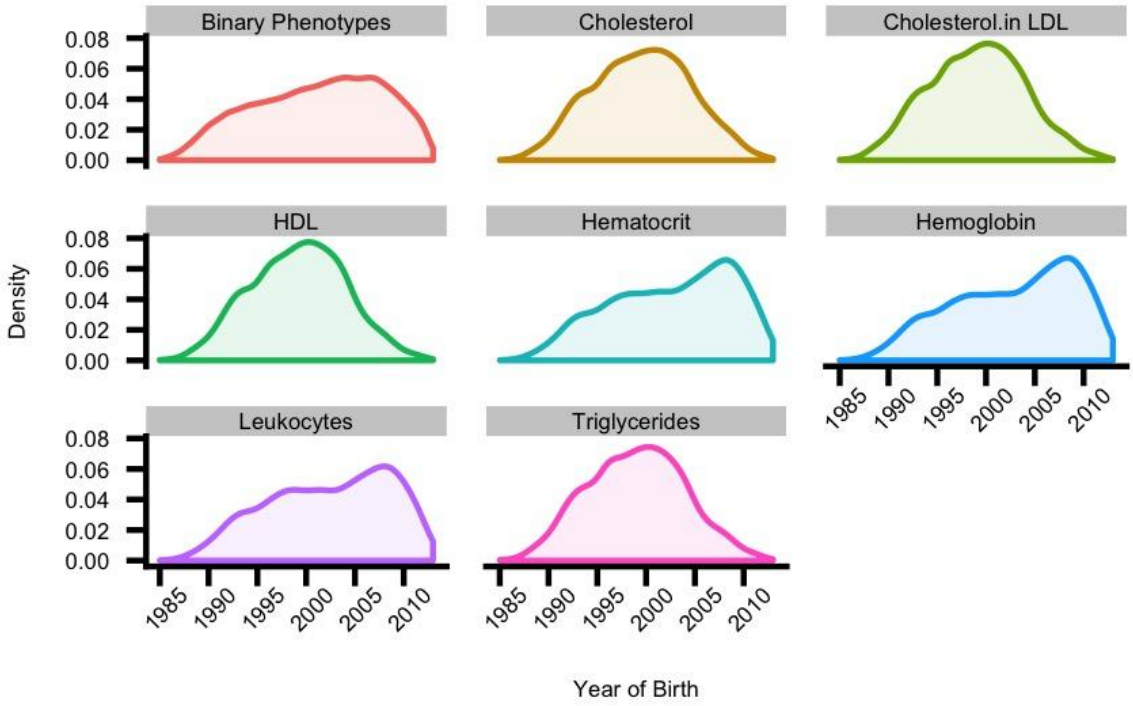
Supplementary Figure 4: Comparison of sibling versus twin correlations. (a) Scatterplot comparing estimate of opposite sex sibling correlation (r_{sibOS}) based on 724,513 sibling pairs versus estimate of opposite sex twin correlation(r_{twinOS}) based on 56,396 twin pairs for 551 binary phenotypes. Black line represents line with slope 1 and intercept 0 and horizontal and error bars represent 95% CI for r_{sibOS} and r_{twinOS}, respectively. **(b)** Scatterplot comparing estimate of same sex sibling correlation (r_{sibSS}) based on 724,513 sibling pairs versus estimate of opposite sex twin correlation(r_{twinOS}) based on 56,396 twin pairs for 551 binary phenotypes. Black line represents line with slope 1 and intercept 0 and horizontal and error bars represent 95% CI for r_{sibSS} and r_{twinOS}, respectively.



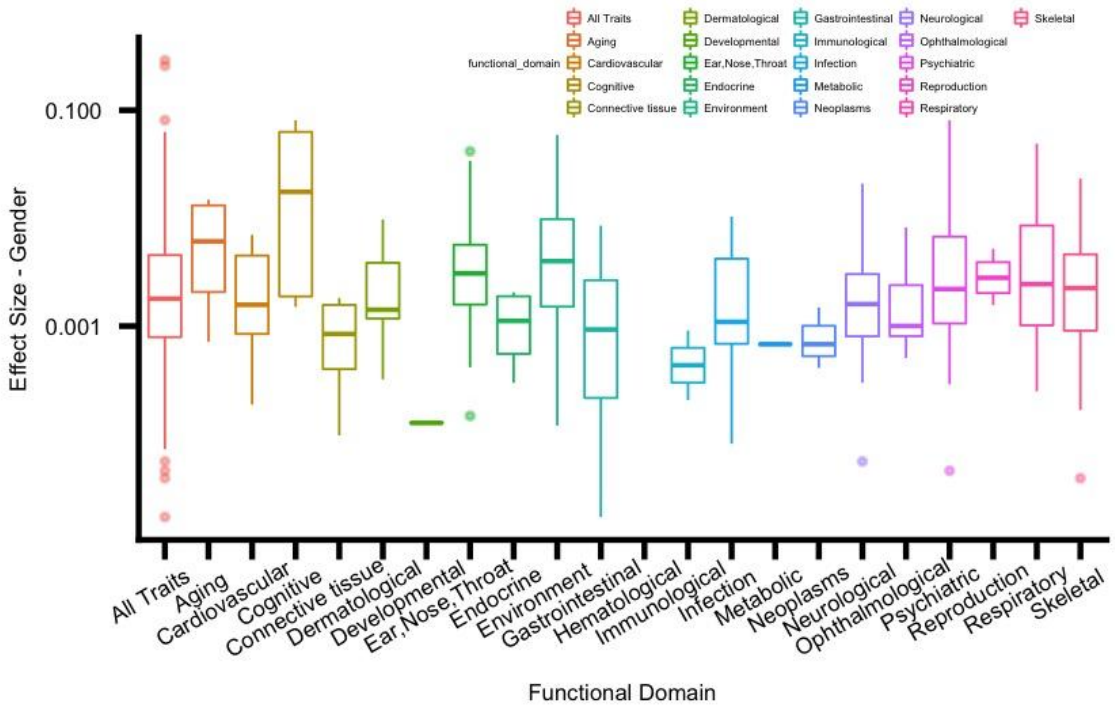
Supplementary Figure 5: Estimates of twin statistics across functional domains for 56,396 twin pairs in CaTCH among all 560 phenotypes. Barplot of meta-analytic estimates of h^2, c^2, var_{SES}, var_{AQI}, and var_{temp} among all 560 phenotypes and within functional domains, error bars represent 95% CI.



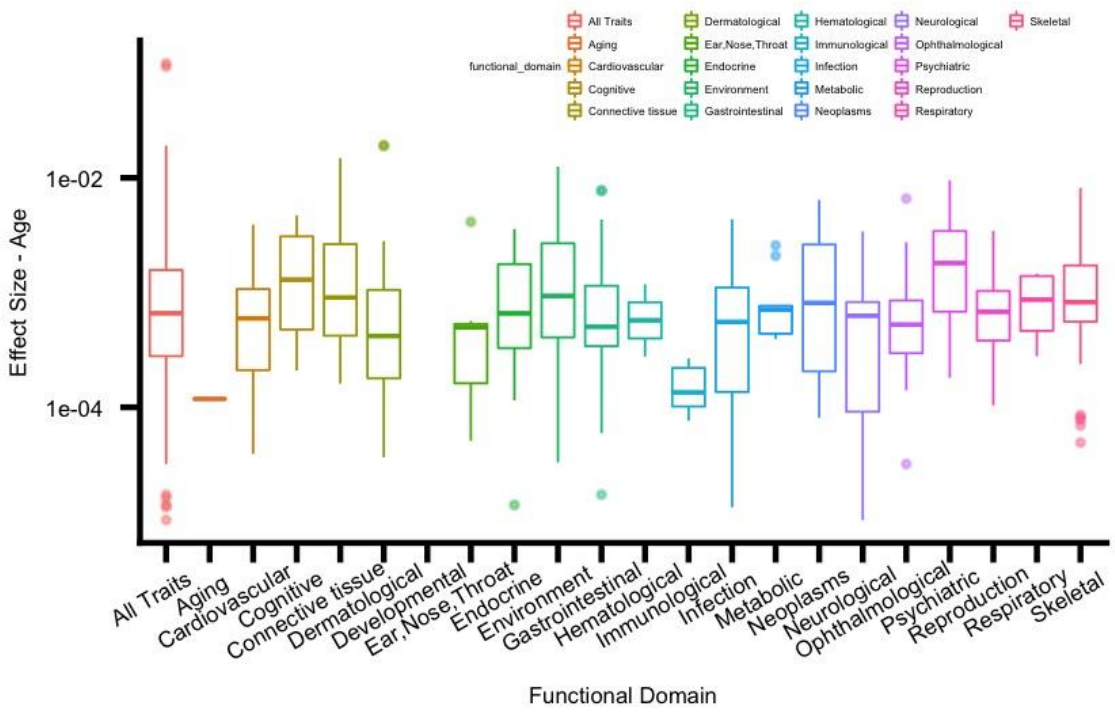
Supplementary Figure 6: Data Processing Flowchart. Flowchart describing steps for the creation of twin and sibling cohort.



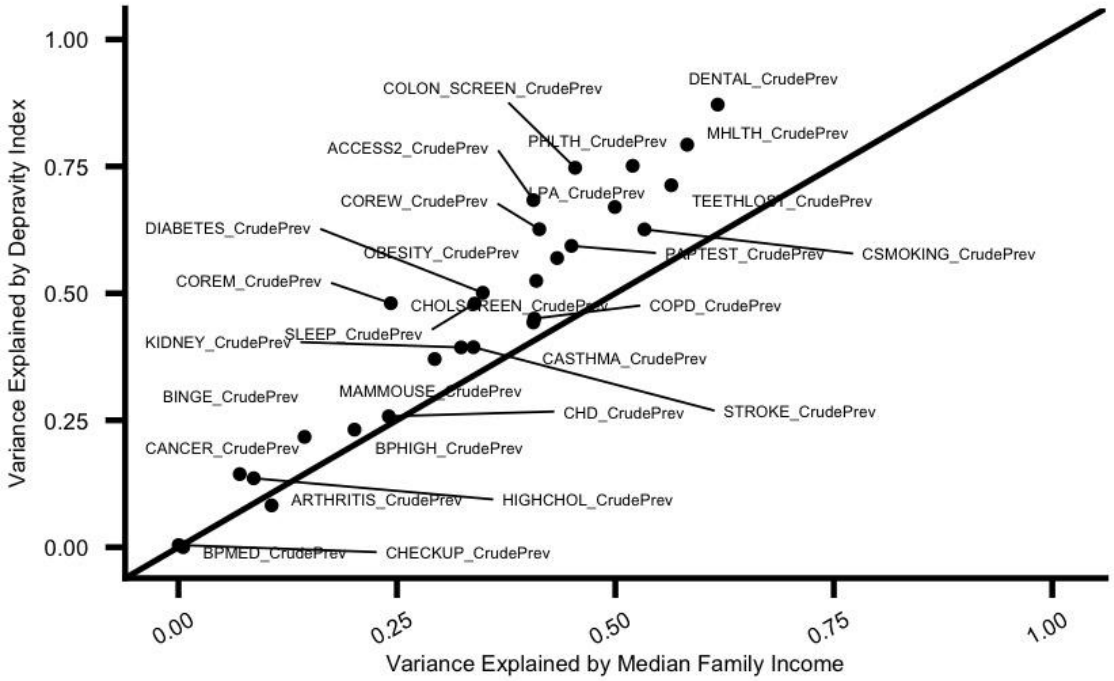
Supplementary Figure 7: Distribution of year of birth. Distribution of year of birth for 56,396 twin pairs in CaTCH (Binary Phenotype) and subsets of twin pairs that have specific lab data in insurance claims dataset.



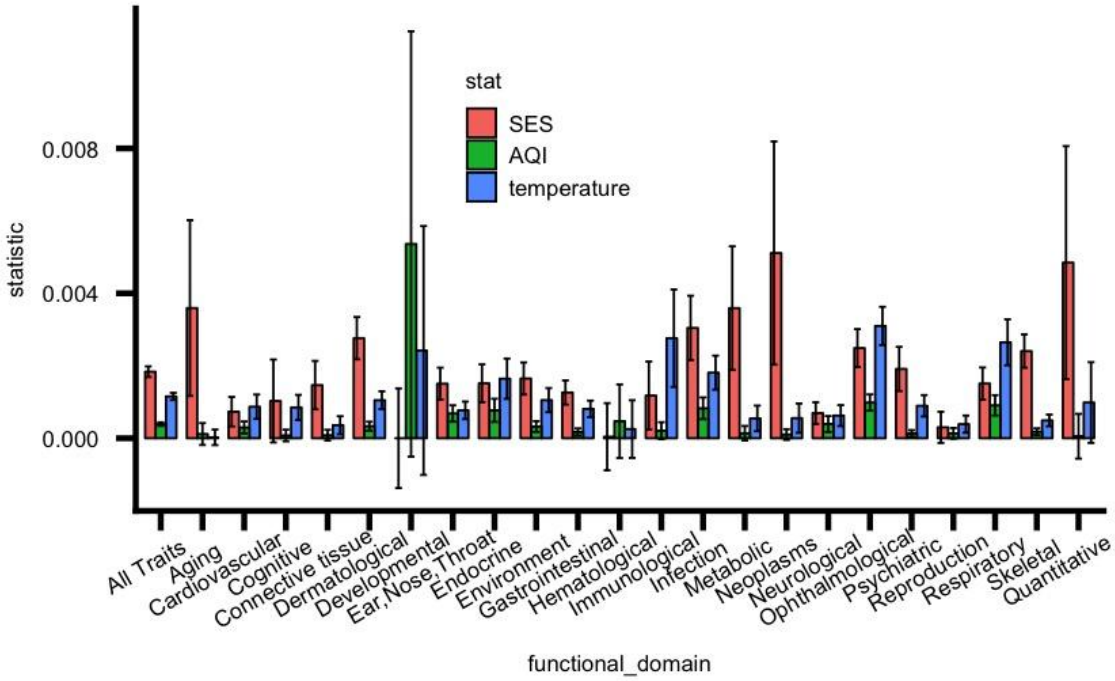
Supplementary Figure 8: Distribution of effect sizes for gender among all functional domain traits for 56,396 twin pairs in CaTCH among 560 phenotypes. Boxplot of effect sizes for gender coefficient (β_{sex}) (female is used as reference variable) from linear mixed model for estimation of h^2 and c^2 **(Online Methods Eq. 3)**. Bottom line, center line, and top line represent 25th percentile, median, and 75th percentile of effect size of β_{sex} for each functional domain, respectively. Bottom and top whiskers represent smallest and largest value of β_{sex} within 1.58 times IQR range for functional domain, respectively. Dots represents phenotypes where β_{sex} is outside 1.58 times IQR range for each functional domain.



Supplementary Figure 9: Distribution of effect sizes for age among all functional domain traits for 56,396 twin pairs in CaTCH among 560 phenotypes. Boxplot of effect sizes for age coefficient (β_{age}) from linear mixed model for estimation of h^2 and c^2 **(Online Methods Eq. 3)**. Bottom line, center line, and top line represent 25th percentile, median, and 75th percentile of effect size of β_{age} for each functional domain, respectively. Bottom and top whiskers represent smallest and largest value of β_{age} within 1.58 times IQR range for functional domain, respectively. Dots represents phenotypes where β_{age} is outside 1.58 times IQR range for each functional domain.



Supplementary Figure 10: Comparison of Variance Explained by Depravity Index versus Median Family Income. Scatterplot of variance explained from model using median family income as covariate versus variance explained from model using depravity index for 21 healthcare measures from CDC 500 cities dataset. Variance explained estimates were based on 30,675 census tracts from 500 Cities CDC project in 2015.



Supplementary Figure 11: Estimates of environmental twin statistics across functional domains for 56,396 twin pairs in CaTCH among all 560 phenotypes. Barplot of meta-analytic estimates of var_{SES}, var_{AQI}, and var_{temp} among all 560 phenotypes and within functional domains, error bars represent 95% CI.

Supplementary Notes

1. Testing the Modeling Assumptions for the Estimation of Heritability and Shared Environmental Variance

Dizygotic Twinning Rate Changes Over Time

We estimated the proportion of monozygotic pairs among SS twins using eq. 29 **(Online Methods)** both using the entire twin cohort and for subset of twin pairs born in either 1985 to 1995, 1996 to 2005, or 2006 to 2015 **(Supplementary Table 2)**. The parameter p showed a clear decrease as the subset of twin pairs became younger with no overlap of their 95% CI. One possible explanation is the increasing availability of IVF treatments. For each lab test the twin pairs measured were only a subset of the full twin cohort, which can have different birthdate distributions compared to the full populations **(Supplementary Figure 7)**. Because p was not constant along different birthdate subsets of our data and each lab test had a different birthdate distribution the p was estimated based on the twin pairs ascertained for that particular lab phenotype. The trend of increasing DZ twinning rate was also seen in other US-based twin registries[1].

Impact of In-Vitro Fertilization on Modeling

Using a cohort of twin pairs born during the period of surveillance (PS twin pairs), we were able to find billing claims codes that denoted procedures for in-vitro fertilization for the mother (IVF, ICD 9 code V26.81, ICD 10 code Z31.83, or CPT code S4015). Overall, we found that for 8.06% of the 5185 PS twin pairs, the mother had an IVF treatment code. We also observed that IVF prevalence increased with year of birth among PS twin pairs **(Supplementary Table 3)**.

For the PS twin pairs, we removed twins with and without IVF information. When IVF twin pairs were included, we found that the proportion of monozygotic twins did not change markedly (p=0.329 when removing IVF twins vs 0.312), resulting in a 5.6% increase of h^2 when these IVF twin pairs were removed. We concluded IVF does influence p and subsequently both h^2 and c^2 estimates, but this was estimated based on the youngest twins in our cohort. Assuming IVF usage trends persist then the inflation of h^2 estimates will be smaller when older twin pairs are included in the analysis.

The assumption of Weinberg's Rule

In our modeling, we are utilizing Weinberg's rule which assumed the sex of the dizygotic twin pairs is equally likely to be opposite sex or same sex. Using United States Center for Disease Control (CDC) natality data from 1995-2016[2] we estimate the male birth ratio is 51.2%, therefore, assuming independent effects of sex for dizygotic pairs, the true proportion of opposite sex (OS) pairs among dizygotic pairs is 49.97%, which is very close to 50%.

In order to test this assumption, we used the East Flanders Prospective Twin Survey[3] the largest, to our knowledge, prospective twin cohort study with zygosity information. This cohort

included twin pairs born between 1969-2011. We estimated the percent of OS twin pairs among dizygotic twin pairs and found this value to also be very close to 50% **(Supplementary Table 4).**

We concluded that Weinberg's rule provided a close estimate to the percent of OS twin pairs among dizygotic pairs. Second, assuming sex of dizygotic pairs are independent effects, we used the male birth rate in the United States from CDC Natality data and found that the proportion of OS DZ pairs was very close to 50% (49.9%), lending support to our assumptions.

2. Distribution of Prevalence among Functional Domains

The functional domains (with at least 5 PheWAS codes) with highest median prevalence were cognitive (median prevalence = 0.029) and environment (median prevalence = 0.020), while the lowest median prevalences occurred for phenotypes in the reproduction (median prevalence = 0.007) and immunological domains (median prevalence = 0.006) **(Supplementary Table 1 and Supplementary Figure 1)**. The PheWAS codes with highest prevalence were acute respiratory infection (prevalence = 0.747), acute pharyngitis (prevalence = 0.532), and ear infection (prevalence = 0.454), whereas PheWAS codes with lowest prevalence were disorders of urethra and urinary tract (prevalence = 0.001), simple goiter (prevalence = 0.002), and chronic thyroiditis (prevalence = 0.002)

3: Distribution of Effect Sizes from Gender and Age Covariates

Median age during surveillance (β_{age}) and gender (β_{gender}) were added as fixed covariates into the variance component model **(Supplementary Figure 8 and Supplementary Figure 9)**. The age covariate is estimated in units of years. For the gender covariate, female was used as the reference variable. Controlling the FDR at 5% there were 488/560 phenotypes for $\beta_{age}$and 284/560 phenotypes for β_{gender} that passed FDR significance. Top traits for gender are symptoms/disorders of urinary system (β_{gender}= -0.082; 95% CI [-0.086, -0.078]), pervasive developmental disorders (β_{gender}= 0.081; 95% CI [0.077, 0.085]), and attention deficit hyperactivity disorder (β_{gender}= 0.063; 95% CI [0.059, 0.066]). Top traits for age are otitis media (β_{age}= -0.038; 95% CI [-0.039, -0.038]), suppurative and otitis media (β_{age}= -0.037; 95% CI [-0.0365, -0.0378]), and fever of unknown origin (β_{age}= -0.0244; 95% CI [-0.0254, -0.0245]).

4. Sibling Cohort

We sought to demonstrate the validity of using twin opposite sex (OS) correlation (r_{twinOS}) as a proxy for dizygotic same sex (SS) correlation ($r_{twinDZSS}$) by estimating the correlation of phenotypes in non-twin siblings, hereafter referred to as "siblings". The sibling cohort selection process is the same as the selection criterion for twins except for a few additional steps:

1. Member must be born before 1985,
2. Member was enrolled for at least 36 months,

3. Sibling pairs are both children of primary subscriber,
4. Each member has at least 1 ICD 9/10 code,
5. Age difference between siblings must be at least 11 months *(**different from twin selection procedure)***
6. Age difference between sibling pairs is at most 36 months *(**different from twin selection procedure**)*
7. After this filter, randomly select one sibling pair from each family *(**different from twin selection procedure**)*

This analysis resulted in a cohort of 724,513 sibling pairs with a mean age difference of 25 months **(Table 1)**. Based on this sibling cohort we estimated the correlation (on the liability threshold scale) between OS and SS sibling pairs for all 551 binary phenotypes.

Estimation of opposite sex and same sex sibling correlation

All sibling estimates relied on the model

$$y = X\beta + u_{pair} + u_{extraSS} + e \quad (1)$$

where var(y) = V_{pair} + $V_{extraSS}$ + V_e. The random effect u_{pair} is common to a pair of both opposite sex (OS) and same sex (SS) sibling pairs, while $u_{extraSS}$ is common to a pair of SS sibling pairs but different for OS sibling pairs, thus the covariance between individuals i and j in a pair is cov(y_i, y_j) = V_{pair} for OS sibling pairs and cov(y_i, y_j) = V_{pair} + $V_{extraSS}$ for SS sibling pairs. SS and OS variance components were estimated as follows:

$$V_{sibSS} = V_{pair} + V_{extraSS} \quad (2)$$

$$V_{sibOS} = V_{pair} \quad (3)$$

$$V_{tot} = V_{pair} + V_{extraSS} + V \quad (4)$$

We used the variance components V_{sibSS} and V_{sibOS} to first estimate correlation the observed scale:

$$r_{sibSS01} = \frac{V_{sibSS}}{V_{tot}} \quad (5)$$

$$r_{sibOS01} = \frac{V_{sibOS}}{V_{tot}} \quad (6)$$

For binary phenotypes we transformed correlation from the observed scale to the liability scale as follows (OS formula is the same as SS formula):

$$T = \Phi^{-1}(1 - K) \quad (7)$$

$$z = \Phi(T) \quad (8)$$

$$i = \frac{z}{K} \quad (9)$$

$$Eb_{sibSS} = K + \frac{V_{sibSS}}{K} \quad (10)$$

$$T_{sibSS} = \Phi^{-1}(1 - Eb_{sibSS}) \quad (11)$$

$$r_{sibSS} = \frac{(T - T_{sibSS})\sqrt{1 - (T^2 - T_{sibSS}^2)(1 - \frac{T}{i})}}{i + T_{sibSS}^2(i - T)} \quad (12)$$

K is the population prevalence for the phenotype (estimated from the filtered population) and Φ was the standard normal distribution. The formulas for r_{sibSS} and r_{sibOS} accounted for the reduction of variance expected from the relatives of proband compared to the general population[4].

Analysis of same sex and opposite sex sibling correlation

Using this procedure, we estimated OS sibling correlation (r_{sibOS}) and SS sibling correlation (r_{sibSS}) for all 551 binary phenotypes **(Supplementary Figure 2).** We see r_{sibOS} and r_{sibSS} were highly correlated (r= 0.978, 95% CI: [0.974,0.981]) **(Supplementary Figure 2)**. Supplementary figure 2 and Supplementary figure 3 show, overall, r_{sibOS} was slightly lower compared to r_{sibSS}. For 95% of traits, r_{sibSS} - r_{sibOS} ranged between -0.012 and 0.051 and r_{sibSS} was, on average, 0.017 higher than r_{sibOS}, but for 23.56% of phenotypes r_{sibSS} - r_{sibOS} followed the null distribution (pi_0 statistic[5]). Despite our phenotypic selection and modeling procedures which helped to minimize any genetic by sex interaction effects, a small bias nevertheless still existed. Figure 12a shows values for r_{sibOS} were generally lower than values of r_{twinOS}. We concluded, for most traits, the lower value of r_{sibOS} compared to r_{twinOS} was due to a lower contribution of shared environment in siblings and inferred the same phenomenon would occur between same sex dizygotic correlation ($r_{twinDZSS}$) to r_{sibSS}.

From the high correlation between r_{sibSS} and r_{sibOS} we concluded the assumption is well-founded, but overall, r_{sibOS} was slightly lower than r_{sibSS} therefore the h^2 and c^2 estimates may have a small bias. We also concluded using r_{sibSS} as a proxy for $r_{twinDZSS}$ would introduce a different bias in the analysis **(Supplementary Figure 4)**, therefore it is best to use r_{twinOS} as a proxy for the unmeasurable $r_{twinDZSS}$.

5. Comparison of Estimates to Published Literature

We systematically compared our estimates of heritability (h^2) to the published literature. This comparison was previously challenging due to difficulty mapping between PheWAS codes and phenotypes reported in the literature. This challenge was overcome in several ways. Specifically, first, using the MaTCH[6] dataset, we estimated h^2 estimates on the ICD 10 subchapter level. For each ICD10 subchapter, we re-estimated h^2 for the twin cohort, where the phenotypes of interest were the presence of any ICD9 or ICD10 belonging to that particular ICD 10 subchapter. In some cases, ICD10 subchapters were a close approximation to a PheWAS code utilized in our analysis. However, in other cases, it is a be a combination of multiple PheWAS codes. From the MaTCH dataset, for each ICD 10 subchapter, we used the meta-analytic (MA) monozygotic and same sex dizygotic twin correlation estimates (r_{twinMZ} and $r_{twinDZSS}$) to estimate h^2 using Falconer's formula

$$h^2=2*(r_{twinMZ} - r_{twinDZSS}).$$

We combined the standard errors for r_{twinMZ} and $r_{twinDZSS}$ to estimate the SE for h^2

$$SE_{h^2} = \sqrt{SE^2_{r_{twinMZ}} + SE^2_{r_{twinDZSS}}}.$$

For ICD 10 subchapters where correlation estimates are not available, we ascertained h^2 estimates based on the ACE model. In the end, we were able to estimate the concordance with 45 MaTCH phenotypes harmonized on the ICD10 subchapter level (43 correlation-based estimates and 2 ACE-based estimates).

We then made an additional 36 comparisons between the published MaTCH h^2 estimates and PheWAS codes/quantitative traits from our twin cohort. Our procedure was as follows (1) We matched 13 PheWAS codes to a list published ACE h^2 estimates generated by Wang et al[7] (Note: their list also contains some MaTCH estimates but we focus on non-MaTCH studies) resulting in 13 matches between our PheWAS codes and published ACE h^2 estimates. (2) We found another 23 phenotypes from our own literature search that matched CaTCH phenotypes and quantitative traits. When both correlation-based and ACE-based estimates of h^2 were provided, we used the correlation-based estimates **(Supplementary Table 5)**. Supplementary Figure 12 shows a scatterplot comparing h^2 estimates from the literature and our twin cohort for all 81 phenotypes.

Using a correlation measure that accounts for standard error[8], we found the correlation between published estimates and the twin cohort (CaTCH) is high, r= .817 (95% CI: [0.493, 1.14]). We found 67/81 (82.7%) of phenotypes have overlapping 95% confidence intervals **(Figure 3a)**.

6. Deprivation Index as a measure of SES status

In this analysis, we utilize an area-based indicator of socioeconomic status called the deprivation index[9], which is also used by US Department of Health and Human Services for analysis of their Medicare population[10]. This score uses the following variables (found in the American Community Survey)

1. Unemployment - Percentage of persons aged 16 years or older in the labor force who are unemployed (and actively seeking work)
2. Below US poverty line - Percentage of persons below the federally defined poverty line
3. Median income - Median household income
4. Property values - Median value of owner-occupied homes
5. Low education - Percentage of persons aged > 25 years with less than a 12th-grade education
6. High education - Percentage of persons aged > 25 years with at least 4 years of college
7. Crowded households - Percentage of households containing one or more person per room

For any geographic area type (in our case zipcode) we ascertained the values of these variables for all of these areas in the US, performed PCA, and extracted the 1st PC score for each geographic area. This PC score the depravity index score for that area. High depravity index corresponds to high SES status.

We then tested whether depravity index is better correlated in prevalence of health indicators versus just median family income. Using the above procedure, we created a depravity index on the American Community Survey Census tract level. Next, we correlated the depravity index with prevalence rates of disease and health outcomes data from the US Center for Disease Control (CDC) 500 cities project (https://www.cdc.gov/500cities/). The measures are listed below:

1. ARTHRITIS_CrudePrev - Arthritis prevalence among adults aged≥18 years
2. BINGE_CrudePrev - Binge drinking among adults aged ≥ 18 years
3. BPHIGH_CrudePrev - High blood pressure among adults aged≥18 years
4. BPMED_CrudePrev - Taking medicine for high blood pressure control among adults aged≥18 years with high blood pressure
5. CANCER_CrudePrev - Cancer (excluding skin cancer) among adults aged≥18 years
6. CASTHMA_CrudePrev - Current asthma among adults aged≥18 years
7. CHD_CrudePrev - Coronary heart disease among adults aged≥18 years
8. CHECKUP_CrudePrev - Visits to doctor for routine checkup within the past year among adults aged≥18 years
9. CHOLSCREEN_CrudePrev - Cholesterol screening among adults aged≥18 years
10. COPD_CrudePrev - Chronic obstructive pulmonary disease among adults aged≥18 years
11. CSMOKING_CrudePrev - Current smoking among adults aged≥18 years
12. DENTAL_CrudePrev - Visit to dentist or dental clinic among adults aged ≥18 years
13. DIABETES_CrudePrev - Diagnosed diabetes among adults aged≥18 years
14. HIGHCHOL_CrudePrev - High cholesterol among adults aged≥18 years who have been screened in the past 5 years
15. KIDNEY_CrudePrev - Chronic kidney disease among adults aged≥18 years
16. LPA_CrudePrev - No leisure-time physical activity among adults aged≥18 years
17. MHLTH_CrudePrev - Mental health not good for≥14 days among adults aged≥18 years
18. OBESITY_CrudePrev - Obesity among adults aged≥18 years
19. PHLTH_CrudePrev - Physical health not good for≥14 days among adults aged≥18 years
20. SLEEP_CrudePrev - Sleeping less than 7 hours among adults aged ≥18 years

21. STROKE_CrudePrev - Stroke among adults aged≥18 years

For each health measure, we built two linear regressions predicting the incidence as a function of median family income and depravity index respectively. Below we show a scatterplot comparing the variance explained by each model for each health measure.

On average, the variance explained by depravity index is 0.119 higher than median family income **(Supplementary Figure 10)**. We found DENTAL_CrudePrev (r=0.948) and CHOLSCREEN_CrudePrev (r=0.758) are positively correlated with the depravity index, while MHLTH_CrudePrev (r=-0.911) and PHLTH_CrudePrev (r=-0.894) are negatively correlated **(Supplementary Figure 10)**. These correlations are consistent with prior expectations of the relationship between SES status and health measures. We concluded that depravity index is a better measure of SES status compared to median family income and was able to validate it matches prior expectations between SES status and health using an external dataset, the 500 cities.

References

1. Rhea, S. A. *et al.* Higher Rates of DZ Twinning in a Twenty-First Century Birth Cohort. *Behav. Genet.* **47,** 581–584 (2017).
2. United States Department of Health and Human Services. National Center for Health Statistics. Natality Detail File, 2002 [United States]. *ICPSR Data Holdings* (2007). doi:10.3886/icpsr04705
3. Derom, C. *et al.* The East Flanders Prospective Twin Survey (EFPTS): an actual perception. *Twin Res. Hum. Genet.* **16,** 58–63 (2013).
4. Reich, T., James, J. W. & Morris, C. A. The use of multiple thresholds in determining the mode of transmission of semi-continuous traits. *Ann. Hum. Genet.* **36,** 163–184 (1972).
5. Storey, J. D. A direct approach to false discovery rates. *J. R. Stat. Soc. Series B Stat. Methodol.* **64,** 479–498 (2002).
6. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47,** 702–709 (2015).
7. Wang, K., Gaitsch, H., Poon, H., Cox, N. J. & Rzhetsky, A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat. Genet.*

49, 1319–1325 (2017).

8. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9,** 2282 (2018).
9. Krieger, N. *et al.* Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *Journal of Epidemiology & Community Health* **57,** 186–199 (2003).
10. Bonito, A. J., Bann, C., Eicheldinger, C. & Carpenter, L. Creation of new race-ethnicity codes and socioeconomic status (SES) indicators for Medicare beneficiaries. *RTI International, Agency for Healthcare Research and Quality* (2008).

Supplementary Tables

Functional category	*n* phenotypes	% phenotypes	Median all prevalence (IQR)	Median female prevalence (IQR)	Median male prevalence (IQR)	*n* MaTCH phenotypes	*% MaTCH phenotypes*
All	560	100	0.010 (0.0054-0.027)	0.0099 (0.0055-0.026)	0.010 (0.0052-0.027)	9568	100%
Aging	4	0.72%	0.024 (0.020-0.032)	0.020 (0.016-0.028)	0.028 (0.021-0.038)	37	0.21%
Cardiovascular	35	6.34%	0.010 (0.0054-0.017)	0.0096 (0.0054-0.016)	0.010 (0.0049 - 0.016)	754	4.24%
Cognitive	5	0.91%	0.029 (0.024-0.095)	0.024 (0.020-0.065)	0.038 (0.025-0.13)	1507	8.46%
Connective tissue	11	1.99%	0.0089 (0.0046-0.029)	0.011 (0.0050-0.029)	0.0072 (0.0044-0.028)	8	0.04%
Dermatological	61	11.05%	0.016 (0.0063-0.029)	0.017 (0.0065-0.034)	0.015 (0.0056-0.025)	181	1.02%
Developmental	1	0.18%	0.0056* (0.0056-0.0056)	0.0055 *(0.0055-0.0055)	0.0058 *(0.0058-0.0058)	12	0.07%
Ear, Nose, Throat	39	7.07%	0.017 (0.0094-0.062)	0.017 (0.0088-0.053)	0.019 (0.010-0.072)	371	2.08%
Endocrine	32	5.80%	0.0070 (0.0044-0.017)	0.0072 (0.0048-0.020)	0.0063 (0.0039-0.014)	399	2.24%
Environment	45	8.15%	0.020 (0.010-0.047)	0.017 (0.0088-0.046)	0.024 (0.012-0.051)	487	2.74%
Gastrointestinal	59	10.69%	0.0092 (0.0050-0.016)	0.0090 (0.0051-0.017)	0.0085 (0.0052-0.015)	144	0.81%
Hematological	3	0.54%	0.0042 (0.0042-0.0076)	0.0048 (0.0045-0.0080)	0.0040 (0.0038-0.0075)	110	0.62%
Immunological	15	2.72%	0.0060 (0.0041-0.012)	0.0061 (0.0041-0.012)	0.0059 (0.0040-0.010)	280	2.05%
Infection	38	6.88%	0.017 (0.0066-0.052)	0.020 (0.0067-0.049)	0.013 (0.0067-0.056)	8	0.04%
Metabolic	13	2.36%	0.0073 (0.0040-0.034)	0.0070 (0.0042-0.037)	0.0067 (0.0034-0.030)	1750	9.83%
Neoplasms	13	2.36%	0.0093 (0.0053-0.016)	0.012 (0.0045-0.017)	0.0077 (0.0055-0.015)	54	0.30%
Neurological	36	6.52%	0.0079 (0.0057-0.023)	0.0085 (0.0053-0.020)	0.0085 (0.0053-0.023)	3371	18.93%
Ophthalmological	42	7.61%	0.012 (0.0071-0.056)	0.013 (0.0071-0.058)	0.012 (0.0069-0.054)	305	1.71%
Psychiatric	37	6.70%	0.013 (0.0071-0.048)	0.012 (0.0073-0.042)	0.014 (0.0065-0.041)	5178	29.08%
Reproduction	18	3.26%	0.0070 (0.0050-0.014)	0.0078 (0.0052-0.012)	0.0062 (0.0041-0.014)	283	1.59%
Respiratory	48	8.70%	0.019 (0.0067-0.088)	0.018 (0.0062-0.082)	0.020 (0.0072-0.092)	252	1.42%
Skeletal	64	11.59%	0.010 (0.0052-0.021)	0.011 (0.0052-0.019)	0.011 (0.0052-0.024)	895	5.03%

Supplementary Table 1: Prevalence Among Phenotypic Categories and Comparisons to MaTCH

Year of Birth	Number of SS Pairs	Number of OS Pairs	Total Pairs	p	p [95% CI]
All	35754	20642	56396	0.423	[0.413,0.433]
1985 -1995	8686	3941	12627	0.546	[0.529,0.563]
1996 - 2005	16729	9803	26532	0.414	[0.399,0.429]
2006 - 2015	10339	6898	17237	0.333	[0.312, 0.353]

Supplementary Table 2: Estimate of the proportion of monozygotic given same sex (p) among twins based on year of birth and corresponding 95% confidence intervals.

Year of Birth	OS Pairs	SS Pairs	Total Pairs	Pairs with IVF Code	Percentage with IVF code
2009	553	768	1321	88	6.66%
2010	529	781	1310	100	7.63%
2011	471	733	1204	86	7.14%
2012	444	641	1085	117	10.78%
2013	123	142	265	27	10.19%

Supplementary Table 3: Proportion of twin pairs born due to IVF treatment for years when information was available.

MZ Pairs	DZ SS Pairs	DZ OS Pairs	Percent of OS Twin Pairs among Dizygotic Pairs
2734	2838	2805	50.03%

Supplementary Table 4: Number of twin pairs by zygosity status in East Flanders Prospective Twin survey and estimate of proportion of OS twin pairs among dizygotic twin pairs.

Year	Percentage of Cost Attributed to Costliest 5% of Twins
2008	73.3%

2009	83.8%
2010	52.3%
2011	53.6%
2012	54.0%
2013	56.0%
2014	56.3%
2015	55.9%

Supplementary Table 6: Percentage of Total Yearly Cost Attributed to Costliest 5% of Twins