LEVERAGING BIOBANKS TO CHARACTERIZE THE GENETIC ARCHITECTURE OF SUICIDE ATTEMPT, SUICIDAL IDEATION, AND TREATMENT-RESISTANT DEPRESSION

By

JooEun Kang

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in

Human Genetics
May 13th, 2022
Nashville, Tennessee

Approved:
Douglas Ruderfer, PhD, Advisor

Lea K Davis, PhD, Committee Chair

Colin Walsh, MD

Jonathan Mosley, MD, PhD

Dan Roden, MD

To those who seek meaning, worth, and hope in life.

Acknowledgements

This work would not have been possible without the financial support of the Vanderbilt Medical Scientist Training Program, the Vanderbilt Dean's Award, or the Vanderbilt Genetics Institute. I am particularly grateful to Dr. Douglas Ruderfer, who took a chance with me as his very first graduate student although I had zero experience with computational research. I am very blessed to have a mentor whom I felt I could share my stupidest, brightest, or darkest moments with. His patience and good humor have carried me through this whirlwind that is graduate school which seem both infinitely long and short. He has been a phenomenal role model for what a good scientist, and person, should be, and I wouldn't be where I am without him.

I am also grateful to all my collaborators with whom I have had the pleasure of working on my thesis project. Although the pandemic has limited all our interactions online, it has been an honor and a delight to meet weekly to share ideas and receive feedback with incredibly smart and passionate individuals. I will miss starting every Zoom meeting with anything but science and ending every meeting with a dash (or a dollop) of healthy skepticism of the field. Thank you to all my lab members for their constant help, support, and friendship. Thank you also to my Dissertation Committee, who have always provided me timely and astute personal and professional guidance throughout my PhD.

Thank you to my family, loved ones, and friends, who have been my most ardent cheerleaders throughout the pursuit of my Ph.D. I am particularly indebted to my quarantine kitty Minerva and my husband Ben who have offered much love and support throughout this whole journey.

Table of Contents

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Major Depressive Disorder (MDD)

Depression is common and heterogeneous

Depression is a heterogenous term that could designate 1) a mood state indicated by having feelings of sadness, despair, anxiety, emptiness, discouragement, or hopelessness (also summarized as dysphoria); or having no feelings, 2) a syndrome of symptoms that may include depressed mood, or 3) a distinct clinical condition of major depression that is featured in several mental disorders including major depressive disorder, bipolar disorder, and schizophrenia. For the rest of this thesis, the definition of depression as a clinical condition will be used unless specified. Depression is one of the most common mental illnesses in the world, with a lifetime prevalence of major depressive disorder (MDD) of 21% among adults in the United States[1]. There is a significant difference in prevalence between age groups, with the prevalence of individuals aged 18-29 being threefold that of individuals of age 60 and above. There are gender differences in prevalence as well, with female prevalence being up to threefold higher than that of males after early adolescence.

Depression is associated with many psychiatric and non-psychiatric comorbidities. In a national study of German insurance claims data, depression cases were found to have twice as many psychiatric comorbidities than age and sex-matched controls[2]. The most prevalent psychiatric comorbidities found in that study included neurotic, stress-related and somatoform disorders, substance use disorders, and personality disorders. In a US study of Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-V) defined MDD, generalized anxiety disorder and borderline personality disorders were the most strongly associated among substance use, anxiety, and personality disorder comorbidities[1]. In a Scottish study of more than 140,000 individuals in primary care with depression, significant non-psychiatric comorbidities of depression included pain, constipation, multiple sclerosis, viral hepatitis, Parkinson's

disease, and migraine[3]. The study also found that medical comorbidities of depression are influenced by socioeconomic factors, as chronic conditions such as pain, dyspepsia, asthma, coronary heart disease, diabetes, and chronic obstructive pulmonary disease were significantly more prevalent among the highest quintile of deprivation compared to the lowest quintile of deprivation. The relationship between depression and its comorbidities is complex. Prognosis of depression is worsened by the presence of comorbidities[4], and conversely depression worsens the prognosis of many psychiatric and non-psychiatric illnesses, including ischemic stroke[5,6], diabetes[7], and cancer[8,9].

Table 1. Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-V) definition of major depressive episode

DSM-V criteria for major depressive episode
5 or more of 9 symptoms (including at least 1 of depressed mood and loss of interest or pleasure) present nearly every day in the same 2-week period; each of these symptoms represents a change from previous functioning: 1. Depressed mood (subjective or observed) 2. Loss of interest or pleasure 3. Significant change in weight or appetite 4. Insomnia or hypersomnia 5. Psychomotor retardation or agitation (observed) 6. Loss of energy or fatigue 7. Worthlessness or guilt 8. Impaired concentration or indecisiveness 9. Thoughts of death or suicidal ideation or suicide attempt *** Symptoms cause clinically significant distress or impairment in function** *** Episode is not attributable to the physiological effect of substance or another medical condition**

Depression has heterogeneous clinical presentation with subtypes defined based on symptoms, etiology, onset time, course, and duration. In the clinical criteria for major depressive disorder in the DSM-V, both increases and decreases in weight, sleep, and psychomotor function are all symptoms of depression (**Table 1**). As such, efforts to quantify depression symptoms have had limited overlap. Fried et al[10] assessed the degree of overlap between seven common depression scales, with only 6 symptoms among 52 depression symptoms across seven depression scales being featured across all instruments: sad mood, appetite decrease, fatigue, and 3 insomnia items. Rather than quantifying major depressive disorder, the DSM-V included specifiers of notable symptoms to define different subtypes in depression, such as symptoms of anxious distress, psychotic features, and weight gain and hypersomnia defining atypical

depression (**Table 2**). These subtypes are meant to provide greater diagnostic specificity and are not mutually exclusive. Many studies have examined the clinical utility of defining subtypes as predictors of antidepressant outcome[11–13] but found no robust evidence for it.

Table 2. Subtypes of major depressive episodes (MDEs)

Subtype	Features	Notes
Anxious distress	Tension, restlessness, rumination, panic attacks	40-50% of MDEs qualify as anxious depression[11]
Atypical	Reactive to pleasurable stimuli, increase in appetite/weight gain, hypersomnia, leaden feeling in limbs	15-50% MDEs
Catatonic	Prominent psychomotor disturbances	
Melancholic	Affect unresponsive to improved circumstances, anhedonia, insomnia, loss of appetite, neurocognitive impairment	15-30% of MDEs
Mixed features	Manic/hypomanic symptoms (elevated mood, grandiosity, talkative, flight of ideas, decreased need for sleep)	Higher comorbidity with panic disorder and substance use disorders[14]
Peripartum	Begins during pregnancy or within four weeks of childbirth	
Psychotic	Delusions and hallucinations	
Seasonal	Recurrent mood episodes that begin during a particular season and remit during another season	

Depression is heritable and variable genetic architecture is observed within subtypes

Depression is a heritable trait, with twin heritability estimates of around 37%[15] and SNP heritability estimate of around 8.7% (SE 0.004) or 8.9% (SE 0.003) on the liability scale[16,17] assuming lifetime risk of 0.15 or 0.30, respectively. Depression is also heterogeneous, and several studies have examined the impact of variable depression phenotyping on the resulting genetic architecture. The following three studies demonstrate that there are differences in heritability as well as genetic characteristics of subtypes within the clinical diagnosis of MDD. These subtypes require different amounts of symptomatic descriptors to depression, such as length and recurrence, or subtypes as defined by the DSM-V based on the presence of distinct symptoms.

Cai et al.[18] devised five definitions of MDD of varying strictness in UK Biobank, ranging from the minimal definition of self-report of seeking help for depression, to the strictest definition requiring DSM-V symptoms of MDD. They observed significant differences in age (strict MDD population are younger

than minimal MDD), experience of traumatic life events and recent stress (higher in strict MDD population). Those differences translated to differences in SNP-heritability estimates, where heritability of minimal phenotyping MDD was lower (14% on liability scale, SE=0.8%) than strict MDD (26%, SE=2.2%). There were strong genetic correlations between the definitions indicating shared genetics among the MDD phenotypes. Variable phenotyping did not affect genetic correlations with other diseases, but associated SNPs were more pleiotropic for minimal MDD.

Jermy et al.[19] aimed to find which aspects of MDD led to the differences between minimal and strict phenotyping. They looked at five components of MDD in addition to the cardinal symptoms of anhedonia or depressed mood, 1) presence of five or more symptoms, 2) episode duration, 3) functional impairment, 4) persistence of symptoms during episodes (symptoms present nearly every day), 5) and recurrence (two or more depressive episodes in lifetime). The authors then generated thirty-two phenotypes of varying combination of the five phenotypic components, ranging from the minimal phenotype requiring just cardinal symptoms, to the more complex phenotype requiring all five phenotypic components in addition to cardinal symptoms. Among the five phenotypic components, the authors observed a significant increase in SNP-heritability estimate only with the additional presence of five or more symptoms in addition to cardinal symptoms (increase of 2.7%, SE=0.008). Overall, SNP-heritability estimates decreased with the addition of symptom components to cardinal symptoms. There were no differences in genetic correlations with existing genome-wide association studies (GWAS) of MDD with additional symptom components, and overall, this suggests that a large portion of the heritable aspect of MDD is captured by the cardinal symptoms of anhedonia and depressed mood.

Nguyen[20] et al. specifically looked at the genetics of sixteen different depression subtypes across eight different domains: vegetative symptoms, symptom severity, comorbid anxiety disorder, age of onset, recurrence, suicidality, functional impairment, and postpartum depression. Each GWAS was conducted against controls with no history of lifetime major depression. In a comparison of heritability estimates of subtypes within each domain (e.g., mild/moderate vs severe depression), the subtype with the more severe clinical manifestation had higher heritability estimates. In a comparison of heritability and genetic

correlation between all sixteen subtypes, the atypical depression subtype defined by hypersomnia and weight gain showed the biggest difference in heritability estimates and genetic correlations. The atypical subtype had the lowest significant genetic correlation with other subtypes (rg=0.55), and atypical depression had the highest heritability estimate at ~ 19% on the observed scale, which was double the estimate for non-atypical depression that do not report both hypersomnia and weight gain. When comparing genetic correlation differences in depression subtypes within each domain, most significant differences were observed between the atypical and non-atypical depression subtypes. Atypical depression showed a stronger positive genetic correlation with BMI and ADHD, while non-atypical depression showed a stronger positive genetic correlation with anorexia nervosa and cognitive traits of intelligence and years of schooling. The study also confirmed general pattern of genetic overlap with psychiatric diseases across different depression subtypes. The genetic correlations with schizophrenia and bipolar disorders were common across depression subtypes, but higher in the clinically challenging depression subtypes (early-onset, recurrent, suicidal, severe functional impairment).

In summary, these three studies suggest that phenotypic heterogeneity in depression presents as genotypic heterogeneity in the following ways: 1) strict phenotyping results in higher heritability estimates and identification of genetic associations that are more specific to depression, 2) cardinal symptoms of anhedonia and depressed mood is able to capture a large portion of the genetics of depression, and 3) there are definite differences in heritability estimates and genetic correlation among different depression subtypes, in particular the atypical depression subtype, but all subtypes share a common genetic architecture and genetic overlap with psychiatric diseases. It is worth noting that all three studies have been conducted in UK Biobank, which is unique in its variety in data types and availability that enables identifying various depression subtypes. However, the reliance on survey data of UK Biobank also signifies that such subtype phenotyping is not replicable in other biobanks where available structured data is limited to medications and diagnostic codes. It is for this reason that some studies identify patients with more severe disease burden of depression using treatment response rather than subjective symptoms.

Treatment-resistant Depression (TRD)

Depression is often chronic

In addition to its high prevalence, major depressive disorder poses a high mental and financial burden to patients because of its chronicity. About 27% of MDD patients develop a chronic depressive disorder with an illness duration of at least two years[21], and about half of patients with chronic depressive disorders do not recover despite treatment[22]. Older age of onset, number of depressive episodes, comorbid psychiatric disorders, and family history of psychopathology have been identified as risk factors for recurrence[23]. One possible reason for recurrence is the lack of response to treatment. Major depressive disorder is managed with psychotherapy (e.g. cognitive behavioral therapy), pharmacotherapy (i.e. selective serotonin reuptake inhibitor), or a combination of the two, and multiple randomized controlled trials suggest the benefit of combined therapy compared to pharmacotherapy alone[24]. While mild/moderate MDD can be managed with psychotherapy alone, pharmacotherapy is preferable for most MDD cases, especially for severe MDD[25]. Selective serotonin reuptake inhibitors (SSRIs) such as escitalopram and sertraline are often the first-line pharmacotherapies for patients and are the most widely prescribed antidepressants overall due to their efficacy and tolerability in randomized trials[26]. A trial length of antidepressant can range from 6-12 weeks, but for patients who do not experience symptom improvement after 4-6 weeks, it is recommended to add on (augmentation) or switch to a second antidepressant[27]. When patients do not experience relief with pharmacotherapy, this could be due to pharmacokinetics and pharmacodynamics, where the drug fails to reach a potent level in the brain due to metabolism or fails to achieve the desired effect in the brain due to differences in receptors in the brain. A patient may stop pharmacotherapy before symptom relief due to undesirable side effects such as weight gain, decreased libido and sexual function, gastrointestinal discomfort, and insomnia. Recurrent depression despite treatment also represents a subset of MDD individuals with severe symptoms, potentially due to high genetic risk of MDD, and for this reason the field has focused on studying patients with treatment-resistant depression.

Definition of treatment resistant depression

Lack of response to treatment measured by number and duration of unique antidepressant trials is one approach to stratify individuals with worse outcomes. Treatment-resistant depression (TRD) is defined by absence of remission following at least two adequate antidepressant treatment trials, although there is no consensus in the field on the measures of remission, length of adequate treatment trial duration, and adequate dose[28]. TRD is estimated to impact at least a third of all individuals with major depressive disorder[29,30] while accounting for nearly half of incremental health costs associated with MDD[30]. TRD patients are also at a higher risk of various mental health outcomes, including suicide, with 30% of TRD patients attempting suicide at least once in their lifetime[31], 15 times the lifetime rate of the general population (~2%)[32].

ECT candidacy and efficacy

Several treatment options are indicated for TRD patients such as repetitive transcranial magnetic stimulation (rTMS) and ketamine, however the gold standard intervention for TRD has remained electroconvulsive therapy (ECT) for decades[33]. ECT is a procedure where seizure activity is electrically induced and is typically reserved for individuals with TRD or refractory bipolar disorder who need a rapid reversal of severe symptoms such as suicidality[34,35]. ECT is the most effective treatment of severe TRD[34] with treatment response of 50-75%[35], which is much higher than the 10-40% remission rates for pharmacotherapy and psychotherapy[29]. ECT efficacy has been associated with increased depression severity, along with older age, and presence of psychotic features[36]. Therefore, it is generally agreed that ECT among individuals with depression is indicative of TRD. However, it is not yet known whether individuals with TRD having received ECT define a generalizable subtype of TRD more broadly. Despite new treatment strategies being available for patients with severe MDD such as ketamine, recent clinical trials confirm the superior remission rates of ECT especially among older patients (63% remission compared to 46% remission for ketamine, n=~90 each group, p=0.026)[37], and as long as patients can tolerate side effects of ECT which include headache, muscle pain, and amnesia, ECT will remain an important and effective treatment option for TRD patients. However, some limitations in using ECT as an ascertainment

method for TRD include the exclusion of patients who may not elect ECT because of personal preference, patients with medical conditions such as extreme obesity that preclude them from the anesthesia requirement of ECT, or socioeconomic factors that influence access to medical centers that provide ECT and availability of a caregiver to accompany patients after anesthesia.

Genetic studies of treatment resistant depression

TRD is a heritable trait with heritability estimates up to 17%[38] when compared to controls and ~8% when compared to non-TRD MDD[39,40]. Three genome-wide significant SNPs for TRD have been identified but none have replicated (**Table 4**). This is in part due to the various methods of ascertainment that has been used in TRD studies, including TRD definitions based on antidepressant prescriptions[41,42], self-reported antidepressant efficacy and side effects[38,39], and ECT treatment[43], and remission of depressive symptoms[44]. Another limitation in studying the genetics of TRD has been a paucity of adequately powered cohorts. In the past three years, there has been substantial improvements in power using large biobanks such as 23andMe and UK Biobank. A GWAS comparing ECT recipients to non-MDD controls in Sweden[45] found higher heritability estimates of ECT in the context of MDD (liability scale SNP h2=31%, SE=0.06. prevalence = 0.01) compared to PGC-MDD (6-8%) (**Table 3**). Patients with ECT-defined TRD had higher PRS of MDD, bipolar disorder, and cognitive traits (educational attainment and IQ), compared to MDD patients with moderate symptoms. In TRD studies where TRD cases ascertained using medication data against non-TRD MDD controls, there were significant heritability estimates around 7-8%[39,40] (**Table 3**). Medication-defined TRD patients showed positive associations with ADHD PRS and negative association with intelligence. These studies suggest that there is a genetic architecture of TRD even when comparing to non-TRD MDD controls, with some differences in genetic overlap with psychiatric and cognitive traits but none that are sufficiently powered or replicated.

Table 3. Summary of SNP-heritability estimates of treatment-resistant depression when compared to MDD controls or non-MDD controls.

Heritability	TRD definition	non-TRD definition	Sample size	Case source	Source
7.8% (SE 4%) p=0.03	self reported antidepressant efficacy and side effects	subjects reported fair or great efficacy to at least one antidepressant and never reported little or no efficacy to any antidepressant	5714 TRD 31,068 non-TRD MDD	23andMe	Li et al. 2020 (PMID 33106475)
7.7% (SE 2.7%) p=0.002	at least 2 switches between antidepressant drug, each prescribed for at least 6 weeks	at least 2 diagnostic code for unipolar disorder. Exclude bipolar disorder, pyschotic disorder, or SUD	2165 TRD 14,207 non-TRD MDD	UK Biobank	Fabbri et al. 2021 (PMID 33753889)
4.2% (SE 8%) p=8.8E-8	ECT for an major depressive episode in the context of MDD	Exclude individuals with self-reported lifetime hx of MDD, BIP, SCZ, or schizoaffective disorder	1796 TRD 3290 healthy control	Predictors for ECT study Swedish National Quality Register for ECT	Clements et al. 2020 (PMID 33483693)

Table 4: Summary of genome-wide significant SNPs identified with TRD

Trait	GWS snp	CHR	A1/A2	OR	P	Source	N
Non-TRD vs TRD	rs150245813	10	T/G	0.8	8.07E-09	Li et al. 2020	29488
SNRI Responder vs Non-Responder	rs4955665	3	G/A	1.25	1.62E-09	Li et al. 2020	8119
ECT for depression, bipolar disorder, and schizoaffective disorder	rs114583506	6 (MHC)	G/T	0.6	3.60E-08	Clements et al. 2020	6015

Suicide

Epidemiology of suicidal thoughts and behaviors[*]

Suicidal thoughts and behaviors (STBs) co-occur with multiple psychiatric and medical conditions and cross many diagnostic boundaries. Currently, STBs are not defined as a discrete psychiatric disorder, and diagnostic criteria only exist within the context of major depressive and borderline personality disorders[46]. Despite the importance of decreasing STBs in improving public health, the multifaceted nature of STBs has led to a historical heterogeneity of both suicide terminology and suicide risk measures, making it difficult to compare findings across epidemiological studies[47]. This relative lack of distinction across outcomes has a complicated interpretation of results. There are three primary phenotypes comprising STBs – suicidal ideation (SI), suicide attempt (SA), and suicide. Consistent with accepted terminologies, SI is defined as thoughts about ending one's own life, SA is defined as self-injurious, non-fatal behavior with the intent to die, and suicide is defined as a fatal behavior with intent to die[48].

Rates of SI and SA are higher than that of suicide, in part because the stigma associated with dying by suicide resulting in misclassification of deaths[49] and negatively influencing reporting rates, and the ethical and legal complexities of obtaining post-mortem suicide data[50]. The World Health Organization (WHO) estimates that ~800,000 individuals take their own life each year, and for every suicide death there are approximately 20 individuals with SAs, and many more with SIs[51]. Broadly, suicide rates have been decreasing in many countries, with a notable increase in the United States[52]. However, among people who attempt suicide, only 10–15% eventually go on to die from suicide, with 1.6% of suicides occurring within 1 year and 3.9% within 5 years of an attempt[53,54]. In other words, SA and SI can lead to an eventual suicide death, but for the vast majority of people, it does not.

[*] Adapted from DiBlasi E, Kang J, and Docherty AR, Psychological Medicine, 2021

Theory of suicide and transition from suicidal ideation to suicide attempt

French sociologist Émile Durkheim is often credited as the first to create a systematic framework in which to study suicide that shaped American suicide research of the twentieth century. His Sociological Theory of suicide[55] derived four types of suicide from the intersection of two major axes: social integration and regulation. Social integration denotes the sense of belonging and inclusion from social ties and social regulation denotes the regulation and guidance that also come from social ties. The egoistic type of suicide is defined by the lack of social integration leading to isolation and lack of a sense of belonging. The altruistic type of suicide is defined by excess of social integration where individuals value the needs of the group over their own need to survive. The anomic type of suicide is defined by the lack of social regulation where society fails to provide a moral framework for individuals. The fatalistic type of suicide is defined by extreme social regulation, leading to a desire to escape oppressive and controlling environments.

Other theories on the motivation of suicide include the Escape Theory of Suicide[56] by social psychologist Roy Baumeister which suggests that the primary motivation of suicide is escape from painful self-awareness. This theory has been particularly influential in explaining adult male suicides. Clinical psychologist Edwin Shneidman[57] pointed to psychache as the central factor of suicide, where psychache is the psychological pain from four causes: thwarted love, acceptance or belonging; excessive hopelessness; damaged self-image causing feeling of shame, defeat, and humiliation; and damaged relationships causing feelings of grief. Shneidman presented a cubic model of suicide where suicide risk is a combination of press (stressors), pain (psychache) and perturbation (restlessness and inclination to act). This reframed suicide from a willingness to die to a means to end psychological pain, where individuals have different thresholds for enduring this pain and attempt suicide when the threshold is reached as the most drastic measure to reduce psychache.

To examine how the rates of suicidal ideation are many folds higher than suicide attempt, psychologists David Klonsky and Alexis May introduced the ideation-to-action framework[58] to differentiate risk factors between suicidal ideation and suicide attempt, and understand the transition from ideation to attempt. Three theories of suicide fit to this ideation-to-action framework and have been summarized in

Table 5: the Interpersonal Psychological Theory of Suicide[59] by Thomas Joiner, the Integrated Motivational Volitional Theory[60] by Rory O'Connor and the Three-Step Theory[61] by Klonsky and May. While there are discrepancies in factors that explain the transition from passive to active suicidal ideation, all three theories have the following factors in common for suicidal ideation: thwarted belongingness and hopelessness. For suicide attempt, reduced fear of death, increased tolerance of pain, and access to lethal means are common factors across the three theories of suicide.

Table 5. Table of theories of suicide with the Ideation-to-Action Framework.
Recurrent terms among the theories have been bolded.

	Theories of Suicide with Ideation-to-Action Framework		
	Interpersonal Theory (Joiner)	**Integrated Motivational-Volitional Theory (O'Connor)**	**Three-Step Theory (Klonsky & May)**
Passive Suicidal Ideation	**Thwarted belongingness** + Perceived burdensomeness	Defeat and humiliation, Entrapment	Pain (various source) + **Hopelessness**
Active Suicidal Ideation	+ **Hopelessness**	Threat-to-self moderators (in/decreases defeat & humiliation): social problem-solving, over-general autobiography, brooding rumination Motivational moderators (in/decreases likelihood that entrapment will lead to suicidal ideation and intent): **thwarted belongingness**, absence of positive future thinking	**Disconnectedness** (Connectedness is protective)
Suicide Attempt	+ **Fearlessness regarding death** + **Elevated pain tolerance** (i.e. past suicide attempt)	Volitional moderators (any factor that bridges suicidal ideation–attempt gap): impulsivity, intent, **access to the lethal means**, exposure to self-harm by friends or family, **fearlessness about death**	Acquired capability that increase suicide capacity: Dispositional: genetics, **pain sensitivity** Acquired: **habituation to experiences associated with pain, injury, fear, and death** Practical: concrete factors that make SA easier – **access to lethal means**

Environmental and Cultural Risk Factors of Suicide

Age is one of the major demographic factors to consider with suicide. According to the 2014 World Health Organization report on Suicide Prevention[62], adults aged 70 and older have the highest suicide rates. Rates of suicide are lower in children and young adults, but accounts for a disproportionately large number of deaths in those age groups. In a cross-national study derived from the World Mental Health Survey conducted in around 85,000 individuals across 17 countries, earlier age of onset of suicidal ideation was highly associated with a higher risk of suicide plan and attempt[63]. Across all 17 countries, risk of the first onset of suicidal ideation increased during adolescence and young adulthood and ages 18-34 years had the highest odds ratio of 9.5-12.4.

Table 6 summarizes the risk factors of by stage of life[64], organized by static and dynamic risk factors, where static risk factors are fixed attributes that establish a baseline risk of suicide and dynamic risk factors fluctuate throughout life.

Gender is another major demographic factor in suicide. Suicide rates of men are about three times higher than those of women, and this imbalance is greater in high-income countries[62]. However, the lifetime rates of suicidal ideation, plan, and attempt are higher in women than men[32,63]. Different cultural expectations of gender roles is a possible sociological explanation for these findings, such as men facing a cultural emphasis to be competitive and strong while women have a higher level of religiosity and extensive social support system thus providing them with better coping mechanisms[65]. Higher rates of alcohol abuse[66] and access to lethal methods[67] in men are other explanations to higher rates of suicide in men.

Race is a demographic factor in suicide. In the United States, suicide rates of non-Hispanic Blacks and Hispanics are less than half of those of non-Hispanic Whites[68]. Some however hypothesize that this is primarily explained by health-data disparities, where suicide data quality for Black and Hispanic individuals are lower because they are more likely to receive a potential suicide misclassification[69].

Income differences affect rates of suicide and alters sex and age patterns. Suicide rates are higher in high-income countries (12.7 vs 11.2 per 100,000 compared to low and middle-income countries) but deaths by suicide among high-income countries account for less than 25% of all suicides worldwide[62].

When comparing demographics between high-income countries with low and middle-income countries, middle-aged men of high-income countries have higher rates of suicide compared to the other countries, while young adults and elderly women of low and middle-income countries have higher rates of suicide than those in high-income countries[62].

Stigma is a cultural factor that can affect rates of suicide. In a study based on the Eurobarometer survey across 25 European countries, stigma was quantified with the survey question on social distance from a person with mental health problems[70]. Social acceptance of someone with mental health problem was shown to be negatively correlated with age standardized national suicide rates in the same year (β=0.46, p=0.014) when a linear regression model was applied controlling for socio-economic indicators.

Table 6: Risk factors by stage of life adapted from Steele et al.[64]
NSSI – non-suicidal self injury

Life Stage	Static Risk Factors	Dynamic Risk Factors
Across all ages	Male gender Personal history of prior suicide attempt; NSSI; physical or sexual abuse Family history of suicide	Current psychiatric disorder (depression most common) Psychological symptoms: insomnia, impulsivity Access to lethal means
Child/Adolescent (age 5-19)	Ages 12-19 LGBTQ sexual orientation Witness to violence, suicidal behavior or suicide Family history of psychiatric illness	Psychological symptoms: burdensomeness, active suicidal ideation Interpersonal conflicts with parents (children) or romantic partner (adolescents) Bullying Legal trouble/incarceration Current substance abuse Social isolation
Adult (age 20-64)	Caucasian Any diagnosed psychiatric disorder Military service History of arrest (additive risk) Less than high school education	Psychological symptoms: agitation, hopelessness Nonmarried status: single, divorced, widowed Active military: army, lower rank, current psychiatric illness, history of TBI, substance use Psychiatric hospitalization course: recently discharged, suicide attempt/self-harm during stay, unplanned discharge/short length of stay, attempted elopement Recent arrests or incarceration Recent loss of job/financial distress Current conflicts with romantic relationship
Geriatric (age 65+)	Increase risk with age Caucasian Chronic medical illness(es)	Psychological symptoms: burdensomeness, guilt, hopelessness, poor perception of health Acute medical illness(es) Current substance use Financial stress Social isolation

Psychiatric and non-psychiatric health risk factors of suicide

Currently, risk factors with the strongest evidence of epidemiological association with suicide include drug and alcohol misuse, the presence of a neuropsychiatric disorder, and a family history of STBs. Other significant risk factors include access to lethal means, adverse life events, diagnoses of chronic and/or terminal illness, previous SAs, and adverse childhood experiences.

Aside from a family history of suicidal thoughts and behaviors, presence of a neuropsychiatric disorders is an important risk factor for suicide[71]. The most common psychiatric disorders in people who die by suicide include major depressive disorder, bipolar disorder, schizophrenia, or substance use disorders[72–75]. After discharge from psychiatric hospitalization, the suicide rate among people with neuropsychiatric disorders is significantly higher, at 88 per 100,000[76], declining slowly over time. This rate is much higher than that of the general population and represents an opportunity for potential targeted treatment and prevention. Yet overall, 98% of people with psychiatric disorders do not die by suicide[77]. Thus, studying all suicide phenotypes, inside and outside the context of psychopathology, is going to be important for understanding the nature of suicide risk and related factors.

While psychiatric symptoms and diagnoses are major risk factors in suicide, non-psychiatric health risk factors also contribute significantly. In a review that calculated weighted odds ratios of major risk factors for suicide attempt, the mean odds ratio of physical illness was one of the risk factors categories that exceeded 2, and was comparable to the odds ratio of prior self-injurious thoughts and behaviors[78]. In a retrospective review of around 2,700 suicide cases that examined physical health conditions previously associated with suicide using diagnostic codes, authors found that nearly all physical health conditions increased suicide risk even after adjusting for potential confounders[79]. Having multiple physical health conditions further increased suicide risk, where having two or more diagnosed physical conditions had an Adjusted OR = 1.70 ($p < 0.001$) after adjusting for age, sex, and psychiatric condition[79]. After adjusting for sex, age, mental health, and substance use diagnoses, three individual diagnoses had odds ratios > 2: traumatic brain injury (AOR=8.80, $p< 0.001$); sleep disorders (AOR = 2.08, $p< 0.001$); and HIV/AIDS

(AOR = 2.14, p< 0.001)[79]. For the rest of this section, I will focus on the relationship between these three diagnoses and suicide.

Traumatic brain injury (TBI) is defined by the CDC as "a disruption in the normal function of the brain that can be caused by a bump, blow, or jolt to the head, or a penetrating head injury"[80]. The association between TBI and suicide rose to prominence with reports of higher rates of suicide in professional athletes and military personnel who sustained concussions[81] and has since been well recognized as a major risk factor of suicide[82]. In a retrospective Swedish national registry cohort study, the incidence rate ratio (IRR) of suicide was approximately two times greater in patients with TBI than individuals without TBI, regardless of TBI severity and after adjusting for potential confounding factors (i.e. sociodemographic factors, pre-TBI psychiatric illness, and pre-TBI deliberate self-harm)[83]. Suicide risk conferred by TBI was additive (risk is higher in those with multiple TBIs compared to those with one) and lasting. While the risk of suicide was higher in the first six months after TBI (IRR 3.67), the risk remained elevated even after 7 years (IRR 1.76)[83]. The increased suicide risk is observed even with lower TBI severity. A recent meta-analysis has shown that individuals diagnosed with mild TBI/concussion had around a two-fold higher relative risk (RR = 2.03, p < 0.001) compared to individuals without a diagnosis of mild TBI/concussion[84]. Many studies have linked affected brain regions with functional regions to provide a neuroanatomical explanation to the link between TBI and suicide. Linked brain regions include the orbitofrontal region, thought to affect impulsivity and disinhibition that are features of the neuropsychosocial syndrome associated with TBI; anterior and posterior medial prefrontal cortex, thought to affect rumination; and the anterior cingulate cortex known to be important for emotional regulation[85]. In addition to direct injury during impact, neuroinflammation is also thought to contribute to the increased suicide risk among TBI patients, as evidenced by the growing attention of suicide risk in patients with chronic traumatic encephalopathy[86].

Sleep disturbance and sleep disorder are another risk factor of suicidal behavior, however findings have been inconsistent in part due to the variable definitions and measurements of insomnia and suicide-related behaviors[87–89]. The field is divisive in terms of whether the association between insomnia and suicide

is directly causal or increased risk is primarily mediated by another mechanism, such as serotonergic dysfunction, or mood dysregulation including depression[88].

HIV/AIDS is not a well-established risk factor of suicide and studies of this association have low power[90]. Bigger sample sizes and comparison of matched controls accounting for potential confounders such as substance use will be required to better understand the association between HIV/AIDS and suicide.

Genetic studies of suicide attempt and suicidal ideation

In the past five years, there has been several well-powered and notable GWAS of suicide attempt and suicidal ideation that shed insight to the genetic architecture of suicide-related traits and genetic overlap with psychiatric and non-psychiatric comorbidities.

Erlangesen et al.[91] performed a GWAS on 6,024 suicide attempt cases ascertained using ICD-10 codes of suicide attempt among individuals with various severe psychiatric disorders including affective disorders, bipolar disorder, and schizophrenia in a Danish national cohort. The GWAS was performed against controls without psychiatric disorders, and to account for and quantify the genetic contribution of multiple psychiatric disorders, another GWAS was performed with binary covariates for diagnoses of psychiatric disorders were included. Three genome-wide significant SNPs were identified (**Table 7**). SNP-heritability estimate of suicide attempt with psychiatric disorders was 4.6% [CI-95 2.9-6.3%], and this estimate decreased to 1.9% [CI-95 0.3-2.5%] when psychiatric disorders were added as binary covariates. These results demonstrate that while there is a large genetic contribution of psychiatric disorders to suicide attempt, genetic risk of suicide attempt is not explained by psychiatric disorders alone.

Strawbridge et al.[92] performed a GWAS on suicidal ideation and attempt in the UK Biobank population, ascertained using self-reported suicide-related traits through their online mental health survey. An ordinal GWAS was performed comparing controls with phenotypes ranging from “thoughts that life is not worth living”, self-harm ideation, self-harm behavior, to suicide attempt. Most of the cases were self-harm ideation cases, and suicide attempt comprised only 6.8% of the ordinal phenotype. Three genome-wide significant loci were identified in chromosomes 9, 11, and 13 (**Table 7**). Observed scale heritability estimate of suicide-related traits was 7.6% (SE 0.006). Significant genetic correlations were observed with

MDD (rg=0.81), anxiety disorder (rg=0.75), neuroticism (rg=0.63), and mood instability (rg=0.50). PRS of suicide-related traits was also significantly associated with mood disorders (bipolar disorder and MDD), as well as mood instability, neuroticism, and risk-taking behavior.

Campos et al.[93] performed a GWAS on 23,192 self-harm ideation and 6,872 self-harm behavior cases ascertained in UK Biobank. One genome-wide significant locus was each identified for self-harm ideation and self-harm behavior (**Table 7**). SNP heritability estimates on the liability scale was 11.1% for self-harm ideation (SE=1.7%, population prevalence = 14.8%) and 10.1% for self-harm behavior (SE=1.0%, population prevalence =4.4%). Self-harm ideation and behavior showed significant genetic correlation with each other (rg=0.85, $p=7.8x10^{-53}$) and both traits showed significant positive genetic correlations with multiple psychiatric disorders including anxiety, depression and schizophrenia, and non-psychiatric traits such as insomnia and risk-taking behavior. PRS analysis on an independent Australian cohort showed significant association of self-harm ideation PRS with non-suicidal self-harm ($p=3.6x10^{-4}$) and suicidal ideation($p=4.5x10^{-6}$), while nominally significant ($p<0.05$) association was observed with self-harm behavior PRS with suicidal ideation and suicide attempt of the target population.

Ashley-Koch et al. [in review] studied suicidal ideation in the US veteran population in the Million Veterans Program (MVP) cohort. Cases were ascertained using ICD9 and ICD10 codes for intentional self-harm in the electronic health record, suicide behavior reports, mental health survey responses, and the National Death Index. Among these, mental health surveys and diagnostic codes were the major sources of ascertainment. GWAS of cases for multiple ancestries were performed as well as a trans-meta-analyses. Two genome-wide significant loci were identified for the GWAS with individuals of European ancestry, and five additional genome-wide significant loci were identified for the pan-ancestry meta-analysis (**Table 7**). SNP-heritability of suicidal ideation among individuals of European ancestry was estimated to be 2.2% (SE 0.0016) in the observed scale and 1.2% (SE 0.0009) in the liability scale. There was high and significant genetic correlation of suicidal ideation and suicide attempt within MVP (rg=0.77, SE=0.05, $p=2.15x10^{-53}$).

These studies of suicidal ideation and suicide attempt highlight the shared genetic architecture between suicide-related traits as well as with psychiatric and non-psychiatric risk factors. Further

investigation is needed to understand how the genetic contribution of psychiatric traits affect the genetic architecture of suicide-related traits, and whether genetic risk of suicidal ideation and suicide attempt is associated with individuals who present with multiple suicide-related traits.

Table 7: Summary of genome-wide significant loci identified in studies of suicide-related behaviors.

Trait	GWS snp	CHR	A1/A2	BETA	SE	P	Nca	Nco	N	Source
Suicide attempt adjusting for psychiatric disorder	rs4809706	20	A/G	-0.052	0.042	2.80E-08	6,024	44,240	50,264	Erlangsen et al. 2018
Ordinal suicidal ideation and suicide attempt	rs62535711	9	T/C	0.105	0.018	1.29E-08	39,378	83,557	122,935	Strawbridge et al. 2019
	rs598046	11	T/G	0.053	0.009	1.07E-08				
	rs7989250	13	A/C	-0.052	0.009	3.49E-08				
Self-harm ideation	rs4865733	5	T/C	-0.008	0.001	1.90E-08	133,524	23,192	156,716	Campos et al. 2020
Self-harm behavior	rs567805973	9	C/T	-0.046	0.008	2.10E-08	6,872	150,008	156,880	
Suicidal ideation (European ancestry)	rs13211166	6	A/T	-0.058	0.010	2.34E-09	62,023	376,826	438,849	Ashley-Koch et al. 2022
	rs73581580	9	A/G	0.065	0.010	1.08E-10				

Trait	GWS snp	CHR	A1/A2	Z score	P	Nca	Nco	N	Source
Suicidal ideation (pan-ancestry)	rs112982323	8	G/GTT	-6.086	1.16E-09	99,814	512,567	612,381	Ashley-Koch et al. 2022
	rs77641763	9	C/T	5.925	3.12E-09				
	rs7185007	16	C/T	-5.835	5.37E-09				
	rs6557168	6	C/T	-5.708	1.14E-08				
	rs142785607	2	T/G	5.620	1.91E-08				

CHAPTER II

Genetic Architecture of Suicide Attempt[†]

Introduction

Suicide is a worldwide public health problem, accounting for nearly 800,000 deaths per year[94]. Suicide attempt (SA), defined as a non-fatal self-injurious behavior with the intent to die, has been estimated to occur over 20 times more frequently and is a major source of disability, reduced quality of life, and social and economic burden[68,94]. The lifetime prevalence of SA in adults ranges from 0.5-5% worldwide[63]. There are several well-established comorbidities and risk factors for SA, with the presence of a psychiatric disorder having the strongest effect on lifetime suicide rates[71,95]. However, the vast majority of patients with psychiatric disorders never attempt suicide[96–99]. Other major risk factors for SA include prior self-injurious thoughts and behaviors[100], physical illness or disability[79,101], sleep disorders[87–89,102], family history of psychiatric disorders[103], substance abuse[73], smoking[104–106], impulsivity and social factors such as childhood maltreatment[107], isolation[108], and stressful life events[109].

Both suicide and SA are moderately heritable, with estimates from genetic epidemiology studies in the range of 17-55%[110–112]. Several genome-wide association studies (GWAS) of SA have reported significant SNP-heritability estimates of ~4%, pointing to an underlying polygenic architecture[91,92,113,114]. Using polygenic risk scoring or genetic correlation analyses, these studies have also demonstrated shared genetic etiology between SA and psychiatric disorders, with major depressive disorder (MDD) showing the largest genetic overlap[92,113,115]. This genetic overlap, along with the prevalence of MDD in the population[116] make it a particularly salient risk factor. Importantly, genetic epidemiology studies have consistently indicated a genetic component of SA which is partially distinct from that of psychiatric disorders[111]. One

[†] Adapted from Mullins N, Kang J et al., Biological Psychiatry, 2021

GWAS of SA adjusted for the presence of a psychiatric disorder and estimated a SNP heritability of 1.9%[91], suggesting that the genetic etiology of SA is likely to comprise genetic variants which confer risk more strongly to SA than psychiatric disorders, as well as variants that confer risk more strongly to psychiatric disorders than SA.

Few genetic samples have been collected specifically for SA, with studies often relying on individuals ascertained for psychiatric disorders. For example, a large GWAS of SA included over 6,500 cases from clinical cohorts of depression, bipolar disorder and schizophrenia cases, within the Psychiatric Genomics Consortium[115]. In a "SA within psychiatric diagnosis" study design, SA cases were compared with cases of the same psychiatric disorder without SA, in order to disentangle the genetic etiology of SA and psychiatric disorders. While GWAS of SA have found genome-wide significant associations[91,92,113–115], thus far none have been replicated, possibly due to limited statistical power or different study designs which may probe varying components of the genetic etiology of SA. Depending on the method of ascertainment, the prevalence of psychiatric disorders may be much higher in SA cases than controls in these studies, which may confound the genetics of SA. Well-powered and carefully designed studies are necessary to advance our understanding of the genetics of SA and dissect the contribution of genetic variation to SA versus psychiatric disorders.

Here, we present the first collaborative GWAS meta-analysis of SA from the International Suicide Genetics Consortium, including over 29,000 cases of suicide or SA from 15 institutes or consortia worldwide. We identify novel loci implicated in SA, disentangle the genetic etiology of SA from that of MDD and psychiatric disorders and characterize the genetic relationship between SA, psychiatric disorders, and a range of known risk factors.

Methods

Cohorts and case definition

This study included 21 cohorts worldwide, many of which have been described previously (Supplementary Table 1, Supplementary Note). These included cohorts ascertained for psychiatric disorders, including substance use (15 cohorts), studies of suicide or SA (4 cohorts), and population-based biobanks (2 cohorts). Cases were individuals who died by suicide (2 cohorts) or made a non-fatal suicide attempt (19 cohorts). A non-fatal suicide attempt was defined as a lifetime act of deliberate self-harm with intent to result in death. Information on SA was ascertained using structured clinical interviews for 15 cohorts, self-report questionnaires for 2 cohorts, and hospital records or International Classification of Diseases (ICD) codes for 2 cohorts. Cases of death by suicide (2 cohorts) were ascertained from the Utah State Office of the Medical Examiner or the Medical Examiner's Office of the Hyogo Prefecture and the Division of Legal Medicine, at the Kobe University Graduate School of Medicine in Japan. A proportion of cases from the iPSYCH and Columbia University cohorts were also individuals who had died by suicide, determined using the Cause of Death Register in Denmark and The Columbia Classification Algorithm for Suicide Assessment respectively[117]. Individuals endorsing suicidal ideation only were not included as cases. There were 14 cohorts of European ancestry, 2 cohorts of admixed African American ancestry, and 5 cohorts of East Asian ancestry. All individual studies received institutional and ethical approval from their local institutional review board (Supplementary Table 1). Detailed information on the ascertainment and case definition for each cohort is included in the Supplementary Note. Supplementary Table 1 contains an overview of cohort characteristics.

Control definition

For the primary GWAS, controls included all individuals with no evidence of SA, including those ascertained for having a psychiatric disorder. Controls from the general population were screened for the absence of SA if such information was available; however since the prevalence of SA in the general population is low (~2%)[63], some cohorts included unscreened controls. Amongst controls ascertained for having a psychiatric disorder, all were screened for the absence of lifetime SA. Controls from the general

population were not screened for the absence of psychiatric disorders and no controls were screened for suicidal ideation. A GWAS of SA within psychiatric diagnosis was also conducted, where controls were individuals with the same psychiatric disorder as the SA cases in each cohort, and were all screened for the absence of lifetime SA. Cohorts were included in the GWAS of SA in the general population and/or the GWAS of SA within psychiatric diagnosis, depending on the characteristics of the controls available, and therefore there is some overlap of individuals and cohorts between the GWAS. The primary GWAS of SA included 29,782 cases and 519,961 controls from 18 cohorts and the GWAS of SA within psychiatric diagnosis included 14,847 cases and 69,951 controls from 13 cohorts (**Table 8**).

Genotyping, quality control and imputation

Cohorts were required to have a minimum of 200 cases prior to quality control for inclusion in the GWAS meta-analysis. Samples underwent standard genotyping, quality control and imputation, according to the local protocol for each study. Briefly, samples were genotyped on microarrays with the exception of one study (CONVERGE) that used low-coverage sequencing. Parameters used to retain individuals and SNPs after quality control for missingness, relatedness and Hardy-Weinberg equilibrium are outlined in the Supplementary Note. Imputation was performed using the appropriate ancestry reference panels, resulting in > 7.7 million SNPs that were well-represented across cohorts. Full details of the genotyping, quality control and imputation for each cohort are available in the Supplementary Note. Identical individuals between the Psychiatric Genomics Consortium (PGC) and UK Biobank cohorts were detected using genotype-based checksums[‡] and removed from PGC cohorts. There was no other known overlap of controls remaining between any of the 21 cohorts after QC.

Genome-wide association study

GWAS were performed in each cohort separately by the collaborating research team and analysis procedures are outlined in the Supplementary Note. GWAS were conducted within ancestry group,

[‡]https://personal.broadinstitute.org/sripke/share_links/zpXkV8INxUg9bayDpLToG4g58TMtjN_PGC_SCZ_w3.0718d.76

covarying for genetic ancestry-informative principal components (PCs), genomic relatedness matrices or factors capturing site of recruitment or genotyping batch, as required. The LD Score regression (LDSC) intercept was calculated for all GWAS results to assess potential confounding from cryptic relatedness or population stratification[118]. For any studies with a significant intercept ($P<0.05$), the GWAS summary statistics were corrected for confounding by multiplying the standard error per SNP by the square root of the LDSC intercept[118]. A meta-analysis of GWAS summary statistics was conducted using an inverse variance-weighted fixed effects model (standard error) in METAL[119], implemented using the Rapid Imputation for COnsortias PIpeLIne (RICOPILI)[120], for the GWAS of SA in the general population, and the GWAS of SA within psychiatric diagnosis. The meta-analyses were performed across all cohorts regardless of ancestry. The weighted mean allele frequency and imputation INFO score per SNP was calculated, weighted by the effective sample size per cohort. SNPs with a weighted minor allele frequency of < 1%, weighted imputation INFO score < 0.6 or SNPs present in < 80% of total effective sample size were removed from the meta-analysis results. A genome-wide significant locus was defined as the region around a SNP with $P<5.0x10^{-8}$ with linkage disequilibrium (LD) $r^2 > 0.1$, within a 3,000 kilobase (kb) window, based on the LD structure of the Haplotype Reference Consortium (HRC) European ancestry reference panel v1.0[121].

mtCOJO

The results of the GWAS of SA were conditioned on the genetics of MDD using mtCOJO (multi-trait-based conditional & joint analysis using GWAS summary data)[122], implemented in GCTA software[123]. mtCOJO[122] estimates the effect size of a SNP on an outcome trait (e.g., SA) conditioned on exposure trait(s) (e.g., MDD). It first uses the genome-wide significant SNPs for the exposure trait as instruments to estimate the effect of the exposure on the outcome, and then performs a genome-wide conditioning of the estimated effect from the exposure, resulting in conditioned effect sizes and P values for the outcome trait. We conditioned SA on MDD, since MDD is the most prevalent psychiatric disorder among individuals who die by suicide[124] and has the highest genetic correlation with SA among psychiatric disorders ($rg=0.44$)[113]. mtCOJO analysis was performed on the SA as the outcome trait. For this, GWAS summary statistics from

the European-only subset of the SA meta-analysis were used (26,590 cases and 492,022 controls), since mtCOJO requires an ancestry-matched LD reference panel. The PGC MDD GWAS summary statistics (excluding 23andMe)[17] were used for the exposure trait. mtCOJO is robust to overlap in samples contributing to the GWAS of the exposure and outcome. In the selection of SNPs as instruments, independence was defined as SNPs more than 1 megabase (Mb) apart or with an LD r^2 value < 0.05 based on the 1000 Genomes Project Phase 3 European reference panel[125]. To obtain at least 10 independent instruments for MDD, the genome-wide significance threshold was adjusted to $P<5.0x10^{-7}$, leading to 15 SNPs used. In a further sensitivity analysis, GWAS summary statistics for bipolar disorder (BIP)[126] and schizophrenia (SCZ)[127] were additionally included as exposure traits.

LD Score regression (LDSC)

LDSC[118] was used to estimate the phenotypic variance in SA explained by common SNPs (SNP-heritability, h^2_{SNP}) from GWAS summary statistics. $h^2_{SNP}$was calculated on the liability scale assuming a lifetime prevalence of SA in the general population of 2%, which is the middle of the range reported worldwide[63]. For the GWAS of SA within psychiatric diagnosis, $h^2_{SNP}$was calculated on the liability scale using a prevalence of SA in psychiatric populations ranging from 10-20%. LDSC bivariate genetic correlations attributable to genome-wide SNPs (rg) were estimated between all GWAS of SA and between each GWAS of SA and a range of psychiatric disorders, self-harm ideation and propensity towards risk-taking behavior (risk tolerance), using the largest available GWAS summary statistics (Supplementary Table 11). The Bonferroni corrected significance threshold was P<0.0042, adjusting for 12 traits tested. The difference between the rg of SA before and after conditioning on MDD was tested for deviation from 0, using the block jackknife method, implemented by the LDSC software[128]. The rg of each SA GWAS with 768 other non-overlapping human diseases and traits was calculated on LD Hub (http://ldsc.broadinstitute.org)[129] (Bonferroni corrected significance threshold $P<6.51x10^{-5}$ for each GWAS). Before analysis, traits were categorized manually into risk factor groups previously ascribed to SA[71,78,79]: autoimmune disease, neurologic disease, heart disease, hypertension, diabetes, kidney disease, cancer, alcohol use, smoking, pain, psychiatric, sleep, life stressors, socioeconomic, and education/cognition (Supplementary Table 12).

A second reviewer validated the categories assigned to traits and their relevance to SA risk. Overlapping traits were appended.

Gene-based, gene-set and tissue-set enrichment analyses

P values quantifying the degree of association of genes and gene-sets with SA based on the GWAS of SA in the general population were generated using MAGMA (v1.08), implemented in FUMA (v1.3.6a) (https://fuma.ctglab.nl)[130,131]. Gene-based tests were performed for 18,517 genes and a Bonferroni correction was applied for the number of genes tested ($P<2.70x10^{-6}$). A total of 11,638 curated gene sets from MSigDB V7.0 were also tested for association with SA (Bonferroni-corrected significance threshold $P<4.30x10^{-6}$). Competitive gene-set tests were conducted correcting for gene size, variant density and LD within and between genes. Gene-sets including < 10 genes were excluded. Finally, tissue-set enrichment analyses were performed using MAGMA[131] implemented in FUMA[130], to test for enrichment of genetic associations with SA in genes expressed in 54 tissues from the Genotype-Tissue Expression (GTEx) project V8[132]. The significance threshold was $P<9.26x10^{-4}$, adjusting for the number of tissues tested.

Integrative eQTL analysis

A transcriptome-wide association study (TWAS) was conducted using FUSION software[133] and precomputed expression reference weights from PsychENCODE data[134]. The PsychENCODE Consortium has conducted a genome-wide eQTL analysis using 1,321 brain samples, predominantly from the dorsolateral prefrontal cortex[134]. For genes with significant *cis*-SNP heritability (13,435 genes), a TWAS was performed to test whether SNPs influencing brain gene expression are also associated with SA, using the meta-analysis results from the GWAS of SA in the general population (Bonferroni corrected significance threshold $P<4.28x10^{-6}$).

Polygenic risk scoring analysis

Polygenic risk scores (PRS) for SA were tested for association with SA or death by suicide in independent target cohorts. The target cohorts used were PGC MDD, PGC BIP, PGC SCZ, CONVERGE (a cohort of East Asian ancestry), and the University of Utah cohort (a sample of individuals who died by suicide). The meta-analysis of SA was repeated excluding each of these cohorts in turn, to create independent discovery

and target datasets. PRS were tested for association with SA versus controls in all five of the target samples. PRS were also tested for association with SA within psychiatric diagnosis in the PGC MDD, BIP and SCZ samples. Analyses in the PGC datasets were repeated using the PRS for SA in the general population generated from the GWAS results after conditioning on MDD. The Bonferroni corrected significance threshold is $P<3.57x10^{-3}$, correcting for 14 tests. The analyses performed are summarized in Supplementary Table 2.

PRS analyses were performed using PRS-CS[135], a method which uses a Bayesian regression framework and places a continuous shrinkage prior on the effect sizes of all SNPs in the discovery GWAS summary statistics. Continuous shrinkage priors allow a specific amount of shrinkage to be applied to each SNP, which is adaptive to the strength of its association signal in the discovery GWAS and the LD structure from an external reference panel[135]. The 1000 Genomes European or East Asian reference panels[125], as appropriate, were used to estimate LD between SNPs. PRS were calculated for each individual in the target cohorts using standard procedures. PLINK 1.9[136] was used to weight all SNPs by their effect sizes calculated using PRS-CS and sum all SNPs into PRS for each individual in the target cohort. PRS were tested for association with case versus control status in the target cohort using a logistic regression model, covarying for PCs, genomic relatedness matrices or factors capturing site of recruitment or batch effects, as required. The amount of phenotypic variance explained by the PRS (R^2) was calculated on the liability scale, which accounts for the proportion of cases in the target sample and the proportion of cases in the population[137]. Calculations assumed a lifetime prevalence of SA in the general population of 2%[63] and a lifetime prevalence of SA in MDD, BIP, and SCZ of 16%, 37% and 36% respectively. These numbers represent the observed prevalence of SA in these disorders in the PGC cohorts.

Pairwise GWAS

Pairwise GWAS[138] was used to investigate genome-wide significant loci for SA and overlapping causal variants with propensity towards risk-taking behavior[139] and lifetime smoking index[140]. These phenotypes were chosen because they share genome-wide significant loci in the same region as the genome-wide significant locus on chromosome 7 in the GWAS of SA and SA conditioned on MDD. The genome-wide

significant locus on chromosome 6 is in the major histocompatibility complex and due to the complex long-range LD of this region, it was not included for this analysis. Pairwise GWAS uses association statistics from two GWAS to estimate the probability that a genomic region 1) contains a genetic variant that influences only the first trait, 2) contains a genetic variant that influences only the second trait, 3) contains a shared causal or pleiotropic variant, and 4) contains two independent variants in the same region, one influencing the first trait and the other influencing the second trait. The GWAS summary statistics from the European-only subset of the SA meta-analysis (26,590 cases and 492,022 controls) were used for Pairwise GWAS as the method requires an ancestry-matched LD reference panel. The genome was divided into approximately independent LD blocks based on patterns of LD in the European population in Phase 1 of the 1000 Genomes Project, as previously described[138]. We divided the 3 Mb-wide genome block containing the genome-wide significant locus for SA on chromosome 7 into two blocks to separate the two independent causal variants for risk-taking behavior in that region (rs8180817 and rs4275159, LD r^2=0.001)[139]. The fgwas package[141] was used to determine the baseline correlation between the two GWAS by extracting all genomic regions with a posterior probability of containing an association less than 0.2 and calculating the correlation in the Z-scores between the two GWAS. This summary statistic-level correlation was used as a correction factor to each Pairwise GWAS analysis.

Results

Study description and samples analyzed

We conducted a primary GWAS meta-analysis of SA (29,782 cases, 519,961 controls) from 18 cohorts (**Table 8**), which included both population-based and clinically ascertained samples for psychiatric disorders (see Methods). The majority (n=26,590) of cases were individuals of European ancestries but cases also included 1,894 individuals of East Asian ancestries and 1,298 individuals of admixed African American ancestries. Case definition was lifetime SA, with ~20% (n=5,438) of cases having died by suicide (see Methods). To investigate the shared and divergent genetic architectures of SA and psychiatric disorders, we performed two additional analyses. We conditioned our primary GWAS results using GWAS summary statistics for MDD, to remove the genetic effects mediated by MDD, the most commonly comorbid psychiatric disorder with SA. Furthermore, we conducted a GWAS of SA versus no SA among individuals with a psychiatric diagnosis in 14,847 cases and 69,951 controls from 13 cohorts.

Table 8: Numbers of cases and controls for 21 cohorts in the International Suicide Genetics Consortium

	SA*		SA within psychiatric diagnosis*	
Cohort	**Cases**	**Controls**	**Cases**	**Controls**
Psychiatric Genomics Consortium MDD	1,528	16,626	1,677	8,793
Psychiatric Genomics Consortium BIP	3,214	17,642	3,214	5,408
Psychiatric Genomics Consortium SCZ	1,640	7,112	1,668	2,928
Psychiatric Genomics Consortium ED	170	5,070	198	583
Army STARRs	670	10,637	376	3,447
German Borderline Genomics Consortium	481	1,653	481	144
UK Biobank	2,433	334,766	2,149	35,912
iPSYCH	7,003	52,227	-	-
Janssen	255	1,684	-	-
Yale-Penn	475	1,817	-	-
GISS Ukraine	660	660	-	-
Columbia University	577	1,233	-	-
Australian Genetics of Depression Study and QSkin Study	2,792	20,193	2,792	8,718
University of Utah	4,692	20,702	-	-
Japan (EAS)	746	14,049	-	-
CONVERGE (EAS)	1,148	6,515	1,148	1,183
Taiwan MDD (EAS)	-	-	222	318
Taiwan BIP (EAS)	-	-	235	397
Taiwan SCZ (EAS)	-	-	332	1,004
Grady Trauma Project (admixed AA)	669	4,473	355	1,116
Yale-Penn (admixed AA)	629	2,902	-	-
Total	**29,782**	**519,961**	**14,847**	**69,951**

SA - suicide attempt, MDD - major depressive disorder, BIP - bipolar disorder, SCZ - schizophrenia, ED - eating disorder, EAS - East Asian ancestry, AA - African American ancestry. All other samples are of European ancestries. *GWAS contain some overlapping individuals and cohorts.

SA shows significant SNP-heritability and polygenic risk association with death by suicide

In the primary GWAS of SA, h^2_{SNP}estimated using LDSC was 6.8% (SE=0.005, P=2.00x10^{-42}) on the liability scale. The genomic inflation factor (λ_{GC}) was 1.23, the LDSC intercept was 1.04 (SE=0.01, P=2.84x10^{-4}) and the LDSC attenuation ratio was 0.14 (SE=0.04), indicating that the majority of inflation of the GWAS test statistics was due to polygenicity. PRS for SA were tested in five target SA cohorts, which were each excluded in turn from the discovery GWAS to ensure independent discovery and target samples (Supplementary Table 2). SA PRS were significantly associated with SA in the PGC MDD, PGC BIP and PGC SCZ cohorts, with a phenotypic variance explained (R^2) of 0.69% (P=7.17x10^{-15}), 0.68% (P=8.11x10^{-28}) and 0.88% (P=1.24x10^{-17}) respectively, on the liability scale. PRS for SA were also associated with death by suicide in the University of Utah cohort, explaining slightly more phenotypic variance (R^2=1.08%, P=9.79x10^{-81}). The genetic correlation between the University of Utah GWAS of

suicide death and SA from a meta-analysis of the remaining cohorts in our study was 0.77 (SE=0.08, $P=1.54x10^{-20}$). Examining the performance of SA PRS across ancestry showed a significant association with SA in the CONVERGE East Asian cohort, although with a lower variance explained (R^2=0.25%, $P=3.06x10^{-3}$) (Supplementary Table 2).

GWAS of SA identifies locus with stronger effect on SA than psychiatric disorders

The GWAS of SA identified two genome-wide significant loci ($P<5x10^{-8}$) (Figure 1a, Supplementary Table 3). The locus most highly associated with SA was in an intergenic region on chromosome 7 (index SNP rs62474683, OR for A allele = 1.06 [1.04-1.08], $P=1.91x10^{-10}$, frequency in SA cases = 0.52). The second genome-wide significant locus was in the major histocompatibility complex (MHC) (index SNP rs71557378, OR for T allele = 1.10 [1.06-1.13], $P=1.97x10^{-8}$, frequency in SA cases = 0.91). After conditioning the genetic effects of SA (European-only subset) on the genetic effects of MDD using mtCOJO, only the chromosome 7 locus remained genome-wide significant (index SNP = rs62474683, OR for A allele = 1.06 [1.04-1.08], $P=1.33x10^{-8}$, **Figure 1a**). In the GWAS of SA within psychiatric diagnosis, this index SNP had a slightly smaller effect size on SA (index SNP = rs62474683, OR for A allele = 1.04 [1.01-1.07], P=0.007), but no SNPs reached genome-wide significance in this analysis. Examining the intergenic locus on chromosome 7 in published GWAS results using Open Targets Genetics web portal[142] (https://genetics.opentargets.org), showed smaller and non-significant effects on all psychiatric disorders tested (**Figure 1b**). However, the index SNP from our SA GWAS has been implicated at genome-wide significance in lifetime smoking index[140] (which accounts for duration and amount of smoking), and propensity towards risk-taking behavior[139], although with smaller effect sizes than on SA (**Figure 1b**, Supplementary Table 4, Supplementary Table 5). Pairwise GWAS analysis on the genomic region containing the chromosome 7 locus indicated that the causal variant is most likely shared between SA and these phenotypes (lifetime smoking index: posterior probability = 0.997, risk-taking behavior: posterior probability = 1) (Supplementary Table 13). Furthermore, a variant in high LD with the index SNP on chromosome 7 (rs12666306, LD r^2=0.94) has a positive genome-wide significant effect on insomnia (reported in GWAS catalog, full summary statistics not available) (**Figure 1b**, Supplementary Table 4,

Supplementary Table 5). The index SNP for SA has also been implicated in self-harm ideation[93], although with a smaller effect size than on SA (**Figure 1b**).

Enrichment analyses using MAGMA[131] and the GWAS results for SA indicated significant enrichment of SA associations in 7 genes (Supplementary Table 6), including *BTN2A1* which is a brain-expressed gene[132] located within the MHC, that encodes a plasma membrane protein. There was no enrichment of SA association signal in any of the biological gene sets (Supplementary Table 7) or in the set of genes expressed in any of the 54 GTEx tissues tested (Supplementary Table 8). Examining individual genes, our transcriptome-wide association study (TWAS) found 5 genes for which SA risk alleles were significantly associated with brain gene expression: *ERC2, RP11−266A24.1, TIAF1, BACE2, NUFIP2* ($P<4.28x10^{-6}$) (Supplementary Table 9). None of the TWAS significant genes were located in genome-wide significant loci.

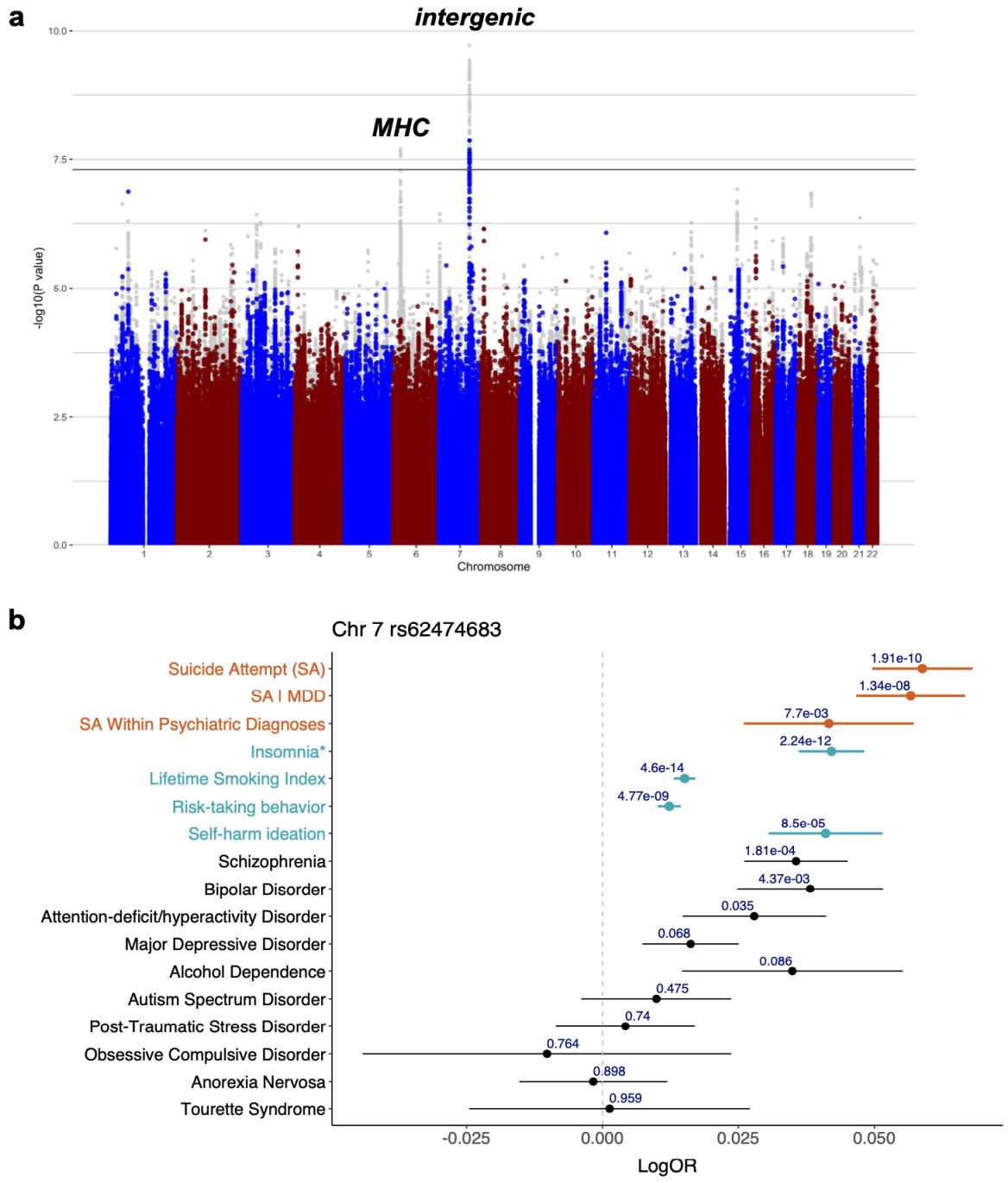

Figure 1: Genome-wide significant locus contributes to suicide attempt more strongly than psychiatric disorders and other traits

a) Manhattan plot: The *x*-axis shows genomic position and the *y*-axis shows statistical significance as $-\log_{10}$(P value). The grey points in the background depict the GWAS results for SA and the colored points in the foreground depict the results after conditioning SA on major depressive disorder (MDD), which was performed on the European meta-analysis results. The horizontal line shows the genome-wide significance threshold ($P<5.0\times10^{-8}$). b) Forest plot: The points indicate the log odds ratio of the A allele at rs62474683 (index SNP for SA on chromosome 7) on each phenotype and the error bars show the standard error. The P value of association with each phenotype is shown above the error bars. For insomnia, the effect size of a variant in high LD with the index SNP is shown instead (rs12666306 A allele, LD $r^2=0.94$ with SA index SNP).

Evidence for substantial proportion of SNP-heritability of SA not mediated by psychiatric disorders

We employed two approaches to assess the genetic architecture of SA after accounting for psychiatric disorders: 1) we statistically conditioned out genetic effects mediated by MDD and 2) we directly analyzed SA versus no SA among psychiatric disorder cases (see Methods). The statistical conditioning was performed on the European-only subset of the meta-analysis, in which the h^2_{SNP} of SA was 7.5% (SE=0.006, P=3.02x10^{-40}) on the liability scale (Supplementary Table 10). Conditioning these SA GWAS results on MDD resulted in a 45% decrease in the h^2_{SNP} of SA to 4.1% (SE=0.005, P=1.20x10^{-16}) on the liability scale (Supplementary Table 10). This conditioned estimate was comparable with estimates of the h^2_{SNP} of SA within psychiatric diagnosis, which ranged from 3.7% to 4.6%, using a prevalence of SA in psychiatric populations from 10-20% (P<1.35x10^{-3}). Conditioning SA on BIP and SCZ in addition to MDD did not change the h^2_{SNP} estimate (h^2_{SNP}=4.1%, SE=0.005, P=1.20x10^{-16}).

The genetic correlation between the GWAS of SA and SA within psychiatric diagnosis was 0.93 (SE=0.09, P=5.35x10^{-24}). PRS for SA were significantly associated with SA within psychiatric diagnosis in the PGC cohorts, with an R^2 of 0.43% (P=5.83x10^{-6}), 0.81% (P=2.33x10^{-11}) and 0.71% (P=5.78x10^{-6}) on the liability scale for SA in MDD, BIP and SCZ respectively (Supplementary Table 2). After conditioning the GWAS of SA on MDD, the genetic correlation with the GWAS of SA within psychiatric diagnosis was not significantly different from 1 (rg=1.13, SE=0.13) (Supplementary Table 10). After conditioning on MDD, PRS for SA remained significantly associated with SA within psychiatric diagnosis in the PGC cohorts, with slightly lower phenotypic variance explained (0.32%, 0.67% and 0.46% for SA in MDD, BIP and SCZ respectively), consistent with the reduction in h^2_{SNP} (Supplementary Table 2).

Significant genetic overlap between SA and psychiatric traits or disorders

Genetic correlations were calculated to explore the genetic overlap between SA and 12 psychiatric traits or disorders, before and after conditioning on MDD. SA showed a significant genetic correlation with 11 traits or disorders tested, most strongly with self-harm ideation (rg=0.81, SE=0.06, P=3.52x10^{-36}) and MDD (rg=0.78, SE=0.03, P=5.82x10^{-112}) (**Figure 2**, Supplementary Table 11). Significant genetic correlations

were also observed between SA and SCZ, attention-deficit/hyperactivity disorder (ADHD), BIP, post-traumatic stress disorder (PTSD) and alcohol dependency (rg=0.46-0.73) (**Figure 2**, Supplementary Table 11).

To investigate whether these genetic correlations were mediated by the genetics of MDD, we estimated genetic correlations with the same traits and disorders after conditioning the GWAS of SA on MDD (SA|MDD). Genetic correlations with all psychiatric disorders remained significant after conditioning except for autism spectrum disorder (ASD) and Tourette syndrome (Figure 2, Supplementary Table 11). As expected, the rg with MDD significantly decreased after conditioning ($P=2.3x10^{-16}$ block jackknife), as well as the rg with self-harm ideation ($P=1.3x10^{-4}$ block jackknife) and ASD ($P=1.8x10^{-5}$ block jackknife) (Figure 2, Supplementary Table 11). The remaining psychiatric disorders did not show significant differences in rg after conditioning on MDD, after Bonferroni correction. Since conditional analysis only removes SNP effects on SA mediated by MDD, the remaining genetic correlation between SA|MDD and MDD (rg=0.53, SE=0.06, $P=8.9x10^{-19}$) indicates pleiotropic SNP effects.

Examining the genetic correlations between SA within psychiatric diagnosis and psychiatric disorders, most genetic correlations were comparable to those observed with SA|MDD (Supplementary Table 11). Genetic correlations of SA within psychiatric diagnosis and MDD (rg=0.52, SE=0.11, $P=4.48x10^{-6}$), ADHD (rg=0.60, SE=0.12, $P=7.08x10^{-7}$), and PTSD (rg=0.56, SE=0.19, $P=3.41x10^{-3}$) were significant after Bonferroni correction. As exceptions, BIP and SCZ had non-significant genetic correlations with SA within psychiatric diagnosis (SCZ: rg=-0.07, SE=0.075, $P=3.24x10^{-1}$, BIP: rg=-0.08, SE=0.10, $P=4.38x10^{-1}$). This is consistent with a previous report that BIP and SCZ cases who had attempted suicide did not have higher BIP or SCZ PRS, compared with cases who did not attempt suicide[115].

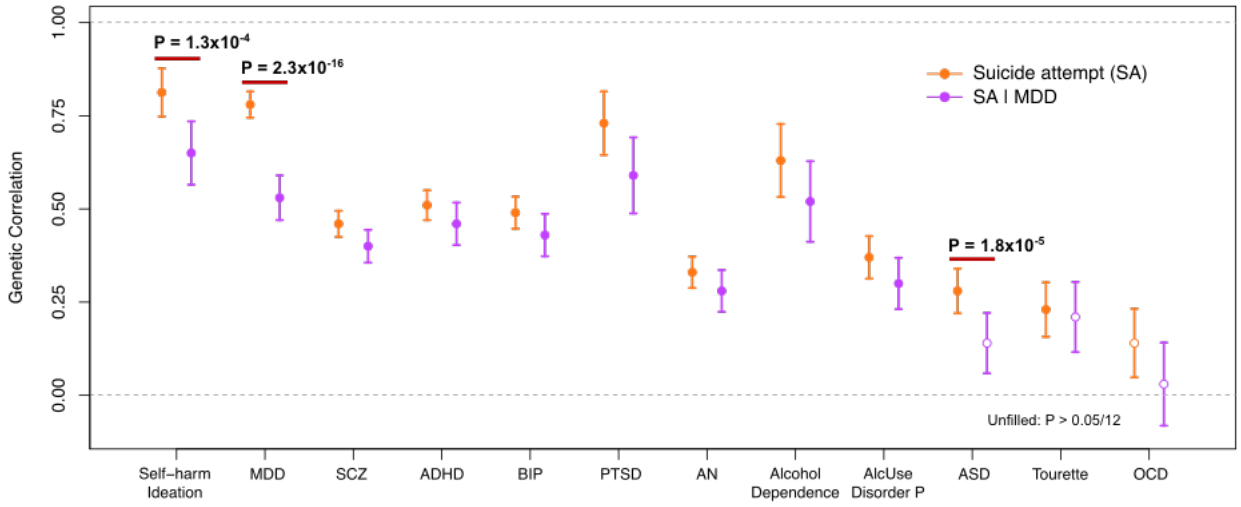

Figure 2: Substantial genetic correlation of suicide attempt with psychiatric traits or disorders before and after conditioning on major depressive disorder
Unfilled points indicate genetic correlations that did not pass the Bonferroni-corrected significance threshold $P<4.17x10^{-3}$ (12 traits tested). Error bars represent the standard error. P values indicate significant differences in genetic correlation after conditioning, that pass the Bonferroni correction. SA|MDD-suicide attempt conditioned on major depressive disorder, MDD-major depressive disorder, SCZ-schizophrenia, ADHD-attention-deficit/hyperactivity disorder, BIP-bipolar disorder, PTSD-post-traumatic stress disorder, AN-anorexia nervosa, AlcUse Disorder P-Alcohol Use Disorders Identification Test-P (AUDIT-P, measure of problematic consequences of drinking), ASD-autism spectrum disorder, OCD-obsessive compulsive disorder.

Substantial shared genetic architecture of SA and non-psychiatric risk factors not mediated by MDD

To assess the shared genetic architecture of SA, psychiatric, and non-psychiatric phenotypes, we calculated genetic correlations of our three GWAS (SA, SA|MDD and SA within psychiatric diagnosis) with 768 non-overlapping phenotypes[129]. We grouped 269 of these phenotypes into 15 categories of previously identified risk factors for SA[71,78,79] (see Methods). There were 194 phenotypes which showed a significant rg with SA, 133 of which were in one of the pre-defined SA risk categories (**Figure 3a**, Supplementary Table 12). The most significant genetic correlations were predominantly with traits related to depressive symptoms, smoking, and socioeconomic status. Examining phenotypes in the risk categories after conditioning on MDD, 81 phenotypes retained a significant genetic correlation with SA (Supplementary Table 12). Within the psychiatric risk category, there was an average decrease in the magnitude of genetic correlation of 33% with SA after conditioning, whereas the genetic correlation values in other risk categories were much less affected by conditioning (smoking: 3% decrease, education/cognition: 0.74% increase, alcohol: 12.5% decrease, and socioeconomic: 9.7% decrease) (**Figure 3a**). Genetic correlations of SA within psychiatric

diagnosis were similar to those of SA|MDD: of the 39 phenotypes with significant genetic correlation after Bonferroni correction, 21 phenotypes were in the smoking, education/cognition or socioeconomic risk categories (**Figure 3b**, Supplementary Table 12).

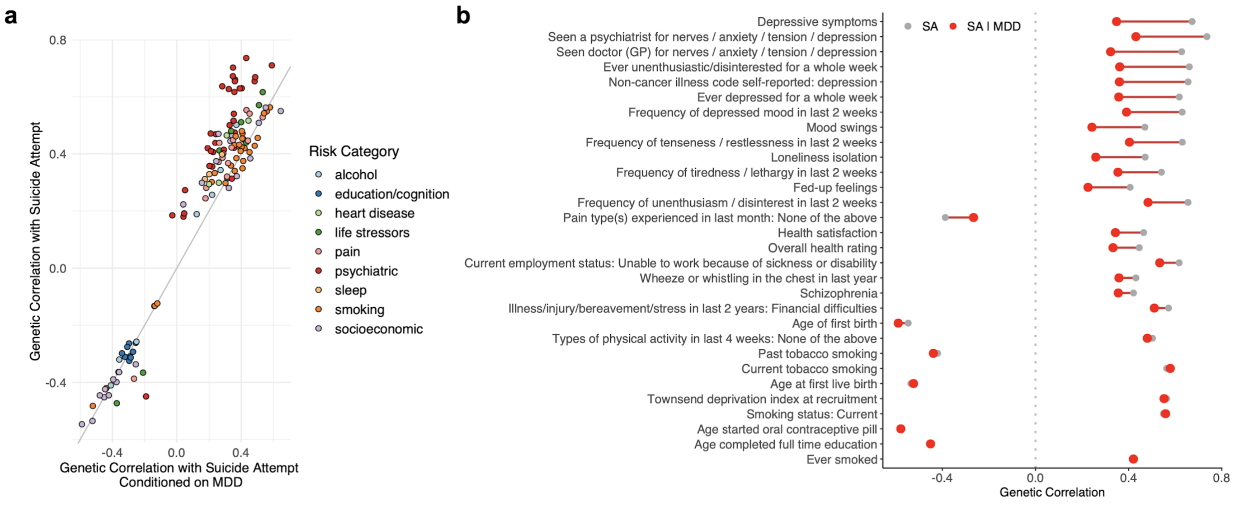

Figure 3: Conditioning suicide attempt on major depressive disorder reduces genetic correlation with psychiatric phenotypes but has limited effect on other traits
a) Comparison of significant genetic correlations with suicide attempt (SA) versus genetic correlations with SA conditioned on MDD (SA|MDD). Data include 133 significant genetic correlations after Bonferroni correction ($P<0.05/768=6.51x10^{-5}$) annotated by risk category. b) Top 30 phenotypes with the most significant genetic correlations with SA before (in gray) and after conditioning on MDD (SA|MDD) (in red). Full genetic correlation results, including standard errors, are provided in Supplementary Table 12.

Discussion

We present a GWAS of suicide attempt in over 29,000 cases, identifying 2 genome-wide significant loci, including one locus more strongly associated with SA than with psychiatric disorders or other related traits. We demonstrate that a substantial proportion of the SNP-heritability of SA is independent of psychiatric diagnosis, by conditioning our GWAS results on the genetics of MDD and by examining the genetics of SA among individuals with a psychiatric diagnosis. Finally, we determine that the genetic liability to SA not mediated by psychiatric disorders is shared with the genetic architecture of traits related to smoking, socioeconomic traits, and poorer overall health.

The locus most strongly associated with SA was in an intergenic region on chromosome 7. The index SNP had a larger effect on SA than any common psychiatric disorder, remained genome-wide significant after conditioning on MDD and had a comparable effect size on SA within psychiatric diagnosis and self-harm ideation. Taken together, these results suggest that the genetic association with SA at this locus is not mediated through risk for psychiatric disorders. Current functional genomic data does not clearly link this variant to any gene, with the nearest gene being a long-non-coding RNA (*LINC01392*) located 149 kb away. The index SNP (rs62474683) is a methylation quantitative trait locus (mQTL), with the SA risk allele associated with decreased methylation of a nearby DNA methylation site (probe cg04544267) in blood[143]. However, this methylation site has not been linked to any gene transcript. Intriguingly, SA-risk alleles at this locus have previously been implicated at genome-wide significance in risk-taking behaviors[139], smoking[140], and insomnia[144]. While variants in the MHC also reached genome-wide significance for SA, this effect did not remain after conditioning the GWAS results on MDD. Indeed, variants in the MHC have previously been associated with risk for a range of psychiatric disorders including MDD[145]. This suggests that the association between the MHC and SA may be pleiotropic or potentially a byproduct of psychiatric diagnosis. Further investigation is needed to determine causality or direction for both of these loci.

Our GWAS results provide robust evidence of the $h^2_{SNP}$of SA, with an estimate of 6.8% on the liability scale (7.5% in the European-only subset). Importantly, conditioning on MDD resulted in a smaller but significant h^2_{SNP}estimate (4.1%), which was on par with estimates from the GWAS of SA within psychiatric diagnosis (h^2_{SNP}3.7-4.6% on the liability scale, using a prevalence of SA in psychiatric populations from 10-20%). These results corroborate previous reports[91,111] of the independent genetic contribution to SA from genetic epidemiology studies and GWAS, and illustrate the importance of accounting for potential bias from the genetics of psychiatric disorders. Traditionally, GWAS of SA have sought to dissect this specific genetic component by conducting GWAS of SA within psychiatric diagnosis. More recently, a GWAS of SA in the iPSYCH Danish Registry took the approach of including a covariate for cases' psychiatric diagnoses[91]. Here, we found complete genetic correlation between the GWAS of SA after conditioning on MDD and the GWAS of SA within psychiatric diagnosis (rg=1.13, SE=0.13), thus demonstrating that comparable results can be achieved via a statistical genetics approach. Since conditioning only requires GWAS summary statistics, this approach is readily applicable to different types of cohort and circumvents the need for samples with specific psychiatric diagnoses, detailed phenotypic information or individual-level genotype data available.

SA showed substantial positive genetic correlation with many psychiatric disorders, the highest being with MDD (rg=0.78, SE=0.03), consistent with previous reports[92,113,115]. Genetic overlap was also particularly strong with PTSD, ADHD, SCZ, and BIP (rg=0.44-0.74). After conditioning on MDD, there was a modest decrease in the genetic correlation of SA with most psychiatric disorders, but only significant decreases were observed with MDD, ASD, and self-harm ideation. Notably, after conditioning, SA was still strongly genetically correlated with MDD (rg=0.53, SE=0.06, P=8.85×10^{-19}), representing pleiotropic effects between them. This genetic correlation would only be completely eliminated if all SNP effects on SA were mediated by MDD. Many studies have demonstrated extensive pleiotropy between psychiatric disorders[146,147], and accordingly genetic overlap between SA and related disorders is anticipated. Our findings suggest that many pleiotropic genetic variants increase risk for SA directly, independent of their

effects on psychiatric disorders. Examining the genetic liability to SA in a group of cases without psychiatric disorders would be a valuable future endeavor to corroborate these findings, however such individuals are a minority.

Genetic correlations were also examined between SA and 768 traits, with a focus on known risk factors and comorbidities. There was significant genetic correlation between SA and many other traits, including smoking, lower socioeconomic status, pain, lower educational attainment, reproductive traits, risk-taking behavior, sleep disturbances and poorer overall general health. While conditioning on MDD reduced the genetic correlations between SA and psychiatric disorders, in contrast, the genetic correlation of SA with most non-psychiatric traits remained unchanged. These results were largely corroborated using the GWAS of SA within psychiatric diagnosis, pointing to a consistent picture of shared genetic architecture between SA and these risk factors that is not a byproduct of psychiatric illness. There is substantial epidemiological literature on the relationship of risk factors including sleep disorders[87–89,102], smoking[104–106] and socioeconomic factors[148–150] on SA, but less on the role of genetics. We have not assessed any causal role between the genetic risk of these traits and SA, but additional work on this topic will provide important insights and potentially highlight opportunities for risk stratification.

This first collaborative study by the International Suicide Genetics Consortium is almost 5-fold larger than any previous GWAS of SA, providing a substantial increase in statistical power. Furthermore, we have assessed the specificity of our findings to SA using two approaches. Nevertheless, several limitations must be acknowledged. Cases were defined across cohorts using a variety of diagnostic interviews, self-report, or hospital records, which may result in heterogeneity in the phenotype definition. Standard diagnostic criteria for SA are lacking and here sample sizes prohibited calculating genetic correlations across pairs of cohorts. Our GWAS included both cases of non-fatal SA and death by suicide which are imperfectly although highly genetically correlated (rg=0.77 between the University of Utah GWAS of suicide death and a meta-analysis of the remaining cohorts in our study). There is potential for misclassification of controls in the GWAS of SA within psychiatric diagnosis, as some patients may go on to make a suicide attempt later in life. We examined the genetic correlation between our GWAS of SA and

psychiatric disorders, using publicly available GWAS summary statistics, however we note that the prevalence of SA amongst the cases in these GWAS are unknown. Finally, population, demographic and environmental factors are always present in genetic analyses and while our sample is large and diverse we did not have expansive data to stratify our analyses, to assess their possible contribution or confounding effects.

This work establishes the best-powered genetic analysis of SA to date. We identify SA risk loci and demonstrate a genetic component of SA that is not mediated through psychiatric disorders, but is shared with known risk factors. At present, PRS for SA do not have meaningful predictive utility and their premature use in either clinical or direct-to-consumer settings could be harmful. Dissecting the shared genetic architecture of SA, psychiatric disorders and other risk factors will be crucial to understanding the biological mechanisms of risk and assessing whether genetics can inform risk stratification or treatment.

CHAPTER III

Quantifying Genetic and Clinical Risk of Treatment-resistant Depression[§]

Introduction

Depression is a common, disabling mental illness, with a lifetime prevalence of up to 16.9% worldwide[116]. Treatment-resistant depression (TRD) is commonly defined by the absence of symptomatic remission following at least two adequate antidepressant treatment trials. However, limited consensus exists on the exact measures of remission, length of adequate treatment trial duration, and adequate treatment dose needed to define TRD[28]. Around a third of all individuals with depression are estimated to have TRD[29,30], although estimates vary widely[151]. TRD accounts for nearly half of incremental health costs associated with depression[30]. Individuals with TRD are also at a higher risk of other negative outcomes including suicide, with 30% attempting suicide at least once in their lifetime[31], 15 times the lifetime rate of the general population (~2%)[32].

Prior work has suggested a significant genetic component of TRD, with heritability estimates up to 17%[38] when compared to controls and ~8% when compared to non-TRD MDD[39,40]. Despite these estimates, no replicated genetic loci have been identified. Limitations of previous genome-wide association studies (GWAS) include various methods of ascertaining individuals with TRD, which include definitions based on antidepressant prescriptions[41,42], self-reported antidepressant efficacy and side effects[38,39], and electroconvulsive therapy (ECT) treatment[43], and remission of depressive symptoms[44]. One barrier to this work has been a paucity of adequately powered cohorts due to difficulty and data required to define TRD. One approach to increasing power for genetic studies is to leverage large-scale clinical data to build risk prediction models where quantitative phenotypes can be generated for genetic samples in associated

[§] Adapted from Kang J et al., manuscript in preparation

biobanks. Previous work has demonstrated substantial power increases from this approach in phenotypes like suicide attempt[152]. Studies using large national biobanks such as UK Biobank can also improve power, and I summarize three recent notable genetic studies of TRD below.

Clements et al.[43] compared ECT recipients to non-MDD controls with no history of other psychiatric disorders in a Swedish cohort from the Predictors for ECT (PREFECT) study. Among ECT recipients, cases included patients who received ECT in the context of MDD (narrow case set) as well as other mood disorders including bipolar disorder and schizoaffective disorder (broad case set). SNP-heritability estimates of either broad or narrow ECT (broad 35%, SE=0.05; narrow: 31%, SE=0.06; lifetime prevalence = 0.01) were higher than that of PGC-MDD[153,154] (6-8%). PRS of psychiatric diseases were compared between ECT patients and patients with moderate MDD who received psychotherapy (iCBT). Narrow ECT patients had higher MDD PRS (p=0.02) compared to moderate MDD patients, and both narrow and broad ECT patients had higher PRS for bipolar and cognitive traits (educational attainment and IQ).

Li et al.[39] studied TRD among 23andMe participants using self-reported survey data on the use of antidepressants in the last five years and qualitative effect from treatment of the current depressive episode overall. TRD was defined as having at least two antidepressants over 5-6 weeks, and whether the patient responded that the effect of treatment was not "helpful or very helpful". Non-TRD was defined as having at least two antidepressants over 3-4 weeks and the effect was rated as "helpful or very helpful". One genome-wide significant locus was identified in their TRD vs non-TRD GWAS in chromosome 10 (lead SNP rs150245813, OR=0.80, $p=8.07\times10^{-9}$, N=29,488). Another genome-wide significant locus was identified with a serotonin and norepinephrine reuptake inhibitor (SNRI) responder vs non-responder analysis in chromosome 3 (lead SNP rs4955665, OR 1.25, $p=1.62\text{x}10^{-9}$, N=8,119).

Fabbri et al.[40] studied TRD using medication data in the electronic health records of UK Biobank participants who had at least two diagnostic codes for unipolar depressive disorder, excluding patients with bipolar disorder, psychotic disorder, and substance use disorder. TRD was defined as having at least two switches between antidepressant drug with each antidepressant prescribed for at least six weeks. Notable

demographic differences between TRD and non-TRD MDD was that TRD patients were younger at first diagnosis of depression and prescription of antidepressants, they had higher BMI and risk of obesity, and had an increased risk in comorbidity with all psychiatric disorders compared to non-TRD MDD. In the GWAS comparing TRD cases with non-TRD MDD controls, no genome-wide significant locus was identified but there was significant SNP-heritability of 7.7% (SE=0.027, $p=2x10^{-3}$). There was a strong positive association with ADHD PRS and negative association with intelligence PRS with TRD.

While these studies suggest that there is a genetic architecture of TRD even when comparing to non-TRD MDD controls, symptom or medication efficacy surveys and longitudinal medication prescription records are not phenotypes that are readily replicable in diverse clinical settings. For this reason, we used ECT for our TRD definition which has remained the gold standard intervention for TRD for decades[33], despite recent US FDA approval of pharmacologic interventions[37]. Multiple comparative studies suggest that ECT is the most effective treatment for TRD[34]. In this study, we used ECT as a surrogate for TRD, and applied prediction models to electronic health record (EHR) data to derive posterior probabilities of receiving ECT, as absolute numbers of ECT cases in individual health systems were modest. We used these probabilities as quantitative traits to perform genome wide association studies on over 152,000 genotyped patients with MDD across four large biobanks to provide insight into the genetic architecture of TRD as defined by ECT.

Methods

Study settings

Clinical and genetic data were used from the biobanks of Mass General Brigham (MGB), Vanderbilt University Medical Center (VUMC), Geisinger Health System (Geisinger), and Million Veteran Program (MVP). MGB consists of 2 academic medical centers and 4 community and psychiatric hospitals in Eastern Massachusetts that serve over 6.5 million patients, and electronic health data were extracted from the Mass General Brigham Research Patient Data Registry[155] and the Enterprise Data Warehouse. VUMC is an academic medical center in Nashville Tennessee that manage over 2 million patient visits every year across Tennessee and its neighboring states. Its deidentified clinical EHR data is stored in the BioVU Synthetic Derivative[156]. Geisinger Health System is an academic medical center in Danville Pennsylvania and serve over 3 million patients in Pennsylvania. Million Veteran Program[157] study has over 825,000 veteran participants.

Clinical prediction model of TRD (MGB and VUMC)

Clinical data were collected from the de-identified repository VUMC Synthetic Derivative (SD) and MGB Healthcare System (**Figure 4A**). Only individuals with age 18-90 at time of data extraction were included for analyses. Cases with depression were identified using International Classification of Diseases, version 9 (ICD-9) codes (311.*, 296.2*, 296.3*, 300.4, * as wildcard digits ranging 0-9) and ICD-10 codes (F32.**, F33.**, F34.1, * as wildcard digits 0-9) for all adults. Individuals with one or more ICD-9 or ICD-10 codes for bipolar disorders, schizophrenia, and psychotic disorders were excluded from analyses. TRD cases were identified using the CPT code for ECT (90870), and all data 24 hours before the date of ECT were censored to avoid surrogates for the outcome (right-censoring). MDD controls were similarly censored using the last MDD code as the censoring point. A minimum of at least two visits or fact dates over four weeks before censoring date was required for study in inclusion for both ECT cases and MDD controls. With the inclusion and exclusion criteria, 106,565 MDD cases and 225 ECT cases were identified in VUMC, and 78,378 MDD and 242 ECT cases were identified in MGB.

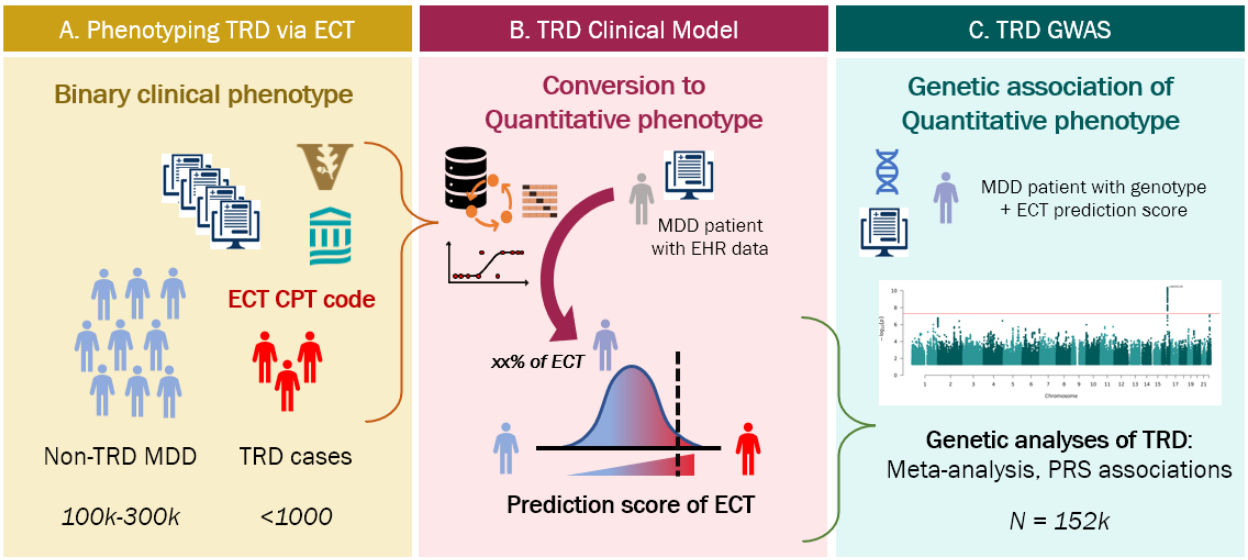

Figure 4: Schematic of the TRD clinical model generation and the genome-wide association study of the quantitative ECT prediction scores.
A. TRD cases were identified using the CPT code for ECT (90870) and non-TRD MDD controls were identified with ICD-9/10 codes for depression among adults (age 18-90) in the de-identified repository VUMC Synthetic Derivative and MGB Healthcare System. Individuals with ICD codes for bipolar disorders, schizophrenia, and psychotic disorders were excluded from analyses. **B.** Structured clinical data from the EHR such as demographics, diagnostic codes, and medications were included as predictors for the LASSO model trained and tested at VUMC and MGB. In addition to internal validation, features and weighs identified by the each of the two LASSO clinical models were applied to other partner sites for external validation (i.e., VUMC model applied to MGB, Geisinger, and Million Veterans Program). The two clinical models each produced a quantitative ECT prediction score among MDD individuals in VUMC, MGB, MVP, and Geisinger. **C.** Genome-wide association studies were conducted on genotyped MDD individuals with the ECT prediction scores using the VUMC or the MGB clinical as the quantitative phenotype. Meta-analyses of the VUMC and MGB model GWAS across four clinical sites (N=152,113) was used for addition post-GWAS genetic analyses, including heritability estimation, genetic correlation, and polygenic risk score associations.

Structured clinical data were included as predictors for the clinical model, including: demographics (age in years, categorical sex [Male, Female, Unknown], categorical race [White, Black, Asian, Hispanic, Other]), area deprivation index (ADI), diagnostic codes (log-transformed counts of historical CCS counts)[158], and medication (log-transformed counts of RXNORM-mapped ingredients). Of note, the VUMC ADI uses six features from the American Community Survey on the census tract level[159], while MGB ADI includes 21 socioeconomic factors from the census on the zip-code level[160].

The VUMC dataset was split into training, validation, and test sets where the test sample was comprised of only patients in the biobank which could have available genetic information (VUMC genotyped: ECT case: 35, MDD control: 14,713). The remaining sample separate from the genotyped test set was then randomly

split into 80% for training (ECT case: 131, MDD control: 58,604) and 20% for testing (ECT case: 59, MDD control: 33,247). In MGB, the dataset was randomly split into 80% for training (ECT case: 904, MDD control: 174,085) and 20% for testing (ECT case: 207, MDD control: 43,520) regardless of genotyping status. A LASSO model[161] was trained separately at each site using Glmnet[162] and hyperparameters were trained via a 10-fold cross-validation on the training data set (**Figure 4B**).

Each clinical model was validated internally using an 80/20 train/test split and externally at the other partner site (**Figure 4B**). Both MGB and VUMC clinical models were further validated at Geisinger and MVP. Model performance was evaluated with discrimination metrics: Area Under the Receiver Operating Characteristic (AUROC); Area Under the Precision-Recall Curve (AUPR); sensitivity/recall; specificity; precision/positive predictive value; and with calibration metrics: calibration-in-the-large, calibration slope/intercept. Predicted probabilities of ECT of 33,306 individuals in VUMC and 7,443 individuals in MGB representing clinical risk of needing ECT among MDD patients were used as quantitative traits for genetic association analyses (**Figure 4C**).

Medication-based definition of TRD

To compare cases ascertained using different TRD definitions, medication-based TRD was defined using first occurrences of unique antidepressants. Individuals with MDD code with three or more unique antidepressants were included, and time interval between the third and first antidepressant had to be between 16 weeks and 2 years to account for adequate and consecutive trial for each antidepressant.

Phenome-wide association study (PheWAS)

Phecodes were mapped from ICD-9 and ICD-10 codes. Phecodes were binarized designating more than one phecodes as cases and used as the outcome variable as part of the generalized linear regression. Associations were tested only when there were more than 100 phecode cases with at least one ECT cases among phecode cases. PheWAS R package[163] was used to visualize results.

Genotyping and quality control of the MGB sample

Genotyping of MGB samples was performed using the three versions of the Illumina Multi-Ethnic Global (MEG) array (Illumina, Inc., San Diego, CA): MEGA (N = 4927; 1,411,334 SNPs), MEGAEX (N = 5353; 1,710,339 SNPs), and MEG (N = 4784; 1,747,639 SNPs). Quality control steps of each cohort was performed separately to avoid batch effects. Individuals with genotypic call rates exceeding 99% were included, and related individuals based on identity by descent (IBD) were removed. From these individuals, SNPs with < 95% call rate or Hardy-Weinberg equilibrium test P value < 10^{-6} were excluded. Samples were imputed using the Michigan Imputation Server implementing Minimac3[164] for imputation, SHAPEIT[165] for phasing, and using all population subsets from 1000 Genomes Project Phase v5 as reference panel.

In each batch, population structure was characterized via principal component analysis of genotype SNPs after linkage-disequilibrium-pruning. Northern European ancestry was determined by plotting principal components of MGB samples to those of HapMap samples. Further analysis was performed only among individuals of Northern European genomic ancestry to minimize confounding due to population stratification.

Genotyping and quality control of the VUMC BioVU sample

The 94,474 individuals' genetic data were genotyped by the BioVU Infinium expanded multi-ethnic genotyping array (MEGAEX), which contains 2,038,233 SNPs. SNP quality control steps include excluding SNPs with MAF <0.01 or Hardy-Weinberg equilibrium test P value $\leq 10^{-6}$ within each self-reported ancestry, or MAF<0.005 within whole samples, or call rate <98%. Individuals were removed if they had a mismatch between genetically inferred sex and self-reported sex, or excess heterozygosity rate within each self-reported ancestry, or missing rate ≥ 0.02, or potentially cross-contaminated samples (proportion IBD >0.8). 90,313 samples and 887,250 high-quality autosomal SNPs remained.

Ancestry was determined with 1000 Genomes phase 3 (1000GP3) data. 1000 Genomes phase 3 (1000GP3) consists of 2,504 unrelated samples from 5 super populations African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), South Asian (SAS). 887,250 genotyped autosomal SNPs from the BioVU MEGAEX array was merged with 1000GP3 after removing C/G and A/T SNPs to avoid unresolvable strand mismatches in MEGA samples. Regions with known high LD[166] (Chr 5 44–51.5 Mb,

Chr 6 25–33.5 Mb, Chr 8 8-12Mb, Chr 11 45–57 Mb) were excluded and the common variants were then pruned (r2 < 0.05) using PLINK 1.9[167] (–indep-pairwise 1000 50 0.05) to yield 71,339 SNPs in relative linkage equilibrium for ancestry analyses. Principal components (PCs) were generated using flashpca version 2.0. By using K nearest neighbors (KNN, k=5) clustering, we inferred MEGA samples' ancestries. We treated 1000GP3 samples' PCs as the training set and MEGA samples' PCs as the test set. For each individual within the MEGA sample, we calculated its Euclidean distance from each training sample from the 1000GP3 based on the 2 leading PCs and then identified the 5 closest individuals. If all the 5 closest 1000GP3 individuals are from the same super population, we inferred that the MEGA individuals belonged to that super population based on the full vote of its neighbors. If they are from more than one super population, we clustered the sample's ancestry as admixed one. Among the 90,313 MEGA individuals, 87,558 (96.5%) were assigned to a homogeneous super-population, with the following breakdown: AFR=13,752, AMR=2,446, EUR=70,107, EAS=441, SAS=390. A subset of individuals of EUR ancestry (MEGA-EUR) were selected for further analysis.

90,313 samples were imputed on the Michigan Imputation Server v.1.2.4 using Eagle (V2.4.1) for phasing, Minimac4 for imputation and the Haplotype Reference Consortium (HRC) reference v1.1 panel in build GRCh37 as reference. Genotype probabilities were converted to hard-call genotypes using PLINK2 (hard-call>= 0.1). SNPs were filtered with imputation info score in any of the batches < 0.3, missing genotype rate > 0.02, or multi-allelic states (>2). Within EUR super populations, SNPs with MAF<0.005 and Hardy-Weinberg equilibrium test P value < 1 × 10–6 were excluded, and individuals with missing rate ≥ 0.02, excess heterozygosity rate over 3* interquartile range (IQR) of the upper heterozygosity quartile (Q3) for each sample were removed.

Genome-wise association study and meta-analysis

TRD posterior probabilities were inverse rank normalized to a mean of 0 and a standard deviation of 1. Quantitative GWAS of TRD posterior probabilities was performed using covariates of sex, age and PC1-PC20 (22 covariates) using Regenie v1.0.7, a computationally efficient method of whole genome regression

modeling for genome-wide association analyses[168] (**Figure 4C**). Default settings of block size 200 and 20 threads were used. GWAS of the VUMC and MGB clinical model output in the individuals of European ancestry in 4 different clinical sites (VUMC, MGB, Geisinger, MVP) were meta-analyzed using inverse variance-weighted fixed effects model in METAL[169]. The weighted mean allele frequency was calculated weighted by the effective sample size per cohort. SNPs with a weighted minor allele frequency of <1% or SNPs present in <80% of total effective sample size were removed from the meta-analysis results. A genome-wide significant locus was defined as the region around a SNP with $P<5.0x10^{-8}$ with linkage disequilibrium (LD) $r^2>0.1$, within a 3,000 kilobase (kb) window, based on the LD structure of the Haplotype Reference Consortium European ancestries reference panel v1.0[170].

Heritability estimates and genetic correlation

LD score regression[171] was used to estimate the phenotypic variance in TRD explained by common SNPs (SNP-heritability, h^2_{SNP}) from GWAS summary statistics. h^2_{SNP} was calculated on the observed scale. LDSC bivariate genetic correlations attributable to genome-wide SNPs (rg) were estimated between GWAS of quantitative TRD and previously published GWAS of ECT[43] or medication-defined TRD[172], as well as other psychiatric and non-psychiatric risk factors of depression. For previously noted epidemiological risk factors of TRD, the rg of TRD GWAS with 29 other non-overlapping human diseases and traits was calculated using publicly available summary statistics (PMID listed in Supplementary Table 11). The Bonferroni corrected significance threshold was $P < 1.72x10^{-3}$, adjusting for 29 traits tested. Differences in rg with VUMC TRD versus MGB TRD and differences in heritability between TRD meta-analyses before and after mtCOJO conditioning for BMI were tested for deviation from 0, using the block jackknife method, implemented in LDSC software[173].

mtCOJO

The results of the GWAS of TRD were conditioned on the genetics of BMI using mtCOJO (multi-trait-based conditional & joint analysis using GWAS summary data)[122], implemented in GCTA software[123]. mtCOJO estimates the effect size of a SNP on an outcome trait (eg. TRD) conditioned on exposure trait(s)

(eg. BMI), using the genome-wide significant SNPs for the exposure trait as instruments to estimate the effect of the exposure on the outcome. It then performs a genome-wide conditioning of the estimated effect from the exposure, which provides conditioned effect sizes and P values for the outcome trait. We conditioned TRD on BMI, since higher BMI among TRD cases have been previously reported[174]. mtCOJO analysis was performed on TRD GWAS as the outcome trait. The GIANT European ancestry GWAS summary statistics[175] was used for the exposure trait since mtCOJO requires an ancestry-matched LD reference panel. mtCOJO is robust to overlap in samples contributing to the GWAS of the exposure and outcome. In the selection of SNPs as instruments, independence was defined as SNPs more than 1 megabase (Mb) apart or with an LD r^2 value < 0.05 based on the 1000 Genomes Project Phase 3 European reference panel[125].

Polygenic risk scoring

PRS of quantitative TRD was tested for association with ECT CPT code as well as posterior probabilities generated with the clinical prediction model in independent target cohorts. The target cohorts were BioVU, MGB, Geisinger, and MVP cohorts. The meta-analysis of quantitative TRD was repeated excluding each cohort in turn to create independent discovery and target datasets. PRS was tested for association with ECT among MDD patients in all four target datasets. PRS was additionally tested for association with VUMC and MGB generated clinical models in all four target datasets. In total we estimate three independent hypotheses tested using polygenic risk scoring and applied a Bonferroni corrected significance threshold of $P < 0.05/3=0.0167$. PRS analyses were performed using PRS-CS which places a continuous shrinkage prior to SNP effect sizes using a Bayesian regression framework[176]. The continuous shrinkage priors adapt the amount of shrinkage applied to each SNP to the strength of the associated GWAS signal based on the LD structure estimated from an external reference panel. PRS were generated in each cohort using PRS-CS and the 1000 Genomes European reference panel was used to estimate LD between SNPs. The PRS were summed for each individual of the target cohort using Plink 1.9. PRS was tested for association with ECT cases vs control status in the target cohort using logistic regression model, covarying with PC1-PC10, sex,

and age. PRS was also tested for association with TRD posterior probabilities of VUMC or MGB clinical model using the linear regression model, covarying with PC1-PC10.

Results

Patients receiving ECT show characteristic TRD phenotypic presentation across two healthcare systems

Leveraging longitudinal clinical data from EHRs at MGB and VUMC (see Methods), we identified 185,167 patients (MGB: 78,378, VUMC: 106,789) with a diagnostic code of MDD or depressive disorder. Depressive disorder was included as prior work in these health systems and others indicated that it is commonly applied by non-psychiatrists to capture MDD symptoms. Among those patients, 467 (MGB: 242, VUMC: 225) had at least one procedure code billed for ECT (CPT code: 90870). The prevalence of ECT among individuals with MDD was 0.26% (MGB: 0.31%, VUMC: 0.21%) which represents a very small fraction of the expected ~30% prevalence of medication-trial defined TRD[177] but is similar to the published prevalence of ECT of ~0.25% among individuals with mood disorders[178]. The mean age at which cases received their first ECT CPT code was 53.8 ± 17.4 years, with a median ECT trial number of 15 (SD = 16); at MGB mean age was 57 ± 17 years and mean ECT trial number of 16 (SD=19).

In descriptive analyses, we identified several demographic differences between MDD patients with ECT and those without across both healthcare systems (**Table 9**). ECT cases on average were 5 years older, 12% more likely to be male but still more common in women, and 8.8% more likely to be white (Table 1). Further, ECT cases had a 5% lower body mass index (BMI) and BMI as measured closest to earliest ECT was even lower in VUMC but no differences were observed in the MGB cohort (VUMC: 27.1 ± 6.86, MGB: 28.0 ± 6.7 kg/m^2) (**Table 9**). Several of these demographic differences were significant, after Bonferroni correction for the 13 tests ($p < 3.85 \times 10^{-3}$) including age, gender, Black race, and BMI.

Table 9: Demographic characteristics of both sites samples.
In parentheses are percentages, standard deviations are reported after ±. Age is defined as years between birth date and last EHR event. BMI uses BMI cleaned for extreme outliers and unit mismatch and further filtered to exclude individuals of age < 18 and BMI > 80. For ECT cases, the BMI measurement closest and withing six months to the earliest ECT CPT code. Deprivation index refers to the normalized score ranging 0-1 of six different measures of American Community Survey (includes measure of poverty, income, education, health insurance coverage, and housing) for each census tract, with higher index indicating more deprivation[159].

Demographic		VUMC ECT case (N=225)	VUMC MDD control (N=106,564)	VUMC Significance testing*	MGB ECT case (N=242)	MGB MDD control (N=78,156)	MGB Significance testing*
Age	Mean	56.7 ± 17.6	51.6 ± 19.5	**2.22E-05**	56.7 ± 16.3	51.2 ± 17.1	**<0.001**
	Mean age at earliest ECT	53.8 ± 17.4			57 ± 17		
Gender	Female	121 (53.8%)	70979 (66.6%)	**6.01E-05**	145 (60%)	55,334 (71%)	**<0.001**
	Male	104 (46.2%)	35548 (33.4%)	**6.00E-05**	97 (40%)	22,819 (29%)	**<0.001**
	Unknown	0 (0%)	1 (0%)	1	0 (0%)	3 (<0.1%)	>0.9
Race	White	200 (88.9%)	88741 (83.3%)	**0.031**	220 (91%)	62,013 (79%)	**<0.001**
	Black	9 (4.0%)	11552 (10.8%)	**1.41E-03**	5 (2.1%)	4,141 (5.3%)	**0.025**
	Asian	3 (1.3%)	1064 (1.0%)	0.87	5 (2.1%)	1,664 (2.1%)	> 0.9
	**Other	13 (5.8%)	5171 (4.8%)	0.63	12 (5.0%)	10,338 (13.2%)	**2.17E-04**
Ethnicity	Hispanic	3 (1.3%)	2189 (2.1%)	0.60	4 (1.7%)	6,883 (8.8%)	**1.38E-04**
	Non-Hispanic	213 (94.7%)	99769 (93.7%)	0.63	238 (98%)	71,273 (91%)	**1.38E-04**
	Unknown	9 (4.0%)	4570 (4.3%)	0.96	0 (0%)	0 (0%)	NA
BMI **	Mean of mean BMI	26.9 ± 6.81	28.2 ± 7.81	**4.77E-04**	28.1 ± 6.3	28.5 ± 6.6	**0.4**
	Mean within 6 months of ECT	27.1 ± 6.86					
	Mean of last BMI measurement	30.9 ± 7.63	32.9 ± 9.08	**1.24E-04**	28.0 ± 6.7	28.5 ± 6.8	**0.3**

*Significance testing: t-test for quantitative values, 2 proportions Z-test for categorical variables

**Other includes all other races including unknowns and combination of races

***BMI measurements exclude individuals < 18 years old

When testing association of comorbid phenotypes with ECT, the most significantly associated phenotype in both VUMC and MGB was suicidal ideation (VUMC: BETA=3.62, SE=0.15, $p=2.67x10^{-128}$; MGB: BETA=2.57, SE=0.18, $p=2.58x10^{-46}$) (**Figure 5**, **Figure 6**). Other significantly associated phenotypes include psychiatric diseases like major depressive disorder (VUMC: BETA=3.31, SE=0.34, $p=2.72x10^{-22}$; MGB: BETA=2.73, SE=0.24, $p=4.04x10^{-30}$) and generalized anxiety disorder (VUMC: BETA=1.39, SE=0.15, $p=1.10x10^{-20}$; MGB: BETA=0.82, SE=0.16, $p=2.49x10^{-7}$), and other suicide-related traits like poisoning by psychotropic agents (VUMC: BETA=2.26, SE=0.25, $p=4.52x10^{-20}$; MGB: BETA=2.09. SE=0.35, $p=1.45x10^{-9}$) and suicide or self-inflicted injury (VUMC: BETA=1.92, SE=0.25, $p=1.03x10^{-14}$; MGB: BETA=2.59. SE=0.35, $1.48x10^{-13}$) (**Figure 5**, **Figure 6**, Supplementary Table 17, Supplementary Table 18). PheWAS results from VUMC and MGB were significantly correlated (r=0.70. $p=4.19x10^{-54}$).

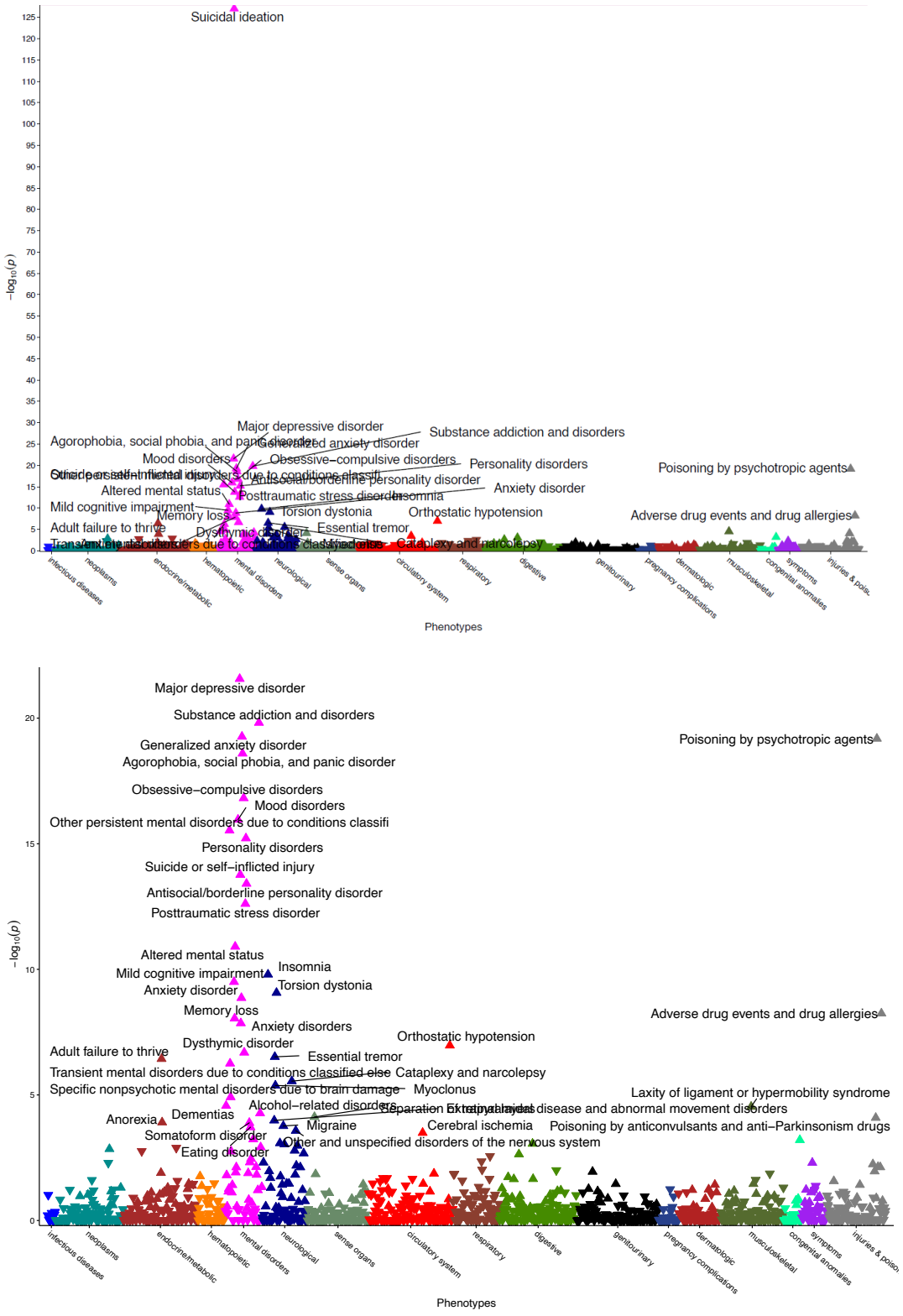

Figure 5: Phenome-wide association study of ECT CPT code among all MDD patients in VUMC.
For power, phecodes with counts over 100 were included for analysis. Covariates of the regression included sex, age, and race. In the second plot, the strongest association, suicidal ideation ($p=2.67x10^{-128}$) was omitted for scale

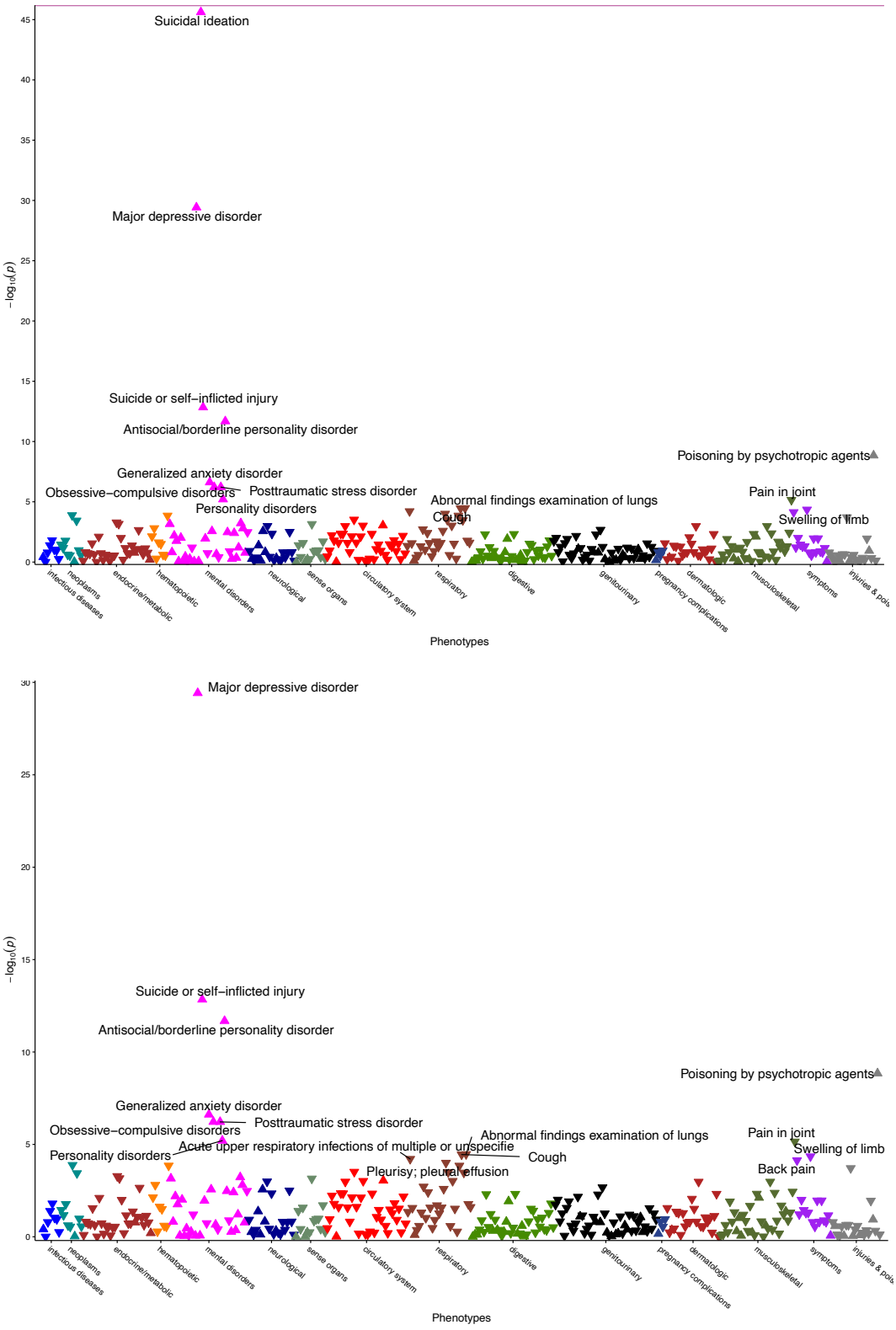

Figure 6: Phenome-wide association study of ECT CPT code among all MDD patients in MGB.
For power, phecodes with counts over 100 were included for analysis. Covariates of the regression included sex, age, and race. In the second plot, the strongest association, suicidal ideation ($p=2.39x10^{-46}$) was omitted for scale

Clinical prediction model for treatment-resistant depression is robust internally and externally across different sites

We next built prediction models of ECT to generate quantitative phenotypes representing clinical risk of needing ECT among MDD patients (**Figure 4A**). The MGB and VUMC datasets were each randomly split into training and test sets, and a LASSO model was trained separately at each site using identically mapped EHR features including diagnostic codes, medications, procedural codes and demographic information (see Methods). Features selected by LASSO with the highest weights included prescriptions of antipsychotics, diagnosis of mood disorders, and suicide in both models (**Figure 7**, **Figure 8**, Supplementary Table 16, Supplementary Table 17). Internal prediction performance as defined by the area under the receiver operator curve (AUROC) was high on both the test and validation sets at MGB (validation: AUROC=0.91; test: AUROC=0.81) and VUMC (validation: AUROC=0.93; test: AUROC=0.93) (**Table 10**). Area Under the Precision-Recall Curve (AUPRC) ranged from 2-3% in both MGB (validation: AUPRC=0.08; test: AUPRC=0.03) and VUMC (validation: AUPRC=0.08; test: AUPRC=0.08) largely owing to the challenge in predicting a rare event, with case frequency ranging from 0.21-0.31%. Applying each model to the samples from the other site (external validation) retained high prediction performance at MGB (AUROC=0.78, AUPRC=0.03) and VUMC (AUROC=0.83, AUPRC=0.03). To increase sample size and power for genetic analysis, both models were applied to biobank samples at two additional sites (**Table 10**), the Geisinger Health System (GHS, 353 cases, 190,841 controls) and the Million Veteran Program (MVP, 600 cases, 259,925 controls). Prediction performance remained consistently high for both models at GHS (VUMC model: AUROC: 0.84, AUPRC: 0.021; MGB model: AUROC: 0.78, AUPRC: 0.023) and MVP (VUMC model: AUROC: 0.81, AUPRC: 0.024; MGB model: AUROC: 0.81, AUPRC: 0.04).

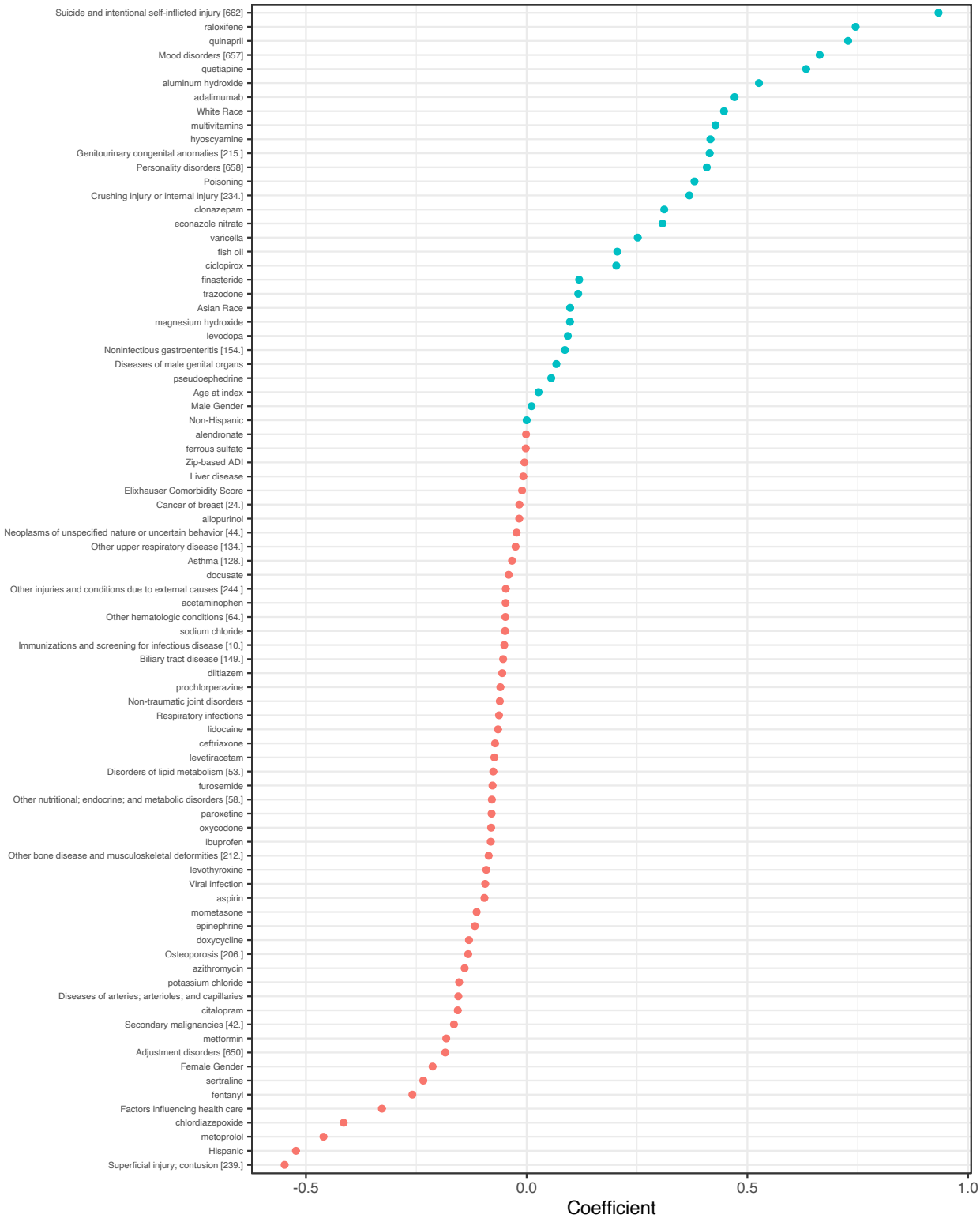

Figure 7: MGB TRD model LASSO features and weights

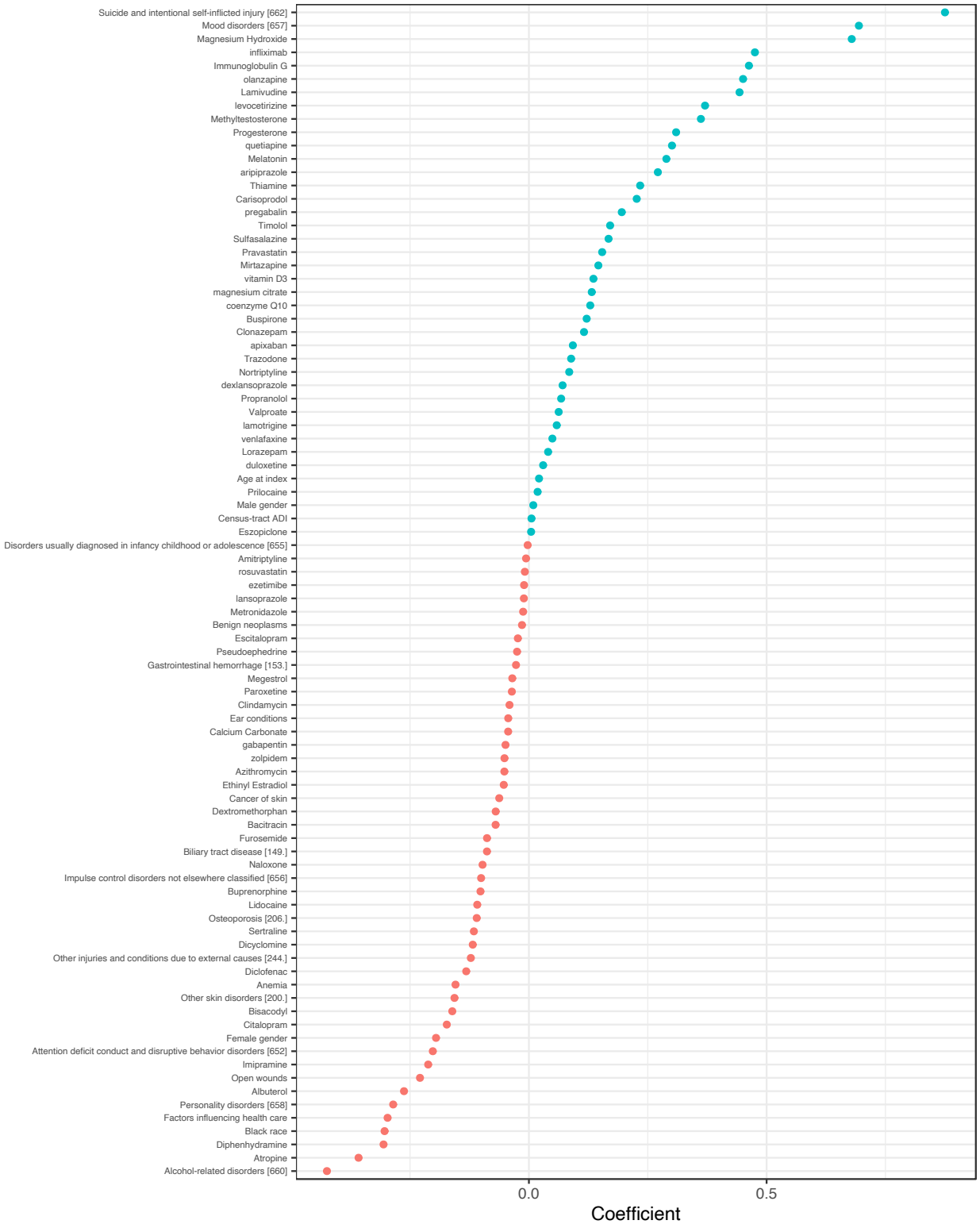

Figure 8: VUMC TRD model LASSO features and weights

Table 10: Prediction model sample size and ECT prevalence per partner site, and performance metrics of VUMC and MGB prediction models in each partner site.
ROC: area under the receiver operator curve; AP: average precision. Bolded numbers are performance measures of internal validation.

	Test Set			VUMC model		MGB model	
Site	**ECT cases**	**MDD controls**	**Prevalence**	**AUROC**	**AUPRC**	**AUROC**	**AUPRC**
VUMC	59	33,248	0.18%	**0.93**	**0.077**	0.83	0.030
MGB	57	15,676	0.36%	0.78	0.028	**0.81**	**0.028**
GHS	353	190,841	0.18%	0.84	0.021	0.78	0.023
MVP	600	259,925	0.23%	0.81	0.024	0.81	0.040

TRD models shows significant heritability and shared genetic architecture across models but not with other TRD phenotypes

The posterior probabilities from the ECT prediction model were rank normalized to generate a quantitative phenotype. Logistic regression of this phenotype on imputed dosage was performed separately on 11,240 samples of European ancestries at VUMC, 5,131 samples of European ancestries at MGB, 39,353 samples of European ancestries at GHS, and 96,389 samples of European ancestries at MVP. We then meta-analyzed the four datasets across 152,113 samples using a variance-weighted fixed effect model. Significant heritability of 0.04 (SE 0.004, $P=8.65 \times 10^{-18}$) for the MGB clinical model meta-analysis and 0.023 (SE 0.01, $P=4.5 \times 10^{-9}$) for the VUMC clinical model meta-analyses were estimated from LD-score regression[171] (**Table 11**). The meta-analyses of the two clinical models were significantly but not completely genetically correlated with each other (rg = 0.72, SE 0.05, $P=6.8 \times 10^{-44}$) (**Table 12**). The rg value reflects the highly overlapping but non-identical phenotypes generated by the two models.

We then examined the genetic correlation of our TRD phenotype with two prior GWAS of TRD (**Table 12**). The first defined TRD based on antidepressant prescriptions in the UK Biobank (UKB)[179] and the second used ECT to define TRD but compared them to healthy controls as opposed to only those with MDD (PREFECT)[43]. No significant genetic correlation was observed between the MGB model or the VUMC model with either PREFECT TRD (VUMC: rg = 0.20, SE 0.13, P=0.12; MGB: rg = 0.09, SE 0.10, P=0.38) or UKB TRD (VUMC: rg = 0.023, SE 0.19, P=0.91; MGB: rg = 0.020, SE 0.13, P=0.12). Notably, UKB

TRD and PREFECT TRD are significantly correlated with each other (rg=0.75, SE=0.24, P=0.003). Further, genome-wide significant loci from prior TRD GWAS were not genome-wide significant in either TRD model meta-analysis (**Table 13**).

Table 11: Heritability estimates of TRD GWAS meta-analyses using LD-score regression.
Heritability estimates are of the inverse-rank normalized predicted probability of TRD within each biobank site (first four rows) and the meta-analysis (fifth row). All heritability estimates are in observed scale.

		Normalized VUMC			Normalized MGB		
Cohort	**N**	**SNP h2**	**SE**	**p**	**SNP h2**	**SE**	**p**
GHS	39,353	0.045	0.013	**2.39E-04**	0.029	0.014	**0.020**
VUMC	11,240	-0.023	0.040	0.277	0.011	0.042	0.395
MGB	5,131	-0.080	0.080	0.157	0.066	0.077	0.194
MVP	96,389	0.013	0.0047	**2.84E-03**	0.035	0.0049	**6.15E-13**
Meta-analysis	152,113	0.023	0.004	**4.46E-09**	0.04	0.0047	**8.65E-18**

Table 12: Genetic correlations of TRD meta-analysis with other GWAS of TRD

Genetic correlations	MGB model TRD meta	UKB medication-TRD	PREFECT ECT-TRD
VUMC model TRD meta	**0.72** (SE 0.05 **P 6.8e-44**)	0.023 (SE 0.19 P 0.91)	0.20 (SE 0.13 P 0.12)
UKB medication-based TRD (Fabbri et al)	-0.18 (SE 0.14 P 0.2)		
PREFECT ECT-based TRD (Clements et al)	0.09 (SE 0.10 P 0.38)	**0.75** (SE 0.25 **P 2.5e-3**)	

Table 13: Genome-wide significant SNPs in prior TRD studies in our meta-analysis.
SNRI – serotonin and norepinephrine reuptake inhibitor

Trait	GWS snp	CHR	POS	A1	A2	BETA	SE	P	Freq	N	r2	PMID
Non-TRD vs TRD	rs150245813	10	38592780	T	G	-0.228		8.07E-09		29488		33106475
TRD MGB model meta-analysis	rs7090978	10	38505552	G	A	0.011	0.0054	4.55E-02	0.140	152186	0.961	
TRD VUMC model meta-analysis	rs7090978	10	38505552	G	A	0.005	0.0054	3.92E-01	0.140	152113	0.961	
SNRI Responder vs Non-Responder	rs4955665	3	169355019	T	C	0.219		1.62E-09		8119		33106475
TRD MGB model meta-analysis	rs34781085	3	169364599	T	C	0.005	0.0038	0.1748	0.347	152186	0.608	
TRD VUMC model meta-analysis	rs34781085	3	169364599	T	C	0.000	0.0039	0.9078	0.347	152113	0.608	
ECT in mood disorders* vs non-MDD controls	rs114583506	6	31263801	G	T	-0.511		3.60E-08		6015		33483693
TRD MGB model meta-analysis	rs114583506	6	31263801	G	T	0.006	0.0089	0.5254	0.043	152186	1	
TRD VUMC model meta-analysis	rs114583506	6	31263801	G	T	0.020	0.009	0.02626	0.043	152113	1	

* Mood disorders included major depressive disorder, bipolar disorder, and schizoaffective disorder

GWAS of quantitative TRD identifies intronic locus in weight-associated gene FTO

One genome-wide significant locus was identified in the MGB model located on chromosome 16 in the intronic region of *FTO* (index SNP = rs8050136, Beta for A allele=-0.0243, SE=0.0037, MAF=0.4, p=4.3x10-11, Cochran's Q: 0.55, I2 heterogeneity index=0) (**Figure 9, Table 14**). The same locus was not significantly associated with TRD in the VUMC TRD GWAS (BETA=-0.045, SE=0.0037, p=0.22) (**Figure 10**) or the prior published TRD GWAS based on medication data[40] or ECT cases against non-psychiatric controls[43]. The TRD index SNP was in high LD (R2=0.992) with the SNP rs9939609 shown to be strongly associated with BMI[180] (BETA=0.075, SE=2.9x10-3, P=1.95x10-145) and weight[181] via its regulation of IRX3 expression[182]. That is, lower BMI is associated with higher risk of TRD. We tested for inflation looking at the lambda GC measured using LDSC, and genomic inflation factor (λ_{GC}) estimate was 1.114 for the MGB meta-analysis and 1.079 for the VUMC meta-analysis. Intercepts were 1.0087 (0.0086) for MGB and 0.9998 (0.008) for VUMC, and these intercepts near 1 suggest there are no confounding factors leading to inflation of summary statistics.

To investigate whether the GWS loci in the *FTO* region is mediated by genetics of BMI, we conditioned the TRD meta-analyses with BMI (TRD | BMI). After conditioning for the genetic contribution of BMI to TRD meta-analyses, the GWS locus was no longer significantly associated for MGB TRD meta-analysis (BETA=-0.006, SE=0.003, P=0.13) and the effect size of VUMC TRD meta-analysis also decreased after conditioning for BMI (BETA=0.002, SE=0.004, P=0.22) (**Figure 9**). Conditioning TRD meta-analyses for BMI also significantly decreased SNP-heritability in both models, resulting in heritability estimates of ~2% for both models after conditioning (VUMC: 0.021, SE=0.0038, P=2.56×10^{-8}, heritability difference p=4.02×10^{-18} block jackknife; MGB: 0.024, SE=0.004, P=1.82×10^{-9}, p=2.38×10^{-10} block jackknife).

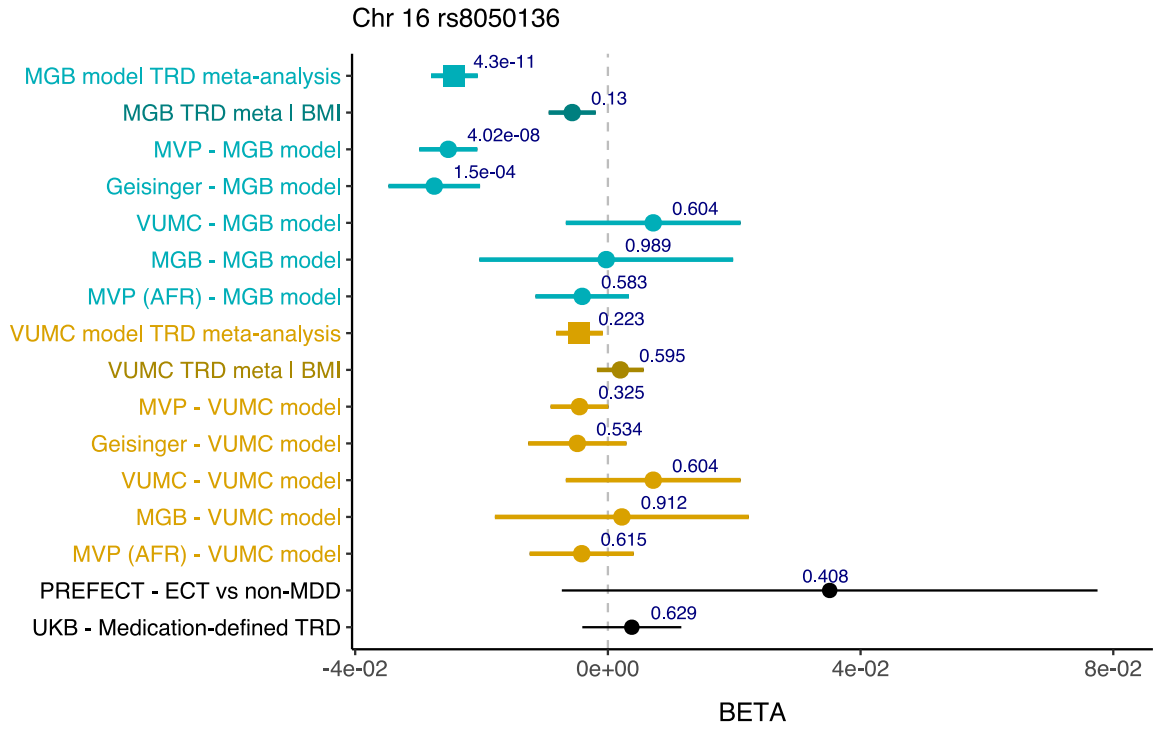

Figure 9: Forest plot of the GWS locus rs8050136 in chromosome 16
Genome-wide significant (GWS) locus of MGB TRD model is not replicated in the VUMC model or TRD GWAS among individuals of African ancestry. The points indicate the log odds ratio of the A allele on each phenotype and the error bars show the standard error. The P value of association with each phenotype is shown above the error bars.

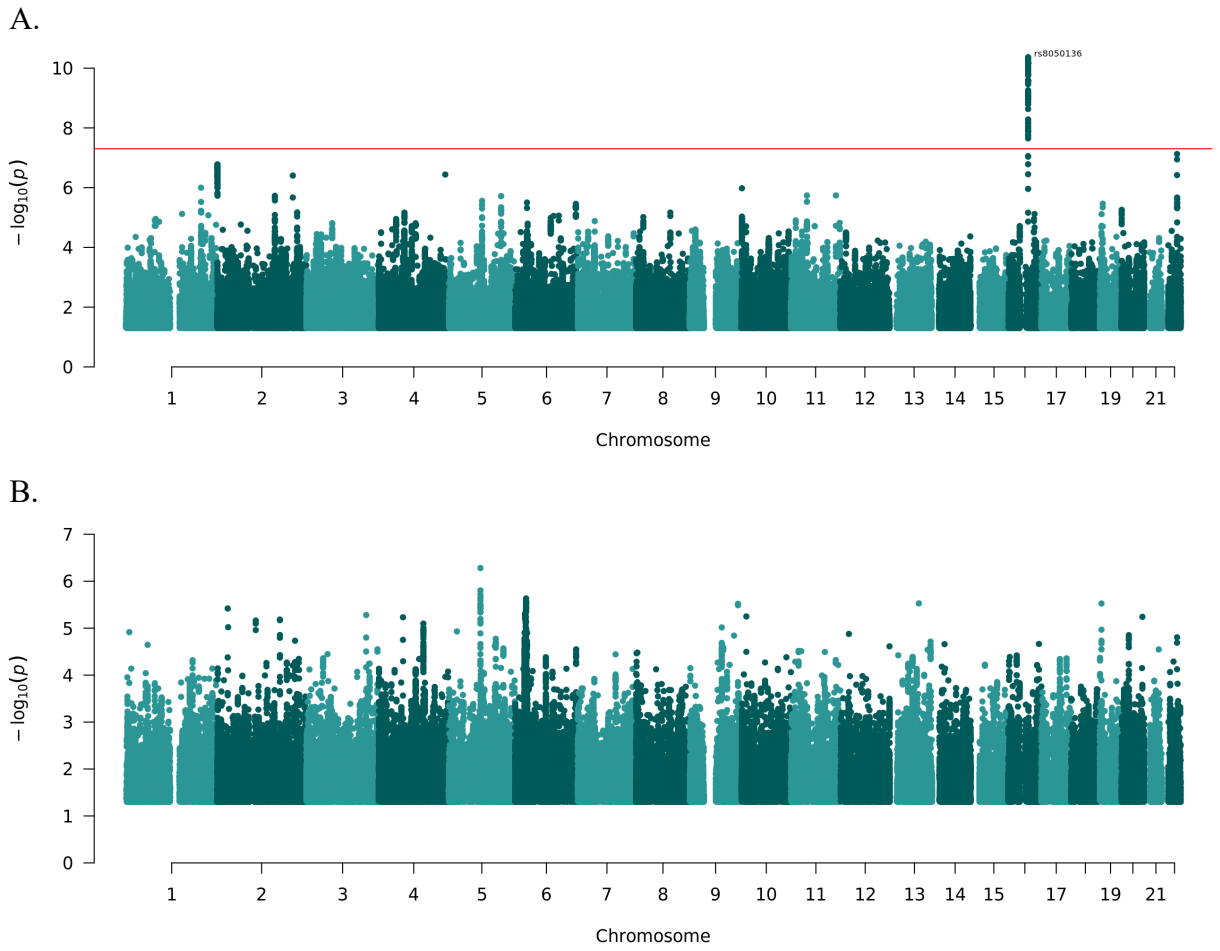

Figure 10: Manhattan plots of A. MGB model meta-analysis and B. VUMC model meta-analyses (N=152,113)

Table 14: Effect size of genome-wide significant loci rs8050136 across individual cohorts and other TRD GWAS

Ancestry	Model	study site	CHR	POS	SNP	A1	A2	FREQ	BETA	SE	P	NMISS
AFR	MGB	MVP	16	53816275	rs8050136	A	C	0.440	-0.0041	7.40E-03	0.583	30235
	VUMC	MVP	16	53816275	rs8050136	A	C	0.440	-0.0041	8.24E-03	0.615	30235
EUR	MGB	**MVP**	16	53816275	rs8050136	A	C	0.398	**-0.0253**	4.60E-03	**4.02E-08**	96389
		Geisinger	16	53816275	rs8050136	A	C	0.413	**-0.0275**	7.25E-03	**1.50E-04**	39426
		VUMC	16	53816275	rs8050136	A	C	0.593	-0.0072	1.38E-02	0.604	11240
		MGB	16	53816275	rs8050136	A	C	0.404	-0.0003	2.01E-02	0.989	5131
	VUMC	MVP	16	53816275	rs8050136	A	C	0.398	-0.0045	4.57E-03	0.325	96389
		Geisinger	16	53816275	rs8050136	A	C	0.413	-0.0048	7.78E-03	0.534	39353
		VUMC	16	53816275	rs8050136	A	C	0.593	-0.0072	1.38E-02	0.604	11240
		MGB	16	53816275	rs8050136	A	C	0.404	0.0022	2.01E-02	0.912	5131
		PREFECT ECT -TRD	16	53816275	rs8050136	A	C	0.405	0.0351	4.24E-02	0.408	5086
		UKB medication-TRD	16	53816275	rs8050136	A	C	0.393	0.0038	7.84E-03	0.629	16372

TRD polygenic risk score association with TRD prediction scores

Polygenic risk scores are a standard approach to collapsing aggregated risk from genome-wide association studies[183]. We tested for association of polygenic risk scores generated using our TRD meta-analysis and our model probabilities in the VUMC or MGB samples after excluding them from the meta-analysis (i.e., leave one out). Among VUMC patients, PRS generated from MGB TRD meta-analysis was significantly associated with both VUMC and MGB TRD model prediction scores (VUMC $p=9.74x10^{-5}$, MGB: $p=1.38x10^{-9}$) and VUMC TRD PRS was also significantly associated with MGB and VUMC TRD model prediction scores although the latter did not survive correction for the 30 total tests (MGB $p=2.78x10^{-5}$, VUMC p=0.0167) (**Table 15**). Among MGB patients, neither TRD PRS was significantly associated with TRD prediction scores, but this in part due to the much smaller sample size of MGB, which is less than 20% of the VUMC genotyped samples.

We next looked at whether PRS derived from relevant psychiatric traits including depression[16], schizophrenia and bipolar disorder associated with our TRD models across VUMC and MGB patients. We identified the depression PRS was significantly associated with VUMC model TRD prediction scores ($p=8.09x10^{-4}$) and nominally associated with MGB model TRD prediction scores ($p=6.6x10^{-2}$). Despite excluding patients with diagnoses of bipolar disorder or schizophrenia defined by at least one diagnostic code, we found that schizophrenia[98] PRS was significantly associated with both MGB and VUMC model ECT prediction scores (**Table 15**, MGB: linear regression $p=1.07x10^{-9}$, VUMC: $p=7.89x10^{-6}$), and bipolar disorder[184] PRS was significantly associated with MGB TRD prediction scores ($p=1.96x10^{-3}$) and nominally associated with VUMC TRD prediction scores ($p=3.30x10^{-2}$). Among MDD patients in MGB, schizophrenia and bipolar disorder PRS were not significantly associated with TRD prediction scores of either model.

Table 15: Polygenic risk score association results of PRS generated using psychiatric traits and TRD meta-analyses as discovery GWAS and the two clinical TRD model prediction scores and medication-defined TRD as target traits.

		VUMC patients (n = 11,240)			*MGB patients (n = 2,126)*		
Discovery GWAS	**Target trait**	**BETA**	**SE**	**P**	**BETA**	**SE**	**P**
ECT TRD (VUMC model)	ECT TRD probabilities (VUMC model)	0.0235	9.80E-03	0.0166	0.0303	2.21E-02	0.171
ECT TRD (MGB model)		0.0383	9.82E-03	**9.74E-05**	0.0394	2.21E-02	0.075
Depression		0.0345	1.03E-02	**8.09E-04**	-5.73E-04	2.19E-02	0.979
Bipolar Disorder		0.0222	1.04E-02	0.033	-0.0125	2.22E-02	0.574
Schizophrenia		0.0461	1.03E-02	**7.89E-06**	0.0202	2.52E-02	0.424
ECT TRD (VUMC model)	ECT TRD probabilities (MGB model)	0.0416	9.92E-03	**2.78E-05**	0.0191	2.19E-02	0.382
ECT TRD (MGB model)		0.0602	9.94E-03	**1.38E-09**	0.0420	2.19E-02	0.055
Depression		0.0190	1.03E-02	0.066	0.0019	2.19E-02	0.931
Bipolar Disorder		0.0324	1.05E-02	**1.96E-03**	0.0068	2.23E-02	0.761
Schizophrenia		0.0633	1.04E-02	**1.07E-09**	-4.16E-04	2.52E-02	0.987
ECT TRD (VUMC model)	Medication based TRD	-0.0246	4.76E-02	0.604	-0.0234	8.50E-02	0.783
ECT TRD (MGB model)		-0.0027	4.74E-02	0.954	-0.0467	8.47E-02	0.581
Depression		-0.0242	4.78E-02	0.613	0.0282	8.39E-02	0.737
Bipolar Disorder		0.0926	4.82E-02	0.055	0.1047	8.64E-02	0.226
Schizophrenia		0.0642	4.76E-02	0.177	-0.1512	9.70E-02	0.119

Medication-defined treatment-resistant depression has higher ECT clinical risk

To compare results from the ECT model with a commonly used alternative definition of TRD, we defined case status based on antidepressant medication trial numbers and length, where cases were defined as having at least three unique antidepressants prescribed, requiring the time interval between the third and first antidepressant had to be between 16 weeks and 2 years to account for adequate and consecutive trial for each antidepressant. MGB and VUMC prediction scores were compared among MDD individuals with or without medication-defined TRD (med-TRD). In the VUMC MDD cohort, individuals with medication-defined TRD (N=1181) had higher normalized prediction scores than nonmed-TRD patients (N=21,400) for both VUMC model (Med-TRD prediction score: 0.358±1.22, non med-TRD prediction score: 0.015±1.04, t-test $p=1.28x10^{-20}$) and MGB model (Med-TRD prediction score: 0.056±1.21, non med-TRD prediction score: -0.034±1.04, t-test p=0.013). In the MGB MDD cohort (N=7,443), there were no differences in the VUMC model prediction scores (Med-TRD prediction score: -0.027±0.95, non med-TRD prediction score: 0.002±1.00, t-test p=0.52) or MGB model prediction scores (Med-TRD prediction score: -0.030±0.96, non med-TRD prediction score: 0.002±1.00, t-test p=0.48) between med-TRD patients (N=501) compared to non-med TRD patients (N=6942) .

We then tested for association of TRD and psychiatric diagnoses PRS with this medication-defined TRD status using a logistic regression. In either MGB or VUMC MDD cohorts, neither TRD meta-analysis PRS was significantly associated with medication-defined TRD (MGB: p=0.954, VUMC: p=0.604), and there were no significant associations with PRS of the psychiatric traits.

Significant genetic overlap is observed with psychiatric traits, substance use traits, and BMI

To complement PRS analyses, to study the genetic overlap between TRD and psychiatric and non-psychiatric traits previously associated to TRD, genetic correlations were estimated. Both TRD models showed significant positive genetic correlations, after multiple test correction, with cognitive traits including years of education (VUMC: rg=0.21; MGB: rg=0.47) and intelligence (VUMC: rg=0.19; MGB: rg=0.29), and significant negative genetic correlations with ADHD (VUMC: rg=-0.30 MGB: rg=-0.40), alcohol dependence (VUMC: rg=-0.45; MGB: rg=-0.41) and smoking traits (VUMC: rg=-0.24; MGB: rg=-0.38) (**Figure 11,** Supplementary Table 19). Both TRD models also showed significant negative genetic correlations with weight-related traits of BMI (VUMC: rg=-0.27; MGB: rg=-0.63) and waist-hip-ratio (VUMC: rg=-0.16; MGB: rg=-0.21). While the models shared substantial genetic architecture, there were noticeable difference in genetic correlation across a subset of traits. Traits with significantly stronger genetic correlations with the MGB model, based on a block jack knife approach in LD score regression[173] included BMI ($P=1.15x10^{-14}$), type 2 diabetes ($P=3.94x10^{-9}$), educational attainment ($P=7.34x10^{-9}$), and marijuana use ($P=1.39x10^{-6}$). Traits that had a significantly stronger genetic correlation with the VUMC model were neuroticism ($P=2.42x10^{-4}$), and multiple measures of alcohol use disorders, AUDIT-C ($P=9.55x10^{-7}$), and AUDIT-T ($P=6.02x10^{-7}$) (**Figure 11,** Supplementary Table 19).

Conditioning TRD for BMI only changes genetic correlation with weight-related traits

We also examined for changes in genetic correlations after conditioning for the genetic contribution of BMI to TRD meta-analyses. Overall, there were no significant differences in genetic correlations with TRD after conditioning for BMI except for obesity-related traits. In the MGB meta-analyses, the significant differences in genetic correlations were observed with BMI (rg=-0.03, SE=0.05, $p=6.79x10^{-60}$ block jackknife), Type 2 diabetes (rg=-0.42, SE=0.08, $p=1.40x10^{-3}$ block jackknife), and anorexia nervosa

(rg=0.09, SE=0.07, p=1.41x10^{-3} block jackknife) (**Figure 12**, Supplementary Table 19). In the VUMC meta-analyses, significant differences in genetic correlations after conditioning for BMI was observed with BMI (rg=0.007, SE=0.04, p=7.49x10^{-28} block jackknife) (**Figure 13**, Supplementary Table 19).

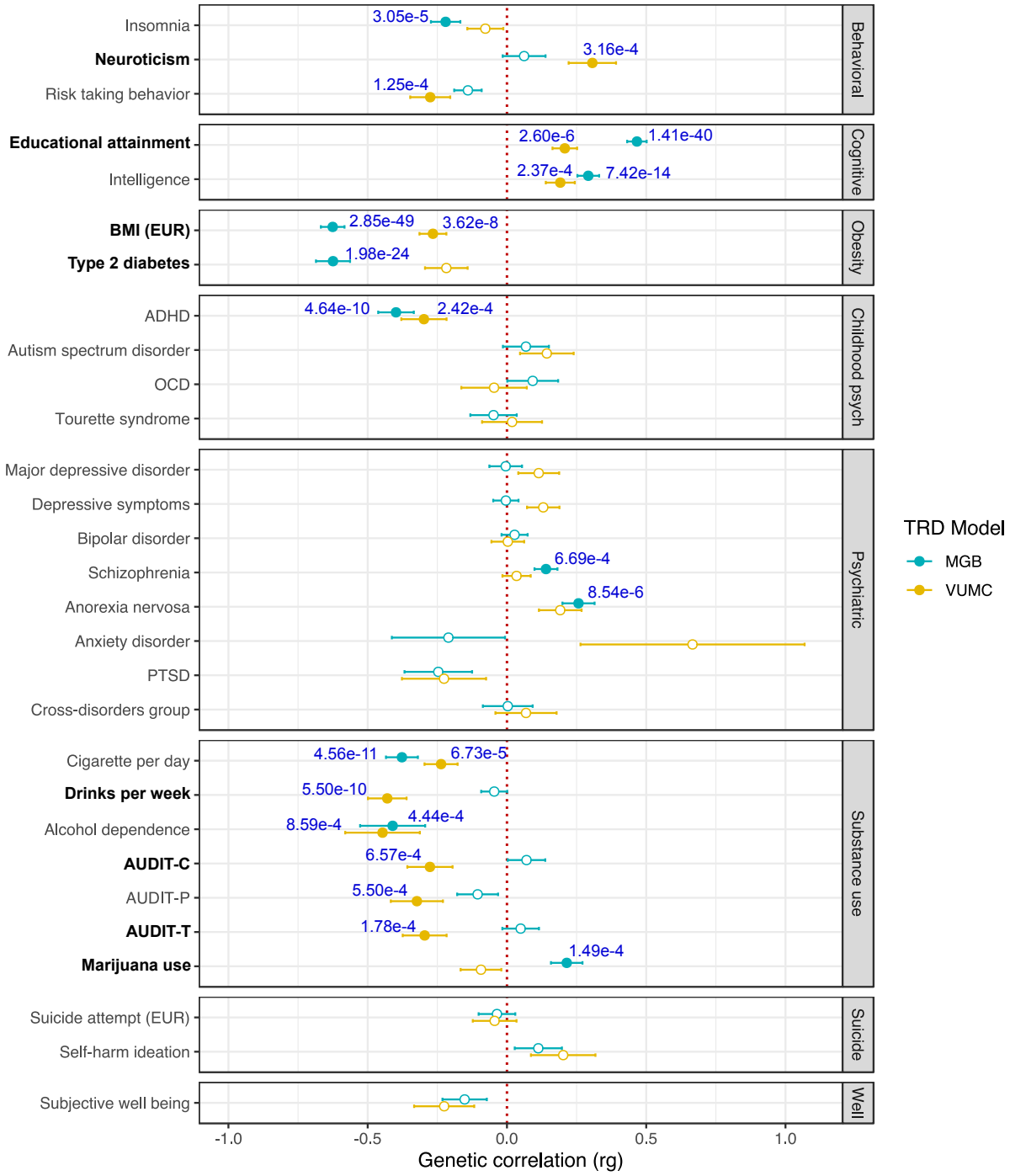

Figure 11: Genetic correlations of VUMC and MGB TRD models with psychiatric and non-psychiatric traits. Unfilled points indicate genetic correlations that did not pass the Bonferroni-corrected significance threshold $P<1.72x10^{-3}$ (29 traits tested). Error bars represent the standard error. P values indicate significant differences in genetic correlation after conditioning, that pass the Bonferroni correction. Bolded traits show significant differences in genetic correlations between the two models. BMI-body mass index, ADHD-attention-deficit/hyperactivity disorder, OCD-obsessive compulsive disorder, PTSD-post-traumatic stress disorder, AUDIT-C-Alcohol Use Disorders Identification Test-C (measure of quantity of alcohol consumption), AUDIT-P- measure of problematic consequences of drinking, AUDIT-T-total score of AUDIT.

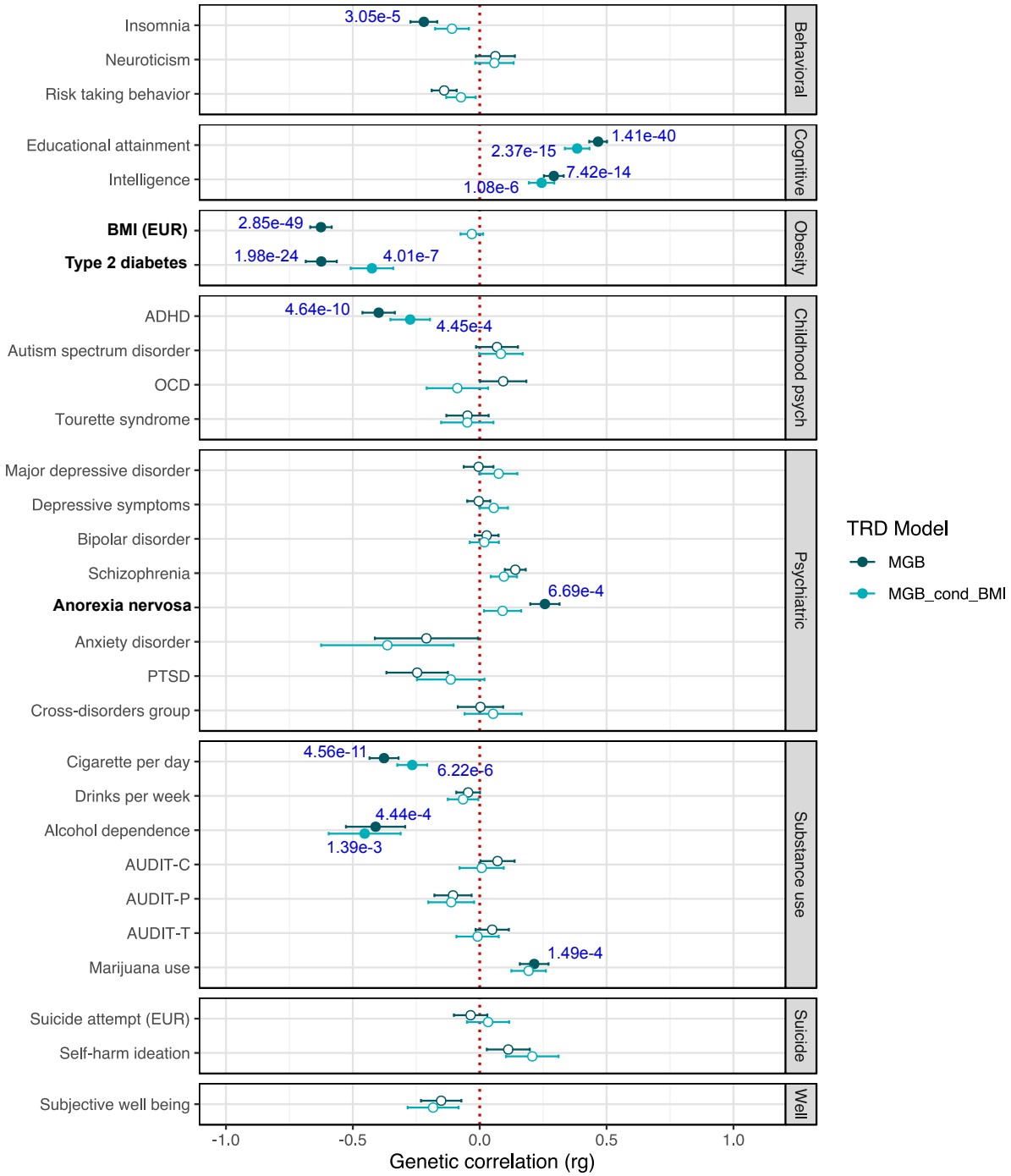

Figure 12: Genetic correlations of MGB TRD meta-analysis GWAS with psychiatric and non-psychiatric traits before and after conditioning for BMI.
Unfilled points indicate genetic correlations that did not pass the Bonferroni-corrected significance threshold $P<1.72x10^{-3}$ (29 traits tested). Error bars represent the standard error. P values indicate significant differences in genetic correlation after conditioning, that pass the Bonferroni correction. Bolded traits show significant differences in genetic correlations between the TRD meta-analysis before and after BMI conditioning with mtCOJO. BMI-body mass index, ADHD-attention-deficit/hyperactivity disorder, OCD-obsessive compulsive disorder, PTSD-post-traumatic stress disorder, AUDIT-C-Alcohol Use Disorders Identification Test-C (measure of quantity of alcohol consumption), AUDIT-P- measure of problematic consequences of drinking, AUDIT-T-total score of AUDIT.

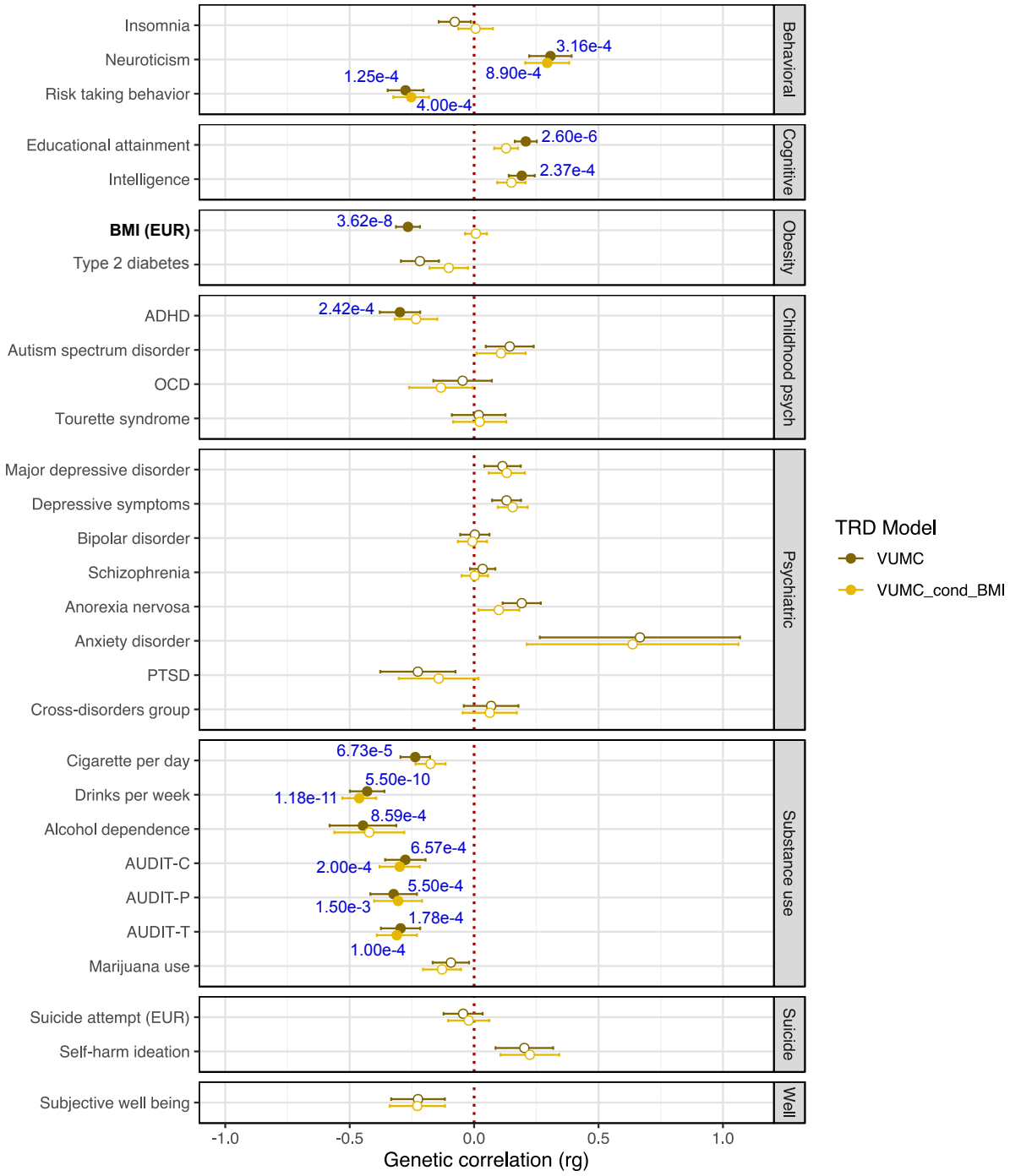

Figure 13: Genetic correlations of VUMC TRD meta-analysis GWAS with psychiatric and non-psychiatric traits before and after conditioning for BMI.
Unfilled points indicate genetic correlations that did not pass the Bonferroni-corrected significance threshold $P<1.72x10^{-3}$ (29 traits tested). Error bars represent the standard error. P values indicate significant differences in genetic correlation after conditioning, that pass the Bonferroni correction. Bolded traits show significant differences in genetic correlations between the TRD meta-analysis before and after BMI conditioning with mtCOJO. BMI-body mass index, ADHD-attention-deficit/hyperactivity disorder, OCD-obsessive compulsive disorder, PTSD-post-traumatic stress disorder, AUDIT-C-Alcohol Use Disorders Identification Test-C (measure of quantity of alcohol consumption), AUDIT-P- measure of problematic consequences of drinking, AUDIT-T-total score of AUDIT.

Discussion

In this study of a quantitative trait reflecting probability of receiving ECT, we found low but significant heritability of TRD, with a single genome-wide significant locus associated with BMI and significant genetic overlap with schizophrenia, cognitive and substance abuse traits, as well as BMI. Application of a computed phenotype from biobank-linked electronic health records allowed detection of these effects in a total of ~152,000 individuals across 4 data sets.

Understanding the genetic architecture of TRD is important for quantifying the role of genetics in treatment response in an effort to move beyond decades-old pharmacogenomic studies. Further, identifying risk loci could facilitate efforts to identify novel treatments in light of the modest response rates observed for interventions other than ECT.

While ECT has been shown by us and others to be a good proxy for TRD, it remains a rare occurrence with prevalence among depression patients much lower than 1%. Even with 152,000 patients, a case-control approach comparing ECT cases to depression controls would have been underpowered with ECT case numbers of ~1400 across all four clinical sites. Leveraging models that can predict ECT from large repositories of clinical data and assign probabilities as quantitative phenotypes allows for substantial increase in power in genetic studies. We showed that that our ECT based prediction models trained both at MGB and VUMC were robust to external validation across three independent healthcare systems. We were also able to show that patients with TRD defined by prescription data had higher probabilities from the ECT models. With quantitative phenotypes, we can benefit from the entire genotyped cohort of 152,000 patients. The increase in power resulted in a significant SNP heritability of 2-4% on the observed scale and significant genetic correlation between the two TRD models. Both TRD models showed significant positive genetic overlap with cognitive traits, and significant negative genetic correlations with ADHD, alcohol and smoking traits, and BMI. We also saw evidence of genetic risk of severe illness of schizophrenia and bipolar disorder even after removing patients with any diagnostic or pharmacologic evidence of schizophrenia or bipolar disorder in our TRD clinical model. However, despite high genetic correlation with each other there were significantly differing genetic overlap with other traits representing potential differences in clinical

population, general population and/or clinical decision making around ECT. We did not see any genetic correlation with other genetic studies of TRD or ECT. However, the comparable ECT study used healthy controls and these two studies were highly genetically correlated with each other pointing to potential that they are predominantly capturing depression genetic architecture as opposed to TRD genetic architecture. Our work shows there is a significant but small contribution of genetics to TRD as defined by ECT. Large studies are currently underway to collect tens of thousands of ECT cases for a case-control study[185] and the comparison to this more timely and efficient approach will be important.

We discovered a single genome-wide significant locus from the MGB model in the intergenic region of the obesity and BMI-related *FTO* gene on chromosome 16. Combined with the highly significant negative genetic correlation with BMI in both the VUMC and MGB models, this suggests the importance of investigating a potential causal role of BMI genetics in TRD or vice versa. However, there are several reasons to be cautious is interpretation of this result. The significant locus was only seen in the MGB model and not the VUMC model. The MGB model had stronger genetic correlation with BMI and the locus did not remain genome-wide significant after statistically conditioning for BMI which also reduced the SNP-heritability from 4% to 2.4% (41% reduction). These results suggest that the genetic association with TRD at this locus is mediated primarily through risk of BMI. One related hypothesis of interest is that there is an overarching reward system pathway that results in anhedonia that increases risk of TRD but also is associated with weight loss and decrease in BMI, and lower risk of substance abuse because disruption to the reward-seek behavior. Future studies looking specifically into anhedonia symptoms among MDD and TRD patients and the comparison of weight and substance use is warranted to test this hypothesis.

We note several additional limitations of our study, particularly the potential confounding of ECT population characteristics in our TRD clinical models. There were significant demographic differences between cases and controls in both sites where a typical ECT case tended to be older, white, male individual with a lower mean BMI and lower level of deprivation compared to MDD controls. These demographic differences could be driven by ascertainment in the medical decisions leading to a patient receiving ECT, such as anesthesia requirements which may exclude patients of extreme weight, and socioeconomic factors

like access to a caregiver as patients need accompaniment after the inpatient ECT procedure. Demographic differences between Nashville and Boston could also contribute to the differences we saw in the VUMC and MGB model, however the two models showed significant genetic overlap (rg=0.72), and both models performed robustly in independent clinical sites with different demographics, especially in the Million Veterans Program cohort which are significantly more male and younger than the other cohorts. Phenotypes based on prediction models are always representative of the original phenotype and could differ in important ways that modify genetic architecture and power. We were able to identify significant but low SNP-heritability meaning that even with our substantial improvements in power many more patients will be required to enable identification of additional genome-wide significant loci. Given such low genetic contribution, an important question is whether ECT represents a generalizable form of TRD such that genetic contribution of TRD broadly is likely as low or whether there is a more biologically homogenous form of TRD. We note that previous estimates of SNP-heritability of TRD within depression patients using prescription data was only slightly higher.

Despite these limitations, this study indicates the utility of investigating a proxy for TRD that can be readily extracted from electronic health records or administrative claims. We confirm a significant but modest genetic contribution to TRD and provide insights into its overlap with other psychiatric and non-psychiatric phenotypes. This effort lays the groundwork for future efforts to apply genomic data for biomarker development, and potentially to identify treatment targets.

CHAPTER IV

Genetic Risk of Suicidal Ideation

Introduction

Suicide has been the leading cause of death for individuals of age 18-45 and rates of suicide attempt (SA) and ideation (SI) are much higher. Lifetime prevalence of suicidal ideation is estimated to be 9.2%. However, the actual prevalence is likely higher, as suicide-related traits of suicidal ideation attempt and death are underreported, in part due to stigma associated with suicide. Suicidal ideation is a major risk factor for suicide attempt and death, but only a subset of individuals with suicidal ideation attempt suicide and even fewer die from suicide. In survey-based studies, 15.6% of individuals with SI attempted suicide within 12 months[186], and 31.8% attempted suicide at some point in their lifetime[63]. Therefore, there has been substantial interest in identifying the phenotypic as well as genetic risks shared between the suicidal thoughts and suicidal behaviors to better understand the similarities and differences in their etiology and potentially contribute to prevention efforts.

Family studies have estimated the heritability of suicidal ideation to be 36% after adjusting for psychopathology[112], and a SNP-heritability estimate of SI measured in a veteran population was 1.2% on the liability scale [MVP SI; Ashley-Koch et al., in review]. In the most recent SI GWAS in veteran populations, genome-wide significant loci in *ESR1* on chromosome 6 and *EXD3* on chromosome 9 were replicated in the independent GWAS of suicide attempt. *EXD3* have been previously been linked with insomnia[144] while *ESR1* has been linked with PTSD and major depressive disorder[187], which are all known risk factors with significant genetic overlap with suicide attempt[188]. The replication of SI GWS SNPs within SA cohort could indicate shared genetics between SI, SA, and psychiatric diseases comorbid with suicide-related behaviors, but there have yet been studies that specifically assess differences in genetic overlap of suicidal ideation and suicide attempt with other psychiatric diseases.

Current methods of ascertaining individuals with suicide-related traits rely on structured data on the electronic health record, including psychiatric assessment questionnaires and diagnostic codes. The current limitations on those methods are they are limited to records of care given to patient directly related to suicide-related traits and administered at the medical center. Freeform clinical text is better able to capture the past medical history of suicide-related traits and psychiatric care given in outside hospitals. Natural language processing (NLP) collectively refers to methods that extract structured information from unstructured text, such as narrative clinical notes. NLP has been demonstrated to improve ascertainment and increase sample size for clinical modeling of phenotypes that lack reliable structured data representation such as adverse child events or homelessness[189] and social determinants of health[190]. Leveraging NLP in clinical notes will help increase ascertainment of suicide-related traits including suicidal ideation in the EHR by incorporating information uniquely available in unstructured clinical text.

Improved ascertainment of individuals with suicide-related traits is necessary for the comparison of those who present with multiple suicide-related traits with those with only suicidal ideation. For the overlap of suicidal ideation and attempt, prior studies have either studied them without distinction[92] under the umbrella of suicidality, or studied for suicidal ideation while excluding co-occurrence of suicide attempt and suicide death using multiple sources of information in the EHR including international classification of diseases (ICD9 or ICD10) codes, mental health surveys, and death registries [MVP SI].

In this study, we conduct a GWAS of SI with 1,849 cases and 62,911 controls, where the SI cases were ascertained using both structured EHR data and NLP. We assessed for genetic overlap with an external SI GWAS in the US veteran population and with the genetics of suicide attempt and tested for differences in genetic correlation with psychiatric traits between suicidal ideation and suicide attempt.

Methods

Sample site

Clinical and genetic data were used the BioVU Synthetic Derivative[156], which stores deidentified clinical electronic health record data from over 3.4 million patients receiving care at Vanderbilt University Medical Center (VUMC). VUMC is an academic medical center in Nashville Tennessee that manage over 2 million patient visits every year across Tennessee and its neighboring states.

Case definition of suicidal ideation

Deidentified clinical data were extracted from the VUMC Synthetic Derivative[156]. Cases of suicidal ideation (SI) were ascertained using 4 sources: 1) patients who said yes/confirmed suicidal ideation in psychiatric hospital screener questionnaire, 2) Patients with International Classification of Diseases, 9th/10th Revision, Clinical Modification (ICD-9/10-CM) of suicidal ideation (ICD9: V62.84, ICD10: R45.851), 3) SI cases from manual review, and 4) SI cases from natural language-processing (NLP) of EHR notes using a 80% positive predictive value (PPV) cutoff[191]. The NLP scores are from Bejan et al.[191] where word2vec method[192] was used to generate list of seed words to identify suicidal ideation, and patients were ranked based on similarity of patients' notes and suicidal ideation query vector. For a PPV cutoff, precision of top K ranked results (P@K) were used, where threshold K was determined to get 80% precision where top K ranked patients resulted in 80% precision (P@K=80%). Individuals with any evidence of SI across all four sources were considered a SI case. Manual validation of SI cases was performed as part of a validation effort of various ascertainment methods of suicidal ideation and attempt by Bejan et al., and was performed for individuals in the top 200 highest ranked patients of suicidal ideation, 200 randomly selected individuals with SI ICD10 codes, and 10 randomly selected individuals with psychiatric screener form data.

To identify individuals with both evidence of SI and SA, SA case status was determined by ICD codes, psychiatric hospital screener questionnaire response, and manual review. ICD codes used to identify SA were ICD9/10 codes of suicide attempt (ICD9: E95*.*, E98*.*; ICD10: T14.91, T14.91*; * denoting wildcard digits), history of self-harm (ICD10 Z91.5), or intentional self-poisoning/self-harm (ICD10 X6*, X7*, X8*, T36*-T71*).

Control definition

Controls were defined as individuals who matched any of the following criteria: 1) negative or absence of positive assertions to SI in psychiatric forms of suicide assessment, 2) absence of SI ICD codes, 3) negative manual review, or 4) individuals that were not included in the 80% PPV cutoff for NLP ascertainment of SI.

Genotyping, quality control, imputation

Standard quality control procedures were applied to the genotype data of BioVU individuals genotyped by the BioVU Infinium expanded multi-ethnic genotyping array (MEGAEX), as described previously in Chapter III. Only individuals of European ancestry were included for genetic analyses.

Genome-wide association study (GWAS)

SI GWAS was conducted using the SI cases ascertained based on psychiatric forms for suicide assessment, ICD-9/10 codes, manual review, and NLP. Firth regression of the binary SI phenotype was performed on 1,849 SI cases and 62,911 controls using Regenie v2.2[193], with age, sex and genetic ancestry-informative principal components 1-20 as covariates (22 covariates) to account for population stratification. Default settings of block size 200 and 20 threads were used. Variants with minor allele frequency < 0.01 were excluded.

LD score regression (LDSC)

LD score regression[171] was used to estimate the phenotypic variance of SI explained by common SNPs (SNP-heritability, h^2_{SNP}) from GWAS summary statistics. $h^2_{SNP}$was calculated on the liability scale using population prevalence of k=0.09[63]. LDSC bivariate genetic correlations attributable to genome-wide SNPs (rg) were estimated between GWAS of suicide attempt from the International Suicide Genetics Consortium (ISGC)[188] and the GWAS of suicidal ideation from MVP [Ashley-Koch et al., in review], as well as other psychiatric and non-psychiatric traits. The Bonferroni corrected significance threshold was $P < 2.1x10^{-3}$, adjusting for 24 traits tested. Differences in genetic correlation between suicidal ideation and suicide attempt were tested using the block jackknife method, implemented in LDSC software[173].

Polygenic risk score (PRS) analysis

PRS of suicidal ideation calculated using the summary statistics from the MVP study was tested for association with SI in our genotyped samples. PRS analyses were performed using PRS-CS which places a continuous shrinkage prior to SNP effect sizes using a Bayesian regression framework[176]. The continuous shrinkage priors adapt the amount of shrinkage applied to each SNP to the strength of the associated GWAS signal based on the LD structure estimated from an external reference panel. PRS were generated in each cohort using PRS-CS and the 1000 Genomes European reference panel was used to estimate LD between SNPs. The PRS were summed for each individual of the target cohort using PLINK1.9 build 3.42[136]. Polygenic risk score using MVP SI GWAS was tested for association with binary suicidal ideation status in the VUMC target cohort using logistic regression model, covarying with PC1-PC10, sex, and age.

Results

Demographics of SI cases by ascertainment method

Leveraging clinical data from the EHR at VUMC, we identified 34,642 patients with evidence of suicidal ideation. In descriptive analyses, we examined the demographic characteristics of SI cases across different ascertainment methods (**Table 16**). Age showed the biggest differences across ascertainment methods, where chart validated SI cases were the oldest (40.9 ± 19.9) and ICD-10 SI cases were the youngest (31.5 ± 17.2). Given the large proportion of ICD-10 cases in the total SI case sample, the mean age of all SI cases is closer to that of ICD-10 SI cases (33.8 ± 17.5). The proportions of gender, race, and ethnicity were similar across ascertainment methods, where SI cases are mostly White (76.9-80.8%), non-Hispanic (89.5-94.4%), and female (53-56.8%). There were no differences in socioeconomic status as measured by area deprivation index[160] which is a composite measure incorporating poverty, income, education, insurance coverage and housing.

Table 16: Demographics of all SI cases across different ascertainment methods.
Deprivation index refers to the normalized score ranging 0-1 of six different measures of American Community Survey (includes measure of poverty, income, education, health insurance coverage, and housing) for each census tract, with higher index indicating more deprivation[194].

Demographic		All SI cases (N=34,642)	Chart validated (N=991)	Screener (N=5,232)	SI ICD9 (N=17,715)	SI ICD10 (N=19,958)	NLP PPV > 80% (N=5,325)
Age	Mean	33.8 ± 17.5	40.9 ± 19.9	34.6 ± 15.7	35.0 ± 16.6	31.5 ± 17.2	36.6 ± 19.2
	Median	29	40	31	31	26	32
Gender	Female	18749 (54.1%)	563 (56.8%)	2952 (56.4%)	9383 (53%)	10848 (54.4%)	2983 (56%)
	Male	15891 (43.2%)	428 (43.2%)	2280 (43.6%)	8331 (47%)	9109 (45.6%)	2342 (44%)
	Unknown	2 (0.01%)	0 (0%)	0 (0%)	1 (0%)	0 (0.01%)	0 (0%)
Race	White	27268 (78.7%)	787 (79.4%)	4229 (80.8%)	14231 (80.3%)	15340 (76.9%)	4110 (77.2%)
	Black	5156 (14.9%)	152 (15.3%)	717 (13.7%)	2711 (15.3%)	2978 (14.9%)	862 (16.2%)
	Asian	380 (1.1%)	9 (0.9%)	51 (1.0%)	149 (0.8%)	255 (1.3%)	88 (1.7%)
	*Other	1838 (5.3%)	43 (4.3%)	235 (4.5%)	624 (3.5%)	1385 (6.9%)	265 (5.0%)
Ethnicity	Hispanic	1249 (3.6%)	44 (4.4%)	131 (2.5%)	505 (2.9%)	851 (4.3%)	204 (3.8%)
	Non-Hispanic	31771 (91.7%)	920 (92.8%)	4876 (93.2%)	16721 (94.4%)	17865 (89.5%)	4943 (92.8%)
	Unknown	1622 (4.7%)	27 (2.7%)	225 (4.3%)	489 (2.8%)	1242 (6.2%)	178 (3.3%)
Area Deprivation Index		0.35 ± 0.12	0.36 ± 0.14	0.35 ± 0.12	0.36 ± 0.12	0.35 ± 0.12	0.35 ± 0.13

*Other includes all other races including unknowns and combination of races

Percentages are reported in parentheses, valules are reported as mean ± standard deviation

Demographic differences of SI only and SI with SA cases

We also examined demographic differences between SI cases who did or did not also have evidence of SA. Compared to SI only cases, SI cases with also evidence of SA were 4 years younger, 5.4% more likely to be female, and 0.8% less likely to be non-Hispanic (**Table 17**). There were no differences in the three largest race categories or area deprivation index between the two groups.

Table 17: Demographics of SI cases compared to SI with SA cases

Demographic		Only SI (N=21,867)	SI with SA (N=12,775)	Significance testing*
Age	Mean	35.4 ± 18.2	31.1 ± 15.8	**4.81E-117**
	Median	31	26	
Gender	Female	11248 (51.4%)	7501 (56.8%)	**3.10E-39**
	Male	10617 (48.6%)	5274 (43.2%)	**3.10E-39**
	Unknown	2 (0%)	0 (0%)	0.728
Race	White	17267 (79.0%)	10001 (78.3%)	0.141
	Black	3253 (14.9%)	1903 (14.9%)	0.972
	Asian	243 (1.1%)	137 (1.1%)	0.333
	**Other	1104 (5.0%)	734 (5.7%)	**5.66E-03**
Ethnicity	Hispanic	756 (3.5%)	493 (3.9%)	**0.0567**
	Non-Hispanic	20126 (92.0%)	11645 (91.2%)	**4.26E-03**
	Unknown	985 (4.5%)	637 (5.0%)	**0.0432**
Area Deprivation Index		0.35 ± 0.12	0.35 ± 0.14	**3.15E-03**

*Significance testing: t-test for quantitative values, 2 proportions Z-test for categorical variables

**Other includes all other races including unknowns and combination of races

Percentages are reported in parentheses, valules are reported as mean ± standard deviation

Suicidal ideation cases show little convergence among different ascertainment methods

We next examined sample overlap among SI cases across different ascertainment methods (**Figure 14**). ICD codes were the most frequent source of ascertainment, with 93.1% (32,284) of all SI cases being identified by ICD-9 or ICD-10 codes. Most SI cases only had SI ICD code as evidence: among individuals with SI ICD codes, only 24.1% (7780) of SI cases had other evidence of SI in addition to ICD codes. On the other hand, SI cases identified from the psychiatric screener or NLP showed high overlap with other ascertainment methods. 86.3% (4,514) of SI cases identified by the screener and 77.6% (4,134) of SI cases identified via NLP had some other evidence of SI. 26.2% (260) of chart-validated SI cases did not have any other evidence of SI. Ultimately, the NLP method identified 1,191 (3.5% of total SI case set) additional SI cases with no other evidence of SI from ICD, psychiatric screener, or manual validation.

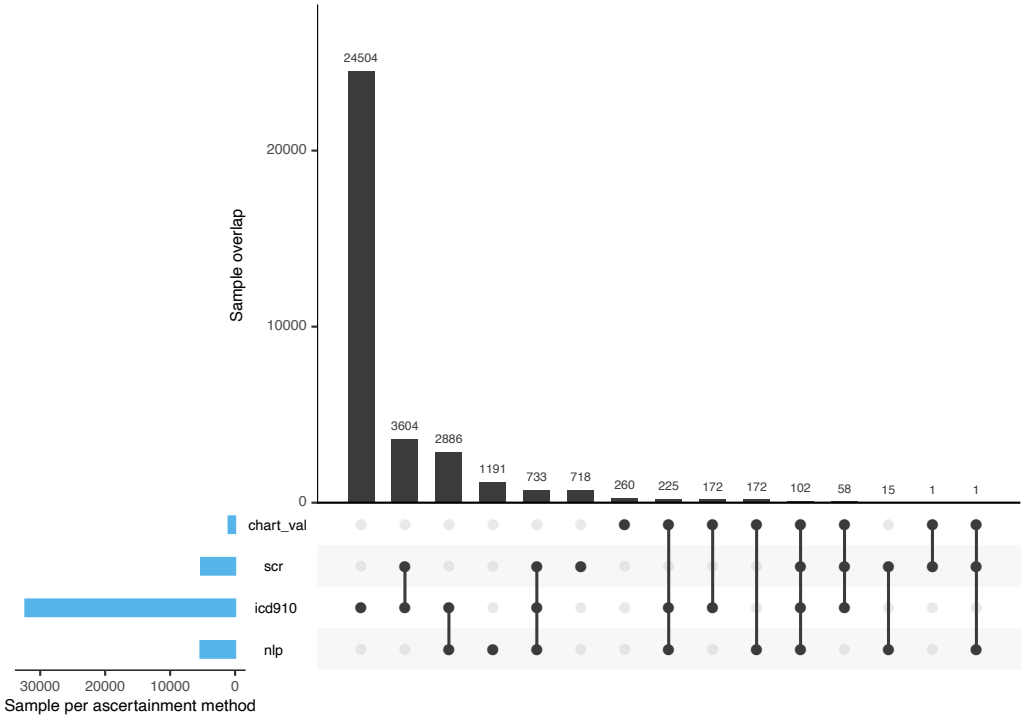

Figure 14: Suicidal ideation cases sample overlap by ascertainment method.
Chart_val: chart validation; scr: psychiatric suicidal ideation screener questionnaire; icd910: ICD-9/ICD-10; nlp: NLP method

Comparison of demographic characteristics of genotyped SI cases and controls

We also performed descriptive analyses of SI cases and controls among genotyped samples of European ancestry and identified several demographic differences (**Table 18**). SI cases on average were 12 years younger, 2.5% more likely to be female, 6.3% more likely to be white, and 5.7% more likely to be non-Hispanic compared to non-SI controls. SI cases had a 3% lower socioeconomic status as measured by higher ADI.

Table 18: Demographic description of genotyped SI cases and non-SI controls.
Deprivation index refers to the normalized score ranging 0-1 of six different measures of American Community Survey (includes measure of poverty, income, education, health insurance coverage, and housing) for each census tract, with higher index indicating more deprivation[194].

Demographic		SI case (N=1849)	non-SI control (N=62,911)	Significance testing*
Age	Mean	42.8 ± 20.1	54.4 ± 22.1	**2.12E-115**
	Median	42	59	
Gender	Female	1077 (58.2%)	35066 (55.7%)	**3.43E-02**
	Male	772 (41.8%)	27845 (44.3%)	**3.43E-02**
Race	White	1833 (99.1%)	58364 (92.8%)	**9.58E-26**
	Black	1 (0.1%)	55 (0.1%)	0.937
	Asian	0 (0%)	23 (0%)	0.844
	**Other	15 (0.8%)	4469 (7.1%)	**1.34E-25**
Ethnicity	Hispanic	9 (0.5%)	290 (0.5%)	1.00
	Non-Hispanic	1827 (98.8%)	58567 (93.1%)	**7.05E-22**
	Unknown	13 (0.7%)	4054 (6.4%)	**1.86E-23**
Area Deprivation Index		0.33 ± 0.12	0.32 ± 0.11	**2.93E-05**

*Significance testing: t-test for quantitative values, 2 proportions Z-test for categorical variables

**Other includes all other races including unknowns and combination of races

Percentages are reported in parentheses, valules are reported as mean ± standard deviation

Heritability estimate of SI and genetic correlations with suicide-related traits

SNP heritability estimated using LDSC was 1.4% (SE=0.0068, P=0.016) on the observed scale and 13.4% (SE=0.0628, P=0.017) on the liability scale using population prevalence of 9%[63]. This heritability estimate is much higher than the 1.2% (SE=0.0009) liability scale heritability estimate measured in the MVP veteran cohort, but comparable compared to the 10.1% (SE=0.01) liability scale heritability estimate of for self-harm ideation in an Australian cohort[93]. The observed heritability of SI was significantly different from 11.6% (SE=0.0088) observed heritability of suicide attempt[188] ($p=6.71x10^{-22}$ block jackknife). The GWAS of suicidal ideation did not identify any genome-wide significant (GWS) loci (**Figure 15**). Using our SI GWAS, we tested replication of the two GWS SNPs identified in the SI GWAS from MVP among individuals of European ancestry. Neither GWS SNP were significant, rs13211166 on chromosome 6 (BETA=-0.079, P= 0.072) and rs73581580 on chromosome 9 (BETA=0.060, P=0.232).

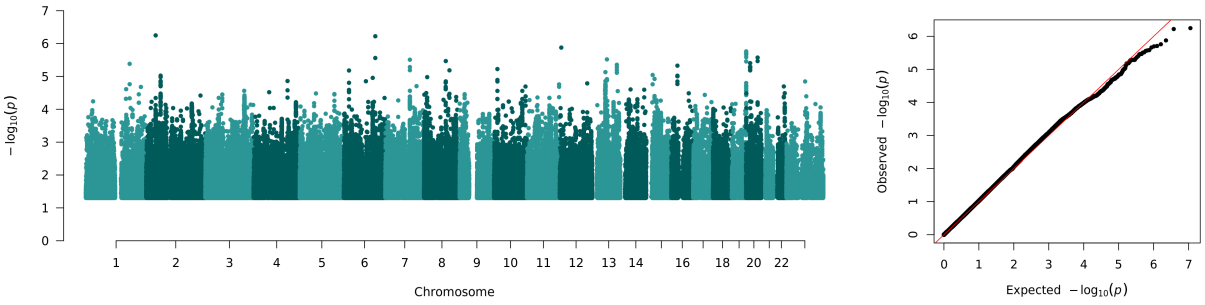

Figure 15: Manhattan plot and QQ plot of the SI GWAS.

Genetic correlations were calculated to test the genetic overlap between SI and other suicidal behaviors as well as other 19 psychiatric traits (**Table 19**). Among suicide-related traits, a significant genetic correlation was observed with suicide attempt (rg=0.80, SE=0.25, P=1.20x10^{-3}) and SI. Genetic correlation with self-harm behavior (rg=0.66, SE=0.29, P=0.024) and self-harm ideation (rg=0.58, SE=0.23, P=0.012) were not significant after multiple testing correction.

Genetic correlations of SI with psychiatric traits and differences in genetic correlation with SA

Among psychiatric traits, SI showed a significant genetic correlation with MDD (rg=0.79, SE=0.26, P=1.87x10^{-3}), depressive symptoms (rg=0.71, SE=0.18, P=7.04x10^{-5}), and bipolar disorder (rg=0.69, SE=0.21, P=9.22x10^{-4}) after multiple testing correction (**Table 19**). Significant genetic correlation was also observed with the cross-disorder group GWAS which represent multiple psychiatric disorders. To investigate whether these genetic correlations with SI were significantly different from those with SA, we examined the same genetic correlations with ISGC SA. No significant differences in genetic correlations were observed between SI and SA except for educational attainment (P=6.99x10^{-3} block jackknife), where a significant negative genetic correlation with educational attainment that is observed with SA (rg =-0.28, SE=0.029, P=4.57x10^{-22}) was no longer significant with SI (rg=-0.06, SE=0.078, P=0.44).

Table 19: Genetic correlations of various psychiatric and suicidal behavior traits with SI GWAS.

Trait	rg	se	z	p	Reference
Major depressive disorder	0.7979	0.2566	3.1095	**1.87E-03**	Wray et al. 2018
Depressive symptoms	0.7111	0.1789	3.9748	**7.04E-05**	Howard et al. 2019
Post-traumatic stress disorder	0.7346	0.3177	2.312	0.0208	Nievergelt et al. 2019
Bipolar disorder	0.6889	0.2079	3.3133	**9.22E-04**	Mullins et al. 2021
Schizophrenia	0.4212	0.1457	2.8906	3.85E-03	Ripke et al. 2020
Anorexia nervosa	0.4177	0.1974	2.1157	0.0344	Watson et al. 2019
Autism spectrum disorder	0.5789	0.1889	3.0649	2.18E-03	Anney et al. 2017
Attention deficit hyperactivity disorder	0.5389	0.184	2.9286	3.40E-03	Demontis et al. 2019
Tourette syndrome	0.4052	0.1856	2.1826	0.029	Yu et al. 2019
Obsessive-compulsive disorder	0.2271	0.2258	1.0059	0.314	Askland et al. 2017
Multiple psychiatric disorders (CDG)	0.8113	0.2245	3.6145	**3.01E-04**	Lee et al. 2019
Alcohol dependency	0.8445	0.318	2.6556	7.92E-03	Walters et al. 2018
Alcohol Use Disorders Identification Test (AUDIT)	-0.063	0.1462	-0.4311	0.666	
AUDIT - consumption	-0.2119	0.1541	-1.3748	0.169	Sanchez-Roige et al.2018
AUDIT - problematic behavior	0.4878	0.1922	2.5373	0.011	
Marijuana use	0.3872	0.1536	2.5215	0.012	Pasman et al. 2019
Risk-taking behavior	0.3504	0.132	2.6536	7.96E-03	Linnér et al. 2019
Insomnia	0.324	0.1475	2.1967	0.028	Jansen et al. 2018
Educational attainment	-0.0599	0.0783	-0.7653	0.444	Lee et al. 2018
Suicide attempt	0.8037	0.2482	3.2382	**1.20E-03**	Mullins et al. 2022
Self-harm behavior	0.6616	0.2931	2.2572	0.024	Campos et al. 2020
Self-harm ideation	0.5793	0.2301	2.5176	0.012	

Genetic overlap of MVP SI and VUMC SI GWAS

We examined the genetic correlation of our SI GWAS with the previously published MVP SI GWAS, a GWAS of SI cases without evidence of SA in the American veteran population. No significant genetic correlation was observed between VUMC SI and MVP SI (rg=0.12, SE=0.10, P=0.25) using LDSC. When genetic overlap was tested using PRS association testing which is more robust to low power, there was a significant association of SI with MVP SI PRS among 62,809 VUMC genotyped samples (BETA=0.053, SE=0.024, P=0.026). To test whether certain ascertainment methods were enriched for SI individuals with high genetic risk of SI, MVP SI PRS was compared among SI cases across different ascertainment methods (**Figure 16**). Compared to the MVP SI PRS of SI controls, nominally significant differences in PRS were observed among ICD9, psychiatric screener questionnaire, NLP, and chart validation methods of ascertainment, but none were significant after multiple testing correction (**Figure 16A**, **Table 20**). When SI cases and controls with evidence of SA were excluded, there were no significant

differences in PRS between SI cases and controls. With the exclusion of individuals with SA, the effect size of MVP SI PRS decreased compared to when SA cases were not excluded, except for SI cases ascertained using screener questionnaires (**Figure 16B**, **Table 20**). There were no significant differences in MVP SI PRS was compared between individuals with only SI to individuals with both SI and SA (Effect size estimate=0.022, SE=0.050, p=0.655, generalized linear regression using sex, age, and PC1-PC10 as covariates).

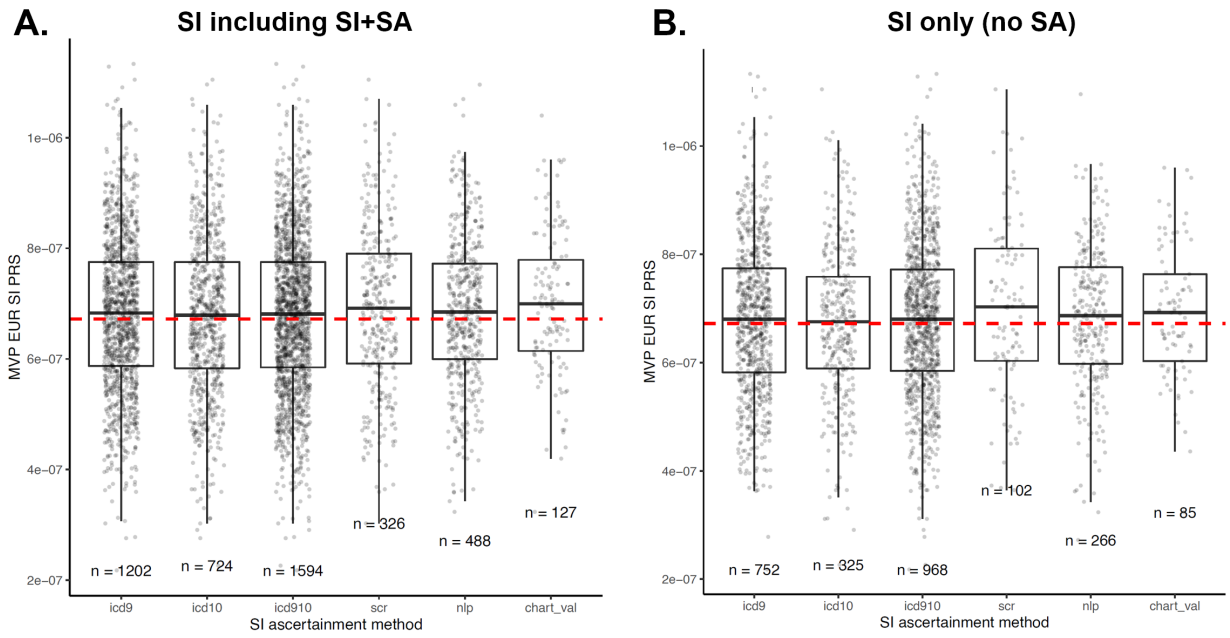

Figure 16: MVP SI PRS distribution of SI cases across different ascertainment methods.
Red dotted line indicates average MVP SI PRS for controls. A. Suicidal ideation case and controls are not filtered for evidence of suicide attempt (using screeners, ICD9/10 codes, and chart validation), and comparison is against 68,243 controls. B. Suicidal ideation cases and controls with evidence of suicide attempt are excluded, and comparison is against 67,642 controls. Chart_val: chart validation; scr: psychiatric screener questionnaire; icd910: ICD-9/ICD-10; nlp: NLP method

Table 20: Regression results of MVP SI PRS of SI cases ascertained via different methods compared to controls. Linear regression was conducted with SI status as outcome, and MVP SI PRS as a predictor, with age, sex, and PC1-PC10 as covariates.

Ascertainment	SI including SI+SA				SI only (no SA)			
Method	Est	SE	P	N	Est	SE	P	N
ICD9	0.058	0.029	**0.046**	1202	0.045	0.037	0.221	752
ICD10	0.030	0.038	0.424	724	0.008	0.056	0.884	325
ICD9/ICD10	0.045	0.026	0.081	1594	0.039	0.033	0.234	968
Screener	0.132	0.056	**0.018**	326	0.184	0.100	0.065	102
NLP	0.091	0.046	**0.045**	488	0.074	0.062	0.229	266
Chart validation	0.193	0.089	**0.030**	127	0.135	0.109	0.214	85

Discussion

In this study we conducted a GWAS of suicidal ideation (SI) ascertained both from structured and unstructured EHR data. Significant heritability of SI was detected along with significant genetic overlap with suicide attempt. Significant genetic correlation was also observed with depression, bipolar disorder, and other psychiatric disorders with comparable estimates to those of suicide attempt. There was no significant genetic correlation between our SI GWAS and an external SI MVP GWAS, but SI cases had significantly higher polygenic risk based on the MVP SI GWAS compared to controls. SI cases with evidence of SA were younger and more likely to be female compared to SI only cases, and there was no difference in MVP SI PRS between those two groups.

We first characterized the overlap of SI cases across different ascertainment methods to assess the variability of SI ascertainment. ICD codes were the primary source for identifying SI cases, signifying that the majority of SI cases ascertained are limited to SI as a primary concern of a visit at VUMC, and past medical history of SI and SI detected at outside hospitals are not being captured. Descriptive analyses of SI cases showed that the SI cases were 12 years younger than controls, this is likely a result of the large proportion of SI cases identified from ICD10 codes which have an overall younger patient population due to its more recent use in the VUMC EHR.

GWAS of binary SI resulted in a significant heritability estimate of 13.4% on the liability scale, and this estimate was comparable to the 10.1% heritability estimate of self-harm ideation but much higher than the 1.2% heritability estimate of MVP SI. One notable difference between the two SI GWAS is that the MVP SI study excludes SI cases with evidence of SA or suicide death, and the higher heritability estimate may be a result of a third of our SI cases having evidence of suicide attempt. The Australian self-harm ideation GWAS used an even broader case definition that includes non-suicidal self-harm which resulted in a higher prevalence than any suicide-related traits (14.8% prevalence in UK Biobank[93]). The inclusion of non-suicidal self-harm ideation in the broad self-harm ideation may result in a lower heritability estimate than suicidal ideation due to increased phenotype heterogeneity and higher misclassification rates, as was observed with the comparison of strict vs minimal definitions of depression[195]. Current heritability

estimates of self-harm ideation and suicidal ideation have overlapping standard errors. Further investigation is needed to assess the effect of broad phenotyping of suicidal ideation.

Comparing liability scale heritability estimates across suicide-related traits, our SI heritability estimate was higher than the 6.8% heritability estimate for SA but similar to the 16% heritability estimate of suicide death. However, using observed scale heritability estimates, heritability of SA was significantly higher than that of SI (11.6% vs 1.4%, $p=6.71x10^{-22}$ block jackknife), which suggests that the difference in heritability in the liability scale is a result of the higher population prevalence of SI (9%) compared to SA (2%)[63]. We observed a high genetic correlation of SI with SA (rg=0.80, SE=0.25) which was not significantly different from 1, and this genetic overlap would explain the similarity of genetic correlations with psychiatric traits, especially with depression and bipolar disorder. The genetic overlap of SI and SA suggest a shared genetic etiology and is in line with the fact that suicidal ideation is necessary to attempt suicide.

We compared our SI with an external SI GWAS conducted in the US veteran population via replication of GWS SNPs and assessment of genetic overlap. No GWS SNPs of MVP SI GWAS were replicated and there was no significant genetic correlation measured using LDSC, but there was a significant positive association of MVP SI PRS and SI status (BETA=0.053, SE=0.024, P=0.026). There was no ascertainment method that resulted in SI cases with significant enrichment of MVP SI PRS. SI with SA likely represents a patient population with more severe and recurrent suicidal ideation symptoms, but there were no differences in MVP SI PRS between SI with SA and SI only cases. Further investigation with larger sample sizes and diverse clinical settings is needed to determine genetic differences among different SI populations, especially between SI only and SI with SA cases.

A limitation of the study is that the NLP method was only able to identify a small fraction of SI cases and likely not fully ascertaining individuals with past medical history of SI especially treated in outside hospitals. This demonstrates that at the 80% PPV threshold there are other structured evidence of SI, and the PPV threshold would need to be lowered to capture additional SI cases. Lowering the PPV threshold would result in an increased false positive rate. Identifying and removing these false positives

remains a challenge and additional effort around accurately identifying negative evidence from screening would help. In this study we did not exclude cases with evidence of suicide attempt, which would affect genetic correlation results with suicide attempt. However, the MVP SI study which excluded SI cases with evidence of SA and suicide death also observed a similar significant genetic correlation with SA (rg=0.77, SE=0.05, $p=2.15x10^{-53}$), suggesting that the genetic overlap with SI and SA would remain similar even when individuals with evidence of SA are excluded.

In this study we conducted a GWAS of suicidal ideation including cases identified using NLP to ascertain cases from unstructured clinical data. We found evidence for high genetic correlation with suicide attempt and significantly higher genetic risk of suicidal ideation among cases. There were no differences in genetic risk of suicidal ideation among individuals with evidence of suicide attempt as well. Genetic correlation with psychiatric traits such as depressive disorders and bipolar disorder were not significantly different from those with suicide attempt. These findings shed light on the significant genetic overlap between suicide attempt and suicidal ideation, especially those that are readily ascertained via electronic health data. More effective ascertainment of past medical histories of suicidal ideation and capturing important aspects of suicidal ideation such as the existence of a plan in clinical notes will be critical in characterizing suicidal ideation. Such advancements in NLP phenotyping will help gain a better understanding progression of suicidal ideation to suicide attempt and assess whether genetics can inform risk stratification and intervention.

CHAPTER V

Discussion

In this dissertation I demonstrate three projects that utilize biobanks to increase power and gain insight into the genetic architecture of complex psychiatric traits: 1) an international consortium GWAS of suicide attempt where biobanks facilitated case ascertainment from multiple sites, 2) clinical prediction modeling of TRD using biobanks to capture the clinical features of individuals receiving electroconvulsive therapy, and 3) a GWAS of suicidal ideation that utilizes both structured and unstructured data in biobanks to ascertain cases.

In Chapter 2, I studied the genetics of a major prevalent psychiatric risk factor and rare but devastating outcome: depression and suicide. This multi-cohort trans-ancestry meta-analysis of suicide attempt GWAS demonstrates the power of an international consortium in dissecting the genetic overlap of two phenotypes with complex genetic architecture. Suicide-related behaviors have been observed with higher prevalence in multiple psychiatric disease, and there is a debate on whether suicide attempt is a manifestation of severe psychiatric illness, or a separate entity. The results of this chapter lend evidence to the latter, where while there is significant genetic sharing between suicide attempt and psychiatric disease, there is also a genetic component that contributes stronger to suicide attempt. The genetic overlap seen with smoking, insomnia, and risk-taking behavior suggest that the genetic architecture of suicide may represent a combination of impulsivity and rumination thinking pattern, which has already been described in multiple theories of suicide. This hypothesis would explain why only a minority of individuals with psychiatric disease attempt suicide, and it is possible that there is an additive effect of psychiatric disease and rumination and impulsivity. It is also important to take into consideration the environmental factors that may trigger suicide-related behaviors. Once we are able to incorporate important environmental covariates including traumatic life experiences such as physical, mental or sexual abuse, diagnosis of a terminal illness, loss of a loved one, or unstable livelihood, we would truly be able to study suicide attempt as a response-

to-stimuli phenotype. Complex phenotypes are a manifestation of genetic and environmental factor, and the closer we are able to mimic that in experimental design would we be able to better understand and model clinical risk.

In Chapter 4, I examined the genetic application of an NLP phenotyping method to extract relevant clinical information from unstructured data, which expanded sources of case ascertainment in the electronic health record. Structured clinical data from current electronic health records rely on billing codes, which limit data collection to information that are directly related to a service provided at a medical center. In countries like the United States which does not have a national health care system, collected clinical data are often fragmented, as patients change health care plans and providers. Therefore, unstructured clinical notes are an important source of valuable information such as past medical history or care provided in external care sites. While in this study only a small fraction of cases was ascertained from NLP, unstructured data remain an important resource especially in extracting environmental factors such as past traumatic experiences, which will be particularly important in studying psychiatric diseases and suicide-related traits. As for the genetics of suicide-related traits, there are currently large GWAS studies that have been conducted on each trait separately, but none that have examined genetic differences in individuals who present with one or multiple suicide-related traits. Granular phenotyping using multiple suicide-related traits in longitudinal patient data will be critical in studying the co-occurrence of suicide-related traits. Furthermore, there are two critical pieces of information that remain to be incorporated into genetic studies of suicide-related traits, which are the existence of a plan for suicidal ideation and means of self-harm for suicide attempt and death by suicide. While there are ICD10 codes that specify means of self-harm, a large proportion of the EHR has ICD9 information, and there may be variability in encoding suicide-related traits in different academic centers, so NLP will be useful in extracting information regarding means of suicide. Data on suicide plan and means of suicide will help quantify sy5mptom severity and establish subphenotypes within suicide-related traits which may present with different genetic overlaps. For example, assuming a suicidal ideation with a plan represents a higher degree of intent-to-self-harm, it is possible that suicidal ideation without a plan has a higher genetic overlap with substance use disorder and smoking,

while suicidal ideation with a plan has a higher genetic correlation with suicide attempt. Means are an important consideration for suicide attempt and death because of varying lethality, and it is possible that there may be genetic differences in suicide attempt cases ascertained in different regions because of the differences in suicide prevention policies (e.g., stricter gun regulations).

In Chapter 3, we demonstrated characterization of a broader phenotype of treatment-resistant depression using electroconvulsive therapy, a treatment option for only a small subset of TRD patients. The goal of studying treatment resistance in depression is identifying genetic and clinical risk factors enriched in patients with the most severe and retractable depressive symptoms, yet significant advancement has been thwarted for decades by the challenge of establishing a set definition of treatment and response that can be replicated across multiple studies. For this reason, we chose electroconvulsive therapy, a rare but established treatment of TRD, to characterize a quantitative predicted probability of TRD. The utilization of clinical prediction models allowed us to harness the genetic information of over ~152,000 individuals by generating prediction scores for all MDD patients rather than a classical ECT case vs control study. Despite using a ECT prediction score to characterize TRD, our study observed similar genetic relationships as other case control studies that ascertained TRD cases using medication data or ECT receipt, such as the relationship with cognitive traits. We also observed genetic overlap with traits that have yet been linked with TRD, such as substance use disorders and BMI. An important caveat of using the ECT clinical model however is that ECT patients are a select subset of individuals with severe depression, and it is possible that the genetic overlap seen with TRD and BMI or cognitive traits are largely explained by clinical decision surrounding ECT administration. Results from large ECT consortiums[196] would have to be carefully interpretated to account for known biases of the ECT patient population, such as higher medical literacy and socioeconomic status.

Biobanks are imperfect resources with several known biases, including demographic differences in the local population and patient population and differences in medical practice across biobanks. They are however a massive and ever-growing source of phenotypic data, and a boon to computational geneticists. Suicide-related traits and depression are phenotypes that are complex and highly interconnected, and

without biobanks we would not have had the power to make the headway that we did in the past five years in understanding the contribution of common genetics to these polygenic traits. A lesson I have learned from these projects is that the full power of biobanks can be harnessed with standardization of phenotyping applied across multiple sites. I have also learned that phenotyping is an iterative process that aims to find the balance between obtaining adequate cases and thus power vs increasing heterogeneity of phenotype which introduces more confounding factors that complicate interpretation of genetic associations. With these lessons, I hope to continue leveraging biobanks to better characterize the clinical and genetic risk of rare outcomes from common risk factors and help develop better risk stratification and clinical decision-making tools.

APPENDIX

Supplementary Table 1: Characteristics of cohorts in the International Suicide Genetics Consortium

Supplementary Table 1: Characteristics of cohorts in the International Suicide Genetics Consortium

Cohort	SA*		SA within psychiatric diagnosis*		Suicide phenotype	Suicide phenotype ascertainment	Psychiatric disorder(s) in SA cases^	Psychiatric disorder for GWAS of SA within psychiatric diagnosis^	Ancestry	Country of ascertainment
	Cases	Controls (general population)	Cases	Controls (psychiatric disorder no SA)						
Psychiatric Genomics Consortium MDD	1528	16626	1677	8793	SA	structured psychiatric interviews	MDD	MDD	EUR	many
Psychiatric Genomics Consortium BIP	3214	17642	3214	5408	SA	structured psychiatric interviews	BIP	BIP	EUR	many
Psychiatric Genomics Consortium SCZ	1640	7112	1668	2928	SA	structured psychiatric interviews	SCZ	SCZ	EUR	many
Psychiatric Genomics Consortium ED	170	5070	198	583	SA	structured psychiatric interviews	AN	AN	EUR	many
CONVERGE	1148	6515	1148	1183	SA	Clinical interview, CIDI	MDD	MDD	EAS	China
Grady Trauma Project	669	4473	355	1116	SA	self-report	PTSD, SCZ, BIP	PTSD	AA	US
Army STARRS	670	10637	376	3447	SA	modified version of Columbia Suicide Severity Rating Scale (C-SSRS)	MDD, PTSD, anxiety	MDD	EUR	US
German BorderlineGenomics Consortium	481	1653	481	144	SA	self-report	Borderline Personality Disorder	Borderline Personality Disorder	EUR	Germany
UK Biobank	2433	334766	2149	35912	SA	self-report: "Have you ever harmed yourself with the intent of ending your life?"	Mood Disorders	Mood Disorders	EUR	UK
Taiwan MDD	-	-	222	318	SA	Interview by well-trained interviewers using Chinese version of the Composite International Diagnostic Interview (CIDI) or Schedule of	MDD	MDD	EAS	Taiwan
Taiwan BIP	-	-	235	397	SA	Affective Disorder and Schizophrenia-Lifetime (SADS-L) or Hamilton Depression Rating Scale (HAM-D)	BIP	BIP	EAS	Taiwan
Taiwan SCZ	-	-	332	1004	SA		SCZ	SCZ	EAS	Taiwan
iPSYCH	7003	52227	-	-	6897 SA (98%), 106 death (2%)	ICD10. Diagnoses of suicide attempts (ICD-10: X60–X84), combinations of diagnoses where the main diagnosis had been recorded as a mental disorder (ICD-10: F chapter) together with a diagnosis of poisoning by drugs or other substances (ICD-10: T36–T50, T52–T60) or injuries to hand, wrist, and forearm (ICD-10: S51, S55, S59, S61, S65, S69).	SCZ, BIP, depression, autism, AN, ADHD	-	EUR	Denmark
Janssen	255	1684	-	-	SA	Clinical interview, medical record/chart review. Some C-SSRS	MDD, schizoaffective, SCZ, BIP	-	EUR	US
Genetic Investigation of Suicide and SA (GISS)	660	660	-	-	SA	Primary selection criteria: Nuclear family trios (all complete with both biological parents and one SA offspring per trio; n = 660), were collected in the Ukraine, by first recruitng the offspring from emergency care due to a severe SA as defined by a score of >=2 on a Medical Damage Rating Scale (MDS)*. Secondary measures of SA: Number of Attempts, Age at first attempt, Method, Intent, Family SB history	SCZ, BIP, MDD	-	EUR	Ukraine
Australian Genetics of Depression Study	2792	20193	2792	8718	SA	online self-report	MDD +/- other, history of at least 4 prescriptions of antidepressants in the last 5 years	MDD	EUR	Australia
Yale-Penn (EUR)	475	1817	-	-	SA	Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)	SUD	-	EUR	US
Yale-Penn (AA)	629	2902	-	-	SA	Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA)	SUD	-	AA	US
Columbia University	577	1233	-	-	260 SA (45%), 317 Death (55%)	Columbia Classification Algorithm for Suicide Assessment (C-CASA)	BIP, MDD, Psychotic disorders, other dx	-	EUR	US, Canada, Germany
University of Utah	4692	20702	-	-	Death	Coroner's report	MDD, BIP, Psychotic disorders, other dx	-	EUR	US
Japan	746	14049	-	-	Death	Coroner's report	-	-	EAS	Japan
Total	**29782**	**519961**	**14847**	**69951**						

SA - suicide attempt, MDD - major depressive disorder, BIP - bipolar disorder, SCZ - schizophrenia, ED - eating disorder, AN - anorexia nervosa, SUD - substance use disorder, PTSD, post-traumatic stress disorder, EAS - East Asian ancestry, AA - admixed African American ancest

*A direct GWAS of suicide attempt within psychiatric diagnosis was conducted (details in Supplementary Note). Samples in the GWAS of SA and SA within psychiatric diagnosis are not independent.

^Many cohorts were ascertained for psychiatric disorders and these columns indicate the disorder present. Where cohorts included cases with multiple psychiatric disorders, the GWAS of SA within psychiatric diagnosis was performed within one psychiatric disorder or w

Supplementary Table 2: Summary and results of polygenic risk scoring analyses with arrows showing direction from discovery GWAS to target cohort

Discovery GWAS		Target cohort and case-control definition	Ancestry of target cohort	Target cohort N cases	Target cohort N controls	R2 % (liability scale)^	P value	Discovery GWAS N cases	Discovery GWAS N controls
a) Test if PRS for SA are associated with SA versus controls									
Primary SA GWAS	→	PGC MDD SA versus controls	European	1,528	16,626	0.689	7.17E-15	28,254	503,335
Primary SA GWAS	→	PGC BIP SA versus controls	European	3,214	17,642	0.682	8.11E-28	26,568	502,319
Primary SA GWAS	→	PGC SCZSA versus controls	European	1,640	7,107	0.878	1.24E-17	28,142	512,849
Primary SA GWAS	→	Utah suicide versus controls	European	4,692	20,702	1.079	9.79E-81	25,090	499,259
Primary SA GWAS	→	CONVERGE SA versus controls	East Asian	1,148	6,515	0.251	3.06E-03	28,634	513,446
Primary SA GWAS	→	Yale-Penn SA versus controls	Admixed African American	629	2,902	0.207	5.28E-01	29,153	517,059
Primary SA GWAS	→	Grady Trauma Project SA versus controls	Admixed African American	669	4,473	0.579	3.44E-03	29,113	515,488
b) Test if PRS for SA are associated with SA within psychiatric diagnosis									
Primary SA GWAS	→	PGC MDD SA within psychiatric diagnosis*	European	1,677	8,793	0.434	5.83E-06	28,254	503,335
Primary SA GWAS	→	PGC BIP SA within psychiatric diagnosis*	European	3,214	5,408	0.806	2.33E-11	26,568	502,319
Primary SA GWAS	→	PGC SCZSA within psychiatric diagnosis*	European	1,668	2,928	0.709	5.78E-06	28,142	512,849
c) Test if PRS for SA conditioned on MDD are associated with SA in the general population									
SA-EUR|MDD	→	PGC MDD SA versus controls	European	1,528	16,626	0.062	1.96E-02	28,254	503,335
SA-EUR|MDD	→	PGC BIP SA versus controls	European	3,214	17,642	0.214	9.03E-10	26,568	502,319
SA-EUR|MDD	→	PGC SCZSA versus controls	European	1,640	7,107	0.214	2.42E-05	28,142	512,849
d) Test if PRS for SA conditioned on MDD are associated with SA within psychiatric diagnosis									
SA-EUR|MDD	→	PGC MDD SA within psychiatric diagnosis*	European	1,677	8,793	0.323	9.37E-05	28,254	503,335
SA-EUR|MDD	→	PGC BIP SA within psychiatric diagnosis*	European	3,214	5,408	0.665	1.30E-09	26,568	502,319
SA-EUR|MDD	→	PGC SCZSA within psychiatric diagnosis*	European	1,668	2,928	0.459	2.67E-04	28,142	512,849

Bonferroni corrected significance threshold is 0.05/16 = P <3.12E-03.

PRS - polygenic risk scores, SA - suicide attempt, GWAS - genome-wide association study, PGC - Psychiatric Genomics Consortium, MDD - major depresisve disorder, BIP - bipolar disorder, SCZ - schizophrenia.

SA-EUR|MDD - European-only suicide attempt meta-anslysis conditioned on major depressive disorder.

^For coversion to the liability scale, the prevalence of SA in the general population was assumed to be 2%, the prevalence of suicide in the general population was assumed to be 0.2% and the prevalence of SA in MDD, BIP and SCZ was assumed to be 16%, 37% and 36% respectively, which represent the observed prevalence of SA in these disorders in the PGC data.

*Comparison of SA cases with a psychiatric disorder versus individuals with the same psychiatric disorder without a history of SA.

Supplementary Table 3: Results of the primary meta-analysis of suicide attempt (trans-ancestry) for loci with P <5x10^{-6}, showing the most significant SNP from each genomic region

SNP	CHR	BP	P	OR (A1 allele)	SE	A1	A2	A1 freq in cases	A1 freq in controls	INFO	Neff	I^2 (%)^	I^2 df	I^2 P	Genes	N.genes.6.50kb
rs62474683	**7**	**115020725**	**1.91E-10**	**1.061**	**0.009**	**A**	**G**	**0.524**	**0.503**	**0.978**	**101456**	**0**	**17**	**0.49**	**-**	**0**
rs71557378	**6**	**26903585**	**1.98E-08**	**1.098**	**0.017**	**T**	**G**	**0.912**	**0.897**	**0.986**	**92411**	**46**	**14**	**0.03**	***SCGN,HIST1H2AA,HIST1H2BA,HIST1H2AP51,SLC17A4,SLC17A1,SLC17A2,SLC17A3,SLC17A12,TRIM38,HIST1H1J***	**138**
rs79950948	15	47644573	1.19E-07	0.935	0.013	T	C	0.822	0.834	0.993	94442	0	15	0.49	*SEMA6D*	1
rs4801131	18	52752700	1.42E-07	0.952	0.009	T	C	0.403	0.412	0.991	101456	49	17	0.01	*LINC01929*	1
rs112149665	18	50738489	1.58E-07	0.882	0.024	A	T	0.044	0.050	0.943	86782	51	13	0.01	*DCC,MIR4528*	2
rs3791129	1	44480093	2.32E-07	0.945	0.011	A	G	0.720	0.723	0.911	92731	0	13	0.55	*ST3GAL3,MIR6079,ARTN,IPO13,DPH2,ATP6V0B,B4GALT2,CCDC24,SLC6A9*	9
rs17133409	7	3515437	3.58E-07	1.069	0.013	A	G	0.862	0.851	0.999	101180	0	17	0.53	*SDK1*	1
rs2873081	3	56029271	3.69E-07	0.953	0.010	T	G	0.376	0.369	0.990	101456	0	17	0.93	*ERC2*	1
rs80119547	21	35226786	4.24E-07	1.111	0.021	A	G	0.063	0.061	0.940	88578	1	13	0.44	*ITSN1,ATP5O*	2
rs58476672	16	15875922	4.53E-07	1.068	0.013	A	G	0.153	0.154	0.952	10105?	0	17	0.68	*NDE1,MYH11*	2
rs2503185	1	66461401	4.91E-07	1.048	0.009	A	G	0.523	0.507	0.990	101456	35	17	0.07	*PDE4B,LOC101927139*	2
rs1864896	3	71469845	5.23E-07	1.049	0.010	A	G	0.488	0.460	0.988	98623	0	16	0.67	*FOXP1*	1
rs9525166	13	96897256	5.23E-07	0.955	0.009	A	T	0.530	0.547	0.984	101456	16	17	0.26	*DIAPC3,UGGT2,HS6ST3*	3
rs12641060	4	15258995	6.27E-07	0.938	0.013	T	G	0.175	0.175	0.941	96186	0	15	0.71	*LOC101929095,C1QTNF7*	2
rs4851921	2	106719833	7.70E-07	1.057	0.011	A	G	0.310	0.299	0.847	96573	0	14	0.55	*C2orf40,UXS1*	2
rs9310709	3	23093574	9.86E-07	0.955	0.009	T	C	0.538	0.552	0.994	101456	0	17	0.70	-	0
rs8072008	17	27407775	1.10E-06	1.050	0.010	C	G	0.372	0.352	0.980	98623	0	16	0.54	*PIPOX,MYO18A,TIAF1,CRYBA1,NUFIP2*	5
rs958361	16	78922488	1.20E-06	1.046	0.009	A	G	0.473	0.463	0.978	101456	0	17	0.92	*WWOX*	1
rs72929066	2	201884260	1.43E-06	1.062	0.013	A	G	0.183	0.179	0.903	99019	11	16	0.32	*LOC101927795,BZW1,CLK1,PPIL3,NIF3L1,ORC2,FAM126B,NDUFB3,CFLAR,CFLAR-AS1,CASP10,CASP8,*	13
rs651350	18	53320738	1.45E-06	1.070	0.014	A	G	0.871	0.861	0.997	94199	0	15	0.86	*TCF4,LINC01415*	2
rs10502971	18	50887961	1.65E-06	1.047	0.010	A	G	0.438	0.412	1.000	101456	19	17	0.23	*DCC,MIR4528*	2
rs4643522	2	207884377	1.69E-06	0.957	0.009	A	G	0.462	0.460	0.988	101456	0	17	0.45	*KLF7,MIR2355*	2
rs60937558	1	21577356	1.70E-06	1.070	0.014	T	C	0.144	0.130	0.951	95800	11	15	0.33	*ECE1,LOC100506801*	2
rs58743764	3	125576909	1.74E-06	0.884	0.026	A	G	0.040	0.046	0.896	90499	0	13	0.62	*MIR5481J,LOC101927056,FAM86JP,ALG1L,ROPN1B,SLC41A3*	6
rs73158760	3	149899955	1.94E-06	1.087	0.018	A	G	0.919	0.912	0.985	92487	7	14	0.38	*LOC105374313*	1
rs12574238	11	133754831	2.10E-06	1.054	0.011	A	G	0.240	0.234	0.981	101373	0	17	0.83	*LOC646522,SPATA19,MIR4697HG,MIR4697,GSG9B*	5
rs2321160	13	33882066	2.12E-06	1.046	0.009	T	C	0.569	0.543	0.976	101456	32	17	0.09	*STARD13,STARD13-AS,LINC02344*	3
rs4800070	18	36551186	2.15E-06	0.954	0.010	A	G	0.341	0.341	0.977	101456	26	17	0.15	-	0
rs944545	9	137831265	2.21E-06	1.056	0.012	T	C	0.232	0.218	0.909	96483	0	15	0.51	*FCN1*	1
rs1012361	14	22615465	2.39E-06	0.948	0.011	T	C	0.288	0.296	0.790	96573	0	14	1.00	-	0
rs2321706	17	28653774	2.40E-06	0.957	0.009	C	G	0.395	0.409	0.987	101456	46	17	0.02	*SLC6A4,BLMH,TMIGD1,CPD*	4
rs7047671	9	83693947	2.53E-06	1.065	0.013	A	C	0.151	0.147	0.953	100914	0	17	0.91	-	0
rs2284000	22	37053338	2.82E-06	1.053	0.011	C	G	0.737	0.722	0.906	99295	13	16	0.30	*CACNG2,LOC105373021*	2
rs2071382	15	91428197	2.88E-06	0.956	0.010	T	C	0.450	0.458	0.954	101456	30	17	0.11	*FURIN,FES,MAN2A2,UNC45A,HDDC3*	5
rs359265	2	60456410	2.89E-06	1.048	0.010	A	G	0.636	0.617	0.958	98623	0	16	0.49	-	0
rs1338443	6	19258898	3.35E-06	1.049	0.010	A	G	0.300	0.290	0.958	101456	0	17	0.54	*LOC101928519*	1
rs28716259	22	49242858	3.39E-06	1.051	0.011	C	G	0.631	0.612	0.828	96573	0	14	0.86	*LINC01310*	1
rs17240756	15	66348983	3.46E-06	1.081	0.017	A	C	0.114	0.109	0.823	88995	8	12	0.37	*MEGF11,MIR4311*	2
rs10810002	9	13656950	3.60E-06	1.046	0.010	A	T	0.427	0.409	0.968	98623	16	16	0.27	-	0
rs75751409	21	40823433	3.92E-06	0.946	0.012	A	T	0.811	0.811	0.989	101278	32	17	0.10	*WRB,WRB-SH3BGR,LCA5L,SH3BGR,MIR6508*	5
rs2249656	10	129201649	3.99E-06	1.046	0.010	T	C	0.493	0.487	0.894	99295	0	16	0.53	*DOCK1*	1
rs6968152	7	127088290	4.03E-06	1.056	0.012	A	C	0.818	0.818	0.976	101456	5	17	0.40	*LOC100506682*	1
rs11629484	15	65019258	4.07E-06	1.077	0.010	A	G	0.132	0.125	0.805	89316	0	12	0.78	*TRIP4,ZNF609,OAZ2,RBPMS2,MIR1272*	5
rs669630	11	103084483	4.19E-06	1.048	0.010	A	C	0.276	0.273	0.966	101456	0	17	0.52	*DCUN1D5,DYNC2H1*	2
rs9704813	11	46333263	4.30E-06	1.069	0.015	C	G	0.135	0.124	0.971	97961	0	16	0.64	*CREB3L1,DGKZ,MIR4688,MDK,CHRM4,AMBRA1,MIR3160-1,MIR3160-2,HARBI1,ATG13,ARHGAP1,ZNF*	16
rs9739327	12	62458724	4.56E-06	1.058	0.012	A	G	0.173	0.163	0.976	101057	19	17	0.23	*FAM19A2*	1
rs4538727	6	129408564	4.63E-06	0.956	0.010	A	G	0.403	0.416	0.944	96565	42	15	0.04	*LAMA2*	1
rs12025314	1	208783606	4.75E-06	1.088	0.018	A	G	0.928	0.917	0.947	96025	43	15	0.03	-	0
rs73521837	1	154252860	4.77E-06	1.052	0.011	T	C	0.773	0.764	0.984	101373	0	17	0.85	*C1orf43,UBAP2L,SNORA58B,HAX1,AQP10,ATP8B2*	6
rs5277255	22	27667235	4.82E-06	0.907	0.021	T	C	0.065	0.068	0.773	94191	0	14	0.75	*LINC01638,LINC02554*	2
rs60402381	13	103343946	4.89E-06	1.100	0.021	T	C	0.937	0.935	0.842	93377	0	15	0.86	*TPP2,METTL21C,CCDC168*	3

SNP, single nucleotide polymorphism; CHR, chromosome; BP, basepair position; OR, odds ratio; SE, standard error; Freq, frequency (average weighted by number of case:or controls per cohort); INFO, imputation information score (average weighted by effective sample size per cohort); Neff, the effective sample size;
I^2 (%), I^2 heterogeneity index; I^2 df, I^2 degrees of freedom; I^2 P, P value for I^{2}heterogeneity index; range, left, r2=0.1, span.r2=0.1(kb), left and right margins and span of region defined based on linkage disequilibrium r2 0.1;
^The I^2 heterogeneity index is a percentage measure of total variation across studies due to heterogeneity beyond chance.
All SNPs have a weighted minor allele frequency of >= 1%, weighted imputation INFO score >= 0.6 or SNPs present in >= 80% of total effective sample size.

Supplementary Table 4: Comparison of results for index SNPs at genome-wide significant loci for suicide attempt across analyses

					Primary SA GWAS (trans-ancestry)					SA-EUR				SA-EUR | MDD			SA in Million Veteran Program^			SA within psychiatric diagnosis*			
SNP	CHR	BP	A1	A2	OR(A1 allele)	SE	P	A1 freq in cases	A1 freq in controls	OR (A1 allele)	SE	P	A1 freq in cases	A1 freq in controls	OR (A1 allele)	SE	P	OR (A1 allele)	SE	P	A1 freq in cases	A1 freq in controls
rs62474683	7	115020725	A	G	1.061	0.009	1.91E-10	0.524	0.503	1.065	0.010	8.32E-11	0.516	0.500	1.058	0.010	1.34E-08	1.039	0.014	6.27E-03	1.042	0.016	7.70E-03	0.503	0.501
rs71557378	6	26903585	T	G	1.098	0.017	1.98E-08	0.912	0.897	1.102	0.017	1.04E-08	0.911	0.896	1.071	0.017	6.78E-05	1.032	0.028	2.57E-01	1.033	0.030	2.81E-01	0.914	0.900

SNP, single nucleotide polymorphism; CHR, chromosome; BP, basepair position; OR, odds ratio; SE, standard error; freq, frequency (average weighted by number of cases or controls per cohort); SA-EUR, European-only meta-analysis of suicide attempt; SA-EUR | MDD, European-only meta-analysis of suicide attempt conditioned on major depressive disorder; *SA within psychiatric diagnosis, a direct GWAS of suicide attempt within psychiatric diagnosis was conducted for comparison with SA_EUR | MDD (details in Supplementary Note)
^Independent replication cohort of 14,089 SA cases versus 395,359 controls (details in Supplementary Note).

Supplementary Table 5: Association of rs62474683 (index SNP for suicide attempt on chromosome 7) with phenotypes from the UK Biobank and other GWAS summary statistics at nominal significance (P < 0.05)

Results are from Open Targets Genetics web portal: https://genetics.opentargets.org/
* Beta and Odds Ratio are for the A allele at rs62474683.

Study ID	Trait	Trait Category	P value	Beta*	Odds Ratio*	PMID	Author	N Cases	N Overall
NEALE2_1239	Current tobacco smoking	Biological process	5.05E-12	0.009			UKB Neale v2		360797
NEALE2_20116_2	Current | smoking status	Biological process	9.75E-12	0.053	1.054		UKB Neale v2	37088	359706
NEALE2_2149	Lifetime number of sexual partners	Uncategorised	1.37E-08	0.011			UKB Neale v2		296609
NEALE2_3062_raw	Forced vital capacity (fvc)	Pulmonary function measurement	1.84E-08	-0.010			UKB Neale v2		329404
NEALE2_709	Number in household	Uncategorised	7.86E-08	-0.008			UKB Neale v2		358963
NEALE2_20151_raw	Forced vital capacity (fvc), best measure	Pulmonary function measurement	1.97E-07	-0.010			UKB Neale v2		272338
NEALE2_2159	Ever had same-sex intercourse	Uncategorised	2.18E-07	0.071	1.073		UKB Neale v2	11109	326849
NEALE2_399_raw	Number of incorrect matches in round	Cognitive function measurement	5.58E-07	-0.021			UKB Neale v2		360686
NEALE2_20075_raw	Home location at assesment - north co-ordinate (rounded)	Uncategorised	2.20E-06	-1452.440			UKB Neale v2		357793
NEALE2_2090	Seen doctor (gp) for nerves, anxiety, tension or depression	Uncategorised	2.29E-06	0.023	1.023		UKB Neale v2	123528	358693
NEALE2_2110	Able to confide	Uncategorised	2.46E-06	-0.021			UKB Neale v2		350618
NEALE2_2040	Risk taking	Mental or behavioural disorder biomarker	2.96E-06	0.025	1.025		UKB Neale v2	90500	348549
NEALE2_6138_1	College or university degree | qualifications	Self reported educational attainment	4.99E-06	0.023	1.023		UKB Neale v2	115981	357549
SAIGE_318	Tobacco use disorder	Nervous system	6.07E-06	0.048	1.050		UKB SAIGE	19780	399135
NEALE2_3063_raw	Forced expiratory volume in 1-second (fev1)	Pulmonary function measurement	7.08E-06	-0.006			UKB Neale v2		329404
NEALE2_1458	Cereal intake	Diet measurement	9.57E-06	-0.009			UKB Neale v2		345019
NEALE2_1787	Maternal smoking around birth	Uncategorised	1.14E-05	0.024	1.024		UKB Neale v2	95182	309942
NEALE2_22704_raw	Home location - north co-ordinate (rounded)	Uncategorised	1.30E-05	-1340.480			UKB Neale v2		361187
NEALE2_1980	Worrier / anxious feelings	Nervous system	1.86E-05	-0.020	0.980		UKB Neale v2	152370	351833
NEALE2_1418_1	Full cream | milk type used	Diet measurement	2.74E-05	0.040	1.041		UKB Neale v2	22902	360806
NEALE2_2335	Chest pain or discomfort	Uncategorised	2.79E-05	0.027	1.028		UKB Neale v2	56233	357507
NEALE2_20453	Ever taken cannabis	Biological process	3.21E-05	0.016			UKB Neale v2		117911
NEALE2_20489	Felt loved as a child	Uncategorised	4.04E-05	-0.016			UKB Neale v2		117624
NEALE2_20116_0	Never | smoking status	Biological process	4.37E-05	-0.019	0.981		UKB Neale v2	164638	359706
NEALE2_2405	Number of children fathered	Reproductive behaviour measurement	6.46E-05	-0.011			UKB Neale v2		165492
NEALE2_728	Number of vehicles in household	Uncategorised	6.68E-05	-0.008			UKB Neale v2		358754
NEALE2_1697	Comparative height size at age 10	Anthropometric measurement	1.04E-04	-0.006			UKB Neale v2		355331
NEALE2_1170	Getting up in morning	Uncategorised	1.13E-04	0.007			UKB Neale v2		360231
SAIGE_386_3	Labyrinthitis	Nervous system	1.17E-04	0.198	1.219		UKB SAIGE	767	403594
NEALE2_2267	Use of sun/uv protection	Uncategorised	1.33E-04	-0.008			UKB Neale v2		358559
GCST005902	Depression (broad)	Nervous system	1.41E-04	0.004	1.004	PMID:29662059	Howard DM	113769	322580
NEALE2_2188	Long-standing illness, disability or infirmity	Disease	1.43E-04	0.019	1.019		UKB Neale v2	114798	352798
NEALE2_2237	Plays computer games	Uncategorised	1.48E-04	0.004			UKB Neale v2		360817
SAIGE_306	Other mental disorder	Psychiatric disorder	1.66E-04	0.034	1.035		UKB SAIGE	28791	394267
SAIGE_571	Chronic liver disease and cirrhosis	Endocrine system	1.96E-04	0.099	1.104		UKB SAIGE	2895	402950
NEALE2_1707_3	Use both right and left hands equally | handedness (chirality/lateralit	Nervous system	2.73E-04	0.067	1.069		UKB Neale v2	6086	360913
NEALE2_30270_raw	Mean sphered cell volume	Uncategorised	2.84E-04	0.046			UKB Neale v2		344729
NEALE2_20150_raw	Forced expiratory volume in 1-second (fev1), best measure	Pulmonary function measurement	2.89E-04	-0.006			UKB Neale v2		272338
NEALE2_20488	Physically abused by family as a child	Uncategorised	3.17E-04	0.010			UKB Neale v2		117838
GCST006620	Self-rated health	Uncategorised	3.36E-04	0.015		PMID:27864402	Harris SE		111749
NEALE2_6139_2	A gas fire that you use regularly in winter time | gas or solid-fuel cook	Uncategorised	4.59E-04	-0.017	0.983		UKB Neale v2	145315	360066
GCST003497	Memory performance	Mental or behavioural disorder biomarker	4.85E-04	-0.015		PMID:27046643	Davies G		112067
NEALE2_189_raw	Townsend deprivation index at recruitment	Uncategorised	4.89E-04	0.024			UKB Neale v2		360763
NEALE2_2316	Wheeze or whistling in the chest in last year	Phenotype	6.24E-04	0.020	1.020		UKB Neale v2	73828	354523
SAIGE_480	Pneumonia	Respiratory system	6.65E-04	0.049	1.050		UKB SAIGE	10059	408597
NEALE2_6146_2	Disability living allowance | attendance/disability/mobility allowan	Uncategorised	6.94E-04	0.039	1.040		UKB Neale v2	15697	358597
NEALE2_2247_1	Yes | hearing difficulty/problems	Nervous system	7.40E-04	0.018	1.018		UKB Neale v2	90710	346635
NEALE2_2247_0	No | hearing difficulty/problems	Diet measurement	7.60E-04	-0.018	0.982		UKB Neale v2	90797	346635
SAIGE_474_2	Chronic tonsillitis and adenoiditis	Uncategorised	7.68E-04	0.142	1.153		UKB SAIGE	1126	391171
NEALE2_20003_1140909	Cod liver oil capsule | treatment/medication code	Drug use measurement	7.69E-04	-0.034	0.967		UKB Neale v2	21034	361141
NEALE2_1448_2	Brown | bread type	Diet measurement	8.01E-04	-0.025	0.976		UKB Neale v2	41518	348424
NEALE2_2178	Overall health rating	Uncategorised	8.51E-04	0.006			UKB Neale v2		359681
SAIGE_418	Nonspecific chest pain	Uncategorised	8.70E-04	0.029	1.029		UKB SAIGE	31429	408961
NEALE2_30000_raw	White blood cell (leukocyte) count	Hematological measurement	8.89E-04	0.016			UKB Neale v2		350470
NEALE2_20531	Victim of sexual assault	Uncategorised	1.01E-03	0.038	1.038		UKB Neale v2	17230	116671
NEALE2_2844	Had other major operations	Uncategorised	1.06E-03	0.022	1.023		UKB Neale v2	63035	192470
GCST005904	Major depressive disorder (probable)	Nervous system	1.08E-03	0.004	1.004	PMID:29662059	Howard DM	30603	174519
NEALE2_102360	Sweet biscuits intake	Diet measurement	1.10E-03	0.034			UKB Neale v2		51427
NEALE2_2257	Hearing difficulty/problems with background noise	Uncategorised	1.15E-03	0.016	1.016		UKB Neale v2	134141	353983
NEALE2_6020_31	Participant wanted to stop early | completion status of test	Uncategorised	1.16E-03	-0.201	0.818		UKB Neale v2	528	54007
NEALE2_30140_raw	Neutrophill count	Hematological measurement	1.17E-03	0.011			UKB Neale v2		349856
NEALE2_20487	Felt hated by family member as a child	Uncategorised	1.21E-03	0.010			UKB Neale v2		117749
NEALE2_3849	Number of pregnancy terminations	Disease	1.32E-03	0.011			UKB Neale v2		60099
NEALE2_20479	Ever thought that life not worth living	Uncategorised	1.40E-03	0.010			UKB Neale v2		117291
SAIGE_381_1	Otitis media	Nervous system	1.48E-03	0.106	1.112		UKB SAIGE	1824	406712
NEALE2_20528	Diagnosed with life-threatening illness	Uncategorised	1.53E-03	0.035	1.036		UKB Neale v2	19291	117617
SAIGE_427_42	Cardiac arrest	Cardiovascular	1.59E-03	0.147	1.158		UKB SAIGE	927	381846
NEALE2_2473	Other serious medical condition/disability diagnosed by doctor	Uncategorised	1.68E-03	0.018	1.019		UKB Neale v2	72965	354588
NEALE2_1080	Time spent using computer	Uncategorised	1.73E-03	0.006			UKB Neale v2		280750
NEALE2_20126_0	No bipolar or depression | bipolar and major depression status	Nervous system	1.75E-03	-0.033	0.967		UKB Neale v2	24070	86895
SAIGE_575_9	Nonspecific abnormal findings on radiological and other examinatio	Uncategorised	1.78E-03	0.237	1.267		UKB SAIGE	349	391656
NEALE2_5135_raw	3mm strong meridian (left)	Eye measurement	1.85E-03	0.025			UKB Neale v2		75398
NEALE2_2506_114	Never smoked | tobacco smoking	Biological process	1.86E-03	-0.029	0.971		UKB Neale v2	36833	91353
NEALE2_6147_1	For short-sightedness, i.e. only or mainly for distance viewing such a	Uncategorised	1.86E-03	0.027	1.027		UKB Neale v2	29318	360677
NEALE2_6151_3	Hip | fractured bone site(s)	Musculoskeletal system	1.88E-03	0.171	1.186		UKB Neale v2	667	359241
SAIGE_721	Spondylosis and allied disorders	Immune system	2.03E-03	0.050	1.051		UKB SAIGE	7930	399847
NEALE2_6179_100	None of the above | mineral and other dietary supplements	Uncategorised	2.07E-03	0.014	1.014		UKB Neale v2	155206	360016
NEALE2_6179_1	Fish oil (including cod liver oil) | mineral and other dietary supplem	Uncategorised	2.19E-03	-0.015	0.985		UKB Neale v2	114131	360016
NEALE2_4620	Number of depression episodes	Mental or behavioural disorder biomarker	2.21E-03	0.018			UKB Neale v2		45695
SAIGE_427	Cardiac dysrhythmias	Cardiovascular	2.23E-03	0.029	1.030		UKB SAIGE	24681	405600
NEALE2_20551_1	A sedative, benzodiazepine or sleeping tablet | substance of prescript	Uncategorised	2.26E-03	-0.207	0.813		UKB Neale v2	439	116746
NEALE2_670_1	Type of accommodation lived in: A house or bungalow	Uncategorised	2.39E-03	-0.025	0.976		UKB Neale v2	33234	360088
NEALE2_20117_2	Current | alcohol drinker status	Biological process	2.43E-03	-0.029	0.972		UKB Neale v2	23807	360726
SAIGE_443	Peripheral vascular disease	Cardiovascular	2.46E-03	0.069	1.072		UKB SAIGE	3927	404522
NEALE2_20002_1421	Meniere's disease | non-cancer illness code, self-reported	Nervous system	2.49E-03	0.134	1.143		UKB Neale v2	1031	361141
NEALE2_4631	Ever unenthusiastic/disinterested for a whole week	Mental or behavioural disorder biomarker	2.51E-03	0.026	1.026		UKB Neale v2	42374	115145
SAIGE_513_3	Hypoventilation	Uncategorised	2.52E-03	0.474	1.606		UKB SAIGE	82	408432
NEALE2_400_raw	Time to complete round	Uncategorised	2.54E-03	-0.759			UKB Neale v2		354739
NEALE2_5325_4	Both eyes | ever had refractive laser eye surgery	Uncategorised	2.57E-03	0.129	1.138		UKB Neale v2	1258	5719
NEALE2_104450	Apple intake	Diet measurement	2.57E-03	-0.020			UKB Neale v2		51427
NEALE2_6179_2	Glucosamine | mineral and other dietary supplements	Uncategorised	2.62E-03	-0.018	0.983		UKB Neale v2	70218	360016
GCST003724	Bipolar disorder	Nervous system	2.66E-03	0.069	1.071	PMID:27329760	Hou L	7647	34950
NEALE2_2000	Worry too long after embarrassment	Mental or behavioural disorder biomarker	2.69E-03	-0.014	0.986		UKB Neale v2	165310	346527
NEALE2_3064_raw	Peak expiratory flow (pef)	Pulmonary function measurement	2.71E-03	-0.758			UKB Neale v2		329404
NEALE2_699_raw	Length of time at current address	Uncategorised	2.71E-03	-0.077			UKB Neale v2		352690
NEALE2_20003_1140888	Fluvastatin | treatment/medication code	Uncategorised	2.81E-03	0.321	1.378		UKB Neale v2	174	361141
NEALE2_20002_1291	Mania/bipolar disorder/manic depression | non-cancer illness code,	Nervous system	2.91E-03	0.133	1.142		UKB Neale v2	1008	361141
NEALE2_20126_4	Probable recurrent major depression (moderate) | bipolar and major	Nervous system	3.03E-03	0.043	1.044		UKB Neale v2	10902	86895
NEALE2_5134_raw	6mm strong meridian (left)	Eye measurement	3.08E-03	0.025			UKB Neale v2		65551
NEALE2_22609_1	Sometimes | workplace very dusty	Uncategorised	3.11E-03	0.029	1.029		UKB Neale v2	31586	89965
NEALE2_6145_1	Serious illness, injury or assault to yourself | illness, injury, bereavem	Disease	3.21E-03	0.024	1.024		UKB Neale v2	33241	358836
NEALE2_5099_raw	3mm weak meridian (right)	Eye measurement	3.25E-03	0.022			UKB Neale v2		75410
NEALE2_20003_1141182	Tiotropium | treatment/medication code	Drug use measurement	3.33E-03	0.137	1.147		UKB Neale v2	922	361141
NEALE2_6164_1	Walking for pleasure (not as a means of transport) | types of physical	Uncategorised	3.37E-03	-0.015	0.985		UKB Neale v2	100689	359263
SAIGE_375_2	Epiphora	Uncategorised	3.57E-03	-0.138	0.871		UKB SAIGE	899	402144
SAIGE_724	Other and unspecified disorders of back	Uncategorised	3.72E-03	0.091	1.095		UKB SAIGE	2077	393994
SAIGE_870_3	Other open wound of head and face	Phenotype	3.79E-03	0.076	1.079		UKB SAIGE	2919	403345
NEALE2_20002_1598	Tonsilitis | non-cancer illness code, self-reported	Uncategorised	3.84E-03	0.081	1.084		UKB Neale v2	2569	361141
NEALE2_20002_1567	Infectious mononucleosis / glandular fever / epstein barr virus (ebv) |	Disease	4.00E-03	0.189	1.209		UKB Neale v2	464	361141
NEALE2_5098_raw	6mm weak meridian (right)	Eye measurement	4.08E-03	0.024			UKB Neale v2		66256
NEALE2_5237	3mm index of best keratometry results (right)	Uncategorised	4.15E-03	-0.010			UKB Neale v2		75410
NEALE2_5133_raw	6mm strong meridian (right)	Eye measurement	4.16E-03	0.024			UKB Neale v2		66256
SAIGE_960	Poisoning by antibiotics	Disease	4.30E-03	0.031	1.032		UKB SAIGE	18430	400227
NEALE2_2100	Seen a psychiatrist for nerves, anxiety, tension or depression	Uncategorised	4.42E-03	0.021	1.021		UKB Neale v2	41233	359535
NEALE2_6146_100	None of the above | attendance/disability/mobility allowance	Uncategorised	4.52E-03	-0.029	0.972		UKB Neale v2	20718	358597
NEALE2_20421	Ever felt worried, tense, or anxious for most of a month or longer	Nervous system	4.57E-03	0.027	1.028		UKB Neale v2	29351	110315
SAIGE_427_4	Cardiac arrest and ventricular fibrillation	Cardiovascular	4.64E-03	0.119	1.126		UKB SAIGE	1137	382056
NEALE2_20003_1140874	Prednisolone | treatment/medication code	Uncategorised	4.72E-03	0.087	1.091		UKB Neale v2	2103	361141
SAIGE_070	Viral hepatitis	Endocrine system	4.77E-03	0.115	1.122		UKB SAIGE	1215	404531
NEALE2_5325_0	No | ever had refractive laser eye surgery	Diet measurement	5.16E-03	-0.111	0.895		UKB Neale v2	1634	5719
SAIGE_496_21	Obstructive chronic bronchitis	Respiratory system	5.20E-03	0.077	1.080		UKB SAIGE	2698	378203
NEALE2_1950	Sensitivity / hurt feelings	Mental or behavioural disorder biomarker	5.28E-03	-0.013	0.987		UKB Neale v2	156066	350821
SAIGE_735_2	Acquired toe deformities	Uncategorised	5.34E-03	0.056	1.058		UKB SAIGE	5144	400058
NEALE2_22506_111	Smokes on most or all days | tobacco smoking	Biological process	5.38E-03	0.086	1.090		UKB Neale v2	2124	91353
NEALE2_20002_1113	Emphysema/chronic bronchitis | non-cancer illness code, self-report	Respiratory system	5.41E-03	0.056	1.057		UKB Neale v2	5031	361141
NEALE2_20084_472	Fish oil | vitamin and/or mineral supplement use	Uncategorised	5.63E-03	-0.040	0.961		UKB Neale v2	12462	51427
NEALE2_4288_raw	Time to answer	Uncategorised	5.85E-03	-1.354			UKB Neale v2		119729
NEALE2_6159_4	Back pain | pain type(s) experienced in last month	Uncategorised	5.90E-03	0.015	1.015		UKB Neale v2	91349	360391
NEALE2_30040_raw	Mean corpuscular volume	Hematological measurement	5.98E-03	0.029			UKB Neale v2		350473
SAIGE_573	Other disorders of liver	Endocrine system	6.34E-03	0.051	1.053		UKB SAIGE	5847	405902
SAIGE_741_4	Joint effusions	Musculoskeletal system	6.36E-03	0.102	1.107		UKB SAIGE	1425	404058
NEALE2_3404	Neck/shoulder pain for 3+ months	Uncategorised	6.64E-03	0.029	1.029		UKB Neale v2	25120	81276
NEALE2_20480	Ever self-harmed	Uncategorised	6.65E-03	0.055	1.056		UKB Neale v2	5099	117733
SAIGE_724_9	Other unspecified back disorders	Uncategorised	6.69E-03	0.096	1.101		UKB SAIGE	1617	393534
SAIGE_979	Adverse drug events and drug allergies	Disease	6.86E-03	0.146	1.157		UKB SAIGE	695	382492

Supplementary Table 6: Phenotypes from the GWAS catalogue which have genome-wide significant SNPs in high linkage disequilibrium with rs62474683 (index SNP for suicide attempt on chromosome 7)

Supplementary Table 6: Phenotypes from the GWAS catalogue which have genome-wide significant SNPs in high linkage disequilibrium with rs62474683 (index SNP for suicide attempt on chromosome 7)

Results are from Open Targets Genetics web portal: https://genetics.opentargets.org/

Trait	Lead Variant rsID	Effect allele*	Study ID	Lead Variant P-value	Beta*	Odds Ratio*	PMID	Author	Study N	LD (r^2)^
General risk tolerance (MTAG)	rs4377898	A	GCST007325	8.00E-31			PMID:30643258	Karlsson Linner R	975353	0.996
Adventurousness	rs10251192	C	GCST007324	8.00E-15	0.015		PMID:30643258	Karlsson Linner R	557923	0.857
Lifetime smoking index	rs2401924	G	GCST009096	3.00E-14			PMID:31689377	Wootton RE	462690	0.992
Smoking status	rs2401924	G	GCST007085	1.00E-12			PMID:30595370	Kichaev G	458000	0.992
Insomnia	rs12666306	A	GCST007988	2.00E-12		1.043	PMID:30804565	Jansen PR	1331010	0.943
Current tobacco smoking	rs201720501	T	NEALE2_1239	2.20E-12	0.009			UKB Neale v2	360797	0.988
Current	smoking status	rs201720501	T	NEALE2_20116_2	4.59E-12		1.055		UKB Neale v2	359706	0.988
Number of sexual partners	rs4377898	A	GCST007326	2.00E-11			PMID:30643258	Karlsson Linner R	370711	0.996
same-sex sexual behaviour	rs10261857	G	GCST008651	1.00E-10	0.003		PMID:31467194	Ganna A	421829	0.631
Self-reported math ability (MTAG) [MTAG]	rs10267100	T	GCST006569	1.00E-10	0.012		PMID:30038396	Lee JJ	670471	0.568
Highest math class taken (MTAG) [MTAG]	rs2106525	G	GCST006568	2.00E-10			PMID:30038396	Lee JJ	811539	0.848
Ever had same-sex intercourse	rs2188541	A	NEALE2_2159	1.16E-09		1.090		UKB Neale v2	326849	0.628
Self-reported math ability	rs10267100	T	GCST006573	3.00E-09	0.012		PMID:30038396	Lee JJ	564698	0.568
Forced vital capacity (fvc), best measure	rs10231331	A	NEALE2_20151_raw	3.21E-09	0.011			UKB Neale v2	272338	
Number of incorrect matches in round	rs62477307	G	NEALE2_399_raw	3.92E-09	-0.027			UKB Neale v2	360686	
Smoking cessation (MTAG)	rs6968125	C	GCST007464	4.00E-09	0.009		PMID:30643251	Liu M	820192	0.623
College or university degree	qualifications	rs10239094	G	NEALE2_6138_1	4.06E-09		1.030		UKB Neale v2	357549	0.856
Number in household	rs62477308	C	NEALE2_709	4.60E-09	-0.009			UKB Neale v2	358963	0.624
Forced vital capacity (fvc)	rs9641536	A	NEALE2_3062_raw	8.27E-09	-0.011			UKB Neale v2	329404	0.980
Risk-taking tendency (4-domain principal component model)	rs2401924	G	GCST007323	2.00E-08			PMID:30643258	Karlsson Linner R	315894	0.992
Cigarettes smoked per day (MTAG)	rs10261857	G	GCST007463	3.00E-08	0.013		PMID:30643251	Liu M	403928	0.631
Insomnia [Female]	rs73201933	T	GCST007988_2	3.00E-08			PMID:30804565	Jansen PR	1331010	0.952

*Effect allele is in LD with the A allele at rs62474683, which is the risk-increasing allele for suicide attempt. Beta and Odds Ratio correspond to the effect allele.
^Linkage disequilibrium r^2 between the lead variant and the index SNP for suicide attempt rs62474683.

Supplementary Table 7: Pairwise GWAS results for genomic region containing genome-wide significant locus for suicide attempt on chromosome 7

Pairwise GWAS uses association statistics from two GWAS to estimate the probability that a genomic region contains a genetic variant that influences only trait 1, only trait 2, both traits (shared causal or pleiotropic variant) or contains two independent variants in the same region, one influencing trait 1 and the other influencing trait 2.

Trait 1	Trait 2	Posterior Probabilities			
		Only trait 1	Only trait 2	Both traits	Two variants in region
SA-EUR	Risk-taking behavior	0.000	0.000	1.000	0.000
SA-EUR	Lifetime Smoking Index	0.000	0.000	0.997	0.003

SA-EUR, European-only meta-analysis of suicide attempt. Results are based on the genomic region containing the index SNP for suicide attempt (rs62474683) on chromosome 7, ranging from 114501142-116780046 base pairs, based on hg19.

Supplementary Table 8: Results for genes with P <1.00E-04 from enrichment analysis of primary suicide attempt meta-analysis results in 18,721 genes conducted using MAGMA

Ensembl ID	CHR	Start BP	Stop BP	N SNPS	P	Gene
ENSG00000184588	**1**	**66258197**	**66840259**	**1026**	**1.02E-07**	***PDE4B***
ENSG00000137872	**15**	**47476298**	**48066420**	**1257**	**2.33E-07**	***SEMA6D***
ENSG00000112763	**6**	**26458150**	**26476849**	**56**	**7.09E-07**	***BTN2A1***
ENSG00000187323	**18**	**49866542**	**51057784**	**3644**	**8.57E-07**	***DCC***
ENSG00000187398	**11**	**24518516**	**25104150**	**1892**	**1.23E-06**	***LUZP2***
ENSG00000221995	**17**	**27400537**	**27418537**	**34**	**2.00E-06**	***TIAF1***
ENSG00000219438	**22**	**48885272**	**49246724**	**1814**	**2.55E-06**	***FAM19A5***
ENSG00000198558	6	27840926	27841289	1	3.37E-06	*HIST1H4L*
ENSG00000149295	11	113280318	113346413	149	3.75E-06	*DRD2*
ENSG00000166862	22	36959968	37099603	249	3.98E-06	*CACNG2*
ENSG00000146555	7	3341080	4308632	2857	4.81E-06	*SDK1*
ENSG00000140564	15	91411822	91426688	24	5.19E-06	*FURIN*
ENSG00000187672	3	55542336	56502391	2132	5.75E-06	*ERC2*
ENSG00000182271	17	28643351	28661077	25	9.56E-06	*TMIGD1*
ENSG00000123836	1	207222801	207254369	33	1.35E-05	*PFKFB2*
ENSG00000102595	13	96453834	96705736	332	1.96E-05	*UGGT2*
ENSG00000114861	3	71003844	71633140	1005	2.00E-05	*FOXP1*
ENSG00000149292	11	113185251	113254266	112	3.15E-05	*TTC12*
ENSG00000184357	6	27834570	27835359	3	3.80E-05	*HIST1H1B*
ENSG00000185352	13	96743093	97485671	1342	3.96E-05	*HS6ST3*
ENSG00000157578	21	40777770	40817731	88	3.99E-05	*LCA5L*
ENSG00000143570	1	153931575	153940188	9	4.20E-05	*SLC39A1*
ENSG00000120658	13	43787654	44361044	1154	4.35E-05	*ENOX1*
ENSG00000117411	1	44444615	44456840	16	4.46E-05	*B4GALT2*
ENSG00000168131	6	27878963	27880174	4	4.58E-05	*OR2B2*
ENSG00000196535	17	27400528	27507430	179	4.87E-05	*MYO18A*
ENSG00000108576	17	28521337	28563020	45	4.92E-05	*SLC6A4*
ENSG00000186472	7	82383329	82792246	859	5.07E-05	*PCLO*
ENSG00000137692	11	102932805	102962944	39	5.34E-05	*DCUN1D5*
ENSG00000170624	5	155297354	156194799	1628	5.39E-05	*SGCD*
ENSG00000213719	6	31698358	31707540	5	5.55E-05	*CLIC1*
ENSG00000233822	6	27806323	27823487	19	5.55E-05	*HIST1H2BN*
ENSG00000106536	7	39017598	39532694	1267	5.65E-05	*POU6F2*
ENSG00000162374	1	50513686	50669458	136	5.79E-05	*ELAVL4*
ENSG00000196569	6	129204342	129837714	1224	6.53E-05	*LAMA2*
ENSG00000168792	17	27887565	27894155	5	7.35E-05	*ABHD15*
ENSG00000117407	1	44398992	44402913	8	7.52E-05	*ARTN*
ENSG00000187626	6	28212401	28227011	26	8.05E-05	*ZKSCAN4*
ENSG00000143578	1	153940010	153946839	9	8.13E-05	*CREB3L4*
ENSG00000256966	9	37512544	37592466	187	8.38E-05	*RP11-613M10.8*
ENSG00000196517	1	44457172	44497139	43	8.64E-05	*SLC6A9*
ENSG00000186470	6	26365387	26378546	84	8.78E-05	*BTN3A2*
ENSG00000147912	9	37510889	37588871	181	9.28E-05	*FBXO10*
ENSG00000166118	11	133710526	133715433	17	9.70E-05	*SPATA19*

CHR, chromosome; BP, basepair position; SNP, single nucletotide polymorphism

Genes with significant enrichment are shown in bold text. Bonferroni corrected significance threshold = 0.05/ 18517 = 2.70e-06

Supplementary Table 9: Results for gene-sets with P <1.00E-03 from enrichment analysis of primary suicide attempt meta-analysis results in 11,638 gene-sets conducted using MAGMA

Gene set	N genes in set	BETA	BETA_STD	SE	P
Curated_gene_sets:boyerinas_oncofetal_targets_of_let7a1	10	1.122	0.026	2.58E-01	7.09E-06
GO_bp:go_synaptic_vesicle_priming	17	0.773	0.023	1.97E-01	4.46E-05
GO_mf:go_c2h2_zinc_finger_domain_binding	12	0.791	0.020	2.07E-01	6.87E-05
GO_bp:go_phospholipid_catabolic_process	38	0.526	0.024	1.41E-01	9.33E-05
GO_bp:go_positive_regulation_of_glutamate_receptor_signaling_pathway	12	1.006	0.026	2.72E-01	1.12E-04
GO_bp:go_organophosphate_catabolic_process	122	0.285	0.023	7.73E-02	1.13E-04
Curated_gene_sets:lopez_mbd_targets	894	0.102	0.022	2.88E-02	1.90E-04
GO_bp:go_regulation_of_cellular_response_to_drug	23	0.644	0.023	1.81E-01	1.95E-04
Curated_gene_sets:park_apl_pathogenesis_up	13	0.816	0.022	2.32E-01	2.18E-04
GO_mf:go_phosphoric_ester_hydrolase_activity	346	0.162	0.022	4.66E-02	2.55E-04
GO_cc:go_exocytic_vesicle	200	0.210	0.022	6.04E-02	2.61E-04
Curated_gene_sets:spielman_lymphoblast_european_vs_asian_2fc_dn	18	0.799	0.025	2.32E-01	2.87E-04
GO_bp:go_synaptic_vesicle_cycle	180	0.218	0.021	6.41E-02	3.44E-04
GO_bp:go_chondrocyte_proliferation	15	0.749	0.021	2.22E-01	3.72E-04
GO_cc:go_beta_catenin_tcf_complex	11	0.865	0.021	2.58E-01	4.11E-04
Curated_gene_sets:reactome_foxo_mediated_transcription_of_oxidative_stress_metabolic_and_neuronal_genes	29	0.493	0.020	1.53E-01	6.48E-04
Curated_gene_sets:kim_wt1_targets_dn	441	0.132	0.020	4.12E-02	6.59E-04
GO_bp:go_mitotic_g2_dna_damage_checkpoint	19	0.582	0.019	1.83E-01	7.31E-04
GO_bp:go_calcium_ion_regulated_exocytosis	146	0.223	0.020	7.02E-02	7.32E-04
Curated_gene_sets:guillaumond_klf10_targets_up	46	0.369	0.018	1.16E-01	7.59E-04
GO_mf:go_translation_initiation_factor_binding	27	0.470	0.018	1.49E-01	7.97E-04
GO_bp:go_b_cell_differentiation	120	0.266	0.021	8.45E-02	8.20E-04
GO_bp:go_antimicrobial_humoral_immune_response_mediated_by_antimicrobial_peptide	60	0.359	0.020	1.15E-01	8.78E-04
Curated_gene_sets:hess_targets_of_hoxa9_and_meis1_up	66	0.296	0.018	9.47E-02	8.79E-04
Curated_gene_sets:caffarel_response_to_thc_24hr_5_dn	53	0.382	0.020	1.22E-01	8.88E-04
Curated_gene_sets:nuytten_ezh2_targets_up	980	0.089	0.020	2.87E-02	9.08E-04
Curated_gene_sets:kyng_environmental_stress_response_not_by_gamma_in_old	30	0.451	0.018	1.45E-01	9.11E-04
GO_bp:go_mitotic_g2_m_transition_checkpoint	29.000	0.491	0.019	1.58E-01	9.27E-04
Curated_gene_sets:reactome_gamma_carboxylation_transport_and_amino_terminal_cleavage_of_proteins	10	0.750	0.017	2.42E-01	9.67E-04

Gene-sets including < 10 genes were excluded. Bonferroni corrected significance threshold = 0.05/11,638 =4.30E-06

Supplementary Table 10: Results for tissue-set enrichment analysis of primary suicide attempt meta-analysis results in 54 GTEx (v8) tissues using MAGMA*

Tissue	BETA	BETA_STD	SE	P
Pituitary	0.022	0.042	0.009	7.57E-03
Pancreas	0.019	0.032	0.009	1.43E-02
Brain_Nucleus_accumbens_basal_gangli	0.017	0.029	0.008	1.71E-02
Brain_Putamen_basal_ganglia	0.017	0.028	0.008	2.17E-02
Brain_Caudate_basal_ganglia	0.016	0.028	0.008	2.33E-02
Brain_Frontal_Cortex_BA9	0.013	0.023	0.007	3.61E-02
Brain_Amygdala	0.014	0.025	0.008	3.71E-02
Brain_Anterior_cingulate_cortex_BA24	0.013	0.023	0.007	3.90E-02
Brain_Cortex	0.013	0.023	0.007	3.97E-02
Kidney_Cortex	0.016	0.028	0.009	4.19E-02
Liver	0.011	0.020	0.007	4.75E-02
Brain_Cerebellum	0.009	0.018	0.006	7.63E-02
Brain_Cerebellar_Hemisphere	0.009	0.017	0.006	8.29E-02
Brain_Hippocampus	0.011	0.019	0.008	8.56E-02
Brain_Hypothalamus	0.011	0.019	0.008	8.75E-02
Brain_Substantia_nigra	0.010	0.017	0.009	1.34E-01
Kidney_Medulla	0.009	0.016	0.010	1.78E-01
Adrenal_Gland	0.009	0.018	0.011	1.89E-01
Muscle_Skeletal	0.006	0.011	0.007	2.01E-01
Thyroid	0.007	0.014	0.010	2.33E-01
Stomach	0.007	0.012	0.012	2.93E-01
Heart_Atrial_Appendage	0.004	0.008	0.010	3.25E-01
Ovary	0.004	0.008	0.010	3.53E-01
Brain_Spinal_cord_cervical_c-1	0.002	0.003	0.009	4.25E-01
Whole_Blood	0.001	0.001	0.006	4.60E-01
Colon_Transverse	-0.002	-0.004	0.012	5.68E-01
Heart_Left_Ventricle	-0.002	-0.003	0.009	5.81E-01
Cells_EBV-transformed_lymphocytes	-0.001	-0.002	0.005	5.84E-01
Small_Intestine_Terminal_Ileum	-0.002	-0.004	0.010	5.91E-01
Breast_Mammary_Tissue	-0.008	-0.015	0.013	7.31E-01
Adipose_Visceral_Omentum	-0.007	-0.014	0.012	7.32E-01
Spleen	-0.005	-0.009	0.008	7.35E-01
Cells_Cultured_fibroblasts	-0.005	-0.010	0.007	7.52E-01
Fallopian_Tube	-0.010	-0.018	0.012	7.83E-01
Testis	-0.005	-0.008	0.006	7.96E-01
Minor_Salivary_Gland	-0.009	-0.017	0.010	8.30E-01
Uterus	-0.014	-0.028	0.011	8.87E-01
Cervix_Endocervix	-0.015	-0.030	0.012	8.98E-01
Esophagus_Mucosa	-0.010	-0.019	0.008	9.02E-01
Lung	-0.013	-0.024	0.009	9.07E-01
Adipose_Subcutaneous	-0.016	-0.031	0.011	9.20E-01
Cervix_Ectocervix	-0.019	-0.037	0.013	9.33E-01
Colon_Sigmoid	-0.021	-0.042	0.013	9.51E-01
Esophagus_Gastroesophageal_Junction	-0.026	-0.052	0.013	9.76E-01
Esophagus_Muscularis	-0.026	-0.051	0.013	9.78E-01
Prostate	-0.025	-0.048	0.012	9.79E-01
Skin_Not_Sun_Exposed_Suprapubic	-0.018	-0.034	0.008	9.87E-01
Vagina	-0.027	-0.051	0.011	9.91E-01
Skin_Sun_Exposed_Lower_leg	-0.020	-0.038	0.008	9.93E-01
Artery_Tibial	-0.028	-0.057	0.010	9.96E-01
Artery_Aorta	-0.028	-0.058	0.011	9.96E-01
Bladder	-0.037	-0.072	0.014	9.97E-01
Artery_Coronary	-0.034	-0.068	0.012	9.97E-01
Nerve_Tibial	-0.032	-0.065	0.010	9.99E-01

STD, standardized; SE, standard error. Bonferroni corrected significance threshold = P < 9.25E-04.

*Number of genes = 16,982

Supplementary Table 11: Results from TWAS FUSION analysis of primary suicide attempt meta-analysis results, using gene expression data from PsychENCODE, for genes with TWAS P value <1E-04

Ensembl ID	Gene	CHR	gene start position	gene end position	HSQ	BEST.GWAS.ID	BEST.GWAS.Z	EQTL.ID	EQTL.R2	EQTL.Z	EQTL.GWAS.Z	NSNP	NWGT	MODEL	MODELCV.R2	MODELCV.PV	TWAS.Z	TWAS.P*
ENSG00000255418	***RP11-266A24.1***	**11**	**23,200,000**	**23,200,000**	**0.033**	11:22897428	**-4.92**	11:23242142	**0.005**	**-5.340**	**-2.124**	**2494**	**2494**	**blup**	**0.012**	**3.84E-05**	**5.472**	**4.44E-08**
ENSG00000221995	***TIAF1***	**17**	**27,400,537**	**27,418,537**	**0.169**	17:27416860	**4.83**	17:27411828	**0.086**	**-11.280**	**4.797**	**878**	7	lasso	**0.103**	**2.16E-33**	**-4.826**	**1.39E-06**
ENSG00000108256	***NUP92***	**17**	**27,582,854**	**27,621,136**	**0.072**	17:27416860	**4.83**	17:27575621	**0.031**	**7.360**	**-3.126**	**939**	14	enet	**0.061**	**5.03E-20**	**-4.709**	**2.49E-06**
ENSG00000182240	***BACE2***	**21**	**42,539,728**	**42,654,445**	**0.037**	21:42094173	**3.32**	21:42520134	**0.001**	**-4.430**	**2.564**	**1949**	**1949**	**bslmm**	**0.005**	**0.00428**	**-4.707**	**2.51E-06**
ENSG00000187672	***ERC2***	**3**	**55,500,000**	**56,500,000**	**0.107**	3:56029271	**-5.08**	3:56028332	**0.022**	**6.160**	**3.308**	**3333**	10	lasso	**0.027**	**1.24E-09**	**4.706**	**2.52E-06**
ENSG00000140564	*FURIN*	15	91,400,000	91,400,000	0.113	15:91428589	-4.53	15:91428589	0.059	9.000	-4.526	1663	3	lasso	0.065	2.47E-21	-4.431	9.36E-06
ENSG00000180667	*YOD1*	1	207,000,000	207,000,000	0.036	1:207171603	4.36	1:207251518	0.031	-6.270	-4.055	1056	2	lasso	0.028	5.19E-10	4.244	2.20E-05
ENSG00000111358	*GTF2H3*	12	124,000,000	124,000,000	0.024	12:124108620	-4.4	12:124204969	0.006	-4.920	-2.840	1729	1729	blup	0.011	6.57E-05	4.231	2.33E-05
ENSG00000108582	*CPD*	17	28,705,923	28,797,007	0.125	17:28652914	-4.65	17:28743156	0.063	9.540	3.239	1259	1259	bslmm	0.070	1.09E-22	4.214	2.51E-05
ENSG00000133606	*MKRN1*	7	140,000,000	140,000,000	0.205	7:139960397	3.6	7:140147902	0.058	-9.310	1.710	1375	1375	bslmm	0.089	6.41E-29	-4.214	2.51E-05
ENSG00000170027	*YWHAG*	7	76,000,000	76,000,000	0.034	7:75819975	-4.11	7:75861673	0.006	-5.160	-4.047	1340	1340	blup	0.015	6.05E-06	4.195	2.73E-05
ENSG00000250131	*RP11-130F10.1*	4	178,000,000	179,000,000	0.136	4:178446844	-3.59	4:178424045	0.060	9.160	-2.519	2338	5	lasso	0.102	7.46E-33	-4.178	2.94E-05
ENSG00000167994	*RAB3IL1*	11	61,700,000	61,700,000	0.101	11:61686336	-4.19	11:61686336	0.056	-9.350	-4.186	1455	4	lasso	0.067	6.14E-22	4.169	3.06E-05
ENSG00000166780	*C16orf45*	16	15,528,152	15,718,885	0.158	16:15875922	5.04	16:15595412	0.032	-7.010	1.805	1253	52	enet	0.049	2.13E-16	-4.128	3.66E-05
ENSG00000187323	*DCC*	18	49,866,542	51,057,784	0.136	18:50531872	-4.85	18:50697120	0.036	8.240	4.407	4843	4843	blup	0.067	8.81E-22	4.088	4.35E-05
ENSG00000170396	*ZNF804A*	2	185,000,000	186,000,000	0.090	2:185798504	-4.22	2:185797159	0.021	-6.730	-4.178	2165	10	lasso	0.039	2.93E-13	4.057	4.98E-05
ENSG00000273026	*RP11-422F24.10*	1	154,000,000	154,000,000	0.032	1:154251514	-4.55	1:153943800	0.034	-7.340	3.915	984	6	lasso	0.034	9.74E-12	-3.977	6.97E-05
ENSG00000181798	*LINC00471*	2	232,000,000	232,000,000	0.344	2:232305077	-5.25	2:232377818	0.316	-20.740	3.215	1486	41	enet	0.387	1.39E-142	-3.951	7.80E-05
ENSG00000100304	*TTLL2*	22	43,562,628	43,583,139	0.443	22:43520258	3.97	22:43559451	0.094	11.470	-2.370	2007	84	enet	0.282	2.79E-97	-3.908	9.32E-05
ENSG00000137054	*PQLC1L*	9	37,500,000	37,500,000	0.128	9:37213641	-4.19	9:37490859	0.106	-12.100	2.256	1456	61	enet	0.135	9.50E-44	-3.896	9.78E-05

CHR, chromosome; HSQ, heritability of the gene's expression; BEST.GWAS.ID, ID of the most significant GWAS SNP in locus; BEST.GWAS.Z, Z-score of the most significant GWAS SNP in locus; EQTL.ID, ID of the best eQTL in the locus; EQTL.R2, cross-validation R2 of the best eQTL in the locus; EQTL.Z, Z-score of the best eQTL in the locus; EQTL.GWAS.Z, GWAS Z-score for this eQTL; NSNP, Number of SNP's in the locus; MODEL, Best performing model (lasso - least absolute shrinkage and selection operator, bslmm - bayesian sparse linear mixed model; enet - elastic net, blup - best linear unbiased prediction); MODELCV.R2, cross-validation R2 of the best performing model; MODELCV.PV, cross-validation P-value of the best performing model; TWAS.Z, TWAS Z-score; TWAS.P, TWAS P value;

*The Bonferroni corrected significance threshold is P <4.28E-6.

Supplementary Table 12: SNP-heritability estimates and genetic correlations between GWAS of suicide attempt

a) Genetic correlations (se) and SNP-heritability (se) on the diagonal. (All heritability estimates are presented on the liability scale, assuming a 2% prevalence of SA in the

GWAS	SA	SA-EUR	SA-EUR | MDD	SA within psychiatric diagnosis*	SA-EUR | MDD, BIP, SCZ
SA	0.068 (0.005)	1.08 (0.007)	1.06 (0.012)	0.93 (0.09)	1.00 (0.01)
SA -EUR		0.075 (0.006)	0.95 (0.007)	1.06 (0.12)	0.89 (0.01)
SA-EUR | MDD			0.041 (0.005)	1.13 (0.13)	0.93 (0.01)
SA within psychiatric diagnosis*				0.044 (0.01)	1.18 (0.14)
SA-EUR | MDD, BIP, SCZ					0.041 (0.005)

b) P value for genetic correlation or SNP-heritability

GWAS	SA	SA-EUR	SA-EUR | MDD	SA within psychiatric diagnosis*	SA-EUR | MDD, BIP, SCZ
SA	2.00E-42	0	0	5.35E-24	0
SA -EUR		3.00E-40	0	5.75E-19	0
SA-EUR | MDD			1.20E-16	1.78E-18	0
SA within psychiatric diagnosis*				5.41E-06	1.75E-17
SA-EUR | MDD, BIP, SCZ					1.20E-16

c) SNP-heritability of SA within psychiatric diagnosis on the liability scale using a range of prevalence estimates for SA in psychiatric disorders

Prevalence of SA in psychiatric disorders	SNP-heritability (se)	P value
0.10	0.037 (0.01)	1.08E-04
0.17	0.044 (0.01)	5.41E-06
0.20	0.046 (0.01)	2.11E-06

GWAS - genome-wide association study, SA - suicide attempt, se - standard error, SA-EUR - European-only meta-analysis of suicide attempt, SA-EUR| MDD - European-only meta-analysis of suicide attempt conditioned on major depressive disorder, SA-EUR | MDD, BIP, SCZ - European-only meta-analysis of suicide attempt conditioned on major depressive disorder, bipolar disorder and schizophrenia. *SA within psychiatric diagnosis - a direct GWAS of suicide attempt within psychiatric diagnosis was conducted for comparison with SA_EUR|MDD (details in Supplementary Note)

Supplementary Table 13: Genetic correlations of suicide attempt with psychiatric traits or disorders

Trait 1	Trait 2	rg	se	p	z	PMID Trait 2	Jackknife rg difference with SA-EUR and SA-EUR|MDD (p-value)
Suicide attempt (SA)	Major depressive disorder (MDD)*	0.780	0.035	5.82E-112	22.485	30718901	
SA-EUR	Major depressive disorder (MDD)*	0.778	0.036	4.11E-106	21.879	30718901	
SA-EUR | MDD	Major depressive disorder (MDD)*	0.530	0.060	8.85E-19	8.849	30718901	8.38E-22
SA within Psychiatric Diagnosis	Major depressive disorder (MDD)*	0.520	0.114	4.48E-06	4.588	30718901	
Suicide attempt (SA)	Schizophrenia (SCZ)	0.460	0.035	1.32E-39	13.169	29483656	
SA-EUR	Schizophrenia (SCZ)	0.445	0.036	4.95E-36	12.533	29483656	
SA-EUR | MDD	Schizophrenia (SCZ)	0.400	0.044	3.71E-20	9.196	29483656	9.66E-03
SA within Psychiatric Diagnosis	Schizophrenia (SCZ)	-0.070	0.075	3.24E-01	-0.987	29483656	
Suicide attempt (SA)	Attention-deficit/hyperactivity disorder (ADHD)	0.510	0.040	3.89E-38	12.911	30478444	
SA-EUR	Attention-deficit/hyperactivity disorder (ADHD)	0.511	0.040	3.89E-38	12.911	30478444	
SA-EUR | MDD	Attention-deficit/hyperactivity disorder (ADHD)	0.460	0.057	3.71E-16	8.148	30478444	1.56E-02
SA within Psychiatric Diagnosis	Attention-deficit/hyperactivity disorder (ADHD)	0.600	0.120	7.08E-07	4.959	30478444	
Suicide attempt (SA)	Self-harm ideation	0.813	0.065	3.52E-36	12.560	32546850	
SA-EUR	Self-harm ideation	0.820	0.065	3.57E-36	12.559	32546850	
SA-EUR | MDD	Self-harm ideation	0.650	0.085	1.72E-14	7.670	32546850	3.95E-08
SA within Psychiatric Diagnosis	Self-harm ideation	0.368	0.172	3.25E-02	2.139	32546850	
Suicide attempt (SA)	Bipolar disorder (BIP)	0.490	0.043	1.16E-30	11.511	31043756	
SA-EUR	Bipolar disorder (BIP)	0.452	0.045	2.82E-24	10.166	31043756	
SA-EUR | MDD	Bipolar disorder (BIP)	0.430	0.057	6.02E-14	7.508	31043756	6.09E-01
SA within Psychiatric Diagnosis	Bipolar disorder (BIP)	-0.080	0.101	4.38E-01	-0.776	31043756	
Suicide attempt (SA)	Post-traumatic stress disorder (PTSD)	0.730	0.085	1.23E-17	8.550	31594949	
SA-EUR	Post-traumatic stress disorder (PTSD)	0.743	0.089	5.29E-17	8.380	31594949	
SA-EUR | MDD	Post-traumatic stress disorder (PTSD)	0.590	0.102	8.59E-09	5.757	31594949	1.71E-04
SA within Psychiatric Diagnosis	Post-traumatic stress disorder (PTSD)	0.560	0.191	3.41E-03	2.929	31594949	
Suicide attempt (SA)	Anorexia nervosa (AN)	0.330	0.042	5.38E-15	7.818	31308545	
SA-EUR	Anorexia nervosa (AN)	0.314	0.042	6.20E-14	7.504	31308545	
SA-EUR | MDD	Anorexia nervosa (AN)	0.280	0.056	7.84E-07	4.939	31308545	6.45E-02
SA within Psychiatric Diagnosis	Anorexia nervosa (AN)	0.270	0.098	7.00E-03	2.697	31308545	
Suicide attempt (SA)	Alcohol dependence	0.630	0.098	8.69E-11	6.488	30482948	
SA-EUR	Alcohol dependence	0.596	0.092	1.11E-10	6.451	30482948	
SA-EUR | MDD	Alcohol dependence	0.520	0.108	1.54E-06	4.806	30482948	2.72E-02
SA within Psychiatric Diagnosis	Alcohol dependence	0.610	0.214	4.50E-03	2.841	30482948	
Suicide attempt (SA)	Alcohol Use Disorders Identification Test-P (AUDIT-P)^	0.370	0.057	1.55E-10	6.401	30336701	
SA-EUR	Alcohol Use Disorders Identification Test-P (AUDIT-P)^	0.351	0.057	8.77E-10	6.130	30336701	
SA-EUR | MDD	Alcohol Use Disorders Identification Test-P (AUDIT-P)^	0.300	0.069	1.24E-05	4.370	30336701	1.63E-02
SA within Psychiatric Diagnosis	Alcohol Use Disorders Identification Test-P (AUDIT-P)^	0.250	0.133	6.15E-02	1.870	30336701	
Suicide attempt (SA)	Ausism spectrum disorder (ASD)	0.280	0.060	3.10E-06	4.664	30804558	
SA-EUR	Ausism spectrum disorder (ASD)	0.254	0.061	2.93E-05	4.179	30804558	
SA-EUR | MDD	Ausism spectrum disorder (ASD)	0.140	0.081	9.14E-02	1.688	30804558	8.79E-06
SA within Psychiatric Diagnosis	Ausism spectrum disorder (ASD)	0.170	0.124	1.61E-01	1.402	30804558	
Suicide attempt (SA)	Tourette syndrome	0.230	0.073	2.14E-03	3.070	30818990	
SA-EUR	Tourette syndrome	0.219	0.074	2.92E-03	2.976	30818990	
SA-EUR | MDD	Tourette syndrome	0.210	0.094	2.83E-02	2.193	30818990	7.33E-01
SA within Psychiatric Diagnosis	Tourette syndrome	-0.020	0.147	8.96E-01	-0.131	30818990	
Suicide attempt (SA)	Obsessive compulsive disorder (OCD)	0.140	0.092	1.21E-01	1.550	28761083	
SA-EUR	Obsessive compulsive disorder (OCD)	0.130	0.092	1.53E-01	1.430	28761083	
SA-EUR | MDD	Obsessive compulsive disorder (OCD)	0.030	0.112	7.90E-01	0.267	28761083	9.68E-03
SA within Psychiatric Diagnosis	Obsessive compulsive disorder (OCD)	-0.130	0.173	4.51E-01	-0.754	28761083	

SA-EUR - European-only meta-analysis of suicide attempt, SA-EUR| MDD - European-only meta-analysis of suicide attempt conditioned on major depressive disorder

SA within Psychiatric Diagnosis - a direct GWAS of suicide attempt within psychiatric diagnosis was conducted for comparison with SA_EUR|MDD (details in Supplementary Note)

*Results excluding 23andMe.

^Measure of problematic consequences of drinking.

Supplementary Table 14: Genetic correlations of suicide attempt with diseases and traits from LD Hub, ranked by significance of genetic correlation with SA-EUR|MDD.
Table is abridges to show the first fifty lines; full table can be found at manuscript (PMID 34861974)

SNPs from the MHC (chr6 26M~34M) region were removed for all traits.
*Caution: genetic correlations with samples of mixed ethnicity may yield less robust results due to minor departure of the LD structure. SA-EUR - European-only meta-analysis of suicide attempt, SA-EUR| MDD - European-only meta-analysis of suicide attempt conditioned on major depressive disorder, SA within Psychiatric Diagnosis - a direct GWAS of suicide attempt within psychiatric diagnosis was conducted for comparison with SA_EUR|MDD (details in Supplementary N
rg, genetic correlation; se, standard error; h2_obs, h2_obs_se, observed scale h2 for trait 2 and standard error; h2_int, h2_int_se, single-trait LD Score regression intercept for trait 2 and standard error; gcov_int, gcov_int_se = cross-trait LD Score regression intercept and standard error. Results in bold text pass the Bonferroni corrected significance threshold of P < 6.54E-05.

				Primary SA GWAS (trans-ancestry)					SA-EUR					SA-EUR | MDD					SA within psychiatric diagnosis				
Trait	PMID^	Ethnicity*	SA Risk Category	rg	SE	P	gcov_int	gcov_int_se	rg	SE	P	gcov_int	gcov_int_se	rg	SE	P	gcov_int	gcov_int_se	rg	SE	P	gcov_int	gcov_int_se
Current tobacco smoking	UK Biobank	European	smoking	0.563	0.038	5.38E-49	0.007	6.50E-03	0.569	0.039	1.28E-48	0.010	0.007	0.579	0.056	3.08E-25	0.005	0.007	0.678	0.129	1.39E-07	0.009	0.0064
Age of first birth	27798627	European	socioeconomic	-0.547	0.041	9.90E-41	-0.005	5.90E-03	-0.561	0.041	2.50E-43	-0.004	0.006	-0.589	0.059	1.01E-23	0.003	0.007	-0.763	0.150	3.75E-07	0.0028	0.0062
Smoking status: Current	UK Biobank	European	smoking	0.549	0.039	2.66E-45	0.007	6.60E-03	0.551	0.039	1.52E-44	0.010	0.007	0.559	0.057	6.85E-23	0.006	0.007	0.646	0.127	3.31E-07	0.0091	0.0065
Age at first live birth	UK Biobank	European	socioeconomic	-0.535	0.037	6.33E-47	-0.012	6.80E-03	-0.555	0.036	1.80E-52	-0.011	0.007	-0.524	0.054	1.27E-22	-0.006	0.008	-0.721	0.129	1.98E-08	-0.0022	0.0065
Past tobacco smoking	UK Biobank	European	smoking	-0.420	0.035	6.30E-34	-0.017	6.30E-03	-0.421	0.033	2.34E-38	-0.019	0.006	-0.439	0.046	8.76E-22	-0.007	0.007	-0.353	0.095	2.00E-04	-0.0115	0.0065
Ever smoked	UK Biobank	European	smoking	0.420	0.037	2.82E-29	0.012	6.30E-03	0.415	0.035	3.33E-32	0.016	0.006	0.421	0.047	4.37E-19	0.008	0.007	0.326	0.098	8.00E-04	0.0147	0.0066
Townsend deprivation index at recruitment	UK Biobank	European	socioeconomic	0.563	0.043	2.77E-39	0.016	6.10E-03	0.592	0.045	2.26E-40	0.013	0.006	0.552	0.064	4.81E-18	0.007	0.007	0.689	0.140	8.34E-07	0.0078	0.0058
Illness/injury/bereavement/stress in last 2 years: Financial difficulties	UK Biobank	European	life stressors	0.571	0.042	7.10E-43	0.014	6.00E-03	0.577	0.043	2.48E-41	0.015	0.006	0.510	0.060	9.68E-18	0.007	0.007	0.676	0.139	1.08E-06	0.0108	0.0066
Age completed full time education	UK Biobank	European	socioeconomic	-0.452	0.035	3.10E-38	-0.005	6.10E-03	-0.478	0.036	1.93E-40	-0.002	0.006	-0.451	0.055	1.85E-16	-0.003	0.007	-0.726	0.125	5.96E-09	0.005	0.0062
Schizophrenia	25056061	Mixed	psychiatric	0.421	0.030	1.15E-45	0.077	7.30E-03	0.411	0.031	2.33E-41	0.083	0.008	0.356	0.043	2.20E-16	0.071	0.008	-0.071	0.082	3.88E-01	0.0229	0.007
Number of full sisters	UK Biobank	European	socioeconomic	0.551	0.059	6.42E-21	-0.005	5.80E-03	0.570	0.057	2.07E-23	-0.006	0.006	0.646	0.084	1.35E-14	-0.014	0.007	0.844	0.171	8.37E-07	-0.0075	0.0067
Current employment status: Unable to work because of sickness or disability	UK Biobank	European	life stressors	0.617	0.054	1.15E-30	0.015	6.40E-03	0.622	0.053	1.17E-31	0.018	0.006	0.534	0.072	1.32E-13	0.012	0.007	0.548	0.146	2.00E-04	0.0152	0.0059
Frequency of unenthusiasm / disinterest in last 2 weeks	UK Biobank	European	psychiatric	0.655	0.042	2.55E-55	0.014	6.00E-03	0.657	0.043	5.23E-53	0.015	0.006	0.483	0.066	1.90E-13	0.005	0.007	0.787	0.143	3.98E-08	0.0048	0.0062
Overall health rating	UK Biobank	European	socioeconomic	0.446	0.031	6.56E-46	0.022	7.10E-03	0.455	0.031	2.64E-50	0.025	0.007	0.333	0.046	3.61E-13	0.015	0.008	0.489	0.097	4.38E-07	0.0136	0.0074
Age started oral contraceptive pill	UK Biobank	European	NA	-0.574	0.053	3.07E-27	0.001	5.40E-03	-0.561	0.054	1.08E-25	-0.003	0.006	-0.578	0.080	4.45E-13	0.002	0.007	-0.559	0.152	2.00E-04	-0.0109	0.0061
Age at last live birth	UK Biobank	European	socioeconomic	-0.424	0.044	1.09E-21	-0.012	6.50E-03	-0.450	0.043	8.28E-26	-0.008	0.006	-0.445	0.062	5.71E-13	-0.005	0.007	-0.518	0.127	4.61E-05	-0.0068	0.0063
Frequency of tenseness / restlessness in last 2 weeks	UK Biobank	European	psychiatric	0.630	0.039	1.10E-59	0.009	6.00E-03	0.606	0.039	1.16E-53	0.012	0.006	0.404	0.056	7.15E-13	0.005	0.007	0.578	0.114	4.37E-07	0.0061	0.0061
Types of physical activity in last 4 weeks: None of the above	UK Biobank	European	NA	0.502	0.047	4.44E-27	0.002	5.20E-03	0.526	0.050	2.58E-26	0.002	0.005	0.480	0.067	9.68E-13	0.004	0.006	0.578	0.136	2.00E-05	0.0039	0.0051
Smoking status: Previous	UK Biobank	European	smoking	0.340	0.039	4.75E-18	0.014	5.80E-03	0.344	0.037	2.26E-20	0.014	0.006	0.360	0.051	1.93E-12	0.004	0.007	0.267	0.099	6.90E-03	0.0059	0.0063
Alcohol usually taken with meals	UK Biobank	European	alcohol	-0.411	0.042	5.63E-23	0.003	6.50E-03	-0.427	0.042	2.61E-24	0.005	0.006	-0.410	0.058	1.95E-12	0.003	0.007	-0.591	0.130	5.80E-06	0.0069	0.0067
Qualifications: College or University degree	UK Biobank	European	education/cognition	-0.264	0.032	6.17E-17	0.010	7.70E-03	-0.289	0.031	4.06E-21	0.014	0.007	-0.297	0.043	7.54E-12	0.021	0.009	-0.455	0.095	1.72E-06	0.0027	0.0074
Job involves heavy manual or physical work	UK Biobank	European	NA	0.339	0.038	4.07E-19	-0.004	5.80E-03	0.356	0.039	6.03E-20	-0.007	0.006	0.370	0.054	8.42E-12	-0.008	0.007	0.642	0.131	9.30E-07	-0.0092	0.0067
Seen a psychiatrist for nerves/anxiety/tension/depression	UK Biobank	European	psychiatric	0.736	0.048	1.07E-53	0.041	6.10E-03	0.695	0.045	6.12E-55	0.048	0.006	0.431	0.064	1.35E-11	0.034	0.007	0.365	0.133	6.10E-03	0.0234	0.0062
Years of schooling 2016	27225129	European	education/cognition	-0.309	0.033	5.93E-21	0.001	6.40E-03	-0.323	0.031	5.50E-25	0.002	0.006	-0.291	0.044	2.90E-11	-0.002	0.008	-0.418	0.095	9.61E-06	-0.006	0.0067
Maternal smoking around birth	UK Biobank	European	smoking	0.385	0.038	4.36E-24	0.018	6.10E-03	0.402	0.040	2.22E-23	0.016	0.006	0.370	0.056	3.41E-11	0.014	0.007	0.418	0.112	2.00E-04	0.0088	0.0064
Pack years adult smoking as proportion of life span exposed to smoking	UK Biobank	European	smoking	0.452	0.044	1.07E-24	0.009	5.00E-03	0.463	0.045	2.50E-25	0.009	0.005	0.414	0.063	3.98E-11	0.010	0.006	0.541	0.132	4.16E-05	0.0061	0.0064
Qualifications: None of the above	UK Biobank	European	education/cognition	0.314	0.034	6.84E-21	-0.004	7.00E-03	0.337	0.034	3.15E-23	-0.007	0.007	0.313	0.048	4.23E-11	-0.006	0.008	0.527	0.108	1.03E-06	-0.0059	0.0074
Wheeze or whistling in the chest in last year	UK Biobank	European	smoking	0.431	0.038	1.20E-29	0.018	6.30E-03	0.441	0.037	4.24E-33	0.019	0.006	0.358	0.055	4.89E-11	0.016	0.007	0.500	0.111	6.24E-06	0.0105	0.0071
Qualifications: A levels/AS levels or equivalent	UK Biobank	European	education/cognition	-0.311	0.033	5.89E-21	0.005	5.90E-03	-0.318	0.033	9.56E-22	0.004	0.006	-0.322	0.049	5.54E-11	0.011	0.008	-0.533	0.103	2.52E-07	0.0104	0.0065
Number of full brothers	UK Biobank	European	socioeconomic	0.509	0.058	1.95E-18	0.004	6.00E-03	0.511	0.054	3.71E-21	0.008	0.006	0.509	0.079	9.04E-11	0.002	0.007	0.772	0.164	2.68E-06	0.0017	0.0058
Number of children ever born	27798627	European	socioeconomic	0.384	0.050	9.17E-15	-0.009	5.10E-03	0.409	0.050	4.96E-16	-0.012	0.005	0.456	0.071	1.14E-10	-0.009	0.007	0.585	0.137	1.89E-05	-0.0077	0.0055
Frequency of tiredness / lethargy in last 2 weeks	UK Biobank	European	psychiatric	0.541	0.036	1.26E-51	0.016	6.40E-03	0.532	0.036	1.42E-50	0.021	0.006	0.354	0.056	2.63E-10	0.010	0.008	0.545	0.121	6.39E-06	0.0118	0.0069
Tobacco smoking: Ex-smoker	UK Biobank	European	smoking	0.429	0.055	1.00E-14	0.013	6.00E-03	0.458	0.056	4.69E-16	0.011	0.006	0.484	0.078	6.73E-10	0.005	0.007	0.384	0.133	0.0039	0.0064	0.0057
Risk taking	UK Biobank	European	psychiatric	0.313	0.038	1.69E-16	0.021	6.40E-03	0.307	0.039	1.93E-15	0.022	0.007	0.342	0.056	1.30E-09	0.014	0.008	0.313	0.099	1.60E-03	0.0054	0.0069
Job involves mainly walking or standing	UK Biobank	European	socioeconomic	0.280	0.040	2.85E-12	-0.007	5.90E-03	0.295	0.041	1.14E-12	-0.010	0.006	0.323	0.053	1.30E-09	-0.016	0.007	0.447	0.109	4.14E-05	-4.20E-03	0.0064
Pack years of smoking	UK Biobank	European	smoking	0.425	0.046	1.72E-20	0.009	5.00E-03	0.438	0.046	1.73E-21	0.009	0.005	0.388	0.066	3.19E-09	0.011	0.006	0.522	0.139	2.00E-04	0.0021	0.0064
Job involves shift work	UK Biobank	European	NA	0.399	0.056	8.57E-13	0.002	4.70E-03	0.417	0.054	1.68E-14	0.000	0.005	0.467	0.079	3.24E-09	-0.005	0.006	0.623	0.167	2.00E-04	-0.0076	0.0059
Seen doctor (GP) for nerves/anxiety/tension/depression	UK Biobank	European	psychiatric	0.628	0.039	1.64E-59	0.032	5.90E-03	0.607	0.039	8.07E-55	0.037	0.006	0.323	0.055	5.47E-09	0.009	0.008	0.354	0.105	7.00E-04	0.0143	0.0062
Types of physical activity in last 4 weeks: Walking for pleasure (not as a means of transpo	UK Biobank	European	NA	-0.355	0.043	9.11E-17	-0.005	5.60E-03	-0.363	0.041	6.64E-19	-0.007	0.005	-0.369	0.064	6.94E-09	-0.005	0.007	-0.589	0.138	2.01E-05	5.00E-04	0.0063
Intelligence	28530673	European	education/cognition	-0.276	0.041	8.95E-12	0.003	5.70E-03	-0.294	0.039	3.40E-14	0.006	0.005	-0.308	0.054	9.41E-09	0.006	0.007	-0.218	0.100	0.0286	-0.008	0.0061
Medication for pain relief constipation heartburn: None of the above	UK Biobank	European	NA	-0.390	0.038	7.97E-25	-0.008	6.20E-03	-0.403	0.038	6.28E-26	-0.008	0.006	-0.295	0.052	1.01E-08	-0.004	0.007	-0.375	0.103	3.00E-04	-0.0077	0.0061
Qualifications: Other professional qualifications eg: nursing teaching	UK Biobank	European	education/cognition	-0.298	0.042	8.95E-13	0.004	6.20E-03	-0.330	0.042	4.73E-15	0.007	0.006	-0.341	0.060	1.15E-08	0.011	0.007	-0.600	0.127	2.48E-06	0.0054	0.0062
Financial situation satisfaction	UK Biobank	European	life stressors	0.440	0.054	2.14E-16	0.015	5.50E-03	0.447	0.053	2.20E-17	0.016	0.005	0.431	0.076	1.18E-08	-0.003	0.006	0.582	0.149	9.55E-05	0.0085	0.0066
Number of live births	UK Biobank	European	socioeconomic	0.322	0.046	2.68E-12	-0.002	6.10E-03	0.326	0.048	1.16E-11	-0.002	0.007	0.372	0.066	1.37E-08	-0.009	0.007	0.512	0.115	9.35E-06	-0.0074	0.0063
Frequency of depressed mood in last 2 weeks	UK Biobank	European	psychiatric	0.630	0.045	8.55E-44	0.016	5.90E-03	0.619	0.047	5.26E-40	0.019	0.006	0.391	0.069	1.55E-08	0.009	0.007	0.601	0.128	2.69E-06	0.0087	0.0058
Pain type(s) experienced in last month: Neck or shoulder pain	UK Biobank	European	pain	0.422	0.045	5.94E-21	0.006	5.40E-03	0.441	0.044	2.55E-23	0.005	0.006	0.340	0.061	2.04E-08	0.000	0.006	0.416	0.118	4.00E-04	0.0024	0.0062
Illness/injury/bereavement/stress in last 2 years: None of the above	UK Biobank	European	life stressors	-0.473	0.047	4.06E-24	-0.014	5.10E-03	-0.484	0.043	7.10E-29	-0.015	0.005	-0.373	0.067	2.28E-08	-0.012	0.006	-0.482	0.146	9.00E-04	-0.0083	0.0057
Mouth/teeth dental problems: Loose teeth	UK Biobank	European	NA	0.412	0.057	3.35E-13	0.002	5.00E-03	0.411	0.059	4.11E-12	0.002	0.005	0.489	0.088	2.94E-08	-0.010	0.006	0.400	0.159	1.16E-02	0.0064	0.0057
Bipolar disorder	21926972	European	psychiatric	0.475	0.048	3.10E-23	0.075	5.90E-03	0.433	0.048	1.67E-19	0.081	0.006	0.359	0.065	3.00E-08	0.070	0.006	-0.151	0.125	2.28E-01	0.01	0.0054
PGC cross-disorder analysis	23453885	European	psychiatric	0.501	0.047	2.55E-26	0.101	8.20E-03	0.459	0.048	9.83E-22	0.111	0.008	0.328	0.060	3.84E-08	0.054	0.008	0.078	0.126	5.34E-01	0.0086	0.0077

Supplementary Table 15: List of VUMC TRD model LASSO features and weights

FEATURE	SCORE	CONCEPT_NAME
ICD_CCS_2__5_13	0.87522048	Suicide and intentional self-inflicted injury [662]
ICD_CCS_2__5_8	0.69413447	Mood disorders [657]
RXNORM_IN__6581	0.67910147	Magnesium Hydroxide
RXNORM_IN__191831	0.47540085	infliximab
RXNORM_IN__5666	0.46283031	Immunoglobulin G
RXNORM_IN__61381	0.45050063	olanzapine
RXNORM_IN__68244	0.44306799	Lamivudine
RXNORM_IN__356887	0.37050246	levocetirizine
RXNORM_IN__6904	0.36178216	Methyltestosterone
RXNORM_IN__8727	0.30967013	Progesterone
RXNORM_IN__51272	0.30099105	quetiapine
RXNORM_IN__6711	0.28915909	Melatonin
RXNORM_IN__89013	0.27126328	aripiprazole
RXNORM_IN__10454	0.23402369	Thiamine
RXNORM_IN__2101	0.22677463	Carisoprodol
RXNORM_IN__187832	0.19534327	pregabalin
RXNORM_IN__10600	0.17082245	Timolol
RXNORM_IN__9524	0.16753337	Sulfasalazine
RXNORM_IN__42463	0.15408274	Pravastatin
RXNORM_IN__15996	0.14606123	Mirtazapine
RXNORM_IN__1244014	0.13574898	vitamin D3
RXNORM_IN__52356	0.13221641	magnesium citrate
RXNORM_IN__21406	0.12900234	coenzyme Q10
RXNORM_IN__1827	0.12147474	Buspirone
RXNORM_IN__2598	0.11589387	Clonazepam
RXNORM_IN__1364430	0.09245822	apixaban
RXNORM_IN__10737	0.08879983	Trazodone
RXNORM_IN__7531	0.08476666	Nortriptyline
RXNORM_IN__816346	0.07085338	dexlansoprazole
RXNORM_IN__8787	0.06780048	Propranolol
RXNORM_IN__40254	0.06263074	Valproate
RXNORM_IN__28439	0.05851764	lamotrigine
RXNORM_IN__39786	0.04936285	venlafaxine
RXNORM_IN__6470	0.04037292	Lorazepam
RXNORM_IN__72625	0.02991102	duloxetine
DEMO__AGE	0.02126898	Age at index
RXNORM_IN__8686	0.01826875	Prilocaine
DEMO__GENDER__M	0.00920184	Male gender
DEMO__ADI	0.00541117	Census-tract ADI

RXNORM_IN__461016	0.00460011	Eszopiclone
ICD_CCS_2__5_6	-0.0026221	Disorders usually diagnosed in infancy childhood or adolescence [655]
RXNORM_IN__704	-0.0060344	Amitriptyline
RXNORM_IN__301542	-0.0085486	rosuvastatin
RXNORM_IN__341248	-0.0102032	ezetimibe
RXNORM_IN__17128	-0.0106024	lansoprazole
RXNORM_IN__6922	-0.0121761	Metronidazole
ICD_CCS_2__2_16	-0.0145582	Benign neoplasms
RXNORM_IN__321988	-0.0232468	Escitalopram
RXNORM_IN__8896	-0.0249526	Pseudoephedrine
ICD_CCS_2__9_10	-0.0271596	Gastrointestinal hemorrhage [153.]
RXNORM_IN__6703	-0.0349746	Megestrol
RXNORM_IN__32937	-0.0358884	Paroxetine
RXNORM_IN__2582	-0.040826	Clindamycin
ICD_CCS_2__6_8	-0.0435291	Ear conditions
RXNORM_IN__1897	-0.0437151	Calcium Carbonate
RXNORM_IN__25480	-0.0492266	gabapentin
RXNORM_IN__39993	-0.0514286	zolpidem
RXNORM_IN__18631	-0.0516464	Azithromycin
RXNORM_IN__4124	-0.0529376	Ethinyl Estradiol
ICD_CCS_2__2_4	-0.0624299	Cancer of skin
RXNORM_IN__3289	-0.0698636	Dextromethorphan
RXNORM_IN__1291	-0.0701002	Bacitracin
RXNORM_IN__4603	-0.0881517	Furosemide
ICD_CCS_2__9_7	-0.0882412	Biliary tract disease [149.]
RXNORM_IN__7242	-0.0975052	Naloxone
ICD_CCS_2__5_7	-0.1005782	Impulse control disorders not elsewhere classified [656]
RXNORM_IN__1819	-0.1018914	Buprenorphine
RXNORM_IN__6387	-0.1087194	Lidocaine
ICD_CCS_2__13_4	-0.1097803	Osteoporosis [206.]
RXNORM_IN__36437	-0.1157892	Sertraline
RXNORM_IN__3361	-0.118119	Dicyclomine
ICD_CCS_2__16_12	-0.1221898	Other injuries and conditions due to external causes [244.]
RXNORM_IN__3355	-0.1318891	Diclofenac
ICD_CCS_2__4_1	-0.1544218	Anemia
ICD_CCS_2__12_4	-0.1565918	Other skin disorders [200.]
RXNORM_IN__1596	-0.1612046	Bisacodyl
RXNORM_IN__2556	-0.172749	Citalopram
DEMO__GENDER__F	-0.1953957	Female gender
ICD_CCS_2__5_3	-0.2021699	Attention deficit conduct and disruptive behavior disorders [652]

RXNORM_IN__5691	-0.2119162	Imipramine
ICD_CCS_2__16_6	-0.2291608	Open wounds
RXNORM_IN__435	-0.2627669	Albuterol
ICD_CCS_2__5_9	-0.2855292	Personality disorders [658]
ICD_CCS_2__17_2	-0.2973178	Factors influencing health care
DEMO__RACE__B	-0.3036093	Black race
RXNORM_IN__3498	-0.3058396	Diphenhydramine
RXNORM_IN__1223	-0.3583579	Atropine
ICD_CCS_2__5_11	-0.4248819	Alcohol-related disorders [660]
(Intercept)	-8.1557044	

Supplementary Table 16: List of MGB TRD model LASSO features and weights

FEATURE_NAME	VALUE	FEATURE_DESC
DEMO__ADI	-0.0050806	Zip-based ADI
DEMO__AGE	0.02686874	Age at index
DEMO__ELIX	-0.0103528	Elixhauser Comorbidity Score
DEMO__ETHNICITY__HL	-0.5227552	Hispanic
DEMO__ETHNICITY__NH	8.10E-12	Non-Hispanic
DEMO__GENDER__F	-0.2131091	Female Gender
DEMO__GENDER__M	0.01104108	Male Gender
DEMO__RACE__A	0.09842257	Asian Race
DEMO__RACE__W	0.44698165	White Race
RXNORM_IN__10582	-0.0914794	levothyroxine
RXNORM_IN__10737	0.11678074	trazodone
RXNORM_IN__108118	-0.1133929	mometasone
RXNORM_IN__11131	0.25155949	varicella
RXNORM_IN__114477	-0.0732036	levetiracetam
RXNORM_IN__1191	-0.0956622	aspirin
RXNORM_IN__142434	0.30780246	econazole nitrate
RXNORM_IN__153970	0.4162128	hyoscyamine
RXNORM_IN__161	-0.0477559	acetaminophen
RXNORM_IN__18631	-0.1405051	azithromycin
RXNORM_IN__21090	0.20295224	ciclopirox
RXNORM_IN__2193	-0.0715914	ceftriaxone
RXNORM_IN__2356	-0.4145198	chlordiazepoxide
RXNORM_IN__24947	-0.0020973	ferrous sulfate
RXNORM_IN__25025	0.11888502	finasteride
RXNORM_IN__2556	-0.1560726	citalopram
RXNORM_IN__2598	0.3116751	clonazepam
RXNORM_IN__327361	0.47098954	adalimumab
RXNORM_IN__32937	-0.0794955	paroxetine
RXNORM_IN__3443	-0.0553176	diltiazem
RXNORM_IN__35208	0.72817704	quinapril
RXNORM_IN__3640	-0.1307273	doxycycline
RXNORM_IN__36437	-0.2341094	sertraline
RXNORM_IN__3992	-0.1174041	epinephrine
RXNORM_IN__4337	-0.2588799	fentanyl
RXNORM_IN__4419	0.20534432	fish oil
RXNORM_IN__4603	-0.0774138	furosemide
RXNORM_IN__46041	-0.0014173	alendronate
RXNORM_IN__51272	0.63306615	quetiapine
RXNORM_IN__519	-0.0168698	allopurinol
RXNORM_IN__5640	-0.0814029	ibuprofen

RXNORM_IN__612	0.52622807	aluminum hydroxide
RXNORM_IN__6375	0.09323907	levodopa
RXNORM_IN__6387	-0.0649814	lidocaine
RXNORM_IN__6581	0.09813547	magnesium hydroxide
RXNORM_IN__6809	-0.1821765	metformin
RXNORM_IN__6918	-0.4603138	metoprolol
RXNORM_IN__72143	0.74512234	raloxifene
RXNORM_IN__7804	-0.0805118	oxycodone
RXNORM_IN__82003	-0.0408666	docusate
RXNORM_IN__8591	-0.1530172	potassium chloride
RXNORM_IN__8704	-0.0596349	prochlorperazine
RXNORM_IN__8896	0.0556722	pseudoephedrine
RXNORM_IN__89905	0.42771109	multivitamins
RXNORM_IN__9863	-0.0489074	sodium chloride
ICD_CCS_2__1_3	-0.0939108	Viral infection
ICD_CCS_2__1_5	-0.0507035	Immunizations and screening for infectious disease [10.]
ICD_CCS_2__10_2	0.06720288	Diseases of male genital organs
ICD_CCS_2__13_2	-0.0608251	Non-traumatic joint disorders
ICD_CCS_2__13_4	-0.132453	Osteoporosis [206.]
ICD_CCS_2__13_9	-0.0861615	Other bone disease and musculoskeletal deformities [212.]
ICD_CCS_2__14_3	0.41443944	Genitourinary congenital anomalies [215.]
ICD_CCS_2__16_11	0.38007415	Poisoning
ICD_CCS_2__16_12	-0.0473065	Other injuries and conditions due to external causes [244.]
ICD_CCS_2__16_5	0.36839068	Crushing injury or internal injury [234.]
ICD_CCS_2__16_8	-0.5483884	Superficial injury; contusion [239.]
ICD_CCS_2__17_2	-0.327971	Factors influencing health care
ICD_CCS_2__2_12	-0.1646711	Secondary malignancies [42.]
ICD_CCS_2__2_14	-0.0227713	Neoplasms of unspecified nature or uncertain behavior [44.]
ICD_CCS_2__2_5	-0.0165791	Cancer of breast [24.]
ICD_CCS_2__3_11	-0.0788667	Other nutritional; endocrine; and metabolic disorders [58.]
ICD_CCS_2__3_6	-0.0755936	Disorders of lipid metabolism [53.]
ICD_CCS_2__4_4	-0.0481203	Other hematologic conditions [64.]
ICD_CCS_2__5_1	-0.1843707	Adjustment disorders [650]
ICD_CCS_2__5_13	0.93294474	Suicide and intentional self-inflicted injury [662]
ICD_CCS_2__5_8	0.66380433	Mood disorders [657]
ICD_CCS_2__5_9	0.40803536	Personality disorders [658]
ICD_CCS_2__7_4	-0.1547601	Diseases of arteries; arterioles; and capillaries
ICD_CCS_2__8_1	-0.0626589	Respiratory infections

ICD_CCS_2__8_3	-0.0331855	Asthma [128.]
ICD_CCS_2__8_9	-0.0251667	Other upper respiratory disease [134.]
ICD_CCS_2__9_11	0.08660664	Noninfectious gastroenteritis [154.]
ICD_CCS_2__9_7	-0.0532803	Biliary tract disease [149.]
ICD_CCS_2__9_8	-0.0076614	Liver disease

Supplementary Table 17: VUMC TRD model PheWAS results that pass multiple testing correction (p=0.05/1103)

Phenotype	PheWAS string	Est	SE	Z	P
297.1	Suicidal ideation	3.618	0.150	24.097	2.67E-128
296.22	Major depressive disorder	3.309	0.341	9.711	2.72E-22
316	Substance addiction and disorders	1.427	0.153	9.326	1.10E-20
969	Poisoning by psychotropic agents	2.260	0.246	9.175	4.52E-20
300.11	Generalized anxiety disorder	1.389	0.152	9.142	6.12E-20
300.12	Agorophobia, social phobia, and panic disorder	1.685	0.187	9.000	2.26E-19
300.3	Obsessive-compulsive disorders	2.149	0.253	8.507	1.78E-17
296	Mood disorders	1.718	0.207	8.304	1.00E-16
301	Personality disorders	2.553	0.314	8.127	4.40E-16
290.3	Other persistent mental disorders due to conditions classified elsewhere	1.950	0.241	8.091	5.93E-16
297.2	Suicide or self-inflicted injury	1.919	0.248	7.736	1.03E-14
301.2	Antisocial/borderline personality disorder	2.649	0.349	7.600	2.95E-14
300.9	Posttraumatic stress disorder	1.434	0.195	7.349	2.00E-13
292.4	Altered mental status	1.122	0.165	6.815	9.43E-12
327.4	Insomnia	0.970	0.151	6.411	1.45E-10
292.2	Mild cognitive impairment	1.576	0.252	6.245	4.24E-10
1005	Other symptoms	1.046	0.168	6.228	4.72E-10
333.4	Torsion dystonia	1.836	0.301	6.101	1.05E-09
300.1	Anxiety disorder	0.817	0.135	6.071	1.27E-09
979	Adverse drug events and drug allergies	1.151	0.198	5.824	5.76E-09
292.3	Memory loss	1.145	0.198	5.782	7.37E-09
300	Anxiety, phobic and dissociative disorders	1.454	0.255	5.698	1.22E-08
458.1	Orthostatic hypotension	1.188	0.225	5.288	1.24E-07
300.4	Dysthymic disorder	1.038	0.200	5.192	2.08E-07
260.3	Adult failure to thrive	1.238	0.242	5.118	3.09E-07
333.1	Essential tremor	1.363	0.266	5.114	3.15E-07
291.1	Transient mental disorders due to conditions classified elsewhere	2.594	0.513	5.058	4.24E-07
347	Cataplexy and narcolepsy	2.149	0.457	4.703	2.57E-06
333.2	Myoclonus	1.788	0.388	4.603	4.16E-06
291.4	Specific nonpsychotic mental disorders due to brain damage	1.424	0.325	4.378	1.20E-05
728.2	Laxity of ligament or hypermobility syndrome	2.170	0.513	4.233	2.31E-05
290.1	Dementias	1.133	0.270	4.198	2.69E-05

Supplementary Table 18: MGB TRD model PheWAS results that pass multiple testing correction (p=0.05/1110)

Phenotype	PheWAS string	Est	SE	Z	P
297.1	Suicidal ideation	2.575	0.180	14.289	2.58E-46
296.22	Major depressive disorder	2.731	0.239	11.403	4.04E-30
1019	Other ill-defined and unknown causes of morbidity and mortality	1.472	0.142	10.350	4.17E-25
297.2	Suicide or self-inflicted injury	2.592	0.351	7.389	1.48E-13
301.2	Antisocial/borderline personality disorder	2.592	0.369	7.026	2.12E-12
969	Poisoning by psychotropic agents	2.094	0.346	6.050	1.45E-09
300.11	Generalized anxiety disorder	0.820	0.159	5.158	2.49E-07
1010	Other tests	-0.694	0.138	-5.027	4.99E-07
300.3	Obsessive-compulsive disorders	1.353	0.271	4.987	6.13E-07
300.9	Posttraumatic stress disorder	1.111	0.223	4.985	6.18E-07
773	Pain in limb	-0.726	0.155	-4.669	3.02E-06
301	Personality disorders	1.752	0.388	4.515	6.32E-06
745	Pain in joint	-0.625	0.139	-4.503	6.70E-06
1009	Injury, NOS	-1.040	0.240	-4.327	1.51E-05
514	Abnormal findings examination of lungs	-0.932	0.225	-4.150	3.32E-05
512.8	Cough	-0.657	0.159	-4.144	3.42E-05
771.1	Swelling of limb	-1.037	0.254	-4.084	4.42E-05

Supplementary Table 19: Genetic correlations with TRD meta-analyses before and after conditioning for BMI using mtCOJO.
Genetic correlations that pass Bonferroni correction for 29 tests are bolded.

P1	P2	RG	SE	Z	P	PMID/DOI
MGB_meta	**insomnia_2019**	**-0.2202**	0.0528	-4.1696	**3.0509E-05**	30804565
MGB_cond_BMI	insomnia_2019	-0.1093	0.0665	-1.6453	0.099902	30804565
VUMC_meta	insomnia_2019	-0.0777	0.0646	-1.2039	0.22863	30804565
VUMC_cond_BMI	insomnia_2019	0.0059	0.0695	0.0851	0.9322	30804565
MGB_meta	neuroticism_2016	0.0616	0.0769	0.8005	0.42341	27089181
MGB_cond_BMI	neuroticism_2016	0.0578	0.0758	0.7633	0.44527	27089181
VUMC_meta	**neuroticism_2016**	**0.3066**	**0.0851**	**3.602**	**3.16E-04**	27089181
VUMC_cond_BMI	**neuroticism_2016**	**0.2938**	**0.0884**	**3.3232**	**8.90E-04**	27089181
MGB_meta	risk_taking_behavior_2019	-0.1401	0.0493	-2.8428	4.47E-03	30643258
MGB_cond_BMI	risk_taking_behavior_2019	-0.0739	0.0582	-1.2701	0.20404	30643258
VUMC_meta	**risk_taking_behavior_2019**	**-0.2755**	**0.0718**	**-3.8357**	**1.25E-04**	30643258
VUMC_cond_BMI	**risk_taking_behavior_2019**	**-0.2529**	**0.0717**	**-3.5284**	**4.00E-04**	30643258
MGB_meta	**edu_attainment_2018**	**0.4666**	0.035	13.337	**1.4096E-40**	30038396
MGB_cond_BMI	**edu_attainment_2018**	**0.3839**	0.0485	7.9205	**2.3656E-15**	30038396
VUMC_meta	**edu_attainment_2018**	**0.2077**	0.0442	4.7	**2.60E-06**	30038396
VUMC_cond_BMI	edu_attainment_2018	0.1284	0.0477	2.6919	7.10E-03	30038396
MGB_meta	**intelligence_2018**	**0.292**	0.039	7.4801	**7.4247E-14**	29942086
MGB_cond_BMI	**intelligence_2018**	**0.2439**	0.05	4.8766	**1.0795E-06**	29942086
VUMC_meta	**intelligence_2018**	**0.1915**	**0.0521**	**3.6763**	**2.37E-04**	29942086
VUMC_cond_BMI	intelligence_2018	0.1499	0.0572	2.6185	8.83E-03	29942086
MGB_meta	**BMI_EUR_2018**	**-0.6257**	0.0424	-14.7553	**2.8451E-49**	30239722
MGB_cond_BMI	BMI_EUR_2018	-0.031	0.0446	-0.6955	0.48675	30239722
VUMC_meta	**BMI_EUR_2018**	**-0.2658**	**0.0483**	**-5.5085**	**3.62E-08**	30239722
VUMC_cond_BMI	BMI_EUR_2018	0.0073	0.044	0.1664	0.86787	30239722
MGB_meta	**t2dm_2017**	**-0.6245**	0.0612	-10.2002	**1.9782E-24**	28566273
MGB_cond_BMI	**t2dm_2017**	**-0.4249**	0.0838	-5.0684	**4.0121E-07**	28566273
VUMC_meta	t2dm_2017	-0.2177	0.0765	-2.8471	4.41E-03	28566273
VUMC_cond_BMI	t2dm_2017	-0.1018	0.0774	-1.3143	0.18875	28566273
MGB_meta	mdd_2018	-0.0045	0.0586	-0.0766	0.93893	29700475
MGB_cond_BMI	mdd_2018	0.0747	0.0737	1.0136	0.31075	29700475
VUMC_meta	mdd_2018	0.1141	0.0734	1.5536	0.1203	29700475
VUMC_cond_BMI	mdd_2018	0.1314	0.0728	1.8065	0.0708	29700475
MGB_meta	depsx_mdd_2019	-0.0039	0.0454	-0.0866	0.93095	30718901
MGB_cond_BMI	depsx_mdd_2019	0.0554	0.0557	0.9944	0.32005	30718901
VUMC_meta	depsx_mdd_2019	0.1303	0.0582	2.2393	0.0251	30718901
VUMC_cond_BMI	depsx_mdd_2019	0.1552	0.0602	2.5768	0.01	30718901

MGB_meta	BIP_pgc3_2021	0.0274	0.0468	0.5865	0.55753	34002096
MGB_cond_BMI	BIP_pgc3_2021	0.0181	0.0575	0.3142	0.75338	34002096
VUMC_meta	BIP_pgc3_2021	0.003	0.0588	0.051	0.9593	34002096
VUMC_cond_BMI	BIP_pgc3_2021	-0.0067	0.0585	-0.1137	0.9095	34002096
MGB_meta	**scz_pgc3_2020**	**0.1401**	0.0412	3.4021	**6.69E-04**	10.1101/2020.09.12.20192922
MGB_cond_BMI	scz_pgc3_2020	0.0954	0.0517	1.844	0.065183	10.1101/2020.09.12.20192923
VUMC_meta	scz_pgc3_2020	0.0345	0.0507	0.68	0.4965	10.1101/2020.09.12.20192924
VUMC_cond_BMI	scz_pgc3_2020	0.0027	0.0535	0.0509	0.9594	10.1101/2020.09.12.20192925
MGB_meta	**ed_AN2_2019**	**0.2569**	0.0577	4.4512	**8.5405E-06**	31308545
MGB_cond_BMI	ed_AN2_2019	0.0901	0.0732	1.2311	0.21828	31308545
VUMC_meta	ed_AN2_2019	0.1913	0.0766	2.4966	0.012538	31308545
VUMC_cond_BMI	ed_AN2_2019	0.0993	0.0823	1.2065	0.2276	31308545
MGB_meta	anxiety_2016	-0.2101	0.2032	-1.0338	0.3012	26754954
MGB_cond_BMI	anxiety_2016	-0.364	0.261	-1.3945	0.1632	26754954
VUMC_meta	anxiety_2016	0.6663	0.4023	1.656	0.0977	26754954
VUMC_cond_BMI	anxiety_2016	0.637	0.4254	1.4976	0.1342	26754954
MGB_meta	ptsd_2019	-0.2465	0.1213	-2.0325	0.042108	31594949
MGB_cond_BMI	ptsd_2019	-0.1143	0.1331	-0.8584	0.39065	31594949
VUMC_meta	ptsd_2019	-0.226	0.1511	-1.4961	0.1346	31594949
VUMC_cond_BMI	ptsd_2019	-0.1424	0.1603	-0.8881	0.3745	31594949
MGB_meta	cdg_2019	0.0029	0.0894	0.0321	0.97437	31835028
MGB_cond_BMI	cdg_2019	0.0529	0.1124	0.4707	0.63788	31835028
VUMC_meta	cdg_2019	0.0684	0.1098	0.6231	0.53324	31835028
VUMC_cond_BMI	cdg_2019	0.0628	0.1088	0.5779	0.5634	31835028
MGB_meta	**adhd_2017**	**-0.3983**	0.0639	-6.2309	**4.638E-10**	30478444
MGB_cond_BMI	**adhd_2017**	**-0.274**	0.078	-3.5116	**4.45E-04**	30478444
VUMC_meta	**adhd_2017**	**-0.2981**	0.0812	-3.6702	**2.42E-04**	30478444
VUMC_cond_BMI	adhd_2017	-0.2333	0.0853	-2.734	6.30E-03	30478444
MGB_meta	asd_2017	0.0681	0.0826	0.8244	0.40969	30804558
MGB_cond_BMI	asd_2017	0.0834	0.0857	0.973	0.33053	30804558
VUMC_meta	asd_2017	0.1432	0.0963	1.4873	0.13694	30804558
VUMC_cond_BMI	asd_2017	0.1081	0.099	1.0915	0.2751	30804558
MGB_meta	ocd_2017	0.0925	0.0913	1.0137	0.31074	28761083
MGB_cond_BMI	ocd_2017	-0.0882	0.1208	-0.7302	0.46524	28761083
VUMC_meta	ocd_2017	-0.0463	0.1175	-0.3941	0.6935	28761083
VUMC_cond_BMI	ocd_2017	-0.1333	0.1278	-1.0424	0.2972	28761083

MGB_meta	ts_tourette_2018	-0.0482	0.0833	-0.5787	0.56278	30818990
MGB_cond_BMI	ts_tourette_2018	-0.049	0.1031	-0.4751	0.63474	30818990
VUMC_meta	ts_tourette_2018	0.0183	0.1077	0.1696	0.8653	30818990
VUMC_cond_BMI	ts_tourette_2018	0.0225	0.1069	0.2107	0.8331	30818990
MGB_meta	**cig_per_day_2019**	**-0.3771**	0.0573	-6.5846	**4.5606E-11**	30643251
MGB_cond_BMI	**cig_per_day_2019**	**-0.266**	0.0589	-4.5186	**6.2248E-06**	30643251
VUMC_meta	**cig_per_day_2019**	**-0.2366**	**0.0594**	**-3.9858**	**6.73E-05**	30643251
VUMC_cond_BMI	cig_per_day_2019	-0.1751	0.0603	-2.9024	3.70E-03	30643251
MGB_meta	drinks_per_week_2019	-0.0457	0.0468	-0.9767	0.3287	30643251
MGB_cond_BMI	drinks_per_week_2019	-0.0658	0.0601	-1.0943	0.27382	30643251
VUMC_meta	**drinks_per_week_2019**	**-0.4297**	**0.0693**	**-6.2042**	**5.50E-10**	30643251
VUMC_cond_BMI	**drinks_per_week_2019**	**-0.4616**	**0.0681**	**-6.7829**	**1.18E-11**	30643251
MGB_meta	**alcdep_2018**	**-0.4104**	0.1168	-3.5124	**4.44E-04**	30482948
MGB_cond_BMI	**alcdep_2018**	**-0.4534**	0.1418	-3.1973	**1.39E-03**	30482948
VUMC_meta	**alcdep_2018**	**-0.4467**	**0.134**	**-3.3331**	**8.59E-04**	30482948
VUMC_cond_BMI	alcdep_2018	-0.4213	0.1406	-2.9963	2.70E-03	30482948
MGB_meta	aud_c_2018	0.0702	0.0673	1.0435	0.29674	30336701
MGB_cond_BMI	aud_c_2018	0.0077	0.087	0.0883	0.92968	30336701
VUMC_meta	**aud_c_2018**	**-0.2765**	**0.0812**	**-3.4069**	**6.57E-04**	30336701
VUMC_cond_BMI	**aud_c_2018**	**-0.299**	**0.0808**	**-3.6998**	**2.00E-04**	30336701
MGB_meta	aud_p_2018	-0.1053	0.0734	-1.4346	0.1514	30336701
MGB_cond_BMI	aud_p_2018	-0.1124	0.0904	-1.2438	0.21358	30336701
VUMC_meta	**aud_p_2018**	**-0.3234**	0.0936	-3.4553	**5.50E-04**	30336701
VUMC_cond_BMI	**aud_p_2018**	**-0.3055**	**0.0961**	**-3.1793**	**1.50E-03**	30336701
MGB_meta	aud_t_2018	0.0492	0.0657	0.7487	0.45404	30336701
MGB_cond_BMI	aud_t_2018	-0.0084	0.0834	-0.1006	0.91988	30336701
VUMC_meta	**aud_t_2018**	**-0.2955**	0.0788	-3.7488	**1.78E-04**	30336701
VUMC_cond_BMI	**aud_t_2018**	**-0.3107**	**0.0807**	**-3.8502**	**1.00E-04**	30336701
MGB_meta	**cannabis_2018**	**0.2146**	0.0566	3.792	**1.49E-04**	33096046
MGB_cond_BMI	cannabis_2018	0.1927	0.0679	2.839	4.53E-03	33096046
VUMC_meta	cannabis_2018	-0.0933	0.0731	-1.2768	0.20169	33096046
VUMC_cond_BMI	cannabis_2018	-0.1289	0.0764	-1.6868	0.0916	33096046
MGB_meta	SA_ISGC_EUR_2021	-0.0359	0.0658	-0.5456	0.5853	34861974
MGB_cond_BMI	SA_ISGC_EUR_2021	0.033	0.0832	0.3965	0.6917	34861974
VUMC_meta	SA_ISGC_EUR_2021	-0.0442	0.0785	-0.5626	0.5737	34861974
VUMC_cond_BMI	SA_ISGC_EUR_2021	-0.0219	0.0829	-0.264	0.7917	34861974
MGB_meta	self-harm_ideation_2020	0.1126	0.0847	1.3291	0.18382	32546850
MGB_cond_BMI	self-harm_ideation_2020	0.2071	0.1036	1.999	0.045604	32546850
VUMC_meta	self-harm_ideation_2020	0.2019	0.1158	1.7445	0.081075	32546850
VUMC_cond_BMI	self-harm_ideation_2020	0.2241	0.1182	1.8962	0.0579	32546850

MGB_meta	subjective_well_being_2016	-0.1519	0.0796	-1.9092	0.0562	27089181
MGB_cond_BMI	subjective_well_being_2016	-0.1837	0.1004	-1.8291	0.0674	27089181
VUMC_meta	subjective_well_being_2016	-0.2254	0.1079	-2.0876	0.0368	27089181
VUMC_cond_BMI	subjective_well_being_2016	-0.2281	0.1111	-2.0539	0.04	27089181

REFERENCES

1. Hasin, D. S. *et al.* Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. *JAMA Psychiatry* **75**, 336–346 (2018).
2. Steffen, A., Nübel, J., Jacobi, F., Bätzing, J. & Holstiege, J. Mental and somatic comorbidity of depression: a comprehensive cross-sectional analysis of 202 diagnosis groups using German nationwide ambulatory claims data. *BMC Psychiatry* **20**, 142 (2020).
3. Smith, D. J. *et al.* Depression and multimorbidity: a cross-sectional study of 1,751,841 patients in primary care. *The Journal of clinical psychiatry* **75**, 1202–8; quiz 1208 (2014).
4. Lamers, F. *et al.* Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). *J Clin Psychiatry* **72**, 341–348 (2011).
5. Williams, L. S., Ghose, S. S. & Swindle, R. W. Depression and other mental health diagnoses increase mortality risk after ischemic stroke. *Am J Psychiatry* **161**, 1090–1095 (2004).
6. Bartoli, F. *et al.* Depression after stroke and risk of mortality: a systematic review and meta-analysis. *Stroke Res Treat* **2013**, 862978 (2013).
7. Katon, W. *et al.* Depression and diabetes: a potentially lethal combination. *J Gen Intern Med* **23**, 1571–1575 (2008).
8. Chida, Y., Hamer, M., Wardle, J. & Steptoe, A. Do stress-related psychosocial factors contribute to cancer incidence and survival? *Nat Clin Pract Oncol* **5**, 466–475 (2008).
9. Reiche, E. M. V., Nunes, S. O. V. & Morimoto, H. K. Stress, depression, the immune system, and cancer. *Lancet Oncol* **5**, 617–625 (2004).

10. Fried, E. I. The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders* **208**, 191–197 (2017).
11. Arnow, B. A. *et al.* Depression Subtypes in Predicting Antidepressant Response: A Report From the iSPOT-D Trial. *Am J Psychiatry* **172**, 743–750 (2015).
12. Uher, R. *et al.* Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. *J Affect Disord* **132**, 112–120 (2011).
13. Cuijpers, P. *et al.* Melancholic and atypical depression as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression. *Depress Anxiety* **34**, 246–256 (2017).
14. Zimmermann, P. *et al.* Heterogeneity of DSM-IV major depressive disorder as a consequence of subthreshold bipolarity. *Arch Gen Psychiatry* **66**, 1341–1352 (2009).
15. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *The American journal of psychiatry* **157**, 1552–62 (2000).
16. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343–352 (2019).
17. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668–681 (2018).
18. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics* **52**, 437–447 (2020).
19. Jermy, B. S., Glanville, K. P., Coleman, J. R. I., Lewis, C. M. & Vassos, E. Exploring the genetic heterogeneity in major depression across diagnostic criteria. *Molecular Psychiatry 2021* 1–9 (2021) doi:10.1038/s41380-021-01231-w.

20. Nguyen, T.-D. *et al.* Genetic heterogeneity and subtypes of major depression. *Molecular Psychiatry* 2021.03.05.21252911 (2022) doi:10.1038/s41380-021-01413-6.
21. Satyanarayana, S., Enns, M. W., Cox, B. J. & Sareen, J. Prevalence and correlates of chronic depression in the canadian community health survey: mental health and well-being. *Can J Psychiatry* **54**, 389–398 (2009).
22. Boschloo, L. *et al.* The four-year course of major depressive disorder: the role of staging and risk factor determination. *Psychother Psychosom* **83**, 279–288 (2014).
23. Burcusa, S. L. & Iacono, W. G. Risk for recurrence in depression. *Clin Psychol Rev* **27**, 959–985 (2007).
24. Cuijpers, P., Dekker, J., Hollon, S. D. & Andersson, G. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. *J Clin Psychiatry* **70**, 1219–1229 (2009).
25. Malhi, G. S. & Mann, J. J. Depression. *The Lancet* vol. 392 2299–2312 (2018).
26. Arroll, B. *et al.* Antidepressants versus placebo for depression in primary care. *Cochrane Database Syst Rev* CD007954 (2009) doi:10.1002/14651858.CD007954.
27. McIntyre, R. S. When should you move beyond first-line therapy for depression? *J Clin Psychiatry* **71 Suppl 1**, 16–20 (2010).
28. Gaynes, B. N. *et al.* Defining treatment-resistant depression. *Depression and Anxiety* **37**, 134–145 (2020).
29. Rush, A. J. *et al.* Acute and Longer-Term Outcomes in Depressed Outpatients Requiring One or Several Treatment Steps: A STAR*D Report. *American Journal of Psychiatry* **163**, 1905–1917 (2006).

30. Zhdanava, M. *et al.* The Prevalence and National Burden of Treatment-Resistant Depression and Major Depressive Disorder in the United States. *The Journal of Clinical Psychiatry* **82**, (2021).
31. Hantouche, E., Angst, J. & Azorin, J.-M. Explained factors of suicide attempts in major depression. *Journal of Affective Disorders* **127**, 305–308 (2010).
32. Nock, M. K. *et al.* Suicide and Suicidal Behavior. *Epidemiologic Reviews* **30**, 133–154 (2008).
33. Ross, E. L., Zivin, K. & Maixner, D. F. Cost-effectiveness of Electroconvulsive Therapy vs Pharmacotherapy/Psychotherapy for Treatment-Resistant Depression in the United States. *JAMA Psychiatry* **75**, 713–722 (2018).
34. Lisanby, S. H. Electroconvulsive Therapy for Depression. *New England Journal of Medicine* **357**, 1939–1945 (2007).
35. Husain, M. M. *et al.* Speed of Response and Remission in Major Depressive Disorder With Acute Electroconvulsive Therapy (ECT): A Consortium for Research in ECT (CORE) Report. *The Journal of Clinical Psychiatry* **65**, 0–0 (2004).
36. Van Diermen, L. *et al.* Prediction of electroconvulsive therapy response and remission in major depression: meta-analysis*. (2018) doi:10.1192/bjp.2017.
37. Ekstrand, J. *et al.* Racemic Ketamine as an Alternative to Electroconvulsive Therapy for Unipolar Depression: A Randomized, Open-Label, Non-Inferiority Trial (KetECT). *Int J Neuropsychopharmacol* pyab088 (2021) doi:10.1093/ijnp/pyab088.
38. Li, Q. S., Tian, C., Seabrook, G. R., Drevets, W. C. & Narayan, V. A. Analysis of 23andMe antidepressant efficacy survey data: implication of circadian rhythm and neuroplasticity in bupropion response. *Translational Psychiatry* **6**, e889–e889 (2016).

39. Li, Q. S. *et al.* Genome-wide association studies of antidepressant class response and treatment-resistant depression. *Translational Psychiatry* **10**, (2020).

40. Fabbri, C. *et al.* Genetic and clinical characteristics of treatment-resistant depression using primary care records in two UK cohorts. *Molecular Psychiatry* (2021) doi:10.1038/s41380-021-01062-9.

41. Wigmore, E. M. *et al.* Genome-wide association study of antidepressant treatment resistance in a population-based cohort using health service prescription data and meta-analysis with GENDEP. *Pharmacogenomics Journal* **20**, 329–341 (2020).

42. Fabbri, C. *et al.* Genetic and clinical characteristics of treatment-resistant depression using primary care records in two UK cohorts. *Molecular Psychiatry* **26**, 3363–3373 (2021).

43. Clements, C. C. *et al.* Genome-wide association study of patients with a severe major depressive episode treated with electroconvulsive therapy. *Molecular Psychiatry* **26**, 2429–2439 (2021).

44. Pain, O. *et al.* Antidepressant Response in Major Depressive Disorder : A Genome-wide Association Study. (2020) doi:https://doi.org/10.1101/2020.12.11.20245035.

45. Clements, C. C. *et al.* Genome-wide association study of patients with a severe major depressive episode treated with electroconvulsive therapy. *Molecular Psychiatry* **26**, 2429–2439 (2021).

46. Sisti, D., Mann, J. J. & Oquendo, M. A. Toward a Distinct Mental Disorder—Suicidal Behavior. *JAMA Psychiatry* **77**, 661 (2020).

47. Klonsky, E. D., May, A. M. & Saffer, B. Y. Suicide, Suicide Attempts, and Suicidal Ideation. *Annual Review of Clinical Psychology* **12**, 307–330 (2016).

48. Turecki, G. *et al.* Suicide and suicide risk. *Nature reviews. Disease primers* **5**, 74 (2019).

49. Klonsky, E. D., May, A. M. & Saffer, B. Y. Suicide, Suicide Attempts, and Suicidal Ideation. *Annual Review of Clinical Psychology* **12**, 307–330 (2016).
50. Nock, M. K., Borges, G. & Ono, Y. *Suicide: Global Perspectives from the WHO World Mental Health Surveys*. (Cambridge University Press, 2012).
51. World Health Organization. *Suicide. A global imperative*. (World Health Organization, 2014).
52. Hedegaard, H., Curtin, S. C. & Warner, M., 1965-. Increase in suicide mortality in the United States, 1999–2018. (2020).
53. Suominen, K. *et al.* Completed suicide after a suicide attempt: a 37-year follow-up study. *Am J Psychiatry* **161**, 562–563 (2004).
54. Olfson, M. *et al.* National Trends in Suicide Attempts Among Adults in the United States. *JAMA Psychiatry* **74**, 1095–1103 (2017).
55. Durkheim, E. *Suicide: A study in sociology*. (Routledge, 2005).
56. Baumeister, R. F. Suicide as escape from self. *Psychological review* **97**, 90 (1990).
57. Shneidman, E. S. *Suicide as psychache: A clinical approach to self-destructive behavior*. *J Nerv Ment Dis* vol. 181 (J. Aronson, 1993).
58. Klonsky, E. D. & May, A. M. Differentiating Suicide Attempters from Suicide Ideators: A Critical Frontier for Suicidology Research. *Suicide and Life-Threatening Behavior* **44**, 1–5 (2014).
59. Joiner, T. *Why people die by suicide*. (Harvard University Press, 2007).
60. O'Connor, R. C. The integratedmotivational-volitional model of suicidal behavior. *Crisis* vol. 32 295–298 (2011).

61. Klonsky, E. D. & May, A. M. The Three-Step Theory (3ST): A New Theory of Suicide Rooted in the "Ideation-to-Action" Framework. *International Journal of Cognitive Therapy* **8**, 114–129 (2015).
62. World Health Organization. *Suicide. A global imperative*. (World Health Organization, 2014).
63. Nock, M. K. *et al.* Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *British Journal of Psychiatry* **192**, 98–105 (2008).
64. Steele, I. H., Thrower, N., Noroian, P. & Saleh, F. M. Understanding Suicide Across the Lifespan: A United States Perspective of Suicide Risk Factors, Assessment & Management. *Journal of Forensic Sciences* **63**, 162–171 (2018).
65. Stack, S. Suicide: a 15-year review of the sociological literature. Part I: cultural and economic factors. *Suicide & life-threatening behavior* **30**, 145–62 (2000).
66. Canetto, S. S. Gender and Suicide in the Elderly. *Suicide and Life-Threatening Behavior* **22**, 80–97 (1992).
67. Kushner, H. I. Women and suicide in historical perspective. *Signs* **10**, 537–552 (1985).
68. CDC. *Web-based Injury Statistics Query and Reporting System (WISQARS)*. (2019).
69. Rockett, I. R. H. *et al.* Race/ethnicity and potential suicide misclassification: Window on a minority suicide paradox? *BMC Psychiatry* **10**, 35 (2010).
70. Schomerus, G. *et al.* Collective levels of stigma and national suicide rates in 25 European countries. *Epidemiology and Psychiatric Sciences* **24**, 166–171 (2015).
71. Fazel, S. & Runeson, B. Suicide. *New England Journal of Medicine* **382**, 266–274 (2020).
72. Arsenault-Lapierre, G., Kim, C. & Turecki, G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry* **4**, 37 (2004).

73. Chesney, E., Goodwin, G. M. & Fazel, S. Risks of all-cause and suicide mortality in mental disorders: A meta-review. *World Psychiatry* **13**, 153–160 (2014).

74. Yeh, H.-H. *et al.* Diagnosed Mental Health Conditions and Risk of Suicide Mortality. *Psychiatr Serv* **70**, 750–757 (2019).

75. Yoshimasu, K., Kiyohara, C., Miyashita, K., & Stress Research Group of the Japanese Society for Hygiene. Suicidal risk factors and completed suicide: meta-analyses based on psychological autopsy studies. *Environ Health Prev Med* **13**, 243–256 (2008).

76. Walter, F. *et al.* Multiple adverse outcomes following first discharge from inpatient psychiatric care: a national cohort study. *Lancet Psychiatry* **6**, 582–589 (2019).

77. Nordentoft, M., Mortensen, P. B. & Pedersen, C. B. Absolute Risk of Suicide After First Hospital Contact in Mental Disorder. *Archives of General Psychiatry* **68**, 1058 (2011).

78. Franklin, J. C. *et al.* Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin* **143**, 187–232 (2017).

79. Ahmedani, B. K. *et al.* Major Physical Health Conditions and Risk of Suicide. *American Journal of Preventive Medicine* **53**, 308–315 (2017).

80. TBI: Get the Facts | Concussion | Traumatic Brain Injury | CDC Injury Center. https://www.cdc.gov/traumaticbraininjury/get_the_facts.html.

81. Webner, D. & Iverson, G. L. Suicide in professional American football players in the past 95 years. *Brain Injury* **30**, 1718–1721 (2016).

82. Bahraini, N. H., Simpson, G. K., Brenner, L. A., Hoffberg, A. S. & Schneider, A. L. Suicidal ideation and behaviours after traumatic brain injury: A systematic review. *Brain Impairment* **14**, 92–112 (2013).

83. Madsen, T. *et al.* Association between traumatic brain injury and risk of suicide. *JAMA - Journal of the American Medical Association* **320**, 580–588 (2018).

84. Fralick, M. *et al.* Association of Concussion with the Risk of Suicide: A Systematic Review and Meta-analysis. *JAMA Neurology* **76**, 144–151 (2019).

85. Bigler, E. D. The lesion(s) in traumatic brain injury: Implications for clinical neuropsychology. *Archives of Clinical Neuropsychology* **16**, 95–131 (2001).

86. Iverson, G. L. Suicide and Chronic Traumatic Encephalopathy. *The Journal of Neuropsychiatry and Clinical Neurosciences* **28**, 9–16 (2016).

87. Bernert, R. A. & Nadorff, M. R. Sleep Disturbances and Suicide Risk. *Sleep Med Clin* **10**, 35–39 (2015).

88. Woznica, A. A., Carney, C. E., Kuo, J. R. & Moss, T. G. The insomnia and suicide link: Toward an enhanced understanding of this relationship. *Sleep Medicine Reviews* **22**, 37–46 (2015).

89. Pigeon, W. R., Pinquart, M. & Conner, K. Meta-Analysis of Sleep Disturbance and Suicidal Thoughts and Behaviors. *The Journal of Clinical Psychiatry* **73**, e1160–e1167 (2012).

90. Bitew, H. *et al.* Suicidal Ideation, Attempt, and Determining Factors among HIV/AIDS Patients, Ethiopia. *Depression Research and Treatment* **2016**, (2016).

91. Erlangsen, A. *et al.* Genetics of suicide attempts in individuals with and without mental disorders: a population-based genome-wide association study. *Mol. Psychiatry* (2018) doi:10.1038/s41380-018-0218-y.

92. Strawbridge, R. J. *et al.* Identification of novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide. *EBioMedicine* **41**, 517–525 (2019).

93. Campos, A. I. *et al.* Genetic aetiology of self-harm ideation and behaviour. *Scientific Reports* **10**, 9713 (2020).

94. WHO. WHO Suicide Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/suicide (2019).

95. Franklin, J. C. *et al.* Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin* **143**, 187–232 (2017).

96. Arsenault-Lapierre, G., Kim, C. & Turecki, G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry* **4**, 37 (2004).

97. Cavanagh, J. T. O., Carson, A. J., Sharpe, M. & Lawrie, S. M. Psychological autopsy studies of suicide: A systematic review. *Psychological Medicine* vol. 33 395–405 (2003).

98. Consortium, T. S. W. G. of the P. G., Ripke, S., Walters, J. T. & O'Donovan, M. C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv* 2020.09.12.20192922 (2020) doi:10.1101/2020.09.12.20192922.

99. Qin, P. The impact of psychiatric illness on suicide: differences by diagnosis of disorders and by sex and age of subjects. *Journal of psychiatric research* **45**, 1445–52 (2011).

100. Ribeiro, J. D. *et al.* Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychological medicine* **46**, (2016).

101. Khazem, L. R. Physical disability and suicide: recent advancements in understanding and future directions for consideration. *Curr Opin Psychol* **22**, 18–22 (2018).

102. Bishop, T. M., Walsh, P. G., Ashrafioun, L., Lavigne, J. E. & Pigeon, W. R. Sleep, suicide behaviors, and the protective role of sleep medicine. *Sleep Med* **66**, 264–270 (2020).

103. Qin, P., Agerbo, E. & Mortensen, P. B. Suicide risk in relation to family history of completed suicide and psychiatric disorders: a nested case-control study based on longitudinal registers. *Lancet* **360**, 1126–1130 (2002).

104. Poorolajal, J. & Darvishi, N. Smoking and Suicide: A Meta-Analysis. *PLoS One* **11**, e0156348 (2016).

105. Harrison, R., Munafò, M. R., Smith, G. D. & Wootton, R. E. Examining the effect of smoking on suicidal ideation and attempts: triangulation of epidemiological approaches. (2020) doi:10.1192/bjp.2020.68.

106. Orri, M. *et al.* A genetically informed study on the association of cannabis, alcohol, and tobacco smoking with suicide attempt. *Molecular Psychiatry* doi:10.1038/s41380-020-0785-6.

107. McMahon, K. *et al.* Childhood maltreatment and impulsivity as predictors of interpersonal violence, self-injury and suicide attempts: A national study. *Psychiatry Res* **269**, 386–393 (2018).

108. Calati, R. *et al.* Suicidal thoughts and behaviors and social isolation: A narrative review of the literature. *J Affect Disord* **245**, 653–667 (2019).

109. Liu, R. T. & Miller, I. Life events and suicidal ideation and behavior: a systematic review. *Clin Psychol Rev* **34**, 181–192 (2014).

110. Voracek, M. & Loibl, L. M. Genetics of suicide: a systematic review of twin studies. *Wien Klin Wochenschr* 463–475 (2007) doi:10.1007/s00508-007-0823-2.

111. Brent, D. A. & Melhem, N. Familial Transmission of Suicidal Behavior. *Psychiatr Clin North Am* **31**, 157–177 (2008).

112. Fu, Q. *et al.* A twin study of genetic and environmental influences on suicidality in men. *Psychological Medicine* **32**, 11–24 (2002).

113. Ruderfer, D. M. *et al.* Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Mol Psychiatry* **25**, 2422–2430 (2020).

114. Docherty, A. R. *et al.* Genome-Wide Association Study of Suicide Death and Polygenic Prediction of Clinical Antecedents. *AJP* **177**, 917–927 (2020).

115. Mullins, N. *et al.* GWAS of Suicide Attempt in Psychiatric Disorders and Association With Major Depression Polygenic Risk Scores. *Am J Psychiatry* **176**, 651–660 (2019).

116. Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annu Rev Public Health* **34**, 119–138 (2013).

117. Posner, K. *et al.* The Columbia-suicide severity rating scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry* **168**, 1266–1277 (2011).

118. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291–295 (2015).

119. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

120. Lam, M. *et al.* RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* **36**, 930–933 (2020).

121. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**, 1279–1283 (2016).

122. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**, 224 (2018).

123. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics* **88**, 76–82 (2011).

124. Bachmann, S. Epidemiology of Suicide and the Psychiatric Perspective. *Int J Environ Res Public Health* **15**, E1425 (2018).

125. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

126. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics* 1 (2019) doi:10.1038/s41588-019-0397-8.

127. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381–389 (2018).

128. Hübel, C. *et al.* Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **180**, 428–438 (2019).

129. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272 (2017).

130. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).

131. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLOS Computational Biology* **11**, e1004219 (2015).

132. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

133. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics* **48**, 245–252 (2016).

134. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).

135. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications 2019 10:1* **10**, 1–10 (2019).

136. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

137. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol* **36**, 214–224 (2012).

138. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* **48**, 709–717 (2016).

139. Karlsson Linnér, R. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics* **51**, 245–257 (2019).

140. Wootton, R. E. *et al.* Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol Med* **50**, 2435–2443 (2020).

141. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *American Journal of Human Genetics* **94**, 559–573 (2014).

142. Mountjoy, E. *et al.* An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet* **53**, 1527–1533 (2021).

143. Min, J. L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53**, 1311–1321 (2021).

144. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet* **51**, 394–403 (2019).

145. Bergen, S. E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol Psychiatry* **17**, 880–886 (2012).

146. Anttila, V. *et al.* Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).

147. Group of the Psychiatric Genomics Consortium, C.-D. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders Article Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders Cross-Disorder Group of the Psychiatric Genomics Cons. *Cell* **179**, (2019).

148. Agerbo, E., Nordentoft, M. & Mortensen, P. B. Familial, psychiatric, and socioeconomic risk factors for suicide in young people: nested case-control study. *BMJ* **325**, 74 (2002).

149. Qin, P., Agerbo, E. & Mortensen, P. B. Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: a national register-based study of all suicides in Denmark, 1981-1997. *Am J Psychiatry* **160**, 765–772 (2003).

150. Lorant, V. *et al.* Socioeconomic inequalities in suicide in Europe: the widening gap. *Br J Psychiatry* **212**, 356–361 (2018).

151. Lage, I., McCoy Jr, T. H., Perlis, R. H. & Doshi-Velez, F. Efficiently identifying individuals at high risk for treatment resistance in major depressive disorder using electronic health records. *Journal of Affective Disorders* (2022) doi:10.1016/j.jad.2022.02.046.

152. Ruderfer, D. M. *et al.* Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Molecular Psychiatry* 1 (2019) doi:10.1038/s41380-018-0326-8.

153. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668–681 (2018).

154. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343–352 (2019).

155. Nalichowski, R., Keogh, D., Chueh, H. C. & Murphy, S. N. Calculating the benefits of a Research Patient Data Repository. *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2006**, 1044 (2006).

156. Roden, D. M. *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical Pharmacology and Therapeutics* **84**, 362–369 (2008).

157. Gaziano, J. M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology* **70**, 214–223 (2016).

158. Elixhauser, A., Steiner, C. & Palmer, L. Clinical classifications software. *US Agency for Healthcare Research and Quality, URL http://www. hcup-us. ahrq. gov/toolssoftware/ccs/ccs. jsp* (2015).

159. Brokamp, C. *et al.* Material community deprivation and hospital utilization during the first year of life: an urban population-based cohort study. *Annals of epidemiology* **30**, 37–43 (2019).

160. Singh, G. K. Area Deprivation and Widening Inequalities in US Mortality, 1969–1998. *Am J Public Health* **93**, 1137–1143 (2003).

161. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288 (1996).

162. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **33**, 1–22 (2010).

163. Carroll, R. J., Bastarache, L. & Denny, J. C. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* **30**, 2375–2376 (2014).

164. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).

165. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179–181 (2012).

166. Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *American Journal of Human Genetics* vol. 83 132–135 (2008).

167. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).

168. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics 2021 53:7* **53**, 1097–1103 (2021).

169. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

170. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48**, 1279–1283 (2016).

171. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291–295 (2015).

172. Fabbri, C. *et al.* Genetic and clinical characteristics of treatment-resistant depression using primary care records in two UK cohorts. *medRxiv* **44**, 2020.08.24.20178715 (2020).

173. Hübel, C. *et al.* Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **180**, 428–438 (2019).

174. Fabbri, C. *et al.* Genetic and clinical characteristics of treatment-resistant depression using primary care records in two UK cohorts. *Molecular Psychiatry* (2021) doi:10.1038/s41380-021-01062-9.

175. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics* **28**, 166–174 (2019).

176. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications 2019 10:1* **10**, 1–10 (2019).

177. Warden, D., Rush, A. J., Trivedi, M. H., Fava, M. & Wisniewski, S. R. The STAR*D project results: A comprehensive review of findings. *Curr Psychiatry Rep* **9**, 449–459 (2007).

178. Wilkinson, S. T., Agbese, E., Leslie, D. L. & Rosenheck, R. A. Identifying recipients of electroconvulsive therapy: Data from privately insured americans. *Psychiatric Services* **69**, 542–548 (2018).

179. Fabbri, C. *et al.* Genome-wide association study of treatment-resistance in depression and meta-analysis of three independent samples. (2018) doi:10.1192/bjp.2018.256.

180. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics* **28**, 166–174 (2019).

181. Church, C. *et al.* Overexpression of Fto leads to increased food intake and results in obesity. *Nature genetics* **42**, 1086–92 (2010).

182. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–5 (2014).

183. The International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

184. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics* **53**, 817–829 (2021).

185. Soda, T. *et al.* International Consortium on the Genetics of Electroconvulsive Therapy and Severe Depressive Disorders (Gen-ECT-ic). *Eur Arch Psychiatry Clin Neurosci* **270**, 921–932 (2020).

186. Borges, G. *et al.* A risk index for 12-month suicide attempts in the National Comorbidity Survey Replication (NCS-R). *Psychological Medicine* **36**, 1747–1757 (2006).

187. Wuchty, S. *et al.* Integration of peripheral transcriptomics, genomics, and interactomics following trauma identifies causal genes for symptoms of post-traumatic stress and major depression. *Mol Psychiatry* **26**, 3077–3092 (2021).

188. Mullins, N. *et al.* Dissecting the Shared Genetic Architecture of Suicide Attempt, Psychiatric Disorders, and Known Risk Factors. *Biological Psychiatry* **91**, 313–327 (2022).

189. Bejan, C. A. *et al.* Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. *J Am Med Inform Assoc* **25**, 61–71 (2018).

190. Dorr, D. *et al.* Identifying Patients with Significant Problems Related to Social Determinants of Health with Natural Language Processing. *Stud Health Technol Inform* **264**, 1456–1457 (2019).

191. Bejan, C. A. *et al.* Improving ascertainment of suicidal ideation and suicide attempt with natural language processing. 2022.02.25.22271532 (2022) doi:10.1101/2022.02.25.22271532.

192. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. Distributed Representations of Words and Phrases and their Compositionality. in *Advances in Neural Information Processing Systems* vol. 26 (Curran Associates, Inc., 2013).

193. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics 2021 53:7* **53**, 1097–1103 (2021).

194. Brokamp, C. *et al.* Material community deprivation and hospital utilization during the first year of life: an urban population-based cohort study. *Annals of epidemiology* **30**, 37–43 (2019).

195. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics* **52**, 437–447 (2020).

196. Soda, T. *et al.* International Consortium on the Genetics of Electroconvulsive Therapy and Severe Depressive Disorders (Gen-ECT-ic). *European Archives of Psychiatry and Clinical Neuroscience* **270**, 44 (2020).