**Phenotype Algorithm for Type 1 Diabetes – eMERGE Phase-IV Program**

Huiqi Qu, Jeff Roizen, Frank Mentch, John Connolly, Heather Hain, Patrick Sleiman, Hakon Hakonarson

The pseudocode for BMI is listed below. Questions regarding this algorithm can be directed to Frank Mentch ([mentchf@chop.edu](mailto:mentchf@chop.edu)).

|  |  |
| --- | --- |
| **Pseudocode for cases** | |
| **Step** | **Description** |
| 0 | Include patients with a T1D diagnosis and that have been prescribed insulin. ICD9 codes 250.\_1\* | 250.\_3\*; ICD10 codes E10\*; see Table 1 for insulin RXNorm codes. |
| 1 | Exclude patients that have been prescribed T2D medications. See Table 2 for RXNorm codes. |
| 2 | Exclude patients with malignant cancer, identified by 4 or more diagnoses on different days with any ICD9 code between 140.\* and 209\* and any C\* ICD10 code |
| 3 | Exclude patients with a cystic fibrosis diagnosis. ICD9 codes 277.0\* and ICD10 codes E84\*. |
| 4 | Exclude patients with drug induced diabetes. ICD9 codes 249\*; ICD10 codes E09\*. |

|  |  |
| --- | --- |
| **Pseudocode for controls** | |
| **Step** | **Description** |
| 0 | Age > 4 years; 2 of any diagnoses 2 or more years apart |
| 1 | Exclude patients with any diabetes diagnosis or that have been prescribed insulin. ICD9 codes 249\* and 250\*; ICD10 codes E09\*, E10\*, E11\*; see Table 1 for insulin RXNorm codes. |
| 2 | Exclude patients that have been prescribed T2D medications. See Table 2 for RXNorm codes |
| 3 | Exclude patients with malignant cancer, identified by 4 or more inpatient or outpatient diagnoses on different days with any ICD9 code between 140.\* and 209\* and any C\* ICD10 code |
| 4 | Exclude patients with a cystic fibrosis diagnosis. ICD9 codes 277.0\* and ICD10 codes E84\*. |

Table 1. Insulin medications

|  |  |
| --- | --- |
| **Generic name** | **RXNORM RxCUI for generic name** |
| insulin lispro | 86009 |
| insulin glulisine | 400008 |
| insulin glargine, human recombin. analog | 274783 |
| insulin aspart | 51428 |
| humulin | 5459 |
| iletin | 217706 |
| insulin degludec | 1324709 |
| insulin detemir | 139825 |
| insulin isophane nph | 5865 |
| insulin human | 253182 |
| novolin | 1372723 |
| insulin degludec | 1670007 |

Table 2. T2D medications

|  |  |
| --- | --- |
| **Generic name** | **RXNORM RxCUI for generic name** |
| Acarbose | 16681 |
| acetohexamide | 173 |
| albiglutide | 1534763 |
| alogliptin | 1368001 |
| canagliflozin | 1373458 |
| chlorpropamide | 2404 |
| dapagliflozin | 1375326 |
| dapagliflozin | 1488564 |
| dulaglutide | 1551291 |
| empagliflozin | 1545653 |
| ertugliflozin | 1992672 |
| exenatide | 60548 |
| glibenclamide | 4815 |
| glimepiride | 25789 |
| glipizide | 4821 |
| glucotrol xl | 217360 |
| Glyburide | 4815 |
| invokana | 1422507 |
| linagliptin | 1100699 |
| liraglutide | 475968 |
| lixisenatide | 1440051 |
| metformin | 6809 |
| miglitol | 30009 |
| nateglinide | 274332 |
| pioglitazone | 33738 |
| prandimet | 802660 |
| repaglinide | 73044 |
| rosiglitazone | 84108 |
| saxagliptin | 857974 |
| semaglutide | 1991302 |
| sitagliptin | 593411 |
| tolazamide | 10633 |
| tolbutamide | 10635 |
| troglitazone | 72610 |

**Data Dictionary**

Return subjids for cases and controls with age at first T1D diagnosis for cases.

**Appendix: PRS/GRS models**

1. The PRS/GRS models were developed in two T1D cohorts:

Cohort A: 3,302 T1D cases (1,739 males, 1,560 females, and 3 cases with undetermined sex) and 6,181 controls(3,326 males, 2,840 females, and 15 cases with undetermined sex).\

Cohort B: 3,297 T1D cases (1,744 males, 1,549 females, and 4 cases with undetermined sex) and 6,169 controls (3,339 males, 2,818 females, and 12 cases with undetermined sex).

1. GRS scoring: The GRS scoring was based on the method developed by Sharp et al(2). To acquire the genotype information of all the T1D-GRS2 SNPs, the *HLA* region was additionally imputed by the SNP2HLA software(3). The overlapped SNPs covered across the imputation methods were highly consistent. Consequently, the GRS scores were assessed for its predictive performance by the Area Under the ROC Curve (AUC).
2. Polygenic risk scores (PRS): PRSs were calculated using the Polygenic Risk Score software (PRSice-2)(4). The performance of a series of cutoff of T1D association P-values (including 10-10, 10-9 , 10-8, 10-7, 10-6, 10-5, 10-4, 0.001, 0.01, 0.05, 0.1, 0.2, and 1) for selection of SNP markers was assessed by the Area Under the ROC Curve (AUC). The P-value cutoff with the largest AUC was adopted. The best AUC (0.8607) is seen at the cutoff of P-value≤1E-05, which suggests that stricter cutoff may cause the missing of informative SNPs, while looser may introduce noise by including SNPs with spurious T1D association. Based on the SNP markers with T1D association P-value≤1E-05, a PRS score was acquired for each individual in the independent test cohort.
3. Assessment of GRS/PRS:

GRS/PRS associated with T1D independently.

**Standardized Coefficients (Beta) with T1D**

|  |  |  |
| --- | --- | --- |
| T1D | CohortA | CohortB |
| GRS | 0.501 | 0.5 |
| PRS | 0.283 | 0.28 |

*Cohort A:*

GRS vs PRS: Pearson Correlation 0.331

GRS with T1D: Pearson Correlation 0.595 // ROC AUC= 0.8699188404001987

PRS with T1D: Pearson Correlation 0.449 // ROC AUC= 0.8646341374982104

Linear regression model of both GRS and PRS with T1D: Pearson Correlation 0.671 (better than GRS alone) // ROC AUC= 0.90290338958087

*Cohort B:*

GRS vs PRS: Pearson Correlation 0.305

GRS with T1D: Pearson Correlation 0.585 // ROC AUC= 0.8699188404001987

PRS with T1D: Pearson Correlation 0.432 // ROC AUC= 0.8646341374982104

Linear regression model of both GRS and PRS with T1D: Pearson Correlation 0.658 (better than GRS alone) // ROC AUC= 0.902903389580876

**Issue/limitation:** Models based on European population.

**References:**

1. Upadhyaya SG, Murphree Jr DH, Ngufor CG, Knight AM, Cronk DJ, Cima RR, Curry TB, Pathak J, Carter RE, Kor DJ. Automated diabetes case identification using electronic health record data at a tertiary care facility. Mayo Clinic Proceedings: Innovations, Quality & Outcomes. 2017;1(1):100-10.

2. Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, Schneider DA, Locke JM, Tyrrell J, Weedon MN. Development and standardization of an improved type 1 diabetes genetic risk score for use in newborn screening and incident diagnosis. Diabetes Care. 2019;42(2):200-7.

3. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, Raychaudhuri S, de Bakker PI. Imputing amino acid polymorphisms in human leukocyte antigens. PloS one. 2013;8(6):e64683.

4. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. GigaScience. 2019;8(7). doi: 10.1093/gigascience/giz082.