TYPE 2 DIABETES
KNOWLEDGE PORTAL
type2diabetesgenetics.org

Providing data and tools
to promote understanding and treatment
of type 2 diabetes and its complications

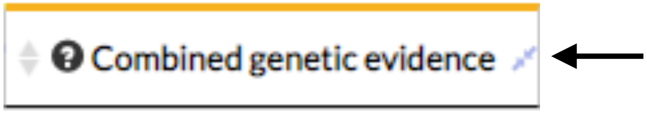
Predicted Type 2 Diabetes Effector Genes

The predictions in this table synthesize multiple kinds of biological evidence to identify and classify genes that are most likely to have direct roles in the development of T2D and may represent potential drug targets. To generate these predictions, Anubha Mahajan and Mark McCarthy (manuscript in preparation) surveyed three types of evidence.

1. GENETIC EVIDENCE

The genetic evidence considered includes evidence about coding variants derived from three recent, comprehensive T2D genetic association studies; evidence from the Online Mendelian Inheritance in Man (OMIM) database on genes involved in monogenic forms of diabetes or related conditions; and evidence from the literature supporting involvement of the gene in diabetes or glycemic traits.

Genetic evidence is combined into an overall score in the **Combined genetic evidence** column of the table. Click the expansion arrows to see the individual types of evidence, described below.


Combined genetic evidence

expansion arrows

The **GWAS coding evidence** column contains a classification of the cumulative posterior probability of association (PPA) for coding variants in credible sets from a GWAS of nearly 900,000 individuals of European ancestry (Mahajan, A., et al. 2018b). **Strong** denotes $\text{PPA} \geq 80\%$; **Medium** denotes $50\% \leq \text{PPA} < 80\%$; **Low** denotes $20\% \leq \text{PPA} < 50\%$. The addition of "glycemic" to a classification indicates that genetic evidence for a non-T2D glycemic trait was found in glycemic trait analyses.

The **Exome array evidence** column contains a classification of the cumulative posterior probability of association (PPA) for coding variants in credible sets from a multi-ancestry exome chip study of nearly 230,000 individuals (Mahajan, A., et al. 2018a). **Strong** denotes $\text{PPA} \geq 80\%$; **Medium** denotes $50\% \leq \text{PPA} < 80\%$; **Low** denotes $20\% \leq \text{PPA} < 50\%$. The addition of "glycemic" to a classification indicates that genetic evidence for a non-T2D glycemic trait was found in glycemic trait analyses.

The **Burden test evidence** column contains a classification of the best gene-level p-value from the extreme p-value aggregation test or the weighted aggregation test performed in an exome sequence analysis study of over 49,000 individuals (Flannick, J., et al. 2019). **Strong** denotes a $\text{p-value} \leq 5x10^{-5}$, while **Medium** denotes $0.05 > \text{p-value} > 5x10^{-5}$.

The **Monogenic associations** column contains evidence from the Online Mendelian Inheritance in Man (OMIM) database that mutation of the gene can confer monogenic diabetes-related phenotypes: monogenic diabetes of the young (MODY); familial early-onset diabetes; juvenile-onset diabetes; neonatal diabetes mellitus (NDM); Wolfram syndrome; neonatal diabetes mellitus with congenital hypothyroidism (NDH); lipodystrophy; insulin resistance; and other diabetes-relevant phenotypes.

The **Other genetic evidence column** contains the PubMed IDs of published papers suggesting the involvement of the gene in diabetes or glycemic traits, many from studies of isolated populations or single ancestries.

Individual pieces of genetic evidence were combined into an overall score.

The combined genetic evidence was considered to be strong (**1C** classification) if any of the following criteria were met:

- **Strong** evidence in the **GWAS coding evidence** or **Exome array evidence** columns
- **Strong** evidence in the **Burden test evidence** column
- strongly associated coding variants are reported in the literature, listed in the **Other genetic evidence** column
- **Medium** or **Low** evidence in the **GWAS coding evidence** or **Exome array evidence** columns **AND** evidence in the **Monogenic associations** column
- Medium evidence in the **Burden test evidence** column **AND** evidence in the **Monogenic associations** column
- gene is within a T2D associated region **AND** there is evidence in the **Monogenic associations** column

The combined genetic evidence was considered to be intermediate (**2C** classification) if any of the following criteria were met:

- **Low** or **Medium** evidence in the **GWAS coding evidence** or **Exome array evidence** columns but no evidence in the **Monogenic associations** column
- **Medium** evidence in the **Burden test evidence** column but no evidence in the **Monogenic associations** column
- evidence in the **Monogenic associations** column, but no coding variant T2D associations

2. REGULATORY EVIDENCE

The regulatory evidence considered includes evidence from several sources bearing on whether there is a T2D- or glycemic trait-associated noncoding variant that influences expression of the gene in a T2D-relevant tissue.

Regulatory evidence is combined into an overall score in the **Combined regulatory evidence** column of the table. Click the expansion arrows to see the individual types of evidence, described below.

The **Islet cis-eQTLs** column lists the PubMed IDs or BioRxiv links for papers containing evidence that a cis-eQTL influences expression of the gene in pancreatic islet tissue.

The **Other relevant cis-eQTLs** column lists the PubMed IDs or BioRxiv links for papers containing evidence that a cis-eQTL influences expression of the gene in fat, muscle, or liver tissue.

The **Islet chromatin conformation** column lists the PubMed IDs or BioRxiv links for papers containing evidence from Capture-C or Hi-C experiments that a T2D-associated variant physically contacts the gene in pancreatic islet tissue.

The **Allelic imbalance** column lists the PubMed IDs or BioRxiv links for papers containing evidence that a T2D-associated variant affects chromatin accessibility in or near the gene.

The **Glucose regulation** column indicates whether expression of the gene was found to be regulated by glucose levels (Ottosson-Laakso et al., 2017). The number 1 indicates glucose regulation, while 0 indicates no glucose regulation.

The **Other regulatory evidence** column lists the PubMed IDs or BioRxiv links for papers containing any other evidence that expression of the gene is responsive to T2D-relevant factors.

Individual pieces of regulatory evidence were combined into an overall score.

The combined regulatory evidence was considered to be intermediate (**2R** classification) if at least 2 pieces of regulatory evidence exist. Evidence within each column was considered to be one piece of evidence, whether or not there are multiple citations, and any cis-eQTL evidence was considered to be one piece of evidence, whether the eQTLs were seen in islets, in other relevant tissues, or in both.

The combined regulatory evidence was considered to be weak (**3R** classification) if one piece of regulatory evidence exists. Evidence within each column was considered to be one piece of evidence, whether or not there are multiple citations.

3. PERTURBATION EVIDENCE

The perturbation evidence considered includes evidence that perturbation of the gene or of its homolog in a model organism confers phenotypes that are relevant to T2D.

Perturbation evidence is combined into an overall score in the **Combined perturbation evidence** column of the table. Click the expansion arrows to see the individual types of evidence, described below.

The **RNA interference evidence** column lists the phenotypes observed by Thomsen et al., 2016 in a human beta-cell line upon using RNA interference to silence the gene.

The **Zebrafish mutant phenotype** column lists any T2D-relevant mutant phenotypes conferred by mutation of the zebrafish homolog of the gene, as curated from the literature by the Zebrafish Information Network curators and retrieved using ZebrafishMine.

The **Mouse mutant phenotype** column lists any T2D-relevant mutant phenotypes conferred by mutation of the mouse homolog of the gene, as curated from the literature at the Mouse Genome Database.

The **Drosophila mutant phenotype** column lists any insulin output-related mutant phenotypes observed upon silencing of the *Drosophila melanogaster* ortholog of the gene, as reported by Peiris et al., 2018.

The **Rat mutant phenotype** column lists any T2D-relevant mutant phenotypes conferred by mutation of the rat homolog of the gene, as curated from the literature at the Rat Genome Database.

The **Other perturbation evidence** column lists the PubMed IDs or BioRxiv links for papers containing evidence that perturbation of the gene or its homolog confers T2D-related phenotypes.

Individual pieces of perturbation evidence were combined into an overall score.

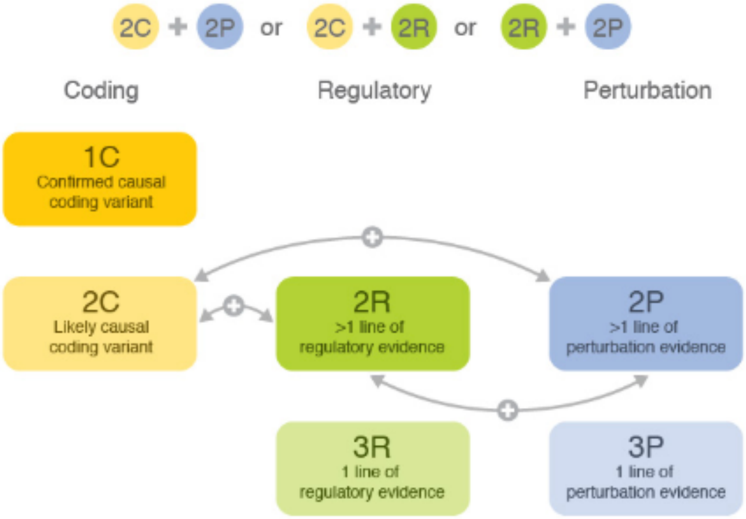
The combined perturbation evidence was considered to be intermediate (**2P** classification) if at least 2 pieces of perturbation evidence exist. Evidence within each column was considered to be one piece of evidence, whether or not there are multiple citations, and equal weight was given to each type of perturbation evidence.

The combined perturbation evidence was considered to be weak (**3P** classification) if one piece of perturbation evidence exists. Evidence within each column was considered to be one piece of evidence, whether or not there are multiple citations.

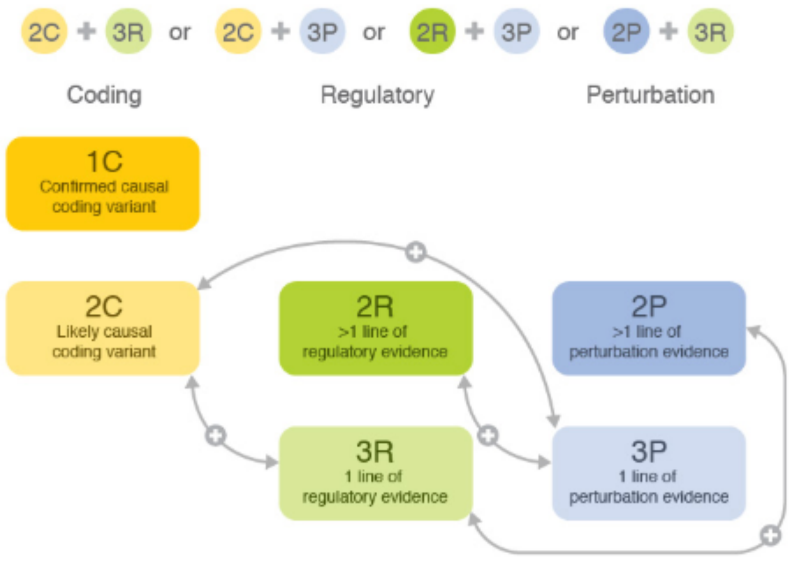
Finally, the classifications for genetic, regulatory, and perturbation evidence were combined into a single classification indicating the likelihood that the gene is an effector for T2D.

CAUSAL genes are those meeting the criteria for **1C** classification (see above).

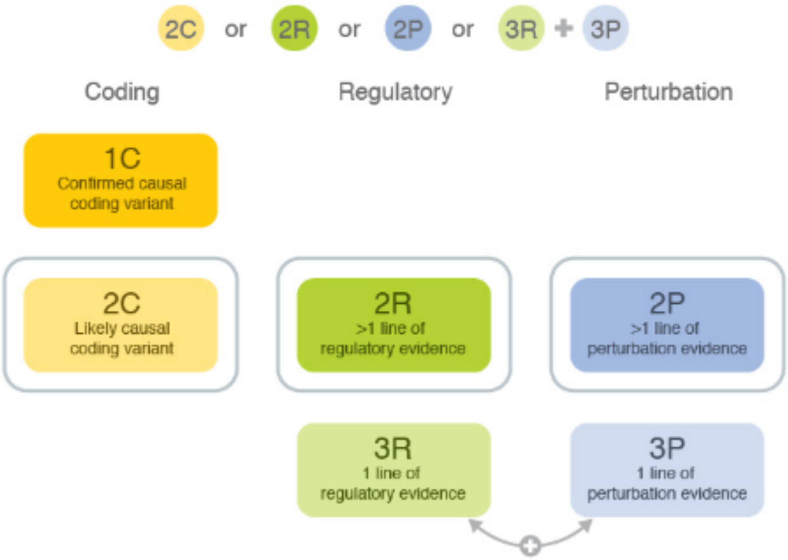
STRONG genes are those with multiple categories of level 2 evidence **(2C+2P**; **2C+2R**; **2R+2P**; see above).



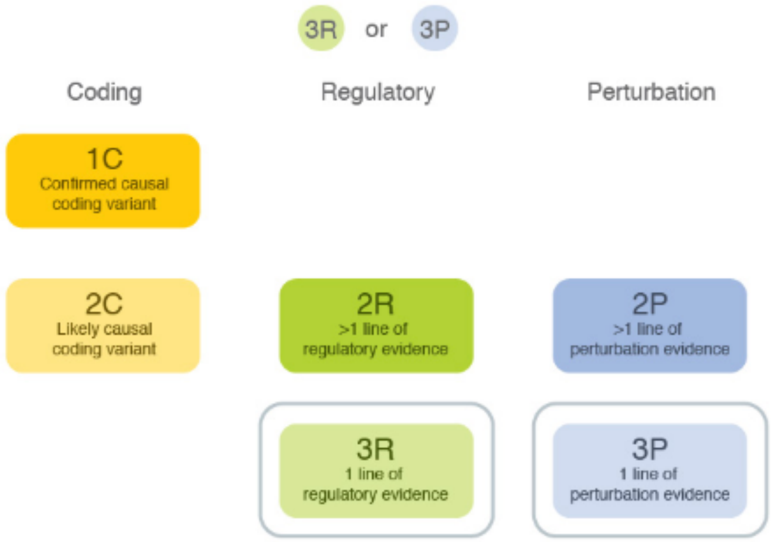
MODERATE genes have some level 2 evidence and have level 3 evidence from a different domain **(2C+3R**; **2C+3P**; **2R+3P**; **2P+3R**; see above).



POSSIBLE genes have level 2 evidence from one domain **OR** have both regulatory and perturbation level 3 evidence **(2C**; **2R**; **2P**; **3R+3P**; see above).



WEAK genes have level 3 evidence from one domain **(3R**; **3P**; see above).



An additional category, **T2D_related**, classifies genes that are not genetically associated with T2D but instead have strong associations with glycemic traits. Evidence for T2D_related genes is collected and combined within each category (genetic, regulatory, perturbation) in the same way as for T2D-associated genes.

These classifications and the evidence behind them will be updated periodically as new information becomes available. Please contact us with feedback on or questions about the table.