Structural Properties of Gene Promoters Highlight More than Two Phenotypes of Diabetes

Constantin Ionescu-Tîrgovişte[1☯], Paul Aurelian Gagniuc[1,2,3☯¤]*, Cristian Guja[1]

1 National Institute of Diabetes, Nutrition and Metabolic Diseases "N.C. Paulescu", Bucharest, Romania, 2 National Institute of Pathology "Victor Babes", Bucharest, Romania, 3 Department of Genetics, University of Bucharest, Aleea Portocalelor 1–3, Sector 6, Bucharest, Romania

☯ These authors contributed equally to this work.
¤ Current Address: Faculty of Engineering in Foreign Languages, Politehnica University of Bucharest, Bucharest, Romania
* paul_gagniuc@acad.ro

Abstract

Genome-wide association studies (GWAS) published in the last decade raised the number of loci associated with type 1 (T1D) and type 2 diabetes (T2D) to more than 50 for each of these diabetes phenotypes. The environmental factors seem to play an important role in the expression of these genes, acting through transcription factors that bind to promoters. Using the available databases we examined the promoters of various genes classically associated with the two main diabetes phenotypes. Our comparative analyses have revealed significant architectural differences between promoters of genes classically associated with T1D and T2D. Nevertheless, five gene promoters (about 16%) belonging to T1D and six gene promoters (over 19%) belonging to T2D have shown some intermediary structural properties, suggesting a direct relationship to either LADA (Latent Autoimmune Diabetes in Adults) phenotype or to non-autoimmune type 1 phenotype. The distribution of these promoters in at least three separate classes seems to indicate specific pathogenic pathways. The image-based patterns (DNA patterns) generated by promoters of genes associated with these three phenotypes support the clinical observation of a smooth link between specific cases of typical T1D and T2D. In addition, a global distribution of these DNA patterns suggests that promoters of genes associated with T1D appear to be evolutionary more conserved than those associated with T2D. Though, the image based patterns obtained by our method might be a new useful parameter for understanding the pathogenetic mechanism and the diabetogenic gene networks.

Introduction

Diabetes mellitus is a heterogeneous syndrome with an onset that can occur from birth to any point in one's lifetime [1]. The hereditary nature of diabetes is long known, but its genetic basis

started to be unravelled only in the 7th decade of the last century [2,3]. It was found that the common phenotypes of diabetes are polygenic and not monogenic, as previously supposed according to the Mendel laws of hereditary. It is not surprising that ~30 monogenic forms of diabetes could be relatively easy identified [4]. Each of these forms have a different clinical phenotype and, frequently, different therapeutic indications [5]. However, their prevalence does not reach ~5% of the total diabetes cases. The gene-sequencing chips using targeted next-generation sequencing allows their quick and efficient detection [5,6].

The identification of the genetic basis for the common polygenic diabetes phenotypes proved to be a much more difficult issue. A deterrent for this was represented by the incoherence of the diabetes classifications over time. Characterization of diabetes phenotypes begun 150 years ago when Etienne Lancereaux (1829–1910), based only on clinical observation corroborated with forensic studies, reached the conclusion that diabetes is not a simple disease but a complex *syndrome*. Based on their features, Lancereaux identified two main clinical forms. Thus, he described the "thin" diabetes (which appears in young age, and it is characterized by a speedy decrease in weight and rapid evolution towards death) and "fat" diabetes (which appears in adults in the presence of obesity, shows a hereditary nature and usually a slow and torpid evolution) [7,8]. Due to its familial nature, the second phenotype was also named as "constitutional" diabetes. All the official classifications proposed by WHO (1965, 1980, 1985 and 1998) derived from these initial observations [9]. Finally, for the two major forms of diabetes, a neutral designation of type 1 (T1D) and type 2 (T2D) diabetes was adopted.

The first important breakthrough for the elucidation of diabetes pathogenesis was represented by the autoimmune-genetic theory of T1D [3,10]. Thus, it was confirmed that diabetes is a polygenic disease and the mechanism of beta cell destruction is immune in nature. From this point, the genetic studies were planned considering that the two major phenotypes of diabetes were two different diseases. Consequently, some researchers focused on the genetics of T1D while others on the genetics of T2D. Usually, pediatric patients were selected for T1D studies [11,12], while for T2D predominantly adult and obese patients were selected. Such a "black and white" vision of diabetes phenotypes led to a tendency in highlighting mainly the differences between the two phenotypes. Moreover, the restrictive selection of patients enrolled in these genetic studies excluded almost all patients with diabetes onset between 20 and 40 years, whose separate analysis could have provided some useful information for a new thinking regarding the classification of diabetes. "Intermediary" or "secondary insulin dependent" diabetes [13], known better as "Latent Autoimmune Diabetes in Adults–LADA" [14–21], placed a *grey zone* in-between the two major phenotypes, which later proved to be associated both with classic T1D and T2D genes [11,22–27].

Genetics of T2D had a rather slow progression during the decade of candidate gene analysis, perhaps due to a not-inspired focus on the putative insulin resistance and not on the β-cell function, its true cause [28,29].

The genetic landscape of the two major diabetes phenotypes included only a couple of genes at the time of the Genome Wide Association (GWA) Scan emergence, awaited with much interest and optimism. GWAs have been able to establish an extended (but only provisory) inventory of the genes associated with T1D (~50) and with T2D (~60). The number of genes associated with T1D and T2D is expected to rise in the near future [30]. However, the discovery of new genes with a significant contribution to the pathogenesis of these phenotypes is less probable.

The current genetic analysis techniques are mainly based on genotyping. Thus, genomic SNPs are tested for their association with one of the two investigated diabetes phenotypes. One major limitation of this technique resides in identifying the causal gene linked to the identified SNP, which can be placed nearby but also at some distance from that SNP [30,31]. The second

constraint of this method seems to be the difficulty in describing the function of encoded proteins for many of these new genes. There is however a hope that these drawbacks will be eliminated in the future [32–35]. The third limitation is represented by a low contribution of recently identified genes to the genetic risk score of the disease [4,30,36–40]. Finally, the fourth limitation is represented by the GWA scan technique itself. Regardless of a potential higher SNP density in the future, it is hard to believe GWAS could identify some new relevant genes associated with these two phenotypes. However, a more precise localization of genes already associated with these two phenotypes is highly expected in the near future.

The current study proposes a new approach to genetic analysis as well as a complementary method to the classical GWAS analysis. Gene promoters have rarely been studied as a whole in relation to this syndrome, although their key role in the expression of genes associated with diabetes may be the root of the issue.

Results and Discussion

Our new analysis method and the available databases enabled us to study a total of 31 promoters of genes associated with T1D (15 promoters) and T2D (16 promoters). The comparative analyses have revealed significant architectural differences between promoters of genes associated with T1D and T2D (p-Value < 0.01). Furthermore, about 16.1% of promoters belonging to T1D and over 19.3% of promoters belonging to T2D have shown some intermediary promoter patterns, suggesting a third diabetes phenotype.

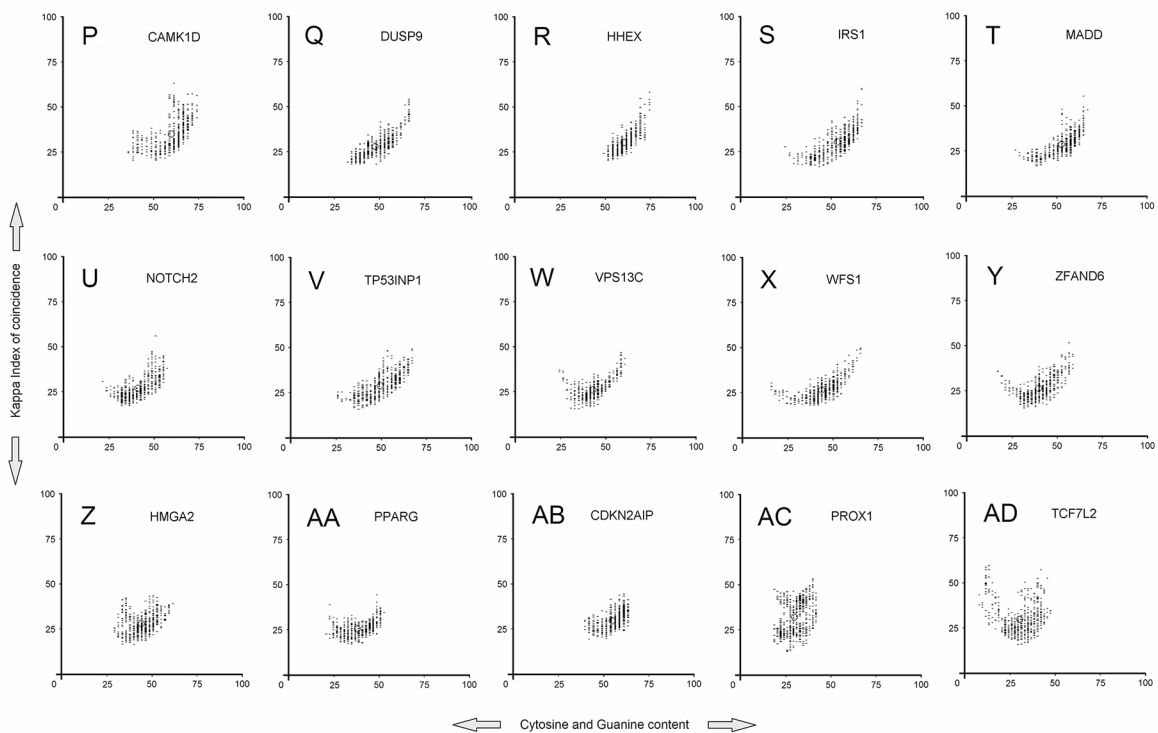### Promoters of genes classically associated with T1D

The T1D associated genes for whom the promoter sequence was available for analysis included *PTPN22* (Protein Tyrosine Phosphatase 22), *TLR7* (Toll Like Receptor 7), *CTLA4* (Cytotoxic T Lynphocyte—assoiciated Antigen 4), *GSDMB* (Gasdermin B), *STAT4* (Signal Transducer and Activator of Transcription 4), *IL7R* (Inter-Leukin 7 Receptor), *C1QTNF6* (C1q - and Tumor Necrosis Factor—related protein 6), *CD55* (complement decay-accelerating factor), CTSH (Cathepsin H), *ERBB3* (V-ERB-B2 Avian Erythroblastic Leukemia Viral Oncogene Homolog 3), *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DPB1* (human leukocyte antigen gene family) and *INS* (Insulin gene) (Fig 1A–1O). The T1D promoters are distinguished by high (A+T)% and low Kappa IC (Kappa Index of Coincidence) values. Accordingly, T1D promoters exhibit image-based patterns characterized by centrally disposed clusters, leading to the formation of narrow shapes (Fig 1A–1O).

The lower region of these patterns contain evenly interspersed nucleotides (A,T,C,G ≈ 25%), while the upper areas show different lengths of short homopolymer tracts (poly(dA), poly(dT), poly(dC), poly(dG)) disposed in tandem in any order. The right side and the left side of these patterns are distributed in a relative 2-fold rotational symmetry. Another general characteristic of these promoters is that the average Kappa IC values are lower than 26%. The exceptions to this apparent rule are represented by *CTSH* (KIC = 31.72%), *INS* (KIC = 28.97%), and *ERBB3* (KIC = 27.35%) gene promoters, which have average Kappa IC values > 26% (S1 File). Also, *CD 55*, *CTSH* and *ERBB3* gene promoters generate slightly more relaxed patterns and show some variations of short poly(dC:dG) homopolymer tracts (Fig 1M–1O). Another interesting observation would be that promoters of *INS* and *C1QTNF6* exhibit different Kappa IC and (C+G)% average values, but have similar pattern shapes, which may suggest a direct connection in their expression (Fig 1K and 1L and S1 File). Moreover, although in accordance with all the properties described above, *HLA-DQA1* promoter also contains an atypical feature, namely additional long poly(dA:dT) homopolymer tracts (Fig 1A).
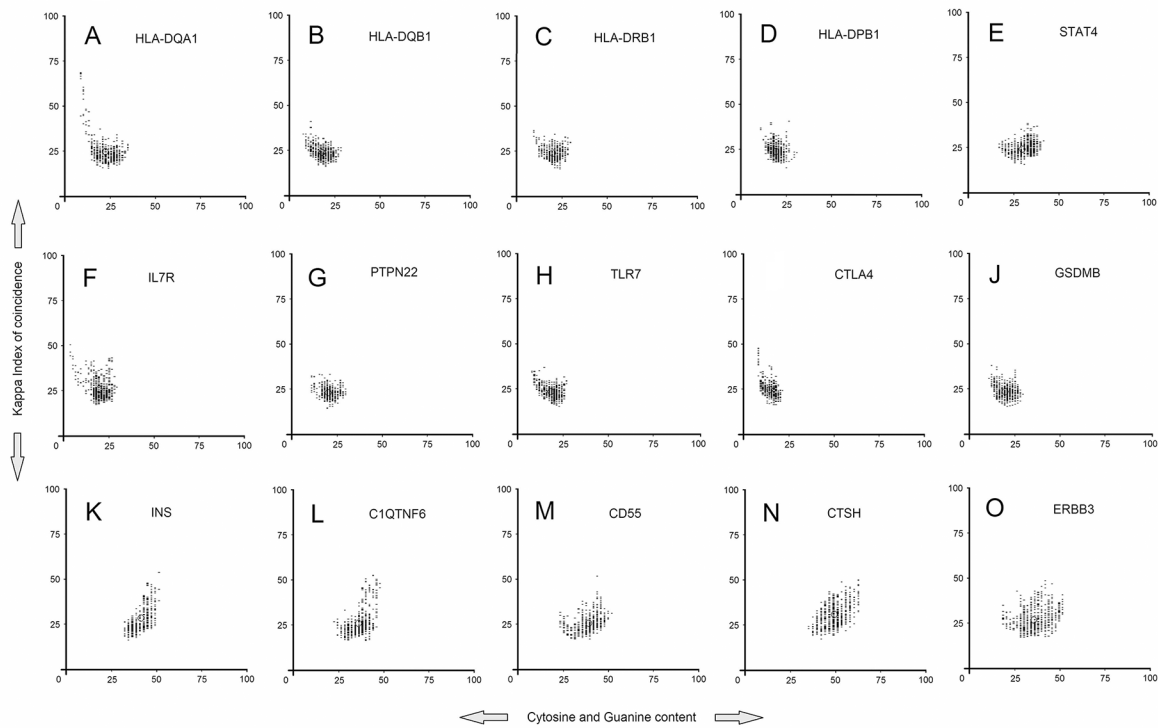
Fig 1. Image-based promoter patterns of genes classically associated with T1D and T2D. (A-O) promoters of genes associated with T1D, (P-AD) promoters of genes associated with T2D. Each black circle represents the center of weight of the promoter pattern.

doi:10.1371/journal.pone.0137950.g001

HLA-DQA1 is also the only promoter that contains sequence areas with Kappa IC values higher than 50%, nevertheless, the average Kappa IC value of the entire promoter remains less than 26% (S1 File). Another feature of T1D gene promoters is represented by specific boundaries, such as (C+G)% values between 50.67% (*CTSH*) and 14.27% (*CTLA4*), and Kappa IC values between 31.72% (*CTSH*) and 23.52% (*PTPN22*). Interestingly, *CTSH* gene promoter contains the highest values, both for Kappa IC and (C+G)% (Fig 2C and S1 File). The average promoter values (Kappa IC = 25.71% and (C+G)% = 27.26) of genes associated with T1D suggests a constant lack of repetitive sequences.

Promoters of genes classically associated with T2D

The T2D associated genes for whom the promoter sequence was available for analysis included *CAMK1*D (Calcium/Calmodulin-Dependent Protein Kinase 1—Delta), *DUSP9* (Dual-Specificity Phosphatase 9), *HHEX* (Hematopoietically Expressed Homeobox), *IRS1* (Insulin Receptor
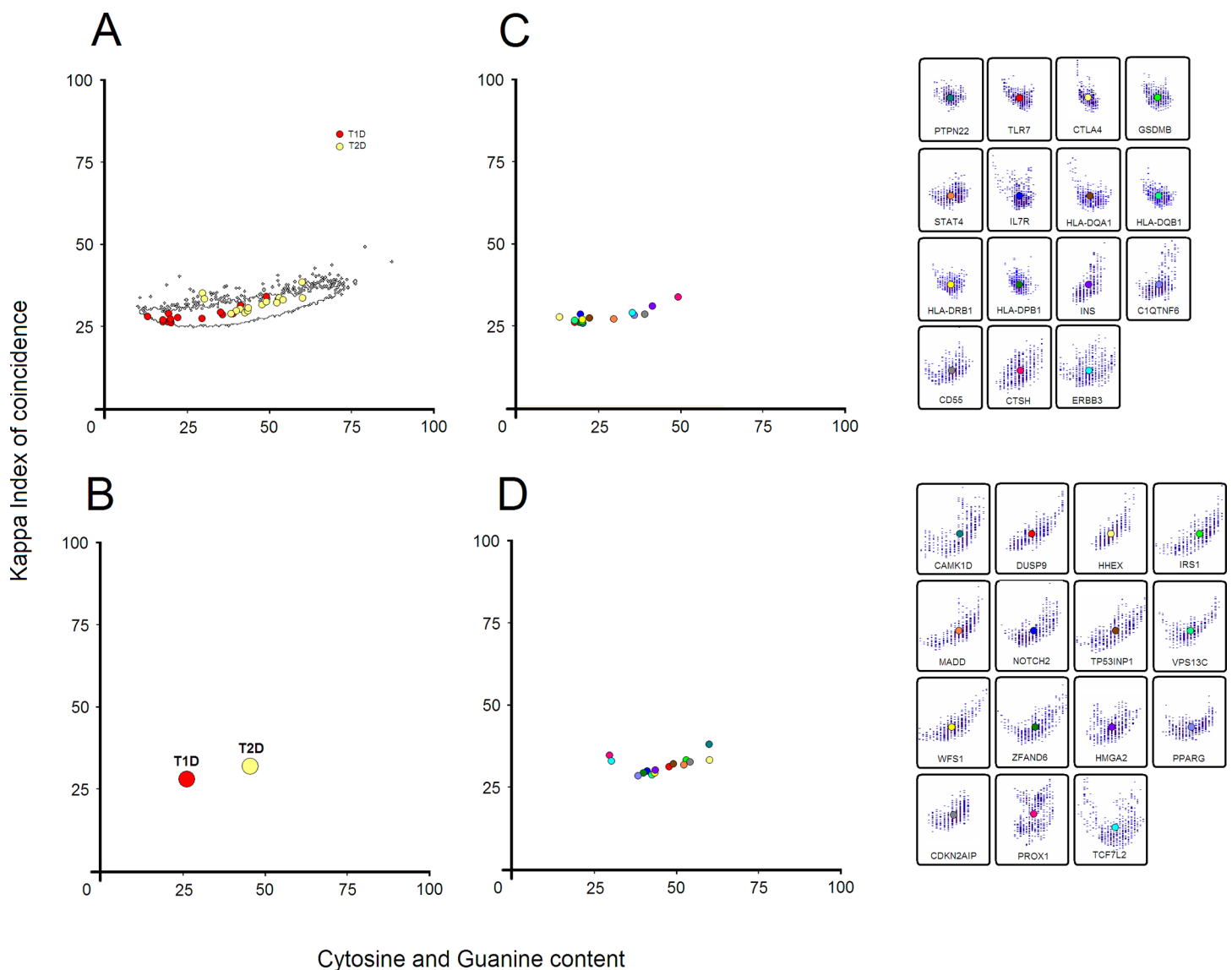
Fig 2. Distribution of T1D and T2D gene promoters. (A) overlapping distribution of T1D and T2D gene promoters on a genome-wide distribution of 8,515 Homo sapiens promoters, (B) mean distribution of T1D and T2D gene promoters, (C) distribution of T1D gene promoters, (D) distribution of T2D gene promoters. Each circle represents the center of weight from a promoter pattern and the circle color is associated with a corresponding gene promoter.

doi:10.1371/journal.pone.0137950.g002

Substrate 1), *MADD* (Multiple Acyl-CoA Dehydrogenase Deficiency), *NOTCH2* (Drosophyla Homolog 2 of *NOTCH*), *TP53INP1* (Tumor Protein p53—Inducible Nuclear Protein 1), *VPS13C* (Vacuolar Protein Sorting 13 C), *WFS1* (Wolfram Syndrome 1), *ZFAND6* (Zinc Finger And Domain-Containing Protein 6), *HMGA2* (High Mobility Group At-Hook 2), *PPARG* (Peroxisome Proliferator Activated Receptor Gamma), *CDKN2A* (Cyclin-Dependent Kinase Inhibitor 2A), *PROX1* (Prospero-Related Homeobox 1) and *TCF7L2* (Transcription Factor 7 Like 2) ([Fig 1P-AD](#)).

The T2D promoters are distinguished by high (C+G)% and high Kappa IC values. T2D promoters are represented by image-based patterns containing a high percentage of C+G, a high CpG content and high Kappa IC values. The right side of the pattern is predominant while the left side is significantly less pronounced. The bi-dimensional shape of these patterns exhibits various different lengths of short poly(dC:dG) homopolymer tracts.

Unlike T1D promoters, T2D promoter patterns have a wide range in C+G content which confer a prolonged shape on the X-axis. In contrast to T1D promoters, the general characteristic of these promoters is that the average Kappa IC values are higher than 26% ([Fig 2B](#)). Overall, the promoters of genes associated with T2D exhibit a significantly higher Kappa IC average value (29.30%) than promoters of genes associated with T1D (25.71%). The T2D promoters that leap well above the genome-wide Kappa IC average belong to *PROX1*, *TCF7L2* and *CAMK1D* genes ([Fig 2A](#)).

Promoters that do not meet the common criteria of the two phenotypes belong to *PROX1* and *TCF7L2* genes. *TCF7L2* and *PROX1* gene promoters are the most interesting ones because they have a very different structure from that described for T2D so far ([Fig 1AC–1AD](#)— bottom right panels). Thus, from a total of 8,515 promoters available in EPD (The Eukaryotic Promoter Database), we found only a few image-based patterns relatively similar (ie. *SOX5* promoter) to *PROX1* promoter ([Fig 1AC](#) and [S2 File](#)). However, promoters of *TCF7L2* and *PROX1* genes comply with the (C+G)% variations found in other gene promoters associated with T2D ([S1 File](#)).

The promoter of *CD55* gene associated with T1D, and the promoters of *PPARG* and *CDKN2AIP* genes associated each with T2D phenotype, appear to have common image-based pattern shapes ([Fig 1M, 1AA and 1AB](#)). The promoters of *CD55* (Kappa IC = 26.03 and (C+G)% = 39.85) and *PPARG* (Kappa IC = 26.22 and (C+G)% = 39.04) genes are even more similar in structure since they share relatively the same Kappa IC (SD = 0.13) and (C+G)% (SD = 0.56) average values ([S1 File](#)). This similarity may suggest common mechanisms of action in their biological pathway, possibly even a direct link in their gene expression. T2D gene promoters located at the extremes are represented by (C+G)% values between 61.90% (*HHEX*) and 30.39% (*PROX1*), whereas Kappa IC values range between 35.74% (*CAMK1D*) and 26.22% (*PPARG*).

The average promoter values (KIC = 29.30% and (C+G)% = 46.91) of genes associated with T2D further show a constant presence of short poly(dC:dG) homopolymer tracts. Nevertheless, perhaps the best noticeable differences are represented by the median values shown by T1D and T2D gene promoters ([Fig 2B](#)). The median Kappa IC values presented by promoters of both phenotypes (T1D = 25.29%, T2D = 29.02%) are relatively less significant than those shown by (C+G)% median values (T1D = 20.50%, T2D = 44.79%).

The "intermediary" phenotype of diabetes

A third, frequently encountered, phenotype of diabetes includes those patients designated along the years as "Lady-like" [41], "Type 1 1/2 diabetes" [42], "type 1.5" [43], etc. [44,45]. From a clinical point of view, this diabetes phenotype makes a smooth transition between

classical "black" (T1D) and "white" (T2D) phenotypes. We designate this phenotype as Intermediary Diabetes Mellitus (IDM).

Several genes investigated by us were reported to be associated with both T1D and T2D or have an unknown function. For instance, although *C1QTNF6* and *CTSH* genes are associated with T1D, they have an uncertain function in the context of diabetes [46]. Our analysis shows that promoter patterns of these two genes belong to an intermediate zone. Genes such as *INS* [47,48,49], *TCF7L2* [22,24,50], *SLC30A8* [23,32] or *PPARG* [12,49] have already been associated with both major phenotypes. Interestingly, promoters of these genes show the same intermediate character (Fig 3C), thus, confirming these associations. Other genes such as *ERBB3* [46,51], *HMGA2* [52,53] or *PROX1* [54–56] also have promoters that generate intermediate patterns.

Considering the previously reported association of these genes with both phenotypes, their uncertain function and the pattern of their promoters, we can distinguish this Intermediary Diabetes Mellitus phenotype (Fig 3ABC). A general distribution of the three phenotypes shows that promoter patterns of T2D and IDM overlap (Fig 3DEF and Fig 4). However, promoter patterns of T1D do not overlap with IDM or T2D, suggesting a different pathological mechanism between T1D and IDM, but a more direct link between T2D and IDM (Fig 4A). On the other hand, extrapolating the general distribution (Fig 3DEF and Fig 2) and considering the promoter pattern shapes (Fig 3ABC), we also suggest that any gene in IDM may be a driver gene for either T1D or T2D.

The structure of promoters and evolution

We previously reported the existence of about 10 different types of promoters in eukaryotes [57]. In general, promoters that belong to older species show lower Kappa IC and C+G values. Although unexpected, promoters of genes associated with T1D show similar structures to those that we generally have found in older eukaryotic species. Thus, a general distribution of these patterns suggests that promoters of genes associated with T1D appear to be evolutionary more conserved than those associated with T2D. This similarity in shape and angle between promoter patterns of older species and T1D is in agreement with this phenotype, which is often incompatible with life from an early age, and it was easily subjected in the past to natural selection. Furthermore, even more interesting is that T1D and T2D promoters fall into only two classes of promoters from a total of 10, namely "ATCG-compact" class (for T1D) and "CG-based" class (for T2D). In our previous studies we concluded that similar promoter patterns use similar transcription factors, which further involve different driver genes for each phenotype [57].

Interestingly, promoters of genes included in the third phenotype (IDM) have patterns that fall into several promoter classes, which may use transcription factors from both phenotypes. Furthermore, our current study suggests that gene promoters prone to methylation are those associated with T2D and IDM, but not those associated with T1D. It seems that diabetogenic influences leading to the increased prevalence of these diabetes phenotypes may act through different epigenetic mechanisms. The lack of CpG sites in the promoters of genes associated with T1D, in contrast with the high frequency of CpGs in the gene promoters associated with T2D or IDM, suggests that the mechanism of their increasing prevalence in the modern society may act through totally different pathways.

Materials and Methods

In our approach we used 31 promoter sequences (15 promoters from T1D and 16 promoters from T2D) obtained from Eukaryotic Promoter Database (EPD) and HomoloGene. To unravel
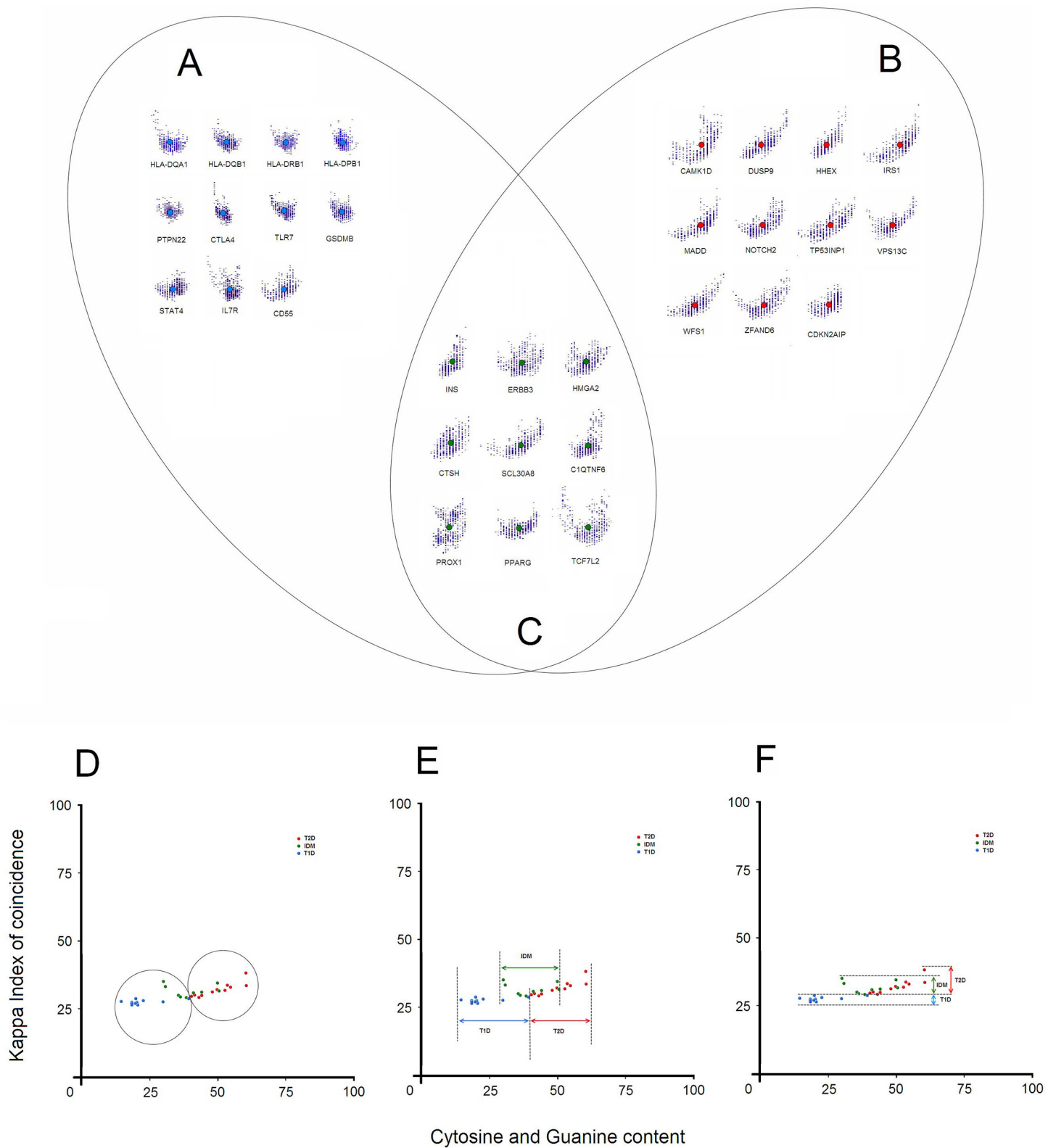
Fig 3. Redistribution of promoters according to T1D, IDM and T2D. (A) T1D promoters (blue dots), (B) T2D promoters (red dots) and (C) promoters from genes associated with the "intermediary" phenotype (green dots). (D) Full separation of T1D and T2D after the inclusion of the intermediary phenotype, (E) C +G values of intermediary phenotype areas overlapping equally in both T1D and T2D regions, (F) Kappa IC values of intermediary phenotype overlapping only with T2D areas.
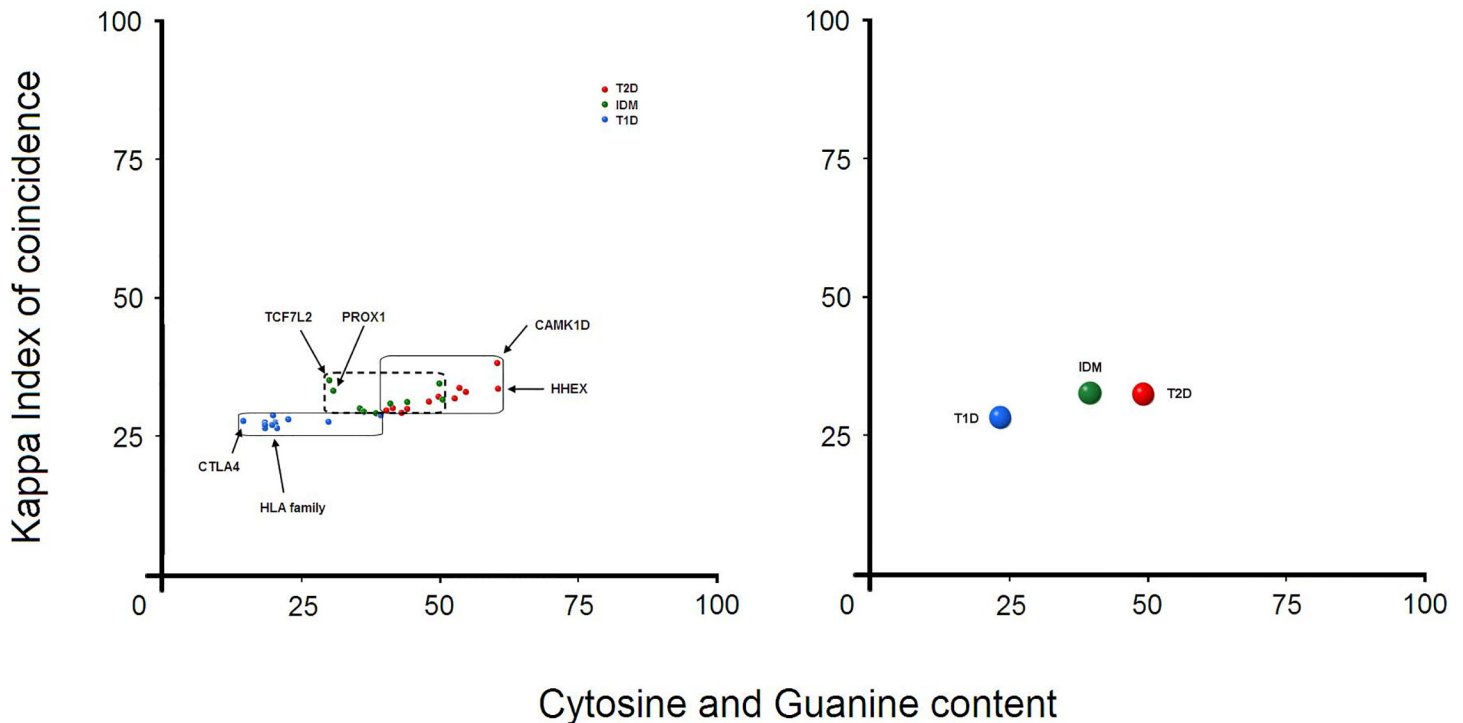
Fig 4. General distribution of the three phenotypes. (A) Complete separation of T1D and T2D after the inclusion of the intermediary phenotype and (B) the global distribution of the three phenotypes.

the design principles of these promoter architectures, we have used Visual Basic to develop a software program for promoter analysis—called PromKappa (Promoter analysis by Kappa), recently published [57–59]. In brief, we used a sliding window approach (window size of 30 nucleotides (nt) and a step of 1 nt) to extract two types of values, namely Kappa IC and (C +G)%. Kappa IC values were plotted on a graph against (C+G)% values, which formed a recognizable promoter pattern for each promoter sequence (S3 File). A promoter pattern is an image that consists of 470 lines, whose coordinates have been plotted observing the two values extracted from each sliding window (Fig 5A, 5B, 5C and 5D). The shape of a pattern is composed from various sized clusters of lines on the y-axis (Fig 5D). The pattern colors range from blue to red according to the number of overlapping lines. Unlike sequence alignment algorithms, our method uses a comparison between the frequency and the nucleotide content of a promoter sequence, thereby measuring the degree of randomization of a DNA sequence [58]. The center of weight of 8,515 promoter patterns were plotted on a second graph in order to show the distribution boundaries of promoters in the human genome (Fig 5D). Next, on this distribution we superimposed the promoter locations of genes associated with T1D and T2D. For a confrontation with the promoters found in genes associated with diabetes, we show a total of 10 possible classes of gene promoters in eukaryotes (Fig 5E), found in our previous study [57].

Promoters from *INS* (insulin gene), *FTO* (fat tissue and obesity associated) and *CTLA4* (cytotoxic T lymphocyte associated antigen 4) genes, which were not found in the EPD database (composed of 8,515 Homo Sapiens promoter sequences), were extracted from HomoloGene (500 bp genomic regions upstream of the gene). Thus, promoters found in HomoloGene were introduced in the EPD data base file. Furthermore, available EPD promoters were
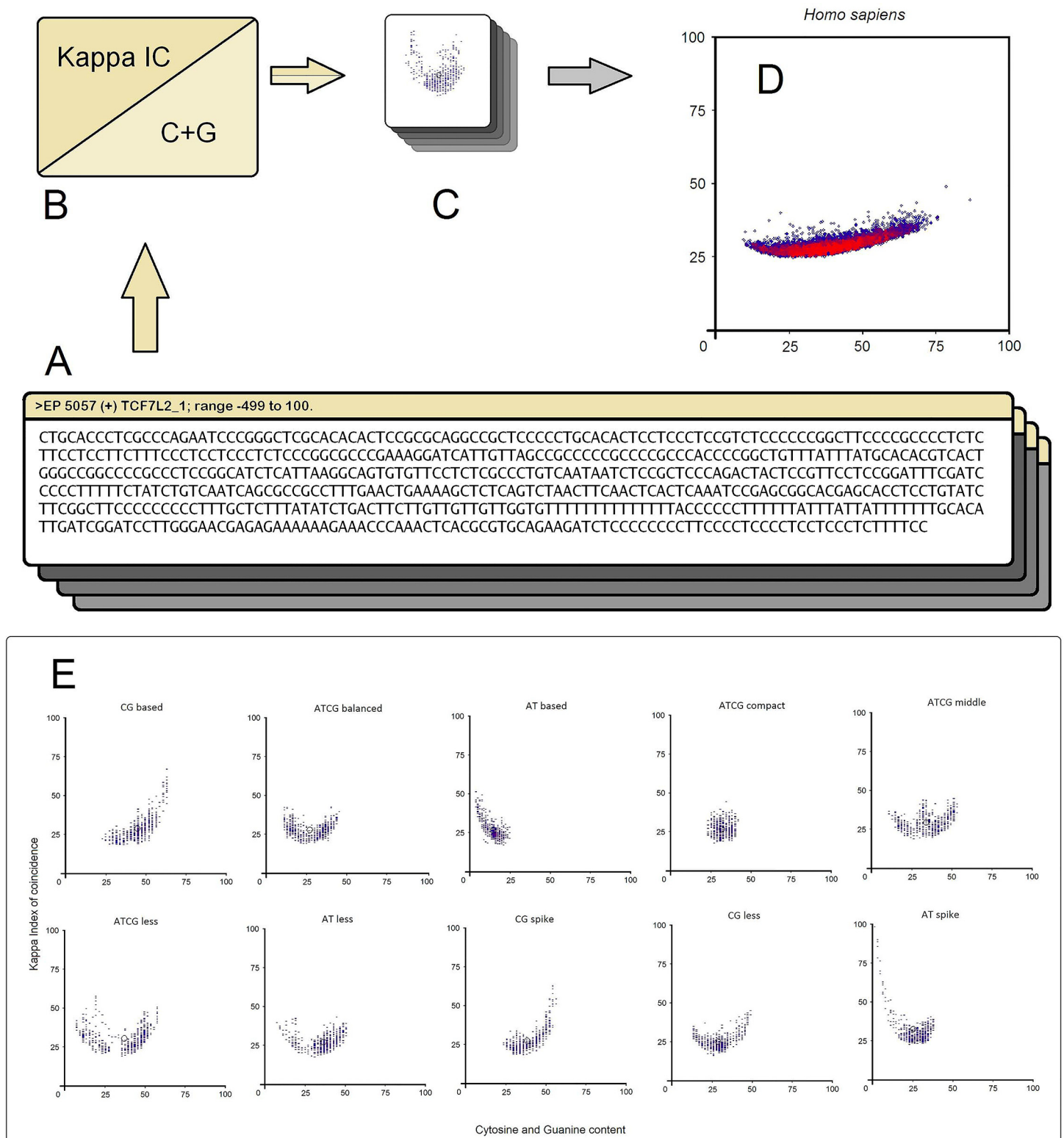
Fig 5. Schematic overview of the promoter analysis. (A) promoter sequences, (B) Kappa IC and (C+G)% values are extracted from each sliding window, (C) sliding window values plotted on a graph which shows a recognizable image-based pattern for each promoter sequence, (D) the center of weight of each promoter pattern plotted on a second graph in order to show the distribution of 8,515 promoters. Red color areas represent denser clusters of promoters. (E) The representative eukaryotic promoter classes are shown in the following sections: AT-based class, CG-based class, ATCG-compact class, ATCG-balanced class, ATCG-middle class, ATCG-less class, AT-less class, CG-spike class, CG-less class and ATspike class [57–59].

doi:10.1371/journal.pone.0137950.g005

confronted with HomoloGene genomic regions (500 bp) upstream of genes associated with T1D and T2D phenotypes in order to ensure their accuracy.

Kappa Index of Coincidence

The Index of Coincidence (IC) principle derives from cryptography and has been used in the analysis of ciphertext. Kappa Index of Coincidence is a modified form of IC, adapted for the analysis of a single DNA sequence [57–59]. Here, Kappa IC algorithm has been used primarily as a unit of measure for the information contained in the DNA of the promoter regions.

Thus, Kappa IC is used for calculating the level of "randomization" of a DNA sequence. Kappa IC is sensitive to various degrees of sequence organization such as simple sequence repeats (SSRs) or short tandem repeats (STRs). The formula for Kappa IC is shown below, where sequences A and B have the same length N. Only if an $A[i]$ nucleotide from sequence A matches the $B[i]$ correspondent from sequence B, then Σ is incremented by 1.

$$Kappa_{IC} = \frac{\sum_{i=1}^{N}[A_i = B_i]}{N/C}$$

The same method for measuring the Index of Coincidence has been applied for only one sequence, in which the sequence was actually compared with itself, as shown below.

```
function KIC(A)
T = 0
N = length(A) - 1
  for u = 1 to N
    B = A[ u + 1] ... A[ N]
    for i = 1 to length(B)
       If A[ i] = B[ i] then C = C + 1
    next i
    T = T + (C / length(B) × 100)
    C = 0
  next u
IC = Round((T / N), 2)
end function
```

Where N is the length of the sliding window, A represents the sliding window content, B contains all variants of sequences generated from A (from u+1 to N), C counts the number of coincidences occurring between B sequence and A sequence and T counts the total number of coincidences between B sequences and A sequence.

C+G content

We extracted $C+G$ values from each sliding window considering the nucleotide frequencies from the entire promoter sequence. In the first stage, to determine the (C+G)% content for the entire (Total = TOT) promoter sequence we used the formula:

$$CG_{TOT} = \left(\frac{100}{(A + T + C + G)_{TOT}}\right) \times (C + G)_{TOT}$$

Where CG_{TOT} represents the percentage of cytosine and guanine from the promoter sequence, $(A+T+C+G)_{TOT}$ represents the sum of the number of occurrences in the promoter sequence of A, T, C and G, and $(C+G)_{TOT}$ represents the sum of the number of occurrences in the promoter sequence of C and G. In the next stage we used the value of CG_{TOT} to calculate the (C+G)%

content from the sliding window (*sw*):

$$CG_{SW} = \left(\frac{CG_{TOT}}{(A + T + C + G)_{SW}} \right) \times (C + G)_{SW}$$

Where CG_{SW} represents the percentage of cytosine and guanine from the sliding window. These promoter patterns are relative to the percentage of C+G of the entire promoter sequence. In this regard, CG_{SW} value is relative to CG_{TOT}. The expression $(A+T+C+G)_{TOT}$ represents the sum of the number of occurrences of A, T, C and G from the sliding window sequence. $(C +G)_{SW}$ represents the sum of the number of occurrences of C and G in the sliding window sequence.

Conclusions

A third diabetes phenotype, known as double diabetes or 1.5 diabetes, is often observed in clinical practice. The results of our genetic analysis objectivly suports this view, showing that this *third phenotype* makes a smooth passage from T1D to T2D. It is interesting to note that Kappa IC values of IDM overlap with T2D but not with T1D. These genetic particularities may explain the difficulties of classifying some diabetic patients in the two "traditional" diabetes phenotypes. We have shown that the number of different phenotypes of diabetes is higher than two and the existence of IDM is objectively supported by our data. The third phenotype has itself two sub-phenotypes corresponding with several clinical particularities. Thus, in the near future the number of diabetes phenotypes is expected to increase, representing a strong impetus for a new classification of diabetes.

Supporting Information

S1 File. Numeric data with reference to the analysis of promoters.
(XLSX)

S2 File. PromKappa Diabetes software.
(ZIP)

S3 File. The sequences of ~ 8000 *Homo Sapiens* promoters.
(ZIP)

Acknowledgments

Author Contributions

Conceived and designed the experiments: CIT PAG CG. Performed the experiments: CIT PAG CG. Analyzed the data: PAG. Contributed reagents/materials/analysis tools: PAG CIT. Wrote the paper: PAG CIT. Conceived of the study and participated in its design and coordination: CIT PAG. Created the algorithms and the software used in the analysis: PAG. Carried out the assembly of promoter files and manually tested the correctness of each promoter sequence: CIT CG. Participated in the promoter sequence analysis and drafted the manuscript: PAG CIT CG. Verified the accuracy of the data and repeated the experiment independently: CIT PAG CG. Discussed the results and commented on the manuscript: PAG CIT CG.

References

1. Hanafusa T, Imagawa A. Fulminant type 1 diabetes: a novel clinical entity requiring special attention by all medical practitioners. Nat Clin Pract Endocrinol Metab. 2007; 3(1):36–45. PMID: 17179928

2. Singal DP, Blajchman MA. Histocompatibility (HL-A) antigens, lymphocytotoxic antibodies and tissue antibodies in patients with diabetes mellitus. Diabetes. 1973; 22(6):429–432. PMID: 4541338

3. Nerup J, Platz P, Andersen OO, Christy M, Lyngsoe J, Poulsen JE, et al. HL-A antigens and diabetes mellitus. Lancet. 1974; 2(7885):864–866.

4. Ellard S, Lango Allen H, De Franco E, Flanagan SE, Hysenaj G, Colclough K, et al. Improved genetic testing for monogenic diabetes using targeted next-generation sequencing. Diabetologia. 2013; 56 (9):1958–1963. doi: 10.1007/s00125-013-2962-5 PMID: 23771172

5. Vaxillaire M, Froguel P. Monogenic diabetes in the young, pharmacogenetics and relevance to multifactorial forms of type 2 diabetes. Endocr Rev. 2008; 29(3):254–264. doi: 10.1210/er.2007-0024 PMID: 18436708

6. Tsai PC, Spector TD, Bell JT. Using epigenome-wide association scans of DNA methylation in age-related complex human traits. Epigenomics. 2012; 4(5):511–526. doi: 10.2217/epi.12.45 PMID: 23130833

7. Lancereaux E. Note et réflexions sur deux cas de diabète sucré avec alteration du pancrèas. Bull. Acad. Méd. 2-e série, tom VI, p. 1215, 1877.

8. Lancereaux E. Le diabète maigre: ses symptomes, son évolution, son prognostic et son traitment; ses rapports avec les altérations du pancréas. Etude comparative du diabéte maigre et du diabéte gras. In: Delahaye A, Lecrosnier E, editors. Leçons de Clinique Médicale. Paris: 1883. pp. 5–16.

9. Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med. 1998; 15(7):539–553. PMID: 9686693

10. Bottazzo GF, Florin-Christensen A, Doniach D. Islet-cell antibodies in diabetes mellitus with autoimmune polyendocrine deficiencies. Lancet. 1974; 2(7892):1279–1283. PMID: 4139522

11. Eftychi C, Howson JM, Barratt BJ, Vella A, Payne F, Smyth DJ, et al. Analysis of the type 2 diabetes-associated single nucleotide polymorphisms in the genes IRS1, KCNJ11, and PPARG2 in type 1 diabetes. Diabetes. 2004; 53(3):870–873. PMID: 14988278

12. Raj SM1, Howson JM, Walker NM, Cooper JD, Smyth DJ, Field SF, et al. No association of multiple type 2 diabetes loci with type 1 diabetes. Diabetologia. 2009; 52(10):2109–2116. doi: 10.1007/s00125-009-1391-y PMID: 19455305

13. Mincu I, Ionescu-Tîrgovişte C, Ghişe-Beer E. Classification of clinically manifest diabetes mellitus as related to etiological factors. Rev Med Interna Neurol Psihiatr Neurochir Dermatovenerol Med Interna. 1975; 27(6):403–414. PMID: 3833

14. Groop LC, Bottazzo GF, Doniach D. Islet cell antibodies identify latent type I diabetes in patients aged 35–75 years at diagnosis. Diabetes. 1986; 35(2):237–241. PMID: 3510930

15. Groop LC, Eriksson J, Ekstrand A, Franssila-Kallunki A, Saloranta C, Miettinen A. Metabolic characteristics of autoimmune diabetes mellitus in adults. Diabetologia. 1991; 34(1):46–51. PMID: 2055339

16. Gale EA. Latent autoimmune diabetes in adults: a guide for the perplexed. Diabetologia. 2005; 48 (11):2195–2199. PMID: 16193287

17. Lin J, Zhou ZG, Wang JP, Zhang C, Huang G. From Type 1, through LADA, to type 2 diabetes: a continuous spectrum?. Ann N Y Acad Sci. 2008; 1150:99–102. doi: 10.1196/annals.1447.036 PMID: 19120276

18. Tuomi T, Groop LC, Zimmet PZ, Rowley MJ, Knowles W, Mackay IR. Antibodies to glutamic acid decarboxylase reveal latent autoimmune diabetes mellitus in adults with a non-insulin-dependent onset of disease. Diabetes. 1993; 42(2):359–362. PMID: 8425674

19. Fourlanos S, Dotta F, Greenbaum CJ, Palmer JP, Rolandsson O, Colman PG, et al. Latent autoimmune diabetes in adults (LADA) should be less latent. Diabetologia. 2005; 48(11):2206–2212. PMID: 16193284

20. Naik RG, Brooks-Worrell BM, Palmer JP. Latent autoimmune diabetes in adults. J Clin Endocrinol Metab. 2009; 94(12):4635–4644. doi: 10.1210/jc.2009-1120 PMID: 19837918

21. Redondo MJ. LADA: time for a new definition. Diabetes. 2013, 62(2):339–340. doi: 10.2337/db12-1171 PMID: 23349539

22. Cervin C, Lyssenko V, Bakhtadze E, Lindholm E, Nilsson P, Tuomi T, et al. Genetic similarities between latent autoimmune diabetes in adults, type 1 diabetes, and type 2 diabetes. Diabetes. 2008; 57 (5):1433–1437. doi: 10.2337/db07-0299 PMID: 18310307

23. Wenzlau JM, Juhl K, Yu L, Moua O, Sarkar SA, Gottlieb P, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. Proc Natl Acad Sci U S A. 2007; 104 (43):17040–17045. PMID: 17942684

24. Pettersen E, Skorpen F, Kvaløy K, Midthjell K, Grill V. Genetic heterogeneity in latent autoimmune diabetes is linked to various degrees of autoimmune activity: results from the Nord-Trøndelag Health Study. Diabetes. 2010; 59(1):302–310. doi: 10.2337/db09-0923 PMID: 19833889

25. Djekic K1, Mouzeyan A, Ipp E. Latent autoimmune diabetes of adults is phenotypically similar to type 1 diabetes in a minority population. J Clin Endocrinol Metab. 2012; 97(3):E409–E413. doi: 10.1210/jc.2011-2544 PMID: 22205710

26. Cleland SJ, Fisher BM, Colhoun HM, Sattar N, Petrie JR. Insulin resistance in type 1 diabetes: what is 'double diabetes' and what are the risks?. Diabetologia. 2013; 56(7):1462–1470. doi: 10.1007/s00125-013-2904-2 PMID: 23613085

27. Teupe B, Bergis K. Epidemiological evidence for "double diabetes". Lancet. 1991; 337(8737):361–362.

28. Ionescu-Tîrgovişte C. Insulin Resistance—What is Myth and What is Reality?. Acta Endo (Buc). 2011; 7 (1): 123–146.

29. Guja C, Gagniuc P, Ionescu-Tîrgovişte C. Genetic Factors Involved In The Pathogenesis Of Type 2 Diabetes. Proc. Rom. Acad. Series B. 2012; 14(1):44–61.

30. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. Diabetes Care. 2013; 36(5):1413–1421. doi: 10.2337/dc12-2211 PMID: 23613601

31. Dayeh TA, Olsson AH, Volkov P, Almgren P, Rönn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. Diabetologia. 2013; 56(5):1036–1046. doi: 10.1007/s00125-012-2815-7 PMID: 23462794

32. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007; 445(7130):881–885. PMID: 17293876

33. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007; 316 (5829):1336–1341. PMID: 17463249

34. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007; 316 (5829):1341–1345. PMID: 17463248

35. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet. 2007; 39(6):770–775. PMID: 17460697

36. Weedon MN, McCarthy MI, Hitman G, Walker M, Groves CJ, Zeggini E, et al. Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med. 2006; 3 (10):e374. PMID: 17020404

37. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, et al. Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes. 2008; 57(11):3122–3128. doi: 10.2337/db08-0425 PMID: 18694974

38. Lango H, UK Type 2 Diabetes Genetics Consortium, Palmer CN, Morris AD, Zeggini E, Hattersley AT, et al. Assessing the combined impact of 18 common genetic variants of modest effect sizes on type 2 diabetes risk. Diabetes. 2008; 57(11):3129–3135. doi: 10.2337/db08-0504 PMID: 18591388

39. Meigs JB. Prediction of type 2 diabetes: the dawn of polygenetic testing for complex disease. Diabetologia. 2009; 52(4):568–570. doi: 10.1007/s00125-009-1296-9 PMID: 19214469

40. Grant RW, O'Brien KE, Waxler JL, Vassy JL, Delahanty LM, Bissett LG, et al. Personalized genetic risk counseling to motivate diabetes prevention: a randomized trial. Diabetes Care. 2013; 36(1):13–19. doi: 10.2337/dc12-0884 PMID: 22933432

41. Lohmann T, Nietzschmann U, Kiess W. "Lady-like": is there a latent autoimmune diabetes in the young?. Diabetes Care. 2000; 23(11):1707–1708. PMID: 11092301

42. Juneja R, Palmer JP. Type 1 1/2 diabetes: myth or reality?. Autoimmunity. 1999; 29(1):65–83. PMID: 10052687

43. Palmer JP, Hirsch IB. What's in a name: latent autoimmune diabetes of adults, type 1.5, adult-onset, and type 1 diabetes. Diabetes Care. 2003; 26(2):536–538. PMID: 12547896

44. Pozzilli P, Di Mario U. Autoimmune diabetes not requiring insulin at diagnosis (latent autoimmune diabetes of the adult): definition, characterization, and potential prevention. Diabetes Care. 2001; 24 (8):1460–1467. PMID: 11473087

45. Leslie RD, Pozzilli P. Type I diabetes masquerading as type II diabetes. Possible implications for prevention and treatment. Diabetes Care. 1994; 17(10):1214–1219. PMID: 7821148

46. Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. N Engl J Med. 2009; 360(16):1646–1654. doi: 10.1056/NEJMra0808284 PMID: 19369670

47. Huxtable SJ, Saker PJ, Haddad L, Walker M, Frayling TM, Levy JC, et al. Analysis of parent-offspring trios provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally transmitted class III variable number tandem repeat alleles. Diabetes. 2000; 49(1):126–130. PMID: 10615960

48. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, et al. Remapping the insulin gene/IDDM2 locus in type 1 diabetes. Diabetes. 2004; 53(7):1884–1889. PMID: 15220214

49. Desai M, Zeggini E, Horton VA, Owen KR, Hattersley AT, Levy JC, et al. The variable number of tandem repeats upstream of the insulin gene is a susceptibility locus for latent autoimmune diabetes in adults. Diabetes. 2006; 55(6):1890–1894. PMID: 16731859

50. Wang J, Kuusisto J, Vänttinen M, Kuulasmaa T, Lindström J, Tuomilehto J, et al. Variants of transcription factor 7-like 2 (TCF7L2) gene predict conversion to type 2 diabetes in the Finnish Diabetes Prevention Study and are associated with impaired glucose regulation and impaired insulin secretion. Diabetologia. 2007; 50(6):1192–1200. PMID: 17437080

51. Keene KL, Quinlan AR, Hou X, Hall IM, Mychaleckyj JC, Onengut-Gumuscu S, et al. Evidence for two independent associations with type 1 diabetes at the 12q13 locus. Genes Immun. 2012; 13(1):66–70. doi: 10.1038/gene.2011.56 PMID: 21850031

52. Saxena R, Saleheen D, Been LF, Garavito ML, Braun T, Bjonnes A, et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. Diabetes. 2013; 62(5):1746–1755. doi: 10.2337/db12-1077 PMID: 23300278

53. Kooner JS1, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet. 2011; 43(10):984–989. doi: 10.1038/ng.921 PMID: 21874001

54. Lecompte S, Pasquetti G, Hermant X, Grenier-Boley B, Gonzalez-Gross M, De Henauw S, et al. Genetic and molecular insights into the role of PROX1 in glucose metabolism. Diabetes. 2013; 62(5):1738–1745. doi: 10.2337/db12-0864 PMID: 23274905

55. Wang J, Kilic G, Aydin M, Burke Z, Oliver G, Sosa-Pineda B. Prox1 activity controls pancreas morphogenesis and participates in the production of "secondary transition" pancreatic endocrine cells. Dev Biol. 2005; 286(1):182–194. PMID: 16122728

56. Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS?. Ann N Y Acad Sci. 2010; 1212:59–77. doi: 10.1111/j.1749-6632.2010.05838.x PMID: 21091714

57. Gagniuc P, Ionescu-Tirgoviste C. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. BMC Genomics. 2012; 13:512. doi: 10.1186/1471-2164-13-512 PMID: 23020586

58. Gagniuc P, Cristea PD, Tuduce R, Ionescu-Tîrgovişte C, Gavrila L. DNA patterns and evolutionary signatures obtained through Kappa Index of Coincidence. Rev Roum Sci Techn Électrotechn et Énerg. 2012; 57(1):100–109.

59. Gagniuc P, Ionescu-Tirgoviste C. Gene promoters show chromosome-specificity and reveal chromosome territories in humans. BMC Genomics. 2013; 14:278. doi: 10.1186/1471-2164-14-278 PMID: 23617842