

RESEARCH ARTICLE

# Using spatio-temporal surveillance data to test the infectious environment of children before type 1 diabetes diagnosis

Pierre Bougnères<sup>1,2\*</sup>, Sophie Le Fur<sup>1,2</sup>, Isis-Diab collaborative group<sup>¶</sup>, Sophie Valtat<sup>1</sup>, Yoichiro Kamatani<sup>3</sup>, Mark Lathrop<sup>4</sup>, Alain-Jacques Valleron<sup>2</sup>

**1** Department of Pediatric Endocrinology, Bicêtre Hospital, Paris Sud University, AP-HP, Le Kremlin Bicêtre, France, **2** Inserm U1169, Paris Sud University, Le Kremlin Bicêtre, France, **3** Center for Integrative Medical Sciences, RIKEN, Laboratory for Statistical Analysis, Kanagawa, Japan, **4** Centre National de Génotypage, Evry, France, and Genome Québec Innovation Centre, McGill University, Montréal (Québec), Canada

¶ Membership of the Isis-Diab collaborative group is provided in the Acknowledgments.

\* [pierre.bougneres@inserm.fr](mailto:pierre.bougneres@inserm.fr)



**OPEN ACCESS**

**Citation:** Bougnères P, Le Fur S, Isis-Diab collaborative group, Valtat S, Kamatani Y, Lathrop M, et al. (2017) Using spatio-temporal surveillance data to test the infectious environment of children before type 1 diabetes diagnosis. PLoS ONE 12(2): e0170658. doi:10.1371/journal.pone.0170658

**Editor:** Marc S. Horwitz, University of British Columbia, CANADA

**Received:** May 19, 2016

**Accepted:** January 9, 2017

**Published:** February 2, 2017

**Copyright:** © 2017 Bougnères et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available on request. Requests will be forwarded by P. Bougnères ([pierre.bougneres@inserm.fr](mailto:pierre.bougneres@inserm.fr)) to the Clinical Research Council of the Isis-Diab cohort which is composed of 7 clinicians representing the heads of the collaborative centers. The confidentiality of the medical data of the Isis-Diab study are protected according to the French regulation, and to the terms of the agreement (DR-2010-0035) given by the National authority on informatics and liberty (CNIL). Investigators using the data must simply fill an “agreement for

## Abstract

The “hygiene hypothesis” postulates that reduced exposure to infections favours the development of autoimmunity and childhood type 1 diabetes (T1D). But on the other side, viruses, notably enteroviruses, are suspected to trigger T1D. The assessment of the possible relationships between infections and T1D still defies the classical tools of epidemiology. We report the methods and results of a geographical approach that maps the addresses of patients to a communicable diseases surveillance database. We mapped the addresses of patients at birth, infancy and T1D diagnosis to the weekly estimates of the regional incidences of 5 frequent communicable diseases routinely collected since 1984 by the French Sentinel network. The pre-diagnostic infectious environment of 3548 patients with T1D diagnosed between 0.5 and 15 years was compared to those of 100 series of age-matched “virtual controls” drawn randomly on the map. Associations were classified as “suggestive” (summer diarrhea, SD, and varicella, V) when  $p < 0.05$ , or “significant” (influenza-like infections, ILI) when they passed the Bonferroni correction for FDR. Exposure to ILI and SD were associated with T1D risk, while V seemed protective. In the subset of 2521 patients for which we had genome wide data, we used a case-only approach to search for interactions between SNPs and the infectious environment as defined by the Sentinel database. Two SNPs, rs116624278 and rs77232854, showed significant interaction with exposure to V between 1 and 3 years of life. The infectious associations found should be taken as possible markers of patients’ environment, not as direct causative factors of T1D. They require replication in other populations. The increasing public availability of geographical environmental databases will expand the present approach to map thousands of environmental factors to the lifeline of patients affected by various diseases.

confidentiality” which states the computer security measures that have to be enforced.

**Funding:** This study was sponsored by ongoing institutional grants from NovoNordisk France and Lilly France, a specific Inserm support lasting from 2008 to 2014, the Programme Hospitalier de Recherche Clinique (PHRC ISIS-DIAB and ISIS-VIRUS), the Agence Nationale de la Recherche (ANR ENVIROGENEPIC), the Association pour la Recherche sur le Diabète (ARD). The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. They also did not have access to the clinical trial database, nor any opportunity to review the manuscript. The corresponding author had full access to all the data in the study and final responsibility for the decision to submit for publication.

**Competing Interests:** NovoNordisk provided an institutional support to the cohort in a joint funding with Inserm, without any matter for employment, consultancy, patents, products in development, or relationships at all with marketed products. Lilly France provided also an institutional support to the cohort without any matter for employment, consultancy, patents, products in development, or relationships at all with marketed products. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

## Introduction

The natural history of T1D makes the search for environmental markers difficult. Indeed, it takes a silent period lasting several months up to several years for the autoimmune reaction to achieve the near-complete destruction of  $\beta$ -cells, as indicated by the appearance of autoantibodies in a child’s serum long before T1D is diagnosed [1–3]. The rate of  $\beta$ -cell destruction varies across patients for unknown reasons, but seems accelerated in young children [4]. Environmental factors can affect susceptibility, triggering of autoimmunity, and possibly preclinical course of T1D. The limited concordance of T1D (30–40%) across monozygotic twin pairs is the main proof of major non-genetic risk factors [5–8] that are thought to interact with the genetic background of predisposed children. The genetic background that predisposes to childhood T1D nowadays may not be the same as 30 years ago [9].

The currently increasing incidence of early forms of T1D in many developed countries [10–14] suggests a deleterious remodeling of infants’ environment, either because ancient protective factors have vanished or because exposures causing T1D have emerged or have become more frequent. Incidence and recent increase in incidence of T1D, however, are different across European countries [15].

Infectious exposures are the main environmental suspects, but there is no direct evidence of their implication in T1D risk or natural history in humans. Based on experimental observations in laboratory rodents, the “hygiene hypothesis” postulates that the reduced exposure to infectious agents that children enjoy nowadays has altered the longstanding balance between the developing immune system and microbiological pressure, favoring the occurrence of allergic or autoimmune diseases [16, 17]. Indeed children in developed countries are less exposed to infections than 50 years ago, a result of antibiotics, vaccination, as well as improved hygiene and sanitation. It is thus tempting to think that this may be an underlying cause for the increase in childhood T1D incidence [18, 19]. An inverse relation between increasing T1D incidence and decreasing tuberculosis, rheumatic fever, pinworm infestation, is found at a population level [18, 20, 21], but the almost complete disappearance of these prototypical diseases decades ago cannot explain the more recent increase in childhood T1D. Widespread treatment with antibiotics has decreased the number of infectious episodes while modifying the physiological immune stimulation afforded by intestinal microbiota [22]. Vaccinations, now applied to most infants, prevent measles, mumps, and rubella that affected the vast majority of them in a recent past. On the other side, viral infections may be favored by community-acquired transmission in day-care centers and pre-school nurseries. Rodent models show that viruses may instead induce autoimmune diabetes through their action on  $\beta$  cells or on the immune system [2, 3, 23–25]. Viruses, notably enteroviruses [26], have been suspected to cause human T1D [27]. Pioneering studies have examined specific viral infections in biological samples from T1D children compared to controls, or have used questionnaires to describe environmental events retrospectively. However, results of these studies are discordant [28–30].

In conclusion, it is possible that early exposure to viruses does protect, or on the contrary trigger T1D, it is also possible that some viral infections protect, while other viruses trigger T1D mechanisms.

The epidemiological search of possible infectious determinants of T1D, and of possible gene-environment associations, is difficult. Indeed, given the low incidence of the disease and its protracted and insidious course [2, 3], cohorts require a long, costly and prospective observation of infectious events in millions of children of the general population. A more promising approach has been developed in the Teddy and BabyDiet cohorts, where specific environmental factors are studied prospectively among children at risk for T1D such as siblings or offspring of T1D patients [31, 32]; those who develop T1D or autoantibodies to  $\beta$  cell antigens

are compared with those remaining disease-free and autoantibody-negative. On the other side, the case-control design requires to retrieve viral events in the past histories of children, with a high risk of recall bias.

An alternative to these classical epidemiological designs can be, as we propose in this paper, to take advantage of the development of the new systems of surveillance of communicable diseases in the general population that provide data with an ever better granularity in time and space (see webpage of the International Society for Disease Surveillance [33] for a list of local and national experiments and resources). Moreover, these databases are increasingly publicly available to researchers without cost. It is then straightforward to map the address of a patient at a given time of his life to know whether he was—or not—living at that time in an area where there was an influenza-like epidemic, or any of the diseases under surveillance. France established a national real time monitoring system called “Sentinel” in 1984 [34], which relies on sentinel general practitioners who report their clinical encounters of 5 frequent communicable diseases. In this work, we show how the geolocation of the patients of a cohort can be used to test if infectious exposures were different in T1D patients and in the general population. This may be done by generating “virtual controls” which provide the reference values. This also enables the search for possible gene-environment associations.

## Patients and methods

### Patients

The patients of this study (n = 3548) were selected from the large multicenter Isis-Diab cohort which was established in 2007 and involves 99 participating centers covering most of continental France. Inclusion criteria for the current study were birth after 01/01/1980 in continental France, Caucasian ethnicity and age at clinical diagnosis between 0.5 and 15 yrs (Table 1). T1D was defined according to the American Diabetes Association [35] and by the presence of at least one class of autoantibodies to glutamic acid decarboxylase, insulin, or islet antigen-2.

### Ethics

The research protocol was approved by the Ethics committee of Ile de France (DC-2008-693) and the computer security and confidentiality guarantees given to patients was approved by the Commission Nationale Informatique et Libertés (DR-2010-0035). The ClinicalTrial.gov identifier was NCT02212522. All patients provided written informed consent for participation in the study and donation of samples. We obtained written informed consent from the next of kin, caretakers, or guardians on behalf of the children enrolled in the study.

### Geocoding

Addresses of patients at birth, between birth and 4 years and at time of T1D diagnosis were collected as part of an environmental questionnaire filled by patients. Geocoding of the patients’ addresses was done using the ArcGIS 9.3.1 system, the ArcView software, and the database BD ADRESSE® V2 database provided by the French National Geographic Institute (<http://professionnels.ign.fr/bdadresse>).

**Table 1. Age distribution of the patients at T1D diagnosis.**

	N	Mean (yrs)	Median (yrs)	SD (yrs)	[0–3] yrs (N)	[3–6] yrs (N)	[6–9] yrs (N)	[9–12] yrs (N)	[12–15] yrs (N)
Males	1808	7.2	7.1	3.7	311	425	458	390	223
Females	1740	7.1	7.1	3.6	297	410	437	421	176
Total	3548	7.2	7.1	3.7	608	835	895	811	399

doi:10.1371/journal.pone.0170658.t001

## Assessing the environment of patients

Geocoding was used to map the addresses of the patients at the 3 times defined above with public geographical databases, in order to characterize the patient’s infectious exposures from birth up to the clinical onset of T1D.

The weekly infectious exposures of patients to a selected set of frequent communicable diseases were assessed using the French Sentinel System [34] created in 1984, which collects in real time the corresponding information in sentinel general practitioners. Weekly incidence data are available on the web <http://www.sentiweb.fr>. We studied the past exposure of patients to three communicable diseases of early childhood (measles, mumps and varicella), and to two communicable diseases which affect children of all ages: influenza like illnesses, and acute diarrheas. The weekly surveillance of Influenza-like illnesses started in 1984. The weekly surveillance of measles and mumps—that were very frequent at that time [36]- started at the same date. The weekly surveillance of Varicella and Acute diarrheas started in 1990.

The association of T1D with past infectious exposures was studied using several different definitions of exposure to an infectious environment (see [S2 File](#)). We assessed the infectious environment of a child during various time-windows in his life ([Table 2](#)) using the regional estimate of the incidences at the places of life. A French region is a Local Administrative Unit of level 1 (LAU1) according to European Union definition (<http://ec.europa.eu/eurostat/web/>

**Table 2. Variables used for the study of the infectious environment (INF-E).** The past exposure of patients was studied for measles, mumps, varicella, influenza like illnesses, and acute diarrheas during different exposure windows. It was abstracted using two methods: “cumulative” is the total of the regional incidences around the case during the window. “high” is the mean of the 10 largest regional incidences to which the patients or the VC were exposed. See Material and Methods, and [S2 File](#) for details on the measures of environmental infectious exposures.

Exposure window (method)	Influenza like illnesses exposure	Measles exposure	Mumps exposure	Varicella exposure	Acute diarrheas exposure
6 months-1yr (“cumulative”)	INF-E1	INF-E21	INF-E33	INF-E45	INF-E57
Birth-1yr (“cumulative”)	INF-E2	INF-E22	INF-E34	INF-E46	INF-E58
1yr-2yrs (“cumulative”)	INF-E3	INF-E23	INF-E35	INF-E47	INF-E59
2yrs-3yrs (“cumulative”)	INF-E4	INF-E24	INF-E36	INF-E48	INF-E60
3yrs-4yrs (“cumulative”)	INF-E5	INF-E25	INF-E37	INF-E49	INF-E61
1yr-T1D diagnosis (“cumulative”)	INF-E6	INF-E26	INF-E38	INF-E50	INF-E62
6 months-1yr (“high”)	INF-E7	INF-E27	INF-E39	INF-E51	INF-E63
Birth-1yr (“high”)	INF-E8	INF-E28	INF-E40	INF-E52	INF-E64
1yr-2yrs (“high”)	INF-E9	INF-E29	INF-E41	INF-E53	INF-E65
2yrs-3yrs (“high”)	INF-E10	INF-E30	INF-E42	INF-E54	INF-E66
3yrs-4yrs (“high”)	INF-E11	INF-E31	INF-E43	INF-E55	INF-E67
1yr-T1D diagnosis (“high”)	INF-E12	INF-E32	INF-E44	INF-E56	INF-E68
Birth-1yr (winter, “cumulative”)	INF-E13	-	-	-	INF-E69
Birth-1yr (summer, “cumulative”)	INF-E14	-	-	-	INF-E70
1yr-2yrs (winter, “cumulative”)	INF-E15	-	-	-	INF-E71
1yr-2yrs (summer, “cumulative”)	INF-E16	-	-	-	INF-E72
2yrs-3yrs (winter, “cumulative”)	INF-E17	-	-	-	INF-E73
2yrs-3yrs (summer, “cumulative”)	INF-E18	-	-	-	INF-E74
3yrs-4yrs (winter, “cumulative”)	INF-E19	-	-	-	INF-E75
3yrs-4yrs (summer, “cumulative”)	INF-E20	-	-	-	INF-E76

doi:10.1371/journal.pone.0170658.t002

[nuts/local-administrative-units](#)); France is divided into 21 regions. We used 2 definitions of the environmental exposures of the children that abstracted the time series of the local incidences during the window: the total of the incidences (noted “cumulative” in [Table 2](#)), and the mean of the 10 largest incidences to focus on the epidemic patterns (noted “high” in [Table 2](#)). For influenza like illnesses, and acute diarrheas, we studied separately the winter and summer exposures, as influenza occurs only in winter, and as acute diarrheas in winter and summer have different etiologies ([Table 2](#)) [[37](#), [38](#)].

## Genome wide genotyping and imputation

A subset of 2521 patients from the Isis-Diab cohort had a whole-genome scan (age at diagnosis:  $6.9 \pm 3.5$  yrs SD). Genotyping was performed at the Centre National de Génotypage (Evry, France) and interpreted at the Genome Québec Innovation Center (Montréal, Canada) with SNP imputation performed at the Riken Institute (Kanagawa, Japan). Technical details are provided in the [S1 File](#).

## Design

**Environmental study.** We used a case-control design with a 1:1 matching in which controls are “virtual controls” (VC), not real persons. The past infectious environment of the T1D patients was compared to that of these VCs. To define the VCs, we followed the key epidemiological principle which is that controls should be obtained from the same source population as the cases [[39](#)]. The difficulty to overcome was that the Isis-Diab cohort does not cover uniformly the whole French territory, and that its coverage has varied with time: in the initial years of the cohort, patients were recruited by the first cooperating centers in a much narrower territory than the current one (which is almost all France). To overcome this difficulty, we performed the VC sampling algorithm within successive 2-year classes of birth (starting in 1941, ending in 2011).

**VC sampling algorithm:** The French territory was partitioned using a grid of  $20 \times 20 = 400$  cells of width of  $0.7425^\circ$  longitude, and height of  $0.495^\circ$  latitude (see [S1 Fig](#)). To model the time variation of the recruitment of patients, we considered the spatial distribution of the cohort by classes of 2 years of birth. For each period of two years, we denoted  $b_i$  the number of patients in the cohort who were born in grid cell  $i$ , and  $B$  the sum over all grid cells namely  $B = \sum(b_i)$ . Thus  $B$  is the number of patients of the cohort born in the period of 2 years of birth to which belongs the case for which a virtual control is searched, and  $b_i$  describes the spatial dispersion of the cases of the cohort at that period. The weight of a square  $i$  in the cohort is  $w_i = b_i / B$ . A small  $w_i$  indicates that few patients of the cohort were born in cell  $i$  during the 2 years period considered. The variation of the  $w_i$  with time models the temporal and spatial variation of the recruitment of patients.

For each case in the cohort, an age-matched VC was sampled thanks to a two-step process. In the first step, the grid cell  $i$  containing the VC was randomly chosen with probability  $w_i$ . In the second step, the precise geographical coordinates of the VC within this “large” grid cell  $i$  is obtained by sampling a “small” 200m x 200m elementary unit within cell  $i$ . This sampling was done using the Proportional Per Size R package by Jack G. Gambino (Functions for PPS sampling: <http://CRANR-project.org/package=pps>). The local density of population needed by the pps algorithm was obtained at the 200m x 200m precision level from the French National Institute of Statistics (INSEE) 2009 database (<https://www.insee.fr/fr/statistiques/2520034>). Finally, the age of the VC was taken randomly (uniform distribution) between the age of the case—6 months, and the age of the case + 6 months.



This algorithm was repeated 3548 times to obtain a first set of 3548 VCs. Then the whole process was repeated 100 times, in order to generate 100 sets of 3548 VCs. The environmental variables were then compared between the group of 3548 patients and each of the 100 groups of 3548 VCs.

### Gene environment analysis

The possibility of gene-infectious environment (GxE) interactions was tested with a case-only method described by Khoury [40] where the exposures to the considered environmental factors are compared among the three genotypic groups at each SNP position.

### Statistics

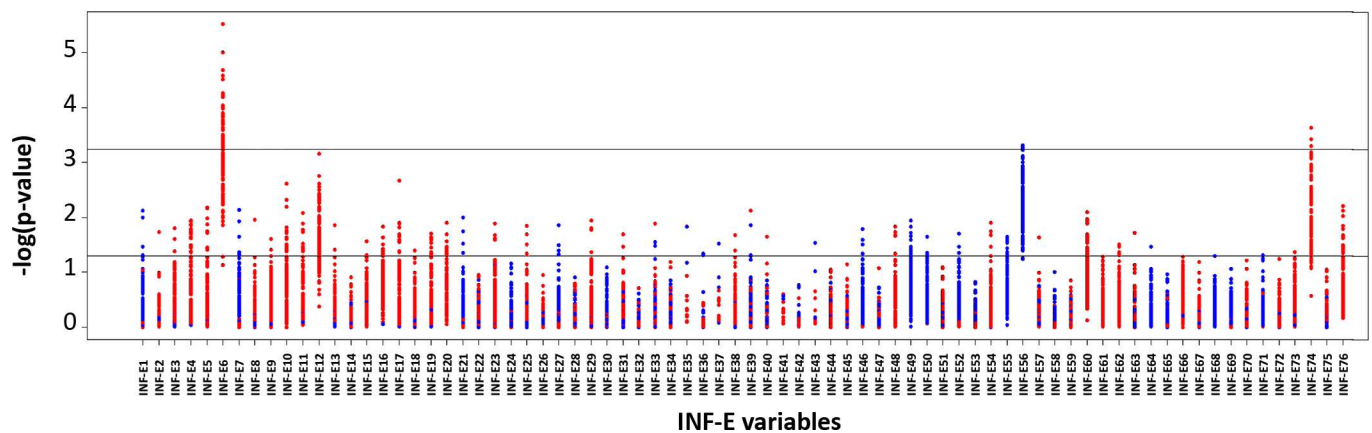
In the environmental study, exposures of cases and controls were compared with the conditional logistic regression R-package CLogit. A test was reported as *indicative of a possible difference* when the median p value computed over the 100 comparisons of the cases with the 100 sets of virtual controls was <0.05. It was reported as *significant* when the median p value was below the Bonferroni limit  $p < 0.05/N$ , where  $N = 76$  (total number of exposure variables in Table 2).

In the gene-environment study, we used the case-only method as implemented in PLINK [41]. We first searched for GxE association by using Wald tests based on the linear regression model. Results were visualized using a Manhattan plot (using ggplot2 package in R). Dominant, recessive, and genotypic analyses were performed. A test was considered as indicative of a possible association when p was between 0.05/N and 1/N, and significant when  $p < 0.05/N$  where  $N = 7,329,768$  (total number of imputed SNPs).

## Results

### Environmental study

Among the variables that we used to measure the potential exposure to infectious environment, the total influenza-like illnesses (ILI) burden between the age of 1 year and time of diagnosis was found significantly larger in T1D children than in controls after Bonferroni correction (Fig 1). T1D children also showed a significantly greater past exposure to acute



**Fig 1. Analysis of the infectious environment of T1D in the Isis-Diab cohort compared with virtual controls.** Red/ blue dots indicate that the patients were more/less exposed than the controls. The two lines stand for tests significance at the 0.05 level without (bottom line) and with (upper line) Bonferroni correction for multiple testing (see Table 2 for the definition of INF-E variables).

doi:10.1371/journal.pone.0170658.g001

summer diarrheas, and a lower past exposure to varicella, but these differences did not survive the Bonferroni correction (Fig 1).

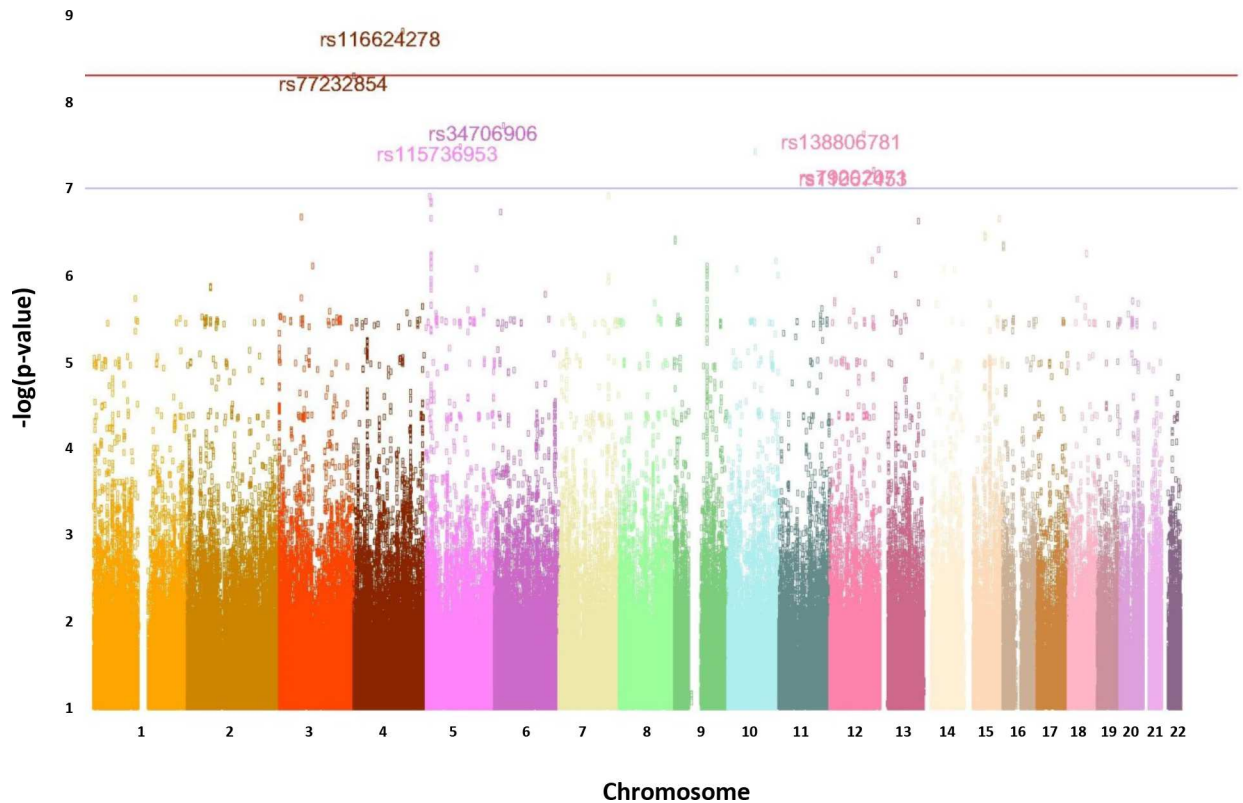
### Gene-infectious exposures

When the infectious exposures were assessed between 1 and 3 years (the child being no longer covered by maternal antibodies), there was an indication of a possible interaction between genetics and infectious exposure. Associations of 2, 1 and 10 SNPs was found with acute diarrheas, ILI, and varicella respectively (Table 3). rs116624278 and rs77232854 survived the Bonferroni correction for an interaction with varicella (Fig 2). rs116624278 is an intergenic SNP located between *PGRMC2* and *JADE1* genes, in a QTL for chronic obstructive pulmonary disease, heart rate and osteoarthritis. rs77232854 is an intergenic SNP located between *PIGG* and *PDE6B*. Both SNPs seem to be in inactive chromatin regions. When cumulative exposures to

**Table 3. SNPs found associated with an environmental exposure to infectious diseases.** The SNPs listed are those for which the p-value of the final test was classified as « indicative » of a possible association ( $p < 10^{-7}$ , see Material and Methods). The 2 SNPs that survived the Bonferroni correction are indicated in bold.

Infectious environment Disease	Exposure window	Chr	SNP	Model	P-value (unadjusted)	P-value (Bonferroni)	
Acute Diarrheas	1yr– 3yrs	11	rs7480811	Recessive	4.57E-8	0.3350	
		11	rs35611816	Recessive	1.34E-7	0.9839	
	Birth– 3yrs	16	rs112398774	General	7.99E-8	0.5858	
		16	rs73500787	General	1.13E-7	0.8273	
		16	rs7191886	General	1.16E-7	0.8491	
		16	rs112398774	Dominant	5.07E-8	0.3714	
		16	rs73500787	Dominant	7.43E-8	0.5448	
		16	rs7191886	Dominant	7.92E-8	0.5806	
		11	rs7480811	Recessive	2.35E-8	0.1719	
		11	rs35611816	Recessive	4.25E-8	0.3111	
		11	rs34123790	Recessive	4.58E-8	0.3354	
		11	rs570887	Recessive	9.05E-8	0.6636	
	ILI	1yr– 3yrs	15	rs77378263	Recessive	3.06E-8	0.2240
		Birth– 3yrs	5	rs115236876	General	1.24E-7	0.9107
			6	rs149303805	Dominant	1.07E-7	0.7871
			15	rs3759862	Recessive	3.27E-8	0.2400
			9	rs11788668	Recessive	4.69E-8	0.3434
4			rs56089258	Recessive	8.15E-8	0.5971	
Chickenpox	1yr– 3yrs	4	<b>rs116624278</b>	<b>Recessive</b>	<b>1.55E-9</b>	<b>0.0114</b>	
		4	<b>rs77232854</b>	<b>Recessive</b>	<b>5.13E-9</b>	<b>0.0376</b>	
		6	rs34706906	Recessive	1.92E-8	0.1408	
		12	rs138806781	Recessive	2.40E-8	0.1759	
		5	rs115736953	Recessive	3.35E-8	0.2455	
		10	chr10:73778256:l	Recessive	3.87E-8	0.2833	
		12	rs79202071	Recessive	6.23E-8	0.4569	
		12	rs11067453	Recessive	6.39E-8	0.4680	
		7	rs3847106	Recessive	1.24E-7	0.9100	
		5	rs140631408	Recessive	1.27E-7	0.9316	
	Birth– 3yrs	7	rs371617	Recessive	3.87E-8	0.2837	
7		rs445489	Recessive	3.92E-08	0.2873		

doi:10.1371/journal.pone.0170658.t003



**Fig 2. Manhattan Plot showing the SNPs associated with an environmental exposure to varicella.** The exposure is the cumulated incidence of chickenpox before age 3yrs computed at the place of birth. The grey line indicates the significance level ( $p = 1.4/10^{-8}$ ) considered as possibly indicating an association at birth. The two SNPs above the red line (located in Chromosome 4) passed the Bonferroni correction ( $p < 7.1 \times 10^{-9}$ ).

doi:10.1371/journal.pone.0170658.g002

infectious diseases were assessed over the entire period of 0 to 3 years of age, there was an indication of a possible association for 7 SNPs located on chromosome 11 and for 6 SNPs located on chromosome 16 with acute diarrheas (Table 3). Two other SNPs showed association with varicella, and 5 SNPs with ILI, but none of these associations survived the Bonferroni correction.

## Discussion

This study detected differences in infectious exposures between future T1D patients and control children that occurred at the location of residence, while social environment was comparable. This observation supports the existence of a relationship between infections and T1D occurrence. A weakness of our study is that it is exclusively based on addresses of residence and brings no proof that the child living in the area of residence of infected individuals has actually developed the infection reported by the sentinel surveillance network, or even has been exposed to infected individuals. This is why in this manuscript said “exposures” should be understood as “potential exposures”.

Our results suggest that T1D is associated with an exposure to more influenza-like illnesses and summer diarrhea during infancy, and to less varicella. These inverse relationships of T1D with different infectious exposures do not support a simple view based on the hygiene hypothesis alone [18]. The development of autoimmunity and T1D may be favored by the interaction of diarrhea with enteroviruses [26] or other members of the gut microbiome [42]. Outside of



the gut, T1D may be favored by the previous interaction of viruses with the immune system [23]. For varicella, recent data regarding the presence of the VZV in the neurons innervating the islets of Langerhans could be used to speculate on a protective mechanism that may result from varicella infection [43, 44]. However, we think that one should resist the temptation to extrapolate biological interpretations from purely statistical observations, unless clear mechanisms can support this speculation, which is not the case here.

It would be very interesting, if feasible, to test if the appearance of autoantibodies to  $\beta$ -cell antigens correlates with infectious exposures, especially because it would make it possible to estimate the time interval between exposure and the earlier detectable manifestation of autoimmunity to  $\beta$  cells. Indeed, the natural history of childhood T1D studied in siblings and offspring of patients with T1D indicates that autoimmunity often declares itself in early childhood [2, 3]. These observations inspired our exploratory approach by focusing our environmental search on early times of life.

There is an almost unlimited number of possible combinations that can be analyzed by our spatio-temporal approach. First, any complex combination of environmental factors could be considered. Second, an environmental factor can target its victim at different possible ages: this is the reason why this work has tested different windows of exposure, but more could be easily considered with the same approach. Even though the number of possibilities is daunting, the good news is that it is possible. We can for example study the consequences of an exposure during the 2<sup>nd</sup> trimester of pregnancy just by matching the dates of this second trimester with the values of the environmental databases at the place of residence of the mother. Finally, this is becoming a simple problem of computing power availability. Everything suggests that the computing power that researchers will be able to use will not be a limit to the ambitions of this kind of research. Until environment can be defined more exhaustively at the individual level, our study should be viewed as an exploratory proof-of-concept approach of a multifactorial disease whose environmental causes are yet largely unknown.

Rappaport and Smith have proposed the “exposome” as a new paradigm in which signatures elicited by environmental factors can be detected in serum or in circulating cells [45]. More specifically, next generation sequencing can be used to search for viruses “blindly” in plasma, a promising approach used in the Teddy cohort [46]. Once significant exposures have been identified in blood, it may be possible to determine their sources in the pre-disease environment. Our approach of children’s exposures could be complementary to this strategy, either by providing suspect factors to future exposome analyses, or by testing exposome-driven information across the environmental landscape. The blood exposome and the external exposures can be combined in a cohort or in a case-control approach. More specifically, serology microarrays may soon provide tools for specific pathogen discovery in serum of patients at time of appearance of  $\beta$ -cell antigen antibodies and T1D diagnosis.

The other part of our study was an attempt to cross genetic and environmental information. For this purpose, we used the case-only method proposed by Khoury [40], which compares the level of exposure to the chosen environmental factors across the genotypes at each SNP position. The strong constraint of the case-only approach is that the tested environmental factors need to be independent from genetics. Indeed, if a given child’s infection is influenced by his genotype, a statistical interaction will be found between infection and genotype, which will have no meaning for T1D research. A handful of SNPs were found associated with the studied infectious exposures, indicating possible gene-environment interactions. It is difficult to interpret these statistical observations in the light of a plausible biological interaction between infectious exposures and genetic factors. Notably, the two SNPs that showed a significant interaction with varicella, rs116624278 and rs77232854, were not located within known coding or regulatory sequences of genes directly relevant to immune processes or beta-cell autoimmunity. Their interaction with

the varicella exposure eludes our understanding, but might be thought of within a protective context for varicella towards T1D. Our lack of biological understanding of the gene-environment interactions should not be felt as too disappointing since it is common to the vast majority of genomic studies in multifactorial diseases [47, 48]. Our blind genome-wide exploration observed no association of infectious exposures with SNPs located within gene loci known to be involved in infectious or immune response, including HLA class II. This may be due to the lack of involvement of these genes versus the studied exposures, or to the heterogeneity of T1D mechanisms [49]. Indeed, the genetic predisposition to T1D, including HLA genotypes, is known to be highly heterogeneous at the individual level since patients' genotypes do not carry the same combination of susceptibility alleles. Thus it is possible that particular groups of genes play a role only in specific subsets of T1D patients by interacting with certain environmental factors. If very distinctive of groups of patients, such interaction could escape statistical detection when a whole-genome search is crossed blindly with markers of infectious environment, as performed in the current study.

In conclusion, the current study supports the contribution of infectious exposures to recent childhood T1D epidemiology, yet should be viewed as a methodological proof-of-concept attempt that awaits for confirmation and replication.

## Supporting information

**S1 Fig. Grid used to generate the virtual controls.** The 20x20 grid constitutes approximately a square of 1000 km x 1000 km. Each cell of the grid has a width of 7425° longitude and a height of 0.7425°. France occupies 190 of the 400 cells of the grid.  
(DOCX)

**S1 File. Technical details on genome wide genotyping and imputation.**  
(DOCX)

**S2 File. Methods used to assess the infectious environment of a patient.**  
(DOCX)

## Acknowledgments

We deeply thank NovoNordisk France for the institutional “Diabetes Alliance” grant running from 2008, Inserm for a specific support to “Diabetes Alliance” 2008–2014, the Programme Hospitalier de Recherche Clinique, the Agence Nationale de la Recherche, Lilly France, the Groupe d'Etude et Thérapeutique de la Croissance and the Association pour la Recherche sur le Diabète (ARD). We thank the patients and parents who participated in the study. We thank D Baker, B Guedj, TN Mai, S Meurisse and A de la Vaissière for their technical help in the statistical analyses. We thank B Aboumradi, P Lucchini and nurses in the Isis-Diab centers for sampling the patients. We thank A Fourreau, G Leprun, A Guégan, V Jauffret, C Mille, D Boudia and A Alenda for technical help. We thank all the investigators of the Isis-Diab collaborative group who co-authored the study (lead author Sophie Le Fur: [sophie.le-fur@inserm.fr](mailto:sophie.le-fur@inserm.fr)): M Akhdar, D Ancelle, C Atlan, E Baechler, P Barat, S Baron, J Bassil, C Ben Signor, D Bendifallah, PY Benhamou, A Benmahammed, AM Bertrand, C Boniface, H Bony Trifunovic, MN Bortoluzzi, N Bouhours Nouet, JP Brossier, JP Cappoen, S Caudrelier, L Chaillous, M Chambon, P Chatelain, S Clavel, J Cohen, ME Combes Moukhovskiy, R Coutant, F Dallavale, M De Kerdanet, B Delemer, D Dervaux, R Desaillood, B Doremus, E Dubosclard, H Dubourg, PH Ducluzeau, D Dufillot, C Dumont, A Fargeot Espaliat, C Fourmy Chatel, S Fournier, V Gajdos, F Gatelais, JF Gautier, CL Gay, A Gourdin, P Gourdy, I Guemas, I Gueorguieva, B Guérin, I Guilhem, A Guillou, P Jan, S Jellimann, MG Joseph, F Joubert, M Joubert, V Kerlan, C

Kerouedan, J Khoury, F Kurtz, F Labay, N Laisney, M Laloï Michelin, E Larger, M Le Bideau, B Le Luyer, C Le Tallec, P Lecomte, AM Leguerrier, A Lehnert, A Lienhardt Roussie, GA Loeuille, C Lukas, E Marcoux, M Matta, C Metz, B Mignot, K Milcent, A Monier, S Nadalon, R Nicolescu, M Nicolino, S Nivot, L Oilleau, F Olivier, A Pambou, L Pantalone, A Paoli, D Paris Bockel, G Parlier, A Penformis, M Pepin Donat, C Petit Bibal, ML Pignol, R Piquemal, C Pons, F Popelard, M Priou Guesdon, C Radet, M Raoulx, B Razafimahefa, R Reynaud, Y Reznic, O Richard Colmant, V Rigalleau, R Salet, G Simonin, E Sonnet, C Stheneur, C Stuckens, P Talon, A Tamboura, D Terral, D Thevenieau, F Tixier, P Trioche, A Valade, V Vautier, F Verax, C Vervel, J Weill, MC Wielickzo, E Yollin, A Zelinsky Gurung, B Zimmermann.

## Author contributions

**Conceptualization:** PB AJV SLF.

**Data curation:** SLF SV.

**Formal analysis:** AJV SV SLF.

**Funding acquisition:** PB.

**Investigation:** PB SLF.

**Methodology:** PB AJV YK ML.

**Project administration:** SLF.

**Resources:** PB AJV SLF the Isis-Diab collaborative group.

**Software:** AJV.

**Supervision:** AJV PB.

**Validation:** SV.

**Visualization:** SV SLF.

**Writing – original draft:** AJV PB SLF.

**Writing – review & editing:** AJV PB SLF.

## References

1. Atkinson MA, Eisenbarth GS. Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Lancet*. 2001; 358(9277):221–9. doi: [10.1016/S0140-6736\(01\)05415-0](https://doi.org/10.1016/S0140-6736(01)05415-0) PMID: [11476858](https://pubmed.ncbi.nlm.nih.gov/11476858/)
2. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014; 383(9911):69–82. PubMed Central PMCID: [PMCPMC4380133](https://pubmed.ncbi.nlm.nih.gov/PMCID/PMC4380133/). doi: [10.1016/S0140-6736\(13\)60591-7](https://doi.org/10.1016/S0140-6736(13)60591-7) PMID: [23890997](https://pubmed.ncbi.nlm.nih.gov/23890997/)
3. Ziegler AG, Nepom GT. Prediction and pathogenesis in type 1 diabetes. *Immunity*. 2010; 32(4):468–78. PubMed Central PMCID: [PMCPMC2861716](https://pubmed.ncbi.nlm.nih.gov/PMCID/PMC2861716/). doi: [10.1016/j.immuni.2010.03.018](https://doi.org/10.1016/j.immuni.2010.03.018) PMID: [20412757](https://pubmed.ncbi.nlm.nih.gov/20412757/)
4. Ziegler AG, Pflueger M, Winkler C, Achenbach P, Akolkar B, Krischer JP, et al. Accelerated progression from islet autoimmunity to diabetes is causing the escalating incidence of type 1 diabetes in young children. *J Autoimmun*. 2011; 37(1):3–7. PubMed Central PMCID: [PMCPMC3110538](https://pubmed.ncbi.nlm.nih.gov/PMCID/PMC3110538/). doi: [10.1016/j.jaut.2011.02.004](https://doi.org/10.1016/j.jaut.2011.02.004) PMID: [21376535](https://pubmed.ncbi.nlm.nih.gov/21376535/)
5. Metcalfe KA, Hitman GA, Rowe RE, Hawa M, Huang X, Stewart T, et al. Concordance for type 1 diabetes in identical twins is affected by insulin genotype. *Diabetes Care*. 2001; 24(5):838–42. PMID: [11347740](https://pubmed.ncbi.nlm.nih.gov/11347740/)
6. Stefan M, Zhang W, Concepcion E, Yi Z, Tomer Y. DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *J Autoimmun*. 2014; 50:33–7. PubMed Central PMCID: [PMCPMC3995844](https://pubmed.ncbi.nlm.nih.gov/PMCID/PMC3995844/). doi: [10.1016/j.jaut.2013.10.001](https://doi.org/10.1016/j.jaut.2013.10.001) PMID: [24210274](https://pubmed.ncbi.nlm.nih.gov/24210274/)

7. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes*. 2003; 52(4):1052–5. PMID: [12663480](#)
8. Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T. Concordance for islet autoimmunity among monozygotic twins. *N Engl J Med*. 2008; 359(26):2849–50. doi: [10.1056/NEJMc0805398](#) PMID: [19109586](#)
9. Gillespie KM, Bain SC, Barnett AH, Bingley PJ, Christie MR, Gill GV, et al. The rising incidence of childhood type 1 diabetes and reduced contribution of high-risk HLA haplotypes. *Lancet*. 2004; 364(9446):1699–700. doi: [10.1016/S0140-6736\(04\)17357-1](#) PMID: [15530631](#)
10. Gale EA. The rise of childhood type 1 diabetes in the 20th century. *Diabetes*. 2002; 51(12):3353–61. PMID: [12453886](#)
11. Patterson CC, Dahlquist GG, Gyurus E, Green A, Soltesz G, Group ES. Incidence trends for childhood type 1 diabetes in Europe during 1989–2003 and predicted new cases 2005–20: a multicentre prospective registration study. *Lancet*. 2009; 373(9680):2027–33. doi: [10.1016/S0140-6736\(09\)60568-7](#) PMID: [19481249](#)
12. Harjutsalo V, Sjoberg L, Tuomilehto J. Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet*. 2008; 371(9626):1777–82. doi: [10.1016/S0140-6736\(08\)60765-5](#) PMID: [18502302](#)
13. Karvonen M, Pitkaniemi J, Tuomilehto J. The onset age of type 1 diabetes in Finnish children has become younger. The Finnish Childhood Diabetes Registry Group. *Diabetes Care*. 1999; 22(7):1066–70. PMID: [10388969](#)
14. Writing Group for the SfdiYSG, Dabelea D, Bell RA, D'Agostino RB Jr., Imperatore G, Johansen JM, et al. Incidence of diabetes in youth in the United States. *JAMA*. 2007; 297(24):2716–24. doi: [10.1001/jama.297.24.2716](#) PMID: [17595272](#)
15. Patterson CC, Gyurus E, Rosenbauer J, Cinek O, Neu A, Schober E, et al. Trends in childhood type 1 diabetes incidence in Europe during 1989–2008: evidence of non-uniformity over time in rates of increase. *Diabetologia*. 2012; 55(8):2142–7. doi: [10.1007/s00125-012-2571-8](#) PMID: [22638547](#)
16. Strachan DP. Hay fever, hygiene, and household size. *BMJ*. 1989; 299(6710):1259–60. PubMed Central PMCID: PMC1838109. PMID: [2513902](#)
17. Zinkernagel RM. Maternal antibodies, childhood infections, and autoimmune diseases. *N Engl J Med*. 2001; 345(18):1331–5. doi: [10.1056/NEJMra012493](#) PMID: [11794153](#)
18. Bach JF. The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med*. 2002; 347(12):911–20. doi: [10.1056/NEJMra020100](#) PMID: [12239261](#)
19. Chapman NM, Coppieters K, von Herrath M, Tracy S. The microbiology of human hygiene and its impact on type 1 diabetes. *Islets*. 2012; 4(4):253–61. PubMed Central PMCID: PMC3496651. doi: [10.4161/isl.21570](#) PMID: [22996796](#)
20. Gale EA. A missing link in the hygiene hypothesis? *Diabetologia*. 2002; 45(4):588–94. doi: [10.1007/s00125-002-0801-1](#) PMID: [12032638](#)
21. Cooke A. Review series on helminths, immune modulation and the hygiene hypothesis: how might infection modulate the onset of type 1 diabetes? *Immunology*. 2009; 126(1):12–7. PubMed Central PMCID: PMC32632708. doi: [10.1111/j.1365-2567.2008.03009.x](#) PMID: [19120494](#)
22. Blaser MJ, Falkow S. What are the consequences of the disappearing human microbiota? *Nat Rev Microbiol*. 2009; 7(12):887–94. doi: [10.1038/nrmicro2245](#) PMID: [19898491](#)
23. von Herrath M, Nepom GT. Animal models of human type 1 diabetes. *Nat Immunol*. 2009; 10(2):129–32. doi: [10.1038/ni0209-129](#) PMID: [19148193](#)
24. Filippi CM, von Herrath MG. Viral trigger for type 1 diabetes: pros and cons. *Diabetes*. 2008; 57(11):2863–71. PubMed Central PMCID: PMC32570378. doi: [10.2337/db07-1023](#) PMID: [18971433](#)
25. Schneider DA, von Herrath MG. Potential viral pathogenic mechanism in human type 1 diabetes. *Diabetologia*. 2014.
26. Yeung WC, Rawlinson WD, Craig ME. Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. *BMJ*. 2011; 342:d35. PubMed Central PMCID: PMC3033438. doi: [10.1136/bmj.d35](#) PMID: [21292721](#)
27. von Herrath M. Can we learn from viruses how to prevent type 1 diabetes?: the role of viral infections in the pathogenesis of type 1 diabetes and the development of novel combination therapies. *Diabetes*. 2009; 58(1):2–11. PubMed Central PMCID: PMC32606872. doi: [10.2337/db08-9027](#) PMID: [19114721](#)
28. Cardwell CR, Carson DJ, Patterson CC. No association between routinely recorded infections in early life and subsequent risk of childhood-onset Type 1 diabetes: a matched case-control study using the

- UK General Practice Research Database. *Diabet Med.* 2008; 25(3):261–7. doi: [10.1111/j.1464-5491.2007.02351.x](https://doi.org/10.1111/j.1464-5491.2007.02351.x) PMID: [18201209](https://pubmed.ncbi.nlm.nih.gov/18201209/)
29. Gibbon C, Smith T, Egger P, Betts P, Phillips D. Early infection and subsequent insulin dependent diabetes. *Arch Dis Child.* 1997; 77(5):384–5. PMID: [9487955](https://pubmed.ncbi.nlm.nih.gov/9487955/)
  30. Beyerlein A, Donnachie E, Jergens S, Ziegler AG. Infections in Early Life and Development of Type 1 Diabetes. *JAMA.* 2016; 315(17):1899–901. doi: [10.1001/jama.2016.2181](https://doi.org/10.1001/jama.2016.2181) PMID: [27139064](https://pubmed.ncbi.nlm.nih.gov/27139064/)
  31. Group TS. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann N Y Acad Sci.* 2008; 1150:1–13. PubMed Central PMCID: [PMCPMC2886800](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC2886800/).
  32. Hummel S, Pfluger M, Hummel M, Bonifacio E, Ziegler AG. Primary dietary intervention study to reduce the risk of islet autoimmunity in children at increased risk for type 1 diabetes: the BABYDIET study. *Diabetes Care.* 2011; 34(6):1301–5. PubMed Central PMCID: [PMCPMC3114350](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC3114350/). doi: [10.2337/dc10-2456](https://doi.org/10.2337/dc10-2456) PMID: [21515839](https://pubmed.ncbi.nlm.nih.gov/21515839/)
  33. International Society for Disease Surveillance 2013. Available from: <http://www.syndromic.org>.
  34. Valleron AJ, Bouvet E, Garnerin P, Menares J, Heard I, Letrait S, et al. A computer network for the surveillance of communicable diseases: the French experiment. *Am J Public Health.* 1986; 76(11):1289–92. PubMed Central PMCID: [PMCPMC1646722](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC1646722/). PMID: [3766824](https://pubmed.ncbi.nlm.nih.gov/3766824/)
  35. American Diabetes A. Diagnosis and classification of diabetes mellitus. *Diabetes Care.* 2012; 35 Suppl 1:S64–71. PubMed Central PMCID: [PMCPMC3632174](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC3632174/).
  36. Chauvin P, Valleron AJ. Persistence of susceptibility to measles in France despite routine immunization: a cohort analysis. *Am J Public Health.* 1999; 89(1):79–81. PMID: [9987470](https://pubmed.ncbi.nlm.nih.gov/9987470/)
  37. Letrilliart L, Desenclos JC, Flahault A. Risk factors for winter outbreak of acute diarrhoea in France: case-control study. *Bmj.* 1997; 315(7123):1645–9. PMID: [9448526](https://pubmed.ncbi.nlm.nih.gov/9448526/)
  38. Yazdanpanah Y, Beaugerie L, Boelle PY, Letrilliart L, Desenclos JC, Flahault A. Risk factors of acute diarrhoea in summer—a nation-wide French case-control study. *Epidemiol Infect.* 2000; 124(3):409–16. PMID: [10982064](https://pubmed.ncbi.nlm.nih.gov/10982064/)
  39. Miettinen OS. The "case-control" study: valid selection of subjects. *J Chronic Dis.* 1985; 38(7):543–48. PMID: [4008595](https://pubmed.ncbi.nlm.nih.gov/4008595/)
  40. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls! *Am J Epidemiol.* 1996; 144(3):207–13. PMID: [8686689](https://pubmed.ncbi.nlm.nih.gov/8686689/)
  41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3):559–75. PubMed Central PMCID: [PMCPMC1950838](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC1950838/). doi: [10.1086/519795](https://doi.org/10.1086/519795) PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)
  42. Vaarala O, Atkinson MA, Neu J. The "perfect storm" for type 1 diabetes: the complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes.* 2008; 57(10):2555–62. PubMed Central PMCID: [PMCPMC2551660](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC2551660/). doi: [10.2337/db08-0331](https://doi.org/10.2337/db08-0331) PMID: [18820210](https://pubmed.ncbi.nlm.nih.gov/18820210/)
  43. Chen JJ, Gershon AA, Li Z, Cowles RA, Gershon MD. Varicella zoster virus (VZV) infects and establishes latency in enteric neurons. *J Neurovirol.* 2011; 17(6):578–89. PubMed Central PMCID: [PMCPMC3324263](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC3324263/). doi: [10.1007/s13365-011-0070-1](https://doi.org/10.1007/s13365-011-0070-1) PMID: [22190254](https://pubmed.ncbi.nlm.nih.gov/22190254/)
  44. Razavi R, Chan Y, Afifiyan FN, Liu XJ, Wan X, Yantha J, et al. TRPV1+ sensory neurons control beta cell stress and islet inflammation in autoimmune diabetes. *Cell.* 2006; 127(6):1123–35. doi: [10.1016/j.cell.2006.10.038](https://doi.org/10.1016/j.cell.2006.10.038) PMID: [17174891](https://pubmed.ncbi.nlm.nih.gov/17174891/)
  45. Rappaport SM, Smith MT. Epidemiology. Environment and disease risks. *Science.* 2010; 330(6003):460–1. doi: [10.1126/science.1192603](https://doi.org/10.1126/science.1192603) PMID: [20966241](https://pubmed.ncbi.nlm.nih.gov/20966241/)
  46. Lee HS, Briese T, Winkler C, Rewers M, Bonifacio E, Hyoty H, et al. Next-generation sequencing for viruses in children with rapid-onset type 1 diabetes. *Diabetologia.* 2013; 56(8):1705–11. PubMed Central PMCID: [PMCPMC4019381](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC4019381/). doi: [10.1007/s00125-013-2924-y](https://doi.org/10.1007/s00125-013-2924-y) PMID: [23657799](https://pubmed.ncbi.nlm.nih.gov/23657799/)
  47. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin.* 2015; 8:57. PubMed Central PMCID: [PMCPMC4696349](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC4696349/). doi: [10.1186/s13072-015-0050-4](https://doi.org/10.1186/s13072-015-0050-4) PMID: [26719772](https://pubmed.ncbi.nlm.nih.gov/26719772/)
  48. Zhang X, Bailey SD, Lupien M. Laying a solid foundation for Manhattan—'setting the functional basis for the post-GWAS era'. *Trends Genet.* 2014; 30(4):140–9. PubMed Central PMCID: [PMCPMC4026049](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC4026049/). doi: [10.1016/j.tig.2014.02.006](https://doi.org/10.1016/j.tig.2014.02.006) PMID: [24661571](https://pubmed.ncbi.nlm.nih.gov/24661571/)
  49. Arif S, Leete P, Nguyen V, Marks K, Nor NM, Estorninho M, et al. Blood and islet phenotypes indicate immunological heterogeneity in type 1 diabetes. *Diabetes.* 2014; 63(11):3835–45. PubMed Central PMCID: [PMCPMC4207393](https://pubmed.ncbi.nlm.nih.gov/pmc/PMC4207393/). doi: [10.2337/db14-0365](https://doi.org/10.2337/db14-0365) PMID: [24939426](https://pubmed.ncbi.nlm.nih.gov/24939426/)