Diabetologia (2025) 68:186–202
https://doi.org/10.1007/s00125-024-06298-y

EXTENDED ARTICLE

The Type 1 Diabetes T Cell Receptor and B Cell Receptor Repository in the AIRR Data Commons: a practical guide for access, use and contributions through the Type 1 Diabetes AIRR Consortium

Stephanie J. Hanna[1] · Rachel H. Bonami[2,3,4,5] · Brian Corrie[6,7] · Monica Westley[8] · Amanda L. Posgai[9] · Eline T. Luning Prak[10] · Felix Breden[6,7] · Aaron W. Michels[11] · Todd M. Brusko[9,12,13] · Type 1 Diabetes AIRR Consortium

Received: 26 May 2024 / Accepted: 19 August 2024 / Published online: 29 October 2024
© The Author(s) 2024

Abstract

Human molecular genetics has brought incredible insights into the variants that confer risk for the development of tissue-specific autoimmune diseases, including type 1 diabetes. The hallmark cell-mediated immune destruction that is characteristic of type 1 diabetes is closely linked with risk conferred by the HLA class II gene locus, in combination with a broad array of additional candidate genes influencing islet-resident beta cells within the pancreas, as well as function, phenotype and trafficking of immune cells to tissues. In addition to the well-studied germline SNP variants, there are critical contributions conferred by T cell receptor (TCR) and B cell receptor (BCR) genes that undergo somatic recombination to yield the Adaptive Immune Receptor Repertoire (AIRR) responsible for autoimmunity in type 1 diabetes. We therefore created the T1D TCR/BCR Repository (The Type 1 Diabetes T Cell Receptor and B Cell Receptor Repository) to study these highly variable and dynamic gene rearrangements. In addition to processed TCR and BCR sequences, the T1D TCR/BCR Repository includes detailed metadata (e.g. participant demographics, disease-associated parameters and tissue type). We introduce the Type 1 Diabetes AIRR Consortium goals and outline methods to use and deposit data to this comprehensive repository. Our ultimate goal is to facilitate research community access to rich, carefully annotated immune AIRR datasets to enable new scientific inquiry and insight into the natural history and pathogenesis of type 1 diabetes.

Keywords AIRR · AIRR Data Commons · Autoantibodies · B cell receptors · FAIR data · Next-generation sequencing · Single-cell RNA-seq · T cell receptors · Type 1 diabetes

Abbreviations

AAb	Autoantibody/autoantibodies
ADC	AIRR Data Commons
AIM	Activation-induced marker
AIRR	Adaptive Immune Receptor Repertoire
AIRR-seq	AIRR sequencing
BCR	B cell receptor
CDR3	Complementarity determining region 3
FAIR	Findable, Accessible, Interoperable, Reusable
GEO	Gene Expression Omnibus
HPAP	Human Pancreas Analysis Program
IEDB	Immune Epitope Database
MiAIRR	Minimal information about AIRR
ML	Machine learning
pLN	Pancreatic lymph node(s)
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
scRNA-seq	Single-cell RNA-seq
SRA	Sequence Read Archive
T1D TCR/BCR Repository	The Type 1 Diabetes T Cell Receptor and B Cell Receptor Repository
TCR	T cell receptor

Stephanie J. Hanna and Rachel H. Bonami contributed equally to this work. Aaron W. Michels and Todd M. Brusko contributed equally as joint senior authors.

Members of the Type 1 Diabetes AIRR Consortium are listed in the Acknowledgements.

Extended author information available on the last page of the article

TCRβ	T cell receptor β chain
Tfh	T follicular helper
Treg	Regulatory T cell(s)
VDJ	Variable, diversity and joining gene segments

Introduction

Recombination of variable (V), diversity (D) and joining (J) gene segments within the T cell receptor (TCR) and B cell receptor (BCR) loci forms a diverse array of antigen receptors that are known as the Adaptive Immune Receptor Repertoire (AIRR). Repertoire diversity is essential for pathogen defence, and recent advances in AIRR sequencing (AIRR-seq) technology and machine learning (ML) analyses have enabled the identification of disease-specific AIRR signatures providing high sensitivity and specificity for certain viral infections [1, 2]. These successes have prompted intense interest in the AIRR as a potential means to improve prediction and monitor the progression of autoimmune diseases, such as type 1 diabetes.

In this methods paper, we first overview the current state of the science of TCR and BCR repertoire studies in human type 1 diabetes, and the specific challenges of developing analytical approaches to harness specific repertoire attributes as biomarkers. We then summarise existing frameworks to access type 1 diabetes AIRR-seq data and their respective limitations. The iReceptor platform [3] and AIRR Data Commons (ADC) [4] will address the needs currently facing the type 1 diabetes research community, as evidenced by how they have been used to share AIRR-seq data in other conditions, including cancer and infectious diseases, as detailed further below. We overview how to upload datasets to share with the community and provide use case examples to demonstrate how the available data may be best leveraged. Finally, we look ahead to future developments and collaborations within the type 1 diabetes field.

Current type 1 diabetes biomarkers to predict disease risk and track progression toward diabetes

Several islet autoantibody (AAb) screening programmes have been developed that predict diabetes onset. For example, Type 1 Diabetes TrialNet screens first-degree relatives of individuals with type 1 diabetes for IAA and GADA, followed by IA-2A, ICA and ZnT8A for those positive in the first screening. These islet AAb predict type 1 diabetes risk and stage disease progression: seropositivity for two or more of these islet AAb (≥2AAb+, stage 1 type 1 diabetes) confers ~70% risk of developing diabetes within 10 years, and a nearly 100% lifetime risk [5–7]. Predicted time to diabetes onset can be further refined by considering which islet AAb are positive; ZnT8A positivity most typically appeared after other islet AAb were detected [8], and was associated with increased risk of developing diabetes on follow-up [6, 8]. Multiple studies, summarised in a systematic review, have demonstrated that IA-2A positivity is particularly associated with disease progression in children [6]. The Environmental Determinants of Diabetes in the Young (TEDDY) study data suggest that very young children are more likely to have IAA as the first AAb, with older children more likely to become GADA positive first [9]. IAA and IA-2A titres are associated with higher type 1 diabetes risk and more rapid progression [6].

Furthermore, type 1 diabetes staging provides a convenient, albeit imperfect, framework to categorise participants, based on evidence of islet autoimmunity and extent of dysglycaemia. Stage 1 type 1 diabetes is defined as the presence of multiple islet AAbs and normal glucose measures. In individuals with ≥2AAb+, impaired glycaemic control, measured through an OGTT to assess beta cell function, defines progression to stage 2 as outlined in Table 1 [7, 10]. Stage 3 is defined as onset of diabetes and stage 4 describes individuals with established diabetes. Other metrics have been proposed, such as continuous glucose monitoring (CGM) data [11] or Index60 (i.e. a composite measure of blood glucose and C-peptide levels across OGTT time points) [12]. However, all of these metrics are imperfect, as islet AAb positivity and OGTT results can vary over time [13]. Even in recently diagnosed stage 3 type 1 diabetes, variations in OGTT results and other metrics make tracking disease progression and success/failure of immunotherapies problematic [14, 15]. Furthermore, these studies have made clear that, when relying on C-peptide measurements as the primary outcome, trials require large numbers of participants and often take 12 months or longer to confirm a positive response to immunotherapy [15]. Genetic risk scores (GRS) and polygenic risk scores (PRS) also support diabetes onset prediction (reviewed in [16]). A combined risk score (CRS), which incorporated genetic, clinical and immunological factors, improved diabetes onset prediction over AAb data alone [17].

Immune repertoire analysis can predict disease progression/severity or pathogen exposure in autoimmunity, cancer and infectious disease

Immune repertoire profiling has potential advantages over current risk stratification assays for type 1 diabetes. With respect to antibody profiling, serological detection of secreted antibodies reflects a state wherein B cell tolerance

Table 1 Type 1 diabetes disease staging criteria

	Stage 1	Stage 2	Stage 3[a]
≥2 islet AAb	Yes	Yes	Yes
Glucose tolerance classification[b]	Normal	Impaired	Diabetes
Fasting OGTT blood glucose, mmol/l (mg/dl)	<6.1 (110)	≥6.1 (110), <7.0 (126)	≥7.0 (126)
30, 60 or 90 min OGTT blood glucose, mmol/l (mg/dl)	<11.1 (200)	≥11.1 (200)	NA
2h OGTT blood glucose, mmol/l (mg/dl)	<7.8 (140)	≥7.8 (140), <11.1 (200)	≥11.1 (200)
Symptoms	No	No	Yes[c]

Staging criteria are in alignment with Insel et al 2015 [7]

[a]Type 1 diabetes diagnosis (stage 3) made if: (1) any OGTT time points are abnormal; (2) HbA_{1c} is ≥48 mmol/mol (6.5%); or (3) casual blood glucose ≥11.1 mmol/l (200 mg/dl) and hyperglycaemia symptoms are present

[b]75 g OGTT: two consecutive normal (stage 1) or impaired (stage 2) OGTT tests are required for these classifications. Impaired glucose tolerance can also be defined as HbA_{1c} ≥39 mmol/mol (5.7%) but <48 mmol/mol (6.5%)

[c]The presence of symptoms may vary

is already broken. Some of the antigen targets are intracellular, implying that destruction of pancreatic beta cells is underway. Furthermore, antibody-forming cells are at a mature stage of B cell development. In the case of BCR gene rearrangement analysis, it may be possible to detect B cell clones with pathogenic potential at earlier stages of maturation, which could provide a means of detecting disease risk at an earlier stage, including the initial emergence of autoimmunity. For TCR analysis, there are already early hints that this type of stratification is possible. For example, analysis of high-throughput TCR β chain (TCRβ) repertoire data showed that anti-insulin T cell clones increased in at-risk individuals who progressed to diabetes relative to those who did not develop diabetes over the same time period [18]. In contrast, clonal changes in influenza A-reactive T cells were not observed in this same cohort [18]. The concept that the repertoire landscape can predict immunological exposures is already well entrenched for infectious diseases and it is not a huge stretch to think that, with adequate databases, similar disease-associated repertoire motifs could be unearthed in type 1 diabetes. With respect to infectious disease, TCR sequencing technology was used to discriminate cytomegalovirus (CMV) seropositive from seronegative individuals with high sensitivity and specificity [1]. An ML approach discriminated between severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-infected vs SARS-CoV-2-naive individuals based on TCR sequencing data [19]. As another example, DeepTCR was used with the ImmuneCODE database to identify TCR sequence motifs associated with recognition of specific SARS-CoV-2 epitopes that tracked increased illness severity [20, 21].

Immune repertoire profiling also has advantages over genetic markers of disease progression in type 1 diabetes. This advantage stems from the fact that BCR and TCR repertoires are influenced not only by inherited differences (e.g. germline allele usage, duplications and deletions, HLA haplotypes) but also by somatic events (recombination and, specifically in the case of BCRs, somatic hypermutation and class switching) and immune selection (e.g. clonal diversification, expansion and contraction). These somatic and selection events provide an opportunity to link immune exposures to disease and monitor clones over time to evaluate the effects of therapy. These concepts are already familiar in haematological malignancies, where clonal BCR heavy chain sequences can be used to monitor B cell acute lymphoblastic leukaemia patients for evidence of minimal residual disease [22]. Furthermore, TCR sequencing of tumour-infiltrating T cells can be used to identify cancer-related TCRs that can be tracked based on their distinctive third complementarity determining region (CDR3) sequences into the circulation and used to differentiate individuals with cancer from healthy control individuals [23]. In autoimmunity, single-cell BCR sequencing identified autoantigen-specific B cells in individuals with the neuromuscular autoimmune disease, myasthenia gravis. These B cell clones persisted in spite of B cell depletion therapy, raising the possibility that clone tracking can be used to predict which patients will respond or fail to respond to certain forms of immunomodulatory therapy [24].

Heterogeneous individual responses to type 1 diabetes immunotherapies complicate clinical trial enrolment and evaluation

A number of immunotherapies targeting B and T cells are in development for type 1 diabetes. Of these, a single course of teplizumab (an anti-CD3 monoclonal antibody) provided a 2.5 year delay in clinical diabetes onset compared with placebo [25, 26], and is now approved by the US Food and Drug Administration (FDA) for use in stage 2 type 1 diabetes. Other experimental T cell immunotherapies include abatacept (cytotoxic

T-lymphocyte associated protein 4 [CTLA4]-Ig), which limits T cell activation by blocking CD28 interaction with costimulatory molecules and preserves C-peptide levels in individuals with recent-onset type 1 diabetes [27], and anti-thymocyte globulin (ATG) [28, 29], which increases the ratio of $CD4^+$ regulatory T cells (Treg) to $CD4^+$ conventional T cells (Tconv) and promotes $CD4^+$ T cell exhaustion. B cells have also been targeted: rituximab (anti-CD20) transiently preserved C-peptide in people with recent-onset type 1 diabetes [30]. However, despite the promise of these and other immunotherapies, challenges remain.

A common theme in all type 1 diabetes immunotherapy trials to date is heterogeneity in individual responses. Several studies have attempted to address this heterogeneity by looking at changes in specific populations of immune cells in responder vs non-responder groups. For example, islet-autoreactive T cells with a distinct cytokine secretion pattern were linked to response to alefacept [31]. Additionally, T follicular helper (Tfh) cell profiles predicted abatacept response, with baseline Tfh-like cell frequencies differentiating responder and non-responder groups [32]. However, approaches that focus on changes in phenotypic subsets of lymphocytes can suffer from a signal-to-noise ratio problem. As an example, the Tfh compartment may contain autoreactive T cell clones, but may also contain T cell clones elicited by protective immune responses, particularly following vaccination or infection. T cell effector function, in addition to autoantigen reactivity, also bears consideration. For example, the overall frequency of insulin-reactive T cells was not different between individuals with and without type 1 diabetes, but insulin peptide responses skewed towards proinflammatory (IFN-γ) and away from regulatory (IL-10) cytokine production among cells isolated from donors with type 1 diabetes relative to those present in healthy control donors [33]. Developing molecular-based approaches to focus on the signal (islet-autoreactive T and B cell clones and their phenotypes) rather than the noise (T and B cell clones unrelated to type 1 diabetes) hold potential to further augment evaluation of experimental and approved immunotherapies in type 1 diabetes. Clinical trial endpoints commonly focus on changes in beta cell function, which is inherently downstream of autoimmune destruction of beta cells. Developing approaches to monitor mechanistic immune changes that track with drug efficacy could accelerate clinical trial evaluation over traditional endpoints that require waiting for beta cell damage (or preservation) to become apparent.

State of the science of AIRR-seq data and type 1 diabetes

Until recently, AIRR-seq datasets in type 1 diabetes were relatively few in number, particularly those examining paired TCRα and TCRβ or BCR heavy and light chain sequences, which are key to truly understanding receptor specificity for antigens (reviewed in [34]). However, with the widespread adoption of single-cell RNA-seq (scRNA-seq) and progress in next-generation sequencing (NGS) of TCRs and BCRs, the available datasets have blossomed (representative examples in [18, 31, 35, 36]). Still, there are difficulties with applying AIRR-seq analysis to type 1 diabetes research. A paucity of 'public' type 1 diabetes autoantigen-specific TCRs (i.e. autoantigen-specific TCRs widely shared between people with type 1 diabetes) makes modelling the type 1 diabetes immune response particularly challenging [37, 38]. There is some evidence for shared TCRα chains [39], whilst a recent paper demonstrated a number of public TCRβ chains in type 1 diabetes [40]. Diabetes autoantigen-specific TCRs have also been identified using different methods (reviewed in [41]). At the most stringent level, this has involved re-expressing the TCR in a cell-based system to identify the peptide epitopes recognised [38]. Useful datasets have also been generated through other assays developed to measure islet-specific immune responses (reviewed in [42]). These include in vitro stimulation assays (commonly known as activation-induced marker [AIM] assays [43, 44]), tetramer/multimer sorting from peripheral blood [45], or sequencing of TCR repertoires from the pancreas and pancreatic lymph nodes (pLNs) of organ donors with type 1 diabetes, which are likely to be enriched in diabetes autoantigen-specific TCRs [41]. However, it is expected that all of these repertoires will also include non-specific TCRs, for example from non-specific binding to multimers, by-stander activation in AIM assays, or incidental location of T cells in the pancreas and pLNs. Furthermore, other datasets of TCR sequences and T cell gene expression from participants in longitudinal studies of type 1 diabetes development [18, 46], and clinical trials of immunotherapies in type 1 diabetes, are also emerging [42]. These may include important information on the overall TCR repertoire dynamics and features in people with and at risk for type 1 diabetes; for example, shorter CDR3 length has been observed in people with type 1 diabetes [47]. Nevertheless, integrating these different datasets presents many challenges. Key to the success of this project will be the ability to make the data available in a standardised but accessible format with all possible metadata included. ML approaches offer the tantalising possibility of being able to predict antigen specificity from TCR sequence and HLA typing [41, 48]. Finally, surveying lymphocytes in peripheral blood or formalin-fixed tissues has created a wealth of data, but these data do not fully capture the impact of the inflammatory milieu and dynamic nature of lymphocyte interactions in tissues. Pancreas slice technology may help to address some of these challenges by enabling dynamic study of lymphocytes within live pancreas tissue [49].

In other autoimmune diseases, the ability to isolate B cells from both the peripheral blood and disease target

tissue has led to fascinating insights [50, 51]. However, relatively few studies have examined antigen-specific BCRs in human type 1 diabetes, with existing studies focused primarily on phenotypic changes among insulin-binding B cells [52–54]. As with T cells, this is due to difficulties sampling from the pancreas and low frequencies of the clones of interest in the peripheral blood. BCR sequencing studies in individuals with type 1 diabetes identified a skewed frequency of *BCR heavy chain* (also known as immunoglobulin heavy chain or *IGH*) V gene occurrence in pancreatic draining lymph nodes relative to control individuals [55]. The specificity of these BCRs was, however, not determined in this bulk BCR sequencing study, as paired heavy and light chain sequences were not generated. Until recently, there were no methods available equivalent to T cell AIM assays and tetramer tools. However, technical developments such as dCODE Klickmers [56] and LIBRA-seq (LInking B cell Receptor to Antigen specificity through SEQuencing [57, 58]), which link single-cell BCR sequencing to antigen specificity and gene expression information, now offer the possibility of enriching diabetes autoantigen-specific B cells from the peripheral blood and facilitating epitope mapping of autoreactive B cell responses.

TCR databases with antigen/epitope specificity

A number of databases exist that contain sequencing information on T cells with known (auto)antigen specificity. These include VDJdb [59] (~92,000 TCR sequences), McPAS [60] (~40,000 TCR sequences) and the Immune Epitope Database (IEDB) [61] (~185,000 TCR sequences). These databases curate data from papers that use TCR specificity assays to study TCR–epitope interaction and store this information with relevant metadata about the TCR, epitope and MHC presentation of that epitope. Although these are valuable resources, with less than 10^6 TCRs, they only cover a small fraction of the possible number of receptors (estimated to be 10^{19}) [62]. More problematic for the study of type 1 diabetes, only a fraction of these TCRs are islet antigen-specific and are therefore relevant to the study of the disease (McPas: ~1500; VDJdb: 12; IEDB: ~1700). In contrast, the iReceptor platform searches AIRR-seq data in the ADC, but the ADC does not currently contain antigen/epitope specificity. The iReceptor Gateway currently queries IEDB to determine antigen/epitope specificity [3]. Until the creation of the T1D TCR/BCR Repository (The Type 1 Diabetes T Cell Receptor and B Cell Receptor Repository), there was no type 1 diabetes data in the ADC.

About the iReceptor platform

The iReceptor platform [3] (https://gateway.ireceptor.org) is based on FAIR (Findable, Accessible, Interoperable, Reusable) data principles [63] and allows access to curated AIRR-seq data and associated metadata in the ADC [4]. The overarching goal of the iReceptor platform is to enable researchers to increase the value of their data through sharing with the community. The ADC is an internationally distributed set of repositories that adhere to standards for sharing and analysing AIRR-seq data [64, 65] developed by the AIRR Community [66]. The ADC currently consists of over 5 billion annotated BCR/TCR sequences from more than 90 studies (Figs 1, 2). Figures 2, 3, 4 and 5 provide a detailed example of how to search the type 1 diabetes repository in the ADC through iReceptor.

What sets the iReceptor platform apart from other AIRR-seq (TCR/BCR) platforms is the ability to search extensive metadata (study, de-identified subject, sample, disease, tissues etc.) and AIRR-seq features (BCR/TCR annotations, cell clonal lineages, cell phenotype and cell–antigen reactivity) across all of the studies and all of the repositories in the ADC. In addition to federating the repositories in the ADC, iReceptor also links to external repositories such as NCBI to allow access to source sequence data, IEDB for immune receptor–antigen specificity [61], ArrayExpress/Gene Expression Omnibus (GEO) for gene expression [67], and Ensembl for definitions of expressed genes [68, 69]. Through this integration, iReceptor allows researchers to find relevant data that will help answer complex research questions at a scale and complexity that is challenging to achieve otherwise. For example, researchers can search for AIRR-seq data with specific characteristics at various levels, including study (e.g. methodology), subject (e.g. age, disease state, HLA genotype), sample (e.g. tissue type), and sample processing (e.g. cell phenotype) (Fig. 3).

In type 1 diabetes, the HLA type is particularly relevant, given the associations between TCRs and HLA and the high degree of HLA association with type 1 diabetes [1, 70]. Searches that filter samples on the basis of these metadata narrow the scope of data being considered from billions of records to a smaller subset of records that are more relevant to the research question and more amenable to detailed analysis. These filtered data can then be searched for BCR/TCR (e.g. CDR3, VDJ annotations) and cell characteristics (e.g. paired chains, expressed genes, cell reactivity) (Fig. 4). Once relevant data are found, the data can then be either downloaded by the researcher for analysis in their custom analysis pipelines or submitted to one of the analysis pipelines built into the iReceptor platform (e.g. CoNGA [71] or CellTypist [72]) (Fig. 5).

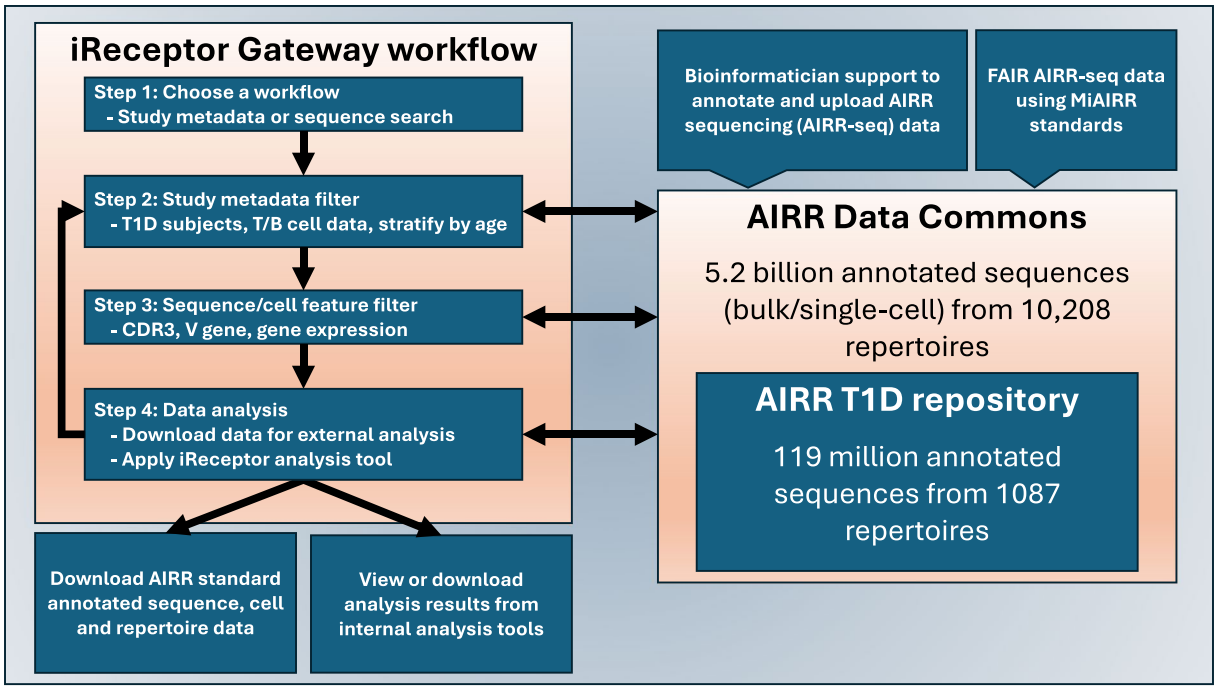


Fig. 1 Overview of how the AIRR type 1 diabetes repository sits within the AIRR Data Commons and is accessed through the iReceptor Gateway. T1D, type 1 diabetes

Goals of the T1D AIRR Consortium

In a collaboration led by the A. Michels lab (University of Colorado), the T. Brusko lab (University of Florida), the iReceptor team (Simon Fraser University), the Human Pancreas Analysis Program (HPAP, [18, 72]), and the(sugar) science group (https://thesugarscience.org/), the aim of the T1D AIRR Consortium is to bring together type 1 diabetes-related AIRR-seq datasets into the T1D TCR/BCR Repository that are freely available, searchable and downloadable as a complete dataset for analysis in researchers' pipeline of choice (e.g. R or Python). In order to provide data that are ML and artificial intelligence (AI) ready, type 1 diabetes datasets in the ADC, accessed through iReceptor, are being curated with optional, customised metadata fields including HLA typing, presence of AAb markers, immunoglobulin and TCR germline gene polymorphisms, disease progression, immune cell source, detailed protocol information for enrichment (e.g. AIM assays or tetramer sorting), gene expression (GEX) data and antigen specificity. It will encompass state-of-the-art scRNA-seq datasets generated by individual investigators, large-scale TCR sequencing projects (e.g. HPAP [18, 73]), and historical datasets [55].

The AIRR Community, including members of the iReceptor team, has published a general methods paper for sharing and reusing AIRR-seq data [74]. Our type 1 diabetes initiative extends this methodology to focus on identifying and onboarding type 1 diabetes research groups with the process of data curation in the ADC. Rigorous data curation that maximises the value of research data is costly and time consuming (see Note 4, page 469 in [74]). One goal of the T1D AIRR Consortium is to increase the capacity to share data. This is primarily accomplished through standardising study and data processing protocols such that data curation into the T1D TCR/BCR Repository is part of the general practice in the labs of the T1D AIRR Consortium members. In our experience, data curation is initially most efficient when study, domain, bioinformatics and technology experts work together to establish a curation pipeline for a research group, but once that pipeline is established the incremental cost of curation to maximise data reuse is greatly reduced. Each subsequent use of the curation pipeline increases its efficiency, making the data curation process easier for the research group. Type 1 diabetes researchers may also wish to include depositing data in the ADC as part of their data management plans in grant applications, as a straightforward way of fulfilling their FAIR data obligations. This is in fact already happening for the HPAP bulk TCR data, with plans to follow for the bulk BCR and single-cell data.

The Minimal information about AIRR (MiAIRR) standard [64] was developed by the AIRR Community as a guide to the documentation and curation of studies that involve AIRR-seq data and is the standard that is used to describe study metadata in the ADC. Up to date information, including detailed descriptions of the MiAIRR fields (Fig. 6), can be found on the AIRR Community

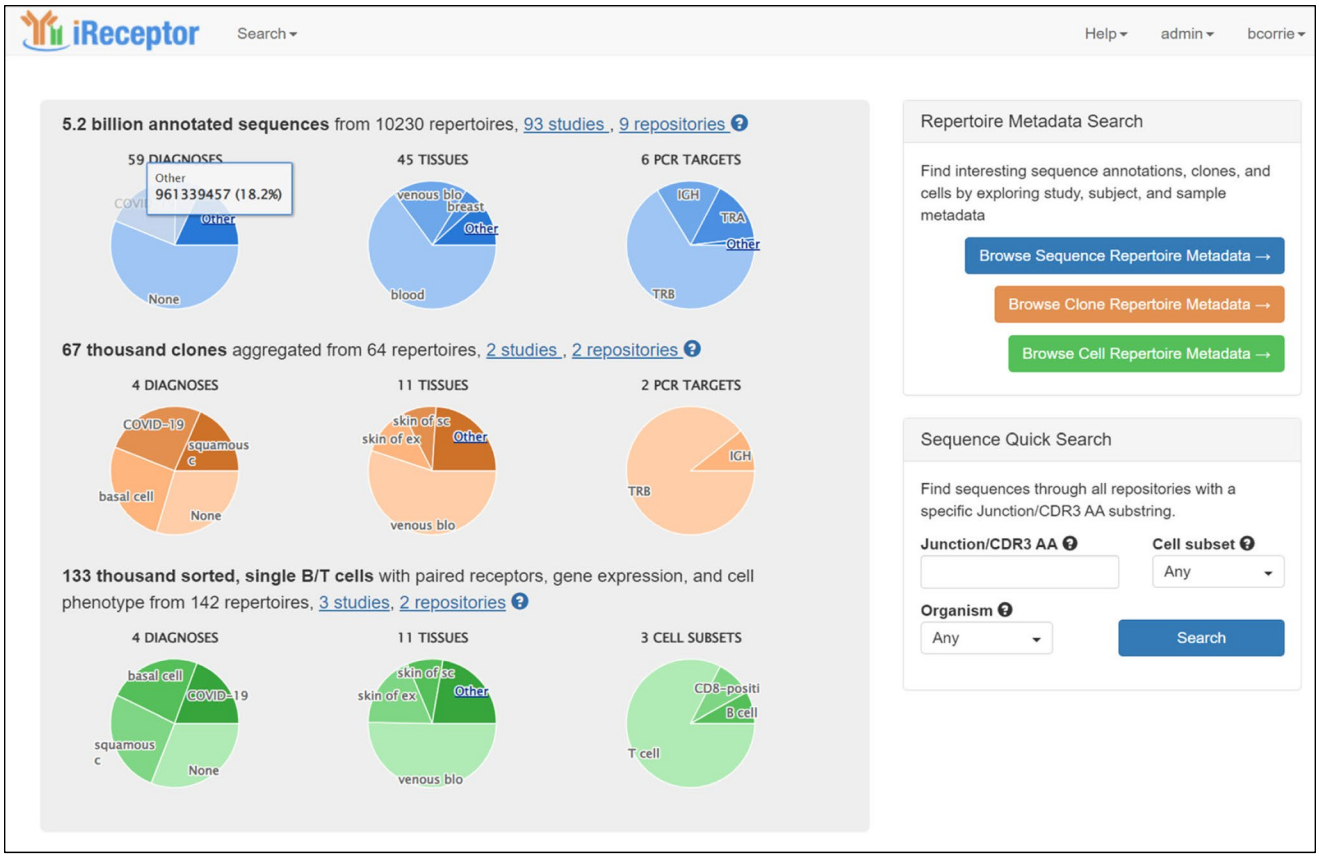


Fig. 2 iReceptor Gateway overview page of the data in the ADC. Summary statistics of the annotated sequence, clone and single-cell data available in the ADC are provided. Users can select a 'Repertoire Metadata Search' or a 'Sequence Quick Search' workflow. The metadata search workflow can be done from the perspective of annotated sequences, clones or single-cell data. This is the user interface for Step 1 in the iReceptor Gateway workflow from Fig. 1. Reproduced with the permission of the iReceptor project

documentation website (https://docs.airr-community.org/en/stable/standards/overview.html). In brief, MiAIRR consists of several classes of metadata:

1. **Study, subject and diagnosis**: describes the study (e.g. title, research group, funding, publications), the study methodology used, the subjects in the study (species, age, sex), genotype characteristics (HLA, BCR and TCR genotype), and clinical characteristics (AAb status, disease diagnosis and duration, vaccinations, treatments such as immunotherapies).
2. **Sample collection:** describes the sample collection process used, tissue (tissue type, anatomical location of sample), disease state of tissue, sample collection timepoints, and details of any enrichment strategies such as tetramer sorting.
3. **Sample processing and sequencing**: describes how the samples were processed in preparation for sequencing and how that sequencing was performed, including sample processing, cell processing, library preparation and the sequencing process (e.g. sequencing platform, sequencing facility, sequencing date).
4. **Raw sequences**: includes Sequence Read Archive (SRA) data.
5. **Data processing**: describes how the sequenced data were processed to produce the AIRR-seq related digital artefacts that are stored in the ADC (annotated BCR/TCR sequences, cells, cell expression, cell reactivity), including quality control processing, primer processing, software packages and versions used, and other data processing protocols.
6. **Processed sequences with annotations**: includes VDJ germline reference databases, V(D)J gene segment calls, IMGT nucleotide and amino acid sequences (www.imgt.org/), and read counts.

Although the MiAIRR standard contains a large number of fields, it does not require that all studies must capture all fields. Rather, it should be considered a guide as to what a research group might want to include in the planning and

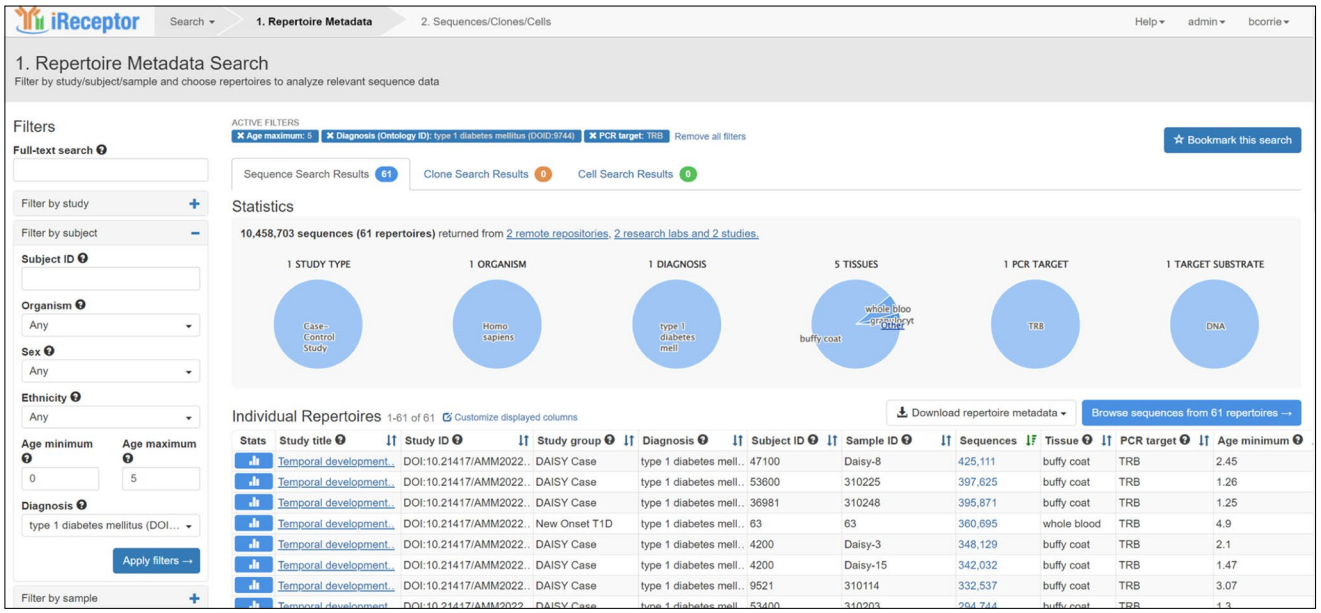


Fig. 3 iReceptor Gateway 'Repertoire Metadata Search' page for annotated sequence data. Repertoires are filtered so that the data are limited to TCRβ sequences from individuals with type 1 diabetes. The data are further stratified to select only data from individuals between 0 and 5 years of age. Over 10 million annotated sequences from 61 repertoires and two studies in the ADC meet this search criteria. All of these type 1 diabetes data are retrieved from the AIRR T1D TCR/BCR Repository. This is the user interface for Step 2 in the iReceptor Gateway workflow from Fig. 1. Reproduced with the permission of the iReceptor project. TRB, alternative name for TCRβ

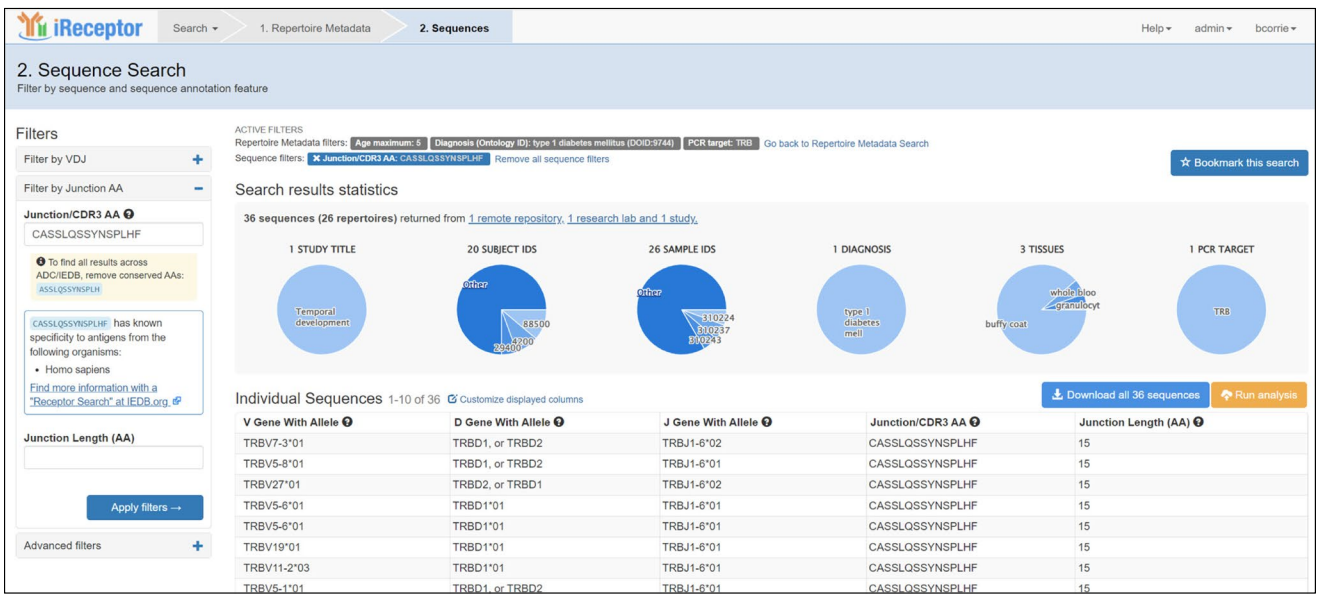


Fig. 4 iReceptor Gateway 'Sequence Search' page. A search for a CDR3 amino acid sequence of interest (CASSLQSSYNSPLHF) is performed on the TCRβ data from an individual with type 1 diabetes between 0 and 5 years of age in the ADC. The iReceptor Gateway reports that it found 36 such sequences across 26 repertoires and 20 subjects, suggesting that this is a 'public' CDR3. It also reports that this CDR3 has a known antigen/epitope specificity and provides a link to external resources (e.g. IEDB) to find out further information about the antigen/epitope. At this stage, the user can either download the data for offline analysis or perform internally supported analyses of these data. This is the user interface for Step 3 in the iReceptor Gateway workflow from Fig. 1. Reproduced with the permission of the iReceptor project. TRB, alternative name for TCRβ

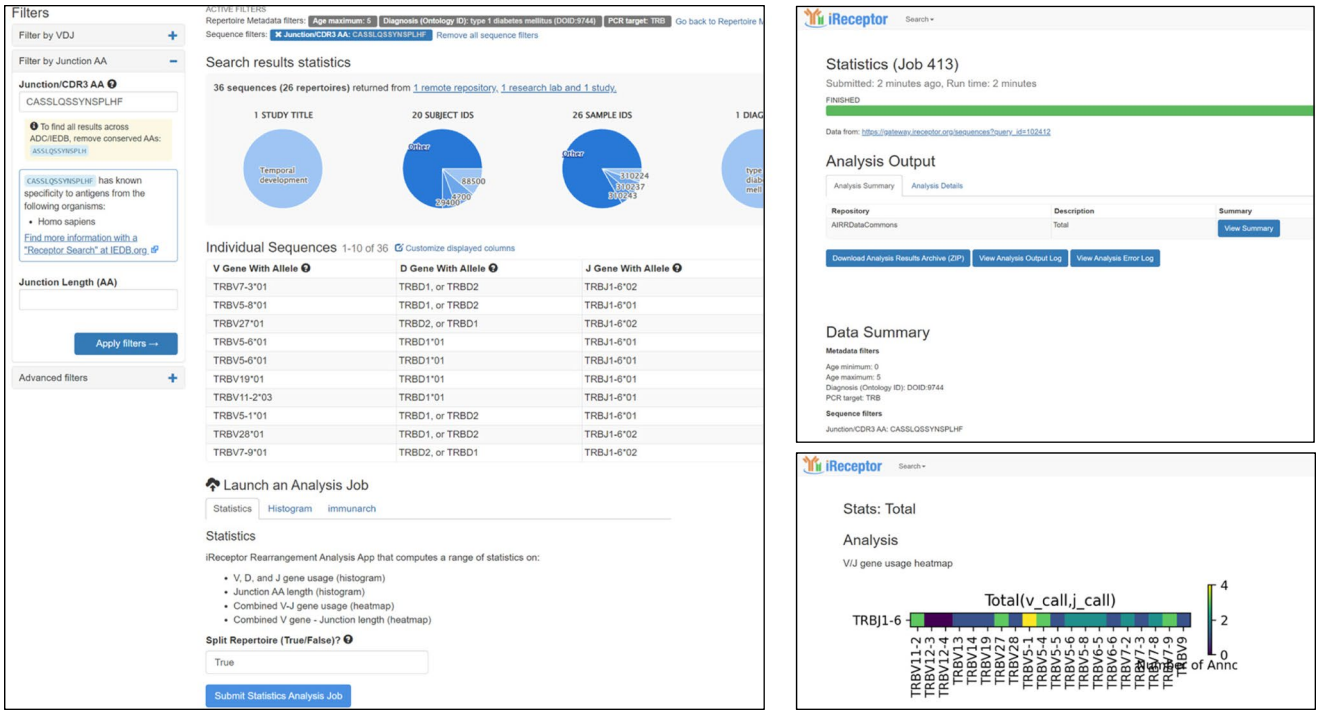


Fig. 5 iReceptor internal statistical analysis of the selected data. The left frame is a view of a subregion of the 'Sequence Search' page, showing the set of analysis tools that are available for the selected data. The top right frame is the output of the 'Statistics' analysis tool. The user can either download the analysis results or view a summary of the analysis output. The search parameters for the repertoire metadata as well as the sequence features that were used to select the analysed data are displayed. The bottom right frame is a view of a subset of the analysis results, in this case a V/J gene heatmap for the CDR3 of interest. The results of this analysis provide the frequency for all TCRβ receptors (V/J gene combinations) for the CDR3 sequence CASSLQSSYNSPLHF from individuals with type 1 diabetes between 0 and 5 years of age in the ADC. This is the user interface for Step 4 in the iReceptor Gateway workflow from Fig. 1. Reproduced with the permission of the iReceptor project. TRB, alternative name for TCRβ

curation of a study. It is worth noting that, even though a metadata field might not be considered important to answer an investigator's specific research question, the field might be important for others to reuse those data. The AIRR Community recommends that as many of the MiAIRR fields be populated as possible in order to maximise data reuse.

Onboarding data

The T1D AIRR Consortium is now a solid community of collaborators who champion the benefit of FAIR research data. Led by the(sugar)science team, outreach to the type 1 diabetes community has resulted in steady growth of consortium members. Each collaborator researching type 1 diabetes who has expressed a willingness to share data is contacted by the data curation team to establish scientific and bioinformatics expertise within the study team. This process of proactive collaboration is critical to the efficient initial curation of study metadata for that research group. Example study metadata worksheets are provided, using the AIRR Community standards, as a basis for this curation, and the study domain experts work closely with the iReceptor curation team to establish a rigorous set of metadata, including type 1 diabetes-specific metadata elements, for each study. This is an iterative process, where the study metadata are refined over time. Once established, this curation pipeline is followed for subsequent studies. Current and new studies use the study metadata curation pipeline up front as part of the planning and study execution process, making it straightforward and efficient to curate new studies in the ADC.

The iReceptor curation team also works with the study bioinformatician to prepare the annotated AIRR-seq or scRNA-seq data for loading. Although the iReceptor platform can load the output of many of the standard VDJ and single-cell annotation tool chains (e.g. IgBLAST [www.ncbi.nlm.nih.gov/igblast/], MiXCR [https://mixcr.com/], 10X CellRanger [www.10xgenomics.com/support/software/cell-ranger/latest], Adaptive ImmuneAccess [https://clients.adaptivebiotech.com/immuneaccess]), it is sometimes necessary to transform these data as part of this preparation.

It is important to note that the key benefit of including the data in the ADC is the value-added curation of standards-based AIRR-seq and scRNA-seq data and their associated

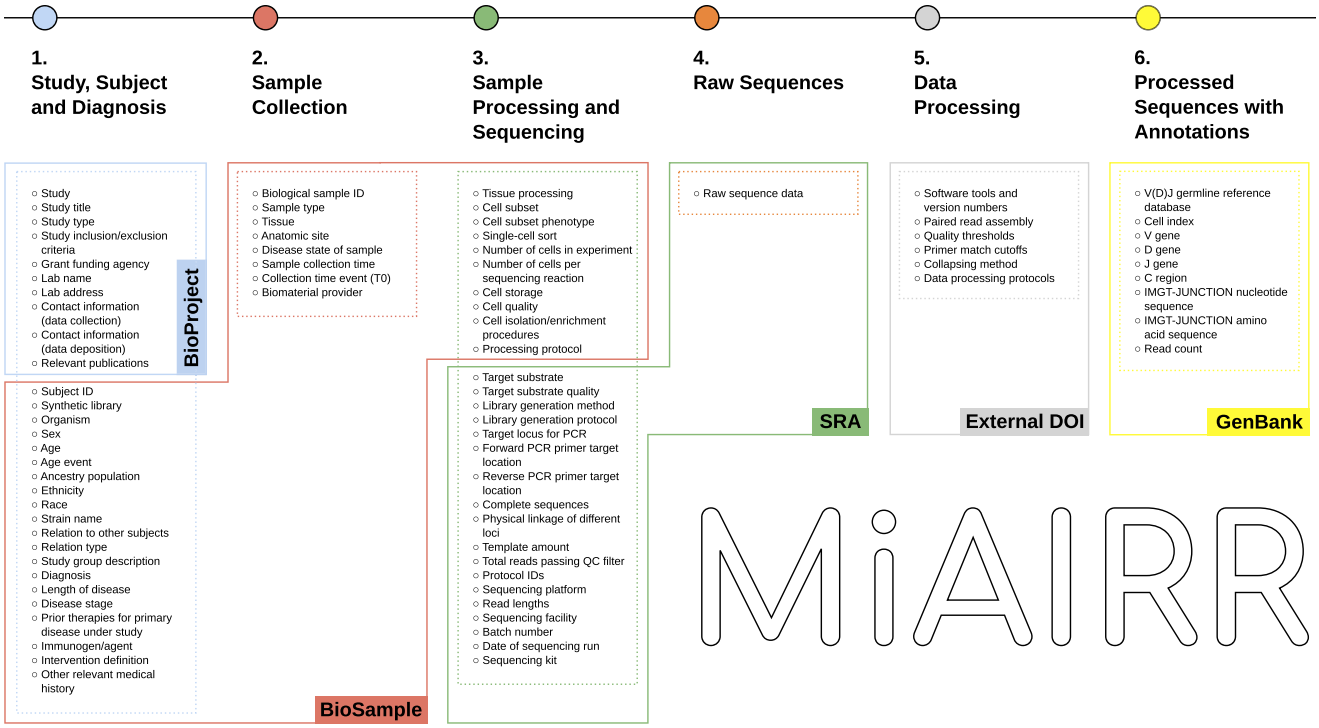


Fig. 6 An overview of the six MiARR [64] standard classes of data and associated data fields. More details on the standard are available at the AIRR Community website: https://docs.airr-community.org/en/stable/standards/overview.html. Copyright AIRR Community, reproduced under a CC BY 4.0 licence. QC, quality control

metadata, which enables data reuse. However, this does not preclude the necessity of maintaining the raw sequence data in a standard sequence archive such as the International Nucleotide Sequence Database Collaboration repositories (INSDC; www.insdc.org/), European Nucleotide Archive (ENA; www.ebi.ac.uk/ena/browser/home) [75], SRA (www.ncbi.nlm.nih.gov/sra), database of Genotypes and Phenotypes (dbGaP; www.ncbi.nlm.nih.gov/gap/), ArrayExpress (www.ebi.ac.uk/biostudies/arrayexpress) or GEO (https://www.ncbi.nlm.nih.gov/geo/).

The ADC has been established as an international resource, with repositories currently spanning Canada, the USA and Germany. The AIRR Community encourages researchers to curate their own data at their home institution, and software has been developed that enables this through the download and installation of the open source iReceptor Turnkey software (https://github.com/sfu-ireceptor/turnkey-service-php). Research groups with large amounts of AIRR-seq data are encouraged to consider running a repository of their own to curate and share their data. The iReceptor team uses the iReceptor Turnkey to run several ADC repositories, including the iReceptor Public Archive (IPA), a COVID-19 repository, as well as other repositories in collaboration with other researchers.

As part of the T1D AIRR Consortium, the iReceptor team runs the T1D TCR/BCR Repository. Once the study metadata and the sequencing data have been prepared for a study from a group in the consortium, the iReceptor curation team loads the data and study metadata into the T1D TCR/BCR Repository. Although documentation is available for loading data into the iReceptor Turnkey (https://github.com/sfu-ireceptor/turnkey-service-php#loading-your-own-data), the iReceptor team performs this role on behalf of collaborators in the consortium. The iReceptor team also supports the repository operationally to ensure reliability. Once the data are loaded, the repository is put into production, and the data are then available to be queried, analysed or downloaded as part of the ADC by any user of the iReceptor platform.

The impact of sharing AIRR-seq data and in particular, the impact of the iReceptor platform in facilitating this data sharing can be demonstrated by the fact that the iReceptor paper [3] has been cited over 100 times in the literature as a source to store, find or reuse AIRR-seq data (Fig. 7).

The first citation of data from the T1D TCR/BCR Repository has just been published, where data from the type 1 diabetes study from Seay et al [55] (loaded into the repository in March 2023) was reused in a recent cancer study from Sammut et al [76] that explored BCR clonal centrality in immunosurveillance across cancer, type 1 diabetes and multiple sclerosis.

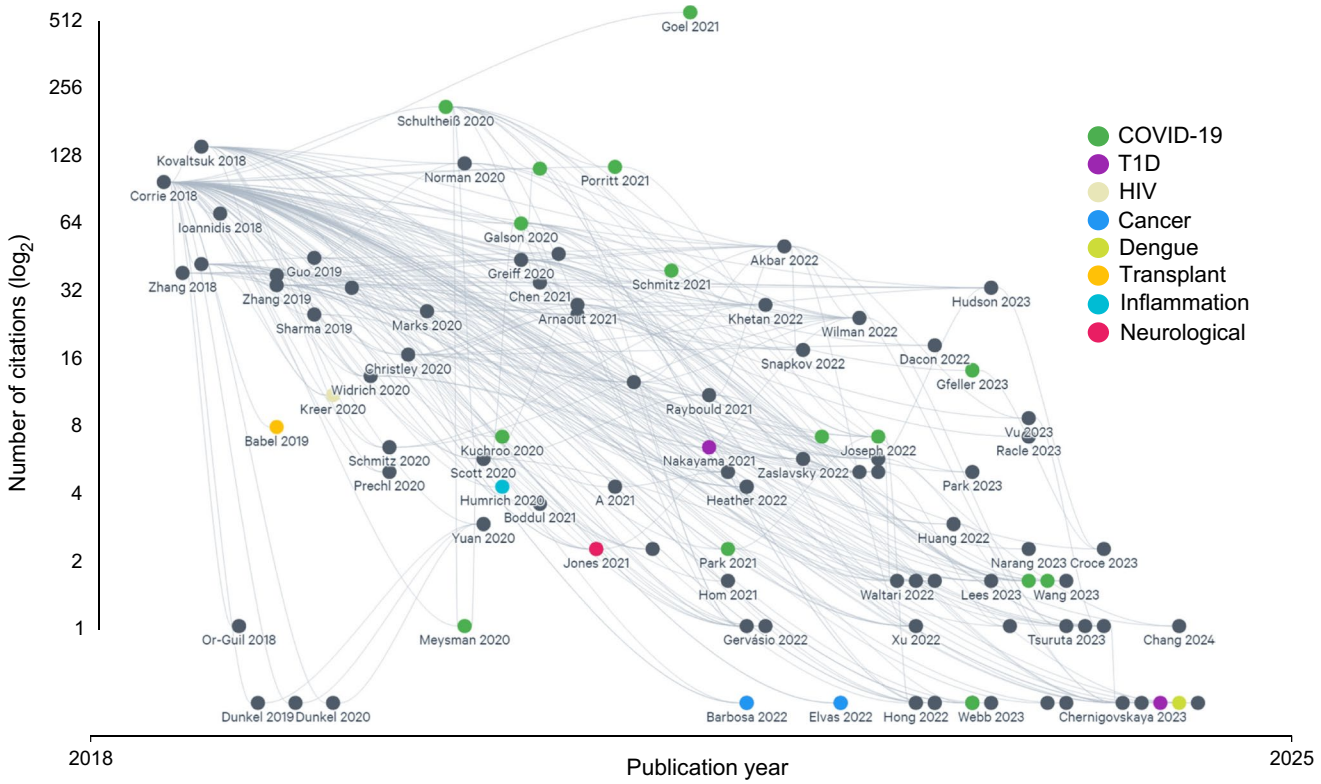


Fig. 7 Citations as a metric of iReceptor manuscript influence, by disease. Citations of the iReceptor paper (Corrie et al 2018 [3], top left corner), coloured by high level disease classification. Grey citations are not disease specific and either represent standards (e.g. AIRR Standards), tools (e.g. ML), or techniques (e.g. repertoire classification) papers. Grey connecting lines indicate manuscripts that were cited in subsequent publications. Interactive visualisation of citation data is available at https://app.litmaps.com/shared/12ead8ba-35d9-4232-91a0-26da9e3c3f59; colour coding in the interactive visualisation is different from the figure above. T1D, type 1 diabetes

Use cases

Gaining insight into type 1 diabetes The overarching aim of the T1D TCR/BCR Repository is to enable development of strategies to track immune repertoire changes that predict type 1 diabetes progression risk and/or individual response to immunotherapy. In order to achieve this, it is important to be able to reach a critical mass of type 1 diabetes data that is available to the research community. To date, six studies have been curated by the T1D AIRR Consortium into the T1D TCR/BCR Repository and are now available as part of the ADC, consisting of over 110 million annotated sequences from 366 individuals, eight tissue types and 12 T cell and B cell subsets. Table 2 provides an overview of publicly available type 1 diabetes repertoire datasets in the ADC.

Once such a critical mass is reached, researchers can reuse these data to explore new hypotheses, which hold potential to drive new developments in type 1 diabetes. For TCRs, this could include analysis of CDR3 length [47], motif sharing [78], modelling of repertoire changes during type 1 diabetes development [18] and during immunotherapy, prediction of TCR specificity [48], understanding the role of immune response to viruses in type 1 diabetes development [79], and understanding features of type 1 diabetes autoantigen-specific TCRs such as positioning of charged amino acids and hydrophobic residues within the CDR3 region. It has recently been demonstrated that specific diabetes autoantigen-specific TCRβ chains showed distinct patterns of accumulation during type 1 diabetes development, suggesting that they could be used as biomarkers [40] and an increase in validated type 1 diabetes-specific TCR sequences will expedite this process. Furthermore, the ADC contains many TCR repertoires from healthy control individuals, which allows immediate comparison of type 1 diabetes results with other repertoires (critical in determining the limits of public clonotypes) and will facilitate the understanding of the differences between people with and without type 1 diabetes.

BCRs in type 1 diabetes remain a largely untapped resource [34], but it is envisaged that identification of specific CDR3s may be additive when combined with islet AAb as predictive tools. The iReceptor platform enables deposition of linked antigen binding, single-cell BCR sequencing, gene and protein expression data types, which can be exported for analysis using immune repertoire tools such as the Immcantation suite

Table 2 Repertoire data from type 1 diabetes studies currently deposited in ADC by the T1D AIRR Consortium

Study	Cell and data type(s)	Tissue type(s)	Single-cell or bulk?	No. of sequences	Disease metadata available
Gomez-Tourino et al [47]	Central memory $CD4^+$ T cells, true naive $CD4^+$ T cells, $CD4^+$ Treg, stem-cell like memory T cells (Tscm), $CD4^+$ T cells (α/β TCR)	Blood	Bulk	13M T1D, 21M NT1D	Age, sex, race, diabetes duration
Mitchell et al [18]	T cells (TCRβ)	Blood	Bulk	46M T1D, 15M NT1D	Age, sex, ethnicity, diabetes duration, islet AAb, HLA class I and II types, longitudinal samples
Seay et al [55]	B cells (BCR heavy chain), $CD4^+$ T cells, $CD8^+$ T cells, $CD4^+$ Treg (TCRβ)	Blood, irrelevant lymph node, islet of Langerhans, pLN, spleen	Bulk	5.2M T1D, 1.7M T2D, 3.6M NT1D	Age, diabetes duration, prior therapies, medical history, HLA, islet AAb
Anderson et al [38]	$CD4^+$ T cells, $CD8^+$T cells (α/β TCR)	Islet of Langerhans	Single-cell	5K T1D, 2K NT1D	Age, diabetes duration, prior therapies, medical history, HLA, islet AAb
Culina et al [77]	Central memory $CD4^+$ T cells, central memory $CD8^+$ T cells, effector memory $CD8^+$ T cells, naive thymus-derived $CD8^+$ T cells, naive thymus-derived $CD4^+$ T cells (α/β TCR)	Blood	Bulk	4.3M T1D, 1.8M T1DR, 3.6M NT1D	Age, ethnicity, diabetes duration, HLA, islet AAb
HPAP T1D TCRβ data [73]	TCRβ	Spleen	Bulk	415K T1D, 107K T2D, 658K NT1D	Age, medical history, islet AAb

K, thousand; M, million; NT1D, control group without type 1 diabetes; T1D, type 1 diabetes; T1DR, type 1 diabetes risk; T2D, type 2 diabetes

of tools (including Change-O, Alakazam and SHazaM [80]) to identify immune repertoire features, including B cell V(D)J identities, CDR composition, clonotypes, isotypes, per cent somatic hypermutation, and BCR physiochemical properties (e.g. CDR charge, length, polarity). Several analysis packages exist that support gene expression and pathway analysis (e.g. Seurat [81], g:Profiler [82] and CytoScape [83]), which can be merged with immune repertoire analysis outputs. These additional data types can also be housed in the T1D iReceptor database, along with AIRR data.

COVID-19 as an example of rapid data sharing that led to improved understanding of immune contributions to disease We demonstrate the value of disease-specific repositories, data sharing and data reuse within the ADC that was catalysed during the COVID-19 pandemic [84] as an example of our vision for the type 1 diabetes community. Early in the COVID-19 pandemic, it was realised that the AIRR Community and, in particular, the ADC, were well positioned to have a significant impact on the understanding and treatment of COVID-19. Like in many areas, COVID-19 emphasised the need to share data, and in March 2020, the AIRR Community made a call for sharing of AIRR-seq data from consented COVID-19 patients (https://www.antibodysociety.org/airr-community/covid-19-demands-increased-public-sharing-of-biomedical-research-data/). In May of 2020, one of the first studies with publicly available AIRR-seq data was published as a pre-print [85], and by June 2020, the iReceptor team had provisioned an ADC COVID-19 repository and curated these data. This was rapidly followed by the curation of a total of 9 COVID-19 studies by the end of 2020. In addition, iReceptor housed data to evaluate SARS-CoV-2 vaccine responses, providing large numbers of spike-binding BCR sequence data to the community (for example [86]).

The impact of the availability of disease-specific data (COVID-19 in this case) for reuse by other researchers can be demonstrated by considering several use cases from the literature. For example, Heming et al [87] studied the immune cell profiles of cerebrospinal fluid from eight neuro-COVID patients, identifying expanded clonotypes in mild and severe neuro-COVID. They utilised iReceptor to discover subjects diagnosed with COVID-19 from seven studies in the ADC. They retrieved >180 million CDR3 amino acid sequences from COVID-19 patients in the ADC and determined the overlap of CDR3 amino acid sequences from their own data, identifying SARS-CoV-2-specific clones from cerebrospinal fluid. They also used TCR sequences from healthy control individuals in the ADC and by comparing the repertoires demonstrated that TCRs were shared between the cerebrospinal fluid and periphery in COVID-19.

It should be noted that the key feature of this analysis is the CDR3 amino acid sequence, which is a fundamental part of the value-added annotation that the ADC provides. If these data were not available in this form from the ADC, the raw sequence data from all the studies listed above would have had to be re-annotated at enormous cost in terms of time, human effort and computational resources. Through the iReceptor platform, the researchers were easily able to find and access relevant data (from individuals both with and without COVID-19), download that data for analysis and comparison with their own data, and directly reuse the identified CDR3 regions in these data for their analysis. Similarly, a paper by Schultheiß et al published one of the early studies on COVID-19 that tracked disease progression over time [2]. The authors worked with the iReceptor team to ensure the data were available in the ADC when the paper was published. This paper is highly cited, and currently, 15 publications cite both the iReceptor paper [3] and the Schultheiß paper, indicating the likely reuse of the Schultheiß data from the ADC.

More recently, we have seen extensive citations of iReceptor from papers discussing the value of the curated and annotated data in the ADC for training and assessing ML algorithms [88]. The power of the iReceptor platform in conjunction with ML approaches is also demonstrated in the COVID-19 arena. Park et al used iReceptor to obtain TCR datasets from individuals with COVID-19 [89]. They then used both standard analysis tools, such as Immunarch [90], OLGA [91], and GLIPH2 [92], as well as a variety of k-mer-based ML models, to gain insight into the TCR repertoires. They identified features of COVID-19-specific TCRs such as gene usage and motif. Moreover, the ML approaches could identify not only whether a person had COVID-19 but also the disease severity, based solely on TCR repertoire features.

COVID-19 BCR datasets using iReceptor have been similarly used to compare and validate protein language models that predict BCR specificity [93] and recognise COVID-19-specific BCR signatures [94]. Use of iReceptor to understand BCR repertoires was demonstrated by Waltari et al, who accessed datasets through iReceptor to develop AIRRscape, an R Shiny-based app and website to visualise BCR repertoires [95]. They downloaded bulk repertoire BCR datasets from COVID-19-infected individuals from iReceptor and complemented them with datasets of known monoclonal antibodies. They demonstrated features of the bulk repertoires that differed between the distinct infections, such as convergent motifs and V-J pairings.

The iReceptor team was able to work directly with many of the COVID-19 researchers to curate and make these data available. It is this model of data sharing that the T1D AIRR Consortium will emulate and promote.

Future collaborations and efforts

Looking ahead, the T1D AIRR Consortium will focus both on uploading historical type 1 diabetes datasets and enabling the type 1 diabetes community to upload

contemporaneous datasets. Outreach through efforts of the(sugar)science will continue to recruit scientists from the type 1 diabetes community to join the T1D AIRR Consortium, enabling continual growth of studies in the T1D TCR/BCR Repository from a range of research groups. The T1D AIRR Consortium encourages researchers to share their valuable type 1 diabetes data and welcomes new collaborators to the consortium. To learn more and join the monthly consortium working group, contact: T1DAIRR@thesugarscience.org

Acknowledgements The authors would like to acknowledge the contributions of the(sugar)science, a type 1 diabetes research advocacy organisation, for initiating, organising and hosting the consortium activities. All screen grabs of the iReceptor Gateway are provided with the permission of the iReceptor project.

Type 1 Diabetes AIRR Consortium members: Erin Baschal [Barbara Davis Center for Diabetes, University of Colorado School of Medicine, Aurora, CO 80045, USA], Karen Cerosaletti [Center for Translational Immunology, Benaroya Research Institute at Virginia Mason, Seattle, WA], Lorissa Corrie [iReceptor Genomic Services, Summerland, BC, Canada], Iria Gomez-Tourino [Centre for Research in Molecular Medicine and Chronic Diseases (CiMUS), University of Santiago de Compostela, Santiago de Compostela, Spain], Lauren Higdon [Diabetes Center, University of California San Francisco, San Francisco, CA, USA], Sally C. Kent [Diabetes Center of Excellence, Department of Medicine, University of Massachusetts Medical Chan School, Worcester, MA, USA], Peter Linsley [Benaroya Research Institute, Seattle, WA 98101, USA], Maki Nakayama [Barbara Davis Center for Diabetes, University of Colorado School of Medicine, Aurora, CO 80045, USA], Kira Neller [iReceptor Genomic Services, Summerland, BC, Canada], William E. Ruff [Repertoire Immune Medicines, Cambridge, MA 02139, USA], Luc Teyton [Department of Immunology and Microbiology, Scripps Research, La Jolla, CA 92037, USA]

Data availability All data in Table 2 are available from https://gateway.ireceptor.org/login

Data from Fig. 7 are available at https://app.litmaps.com/shared/12ead8ba-35d9-4232-91a0-26da9e3c3f59

Funding SJH is funded by the Diabetes Research and Wellness Foundation Professor David Matthews Non-Clinical Research Fellowship, The Leona M. and Harry B. Helmsley Charitable Trust by grant #2101-04969, and this work was supported on behalf of the "Steve Morgan Foundation Type 1 Diabetes Grand Challenge" by Breakthrough T1D UK, formerly JDRF and SMF (grant numbers 2-SRA-2024-1474-M-N and 2-SRA-2024-1473-M-N). This work was also funded by NIH grants R01 DK131070 (RHB), DK108868 (AWM), DK032083 (AWM), DK099317 (AWM), U24 AI177622 (FB, BC) and U01-DK112217 (ETLP). TMB is funded by NIH P01 AI042288 and AWM and TMB by The Leona M. and Harry B. Helmsley Charitable Trust by grant #2301-06562.

Authors' relationships and activities TMB has consulted for Repertoire Immune Medicines and both TMB and AWM have received in-kind sequencing support by Adaptive Biotech. BC and FB are both researchers on the iReceptor team at Simon Fraser University and are partners in iReceptor Genomic Services. The other authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement SJH, RHB, ALP, ETLP, MW, FB, AWM and TMB were responsible for conceptualisation and design of the work. BC was responsible for data curation: MW provided project administration for the Type 1 Diabetes AIRR Consortium. MW, SJH, RHB and BC wrote the original draft. All authors reviewed and edited the original draft and approved the final manuscript for publication. TMB and AWM are guarantors of this work

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Emerson RO, DeWitt WS, Vignali M et al (2017) Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat Genet 49(5):659–665. https://doi.org/10.1038/ng.3822
2. Schultheiß C, Paschold L, Simnica D, et al (2020) Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures associated with severity of disease. Immunity S107476132030279X. https://doi.org/10.1016/j.immuni.2020.06.024
3. Corrie BD, Marthandan N, Zimonja B et al (2018) iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. Immunol Rev 284(1):24–41. https://doi.org/10.1111/imr.12666
4. Christley S, Aguiar A, Blanck G et al (2020) The ADC API: A Web API for the programmatic query of the AIRR data commons. Front Big Data 3:22. https://doi.org/10.3389/fdata.2020.00022
5. Ziegler AG, Rewers M, Simell O et al (2013) Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA 309(23):2473. https://doi.org/10.1001/jama.2013.6285
6. Felton JL, Redondo MJ, Oram RA et al (2024) Islet autoantibodies as precision diagnostic tools to characterize heterogeneity in type 1 diabetes: a systematic review. Commun Med 4(1):66. https://doi.org/10.1038/s43856-024-00478-y
7. Insel RA, Dunne JL, Atkinson MA et al (2015) Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the endocrine society, and the American diabetes association. Diabetes Care 38(10):1964–1974. https://doi.org/10.2337/dc15-1419
8. Achenbach P, Lampasona V, Landherr U et al (2009) Autoantibodies to zinc transporter 8 and SLC30A8 genotype stratify type 1 diabetes risk. Diabetologia 52(9):1881–1888. https://doi.org/10.1007/s00125-009-1438-0
9. Krischer JP, Lynch KF, Lernmark Å et al (2017) Genetic and environmental interactions modify the risk of diabetes-related autoimmunity by 6 years of age: the TEDDY study. Diabetes Care 40(9):1194–1202. https://doi.org/10.2337/dc17-0238
10. Besser REJ, Bell KJ, Couper JJ et al (2022) ISPAD clinical practice consensus guidelines 2022: stages of type 1 diabetes

in children and adolescents. Pediatr Diabetes 23(8):1175–1187. https://doi.org/10.1111/pedi.13410
11. Galderisi A, Carr ALJ, Martino M, Taylor P, Senior P, Dayan C (2023) Quantifying beta cell function in the preclinical stages of type 1 diabetes. Diabetologia 66(12):2189–2199. https://doi.org/10.1007/s00125-023-06011-5
12. Jacobsen LM, Bundy BN, Ismail HM et al (2022) Index60 is superior to HbA1c for identifying individuals at high risk for type 1 diabetes. J Clin Endocrinol Metab 107(10):2784–2792. https://doi.org/10.1210/clinem/dgac440
13. So M, O'Rourke C, Bahnson HT, Greenbaum CJ, Speake C (2020) Autoantibody reversion: changing risk categories in multiple-autoantibody–positive individuals. Diabetes Care 43(4):913–917. https://doi.org/10.2337/dc19-1731
14. Taylor PN, Collins KS, Lam A et al (2023) C-peptide and metabolic outcomes in trials of disease modifying therapy in new-onset type 1 diabetes: an individual participant meta-analysis. Lancet Diabetes Endocrinol 11(12):915–925. https://doi.org/10.1016/S2213-8587(23)00267-X
15. Ylescupidez A, Bahnson HT, O'Rourke C, Lord S, Speake C, Greenbaum CJ (2023) A standardized metric to enhance clinical trial design and outcome interpretation in type 1 diabetes. Nat Commun 14(1):7214. https://doi.org/10.1038/s41467-023-42581-z
16. Luckett AM, Weedon MN, Hawkes G, Leslie RD, Oram RA, Grant SFA (2023) Utility of genetic risk scores in type 1 diabetes. Diabetologia 66(9):1589–1600. https://doi.org/10.1007/s00125-023-05955-y
17. Ferrat LA, Vehik K, Sharp SA et al (2020) A combined risk score enhances prediction of type 1 diabetes among susceptible children. Nat Med 26(8):1247–1255. https://doi.org/10.1038/s41591-020-0930-4
18. Mitchell AM, Baschal EE, McDaniel KA et al (2022) Temporal development of T cell receptor repertoires during childhood in health and disease. JCI Insight 7(18):e161885. https://doi.org/10.1172/jci.insight.161885
19. Shoukat MS, Foers AD, Woodmansey S, Evans SC, Fowler A, Soilleux EJ (2021) Use of machine learning to identify a T cell response to SARS-CoV-2. Cell Rep Med 2(2):100192. https://doi.org/10.1016/j.xcrm.2021.100192
20. Sidhom J-W, Baras AS (2021) Deep learning identifies antigenic determinants of severe SARS-CoV-2 infection within T-cell repertoires. Sci Rep 11(1):14275. https://doi.org/10.1038/s41598-021-93608-8
21. Gittelman RM, Lavezzo E, Snyder TM et al (2022) Longitudinal analysis of T cell receptor repertoires reveals shared patterns of antigen-specific response to SARS-CoV-2 infection. JCI Insight 7(10):e151849. https://doi.org/10.1172/jci.insight.151849
22. Wu D, Emerson RO, Sherwood A et al (2014) Detection of minimal residual disease in B lymphoblastic leukemia by high-throughput sequencing of IGH. Clin Cancer Res Off J Am Assoc Cancer Res 20(17):4540–4548. https://doi.org/10.1158/1078-0432.CCR-13-3231
23. Zhang H, Liu L, Zhang J et al (2020) Investigation of antigen-specific T-cell receptor clusters in human cancers. Clin Cancer Res 26(6):1359–1371. https://doi.org/10.1158/1078-0432.CCR-19-3249
24. Jiang R, Fichtner ML, Hoehn KB et al (2020) Single-cell repertoire tracing identifies rituximab-resistant B cells during myasthenia gravis relapses. JCI Insight 5(14):e136471. https://doi.org/10.1172/jci.insight.136471
25. Herold KC, Bundy BN, Long SA et al (2019) An anti-CD3 antibody, teplizumab, in relatives at risk for type 1 diabetes. N Engl J Med 381(7):603–613. https://doi.org/10.1056/NEJMoa1902226
26. Herold KC, Gitelman SE, Gottlieb PA, Knecht LA, Raymond R, Ramos EL (2023) Teplizumab: a disease-modifying therapy for type 1 diabetes that preserves β-cell function. Diabetes Care 46(10):1848–1856. https://doi.org/10.2337/dc23-0675
27. Orban T, Bundy B, Becker DJ et al (2011) Co-stimulation modulation with abatacept in patients with recent-onset type 1 diabetes: a randomised, double-blind, placebo-controlled trial. Lancet 378(9789):412–419. https://doi.org/10.1016/S0140-6736(11)60886-6
28. Haller MJ, Long SA, Blanchfield JL et al (2019) Low-dose anti-thymocyte globulin preserves C-peptide, reduces HbA1c, and increases regulatory to conventional T-cell ratios in new-onset type 1 diabetes: two-year clinical trial data. Diabetes 68(6):1267–1276. https://doi.org/10.2337/db19-0057
29. Jacobsen LM, Diggins K, Blanchfield L et al (2023) Responders to low-dose ATG induce CD4+ T cell exhaustion in type 1 diabetes. JCI Insight 8(16):e161812. https://doi.org/10.1172/jci.insight.161812
30. Pescovitz MD, Greenbaum CJ, Krause-Steinrauf H et al (2009) Rituximab, B-lymphocyte depletion, and preservation of beta-cell function. N Engl J Med 361(22):2143–2152. https://doi.org/10.1056/NEJMoa0904452
31. Balmas E, Chen J, Hu AK et al (2023) Islet-autoreactive CD4+ T cells are linked with response to alefacept in type 1 diabetes. JCI Insight 8(21):e167881. https://doi.org/10.1172/jci.insight.167881
32. Edner NM, Heuts F, Thomas N et al (2020) Follicular helper T cell profiles predict response to costimulation blockade in type 1 diabetes. Nat Immunol 21(10):1244–1255. https://doi.org/10.1038/s41590-020-0744-z
33. Nakayama M, McDaniel K, Fitzgerald-Miller L et al (2015) Regulatory vs. inflammatory cytokine T-cell responses to mutated insulin peptides in healthy and type 1 diabetic subjects. Proc Natl Acad Sci U S A 112(14):4429–4434. https://doi.org/10.1073/pnas.1502967112
34. Hanna SJ, Tatovic D, Thayer TC, Dayan CM (2021) Insights from single cell RNA sequencing into the immunology of type 1 diabetes- cell phenotypes and antigen specificity. Front Immunol 12:751701. https://doi.org/10.3389/fimmu.2021.751701
35. Hanna SJ, Thayer TC, Robinson EJS et al (2023) Single-cell RNAseq identifies clonally expanded antigen-specific T-cells following intradermal injection of gold nanoparticles loaded with diabetes autoantigen in humans. Front Immunol 14:1276255. https://doi.org/10.3389/fimmu.2023.1276255
36. Sharma S, Tan X, Boyer J et al (2023) Measuring anti-islet autoimmunity in mouse and human by profiling peripheral blood antigen-specific CD4 T cells. Sci Transl Med 15(703):eade3614. https://doi.org/10.1126/scitranslmed.ade3614
37. Michels AW, Landry LG, McDaniel KA et al (2017) Islet-derived CD4 T cells targeting proinsulin in human autoimmune diabetes. Diabetes 66(3):722–734. https://doi.org/10.2337/db16-1025
38. Anderson AM, Landry LG, Alkanani AA et al (2021) Human islet T cells are highly reactive to preproinsulin in type 1 diabetes. Proc Natl Acad Sci 118(41):e2107208118. https://doi.org/10.1073/pnas.2107208118
39. Linsley PS, Barahmand-pour-Whitman F, Balmas E et al (2021) Autoreactive T cell receptors with shared germline-like α chains in type 1 diabetes. JCI Insight 6(22):e151349. https://doi.org/10.1172/jci.insight.151349
40. Mitchell AM, Baschal EE, McDaniel KA et al (2023) Tracking DNA-based antigen-specific T cell receptors during progression to type 1 diabetes. Sci Adv 9(49):eadj6975. https://doi.org/10.1126/sciadv.adj6975
41. Nakayama M, Michels AW (2019) Determining antigen specificity of human islet infiltrating T cells in type 1 diabetes. Front Immunol 10:365. https://doi.org/10.3389/fimmu.2019.00365
42. Arif S, Domingo-Vila C, Pollock E, Christakou E, Williams E, Tree TIM (2023) Monitoring islet specific immune responses in type 1 diabetes clinical immunotherapy trials. Front Immunol 14:1183909. https://doi.org/10.3389/fimmu.2023.1183909

43. Cerosaletti K, Barahmand-pour-Whitman F, Yang J et al (2017) Single-cell RNA sequencing reveals expanded clones of islet antigen-reactive CD4 + T cells in peripheral blood of subjects with type 1 diabetes. J Immunol 199(1):323–335. https://doi.org/10.4049/jimmunol.1700172
44. Arif S, Pujol-Autonell I, Kamra Y et al (2021) Mapping T cell responses to native and neo-islet antigen epitopes in at risk and type 1 diabetes subjects. Front Immunol 12:675746. https://doi.org/10.3389/fimmu.2021.675746
45. Eugster A, Lindner A, Catani M et al (2015) High diversity in the TCR repertoire of GAD65 autoantigen-specific human CD4 + T cells. J Immunol 194(6):2531–2538. https://doi.org/10.4049/jimmunol.1403031
46. Okamura T, Hamaguchi M, Tominaga H et al (2022) Characterization of peripheral blood TCR in patients with type 1 diabetes mellitus by BD rhapsodyTM VDJ CDR3 assay. Cells 11(10):1623. https://doi.org/10.3390/cells11101623
47. Gomez-Tourino I, Kamra Y, Baptista R, Lorenc A, Peakman M (2017) T cell receptor β-chains display abnormal shortening and repertoire sharing in type 1 diabetes. Nat Commun 8(1):1792. https://doi.org/10.1038/s41467-017-01925-2
48. Hudson D, Fernandes RA, Basham M, Ogg G, Koohy H (2023) Can we predict T cell specificity with digital biology and machine learning? Nat Rev Immunol 23(8):511–521. https://doi.org/10.1038/s41577-023-00835-3
49. Panzer JK, Hiller H, Cohrs CM et al (2020) Pancreas tissue slices from organ donors enable in situ analysis of type 1 diabetes pathogenesis. JCI Insight 5(8):e134525. https://doi.org/10.1172/jci.insight.134525
50. Dunlap G, Wagner A, Meednu N et al (2024) Clonal associations between lymphocyte subsets and functional states in rheumatoid arthritis synovium. Nat Commun 15(1):4491. https://doi.org/10.1038/s41467-024-49186-0
51. Lanz TV, Brewer RC, Ho PP et al (2022) Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. Nature 603(7900):321–327. https://doi.org/10.1038/s41586-022-04432-7
52. Catani M, Walther D, Christie MR, McLaughlin KA, Bonifacio E, Eugster A (2016) Isolation of human monoclonal autoantibodies derived from pancreatic lymph node and peripheral blood B cells of islet autoantibody-positive patients. Diabetologia 59(2):294–298. https://doi.org/10.1007/s00125-015-3792-4
53. Smith MJ, Packard TA, O'Neill SK et al (2015) Loss of anergic B cells in prediabetic and new-onset type 1 diabetic patients. Diabetes 64(5):1703–1712. https://doi.org/10.2337/db13-1798
54. Stensland ZC, Magera CA, Broncucia H et al (2023) Identification of an anergic BND cell–derived activated B cell population (BND2) in young-onset type 1 diabetes patients. J Exp Med 220(8):e20221604. https://doi.org/10.1084/jem.20221604
55. Seay HR, Yusko E, Rothweiler SJ et al (2016) Tissue distribution and clonal diversity of the T and B cell repertoire in type 1 diabetes. JCI Insight 1(20):e88242. https://doi.org/10.1172/jci.insight.88242
56. Scharf L, Axelsson H, Emmanouilidi A et al (2023) Longitudinal single-cell analysis of SARS-CoV-2–reactive B cells uncovers persistence of early-formed, antigen-specific clones. JCI Insight 8(1):e165299. https://doi.org/10.1172/jci.insight.165299
57. Setliff I, Shiakolas AR, Pilewski KA et al (2019) High-throughput mapping of B cell receptor sequences to antigen specificity. Cell 179(7):1636-1646.e15. https://doi.org/10.1016/j.cell.2019.11.003
58. Walker LM, Shiakolas AR, Venkat R et al (2022) High-throughput B cell epitope determination by next-generation sequencing. Front Immunol 13:855772. https://doi.org/10.3389/fimmu.2022.855772
59. Bagaev DV, Vroomans RMA, Samir J et al (2020) VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res 48(D1):D1057–D1062. https://doi.org/10.1093/nar/gkz874
60. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N (2017) McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics 33(18):2924–2929. https://doi.org/10.1093/bioinformatics/btx286
61. Vita R, Mahajan S, Overton JA et al (2019) The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res 47(D1):D339–D343. https://doi.org/10.1093/nar/gky1006
62. Mhanna V, Bashour H, Lê Quý K et al (2024) Adaptive immune receptor repertoire analysis. Nat Rev Methods Primer 4(1):6. https://doi.org/10.1038/s43586-023-00284-1
63. Wilkinson MD, Dumontier M, Aalbersberg IJ et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3(1):160018. https://doi.org/10.1038/sdata.2016.18
64. Rubelt F, Busse CE, Ahmad S et al (2017) Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data. Nat Immunol 18(12):1274–1278. https://doi.org/10.1038/ni.3873
65. Vander Heiden JA, Marquez S, Marthandan N et al (2018) AIRR community standardized representations for annotated immune repertoires. Front Immunol 9:2206. https://doi.org/10.3389/fimmu.2018.02206
66. Breden F, Luning Prak ET, Peters B et al (2017) Reproducibility and reuse of adaptive immune receptor repertoire data. Front Immunol 8:1418. https://doi.org/10.3389/fimmu.2017.01418
67. Parkinson H, Kapushesky M, Shojatalab M et al (2007) ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35(Database):D747–D750. https://doi.org/10.1093/nar/gkl995
68. Martin FJ, Amode MR, Aneja A et al (2023) Ensembl 2023. Nucleic Acids Res 51(D1):D933–D941. https://doi.org/10.1093/nar/gkac958
69. Barrett T, Troup DB, Wilhite SE et al (2011) NCBI GEO: archive for functional genomics data sets–10 years on. Nucleic Acids Res 39(1):D1005–D1010. https://doi.org/10.1093/nar/gkq1184
70. Liu S, Bradley P, Sun W (2023) Neural network models for sequence-based TCR and HLA association prediction. PLOS Comput Biol 19(11):e1011664. https://doi.org/10.1371/journal.pcbi.1011664
71. Schattgen SA, Guion K, Crawford JC et al (2022) Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). Nat Biotechnol 40(1):54–63. https://doi.org/10.1038/s41587-021-00989-2
72. Domínguez Conde C, Xu C, Jarvis LB et al (2022) Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science 376(6594):eabl5197. https://doi.org/10.1126/science.abl5197
73. Kaestner KH, Powers AC, Naji A, HPAP Consortium, Atkinson MA (2019) NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: the Human Pancreas Analysis Program (HPAP). Diabetes 68(7):1394–1402. https://doi.org/10.2337/db19-0058
74. Corrie BD, Christley S, Busse CE et al (2022) Data sharing and reuse: a method by the AIRR community. Methods Mol Biol 2453:447–476. https://doi.org/10.1007/978-1-0716-2115-8_23
75. Yuan D, Ahamed A, Burgin J et al (2024) The European Nucleotide Archive in 2023. Nucleic Acids Res 52(D1):D92–D97. https://doi.org/10.1093/nar/gkad1067
76. Sammut S-J, Galson JD, Minter R et al (2024) Predictability of B cell clonal persistence and immunosurveillance in breast cancer. Nat Immunol 25(5):916–924. https://doi.org/10.1038/s41590-024-01821-0
77. Culina S, Lalanne AI, Afonso G et al (2018) Islet-reactive CD8 + T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors. Sci Immunol 3(20):13. https://doi.org/10.1126/sciimmunol.aao4013
78. Fuchs YF, Eugster A, Dietz S et al (2017) CD8+ T cells specific for the islet autoantigen IGRP are restricted in their T cell receptor chain usage. Sci Rep 7(1):44661. https://doi.org/10.1038/srep44661
79. Nakayama M, Michels AW (2021) Using the T cell receptor as a biomarker in type 1 diabetes. Front Immunol 12:777788. https://doi.org/10.3389/fimmu.2021.777788

80. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH (2015) Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data: Table 1. Bioinformatics 31(20):3356–3358. https://doi.org/10.1093/bioinformatics/btv359
81. Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 33(5):495–502. https://doi.org/10.1038/nbt.3192
82. Reimand J, Arak T, Adler P et al (2016) g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res 44(W1):W83-9. https://doi.org/10.1093/nar/gkw199
83. Reimand J, Isserlin R, Voisin V et al (2019) Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc 14(2):482–517. https://doi.org/10.1038/s41596-018-0103-9
84. Scott JK, Breden F (2020) The adaptive immune receptor repertoire community as a model for FAIR stewardship of big immunology data. Curr Opin Syst Biol 24:71–77. https://doi.org/10.1016/j.coisb.2020.10.001
85. Nielsen SCA, Yang F, Jackson KJL et al (2020) Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2. Cell Host Microbe 28(4):516-525.e5. https://doi.org/10.1016/j.chom.2020.09.002
86. Goel RR, Painter MM, Apostolidis SA, et al (2021) mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. Science 374(6572):abm0829. https://doi.org/10.1126/science.abm0829
87. Heming M, Li X, Räuber S et al (2021) Neurological manifestations of COVID-19 feature T cell exhaustion and dedifferentiated monocytes in cerebrospinal fluid. Immunity 54(1):164-175.e6. https://doi.org/10.1016/j.immuni.2020.12.011
88. Pavlović M, Scheffer L, Motwani K et al (2021) The immuneML ecosystem for machine learning analysis of adaptive immune receptor repertoires. Nat Mach Intell 3(11):936–944. https://doi.org/10.1038/s42256-021-00413-z
89. Park JJ, Lee KAV, Lam SZ, Moon KS, Fang Z, Chen S (2023) Machine learning identifies T cell receptor repertoire signatures associated with COVID-19 severity. Commun Biol 6(1):76. https://doi.org/10.1038/s42003-023-04447-4
90. Nazarov V, Tsvetkov V, Fiadziushchanka S et al (2023) Immunarch: Bioinformatics analysis of T-cell and B-cell immune repertoires. Available from https://immunarch.com/ and https://github.com/immunomind/immunarch, accessed 3 Oct 2024
91. Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T (2019) OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. Bioinformatics 35(17):2974–2981. https://doi.org/10.1093/bioinformatics/btz035
92. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM (2020) Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. Nat Biotechnol 38(10):1194–1202. https://doi.org/10.1038/s41587-020-0505-4
93. Wang M, Patsenker J, Li H, Kluger Y, Kleinstein SH (2024) Language model-based B cell receptor sequence embeddings can effectively encode receptor specificity. Nucleic Acids Res 52(2):548–557. https://doi.org/10.1093/nar/gkad1128
94. Zaslavsky ME, Craig E, Michuda JK et al (2024) Disease diagnostics using machine learning of immune receptors. BioRχiv. https://doi.org/10.1101/2022.04.26.489314
95. Waltari E, Nafees S, McCutcheon KM, Wong J, Pak JE (2022) AIRRscape: An interactive tool for exploring B-cell receptor repertoires and antibody responses. PLOS Comput Biol 18(9):e1010052. https://doi.org/10.1371/journal.pcbi.1010052

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Stephanie J. Hanna[1] · Rachel H. Bonami[2,3,4,5] · Brian Corrie[6,7] · Monica Westley[8] · Amanda L. Posgai[9] · Eline T. Luning Prak[10] · Felix Breden[6,7] · Aaron W. Michels[11] · Todd M. Brusko[9,12,13] · Type 1 Diabetes AIRR Consortium

✉ Stephanie J. Hanna
HannaSJ@cardiff.ac.uk

✉ Aaron W. Michels
aaron.michels@cuanschutz.edu

✉ Todd M. Brusko
tbrusko@ufl.edu

1 Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK

2 Department of Medicine, Division of Rheumatology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA

3 Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA

4 Vanderbilt Center for Immunobiology, Nashville, TN, USA

5 Vanderbilt Institute for Infection, Immunology, and Inflammation, Nashville, TN, USA

6 Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada

7 iReceptor Genomic Services, Summerland, BC, Canada

8 the(sugar)science, Redondo Beach, CA, USA

9 Department of Pathology, Immunology, and Laboratory Medicine, College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL, USA

10 Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

11 Barbara Davis Center for Diabetes, University of Colorado School of Medicine, Aurora, CO, USA

12 Department of Pediatrics, College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL, USA

13 Department of Biochemistry and Molecular Biology, College of Medicine, Diabetes Institute, University of Florida, Gainesville, FL, USA