Peripheral tolerance to insulin is encoded by mimicry in the microbiome

Arcadio Rubio García,* Athina Paterou, Mercede Lee, Hubert Sławiński,
Linda S. Wicker, John A. Todd* and Marcin Ł. Pękalski*

*{arc,jatodd,marcin}@well.ox.ac.uk
Wellcome Centre for Human Genetics
University of Oxford
Roosevelt Drive, Oxford OX3 7BN

Abstract

How organisms achieve sustained peripheral tolerance throughout their lifetime, a correct immune discrimination between self and non-self, remains poorly understood. Host-microbiome interactions carry fundamental information that facilitates this process. We hypothesize that commensal microbes are under evolutionary pressure to develop epitopes that, when presented along with other antigens from their own bacterial community, lead to an overall tolerogenic self classification by the host immune system. Hosts, which have co-evolved with commensals, may rely on mimotopes, bacterial epitopes that are indistinguishable from key self epitopes, as a homeostatic feedback mechanism to establish and maintain tolerance. Using a probabilistic sequence model of peptide mimicry, we show that the gut microbiome contains a set of genes that are likely to trigger identical immune responses to insulin B 9–25, a widely distributed self epitope across tissues and the primary autoantigen in type 1 diabetes. Similarities in the antigen receptor sequences determined from CD4 T cells reacting to insulin epitopes and mimotopes provide experimental evidence for mimicry. All predicted high posterior probability mimotopes belong to the transketolase superfamily, an enzyme that allows efficient harvest of commensal-derived sugar polymers and dietary fibre, an advantage during host colonisation. Microbial transketolase upregulation during infant weaning coincides in time with the peak in autoantibody development against insulin. Abundance changes in bacterial genera that carry these mimotopes have also been observed to precede disease diagnosis. Our findings suggest gut dysbiosis followed by immune response to insulin mimotopes as a primary cause of type 1 diabetes, and may contribute towards unraveling similar causal patterns in a wide variety of disorders.

1 Introduction

The human microbiome plays a fundamental role in health and disease. Early twin studies showed deviations from a core set of gut microbial genes are associated with different physiological states [1]. This result has been extended to a variety of tissues [2] and disorders [3] such as inflammatory bowel disease [4], multiple sclerosis [5], type 1 diabetes [6], type 2 diabetes [7], Alzheimer's disease [8], and Parkinson's disease [9].

Westernized societies exhibit dramatic microbiome differences with respect to isolated hunter-gatherer populations, both in terms of diversity and composition, due to the myriad of modern practices that decrease commensal transmission and survival [10]. These include scarce microbiota-accessible carbohydrates related to low-fibre diets [11] or early-life antibiotic treatments [12], providing some cues of a possible causal chain linking environmental changes and increased prevalence of complex disorders via alterations to the microbiome. Germ-free model organisms and fecal transplants provide further mechanistic evidence by inducing [13] or reversing [14], with some limitations, a variety of disease phenotypes.

Dysregulated metabolic pathways do not seem sufficient to explain the numerous associations between microbiome changes and disease, in particular within autoimmune inflammatory disorders. This has prompted revisiting molecular mimicry [15] in the context of symbiosis and the subsequent discovery of several commensal peptides that may mimic key autoantigens in Sjögren's disease [16], systemic lupus erythematosus [17], multiple sclerosis [18] and antiphospholipid syndrome [19]. The role of commensal mimicry in human physiology remains unclear as most candidates are limited to a reduced number of microbial genes, some mimotopes originate from bacterial orthologs of host proteins, or belong to species with possible pathogenic potential [20, 21]. Moreover, the lack of a precise definition of molecular mimicry at sequence level hinders further progress, with sequence homology being a poor proxy [22].

Since commensal proteomes do not seem to influence central thymic selection through any primary pathways, mimicry could be a ubiquitous evolved strategy of the microbiome to avoid potentially harmful immune responses against microbes and their metabolites that are essential for the health of the host. Hosts may have co-evolved a mechanism to maintain immune homeostasis by exploiting continuous presentation of mimotopes obtained in the microbiota-mucosal interface to, for example, induce peripheral regulatory T cells or carry out a controlled apoptosis of T effector cells.

We postulate that molecular mimicry occurs when all antigen receptors that can be constructed for a given epitope, or a set of epitopes digested and presented from a particular antigen, bind either the original epitope or the mimotope with equivalent affinity. This may happen if and only if all contact points in the receptor-epitope or receptor-mimotope structure contain identical amino acids or some closely related residue. The remaining positions are more prone to tolerate substitutions with similar or sometimes distant amino acids. In general, insertions or deletions may break mimicry. Using this probabilistic definition of sequence mimicry we uncover a large set of bacteria that are likely to trigger identical immune responses to insulin B 9–25, a widely expressed self epitope and the primary autoantigen in type 1 diabetes (T1D), by adapting a region of the transketolase protein superfamily—an enzyme related to the metabolism of fibre upregulated during infant gut development [23].

2

2   Results

2.1   The human gut biome harbors potential insulin B mimotopes

In order to assess the existence of commensal mimotopes, we have defined a generative process that models the mixture of sequences that exhibit some homology, or significant similarity [24], plus those with mimicry potential to a self epitope of choice. The core idea is to sample two different motifs for each class of sequences, homologs and mimotopes, from position-specific prior distributions. Then, for each sequence, to sample its class and each residue from the previously constructed motifs, which now act as distributions, conditioned on the class.

Algorithm 1 Hierarchical mixture mimicry model.

1: $\pi_c \sim LogitNormal(\mu_c, \sigma_c^2)$
2: $\pi_i \sim LogitNormal(\mu_i, \sigma_i^2)$
3: **for** $j \in \{1, \dots, J\}$ **do**
4:      $p_j \sim Bernoulli(\pi_i)$
5:      **if** p_j **then**
6:          $M_{cj} \sim Dirichlet(\alpha_{cj})$
7:      **else**
8:          $M_{cj} \sim Dirichlet(\alpha_{hj})$
9:      **end if**
10:     $M_{hj} \sim Dirichlet(\alpha_{hj})$
11: **end for**
12: **for** $k \in \{1, \dots, K\}$ **do**
13:     $c_k \sim Bernoulli(\pi_c)$
14:     **for** $j \in \{1, \dots, J\}$ **do**
15:         **if** c_k **then**
16:             $S_{kj} \sim Categorical(M_{cj})$
17:         **else**
18:             $S_{kj} \sim Categorical(M_{hj})$
19:         **end if**
20:     **end for**
21: **end for**

We performed inference using the generative model described by Algorithm 1 coupled with Metropolis-Hastings sampling on probabilistic program traces [25] using two gut microbiome databases, IGC [26] and our own assembly of the T1D DIABIMMUNE cohort [6]. Metagenes, microbial genes from a metagenome, were filtered to contain a valid open reading frame present in at least 10% of donors. The insulin chain INS B 9–25, a region the primary autoantigen in T1D has been mapped to [27], was used as a target epitope.

Tables 1 and 2 list the top 10 hits for each database assembly, sorted by posterior probability. The homology column uses the notation introduced by Clustal Ω [28], with a colon punctuation mark for highly similar but not identical amino acids, a dot for close substitutions and a space for unrelated amino acids. The BLOSUM45 matrix was used to define a distance metric among pairs of amino acids [29].

Metagene	Sequence	Homology	p
3043114	GHSVEALYAVLAEKGFF	.H VEALY V .E:GFF	0.93
3160029	GHSVEALYAVLAKKGFF	.H VEALY V .::GFF	0.90
3247770	GHTVEALYAVLCQKGFF	.H VEALY V :::GFF	0.89
3169624	GHTVPALYAVLAEKGYF	.H V ALY V .E:G:F	0.77
3161322	GHSVEAYYAVLAEKGFF	.H VEA Y V .E:GFF	0.75
2969430	GHSVEAYYSVLAEKGFF	.H VEA Y V .E:GFF	0.74
3278638	GHSVEALYCILADRGFF	.H VEALY : .:RGFF	0.73
3186877	GHSVETLYCILADRGYF	.H VE:LY : .:RG:F	0.72
3298615	GHIAEALYVTLAKRGFF	.H:.EALY:. .:RGFF	0.52
3205639	GHIAEALYVVLAKKGYF	.H:.EALY:V .:.G:F	0.51
INS B 9–25	SHLVEALYLVCGERGFF	.H . : Y . ::G:F	

Table 1: Top predicted IGC bacterial mimotopes of INS B chain, sorted by posterior probability p.

The same pattern stands out in both databases, with 3 or perhaps 4 blocks of identity and homology, plus a few non-conserved regions. This motif degrades with decreasing posterior probabilities, which might imply imperfect mimotopes are biologically relevant, contributing with additive effects towards tolerance. It may also reflect an evolutionary process towards mimicry.

Tables 3 and 4 list mappings between IGC and DIABIMMUNE, but also to the UniProt protein database [30]. These have been constructed by finding metagene matches with at least 97% overall identity, no gaps and no substitutions within the mimicry region. The species column has been taken from the corresponding UniProt entry. Many sequences are common between both top predicted hits. Others have matching proteins in the alternative assembly which did not enter the corresponding top results due to using a 10% abundance cutoff. This might reflect taxa changes between infant and adult guts, geographical variation, or differences induced by sample collection and sequencing parameters.

Nonetheless, reproducibility is particularly noteworthy. All species listed belong to the Firmicutes phylum, except *Collinsella aerofaciens* which is classified as an Actinobacteria. Other Actinobacteria are present in entries that are located close to the highest posterior probabilities. By inspecting assembled proteins or the corresponding UniProt entry, listed hits appear to be highly variable in terms of sequence length and amino acid composition except within the mimotope region.

2.2 Insulin B mimotopes belong to the transketolase superfamily

In case of the IGC reference gut microbiome, 49 out of 50 hits with highest posterior probabilities are annotated as belonging to the Kyoto Encyclopedia of Genes and Genomes (KEGG) ortholog class K00615 [31]. This corresponds to the transketolase (TKT) enzyme domain, part of the pentose phosphate pathway. Given that K00615 is rare in IGC, with just 6176 entries out of a total 9879896 (0.06%), this result appears to be significant. A formal enrichment test confirms this fact with a p-value estimate $< 10^{-6}$. Reannotating DIABIMMUNE yields equivalent results, given that 48 out of 50 entries containin domain K00615.

Metagene	Sequence	Homology	p
267926_71	GHSVEALYAVLAEKGFF	.H VEALY V .E:GFF	0.96
750513_07	GHSVEALYAVLCDKGFF	.H VEALY V ::GFF	0.88
778380_12	GHTVEALYAVLCQKGFF	.H VEALY V ::GFF	0.80
045379_02	GHTVPALYAVLAEKGYF	.H V ALY V .E:G:F	0.74
611716_02	GHSVEAYYSVLAEKGFF	.H VEA Y V .E:GFF	0.72
846083_03	GHSVEALYCILADRGFF	.H VEALY : .:RGFF	0.70
974697_44	GHTVPALYATLAERGFF	.H V ALY . .ERGFF	0.65
037280_07	GHIAEALYVTLAKRGFF	.H:.EALY:. .:RGFF	0.56
947478_02	GHIAEALYVVLAKKGYF	.H:.EALY:V .::G:F	0.55
822586_01	GHCAPALYAVLAERGFF	.H . ALY V .ERGFF	0.54
INS B 9–25	SHLVEALYLVCGERGFF	.H . A Y . ::G:F	

Table 2: Top predicted DIABIMMUNE bacterial mimotopes of INS B chain, sorted by posterior probability p.

IGC	DIABIM	UniProt	Species
3043114	267926_71	R5X0Y1	*Blautia sp.*
3160029	—	A0A373VWY7	*Blautia sp.*
3247770	778380_12	D4JQT7	*Eubacterium siraeum*
3169624	—	A0A413T8V8	*Eubacterium ventriosum*
3161322	—	B0NC25	*Clostridium scindens*
2969430	611716_02	A0A3A6F383	*Lachnospiraceae bacterium*
3278638	846083_03	—	—
3186877	—	—	—
3298615	037280_07	H1ALH4	*Coprobacillus sp.*
3205639	947478_02	E7GFP6	*Coprobacillus cateniformis*

Table 3: Mapping of top IGC bacterial mimotopes to DIABIMMUNE and UniProt.

DIABIM	IGC	UniProt	Species
267926_71	3043114	R5X0Y1	*Blautia sp.*
750513_07	—	—	—
778380_12	3247770	D4JQT7	*Eubacterium siraeum*
045379_02	—	A0A413RAY3	*Eubacterium ventriosum*
611716_02	2969430	A0A3A6F383	*Lachnospiraceae bacterium*
846083_03	3278638	—	—
974697_44	3268248	A0A1M6LKK1	*Megasphaera elsdenii*
037280_07	3298615	H1ALH4	*Coprobacillus sp.*
947478_02	3205639	E7GFP6	*Coprobacillus cateniformis*
822586_01	2887265	A0A174KQ89	*Collinsella aerofaciens*

Table 4: Mapping of top DIABIMMUNE bacterial mimotopes to IGC and UniProt.

Pfam is a database that contains profile hidden Markov models (pHMMs) [32] derived from curated multiple alignments [33]. Within the Pfam pHMM for TKT, accession no. PF00456, all mimotope regions map to residues 63–79. Since insertions or deletions are very unlikely in that interval, we can treat the model as a simpler position-specific weight matrix (PWM). The latter describes the probability of each amino acid per position independently of other residues [34]. The lack of insertions, deletions and hidden states allows for a much easier mathematical treatment and visualization.

A sequence logo plot is derived by making the PWM plot y axis length at each residue equal to its information content [35]. Positions where all amino acids are equally likely will have a maximal entropy and carry no information at all. The logo plot in Figure 1 shows TKT residues in the region of interest are relatively well conserved across all superfamily members, whereas the others are highly random indicating they might not be required for the correct functioning of the enzyme. Amino acids occupying highly conserved residues and their spacing seem very well suited for evolution to develop mimotopes employing TKT as a template. This is also facilitated by the lack of conservation in the remaining residues, which can be mutated without disrupting the function of the enzyme. Given that insulin mimotopes encoded in TKT are present across two bacterial phylums, Firmicutes and Actinobacteria, horizontal or lateral gene transfer events are probable. Phylogenetic tree reconstruction supports the existence of these events [36], as well as some observations derived from insect models [37].

2.3   Insulin B and transketolase mimotopes are crossreactive

To obtain further experimental evidence that supports these findings, we activated CD4 T cells using monocyte-derived dendritic cells as antigen presenting cells isolated from whole blood belonging to a diagnosed T1D donor. Synthesized peptides matching some of the top predicted IGC and DIABIMMUNE mimotopes, outlined in Table 5, were used. Four CD4 T cell pools were created. Pool U was not stimulated. Pool C was stimulated with a set of control peptides, pool I with all synthesized insulin peptides, and pool T with all mimotopes listed in Table 5.

The intersection of clonotypes present in pools I and T with at least frequency 10, minus any clonotype present in any other pool at any frequency returned the 8 clonotypes listed in Table 7. All other equivalent pairwise intersections across pools returned an empty set or just one public clonotype. Not only a few high frequency clonotypes are shared and exclusive to the insulin and TKT pools, but some do have extremely low predicted recombination probabilities for β chains [38] and exhibit high degrees of sequence similarity suggesting equivalent binding properties.
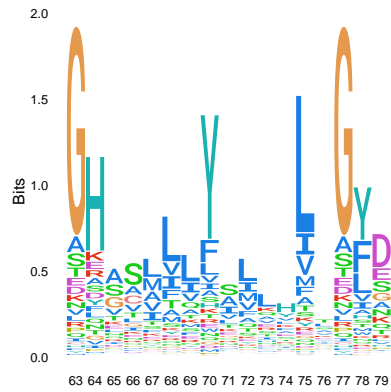
(a) Transketolase N superfamily
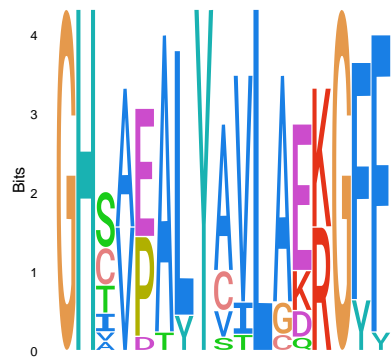
(b) Prioritized IGC mimotopes
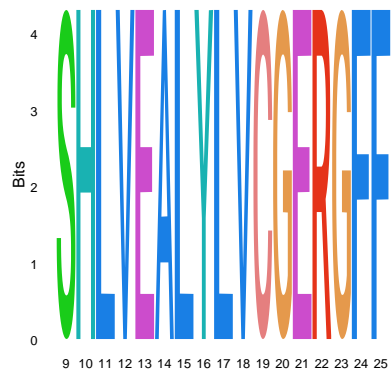
(c) Insulin B 9–25

Figure 1: Sequence logo for transketolase N superfamily, prioritized IGC mimotopes and insulin B 9–25 chain.

7

Source	Begin	End	Sequence
3043114	76	91	HSVEALYAVLAEKGFF
3043114	78	91	VEALYAVLAEKGFF
3160029	68	83	HSVEALYAVLAKKGFF
3278638	66	81	HSVEALYCILADRGFF
3278638	68	81	VEALYCILADRGFF
3269707	66	81	HCADALYCVLGERGFY
INS B	10	25	HLVEALYLVCGERGFF
INS B	12	25	VEALYLVCGERGFF

Table 5: Mimotopes and insulin B chain peptides used to assess commensal crossreactivity.

Pool	Clonotypes	Cells	Peptides
U	910	4716	—
C	1180	5707	Control
I	1659	5618	Insulin
T	1902	7147	Transketolase

Table 6: Monocyte-derived DC assay restimulation pools.

CDR3α	CDR3β	f_I	f_T	p_α	p_β
CAVTGANSKLTF	CSAIRGTSGAYEQYV	31	19	1e-5	1e-20
CAVTGANSKLTF	CSAVHGTSGTYEQYV	43	14	1e-5	6e-20
CAASGDSGGGADGLTF	CASSLGLRPLSYEQYV	26	10	3e-6	2e-19
CAFMKWGGGSNYKLTF	CASSPHIQGPQETQYF	42	16	1e-8	4e-11
CAVNRPGGSGKLIF	CASSSGVTTDTQYF	10	28	3e-12	2e-8
CAVNRGGGNKLTF	CSDSQTGGETQYF	27	56	2e-7	2e-11
CAVVRYGGSQGNLIF	CASS...SGANVLTF	18	15	3e-7	4e-11
CAALSTSGTYKYIF	CASSPDGNSGNTIYF	74	72	5e-7	4e-9

Table 7: High frequency clonotypes stimulated by insulin B and mimotopes.

8

3 Discussion

The discovery that some bacterial TKT enzymes contain functional mimotopes of the key autoantigenic epitope in T1D, INS B 9–25, supports the hypothesis that self is partly encoded in the microbiome. In some scenarios, these peptides are probably mimicking self antigens recognized by a large number of regulatory T cells, as that is an optimal solution that confers bacterial communities maximal fitness. However, in most cases, mimotopes probably have less pronounced effects, as bacteria have not been able to adapt their sequences enough or target key self peptides. Hence, only the combined presentation of many weak mimotopes possibly recruiting few regulatory T cells each, to antigen presenting cells, would be enough for an overall tolerogenic self classification. This could also explain the difficulties in identifying mimotopes, as most sequences would exhibit weak mimicry motifs. Non-peptide mimotopes [39] may also have an underappreciated role in conveying signals currently denoted as immunological context [40].

Dysbiosis, defined as a disruption of commensal communities, alters the proportion of peptides and context signals presented. The mimotopes evolved by selection pressure are no longer sufficient to elicit a tolerogenic self response. This has important potential consequences for the host organism. Repeated presentations of the mimotope in a dysbiotic context could lead to an autoimmune response against the bacterial antigen. Which implies the self epitope it mimics also becomes downgraded, as they are undistinguishable, and no longer has a positive contribution towards tolerance when presented along other self peptides. Changes of epitope status can be thought of, in practical terms, as for example alterations of T conventional versus regulatory cell proportions recognizing them.

Aside from molecular mimicry, there are several independent sources of evidence pointing to the role of TKT in T1D onset. The TEDDY study has shown TKT to be one of the most consistently upregulated microbial enzymes during the first year of life [23]. The TKT enzyme is involved in the non-oxidative branch of the pentose phosphate pathway and has been linked to the fermentation of fibre [41]. This becomes an important metabolic pathway after the transition to solid food, known as weaning, that typically occurs between 4 to 12 months of age. Microbial TKT has been also implicated in the metabolism of exopolysaccharides, high-molecular-weight polymers secreted by microbes into the microbial ecosystem, promoting a shift in the microbial metabolism to adapt to new dietary conditions and render more reducing power [42]. TKT expression confers microbes an important adaptation in sourcing energy, including saccharides produced by other microbes such as bifidobacteria, and thus allows rapid host colonisation [43].

Notwithstanding the benefits of fibre-rich diets are widely accepted, early introduction of fibre especially at time of dysbiosis may also propagate pre-existing TKT-expressing microbes and lead to T1D. Early introduction of root vegetables during infancy has been associated with increased risk of insulin autoimmunity in a cohort of children with high genetic susceptibility to T1D [44].

A recent study has demonstrated the importance of the intestinal microbiome inducing a vigorous immune response known as a weaning reaction [45]. While the weaning reaction is programmed in time to allow healthy immune development, at state of microbial dysbiosis the weaning reaction can lead to responses against mimotope-expressing commensal bacteria.

The TEDDY study has shown autoantibodies are rare before 6 months of life [46]. However, autoantibodies against insulin typically develop first shortly afterwards, with a highest incidence around 12 months of age, and a median to seroconversion of 18 months. Hence, both the time expression pattern of TKT and autoantibody development support the hypothesis of dysbiosis leading to immunity against the microbiome, and eventually against insulin due to mimicry. This process may be initiated before transition to solid food, and occur at a time when undeveloped epithelial barriers can cause microbial TKT leaks into the host [47].

In particular, dysbiosis might start before weaning if infants are fed artificial milk formulas lacking key human milk oligosaccharides that naturally support healthy metabolism, growth of bifidobacteria and gut homeostasis. TKT expression may be an advantage for host colonization as it allows specific microbes to efficiently source energy from polysaccharides present in the diet or from polysaccharides produced by other microbes. Therefore TKT may link microbial metabolism of polysaccharides, dysbiosis and antigen-specific immune responses to insulin.

This is also supported by the lower metagene abundance in T1D children after seroconversion and before diagnosis observed by DIABIMMUNE. Some of these changes seem to come from *Blautia*, *Rikenellaceae*, and the *Ruminococcus* and *Streptococcus* genera [6]. *Blautia* is the species containing the top insulin mimotope in both reference gut microbiome datasets, as shown in Tables 3 and 4. Gut dysbiosis has been associated with dysregulated abundance of essential short chain fatty acids, products of polysaccharide processing by microbes, and increased gut permeability or leaky gut syndrome [48]. Overall, our findings suggest that dysbiosis, linked to colonisation by TKT-expressing microbes, may destabilize epithelial barriers of the gut causing leak of mimotopes to the host followed by an autoimmune response to insulin.

Vitamin B1 (thiamine) deficiency is a frequent complication in T1D that leads to widespread tissue damage [49]. Thiamine is a well known co-factor of TKT [50], and its deficiency and low TKT activity are highly correlated. Notably, changes in gut microbiome taxa abundances and transcribed pathways of T1D patients have been shown to precede thiamine deficiency [51].

Probiotics and synbiotic supplementation improve peanut allergies [52] or reduce mortality in children with sepsis [53] by promoting gut microbial homeostasis. Our results impact future design and implementation of such strategies. It is possible that synbiotic supplementation would enhance tolerance induction to insulin in newborns previously stratified for genetic T1D susceptibility, where transketolase mimotopes would have to be taken into account. Finally, after decades of research, ever since particular HLA class II molecules were implicated as the strongest genetic risk factor [54], we present a possible missing major environmental link—dysbiosis-sensitive microbial mimotopes of insulin B 9–25 as a primary and specific cause of type 1 diabetes.

Acknowledgements

References

[1] Peter J. Turnbaugh et al. "A core gut microbiome in obese and lean twins". In: *Nature* 457.7228 (2009), p. 480.

[2] Curtis Huttenhower et al. "Structure, function and diversity of the healthy human microbiome". In: *Nature* 486.7402 (2012), p. 207.

[3] Matthew A. Jackson et al. "Gut microbiota associations with common diseases and prescription medications in a population-based cohort". In: *Nature Communications* 9 (2018).

[4] Jonas Halfvarson et al. "Dynamics of the human gut microbiome in inflammatory bowel disease". In: *Nature Microbiology* 2.5 (2017), p. 17004.

[5] Kerstin Berer et al. "Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice". In: *Proceedings of the National Academy of Sciences* 114.40 (2017), pp. 10719–10724.

[6] Aleksandar D. Kostic et al. "The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes". In: *Cell Host & Microbe* 17.2 (2015), pp. 260–273.

[7] Helle Krogh Pedersen et al. "Human gut microbes impact host serum metabolome and insulin sensitivity". In: *Nature* 535.7612 (2016), p. 376.

[8] Stephen S. Dominy et al. "Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors". In: *Science Advances* 5.1 (2019), eaau3333.

[9] Anna Heintz-Buschart et al. "The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder". In: *Movement Disorders* 33.1 (2018), pp. 88–98.

[10] José C. Clemente et al. "The microbiome of uncontacted Amerindians". In: *Science Advances* 1.3 (2015), e1500183.

11

[11] Erica D. Sonnenburg et al. "Diet-induced extinctions in the gut microbiota compound over generations". In: *Nature* 529.7585 (2016), p. 212.

[12] Moran Yassour et al. "Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability". In: *Science Translational Medicine* 8.343 (2016), ra81.

[13] Kerstin Berer et al. "Commensal microbiota and myelin autoantigen co-operate to trigger autoimmune demyelination". In: *Nature* 479.7374 (2011), p. 538.

[14] Christina Jorup-Rönström et al. "Fecal transplant against relapsing Clostridium difficile-associated diarrhea in 32 patients". In: *Scandinavian Journal of Gastroenterology* 47.5 (2012), pp. 548–552.

[15] Michael B. Oldstone and Abner L. Notkins. "Molecular mimicry". In: *Concepts in viral pathogenesis II*. New York: Springer, 1986, pp. 195–202.

[16] Agnieszka Szymula et al. "T cell epitope mimicry between Sjögren's syndrome Antigen A (SSA)/Ro60 and oral, gut, skin and vaginal bacteria". In: *Clinical Immunology* 152.1-2 (2014), pp. 1–9.

[17] Teri M. Greiling et al. "Commensal orthologs of the human autoantigen Ro60 as triggers of autoimmunity in lupus". In: *Science Translational Medicine* 10.434 (2018), eaan2306.

[18] Raquel Planas et al. "GDP-l-fucose synthase is a CD4+ T cell–specific autoantigen in DRB3*02:02 patients with multiple sclerosis". In: *Science Translational Medicine* 10.462 (2018), eaat4301.

[19] William E. Ruff et al. "Pathogenic autoreactive T and B cells cross-react with mimotopes expressed by a common human gut commensal to trigger autoimmunity". In: *Cell Host & Microbe* (2019).

[20] Gregg J. Silverman, Doua F. Azzouz, and Alexander V. Alekseyenko. "Systemic Lupus Erythematosus and dysbiosis in the microbiome: cause or effect or both?" In: *Current Opinion in Immunology* 61 (2019), pp. 80–85.

[21] William E. Ruff, Silvio M. Vieira, and Martin A. Kriegel. "The role of the gut microbiota in the pathogenesis of antiphospholipid syndrome". In: *Current Rheumatology Reports* 17.1 (2015), p. 472.

[22] Junbao Yang et al. "Autoreactive T cells specific for insulin B: 11-23 recognize a low-affinity peptide register in human subjects with autoimmune diabetes". In: *Proceedings of the National Academy of Sciences* 111.41 (2014), pp. 14840–14845.

[23] Tommi Vatanen et al. "The human gut microbiome in early-onset type 1 diabetes from the TEDDY study". In: *Nature* 562.7728 (2018), p. 589.

[24] Richard Durbin et al. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press, 1998.

[25] Noah Goodman et al. "Church: a language for generative models". In: *arXiv* 1206.3255 (2012).

[26] Junhua Li et al. "An integrated catalog of reference genes in the human gut microbiome". In: *Nature Biotechnology* 32.8 (2014), p. 834.

[27]   Nanette C. Schloot et al. "Cloned T cells from a recent onset IDDM patient reactive with insulin B-chain". In: *Journal of Autoimmunity* 11.2 (1998), pp. 169–175.

[28]   Julie D. Thompson, Desmond G. Higgins, and Toby J. Gibson. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice". In: *Nucleic Acids Research* 22.22 (1994), pp. 4673–4680.

[29]   Steven Henikoff and Jorja G. Henikoff. "Amino acid substitution matrices from protein blocks". In: *Proceedings of the National Academy of Sciences* 89.22 (1992), pp. 10915–10919.

[30]   Rolf Apweiler et al. "UniProt: the universal protein knowledgebase". In: *Nucleic Acids Research* 32.S1 (2004), pp. D115–D119.

[31]   Minoru Kanehisa and Susumu Goto. "KEGG: Kyoto Encyclopedia of Genes and Genomes". In: *Nucleic Acids Research* 28.1 (2000), pp. 27–30.

[32]   Anders Krogh et al. "Hidden Markov models in computational biology: Applications to protein modeling". In: *Journal of Molecular Biology* 235.5 (1994), pp. 1501–1531.

[33]   Erik L. L. Sonnhammer, Sean R. Eddy, and Richard Durbin. "Pfam: a comprehensive database of protein domain families based on seed alignments". In: *Proteins: Structure, Function, and Bioinformatics* 28.3 (1997), pp. 405–420.

[34]   Gary D. Stormo et al. "Use of the Perceptron algorithm to distinguish translational initiation sites in E. coli". In: *Nucleic Acids Research* 10.9 (1982), pp. 2997–3011.

[35]   Thomas D. Schneider and R. Michael Stephens. "Sequence logos: a new way to display consensus sequences". In: *Nucleic Acids Research* 18.20 (1990), pp. 6097–6100.

[36]   Matthew B. Rogers et al. "A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer". In: *BMC Evolutionary Biology* 7.1 (2007), p. 89.

[37]   Bryan P. Brown and Jennifer J. Wernegreen. "Genomic erosion and extensive horizontal gene transfer in gut-associated Acetobacteraceae". In: *BMC Genomics* 20.1 (2019), p. 472.

[38]   Zachary Sethna et al. "OLGA: fast computation of generation probabilities of B-and T-cell receptor amino acid sequences and motifs". In: *Bioinformatics* 35.17 (2019), pp. 2974–2981.

[39]   Louis J. Cohen et al. "Commensal bacteria make GPCR ligands that mimic human signalling molecules". In: *Nature* 549.7670 (2017), p. 48.

[40]   Polly Matzinger. "Tolerance, danger, and the extended family". In: *Annual Review of Immunology* 12.1 (1994), pp. 991–1045.

[41]   B. Thurston, K. A. Dawson, and H. J. Strobel. "Pentose utilization by the ruminal bacterium Ruminococcus albus". In: *Applied and Environmental Microbiology* 60.4 (1994), pp. 1087–1092.

[42] David Ríos Covián et al. "Different metabolic features of Bacteroides fragilis growing in the presence of glucose and exopolysaccharides of bifidobacteria". In: *Frontiers in Microbiology* 6 (2015), p. 825.

[43] Xin Tan et al. "Transketolase of Staphylococcus aureus in the control of master regulators of stress response during infection". In: *The Journal of Infectious Diseases* 220.12 (2019), pp. 1967–1976.

[44] Suvi M. Virtanen et al. "Early introduction of root vegetables in infancy associated with advanced ß-cell autoimmunity in young children with human leukocyte antigen-conferred susceptibility to Type 1 diabetes". In: *Diabetic Medicine* 28.8 (2011), pp. 965–971.

[45] Ziad Al Nabhani et al. "A weaning reaction to microbiota is required for resistance to immunopathologies in the adult". In: *Immunity* 50.5 (2019), pp. 1276–1288.

[46] Jeffrey P. Krischer et al. "The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study". In: *Diabetologia* 58.5 (2015), pp. 980–987.

[47] Emilie Tourneur and Cecilia Chassin. "Neonatal immune adaptation of the gut and its role during infections". In: *Clinical and Developmental Immunology* 2013 (2013).

[48] James L. Richards et al. "Dietary metabolites and the gut microbiota: an alternative approach to control inflammatory and autoimmune diseases". In: *Clinical & Translational Immunology* 5.5 (2016), e82.

[49] Elizabeth A. Rosner et al. "Low thiamine levels in children with type 1 diabetes and diabetic ketoacidosis: a pilot study". In: *Pediatric Critical Care Medicine* 16.2 (2015), pp. 114–118.

[50] Efraim Racker, G. de la Haba, and I. G. Leder. "Thiamine pyrophosphate, a coenzyme of transketolase". In: *Journal of the American Chemical Society* 75.4 (1953), pp. 1010–1011.

[51] Anna Heintz-Buschart et al. "Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes". In: *Nature Microbiology* 2.1 (2017), p. 16180.

[52] Kuang-Chih Hsiao et al. "Long-term clinical and immunological effects of probiotic and peanut oral immunotherapy after treatment cessation: 4-year follow-up of a randomised, double-blind, placebo-controlled trial". In: *The Lancet Child & Adolescent Health* 1.2 (2017), pp. 97–105.

[53] Pinaki Panigrahi et al. "A randomized synbiotic trial to prevent sepsis among infants in rural India". In: *Nature* 548.7668 (2017), p. 407.

[54] John A. Todd, John I. Bell, and Hugh O. McDevitt. "HLA-DQβ gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus". In: *Nature* 329.6140 (1987), p. 599.