nature genetics

Activating germline mutations in *STAT3* cause early-onset multi-organ autoimmune disease

Sarah E Flanagan[1,16], Emma Haapaniemi[2,3,16], Mark A Russell[1,16], Richard Caswell[1], Hana Lango Allen[1], Elisa De Franco[1], Timothy J McDonald[1], Hanna Rajala[4,5], Anita Ramelius[6,7], John Barton[8], Kaarina Heiskanen[3,9], Tarja Heiskanen-Kosma[10], Merja Kajosaari[9], Nuala P Murphy[11], Tatjana Milenkovic[12], Mikko Seppänen[13], Åke Lernmark[6,7], Satu Mustjoki[4,5], Timo Otonkoski[3,9], Juha Kere[2,3,14,15], Noel G Morgan[1], Sian Ellard[1] & Andrew T Hattersley[1]

Monogenic causes of autoimmunity provide key insights into the complex regulation of the immune system. We report a new monogenic cause of autoimmunity resulting from *de novo* germline activating *STAT3* mutations in five individuals with a spectrum of early-onset autoimmune disease, including type 1 diabetes. These findings emphasize the critical role of STAT3 in autoimmune disease and contrast with the germline inactivating *STAT3* mutations that result in hyper IgE syndrome.

Autoimmune disorders are usually multifactorial in etiology, involving a combination of background risk due to an individual's genetic makeup and environmental exposure. Considerable advances in understanding genetic susceptibility to autoimmune disease have identified common and disease-specific polymorphisms in the human leukocyte antigen (HLA) region and throughout the genome[1–3]. Rarely, monogenic defects can explain a specific clustering of autoimmune conditions and give important biological insights[4–6].

Type 1 diabetes (T1D) can occur in association with other autoimmune conditions because of a shared polygenic predisposition or, rarely, as part of a monogenic polyautoimmune disorder. T1D is a common feature of immunodysregulation, polyendocrinopathy, enteropathy X-linked (IPEX) syndrome, which is caused by *FOXP3* mutations, and autoimmune polyendocrinopathy syndrome 1 (APS1), which is caused by recessive mutations in *AIRE*[4,5]. In these polyendocrinopathy syndromes, autoimmune disease has a very young age of onset, with T1D usually diagnosed before 3 months of age in IPEX syndrome, which is in contrast to polygenic T1D that is very rare before 6 months of age[7]. Investigation of individuals with multiple early-onset autoimmune features might therefore identify new monogenic causes of autoimmunity.

We initially searched for a *de novo* mutation in an individual with early-onset polyautoimmunity (T1D (2 weeks), autoimmune hypothyroidism (3 years) and celiac disease (17 months)) by exome sequencing and comparison with variants identified in the unaffected parents. Heterozygous variants called using the Genome Analysis Toolkit (GATK) were filtered by removing noncoding and synonymous variants, variants in dbSNP131 or the 1000 Genomes Project database, and variants identified in either parent (Online Methods). This filtering reduced the number of potentially pathogenic *de novo* mutations to one encoding a single heterozygous missense substitution, p.Thr716Met (c.2147C>T), in the transactivation domain of STAT3 (**Supplementary Table 1**).

We sequenced the *STAT3* coding exons in 24 individuals with >2 early-onset (diagnosis at <5 years) autoimmune disorders of unknown cause and in 39 individuals with isolated permanent diabetes diagnosed at <6 months, aged <5 years at referral and having 20 known causes of neonatal diabetes excluded (Online Methods and **Supplementary Table 2**). We identified three different heterozygous missense mutations (encoding p.Lys392Arg, p.Asn646Lys and p.Lys658Asn substitutions) in four individuals (three with polyautoimmune disease and one with isolated permanent neonatal diabetes, aged 3 years). Analysis of parental DNA samples established that all mutations had arisen *de novo* (Online Methods). With the exception of p.Lys658Asn, which has been reported as a somatic alteration in large granular lymphocytic (LGL) leukemia[8], all mutations were previously unreported and absent in the 1000 Genomes Project and National Heart, Lung, and Blood Institute (NHLBI) exome databases. All substitutions affected highly conserved residues within the SH2 (Src homology 2), transactivation or DNA-binding domains (**Fig. 1a**). The discovery of *de novo* mutations in *STAT3* in five individuals with early-onset autoimmunity provides overwhelming genetic evidence that these are disease-causing mutations.

We assessed the activity of each mutant using a STAT3-responsive dual-luciferase reporter assay (Online Methods). Constructs encoding the four different *STAT3* mutations identified in individuals with polyautoimmune disease, two dominant-negative hyper IgE syndrome *STAT3* mutations and wild-type *STAT3* were generated and transiently transfected into cultured HEK293 cells (**Supplementary Fig. 1**). Under non-stimulated (basal) conditions, expression of all four newly

[1]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. [2]Folkhälsan Institute of Genetics, University of Helsinki, Helsinki, Finland. [3]Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland. [4]Hematology Research Unit Helsinki, Department of Hematology, University of Helsinki, Helsinki, Finland. [5]Helsinki University Central Hospital Cancer Center, Helsinki, Finland. [6]Department of Clinical Sciences, Lund University, Lund, Sweden. [7]Clinical Research Centre, Skåne University Hospital, Malmö, Sweden. [8]Bristol Royal Hospital for Children, Bristol, UK. [9]Children's Hospital, Helsinki University Central Hospital, Helsinki, Finland. [10]Department of Pediatrics, Kuopio University Hospital, Helsinki, Finland. [11]Department of Diabetes and Endocrinology, Children's University Hospital, Dublin, Ireland. [12]Department of Endocrinology, Institute for Mother and Child Health Care of Serbia 'Dr Vukan Cupic', Belgrade, Serbia. [13]Immunodeficiency Unit, Division of Infectious Diseases, Helsinki University Central Hospital, Helsinki, Finland. [14]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden. [15]Center for Innovative Medicine, Karolinska Institutet, Huddinge, Sweden. [16]These authors contributed equally to this work. Correspondence should be addressed to A.T.H. (a.hattersley@exeter.ac.uk), N.G.M. (n.g.morgan@exeter.ac.uk) or J.K. (juha.kere@ki.se).

Received 28 March; accepted 27 June; published online 20 July 2014; doi:10.1038/ng.3040

© 2014 Nature America, Inc. All rights reserved.

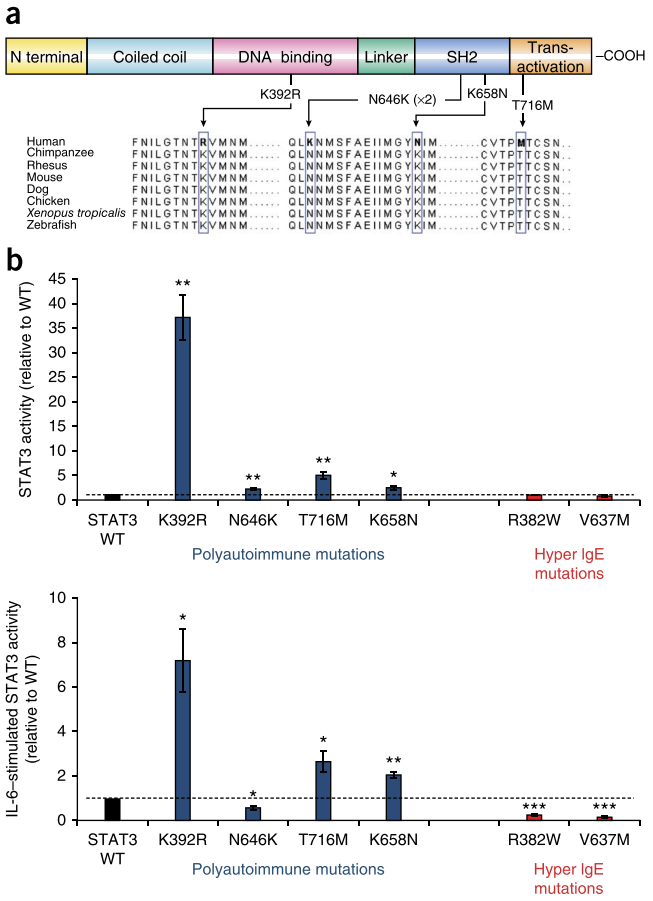
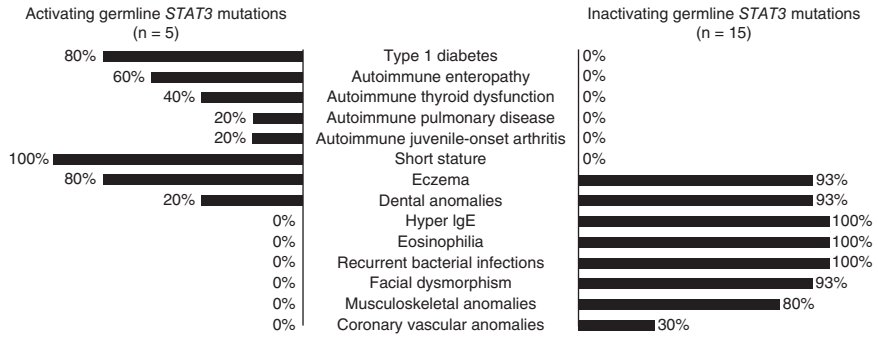


Figure 1 Activating *STAT3* mutations cause early-onset autoimmune disease. (**a**) Schematic of STAT3 (adapted from ref. 19). The positions of the four different *de novo* substitutions identified in five individuals with the polyautoimmune syndrome are shown below the STAT3 domains. The highly conserved sequence at the position of each substitution is provided for various species. (**b**) STAT3 activity of polyautoimmune mutants under non-stimulated (top) and IL-6–stimulated (20 ng/ml; bottom) conditions. The STAT3 reporter activity of four polyautoimmune mutants (Lys392Arg, Asn646Lys, Lys658Asn, Thr716Met) was examined alongside that of two previously described hyper IgE syndrome mutants (Arg382Trp, Val637Met)[20] and wild-type (WT) STAT3 after transient transfection into HEK293 cells. The dotted line indicates the activity of wild-type STAT3 under either basal (top) or IL-6–stimulated (bottom) conditions. Data are presented as average fold change relative to wild-type STAT3 ($n = 3$) under each experimental condition ± s.e.m. Typically, IL-6 caused a 20- to 30-fold increase in activity above basal levels in cells transfected with wild-type STAT3. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, *t* test.

discovered STAT3 mutants resulted in an increase in reporter activity in comparison to wild-type STAT3 or STAT3 with the hyper IgE syndrome inactivating substitutions (**Fig. 1b**). Expression of three of the four mutants also resulted in a less marked increase in reporter activity relative to wild-type STAT3 after stimulation with interleukin (IL)-6: for the Asn646Lys mutant, with the least activating alteration of those tested, although there was a significant increase in reporter activity under non-stimulated (basal) conditions ($P < 0.01$), there was no increase above the levels with wild-type STAT3 after stimulation with IL-6 (**Fig. 1b**). Molecular modeling of the p.Asn646Lys substitution suggests that this alteration is likely to result in higher DNA binding affinity owing to electrostatic interaction with the DNA backbone (**Supplementary Fig. 2**), leading to enhanced STAT3 activity.

To examine underlying gain of cytokine-related function, we performed T cell immunophenotyping and activation assays on samples from two affected individuals (with p.Lys658Asn and p.Lys392Arg alterations). Both showed reduced regulatory T cell numbers, and $CD4^{+}$ cells from case 2 also showed increased cytokine production (interferon (IFN)-γ and tumor necrosis factor (TNF)-α) when stimulated (**Supplementary Fig. 3** and **Supplementary Table 3**). These findings provide evidence for enhanced T helper type 1 (T_H1) cell response *in vivo*. These functional studies support the hypothesis that all of the *STAT3* mutations causing polyautoimmune disease are activating and lead to increased basal STAT3 activity *in vitro*.

Diabetes (in 4/5 individuals) presented early (median of 2.5 weeks, range of 0–43 weeks) and was insulin treated from diagnosis in doses required for total endogenous insulin deficiency. Three of the four affected individuals had detectable islet autoantibodies for T1D (**Supplementary Table 3**), supporting an autoimmune etiology. Autoimmune destruction of beta cells starting during fetal life is suggested by intrauterine growth retardation, a likely consequence of reduced insulin secretion *in utero*, and early-onset diabetes (≤3 weeks).



Additional autoimmune conditions diagnosed during childhood included autoimmune enteropathy, autoimmune interstitial lung disease, juvenile-onset arthritis and primary hypothyroidism. Other common features were short stature (5/5; <2 s.d. relative to the population mean) and eczema (4/5) (**Supplementary Table 3**). The young age at diagnosis of the autoimmune features is consistent with *STAT3* mutations causing accelerated autoimmune disease.

The signal transducer and activator of transcription (STAT) factors are involved in multiple processes, including early development and cellular proliferation, survival and differentiation. It is possible that STAT3 activation leads to autoimmunity by impairing the development of regulatory T cells and promoting the expansion and activation of T helper type 17 (T_H17) cells[9,10]. T_H17 cell expansion, activation and dysregulation are thought to have a critical role in many autoimmune diseases, including T1D[11,12]. In addition, direct cell type–specific effects of constitutively active STAT3 might also have a role in the development of autoimmunity.

Germline dominant-negative *STAT3* mutations in individuals with hyper IgE syndrome cause a primary immunodeficiency disorder characterized by elevated serum IgE levels and recurrent staphylococcal infections resulting from a deficiency in T_H17 cells[12,13].

Figure 2 Clinical characteristics associated with activating germline *STAT3* mutations causing the polyautoimmune syndrome and inactivating germline *STAT3* mutations causing hyper IgE syndrome[21].

© 2014 Nature America, Inc. All rights reserved.

The spectrum of multiple autoimmune features observed in our cohort with germline activating *STAT3* mutations is in contrast to the severe immunodeficiency observed in individuals with hyper IgE syndrome (**Fig. 2**).

Somatic activating *STAT3* mutations have been reported in 40% of individuals with LGL leukemia[8] and in 8% of individuals with inflammatory hepatocellular adenoma[14]. The primary presentation of LGL leukemia is recurrent bacterial infections, fatigue and autoimmune cytopenias typically in the sixth decade of life. Affected individuals frequently (26%) have adult-onset rheumatoid arthritis, but diabetes and other autoimmune disease have not been reported[8]. Therefore, somatic and germline *STAT3* mutations result in different phenotypes.

Common variation in *STAT3* is associated with autoimmune disease, with SNPs conferring altered susceptibility to Crohn's disease[2], psoriasis[15] and multiple sclerosis[16], but, interestingly, these conditions were not seen in our series of affected individuals. Suggestive associations were seen in T1D and rheumatoid arthritis but did not reach genome-wide significance[17,18]. Further studies investigating the role of rare coding variants in susceptibility to common organ-specific autoimmune disease are warranted.

STAT3-associated polyautoimmunity has a different but overlapping clinical phenotype in comparison to the four previously described monogenic autoimmune syndromes (**Supplementary Table 4**). Although these disorders are extremely rare, they offer valuable insights into the biology of the immune system in T1D and other related autoimmune disease. Further large-scale sequencing efforts of individuals with multiple early-onset autoimmune disorders are required to identify additional monogenic conditions, which will shed further light on the complex regulation of the adaptive immune response.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Chilton, A. Damhuis, R. Raman and B. Yang for technical assistance. This work was supported by the UK National Institute for Health Research (NIHR) Exeter Clinical Research Facility through funding for S.E. and A.T.H. A.T.H. is an NIHR Senior Investigator. S.E. and A.T.H. are supported by Wellcome Trust Senior Investigator awards. Further funding was provided by Diabetes UK and the Finnish Medical Foundation.

AUTHOR CONTRIBUTIONS

S.E.F., S.E. and A.T.H. designed the study. N.P.M., T.M., T.O., E.H., K.H., T.H.-K., M.K., A.R., A.L. and J.B. recruited subjects to the study. R.C. and E.D.F. performed the exome sequencing and targeted next-generation sequence analysis. H.L.A. performed the bioinformatics analysis. S.E.F. and E.H. performed the Sanger sequencing analysis and the interpretation of the resulting data. S.E.F., T.J.M., E.H., M.S., J.K. and A.T.H. analyzed the clinical data. M.A.R. and N.G.M. designed and performed the functional studies. H.R. and S.M. performed the T cell assays. S.E.F., M.A.R. and A.T.H. prepared the draft manuscript. All authors contributed to discussion of the results and to manuscript preparation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Barrett, J.C. *et al. Nat. Genet.* **41**, 703–707 (2009).
2. Barrett, J.C. *et al. Nat. Genet.* **40**, 955–962 (2008).
3. Stahl, E.A. *et al. Nat. Genet.* **42**, 508–514 (2010).
4. Bennett, C.L. *et al. Nat. Genet.* **27**, 20–21 (2001).
5. Finnish-German APECED Consortium. *Nat. Genet.* **17**, 399–403 (1997).
6. Sharfe, N. *et al. Proc. Natl. Acad. Sci. USA* **94**, 3168–3171 (1997).
7. Rubio-Cabezas, O. *et al. Diabetes Care* **32**, 111–116 (2009).
8. Koskela, H.L. *et al. N. Engl. J. Med.* **366**, 1905–1913 (2012).
9. Yang, X.O. *et al. J. Biol. Chem.* **282**, 9358–9363 (2007).
10. Harris, T.J. *et al. J. Immunol.* **179**, 4313–4317 (2007).
11. Shao, S. *et al. Cell. Immunol.* **280**, 16–21 (2012).
12. Ma, C.S. *et al. J. Exp. Med.* **205**, 1551–1557 (2008).
13. Minegishi, Y. *et al. Nature* **448**, 1058–1062 (2007).
14. Pilati, C. *et al. J. Exp. Med.* **208**, 1359–1366 (2011).
15. Tsoi, L.C. *et al. Nat. Genet.* **44**, 1341–1348 (2012).
16. Jakkula, E. *et al. Am. J. Hum. Genet.* **86**, 285–291 (2010).
17. Fung, E.Y. *et al. Genes Immun.* **10**, 188–191 (2009).
18. Seddighzadeh, M. *et al. J. Rheumatol.* **39**, 1509–1516 (2012).
19. Levy, D.E. & Darnell, J.E. Jr. *Nat. Rev. Mol. Cell Biol.* **3**, 651–662 (2002).
20. Holland, S.M. *et al. N. Engl. J. Med.* **357**, 1608–1619 (2007).
21. Heimall, J. *et al. Clin. Immunol.* **139**, 75–84 (2011).

© 2014 Nature America, Inc. All rights reserved.

ONLINE METHODS

Cohort selection and sample preparation. Twenty-five individuals with early-onset polyautoimmune disease (diagnosed before 5 years of age) and 39 subjects with isolated permanent diabetes diagnosed before 6 months of age were recruited by their clinicians for molecular genetic analysis in the Exeter Molecular Genetics Laboratory ($n = 63$) or the Folkhälsan Institute of Genetics, University of Helsinki ($n = 1$). Genomic DNA was extracted from peripheral leukocytes using standard procedures. All subjects and/or their parents gave informed consent for genetic testing, and institutional review board approval was received for this study (12/WA/0225 approved by North Wales Research Ethics Committee–West).

Exome sequencing and variant calling. Genomic regions corresponding to the NCBI Consensus Coding Sequence (CCDS) database were captured and amplified using the SureSelect Human All Exon kit (v1; Agilent Technologies). Paired-end sequencing was performed on an Illumina Genome Analyzer IIx, one lane per sample, with 101-bp or 76-bp read length. The resulting reads were aligned to the hg19 reference genome with the Burrows-Wheeler Aligner (BWA), providing mean target coverage of 66.3 reads per base. At least 72% of the targeted bases were covered by at least 20 reads. Variants were called with the GATK UnifiedGenotyper and annotated using Annovar and the SeattleSeq Annotation server, as previously described[22]. Variant filtering steps are shown in **Supplementary Table 1**.

***STAT3* sequencing and microsatellite analysis.** Sanger sequencing was undertaken in case 1 and her unaffected parents to confirm that the *STAT3* mutation encoding the p.Thr716Met variant had arisen *de novo*. Exons 2–24 and intron-exon boundaries of *STAT3* (NM_139276.2) underwent Sanger sequencing in a further 24 individuals with at least 2 early-onset autoimmune features of unknown cause (**Supplementary Table 2**). Primer sequences for *STAT3* exons 2–24 are provided in **Supplementary Table 5**.

Targeted next-generation sequencing of *STAT3* was undertaken on a further 39 individuals with isolated permanent diabetes diagnosed before the age of 6 months of unknown cause without additional autoimmune features. All cases were less than 5 years of age at the time of genetic testing. We adapted our custom Agilent SureSelect exon-capture assay (Agilent Technologies) to include baits for exons 2–24 and intron-exon boundaries of *STAT3* (sequences available upon request from authors)[23]. Samples were fragmented using a Bioruptor (Diagenode), indexed for multiplexing and hybridized (in pools of 12 samples) according to the manufacturer's instructions. Sequencing was performed with an Illumina HiSeq 2000 instrument (48 samples per lane) and 100-bp paired-end reads. Data were processed to identify potential pathogenic mutations located within 50 bp upstream and 10 bp downstream of each exon.

We identified *STAT3* mutations in a further 3 individuals with early-onset autoimmune disease (4 of 25 in total; 16% of the cohort) and 1 individual with permanent neonatal diabetes (1 of 39 in total; 2.6% of the cohort). This brought the total number of subjects positive for *STAT3* mutation to five. Sanger sequencing of parental samples confirmed that all mutations had arisen *de novo*. Biological relationships were confirmed by microsatellite analysis using the PowerPlex kit (PowerPlex 16 System, Promega).

In total, four different mutations were identified in five unrelated individuals. All mutations affected residues in the highly conserved DNA-binding domain (×1), SH2 domain (×2) and transactivation domain (×1) (conserved to zebrafish). None of the mutations were present in dbSNP132, the 1000 Genomes Project database (based on 1,094 individuals) or the Exome Sequencing Project (based on 6,500 individuals).

Functional studies of *STAT3* mutations. Mutations in human *STAT3* (Source Bioscience) were generated using the QuikChange site-directed mutagenesis kit, following the manufacturer's guidelines (Agilent Technology). The sequences of the primer pairs used to generate each mutant are provided in **Supplementary Table 6**. The success of all mutagenesis reactions was confirmed by direct sequencing of the entire *STAT3* insert (Eurofins). After mutagenesis, *STAT3* inserts were subcloned into the multiple-cloning site of a pcDNA5/FRT/TO expression vector between AflII and EcoRV restriction sites.

The transcriptional activity of STAT3 was assessed via a STAT3-responsive dual firefly/Renilla luciferase Cignal reporter system (Qiagen). HEK293 cells were seeded at a density of 1×10^5 cells/well and were transfected after 24 h with a combination of 200 ng of Cignal reporter assay constructs and 400 ng of wild-type or mutant *STAT3* pcDNA5/FRT/TO using the Attractene transfection reagent according to the manufacturer's instructions (Qiagen). Cells were incubated in the transfection mix for 24 h, and, where appropriate, 20 ng/ml IL-6 was also added for the final 18 h of incubation. STAT3 reporter activity was assessed using a dual-luciferase reporter assay system (Promega). To confirm that equivalent amounts of STAT3 protein were expressed after transfection with each construct, cells were lysed and protein was extracted before protein blotting with antibody to STAT3 (Cell Signaling Technology; 1:1,000 dilution) as described previously[24]. HEK293 cells were a gift from A. Demaine (Plymouth University); we did not test them for mycoplasma.

Regulatory T cell immunophenotyping and evaluation of T cell cytokine production. Fresh peripheral blood mononuclear cells (PBMCs) from cases 2 and 5 (**Supplementary Table 3**) and six healthy controls were used. For regulatory T cell immunophenotyping, cells were stained using monoclonal antibodies against the antigens CD3, CD4 and CD25 (333170, BD Biosciences) and FOXP3 (clone259D, eBioscience), and data were collected with four- or six-color flow cytometry. Regulatory T cells were defined as $CD3^+CD4^+CD25^{high}FOXP3^+$. Evaluation of T cell cytokine production is described in detail elsewhere[24]. Briefly, fresh mononuclear cells were stimulated for 6 h with antibodies to CD3, CD28 and CD49d (BD Biosciences). Cells were analyzed using a six-color flow cytometry panel with monoclonal antibodies against the antigens CD3, CD4, CD8, IFN-γ and TNF-α (BD Biosciences). The data were analyzed with the FACSAria II flow cytometer and FACSDiva software (BD Biosciences).

22. Lango Allen, H. *et al. Nat. Genet.* **44**, 20–22 (2012).
23. Ellard, S. *et al. Diabetologia* **56**, 1958–1963 (2013).
24. Russell, M.A. *et al. Islets* **5**, 95–105 (2013).

© 2014 Nature America, Inc. All rights reserved.