nature genetics

Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease

Steven A McCarroll[1–3], James E Bradner[3,4], Hannu Turpeinen[5], Liisa Volin[6], Paul J Martin[7,8], Shannon D Chilewski[3], Joseph H Antin[4], Stephanie J Lee[4,7], Tapani Ruutu[6], Barry Storer[7], Edus H Warren[7,8], Bo Zhang[7], Lue Ping Zhao[7], David Ginsburg[9], Robert J Soiffer[4], Jukka Partanen[5], John A Hansen[7,8], Jerome Ritz[4], Aarno Palotie[3,10,11] & David Altshuler[1–3]

Transplantation and pregnancy, in which two diploid genomes reside in one body, can each lead to diseases in which immune cells from one individual target antigens encoded in the other's genome. One such disease, graft-versus-host disease (GVHD) after hematopoietic stem cell transplantation (HSCT, or bone marrow transplant), is common even after transplants between HLA-identical siblings, indicating that cryptic histocompatibility loci exist outside the HLA locus. The immune system of an individual whose genome is homozygous for a gene deletion could recognize epitopes encoded by that gene as alloantigens. Analyzing common gene deletions in three HSCT cohorts (1,345 HLA-identical sibling donor-recipient pairs), we found that risk of acute GVHD was greater (odds ratio (OR) = 2.5; 95% confidence interval (CI) 1.4–4.6) when donor and recipient were mismatched for homozygous deletion of *UGT2B17*, a gene expressed in GVHD-affected tissues and giving rise to multiple histocompatibility antigens. Human genome structural variation merits investigation as a potential mechanism in diseases of alloimmunity.

GVHD is a serious, common complication of allogeneic HSCT (or bone marrow transplant) in which immune responses by donor-derived lymphocytes target alloantigens in the host. GVHD rarely occurs after transplants between monozygotic twins[1,2] but frequently occurs after transplants between HLA-identical siblings, indicating that additional histocompatibility loci must exist outside the HLA locus.

One clear example of a non-HLA compatibility locus is the Y chromosome, in that there is increased risk of GVHD when HSCT involves a female donor and a male recipient[3]. This effect arises from immune recognition (by donor-derived lymphocytes and antibodies) of antigens encoded by a few Y-linked genes that are expressed in the soma[4–9]. These genes collectively differ in sequence from their X-linked paralogs at only a few hundred amino acids[10]. This observation demonstrates that changes in an individual's antigen repertoire of hundreds of amino acids, the size of many individual autosomal genes, can increase risk of GVHD.

The human genome is increasingly recognized to have extensive structural polymorphism[11,12], including deletions of entire autosomal genes[13,14]. Some of these gene-deletion alleles are sufficiently common that individuals inherit them from both parents and therefore completely lack a protein-coding gene that is expressed in other individuals[13]. Because the immune system of an individual with a homozygous gene deletion presumably has not learned to tolerate the protein encoded by that gene, immune recognition of that protein as an alloantigen[15,16] by immune cells or antibodies from that individual could in principle contribute to risk of alloimmune disease.

To assess whether donor-recipient mismatch for homozygous gene deletions increases the risk of GVHD after transplantation, we first typed a set of common gene deletions in 414 HSCT patients and their HLA-identical sibling donors (cohort A; Online Methods, **Table 1**). This screen involved gene-deletion polymorphisms that we identified from a genome-wide survey of copy number variation[17] as satisfying the following criteria: a deletion allele that (i) removes the gene protein-coding sequence and (ii) segregates in the population with allele frequency >10%; (iii) expression of the gene in one or more of the tissues commonly involved in acute GVHD (liver, intestine, skin); and (iv) significant sequence difference (tens to hundreds of amino acids) from any paralogous gene encoded elsewhere in the human genome (**Table 2**). The following genes were identified from this screen: *UGT2B17*, *UGT2B28*, *GSTM1*, *GSTT1*, *LCE3C* and *OR51A2*. We assessed association with GVHD risk by first typing each deletion in HSCT patients and their sibling donors so as to determine which transplants involved homozygous gene deletion in the donor but not the recipient (**Supplementary**

[1]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [3]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [4]Division of Hematologic Malignancies and Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Research and Development, Finnish Red Cross Blood Service, Helsinki, Finland. [6]Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland. [7]Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA. [8]Department of Medicine, University of Washington, Seattle, Washington, USA. [9]Departments of Human Genetics and Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA. [10]Finnish Institute of Molecular Medicine, University of Helsinki, Helsinki, Finland. [11]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. Correspondence should be directed to S.A.M. (mccarroll@genetics.med.harvard.edu).

Received 23 April; accepted 23 October; published online 22 November 2009; doi:10.1038/ng.490

Table 1 Characteristics of the clinical populations in the GVHD study

	Cohort A (initial screen)		Cohort B	Cohort C
	A1	A2		
Hospital	Helsinki University Central Hospital	Dana-Farber Cancer Institute	Various	Fred Hutchinson Cancer Research Center
Pretransplant conditioning	Myeloablative	Myeloablative	Myeloablative	Myeloablative
Transplant year	1993–2005	1995–2005	1985–1993	1992–2004
Donor type	HLA-identical sibling	HLA-identical sibling	HLA-identical sibling	HLA-identical sibling
Donor-recipient pairs	232	182	336	595
Graft type				
Bone marrow	77	36	336	265
Peripheral blood stem cells	155	146	0	226
Both	0	0	0	4
Sex (donor-patient)				
Female-male	50	45	74	164
Female-female	57	46	86	115
Male-male	66	51	90	203
Male-female	59	40	86	113
Acute GVHD (grade II–IV)				
Yes	39	54	243	410
No	193	128	93	180
Not gradable	0	0	0	5
Disease				
Nonmalignant disease	7	8	0	0
Malignant hematological disease	225	174	0	595
Not recorded	0	0	336	0

Acute GVHD was diagnosed and graded according to standard criteria in use at the time of documented patient care. Note that the ratio of affected to unaffected individuals is not uniform across cohorts, primarily because the criteria for patient selection differed from institution to institution. These selection criteria were defined before genetic analysis and for cohorts B and C reflected the design of earlier studies. Cohort C used a cohort design, whereas cohort B used a case-control design and oversampled patients with severe acute GVHD to maximize power. Diagnostic sensitivity in the grading of GVHD can also differ among clinical institutions[29], although it is generally well harmonized within institutions.

Note). At each locus we then assessed whether such mismatches were associated with increased risk of acute GVHD (**Fig. 1a**).

In this initial screen, donor-recipient mismatch for homozygous deletion of one of these genes, *UGT2B17*, showed a promising potential association with GVHD (**Fig. 1a**; OR = 3.0, 95% CI 1.3–6.9, nominal $P = 0.006$, by Cochran-Mantel-Haenszel test; 0.03 after Bonferroni correction). *UGT2B17* encodes a cell-surface protein comprising 530 amino acids that is highly expressed in the same tissues—liver, intestine and skin—that are affected by clinically apparent GVHD and targeted by donor-derived lymphocytes. For the other five gene deletions tested, we observed no evidence of association of donor-recipient mismatch with acute GVHD (**Fig. 1a**).

We further assessed the contribution of *UGT2B17* mismatches to GVHD in two additional patient cohorts (cohorts B and C, **Table 1**). Outcomes in cohorts B and C also involved an increased risk in transplants involving donor-recipient mismatch at *UGT2B17* (OR = 2.4, 95% CI 1.1–5.1, $P = 0.02$, by Cochran-Mantel-Haenszel test), strengthening the overall evidence for association (**Fig. 1b**, **Supplementary Table 1**) (OR = 2.5, $P = 5 \times 10^{-4}$, by Cochran-Mantel-Haenszel test).

We evaluated alternative models for the association of GVHD with donor-recipient mismatch at *UGT2B17*. In particular, the observed association might in principle be due to donor genotype or patient genotype independent of donor-recipient mismatch, particularly given that variation at *UGT2B17* associates to other clinical phenotypes[18–20]. We therefore evaluated the data from all 1,345 donor-recipient pairs in cohorts A, B and C (Cochran-Mantel-Haenszel test) to assess the risk of GVHD for each combination of donor and recipient *UGT2B17* status, relative to a reference group in which donor and recipient were both *UGT2B17* positive (**Fig. 1c**). Increased risk was confined to the group of transplants for which donors were *UGT2B17* negative and recipients were *UGT2B17* positive (**Fig. 1c**). In particular, the *UGT2B17* status of the hematopoietic stem cell (HSC) donor was not associated with GVHD when HSC recipients were *UGT2B17* negative, and the *UGT2B17* status of the HSC recipient was not associated with GVHD when HSC donors were *UGT2B17* positive (**Fig. 1c**).

To assess the time course of GVHD incidence in patients with *UGT2B17*-mismatched donors, a cumulative incidence analysis was performed (**Fig. 2**). *UGT2B17*-positive patients with *UGT2B17*-negative donors showed an unremarkable incidence of acute GVHD during the first 20 d after transplantation; GVHD then began to increase beyond the level observed in other patients after day 20, a pattern similar to that observed in male patients who receive transplants from female donors.

A corollary of the model in which common gene deletions contribute to GVHD risk is that immune responses to multiple antigens derived from UGT2B17 would be present in patients. Several years ago a cytotoxic T-cell clone derived from a patient with acute GVHD of the intestine was used to screen an expression library to identify the antigen recognized by the T cell; the antigen was determined to be AELLNIPFLY, a peptide derived from UGT2B17 and presented by HLA-A*2902 in this patient[16]. AELLNIPFLY was subsequently found to be presented also by

Table 2 Characteristics of the gene deletions assayed in the GVHD study

Gene	Peptide length	Genomic location	Sites of expression	Deletion size (kb)	Deletion extent
UGT2B17	530	Chr. 4: 69.2 Mb	Liver, intestine, skin, prostate, lymphocytes	129	All exons
UGT2B28	529	Chr. 4: 70.3 Mb	Liver, mammary gland, kidney, salivary gland	104	All exons
GSTT1	240	Chr. 22: 22.7 Mb	All tissues	114	All exons
GSTM1	218	Chr. 1: 109.9 Mb	All tissues	19	All exons
LCE3C	95	Chr. 1: 149.4 Mb	Connective epithelia	30	All exons
OR51A2	313	Chr. 11: 4.9 Mb	Olfactory epithelia	8	All exons

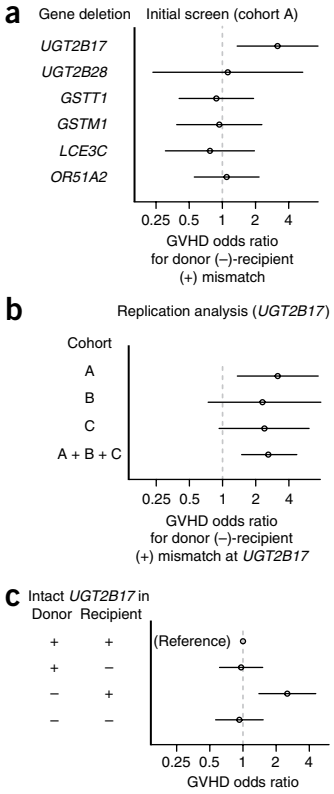

Figure 1 Association analysis of donor-recipient mismatch for common gene deletions in GVHD. (**a**) Initial screen for association of acute GVHD with donor-recipient mismatch for common gene deletions, in cohort A. Six common gene-deletion polymorphisms were screened by typing in donors and recipients. Data represent association of donor (–)-recipient (+) mismatch with the development of GVHD after transplantation. Odds ratios and 95% confidence intervals are shown. (**b**) Analysis in additional patient cohorts of the association of donor (–)-recipient (+) mismatch at *UGT2B17* with acute GVHD. (**c**) Association of *UGT2B17* deletion in donor and recipient with GVHD risk. The group of transplants in which both donor and patient were *UGT2B17* positive is used as the reference group for analysis. Odds ratios and confidence intervals were calculated using the Cochran-Mantel-Haenszel test to combine data from the 1,345 donor-recipient pairs from cohorts A, B and C.

HLA-B*4403 and recognized in this form by a distinct T-cell clone from the same patient[21]. Screening of a cytotoxic T-cell clone from a different patient recently identified CVATMIFMI, a different UGT2B17-derived peptide, as the antigen recognized by the clone, when presented by the common HLA allele HLA-A*0206 (ref. 22). Given evidence that antibody responses can also contribute to GVHD[23–26], and the possibility that UGT2B17's predicted localization to the cell surface would facilitate such responses, we screened sera from 26 GVHD patients (including ten with *UGT2B17*-mismatched donors) against an array of overlapping peptides designed to span the UGT2B17 protein sequence. Serum from one patient (a *UGT2B17*-positive patient whose donor was *UGT2B17* negative) showed a robust antibody response to the peptide LQESKFDVLLADAVNPCGEL ($UGT2B17_{141–160}$); we fine mapped this response to the epitope VLLADAVNP ($UGT2B17_{148–156}$). Antibodies from this patient distinguished this epitope from paralogous peptide sequences encoded by all other UGT2B genes in the human genome. These data collectively indicate that UGT2B17 gives rise to multiple histocompatibility antigens (**Fig. 3**), offering a candidate molecular and cellular mechanism for genetic association of *UGT2B17* mismatch with GVHD.

Several features of the UGT2B17 protein may make *UGT2B17* a more potent histocompatibility locus than other gene deletions: (i) UGT2B17 is a large protein (530 amino acids), increasing the likelihood that it contains multiple antigenic epitopes; (ii) UGT2B17 is abundant in liver, intestine and skin, the tissues in which pre-HSCT conditioning elicits the strongest inflammation and in which immune surveillance for alloantigens may therefore be strongest; (iii) UGT2B17 is expressed on the cell surface, well positioned to contribute to antibody-mediated as well as cell-mediated immune responses; and (iv) UGT2B17 is also abundant in blood, skin, semen and placenta, tissues that give rise to interindividual immune exposures that may pre-expose and immunize *UGT2B17*-negative individuals against UGT2B17, a phenomenon that has been observed in healthy female donors for some of the antigens encoded on the Y chromosome[23].

Although an estimate of effect size for donor-recipient *UGT2B17* mismatches in GVHD based on the cohorts analyzed here (OR = 2.5) is comparable to the established effect of sex mismatch (female donor, male recipient), *UGT2B17* mismatches cannot explain a comparable fraction of GVHD incidence because of the lower frequency at which *UGT2B17* mismatches arise between siblings. This sibling mismatch frequency varies among populations as a result of population variation in frequency of the *UGT2B17*-deletion allele (19–85%, an unusually high level of variation that has been attributed to adaptive evolution of *UGT2B17* copy number[27]); as a result, the expected frequency of sibling mismatches ranges from 2% in African Americans to 5% in most European populations to 9% in Gujarati Indians but does not approach the frequency of female-to-male sex mismatch (25%) in any population (**Supplementary Table 2**). We also caution that the association observed here should be considered preliminary until confirmed by independent, multicenter investigations, and that it may not extend to transplants involving unrelated donors[19], for which the potent effects of donor-recipient mismatches at untyped HLA loci[28] may dominate the effects of mismatches outside the HLA.

The gene-deletion polymorphisms analyzed here could in principle be a leading edge of a larger class of polymorphisms that have many and varied effects on antigen repertoire. Such polymorphisms might include not only deletion alleles but SNPs that introduce early stop codons and frameshifts into protein-coding sequences, alleles that alter transcript splicing, and null regulatory alleles. As large-scale sequence data increasingly make it possible to catalog all the single-nucleotide and structural polymorphisms that segregate in human populations, it will be important to identify the polymorphisms that have multifold effects on antigen repertoire.

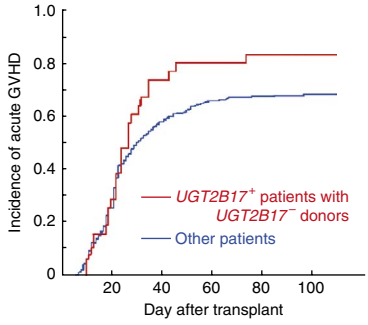

Figure 2 Cumulative incidence of acute GVHD during the first 100 days after HSCT (cohort C). In this analysis, death and recurring malignancy before the onset of grades II–IV GVHD were treated as competing risks.

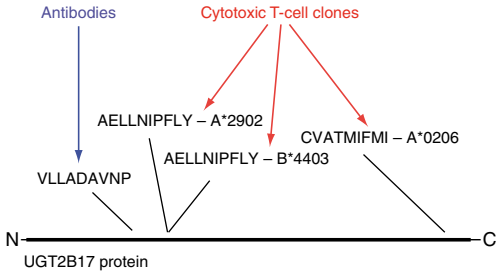

Figure 3 Multiple histocompatibility antigens derived from UGT2B17. An antibody response to the UGT2B17-derived peptide VLLADAVNP was detected in the serum of a *UGT2B17*-positive GVHD patient whose donor was *UGT2B17* negative. Responses of T cell clones against multiple antigens derived from UGT2B17 and presented by three distinct HLA alleles have been detected in other GVHD patients[16,21,22].

Human genome structural variation has been proposed to affect phenotypes by altering gene dosage and by affecting the regulation of nearby genes. Structural polymorphism may also influence disease by a very different mechanism—by giving rise to multifold differences between individuals' antigen repertoires that arise from specific genomic loci. The generality of such relationships—in other transplantation settings and in pregnancy—will be an important subject of investigation.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We wish to thank the patients, donors and clinical care teams at Helsinki University Central Hospital, the Dana-Farber Cancer Institute and the Fred Hutchinson Cancer Research Center. We also wish to thank D. Sese and colleagues at the Brigham and Women's Hospital Tissue Typing Laboratory; the Ted and Eileen Pasquarello Tissue Bank; Q. Yang and colleagues at the Quality Assurance Office for Clinical Trials at Dana-Farber Cancer Institute; E. Lander, Y. Ofran, J. Chien and M. Daly for helpful conversations on the project and data; and B. Blazar, C. Keever-Taylor and D. Senitzer for providing patient samples for the Nichols *et al.* cohort further studied here. This work was supported by the Broad Institute of MIT and Harvard (D.A., S.A.M.), the Academy of Finland (L.V., H.T., J.P.), the Helsinki University Central Hospital Research Fund (L.V.), the Fred Hutchinson Cancer Research Center, a Lilly Life Sciences Research fellowship (S.A.M.), the S. Juselius Foundation (J.P.), the Center of Excellence Program of the Finnish Academy (A.P.) and the National Institutes of Health (HA070149 to J.H.A.; AI29530 to J.R.; AI33484, CA18029 and HL087690 to J.A.H.).

AUTHOR CONTRIBUTIONS

S.A.M., J.E.B. and D.A. developed an initial study plan and refined this plan with insights from R.J.S., J.R., A.P., E.H.W. and P.J.M. Patient collections including DNA samples were established and/or further developed by L.V., H.T., T.R. and J.P. in Helsinki; J.E.B., S.J.L., J.H.A., J.R. and R.J.S. in Boston; D.G. in Michigan; and P.J.M., S.J.L., B.S. and J.A.H. in Seattle. Analyses of patient clinical data were led, performed and/or further analyzed by L.V. and T.R. at Helsinki University Central Hospital; S.J.L., J.E.B., J.H.A., R.J.S. and J.R. at Dana-Farber Cancer Institute; and B.S., P.J.M., S.J.L., E.H.W. and J.A.H. at the Fred Hutchinson Cancer Research Center (FHCRC). Deletion polymorphisms were genotyped by S.A.M., H.T. and S.D.C., using molecular assays developed by S.A.M. B.Z. and L.P.Z. analyzed array-based data to genotype deletions and analyzed association and time course in the FHCRC cohort. S.A.M. performed statistical analyses of genotype-phenotype correlation with feedback from other authors, especially P.J.M., A.P. and D.A. S.D.C. and H.T. performed ELISA experiments. S.A.M. wrote the manuscript with extensive input and feedback from coauthors.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Gale, R.P. & Champlin, R.E. How does bone-marrow transplantation cure leukaemia? *Lancet* **2**, 28–30 (1984).
2. Gale, R.P. *et al.* Identical-twin bone marrow transplants for leukemia. *Ann. Intern. Med.* **120**, 646–652 (1994).
3. Randolph, S.S., Gooley, T.A., Warren, E.H., Appelbaum, F.R. & Riddell, S.R. Female donors contribute to a selective graft-versus-leukemia effect in male recipients of HLA-matched, related hematopoietic stem cell transplants. *Blood* **103**, 347–352 (2004).
4. Wang, W. *et al.* Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science* **269**, 1588–1590 (1995).
5. Meadows, L. *et al.* The HLA-A*0201-restricted H-Y antigen contains a posttranslationally modified cysteine that significantly affects T cell recognition. *Immunity* **6**, 273–281 (1997).
6. Warren, E.H. *et al.* The human UTY gene encodes a novel HLA-B8-restricted H-Y antigen. *J. Immunol.* **164**, 2807–2814 (2000).
7. Vogt, M.H. *et al.* The DBY gene codes for an HLA-DQ5-restricted human male-specific minor histocompatibility antigen involved in graft-versus-host disease. *Blood* **99**, 3027–3032 (2002).
8. Spierings, E. *et al.* Identification of HLA class II-restricted H-Y-specific T-helper epitope evoking $CD4^+$ T-helper cells in H-Y-mismatched transplantation. *Lancet* **362**, 610–615 (2003).
9. Ivanov, R. *et al.* Identification of a 40S ribosomal protein S4-derived H-Y epitope able to elicit a lymphoblast-specific cytotoxic T lymphocyte response. *Clin. Cancer Res.* **11**, 1694–1703 (2005).
10. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
11. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
12. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
13. McCarroll, S.A. *et al.* Common deletion polymorphisms in the human genome. *Nat. Genet.* **38**, 86–92 (2006).
14. Conrad, D.F., Andrews, T.D., Carter, N.P., Hurles, M.E. & Pritchard, J.K. A high-resolution survey of deletion polymorphism in the human genome. *Nat. Genet.* **38**, 75–81 (2006).
15. Colin, Y. *et al.* Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. *Blood* **78**, 2747–2752 (1991).
16. Murata, M., Warren, E.H. & Riddell, S.R. A human minor histocompatibility antigen resulting from differential expression due to a gene deletion. *J. Exp. Med.* **197**, 1279–1289 (2003).
17. McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174 (2008).
18. Schulze, J.J. *et al.* Doping test results dependent on genotype of uridine diphospho-glucuronosyl transferase 2B17, the major enzyme for testosterone glucuronidation. *J. Clin. Endocrinol. Metab.* **93**, 2500–2506 (2008).
19. Terakura, S. *et al.* A *UGT2B17*-positive donor is a risk factor for higher transplant-related mortality and lower survival after bone marrow transplantation. *Br. J. Haematol.* **129**, 221–228 (2005).
20. Yang, T.L. *et al.* Genome-wide copy-number-variation study identified a susceptibility gene, *UGT2B17*, for osteoporosis. *Am. J. Hum. Genet.* **83**, 663–674 (2008).
21. Terakura, S. *et al.* A single minor histocompatibility antigen encoded by *UGT2B17* and presented by human leukocyte antigen-A*2902 and -B*4403. *Transplantation* **83**, 1242–1248 (2007).
22. Kamei, M. *et al.* HapMap scanning of novel human minor histocompatibility antigens. *Blood* **113**, 5041–5048 (2009).
23. Miklos, D.B. *et al.* Antibody response to DBY minor histocompatibility antigen is induced after allogeneic stem cell transplantation and in healthy female donors. *Blood* **103**, 353–359 (2004).
24. Miklos, D.B. *et al.* Antibody responses to H-Y minor histocompatibility antigens correlate with chronic graft-versus-host disease and disease remission. *Blood* **105**, 2973–2978 (2005).
25. Zorn, E. *et al.* Minor histocompatibility antigen DBY elicits a coordinated B and T cell response after allogeneic stem cell transplantation. *J. Exp. Med.* **199**, 1133–1142 (2004).
26. Cutler, C. *et al.* Rituximab for steroid-refractory chronic graft-versus-host disease. *Blood* **108**, 756–762 (2006).
27. Xue, Y. *et al.* Adaptive evolution of UGT2B17 copy-number variation. *Am. J. Hum. Genet.* **83**, 337–346 (2008).
28. Petersdorf, E.W., Malkki, M., Gooley, T.A., Martin, P.J. & Guo, Z. MHC haplotype matching for unrelated hematopoietic cell transplantation. *PLoS Med.* **4**, e8 (2007).
29. Martin, P.J. *et al.* Increasingly frequent diagnosis of acute gastrointestinal graft-versus-host disease after allogeneic hematopoietic cell transplantation. *Biol. Blood Marrow Transplant.* **10**, 320–327 (2004).

ONLINE METHODS

Study populations. The first study population (cohort A) consisted of 414 allogeneic HSCT recipients and their HLA-identical sibling donors undergoing transplant at Helsinki University Central Hospital (HUCH) and the Dana-Farber Cancer Institute (DFCI) for the treatment of a hematological disease (**Table 1**). Analysis was retrospective. Recipients had received myeloablative conditioning followed by a graft from an HLA-identical sibling between 1993 and 2005. All aspects of human subjects research adhered to protocols approved by the institutional review boards of HUCH and DFCI (Protocol 01-206).

A second study cohort (cohort B, **Table 1**) consisted of 336 bone marrow recipients and their HLA-identical sibling donors. This cohort has been described elsewhere[30] and consists of sibling donor-recipient pairs who had allogeneic bone marrow transplantation from 1985 through 1993 at several US institutions. Analysis was retrospective. Samples were obtained from patients with grades III or IV acute GVHD and from patients without clinically apparent acute GVHD. Patients with grades I or II acute GVHD were not included in this cohort.

A third study cohort (cohort C, **Table 1**) consisted of 595 bone marrow recipients and their HLA-identical sibling donors. These individuals were included in a whole-genome scanning study of approximately 1,500 unrelated donor-recipient and HLA-identical sibling pairs randomly selected from patients who had HSCT with myeloablative conditioning regimens at Fred Hutchinson Cancer Research Center (FHCRC) from 1992 through 2004 for treatment of a hematological malignancy or myelodysplasia. Analysis was retrospective. Lack of available DNA was the only reason for exclusion from the study cohort. This human subjects research was performed according to protocols approved by the institutional review board of the FHCRC.

Grading of GVHD. Acute GVHD was diagnosed and graded according to standard criteria in use at the time of documented patient care. For the purpose of the current study, affected individuals were defined as those with grades II–IV acute GVHD; unaffected individuals were those with grades 0–I acute GVHD. Diagnostic sensitivity in the grading of GVHD can differ among clinical institutions[29], although it is generally well harmonized within institutions. Note that the criteria for selection of patients (which were defined before genetic analysis, and which for cohorts B and C reflected the design of other studies) differed from institution to institution; for example, cohort C used a cohort design, whereas cohort B used a case-control design and oversampled patients with severe (grades III–IV) acute GVHD.

Genotyping of deletion polymorphisms. Quantitative PCR assays were developed for typing each deletion polymorphism (**Supplementary Note** and **Supplementary Table 3**). Each internally controlled, two-color fluorescence assay allowed the individuals in a cohort to be assigned to three clear genotype classes, consisting of individuals with 0, 1 or 2 gene copies. To ensure the quality of gene-deletion genotypes, we verified that (i) membership in the three genotype classes (corresponding to 0, 1, 2 copies) showed Hardy-Weinberg equilibrium, and (ii) regression of patient genotypes against the genotypes of their sibling donors yielded a regression coefficient that was not significantly different from the expected value of 0.5. The accuracy of these assays was further evaluated by using them to genotype the gene deletions in the 270 HapMap samples[31]; these results showed 99.7% concordance with results from an independent experimental approach (use of the Affymetrix SNP 6.0 array[17], with analysis by the Canary algorithm in Birdsuite[32]). These assays were used for the initial screen in cohort A and the replication analysis in cohort B.

Genotyping of the *UGT2B17* gene deletion in cohort C used data from the Affymetrix SNP 5.0 array[17], which is being used for an ongoing genome-wide association study of transplant outcomes at the FHCRC. Copy-number probes spanning the deleted segment containing *UGT2B17* were identified[17], and the intensity measurements from these probes were summarized into a single measurement for each deletion polymorphism in each patient (**Supplementary Note**). These measurements showed a trimodal distribution identifying individuals with 0, 1 or 2 copies of each locus (**Supplementary Note**).

Determination of mismatches. Transplants were determined to involve a donor-recipient mismatch for a gene deletion if the donor had a homozygous deletion for that gene (0 copies) and the recipient had 1 or 2 gene copies.

Statistical analysis. The association of donor-recipient mismatch with case-control status within cohort A was evaluated with the use of a Cochran-Mantel-Haenszel test to combine data from the two subcohorts (**Supplementary Table 1**). Separate analyses of cohort B and of cohort C used a χ^2 test. All analyses of multiple study cohorts used a Cochran-Mantel-Haenszel test. The hypothesis tested was directional in two important dimensions: 'mismatch' was defined in the direction prescribed by the antigenicity model (homozygous deletion of a gene in the HSC donor but not the recipient); the direction of effect was also prescribed by this model (mismatch associated with increased risk). The directionality of this hypothesis is not typical of genome-wide association studies. Nonetheless, given the typical use of two-sided hypothesis tests in genome-wide association studies, we report the results of both one- and two-sided statistical tests in **Supplementary Table 3**.

ELISA. ELISA used an approach similar to earlier studies of alloantibodies to Y antigens[25,26]. Overlapping 20-mer peptides (with overlaps of 10 amino acids) across UGT2B17 and UGT2B28 were synthesized. To prepare each ELISA plate, 0.5 μg of each peptide was dissolved in coating buffer (BioFX), added to pretreated ELISA plates (Evergreen), and incubated overnight at 4 °C. Wells were blocked with BSA blocking buffer (BioFX) for 3 h at 37 °C. Stored serum samples from patients diagnosed with GVHD (generally obtained 6–18 months after transplantation) were used. Patient serum (1:1,000) was incubated in each well (1 h at 37 °C). To assay patient sera for antibodies, goat anti-human IgG conjugated to alkaline phosphatase (Abcam) (1:1,000) was incubated for 1 h at 37 °C. Plates were washed with PBS wash solution (BioFX), and the colorimetric substrate *p*-nitrophenyl phosphate (Chemicon) was added and allowed to develop at 20–25 °C for 20 min. Plates were read at 410 nm. Each serum sample was used to screen the peptide library on its own 96-well plate. (Several highly charged peptides yielded signal in all patient sera and were discarded.) A peptide was scored positive if the test result exceeded the median of the signals for all other peptides by at least 6 s.d. Positive results were retested in three replicates and considered confirmed only if they scored positive in retests both (i) relative to other peptides for the same serum sample and (ii) relative to other serum samples for the same peptide.

30. Nichols, W.C. *et al.* Polymorphism of adhesion molecule CD31 is not a significant risk factor for graft-versus-host disease. *Blood* **88**, 4429–4434 (1996).
31. International HapMap Consortium. A haplotype map of the human genome. *Nature* **327**, 1299–1320 (2005).
32. Korn, J.M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.* **40**, 1253–1260 (2008).