DR. SANDY FENG (Orcid ID : 0000-0002-2601-4350)

DR. GEORGE MAZARIEGOS (Orcid ID : 0000-0002-2624-8632)

DR. RYAN HIMES (Orcid ID : 0000-0002-0116-2613)

DR. MERCEDES MARTINEZ (Orcid ID : 0000-0001-6128-0155)

DR. ELIZABETH B. RAND (Orcid ID : 0000-0002-1173-4558)

DR. MICHELLE A WOOD-TRAGESER (Orcid ID : 0000-0002-1880-4926)

Article type : Original

Efficacy and Safety of Immunosuppression Withdrawal in Pediatric Liver Transplant Recipients: Moving Towards Personalized Management

Sandy Feng, M.D., Ph.D.[1], John C. Bucuvalas, M.D.[2], George V. Mazariegos, M.D.[3], John C. Magee, M.D.[4], Alberto Sanchez-Fueyo, M.D.[5], Katharine M. Spain, M.S.[6], Andrew Lesniak[7], Sai Kanaparthi, M.S.[8], Emily Perito, M.D.[9], Veena L. Venkat, M.D.[10], Bryna E. Burrell, Ph.D.[8], Estella M. Alonso M.D.[11], Nancy D. Bridges, M.D.[12], Edward Doo, M.D.[13], Nitika A. Gupta, M.D.[14], Ryan W. Himes, M.D.[15], David Ikle, Ph.D.[6], Annette M. Jackson, Ph.D.[16], Steven J. Lobritto, M.D.[17],

Juan Jose Lozano, Ph.D.[18], Mercedes Martinez, M.D.[17], Vicky L. Ng, M.D.[19] Elizabeth B. Rand, M.D.[20], Averell H. Sherker, M.D.[13], Shikha S. Sundaram, M.D.[21], Yumirle P. Turmelle, M.D.[22], Michele Wood-Trageser, Ph.D.[7], and Anthony J. Demetris, M.D.[7]

Author Affiliations

[1] Division of Transplantation, Department of Surgery, University of California San Francisco, San Francisco, CA

[2] Mount Sinai Kravis Children's Hospital and Recanati/Miller Transplantation Institute, Mount Sinai Health System, New York, NY

[3] Hillman Center for Pediatric Transplantation, Children's Hospital of Pittsburgh, Pittsburgh, PA

[4] Section of Transplant Surgery, Department of Surgery, University of Michigan, Ann Arbor, MI

[5] Institute of Liver Studies, King's College London, London, United Kingdom

[6] Rho, Inc., Chapel Hill, NC

[7] Department of Pathology, University of Pittsburgh, Pittsburgh, PA

[8] The Immune Tolerance Network, Bethesda, MD

[9] Division of Pediatric Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, University of California San Francisco, San Francisco, CA

[10] Division of Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh, Pittsburgh PA

[11] Siragusa Transplantation Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL

[12] Transplantation Branch, Division of Allergy, Immunology, and Transplantation, National Institute of Allergy and Infectious Diseases, Rockville, MD

[13] Division of Digestive Diseases and Nutrition, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD

[14] Division of Pediatric Gastroenterology, Hepatology and Nutrition, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA

[15] Section of Gastroenterology, Hepatology, and Nutrition, Texas Children's Hospital, Houston, TX

[16] Department of Surgery, Duke University, Durham, NC

[17] Center for Liver Diseases and Transplantation, Department of Surgery, Columbia University Irving Medical Center, New York, NY

[18] Bioinformatic Platform, Biomedical Research Center in Hepatic and Digestive Diseases (CIBEREHD), Instituto de Salud Carlos III, Barcelona, Spain.

[19] Division of Pediatric Gastroenterology, Hepatology and Nutrition, Transplant and Regenerative Medicine Center, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada

[20] Liver Transplant Program, The Children's Hospital of Pennsylvania, Philadelphia, PA

[21] Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital Colorado, University of Colorado School of Medicine, Aurora, CO

[22] Division of Gastroenterology, Hepatology, and Nutrition, Washington University School of Medicine, St. Louis, MO

Key words:

Operational tolerance

Drug minimization

Donor specific antibody

Interface activity

Corresponding Author:

Sandy Feng, MD, PhD

Professor of Surgery in Residence

Vice Chair of Research, Department of Surgery

Director, Abdominal Transplant Fellowship

University of California San Francisco

505 Parnassus Avenue

San Francisco, CA 94143-0780

Email: sandy.feng@ucsf.edu

Telephone: (415) 353-8725

Fax: (415) 353 8709

Abbreviations

ISW immunosuppression withdrawal

DSA donor specific antibody

GFR glomerular filtration rate

GSEA gene set enrichment analysis

TCMR T-cell mediated rejection

Financial Support

This research was supported by U01-AI-100807 awarded by the NIAID and NIDDK as well as by the Immune Tolerance Network (UM1AI109565), an international clinical research consortium headquartered at the Benaroya Research Institute and supported by NIAID.

ABSTRACT

Background and Aims

Tolerance is transplantation's holy grail as it denotes allograft health without immunosuppression and its toxicities. Our aim was to determine, among stable long-term pediatric liver transplant recipients, the efficacy and safety of immunosuppression withdrawal to identify operational tolerance.

Approach and Results

We conducted a multi-center, single-arm trial of immunosuppression withdrawal over 36-48 weeks. Liver tests were monitored biweekly (year 1), monthly (year 2) and bimonthly (years 3-4). For-cause biopsies were done at investigators' discretion but mandated when alanine aminotransferase or gamma glutamyl transferase exceeded 100U/L. All subjects underwent final liver biopsy at trial end. The primary efficacy endpoint was operational tolerance, defined by strict biochemical and histological criteria 1 year after stopping immunosuppression. Among 88 subjects (median age 11 years; 39 boys; 57 deceased donor grafts), 33 (37.5%; 95%CI 27.4%, 48.5%) were operationally tolerant, 16 were non-tolerant by histology (met biochemical but failed histological criteria) and 39 were non-tolerant by rejection. Rejection, predicted by subtle liver inflammation in trial entry biopsies, typically (n=32) occurred at $\leq$32% of the trial entry immunosuppression dose and was treated with corticosteroids (n=32) and/or tacrolimus (n=38) with resolution (liver tests within 1.5X baseline) for all but 1 subject. No death, graft loss, or chronic,

severe, or refractory rejection occurred. Neither fibrosis stage nor the expression level of a rejection gene set increased over 4 years for either tolerant or non-tolerant subjects.

Conclusions

Immunosuppression withdrawal showed that 37.5% of selected pediatric liver transplant recipients were operationally tolerant. Allograft histology did not deteriorate for either tolerant or non-tolerant subjects. The timing and reversibility of failed withdrawal justifies future trials exploring the efficacy, safety, and potential benefits of immunosuppression minimization.

Trial Registration: ClinicalTrials.gov NCT01638559

INTRODUCTION

For children with liver transplants, the primary barriers to optimal allograft and patient health are chronic allo-immune graft injury (1-8) and the cumulative toxicities of immunosuppression (9-12). The perception that the liver is "tolerogenic" has spawned intense interest in mitigating the cumulative toxicities of immunosuppression by reducing dosage or discontinuing immunosuppression. Multiple single-center reports have described recipients who, off immunosuppression, have apparently maintained normal allograft function, (summarized in (13, 14). However, studies showing a high prevalence of graft injury in patients with normal liver tests on standard-of-care immunosuppression (1-8) and those refuting the benign nature of rejection (15-18) have raised concerns regarding the wisdom of immunosuppression withdrawal (ISW). These concerns have been partly assuaged by two adult and one pediatric multi-center ISW trials (19-21). With respect to efficacy, only the adult trials estimated the prevalence of operational tolerance. With respect to safety, clinical and histological follow-up of operationally tolerant subjects was limited to 1 and 3 years in the two adult trials and extended to 5 years in the pilot pediatric trial. Neither allograft inflammation nor fibrosis was reported to increase. However, no trials afforded histologic follow-up to those who failed ISW. It remains unknown whether rejection that occurred during ISW resulted in histological sequelae. Thus, the safety of attempted ISW has not been fully elucidated.

We report on a multi-center trial conducted at 12 North American transplant centers to determine the efficacy and safety of ISW in children with stable, long-term liver transplants. Our primary objective was to ascertain the prevalence of operational tolerance with a sufficiently narrow confidence interval to guide

clinicians and patients. A critical secondary objective was the safety of attempted ISW. iWITH mandated that all subjects who initiated ISW, irrespective of outcome, undergo the same duration of follow-up with liver tests, donor-specific antibody (DSA) testing, and liver biopsy. Finally, as an exploratory objective, we aimed to identify predictors of operational tolerance. Elucidating conditions permissive of immunosuppression dose reduction and delineating appropriate monitoring thereafter, inclusive of histological evaluation, may free children with liver transplants from the current impossible dichotomy of "too little" or "too much" immunosuppression.

METHODS

Trial design and subjects

iWITH, "Immunosuppression Withdrawal for Stable Pediatric Liver Transplant Recipients" (NCT01638559) was a prospective, single-arm, multi-center trial to determine the efficacy and safety of ISW in pediatric liver transplant recipients. Each participant and his/her guardian provided informed assent and consent. The trial protocol and statistical analysis plan are submitted as supplementary material.

Inclusion/exclusion criteria

Inclusion and exclusion criteria are detailed in the trial protocol (Supplement); key criteria were:

Inclusion

- Liver transplant recipient at ≤6 years of age;
- ≥4 years after transplant and <18 years at enrollment;
- ALT and GGT consistently <50U/L;
- No acute or chronic rejection within 2 years;
- On calcineurin inhibitor monotherapy for the preceding year without ≥50% dose increase during the preceding 6 months
- Eligible screening biopsy per central pathology (Table S1) (22)

Exclusion

- Transplant secondary to autoimmune etiologies, HBV or HCV;
- Recipient of second organ transplant before, simultaneous to, or after liver transplant;
- Calculated glomerular filtration rate (GFR; modified Schwartz formula) <60 mL/min/1.73m^2.

No subjects participated in a previous ISW trial.

Trial procedures and endpoints

Immunosuppression was reduced in 7 steps of 4 or 6 week duration. The total daily dose was reduced by 25% (steps 1 and 2) followed by the number of dosing days per week (steps 3 through 7) (Figure S1). Step progression could be paused for $\leq$ 4 weeks; each subject was limited to 3 pauses so ISW ranged from 36 to 48 weeks. Liver tests were checked every 2 weeks during ISW and for an additional 12 weeks thereafter. For-cause liver biopsies were performed at investigators' discretion but mandated if ALT or GGT exceeded 100U/L. Local pathology assessment guided clinical decision-making including immunosuppression management; central pathology assessment was utilized for data analysis.

Subjects completing ISW with stable liver tests and without rejection were evaluated 1 year after the last dose for operational tolerance using the primary endpoint criteria of ALT and GGT <50U/L and liver biopsy with no more than minimal change compared to the eligibility biopsy (22) (Table S1). All subjects were required to undergo biopsy at trial end. The primary safety endpoint was absence of histologically severe rejection, refractory rejection (rejection requiring treatment with a lymphocyte depleting drug), chronic rejection, allograft loss, or death. Secondary safety endpoints included histological grade, clinical severity, and time to resolution of acute rejection, overall immunosuppression exposure, and fibrosis progression.

Safety monitoring

Site investigators reported adverse events through 30 days after trial completion. Investigators graded severity according to Common Terminology Criteria for Adverse Events (version 4.03) and assessed seriousness and relatedness to trial procedures (phlebotomy, biopsy, and ISW). Severe adverse events were reviewed by a NIAID medical monitor who determined final severity and attribution assessments.

Mechanistic studies

Detailed methods for mechanistic studies described below are provided in the Supplementary Methods. HLA and eplet mismatch were determined for donor-recipient pairs. Serum specimens were tested for autoantibodies, quantitative IgG, and class II DSA (3).

Formalin-fixed liver biopsies were scored for inflammation and fibrosis. Batched slide sets underwent multiplex immunohistochemistry for (i) leukocytes (CD45+), antigen-presenting cells (MHCII+), and endothelial cells (CD34+); ii) T cells (CD3+) and recent infiltrating monocytes/macrophages (MAC387+); and (iii) T cell subsets (CD4+, CD8+, T-box protein expressed in T cells+ and forkhead box protein P3+. Fully automated tissue-tethered cytometry was performed using image analysis software and applied via automated batch processing without human intervention.

We conducted tissue transcriptional profiling of 303 cryopreserved liver tissue samples employing Affymetrix U219 microarrays and validated results using the NanoString nCounter platform. To assess the probability of T cell mediated rejection (TCMR) based on changes in microarray gene expression, we analyzed transcript levels of a previously published 12-gene TCMR signature (23). To assess over-representation of biological pathways in the microarray dataset, we employed gene set enrichment analysis (GSEA) and three publicly available gene sets. Finally, real-time PCR experiments assessed the transcript levels of a 5-gene predictor of ISW success for adult liver transplant recipients (24).

Statistical analyses

The sample size was based on an estimated operational tolerance rate of 35% and a specified 95% CI half-width of 10%. Categorical and continuous variables were compared using Fisher exact and two-sample t-tests. Logistic regression analyses were used to identify associations between operational tolerance and clinical, serological, histological, immunohistochemical, and transcriptional parameters. Statistical analyses were performed using SAS, version 9.4 (SAS Institute). All data and analyses are available on the Immune Tolerance Network TrialShare analysis portal (https://www.itntrialshare.org/iWITH_primary.url).

Trial oversight

The first two authors designed the trial in collaboration with NIAID, NIDDK, Immune Tolerance Network, and iWITH investigators. The trial was conducted in accordance with the Declaration of Helsinki. Adverse events were reviewed by the NIAID medical monitor and data and safety monitoring board**.** Clinical data and mechanistic data were submitted to and analyzed by a central data coordinating center and by the Immune Tolerance Network. The manuscript was drafted by the first author and edited and reviewed by

all. All authors confirm that the trial was conducted according to the protocol and that the data and analyses presented are accurate and complete.

RESULTS

Patient cohort

Among 2,909 children with liver transplants followed at 12 centers (Figure 1), 1,731 were excluded for transplant within <4 years ago (n=963), at age ≥7 years (n=507), a second liver or other transplant (n=186), or transplant for viral or auto-immune disease (n=75). After additional exclusions (n=823), 355 patients remained: among 276 approached, 161 patients and guardians provided assent and informed consent between August 2012 and June 2014. Final assessment excluded 73 subjects, 69 secondary to histopathology and 4 secondary to abnormal liver tests. Key histological features of the 88 eligible subjects are shown in Figure S2.

Primary endpoint

Among the 88 subjects who initiated ISW, 33 subjects (37.5%; 95% CI: 27.4%-48.5%) were operationally tolerant, meeting biochemical and histological criteria (Figure 2A). The remaining 55 were non-tolerant (Figure 2B). Table 1 shows comparisons of operationally tolerant and non-tolerant subjects. All subjects were assessed for the primary endpoint; 3 did not complete the trial.

Operationally tolerant subjects

For the 33 operationally tolerant subjects, Figure 3A shows baseline, peak, and final ALT and GGT over 4 years: 4% of ALT and 2% of GGT values exceeded 50U/L (Figure S3A). Serial DSA testing demonstrated that 12 (36%) never had, 15 (45%) had at trial entry, and 5 (15%) developed class II DSA (Figure S4A). Biopsies at trial start and end (4-year separation) were compared for inflammation (Figure S5A) and fibrosis (Figure 3B). Liver allograft fibrosis scores (LAFSc)(25) were unchanged or minimally changed ($-1 \leq \Delta \leq 1$) for 25/32, improved for 5/32 (2 with -3 and 3 with -2 change), and worsened for 2/32 (1 each with +2 and +3 change).

Consistent with histological observations, TCMR probability scores were unchanged, comparing trial end to start biopsies (4-year separation; Figure 3C). Consistency was confirmed at the gene pathway level by whole-transcriptome pair-wise comparisons of the 3 longitudinal, protocol-directed biopsies. Pro-

inflammatory pathway enrichment scores did not increase and, in some cases, decreased after immunosuppression discontinuation (Tables S5 and S6).

Non-tolerant subjects by rejection

Among 55 non-tolerant subjects, 39 were non-tolerant by rejection (Figure 2B) and 16 were non-tolerant by histology (following section). There were 37 biopsy-proven and 2 clinical rejection episodes, defined per protocol (Figure 4A). Rejection was diagnosed during ISW (n=33), after ISW but before tolerance adjudication (n=2), or at the time of tolerance adjudication (n=4). Most episodes (33/39; 85%) occurred at ≤32% of the entry immunosuppression dose. Histological severity (26) of the 37 biopsy-proven episodes was predominantly indeterminate (n=16) or mild (n=17). Median (IQR) peak ALT was 136 (101-205)U/L and peak GGT was 63 (42-104)U/L (Figure 4B). During the 4-year trial, 12% of ALT and 9% of GGT values exceeded 50U/L (Figure S3B).

Rejection was treated with corticosteroids in 32 of 39 subjects, with a median (IQR) total dose of 34.7 (16.4-50.7)mg/kg over 63 (42-121)days. Nearly all subjects (38/39) were also treated with reinitiation or increased tacrolimus dosing. In 5 subjects, azathioprine or mycophenolate was added. No rejection episodes were steroid-refractory. For 4 subjects with stable ALT and GGT but whose tolerance adjudication biopsy showed rejection, biochemical resolution could not be assessed. For the remaining 35, biochemical resolution, defined per protocol as ALT and GGT <1.5X baseline, occurred for 34 in a median (IQR) of 13 (7.1-19.1)weeks (Figure 4C); 1 subject did not resolve. All 35 subjects achieved ALT and GGT <50U/L in a median (IQR) of 5.0 (3.3-12.3)weeks.

At trial end, 36 of 39 subjects who rejected were on monotherapy. For those on tacrolimus (n=35), the end-of-trial compared to entry tacrolimus dose was lower, same, and higher for 46%, 26%, and 29%, respectively. Over 4 years, 50% of subjects received less total tacrolimus exposure than if they had been maintained on their entry dose (Figure 4D).

Serial DSA testing demonstrated that 6 (15%) never had, 21 (54%) had pre-existing, and 12 (31%) developed class II DSA (Figure S4B). Biopsies at trial start and end (4-year separation) were compared for

inflammation (Figure S5B) and fibrosis (Figure 4E). LAFSc (25) was unchanged or minimally changed ($-1 \leq \Delta \leq 1$) for 17/37, improved for 14/37 (5 with -3 and 9 with -2 change) and worsened for 6/37 (4 with +2 and 2 with +3 change).

Longitudinal transcriptional profiling revealed a significant increase in rejection-related transcripts at rejection diagnosis (Figure 4F), with the majority (85%; 23/27) of tested samples classified as rejection by the TCMR score. Following treatment, TCMR transcriptional changes resolved. Resolution was confirmed at the functional pathway level using GSEA which showed no differences in either rejection or inflammatory-related transcriptional pathways between trial start and end (4-year separation) biopsies (Tables S5 and S6).

Non-tolerant subjects by histology

Among the 55 non-tolerant subjects, 16 subjects met biochemical but not histological criteria for operational tolerance (Table S1; Figure 2B). The primary criterion for non-tolerant designation was interface hepatitis (n=15); bile duct damage and isolated arteriopathy without other chronic rejection features developed in 1 subject each (Figure 5A). Baseline, peak, and final ALT and GGT are shown in Figure 5B; 2% of ALT and 3% of GGT values collected during the 4-year trial exceeded 50U/L (Figure S3C).

At the discretion of the subjects' physicians, 8 reinitiated and 8 remained off immunosuppression (Supplementary Methods). The latter underwent an additional biopsy 6 to 20 months after tolerance adjudication, prompting immunosuppression re-initiation in 2 subjects (carets in Figures 5A and 5D). The other 6 subjects remained off immunosuppression through trial end. Over 4 years, 12 of 13 received less tacrolimus than if they were maintained on their entry dose (Figure 5C).

Serial DSA testing showed that, of 16 non-tolerant by biopsy subjects, 2 never had, 8 had pre-existing, and 6 developed DSA (Figure S4C). Biopsies at trial start and end (4-year separation) were compared for inflammation (Figure S5C) and fibrosis (Figure 5D). LAFSc (25) was unchanged or minimally changed ($-1 \leq \Delta \leq 1$) for 10/15, improved for 3/15 (1 with -3 and 2 with -2 change) and worsened for 2/15 (1 each +2 and +4 change). The single subject whose fibrosis score increased by +4 was diagnosed with and required treatment for biliary stricture.

Comparison of biopsies performed at trial entry to tolerance adjudication (2-year separation) showed increased expression of TCMR genes (Figures 5E and S6) which did not reach statistical significance. Only 5 of 14 tested adjudication biopsies exhibited a TCMR probability above the suggested rejection threshold (Figure 5E). Furthermore, whole-genome pair-wise analysis of tolerance adjudication compared to trial entry biopsies identified 149 and 107 over- and under-expressed (false discovery rate <0.05) genes. However, none are known to be involved in allograft rejection (https://www.itntrialshare.org/iWITH_primary.url). Similarly, GSEA did not reveal enrichment in pro-inflammatory pathways (Tables S5 and S6).

In contrast, comparison of tolerance adjudication to end-of-trial biopsies (2-year separation) showed significantly decreased TCMR signature gene expression levels (Figure 5E). This decrease was observed predominantly in subjects who re-initiated immunosuppression (Figures S6A and S6B); subjects kept off immunosuppression remained stable. Comparison of trial end to start biopsies (4-year separation) demonstrated no change in either the TCMR score (Figure 5E) or in rejection-associated molecular pathways (Tables S5 and S6).

Safety of ISW

No death, graft loss, or chronic, refractory, or severe rejection occurred. Among 1,023 non-rejection adverse events, 47 were possibly or definitely related trial procedures: phlebotomy (n=4), ISW (n=21), or liver biopsy (n=22; 4 serious: 1 episode each of cholangitis, bile leak, abdominal pain, and skin infection) (Tables S2 and S3). Operationally tolerant and non-tolerant subjects did not differ in the frequency of serious adverse events. Immunosuppression escalation to treat rejection did not increase infectious events (data not shown). Tolerant and non-tolerant subjects did not differ in calculated GFR at either trial start or end. Moreover, they did not differ in the 4-year change in calculated GFR (Table S4). Longitudinal biopsies over 4 years from subjects with *de novo* class II DSA or those with DSA at trial entry did not exhibit increased fibrosis change compared to those who remained DSA-free (Figure S7).

Histological and immunohistochemical but not transcriptional parameters of the entry biopsy predicted ISW outcome.

At trial entry, no clinical or serological variables predicted operational tolerance, including age at or time after transplant, living versus deceased donor, allele or eplet mismatch, and class II DSA status (Table 2). However, prospectively scored histological and immunohistochemical features of the eligibility biopsy were associated with operational tolerance. Operationally tolerant subjects more frequently had no portal inflammation and decreased leukocytes (CD45+), antigen-presenting cells (MHCII+), leukocyte/antigen-presenting cell pairs, infiltrating monocytes/macrophages (MAC387+), and effector T cells (CD8+)(Table 2). Plotting 3 immunohistochemical parameters shows clustering of operationally tolerant subjects (Figure 6A). As a predictor of operational tolerance, immunohistochemical clustering offered sensitivity and specificity of 94% and 66%, respectively, while portal inflammation offered sensitivity and specificity of 70% and 55%, respectively. Differential performance is illustrated when eligibility biopsies with similar portal (and lobular) inflammation grades but different immunohistochemical inflammatory loads are juxtaposed (Figure 6B).

At trial entry, whole-genome microarray tissue transcriptional profiling did not reveal differences between tolerant and non-tolerant subjects: no genes showed significant differential expression at a false discovery rate of <10% (https://www.itntrialshare.org/iWITH_primary.url). Similarly, expression level of neither the TCMR gene set nor a 5-gene tolerance classifier, derived from a European adult ISW trial (24), predicted operational tolerance (Table 2).

The development of *de novo* DSA was associated with ISW failure

In addition to trial entry DSA status, we analyzed DSA development during ISW (year 0 to 1). Among subjects who never exhibited DSA, 64% (14/22) were operationally tolerant (Figure 6C), compared to 34% (15/44) for those with DSA at entry and 15% (3/20) for those who developed DSA. Compared to subjects who never exhibited DSA, subjects with DSA at entry and those who developed DSA were less likely to be operationally tolerant [30% (95% CI 10%-86%) and 10% (95% CI 2%-45%), respectively; (Table 2)].

DISCUSSION

iWITH has shown that, among selected children with liver transplants, 37.5% were operationally tolerant. The 10% CI width on either side of this point estimate is sufficiently narrow to guide clinicians and patients. ISW failure typically occurred after substantial dose reduction, suggesting that modest immunosuppression dose reduction might have been possible and safe for non-tolerant subjects.

Rejection episodes were, with few exceptions, histologically mild, biochemically reversed by treatment and, most importantly, not associated with histological sequelae over 4 years.

The threat of life long immunosuppression is particularly potent for children for whom liver transplantation is expected to secure decades of healthy and productive life (10, 27). Two-thirds of late mortality is directly attributed to immunosuppression complications such as infection or malignancy (27, 28). A population-based study from 4 Nordic countries reported that children with liver transplants, experience nearly 10-fold higher standardized incidence ratio for all cancers (9). The cumulative incidence of cancer rises steeply in young adulthood, increasing from 2% to 6% and 22% at 10, 20, and 25 years, respectively, after transplant. The dramatic change in slope confirms suspicions that morbidity and mortality imposed on children by cumulative immunosuppression exposure manifests ***decades*** after transplant, a timeline well beyond that of a clinical trial. iWITH collected and analyzed longitudinal data relevant to the toxicities of immunosuppression. Successful ISW did not yield perceptible benefit; failed ISW did not result in perceptible harm.

iWITH targeted the prevalence of operational tolerance as the primary endpoint and was powered to provide a prevalence estimate useful to clinicians and patients. iWITH uniquely utilized an active and rigorous definition of operational tolerance, requiring no or minimal change in ALT, GGT, and liver biopsy 1 year after stopping immunosuppression, compared to trial entry 2 years earlier. Any subject failing to meet criteria for operational tolerance was deemed non-tolerant. The validity of histological assessment to adjudicate operational tolerance was corroborated by tissue transcriptional data; biopsies that met operational tolerance criteria had low TCMR probability scores, consistent with immunologic quiescence. In contrast to iWITH's strict criteria, most previous single-center studies considered subjects to be operationally tolerant if liver tests did not escalate and rejection had not occurred. An American multi-center adult trial similarly adjudicated operational tolerance according to clinical and biochemical criteria (21). The European multi-center adult trial included histological assessment (19) but criteria were lax, stipulating absence of rejection rather than a comparison with pre-withdrawal biopsies. The high prevalence of silent allograft injury with histopathological findings non-diagnostic of acute rejection but reflective of an allo-immune response (3, 29) indicates that operational tolerance must be actively determined, with inclusion of rigorous histological assessment.

Our approach to adjudicate operational tolerance led to the identification of a phenotype not previously reported. Twenty of 88 subjects were “biochemically tolerant” but “histologically non-tolerant”, a group almost certainly characterized by previous trials as “tolerant”. Compared to trial entry biopsies, tolerance adjudication biopsies showed distinctly but modestly increased TCMR gene expression, supporting our hypothesis that the histopathological features driving the non-tolerant designation reflected mild T-cell mediated graft injury (3, 23, 29). The transcriptional probability of rejection for the non-tolerant by biopsy subjects strongly mirrored the levels observed for a previously described cluster of iWITH eligibility biopsies (3) that was deemed ineligible for ISW.

Although iWITH’s primary objective was efficacy, many of iWITH’s secondary objectives related to safety. Histologically severe, steroid-refractory, or chronic rejection along with graft loss or patient death did not occur. However, iWITH’s design enabled us to address several novel and subtle metrics of safety. First, iWITH mandated histological follow-up of those who rejected. Previous ISW trials followed those who rejected with laboratory assessments alone, inadequate in the context of subclinical graft injury and fibrosis progression. Standard histological assessment confirmed that, at trial end, liver allografts of those who rejected did not exhibit changes over time in either inflammatory or fibrosis parameters. This interpretation is strengthened by two considerations: i) subjects entered iWITH with nearly pristine biopsies such that, for many histological parameters, only deterioration was possible; ii) tissue transcriptional profiles at trial start and end were comparable. As such, we suggest that rejection precipitated by monitored ISW, promptly diagnosed and treated, did not compromise mid-term allograft health. Although nearly all subjects who rejected were exposed to corticosteroids and approximately half were exposed to additional tacrolimus, we did not identify deterioration of either renal or infectious parameters during the 4-year trial.

In addition to assessing those who rejected, iWITH rigorously assessed allograft health of the other two cohorts, operationally tolerant and non-tolerant by biopsy cohorts. Reassuringly, inflammation and fibrosis parameters did not change over the 4-year trial. Histological stability was confirmed by tissue transcriptional stability in TCMR and immune activation genes.

A final critical metric of safety that we assessed was the impact of pre-existing and *de novo* DSA. The literature is replete with evidence that lowering and/or discontinuing immunosuppression after solid

organ and cellular transplants engenders a humoral allo-immune response (20, 30-32). Although *de novo* DSA is associated with injury and diminished graft survival for non-liver allografts (33-35), a negative impact on liver allograft health is less well-established (36-41). The relative resistance and resiliency of the liver allograft to antibody-mediated injury almost certainly reflects unique, liver-specific, innate and adaptive mechanisms that attenuate potential immunologic insults (42). Our findings that a substantial number of subjects with class II DSA had healthy allografts at trial entry which remained healthy after attempted ISW, irrespective of outcome, provides reassurance that pre-existing or *de novo* DSA does not portend inevitable or aggressive structural deterioration (20, 38). The histological stability almost certainly reflects the robust allograft health of this cohort at trial entry. iWITH excluded patients with any significant necro-inflammatory activity and/or fibrosis as these damaged allografts have up-regulated expression of microvascular, endothelial class II antigens which likely increases their vulnerability to class II DSA (37, 43).

iWITH specifically aimed to identify predictors of operational tolerance. No clinical, biochemical, or serological factors at trial entry, including time after transplant, living or deceased donor, or class II DSA status was associated with ISW outcome. Tissue transcriptional profiling and specifically, the biomarker associated with successful ISW in 2 adult ISW trials (24, 44), did not predict iWITH outcomes, possibly suggesting different mechanisms of operational tolerance in children and adults. However, differences in trial design, inclusion/exclusion criteria, and cohort demographics preclude robust conclusions. Currently, a European multi-center trial (NCT02498977), specifically designed to prospectively assess the biomarker's diagnostic accuracy is ongoing. Although perhaps counter-intuitive, iWITH, along with other trials, have consistently shown that DSA presence, in and of itself, should not preclude ISW and does not contradict operational tolerance (20, 21, 38). However, iWITH has uniquely shown that the development of DSA during ISW was associated with non-tolerance. A trial of ISW involving adults early after transplant reported that *de novo* DSA predicted acute rejection (32). These findings raise the question as to why, when immunosuppression is reduced, some subjects develop DSA while others do not. Understanding the mechanisms that determine whether DSA will emerge is necessary to inform rational approaches to therapy. We are currently utilizing longitudinal biospecimens collected during iWITH to address this critical issue.

While clinical, biochemical, and serological metrics did not predict operational tolerance, several histological and immunohistochemical parameters of the trial entry biopsy did. The negative impact of

prospectively scored portal inflammation emerged, despite excluding subjects whose grafts showed more than minimal or focal mild abnormalities. In a pilot trial, we similarly observed an inverse association between portal inflammation and operational tolerance (20). An adult center has also reported that increased numbers of CD8+ cells correlated with failed ISW (45). These observations recapitulate animal data where pre-existing inflammation promotes liver damage by facilitating effector T cell maturation (46, 47). While some investigators have speculated that allograft infiltrates may be regulatory in nature, facilitating tolerance (48, 49), our data indicate otherwise. ISW appeared to activate even the few scattered inflammatory cells and precipitate rejection. Multiplex immunohistochemistry analyses of allograft biopsies yielded quantitative, objective, and relational data and suggested thresholds of inflammatory cells prohibitive of successful ISW.

Although iWITH was rigorously executed, with crisply defined entry criteria, a detailed ISW algorithm, strict guidelines mandating for-cause biopsies, and comprehensive data and specimen collection, we acknowledge limitations. First, the strict eligibility criteria and moderate sample size limits generalizability to the overall pediatric liver transplant population with respect to immunosuppression reduction. Second, the non-randomized design and, as a result, the lack of a control group rendered it impossible to assess whether ISW yielded benefit. The one-arm design reflects regulatory and equipoise considerations that govern clinical trial participation for children. Finally, the 4-year follow-up, typical for a clinical trial, may be insufficient to support definitive long-term conclusions regarding either efficacy, safety, or durability of operational tolerance.

In conclusion, iWITH has shown that more than one third of selected pediatric liver transplant recipients, clinically stable on a single immunosuppression drug with normal allograft histology, are operationally tolerant. Just as importantly, we have shown that ISW, successful or unsuccessful, was safe according to clinical, biochemical, histological and transcriptional assessment over 4 years. Since withdrawal failure almost always occurred at one third or less of the baseline dose, some subjects may have been receiving more pharmacologic immunosuppression than necessary. Although DSA emerging as immunosuppression is reduced portends non-tolerance, it may not compromise a robustly healthy allograft. Intermittent histological evaluation is, however, critical. The development of interface activity and/or other histopathology, even with stable relative to baseline liver tests, should prompt consideration of stopping further dose reduction and perhaps even dose escalation. As such optimal long-term immunosuppression management requires intermittent histological assessment. Longitudinal decision-making embodies the

enormous challenge facing physicians caring for children with liver transplants with expectations of robust graft function and overall health for many decades. The insights gained from iWITH, the culmination of effort exerted over nearly a decade, should spur future interventional trials and mechanistic investigations that address the steep challenge of improving lifelong graft and patient outcomes.

ACKNOWLEDGEMENTS

This research was supported by U01-AI-100807 awarded by the NIAID and NIDDK as well as by the Immune Tolerance Network (UM1AI109565), an international clinical research consortium headquartered at the Benaroya Research Institute and supported by NIAID. The conduct of this trial at the some of the clinical sites was supported, in part, by awards from the National Center for Advancing Translational Sciences of the National Institutes of Health, through UL1 TR000004 and UL1 TR001872 at University of California San Francisco, 5UL1TR001425-04 at Cincinnati Children's Hospital, UL1TR000003 and UL1TR001878 at Children's Hospital of Philadelphia, UL1TR000454 and UL1TR002378 at Emory University School of Medicine, and UO1AI100807 at St Louis Children's Hospital. The content is solely the responsibility of the authors and does not represent the official views of the NIH.

A data sharing agreement is available with the full text of this article.

Drs. Demetris and Feng reported that they are consultants for Novartis Pharmaceuticals Corporation and Drs. Feng and Sanchez-Fueyo reported being a consultant of Quell Therapeutics. Dr. Jackson reported that she is a member of the Speaker Bureau for One Lambda. No other potential conflict of interest relevant to this article was reported. Disclosure forms provided by the authors are available with the full text of this article.

We are deeply grateful to the patients and their families who participated in the study. We thank Allison Priore, Sharon Blaschka, RN, FNP and Andre Hawkins for their extensive contributions to the management and conduct of the trial and Kalpana Harish, M.P.H. for expert grants administration. We thank all of the co-investigators and research coordinators. Finally, we are indebted to Deborah Phippard, Ph.D. and Allan Kirk, M.D. for wise guidance, particularly during the planning phase of the trial.

REFERENCES

1. Ekong UD, Melin-Aldana H, Seshadri R, Lokar J, Harris D, Whitington PF, Alonso EM. Graft histology characteristics in long-term survivors of pediatric liver transplantation. Liver Transpl 2008;14:1582-1587.

2. Evans HM, Kelly DA, McKiernan PJ, Hubscher S. Progressive histological damage in liver allografts following pediatric liver transplantation. Hepatology 2006;43:1109-1117.

3. Feng S, Bucuvalas JC, Demetris AJ, Burrell BE, Spain KM, Kanaparthi S, Magee JC, et al. Evidence of Chronic Allograft Injury in Liver Biopsies From Long-term Pediatric Recipients of Liver Transplants. Gastroenterology 2018;155:1838-1851 e1837.

4. Guerra MR, Naini BV, Scapa JV, Reed EF, Busuttil RW, Cheng EY, Farmer DG, et al. Obliterative portal venopathy: A histopathologic finding associated with chronic antibody-mediated rejection in pediatric liver allografts. Pediatr Transplant 2018;22.

5. Herzog D, Soglio DB, Fournet JC, Martin S, Marleau D, Alvarez F. Interface hepatitis is associated with a high incidence of late graft fibrosis in a group of tightly monitored pediatric orthotopic liver transplantation patients. Liver Transpl 2008;14:946-955.

6. Miyagawa-Hayashino A, Yoshizawa A, Uchida Y, Egawa H, Yurugi K, Masuda S, Minamiguchi S, et al. Progressive graft fibrosis and donor-specific human leukocyte antigen antibodies in pediatric late liver allografts. Liver Transpl 2012;18:1333-1342.

7. Scheenstra R, Peeters PM, Verkade HJ, Gouw AS. Graft fibrosis after pediatric liver transplantation: ten years of follow-up. Hepatology 2009;49:880-886.

8. Wozniak LJ, Hickey MJ, Venick RS, Vargas JH, Farmer DG, Busuttil RW, McDiarmid SV, et al. Donor-specific HLA Antibodies Are Associated With Late Allograft Dysfunction After Pediatric Liver Transplantation. Transplantation 2015;99:1416-1422.

9. Aberg F, Isoniemi H, Pukkala E, Jalanko H, Rasmussen A, Storm HH, Schultz N, et al. Cancer After Liver Transplantation in Children and Young Adults: A Population-Based Study From 4 Nordic Countries. Liver Transpl 2018;24:1252-1259.

10. Martinelli J, Habes D, Majed L, Guettier C, Gonzales E, Linglart A, Larue C, et al. Long-term outcome of liver transplantation in childhood: A study of 20-year survivors. Am J Transplant 2018;18:1680-1689.

11. Memaran N, Borchert-Morlins B, Schmidt BMW, Sugianto RI, Wilke H, Blote R, Baumann U, et al. High Burden of Subclinical Cardiovascular Target Organ Damage After Pediatric Liver Transplantation. Liver Transpl 2019;25:752-762.

12. Ruebner RL, Reese PP, Denburg MR, Abt PL, Furth SL. End-stage kidney disease after pediatric nonrenal solid organ transplantation. Pediatrics 2013;132:e1319-1326.

13. Feng S, Bucuvalas J. Tolerance after liver transplantation: Where are we? Liver Transpl 2017;23:1601-1614.

14. Levitsky J, Feng S. Tolerance in clinical liver transplantation. Hum Immunol 2018;79:283-287.

15. Uemura T, Ikegami T, Sanchez EQ, Jennings LW, Narasimhan G, McKenna GJ, Randall HB, et al. Late acute rejection after liver transplantation impacts patient survival. Clin Transplant 2008;22:316-323.

16. Thurairajah PH, Carbone M, Bridgestock H, Thomas P, Hebbar S, Gunson BK, Shah T, et al. Late acute liver allograft rejection; a study of its natural history and graft survival in the current era. Transplantation 2013;95:955-959.

17. Levitsky J, Goldberg D, Smith AR, Mansfield SA, Gillespie BW, Merion RM, Lok AS, et al. Acute Rejection Increases Risk of Graft Failure and Death in Recent Liver Transplant Recipients. Clin Gastroenterol Hepatol 2017;15:584-593 e582.

18. Jadlowiec CC, Morgan PE, Nehra AK, Hathcock MA, Kremers WK, Heimbach JK, Wiesner RH, et al. Not All Cellular Rejections Are the Same: Differences in Early and Late Hepatic Allograft Rejection. Liver Transpl 2019;25:425-435.

19. Benitez C, Londono MC, Miquel R, Manzia TM, Abraldes JG, Lozano JJ, Martinez-Llordella M, et al. Prospective multicenter clinical trial of immunosuppressive drug withdrawal in stable adult liver transplant recipients. Hepatology 2013;58:1824-1835.

20. Feng S, Ekong UD, Lobritto SJ, Demetris AJ, Roberts JP, Rosenthal P, Alonso EM, et al. Complete immunosuppression withdrawal and subsequent allograft function among pediatric recipients of parental living donor liver transplants. JAMA 2012;307:283-293.

21. Shaked A, DesMarais MR, Kopetskie H, Feng S, Punch JD, Levitsky J, Reyes J, et al. Outcomes of immunosuppression minimization and withdrawal early after liver transplantation. Am J Transplant 2019;19:1397-1409.

22. Banff Working Group on Liver Allograft P. Importance of liver biopsy findings in immunosuppression management: biopsy monitoring and working criteria for patients with operational tolerance. Liver Transpl 2012;18:1154-1170.

23. Bonaccorsi-Riani E, Pennycuick A, Londono MC, Lozano JJ, Benitez C, Sawitzki B, Martinez-Picola M, et al. Molecular Characterization of Acute Cellular Rejection Occurring During Intentional Immunosuppression Withdrawal in Liver Transplantation. Am J Transplant 2016;16:484-496.

24. Bohne F, Martinez-Llordella M, Lozano JJ, Miquel R, Benitez C, Londono MC, Manzia TM, et al. Intra-graft expression of genes involved in iron homeostasis predicts the development of operational tolerance in human liver transplantation. J Clin Invest 2012;122:368-382.

25. Venturi C, Sempoux C, Bueno J, Ferreres Pinas JC, Bourdeaux C, Abarca-Quinones J, Rahier J, et al. Novel histologic scoring system for long-term allograft fibrosis after liver transplantation in children. Am J Transplant 2012;12:2986-2996.

26. Demetris AJ, Bellamy C, Hubscher SG, O'Leary J, Randhawa PS, Feng S, Neil D, et al. 2016 Comprehensive Update of the Banff Working Group on Liver Allograft Pathology: Introduction of Antibody-Mediated Rejection. Am J Transplant 2016;16:2816-2835.

27. Ng VL, Alonso EM, Bucuvalas JC, Cohen G, Limbers CA, Varni JW, Mazariegos G, et al. Health status of children alive 10 years after pediatric liver transplantation performed in the US and Canada: report of the studies of pediatric liver transplantation experience. J Pediatr 2012;160:820-826 e823.

28. Soltys KA, Mazariegos GV, Squires RH, Sindhi RK, Anand R, Group SR. Late graft loss or death in pediatric liver transplantation: an analysis of the SPLIT database. Am J Transplant 2007;7:2165-2171.

29. Londono MC, Souza LN, Lozano JJ, Miquel R, Abraldes JG, Llovet LP, Quaglia A, et al. Molecular profiling of subclinical inflammatory lesions in long-term surviving adult liver transplant recipients. J Hepatol 2018.

30. Campbell PM, Senior PA, Salam A, Labranche K, Bigam DL, Kneteman NM, Imes S, et al. High risk of sensitization after failed islet transplantation. Am J Transplant 2007;7:2311-2317.

31. Hricik DE, Formica RN, Nickerson P, Rush D, Fairchild RL, Poggio ED, Gibson IW, et al. Adverse Outcomes of Tacrolimus Withdrawal in Immune-Quiescent Kidney Transplant Recipients. J Am Soc Nephrol 2015;26:3114-3122.

32. Jucaud V, Shaked A, DesMarais M, Sayre P, Feng S, Levitsky J, Everly MJ. Prevalence and Impact of De Novo Donor-Specific Antibodies During a Multicenter Immunosuppression Withdrawal Trial in Adult Liver Transplant Recipients. Hepatology 2019;69:1273-1286.

33. Wiebe C, Gibson IW, Blydt-Hansen TD, Pochinco D, Birk PE, Ho J, Karpinski M, et al. Rates and Determinants of Progression to Graft Failure in Kidney Allograft Recipients With De Novo Donor-Specific Antibody. Am J Transplant 2015.

34. Hodges AM, Lyster H, McDermott A, Rice AJ, Smith JD, Rose ML, Banner NR. Late antibody-mediated rejection after heart transplantation following the development of de novo donor-specific human leukocyte antigen antibody. Transplantation 2012;93:650-656.

35. Le Pavec J, Suberbielle C, Lamrani L, Feuillet S, Savale L, Dorfmuller P, Stephan F, et al. De-novo donor-specific anti-HLA antibodies 30 days after lung transplantation are associated with a worse outcome. J Heart Lung Transplant 2016;35:1067-1077.

36. Taner T, Gandhi MJ, Sanderson SO, Poterucha CR, De Goey SR, Stegall MD, Heimbach JK. Prevalence, course and impact of HLA donor-specific antibodies in liver transplantation in the first year. Am J Transplant 2012;12:1504-1510.

37. O'Leary JG, Demetris AJ, Friedman LS, Gebel HM, Halloran PF, Kirk AD, Knechtle SJ, et al. The role of donor-specific HLA alloantibodies in liver transplantation. Am J Transplant 2014;14:779-787.

38. Feng S, Demetris AJ, Spain KM, Kanaparthi S, Burrell BE, Ekong UD, Alonso EM, et al. Five-year histological and serological follow-up of operationally tolerant pediatric liver transplant recipients enrolled in WISP-R. Hepatology 2017;65:647-660.

39. Del Bello A, Congy-Jolivet N, Danjoux M, Muscari F, Kamar N. Donor-specific antibodies and liver transplantation. Hum Immunol 2016;77:1063-1070.

40. Vandevoorde K, Ducreux S, Bosch A, Guillaud O, Hervieu V, Chambon-Augoyard C, Poinsot D, et al. Prevalence, Risk Factors, and Impact of Donor-Specific Alloantibodies After Adult Liver Transplantation. Liver Transpl 2018;24:1091-1100.

41. Todo S, Yamashita K, Goto R, Zaitsu M, Nagatsu A, Oura T, Watanabe M, et al. A Pilot Study of Operational Tolerance with a Regulatory T Cell-Based Cell Therapy in Living Donor Liver Transplantation. Hepatology 2016.

42. Crispe IN. Immune tolerance in liver disease. Hepatology 2014;60:2109-2117.

43. Jackson AM, Kanaparthi S, Burrell BE, Lucas DP, Vega RM, Demetris AJ, Feng S. IgG4 donor-specific HLA antibody profile is associated with subclinical rejection in stable pediatric liver recipients. Am J Transplant 2020;20:513-524.

44. Levitsky J, Burrell BE, Kanaparthi S, Turka LA, Kurian S, Sanchez-Fueyo A, Lozano JJ, et al. Immunosuppression Withdrawal in Liver Transplant Recipients on Sirolimus. Hepatology 2019.

45. Wong T, Nouri-Aria KT, Devlin J, Portmann B, Williams R. Tolerance and latent cellular rejection in long-term liver transplant recipients. Hepatology 1998;28:443-449.

46. Heymann F, Peusquens J, Ludwig-Portugall I, Kohlhepp M, Ergen C, Niemietz P, Martin C, et al. Liver inflammation abrogates immunological tolerance induced by Kupffer cells. Hepatology 2015;62:279-291.

47. Wong YC, Tay SS, McCaughan GW, Bowen DG, Bertolino P. Immune outcomes in the liver: Is CD8 T cell fate determined by the environment? J Hepatol 2015;63:1005-1014.

48. Li Y, Zhao X, Cheng D, Haga H, Tsuruyama T, Wood K, Sakaguchi S, et al. The presence of Foxp3 expressing T cells within grafts of tolerant human liver transplant recipients. Transplantation 2008;86:1837-1843.

49. Taubert R, Danger R, Londono MC, Christakoudi S, Martinez-Picola M, Rimola A, Manns MP, et al. Hepatic Infiltrates in Operational Tolerant Patients After Liver Transplantation Show Enrichment of Regulatory T Cells Before Proinflammatory Genes Are Downregulated. Am J Transplant 2016;16:1285-1293.

FIGURE LEGENDS

FIGURE 1: iWITH enrollment diagram

Beginning with 1,178 potentially eligible recipients, 1,090 patients were sequentially excluded, resulting in 88 subjects who were fully eligible and initiated ISW.

FIGURE 2: iWITH primary endpoint: the outcome of immunosuppression withdrawal

Eighty-eight subjects initiated immunosuppression withdrawal in 7 steps according to a protocol-specified algorithm provided in Figure S2.

A. Thirty-three subjects met biochemical and histological criteria for operational tolerance. Fifty-five subjects were non-tolerant: 39 failed secondary to rejection; 16 failed secondary to histological findings although they met biochemical criteria. Three subjects did not complete trial participation, secondary to being lost to follow-up (operationally tolerant subject, after tolerance adjudication), withdrawal of consent and refusal to travel for the end-of-trial (year 4) biopsy (both subjects who rejected).

B. Among the 55 non-tolerant subjects, 35 rejected prior to tolerance adjudication. The remaining 20 subjects were determined to be non-tolerant based on adjudication biopsy findings: 16 subjects failed to meet histologic criteria of operational tolerance (Table S1) and 4 subjects met biopsy criteria for rejection despite stable relative to baseline serum ALT and GGT levels. We classified 39 subjects as nontolerant by rejection, 35 before the tolerance adjudication biopsy and 4 at the time of the adjudication biopsy.

FIGURE 3: Data regarding 33 tolerant subjects

A. *Trial entry, peak, and end-of-trial ALT and GGT values [mean (interquartile range; IQR)] for operationally tolerant subjects;* ALT and GGT levels over time are shown in Figure S3A.

B. *Change in key features of the final (year 4) compared to the baseline (year 0) biopsy for operationally tolerant subjects;* changes in additional biopsy features are presented in Figure S5A. Each row represents

a single subject. In both figures, subjects are presented in the same order, sorted by change in portal inflammation and then subject identification number. To calculate change over time, absolute scores at year 0 were subtracted from scores at year 4 for the following parameters: portal inflammation, portal, sinusoidal, and perivenular fibrosis and the LAFSc. All score scales ranged from 0 to 3 except the LAFSc scale which ranged from 0 to 9 (25). Pink indicates progression while green indicates regression; increasing intensity of either pink or green indicates larger magnitude of change. Gray indicates missing data; one operationally tolerant subject was lost to follow-up after 3 years of trial participation.

C. *The TCMR probability is plotted for protocol-driven liver biopsies collected at 3 timepoints (yr 0: trial entry; yr 2: tolerance adjudication; yr 4: end-of-trial).* The transcriptional probability of rejection was calculated based on the expression levels of genes in the TCMR signature. The dotted line corresponds to the optimal probability threshold to identify biopsies diagnostic of acute rejection. P-values correspond to an unpaired Mann-Whitney test.

FIGURE 4: Data regarding 39 non-tolerant by rejection subjects

A. *Timing of rejection episodes.* The time of rejection for each subject diagnosed with rejection is represented by a bar. Bar segments represent ISW steps (Figure S2); segment length represents step duration. Time of rejection diagnosis is marked by a circle for biopsy-proven acute rejection (n=37), based on central pathology assessment according to Banff criteria (22) or by a star for clinical rejection (n=2), defined by the trial protocol (Supplementary Appendix) as elevated liver tests treated with increased or re-initiation of immunosuppression but without biopsy confirmation. Rejection occurred during withdrawal for 33 subjects and after stopping immunosuppression for 6 subjects. Of these 6, 4 subjects with tolerance (Table S1) and 4 subjects met biopsy criteria for rejection despite stable relative to baseline liver tests were diagnosed with biopsy-proven acute rejection based on the tolerance adjudication biopsy and are noted with an asterisk.

B. *Trial entry, peak, and end-of-trial ALT and GGT values [mean (IQR)] for non-tolerant by rejection subjects;* ALT and GGT levels over time are shown in Figure S3B.

C. *Time to resolution of rejection for those with elevated liver tests are shown (n=35);* 4 subjects with stable relative to baseline ALT and GGT values but biopsy-proven acute rejection at the tolerance adjudication biopsy are excluded. Two definitions for resolution are presented: i) ALT and GGT values $\leq$ 1.5X baseline as defined in the trial protocol (black); one unresolved episode is censored (O) at the end of the trial; ii) ALT and GGT values <50 units per milliliter (gray).

D. *Immunosuppression exposure over the 4-year trial is shown for those who were non-tolerant by rejection and on tacrolimus (n=38);* one subject that converted to azathioprine monotherapy was excluded. Expected exposure (X axis) was calculated assuming that the subject was maintained on the dose at trial entry and plotted against actual exposure (Y axis). Pink circles (n=19) identify subjects with higher actual than expected exposure while green circles (n=19) identify subjects with lower actual than expected exposures. Color intensity increases with larger differences between actual and expected exposures.

E. *Change in key features of the final (year 4) compared to the baseline (year 0) biopsy for non-tolerant by rejection subjects;* changes in additional biopsy features are presented in Figure S5B. Each row represents a single subject. In both figures, subjects are presented in the same order, sorted by change in portal inflammation and then subject identification number. To calculate change over time, absolute scores at year 0 were subtracted from scores at year 4 for the following parameters: portal inflammation, portal, sinusoidal, and perivenular fibrosis and the LAFSc. All score scales ranged from 0 to 3 except the LAFSc scale which ranged from 0 to 9 (25). Pink indicates progression while green indicates regression; increasing intensity of either pink or green indicates larger magnitude of change. Gray indicates missing data. Two subjects did not complete trial participation one withdrew assent/consent after 3 years; the other refused to travel for the end of trial biopsy.

F. *The TCMR probability is plotted for liver biopsies collected at 3 timepoints (yr 0: trial entry; rej: time of rejection diagnosis; yr 4: end-of-trial).* The transcriptional probability of rejection was calculated based on the expression levels of genes in the TCMR signature. The dotted line corresponds to the optimal probability threshold to identify biopsies diagnostic of acute rejection. P-values correspond to an unpaired Mann-Whitney test.

FIGURE 5: Data regarding 16 non-tolerant by histology subjects

A. *Changes in the specific histological features utilized to adjudicate operational tolerance* (Table S1); each row represents a single subject. The upper 8 rows represent subjects who were kept off immunosuppression; the lower 8 rows represent subjects who were restarted on immunosuppression as a result of the tolerance adjudication biopsy. Two subjects, identified by carets, were re-initiated on immunosuppression prior to the end of the trial. To calculate change over time, absolute scores at year 0 were subtracted from those at year 2 for the following: 3 parameters of inflammation (portal, interface, and perivenular), 2 parameters of fibrosis (Ishak and perivenular), bile duct damage, and isolated arteriopathy. All score scales ranged from 0 to 3 except Ishak fibrosis stage which ranged from 0 to 6 (22).

Pink indicates progression and green indicates regression; increasing intensity indicates larger magnitude of change. Gray indicates missing data. All except one subject failed the primary endpoint due to new onset necro-inflammatory-type interface activity with or without other disqualifying features, such as increase in fibrosis stage of 2 or new onset isolated arteriopathy.

B. *Trial entry, peak, and end-of-trial ALT and GGT values [mean (IQR)] for non-tolerant by histology subjects;* ALT and GGT levels over time are shown in Figure S3B.

C. *Immunosuppression exposure over the 4-year trial for those who were non-tolerant by histology and on tacrolimus (n=13);* 3 subjects on cyclosporine were excluded. Expected exposure (X axis) was calculated assuming that the subject was maintained on the dose at trial entry and plotted against actual exposure (Y axis). The pink circle identifies the single subject with higher actual than expected exposure. The remaining subjects (n=12) with lower actual than expected exposures are identified by green symbols: green circles (n=5) identify subjects who resumed tacrolimus and green stars (n=7) identify subjects who remained off immunosuppression after the tolerance adjudication biopsy (year 2). Color intensity increases with larger differences between actual and expected exposures.

D. *Change in key features of the final (year 4) compared to the baseline (year 0) biopsy for non-tolerant by histology subjects;* changes in additional biopsy features are presented in Figure S5C. Each row represents a single subject. The upper 8 rows represent subjects who were kept off immunosuppression, while the lower 8 rows represent subjects who were restarted on immunosuppression as a result of the biopsy. In both figures, subjects in the groups of 8 are presented in the same order, sorted by change in portal inflammation and then subject identification number. Two subjects, identified by carets, were re-initiated on immunosuppression prior to the end of the trial (Supplementary Methods). To calculate change over time, absolute scores at year 0 were subtracted from scores at year 4 for the following parameters: portal inflammation, portal, sinusoidal, and perivenular fibrosis and the LAFSc. All score scales ranged from 0 to 3 except the LAFSc scale which ranged from 0 to 9(25). Pink indicates progression and green indicates regression; increasing intensity indicates larger magnitude of change. Gray indicates missing data. One subject did not undergo the end of trial biopsy secondary to for-cause within the preceding 6 months.

E. *The TCMR probability for protocol-driven liver biopsies collected at 3 timepoints (yr 0: trial entry; yr 2: tolerance adjudication; yr 4: end-of-trial).* The transcriptional probability of rejection was calculated based on the expression levels of genes in the TCMR signature. The dotted line corresponds to the optimal probability threshold to identify biopsies diagnostic of acute rejection. P-values correspond to an unpaired Mann-Whitney test.

FIGURE 6: Factors associated with operational tolerance

A. *Multiplex immunohistochemistry parameters of the eligibility biopsy separate tolerant from non-tolerant subjects.* Shown is a 3-dimensional scatter plot of tolerant (green circles; n=18) and non-tolerant (red squares; n=35) subjects according to the number of CD8+ cells per mm^2 (T effector cells; X axis), lobular CD45+/MHCII+ pairs per mm^2 (leukocyte/antigen-presenting cell pairs; Y axis), and MAC387+ cells per mm^2 (infiltrating macrophages; Z axis) in the eligibility biopsy. The inner cube identifies thresholds that, simultaneously, maximizes the number of tolerant subjects (17 of 18; 94%) and minimizes the number of non-tolerant subjects (12 of 35; 34%). Subjects within the inner cube are closed symbols; those outside are open symbols. Plots only show subjects for which values of all 3 parameters were available.

B. *Eligibility biopsies with comparable portal and lobular inflammation grade but different immunohistochemical inflammatory loads.* Hematoxylin and eosin sections are shown in the top row while corresponding immunostained sections [CD34 (green) /CD45 (teal) /MHCII (red)] are shown in the bottom row. The left column is the eligibility biopsy from an operationally tolerant subject while the right column is from a non-tolerant subject. Using a scale from 0 to 3, both biopsies were graded as 0 for both portal and lobular inflammation. Algorithmically detected pairings of leukocytes (CD45+) and antigen-presenting cells (MHCII+), shown in high magnification in the inset, are highlighted in yellow circles in the immunostained sections. The number of pairings was 7.6 per mm^2 for the tolerant (lower left) and 15.3 per mm^2 for the non-tolerant (lower right) subject.

C. *Class II DSA presence during ISW (year 0 to 1) is shown in 3 heatmaps.* A minimum mean fluorescence intensity (MFI) threshold of 1,000 was used to identify a positive class II DSA. Heat maps show the maximum MFI for class II DSA with a range from 1,000 to 20,000; white indicates missing data. Immunosuppression was reduced stepwise according to a protocol-specified algorithm (Figure S2). Class II DSA was determined at baseline (year 0), weeks 12, 24, 36, and year 1 as long as subjects continued to withdraw immunosuppression. After diagnosis of rejection, subjects were not tested for class II DSA until the year 1 visit. Hence, non-tolerant subjects have a high frequency of missing data, particularly at the week 24 and 36 timepoints.

Subjects are divided into those who did not show any DSA during year 1 (n=22), those who have detectible DSA at trial entry (n=44), and those who develop DSA as immunosuppression is reduced (n=20); 2 subjects with missing data at trial entry were excluded. Subjects within each group were ordered first by tolerance status, then timepoint, and finally MFI. Univariable logistic regression models were used to explore class

II DSA status during ISW for association with operational tolerance. The accompanying table shows ORs and 95% CIs.

Table 1: Characteristics of 88 subjects undergoing immunosuppression withdrawal by tolerance status

CHARACTERISTIC[a]			Tolerant n=33	Non-tolerant n=55
Donor	**Age** (years)		15 (2-27)	10 (3-31)
	Male gender		18 (55)	26 (47)
	Race	White	22 (67)	39 (71)
		Black	5 (15)	5 (9.1)
		Other	6 (18)	11 (20)
	Deceased		22 (67)	35 (64)
Recipient	**Age at transplant** (years)		1 (1-2)	1 (1-3)
	Male gender		13 (39)	26 (47)
	Race	White	30 (91)	46 (84)
		Black	1 (3.0)	3 (6)
		Other	2 (6)	6 (11)
	Transplant indication	Acute liver failure	2 (6)	5 (9)
		Biliary atresia	20 (61)	31 (56)
		Tumor	2 (6)	3 (6)
		Metabolic liver disease	2 (6)	7 (13)
		Other	7 (21)	9 (16)
Transplant	**Whole graft**		15 (46)	26 (47)
	Previous rejection episodes	0	24 (73)	32 (58)
		1	7 (21)	12 (22)
		2 or more	2 (6)	11 (20)
	Time since last rejection (years)		8 (7-9)	6 (4-8)
	Received induction immunosuppression		8 (24)	8 (15)
At Trial Entry	**Tacrolimus**		29 (88)	52 (95)
	Tacrolimus dose (n=81; mg/kg/day)		0.04 (0.02-0.05)	0.05 (0.04-0.07)
	Age (years)		11 (7-13)	11 (8-13)
	Time since transplant (years)		9 (6-10)	8 (6-11)
	Alanine aminotransferase (U/L)		26 (21-30)	23 (19-30)
	Gamma-glutamyl transferase (U/L)		14 (12-19)	15 (12-19)
	Anti-nuclear antibody (n=77)	Positive (≥ 1:40)	4 (13)	12 (26)
	Anti-smooth muscle antibody (n=77)	Positive (1:80)	3 (10)	0
	Quantitative immunoglobulin G (n=73; mg/dL)		629 (562-822)	701 (616-801)
	Eplet mismatch (n=86)	Total (DR + DQ)	27 (16-45)	28 (20-39)
		DQ only	10 (5-15)	9 (5-15)
	α-Class II DSA (n=87)	Positive	15 (47)	29 (53)
		Maximum MFI[c] >20,000	4 (13)	5 (9)
		MFI[c] sum >20,000	6 (19)	8 (15)

[a] Continuous variables are summarized using median and interquartile range. Categorical variables are summarized by counts and percentages.

[b] Seven subjects were on cyclosporine at trial entry.

[c] MFI: mean fluorescence intensity.

Table 2. Univariable models of factors potentially associated with operational tolerance[a]

FACTOR	Reference group	OR	95% confidence interval
Demographic and clinical			
Age at transplant (per year)		0.92	0.70-1.20
Age at trial entry (per year)		0.98	0.86-1.11
Time since transplant (per year)		1.00	0.87-1.13
Living donor	Deceased donor	0.88	0.35-2.17
Whole liver	Partial liver	0.93	0.39-2.21
Induction at time of transplant	None	1.84	0.62-5.50
History of rejection	None	0.52	0.21-1.33
Alanine aminotransferase (per year)		1.04	0.98-1.10
Gamma glutamyl transferase (per year)		0.98	0.93-1.05
Serological			
Class II eplet mismatch (per unit increment)		1.00	0.97-1.03
Class II DSA present at trial entry	No Class II DSA at trial entry	0.61	0.25-1.48
Class II present at baseline (n=44)	No Class II DSA during year 1 (n=22)	**0.30**	**0.10-0.86**
Class II DSA develops *de novo* (n=20)		**0.10**	**0.02-0.45**
Histological			
Mild portal inflammation	None	**0.36**	**0.14-0.90**
Mild lobular inflammation	None	0.29	0.06-1.42
Mild perivenular inflammation[b]	None	NA	NA
Ishak fibrosis stage (per unit increment)		1.45	0.69-3.06
Portal fibrosis (per unit increment)		1.37	0.57-3.25
Sinusoidal fibrosis (per unit increment)		0.63	0.27-1.47
Perivenular fibrosis (per unit increment)		0.54	0.21-1.39
Liver allograft fibrosis score (per unit increment)		0.87	0.59-1.29
Immunohistochemical[c]			
CD45+ cells (per unit increment; n=70)	Portal	**0.87**	**0.79-0.96**
	Lobular	**0.96**	**0.94-0.99**
	Total	**0.97**	**0.95-0.99**

MHC II+ cells (per unit increment; n=70)	Portal	0.99	0.97-1.01
	Lobular	**0.98**	**0.97-1.00**
	Total	0.99	0.98-1.00
CD45+/ MHC II+ pairs (per unit increment; n=70)	Portal	**0.78**	**0.64-0.94**
	Lobular	**0.80**	**0.70-0.92**
	Total	**0.82**	**0.74-0.92**
MAC 387+ cells (per unit increment; n=68)		**0.91**	**0.85-0.97**
CD4+ cells (per unit increment; n=72)		1.00	1.00-1.00
CD8+ cells (per unit increment; n=72)		0.99	0.97-1.00
Transcriptional			
TCMR probability score[26] (per unit increment; n=75)		0.10	0.002-5.23
5-Gene tolerance biomarker[25] (per unit increment; n=83)		0.96	0.19-4.74

[a] Significant associations are identified in **bold.** Numbers are provided when data is not available for all 88 subjects.

[b] All subjects with mild peri-venular inflammation were in the non-tolerant group.

Figure 1

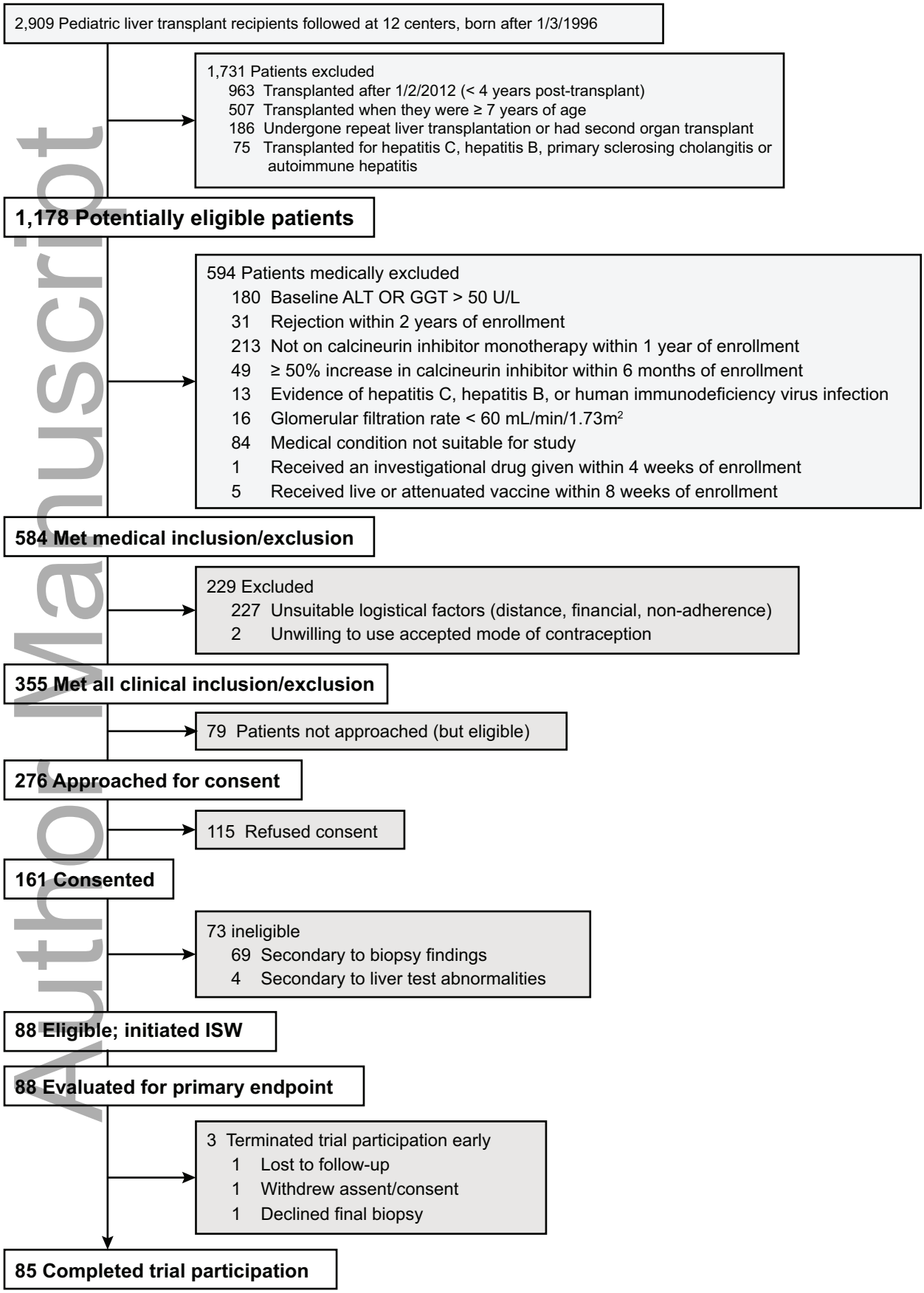


Figure 2

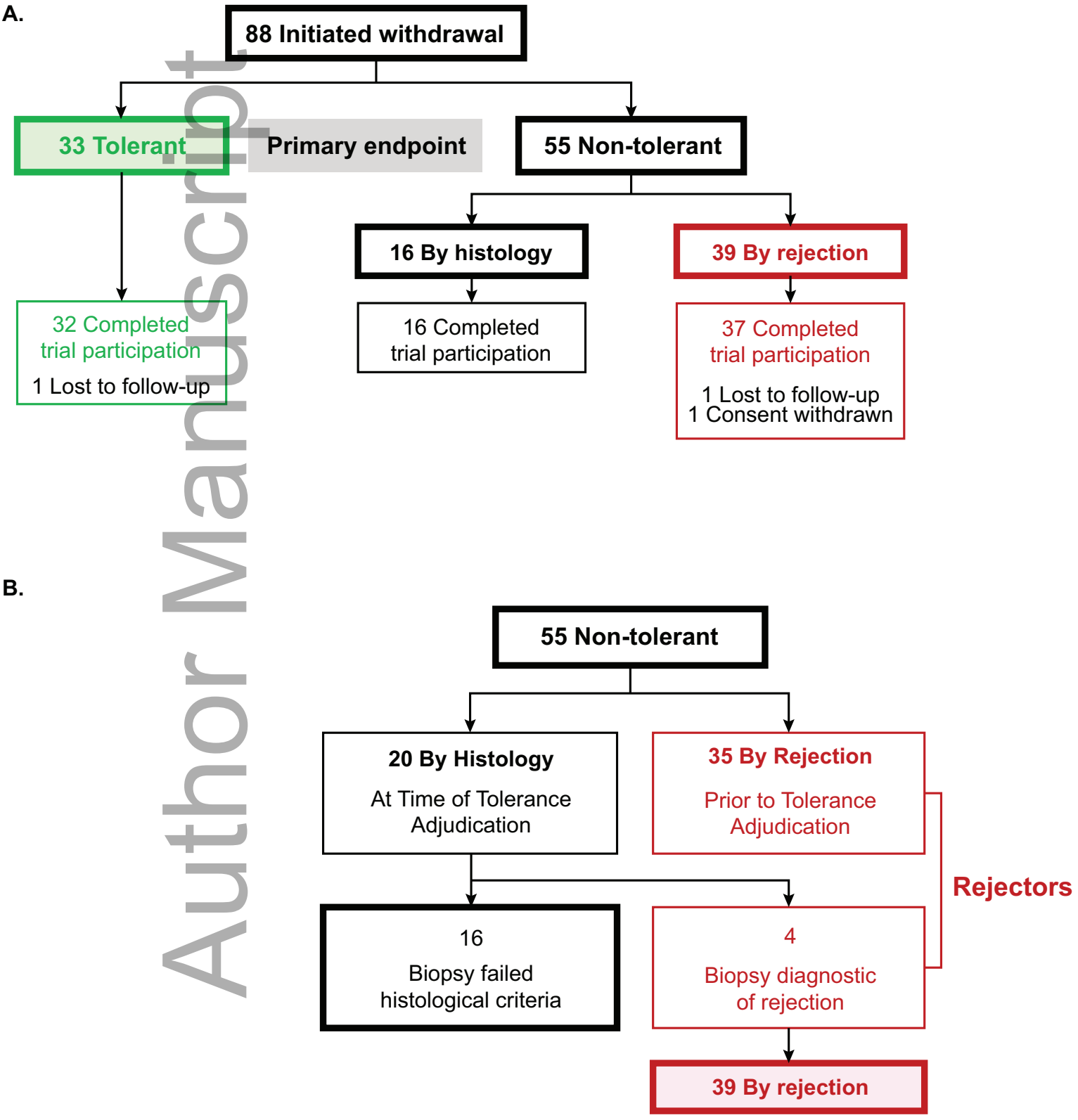
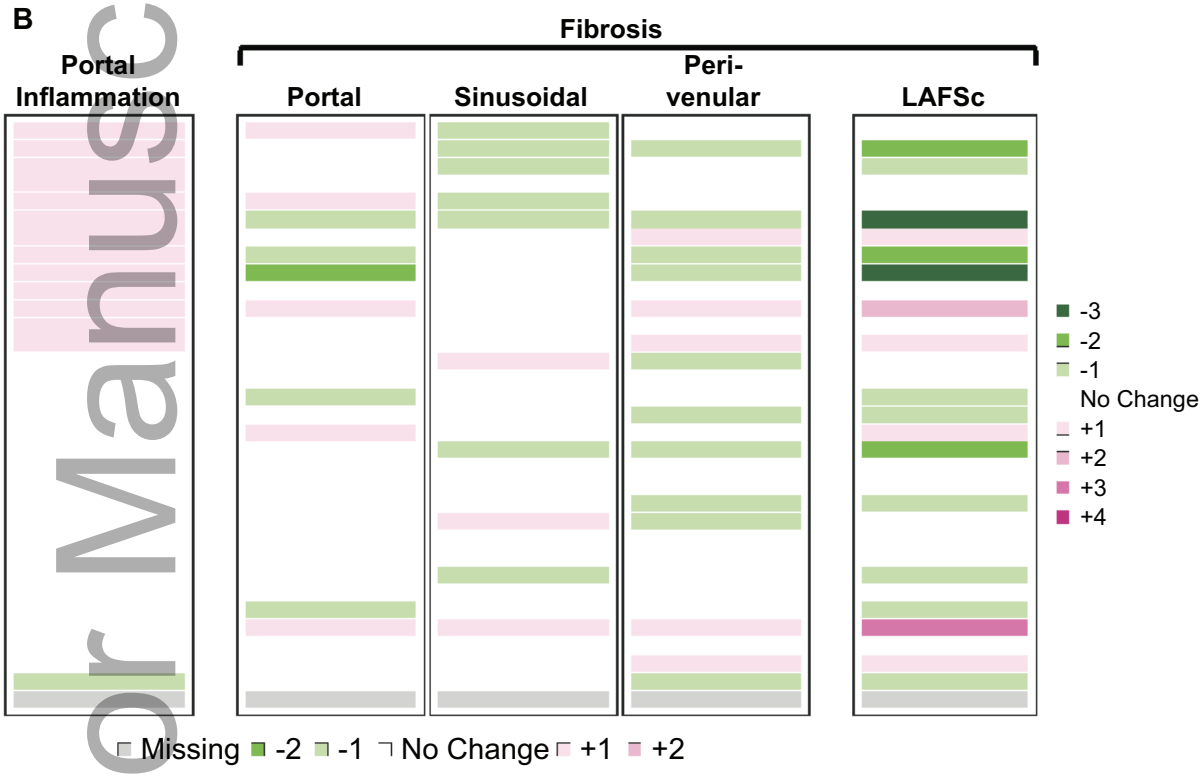


Figure 3

A

	Baseline	Peak	Final
ALT (U/L)	26 (21-30)	54 (44-75)	30 (19-35)
GGT (U/L)	14 (12-19)	34 (19-45)	15 (13-24)

B



C

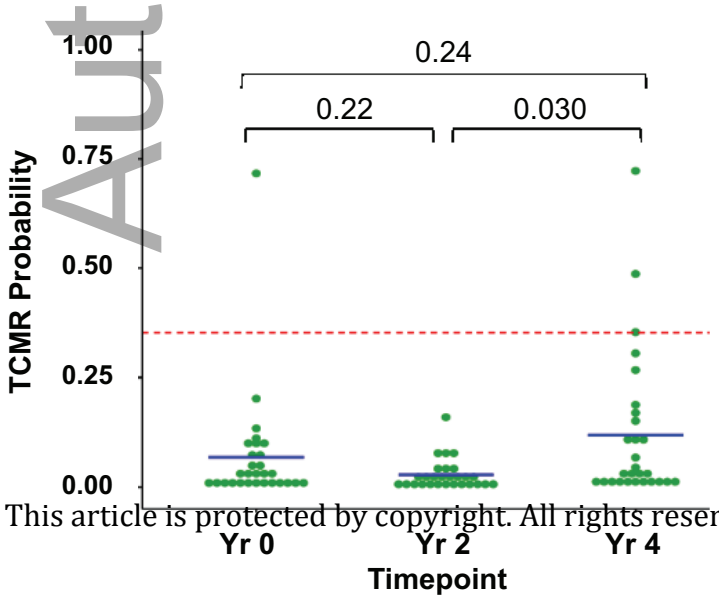
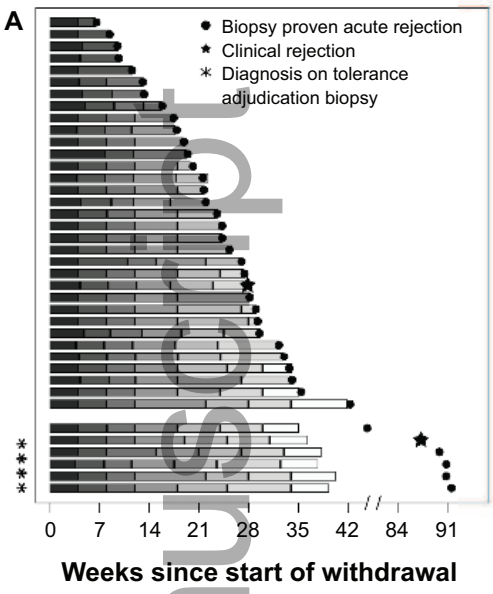
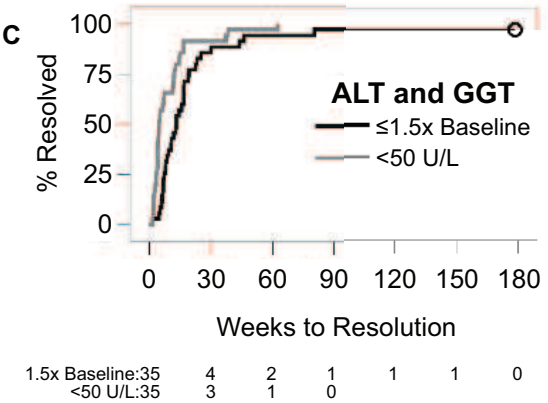


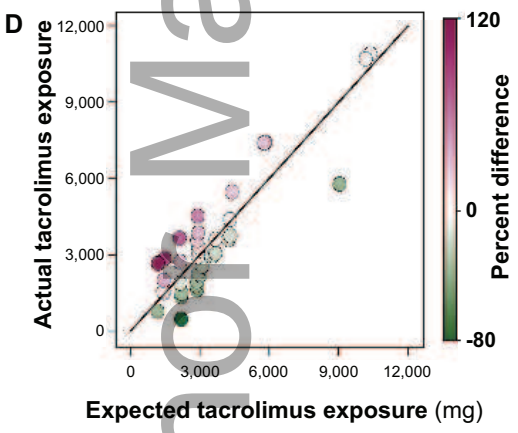
Figure 4

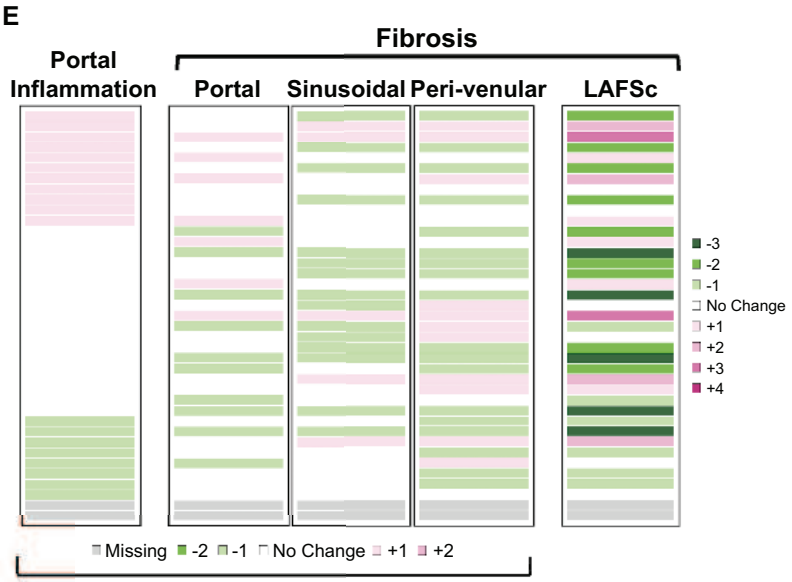


B

	Baseline	Peak	Final
ALT (U/L)	25 (20-31)	136 (101-205)	23 (19-29)
GGT (U/L)	16 (12-19)	63 (42-104)	17 (12-40)







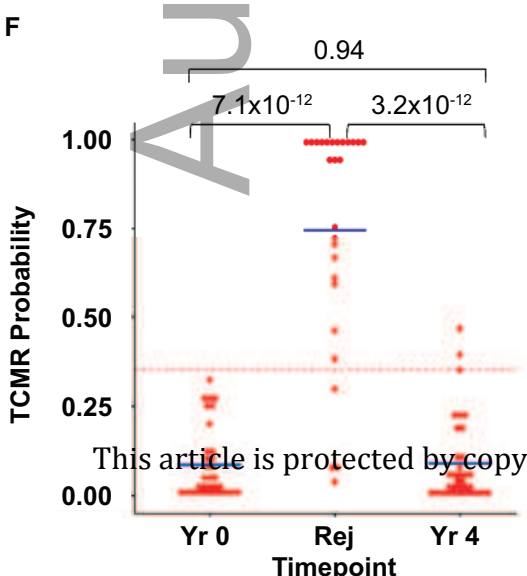
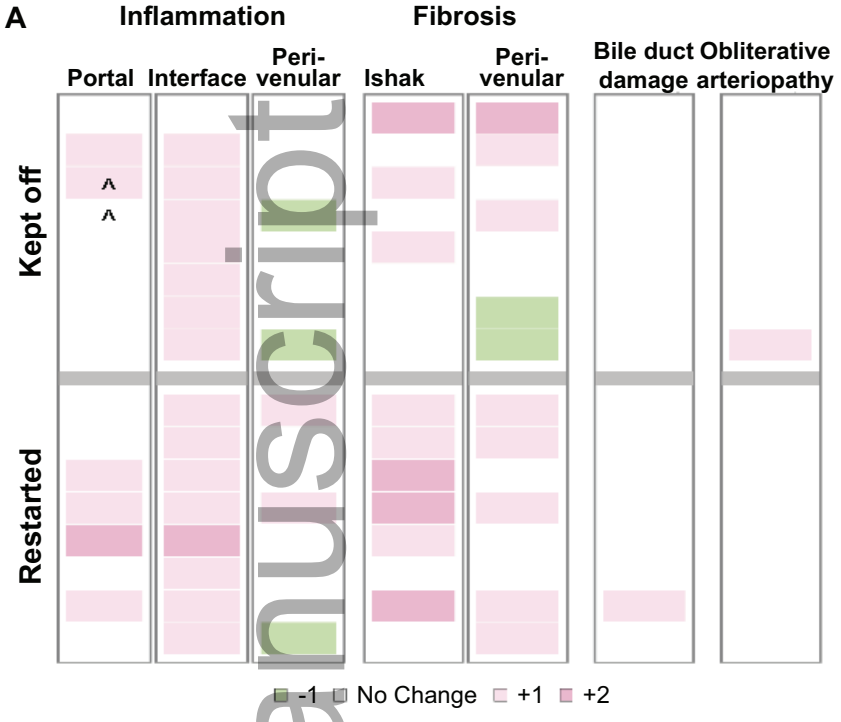
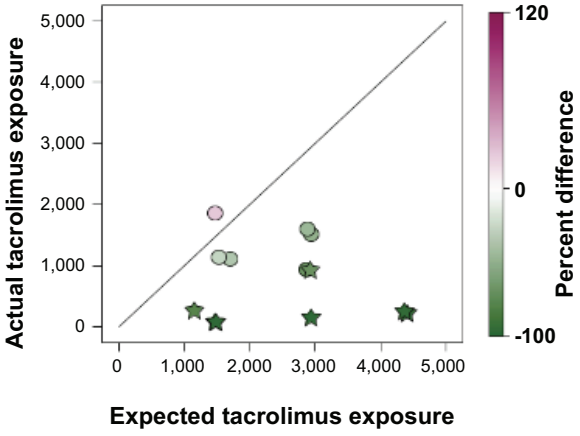


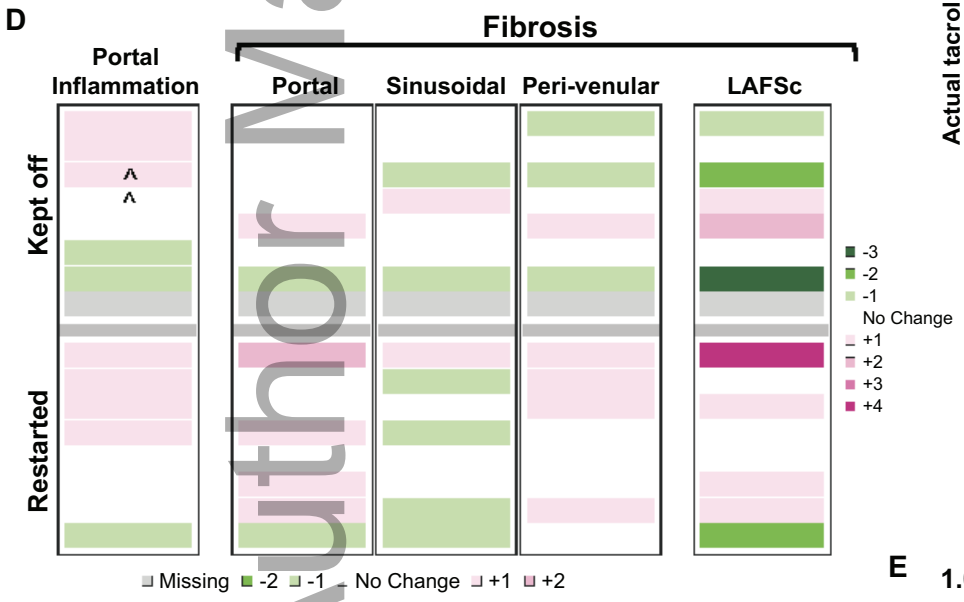
Figure 5



B

	Baseline	Peak	Final
ALT (U/L)	22 (19-25)	44 (34-62)	24 (19-28)
GGT (U/L)	13 (11-19)	24 (16-41)	16 (11-23)





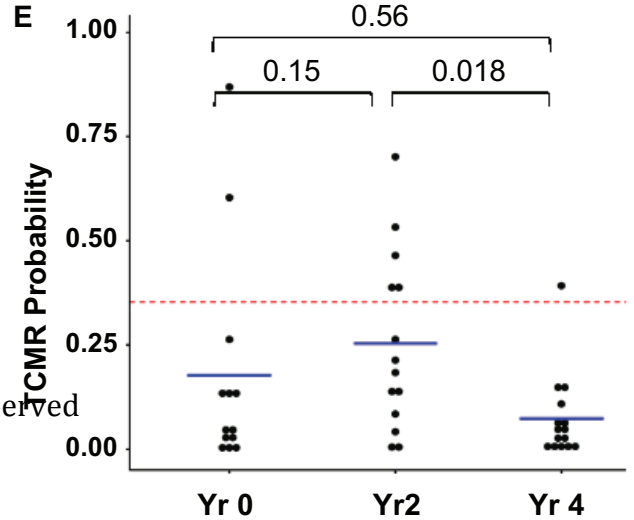


Figure 6