Supplementary information for:

Computational Analysis of Continuous Body Temperature Provides Early Discrimination of Graft-versus-Host Disease in Mice

Short title: Early Discrimination of Graft-versus-Host Disease

Kuang He[1], Zhenke Wu[2,3,4], Hideaki Fujiwara[1], Steven Whitesall[5], Cynthia Zajac[1], Sung Won Choi[6], Pavan Reddy[1], Muneesh Tewari[1,7,8,9,*]

Supplemental Figure Legends

Figure S1. ***Ex vivo*** **accuracy tests of RFID sensors with and without irradiation.** Two implants were submerged in a heated water bath, before and after being exposed to a split dose of ^{137}Cs for a combined total of 8 Gy, which is the dose used in subsequent mouse HSCT experiments. The readings from the sensors were compared with a digital thermometer (Data Logger Thermometer with USB and RS232, Omega) as a gold standard as the temperature of the water bath was ramped up in a stepwise fashion.

Figure S2. **Quality assessment and cleaning of the raw temperature data.** (a) A histogram of recorded temperature shows a wider range than the physiological range. (b) A histogram of signal strength reveals presence of low quality data points (signal strength < 19). (c-d) Temporal plots of two mice that died during the course of monitoring show noise recorded after the death of the animals. (e) Workflow of data cleaning. Three crude filtered were applied to the raw temperature data to remove non-physiological (1) and low quality (2) data points, as well as those recorded after animals died (3).

Figure S3. **Clinical GvHD assessments.** (a) Group average of percentage body weight change from Day 0. (b) Group average of clinical GvHD Scores graded on a scale of 0-10 by five equally weighted parameters: weight loss, hunching, mobility, fur texture, and skin integrity.

Figure S4. **Descriptive statistics of the results for naive, SynLowT, and AlloLowT mice.** (a) Mean hourly temperatures (left y axes) and boxplots of daily variances of all individuals in each of the three groups (right y axes). Outliers in the boxplots are not shown. (b) Boxplots of mean daily temperature of each individual in the SynLowT and AlloLowT groups. Outliers are shown in square points. Asterisks indicate significant differences in the daily means between the SynHighT

and AlloHighT groups (p values were calculated for Days 0-23 using 2-sided t tests and adjusted for multiple testing using the Benjamini-Yekutieli method; *p<0.05, ***p<0.005). (c) Mean daily night time temperature and mean daily day time temperature of all individuals in the naive, SynLowT, and AlloLowT groups. (d) Daily difference of mean daily night time temperature and mean daily day time temperature of all individuals in the naive, SynLowT, and AlloLowT groups.

Figure S5. **Boxplots of temperatures recorded in each hour of the day shown in five time blocks.** Temperature recorded during night time (house lights off) are represented in blue and during day time (house lights on) in red. Pre: pre-transplant; Post: post-transplant.

Figure S6. **Average daily activity counts of the naive, SynLowT, SynHighT, AlloLowT, and AlloHighT mice.**

Supplemental Tables

	Naive	SynLowT	SynHighT	AlloLowT	AlloHighT
pre-Day0	36.4 ± 0.9	37.0 ± 0.9	36.8 ± 0.9	36.8 ± 0.9	37.0 ± 0.9
post-Day0	37.0 ± 0.9	36.2 ± 1.4	36.2 ± 1.3	35.7 ± 1.8	35.5 ± 1.6

Table S1. Mean ± standard deviation for the five groups of mice in this study, pre-Day0 and post-Day0. Day 0 refers to the day of transplant for SynLowT, SynHighT, AlloLowT, and AlloHighT mice.

	Naive	SynLowT	SynHighT	AlloLowT	AlloHighT
pre-transplant	0.7	0.8	0.6	0.6	0.7
Days 0-14	0.7	1.2	0.8	0.6	0.5
Days 15-23	0.9	1.3	1.0	4.5	3.1
Range	[0.4, 1.3]	[0.3, 11.1]	[0.1, 17.8]	[0.1, 19.4]	[0.1, 13.1]

Table S2. Median and range of daily variances in body temperatures of the five mouse groups.

Supplemental Methods

Assessment of GvHD. Systemic GvHD was graded using a clinical GvHD score on a scale of 0-10 by five equally weighted clinical parameters[13]: weight loss, posture (hunching), mobility, fur texture, and skin integrity.

Data pre-processing criteria

1. Physiological range of body temperature. The raw temperature data recorded from the implanted sensors ranged from -73.018 to 99.982 °C . Since the ambient temperature was kept at 22 °C and mice usually maintain thermostability, temperature readings below 20 °C or above 50 °C were considered noise and therefore filtered (**Figure 6a**).
2. Signal strength. The sensors need to be within 12 inches from the receivers, which were placed under the cages. As certain procedures (e.g. irradiation, transplantation, body weight measurements, cage cleaning, etc) required that the mice be moved away from the receivers, there is a significant amount of data points recorded with signal strength below 19, which is a cutoff value recommended by the manufacturer.
3. Date of death. Two of the mice died during the study, therefore the data points that were recorded after the death of the animals should be filtered out. Since the exact times of death are unknown, we filtered all the data points recorded after the time at which each animal was observed having a body temperature of 30 °C, which is a conservative estimate of the time of death.

Data Analysis.

We first binned data into 2-7 day sliding windows and used principal component analysis (PCA) to project these windows of data onto two-dimensional space with principal components (PC) 1 and 2 such that these two dimensions represent the most variances in the data input (**Figure 2a**). Next, we used an exploratory clustering technique called K-means to classify the projected data points (one per animal) into two groups based on euclidean distances (i.e. how far apart the data points are from one another). Since the grouping was performed in an unsupervised fashion, we calculated an adjusted Rand Index (aRI) [11] to assess concordance between K-means assigned grouping and true identities of the data points (**Figure 2a**). The rationale behind this sliding window design is to imitate real-time monitoring of patients, for whom early changes in core temperature may rise to categorize them into high or low risk groups for GvHD. For example, if K-means clusters all the syn mice (non-GvHD) into one cluster, and all the allo mice (GvHD) into the other on Day 7, based on the temperature data from Days 3-7 (a window of 5 days), then GvHD treatments could be initiated timely when they are most effective.

Figure S1

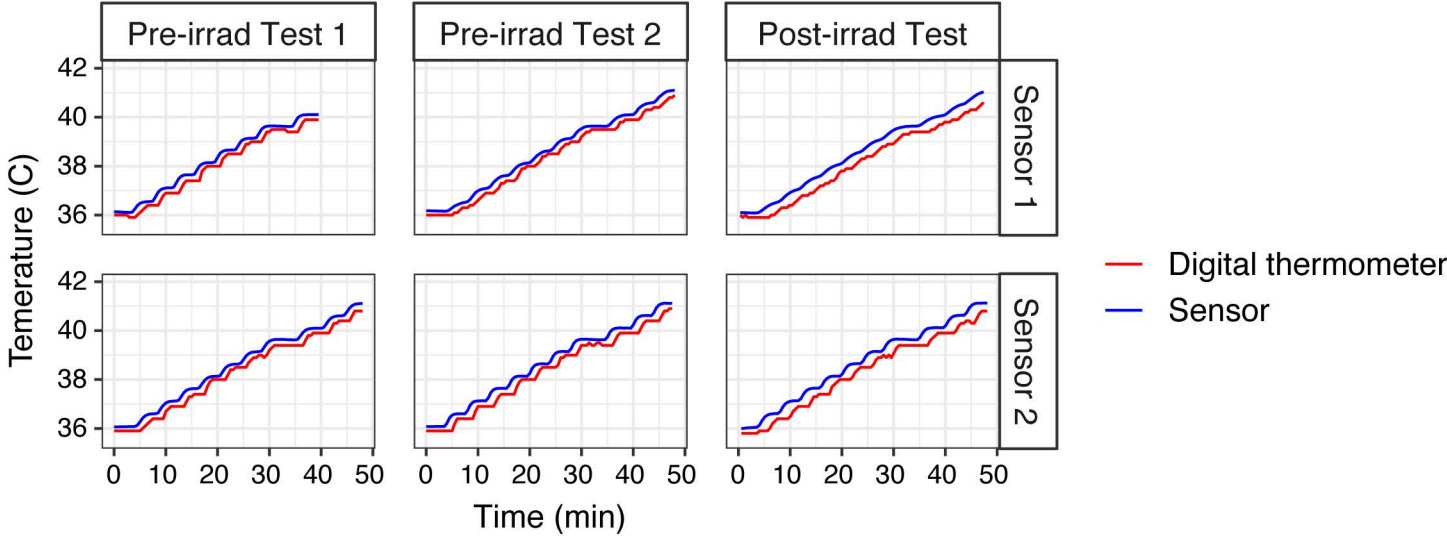
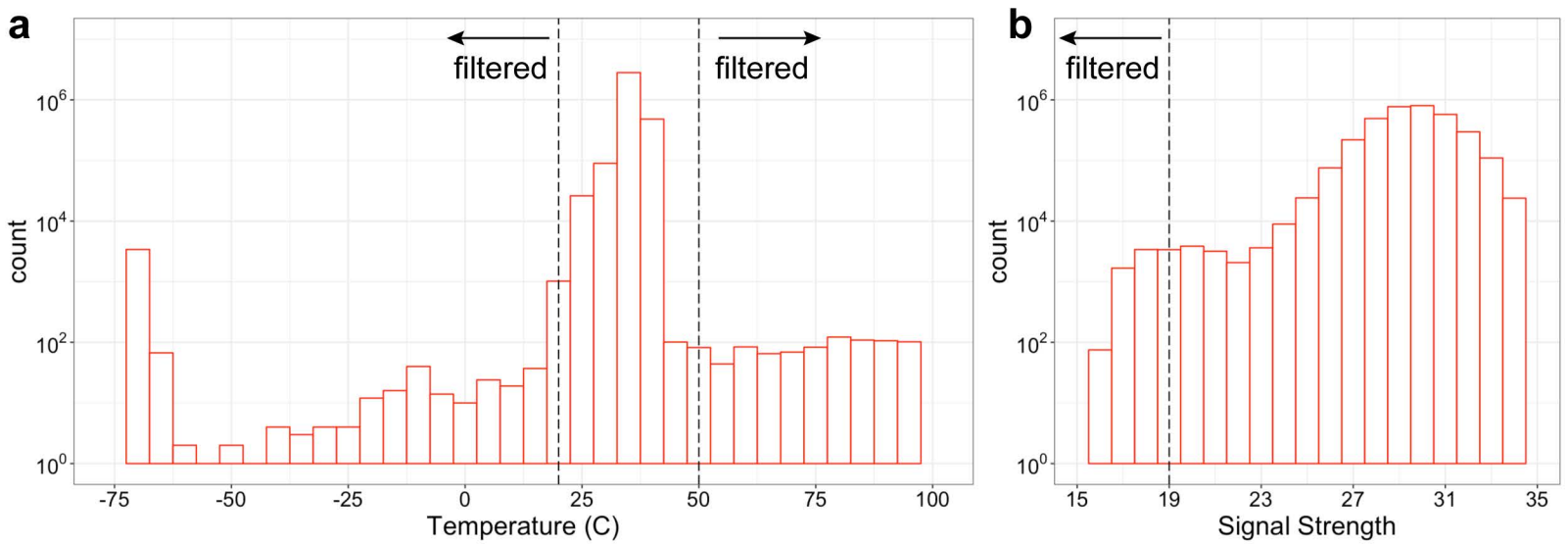
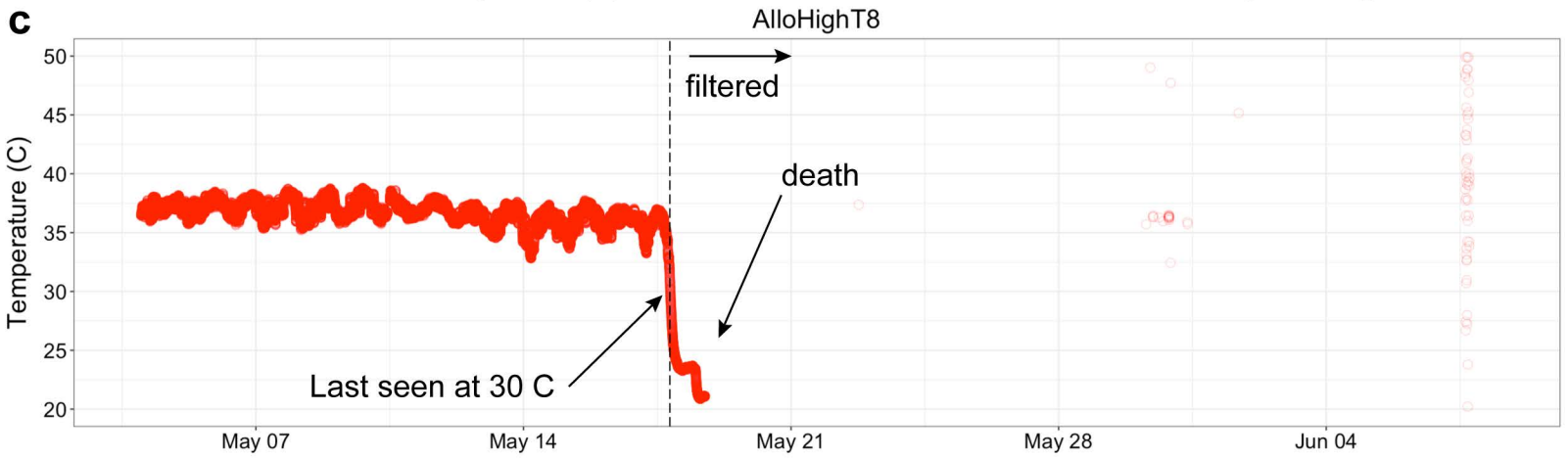
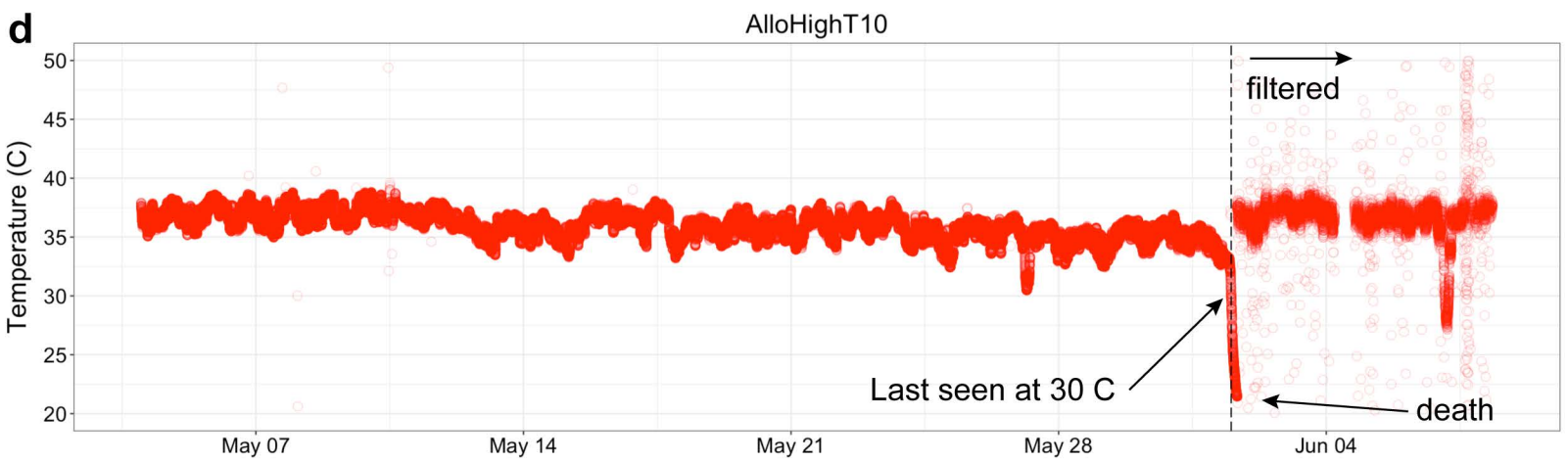

Pre-irrad Test 1
Pre-irrad Test 2
Post-irrad Test
Sensor 1
Sensor 2
Temerature (C)
42
40
38
36
0
10
20
30
40
50
Time (min)
Digital thermometer
Sensor

Figure S2


a
filtered
filtered
count
Temperature (C)
b
filtered
count
Signal Strength


c
AlloHighT8
filtered
death
Last seen at 30 C
Temperature (C)
May 07
May 14
May 21
May 28
Jun 04


d
AlloHighT10
filtered
Last seen at 30 C
death
Temperature (C)
May 07
May 14
May 21
May 28
Jun 04

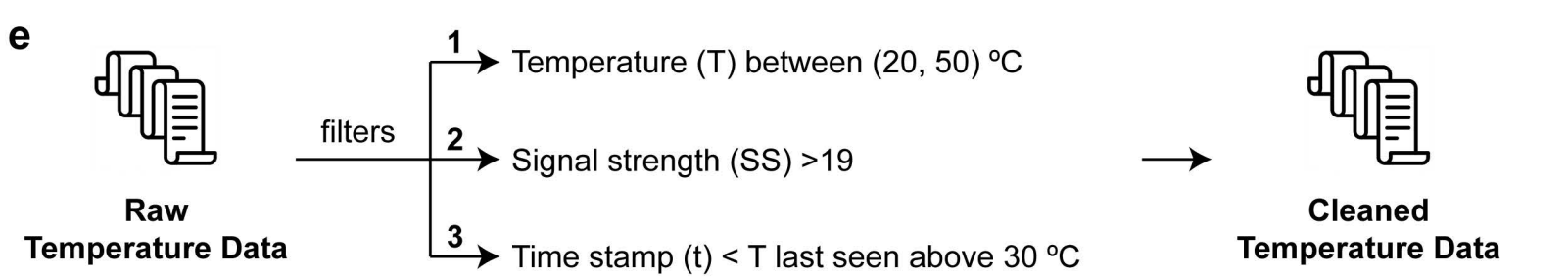

e
filters
1
Temperature (T) between (20, 50) ºC
2
Signal strength (SS) >19
3
Time stamp (t) < T last seen above 30 ºC
Raw Temperature Data
Cleaned Temperature Data

Figure S3

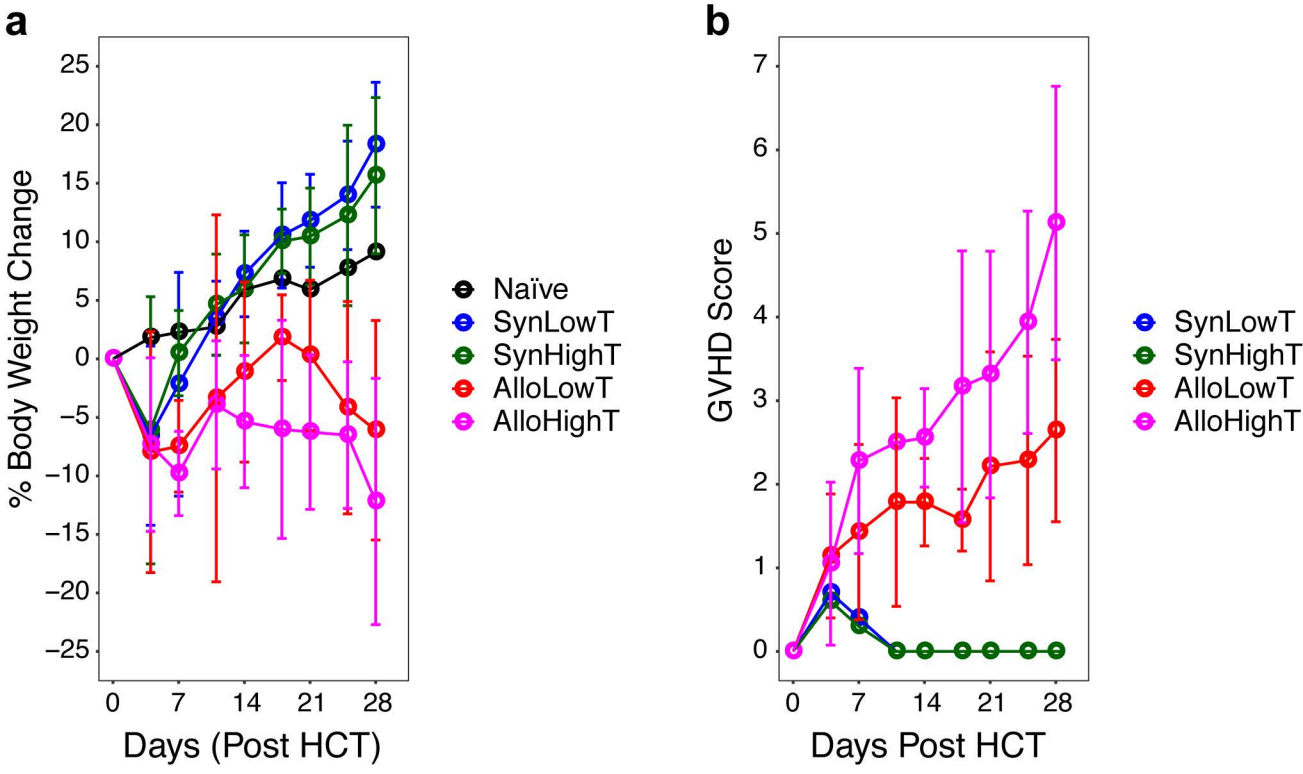

a
% Body Weight Change
25
20
15
10
5
0
−5
−10
−15
−20
−25
0
7
14
21
28
Days (Post HCT)
Naïve
SynLowT
SynHighT
AlloLowT
AlloHighT
b
GVHD Score
7
6
5
4
3
2
1
0
0
7
14
21
28
Days Post HCT
SynLowT
SynHighT
AlloLowT
AlloHighT

Figure S4

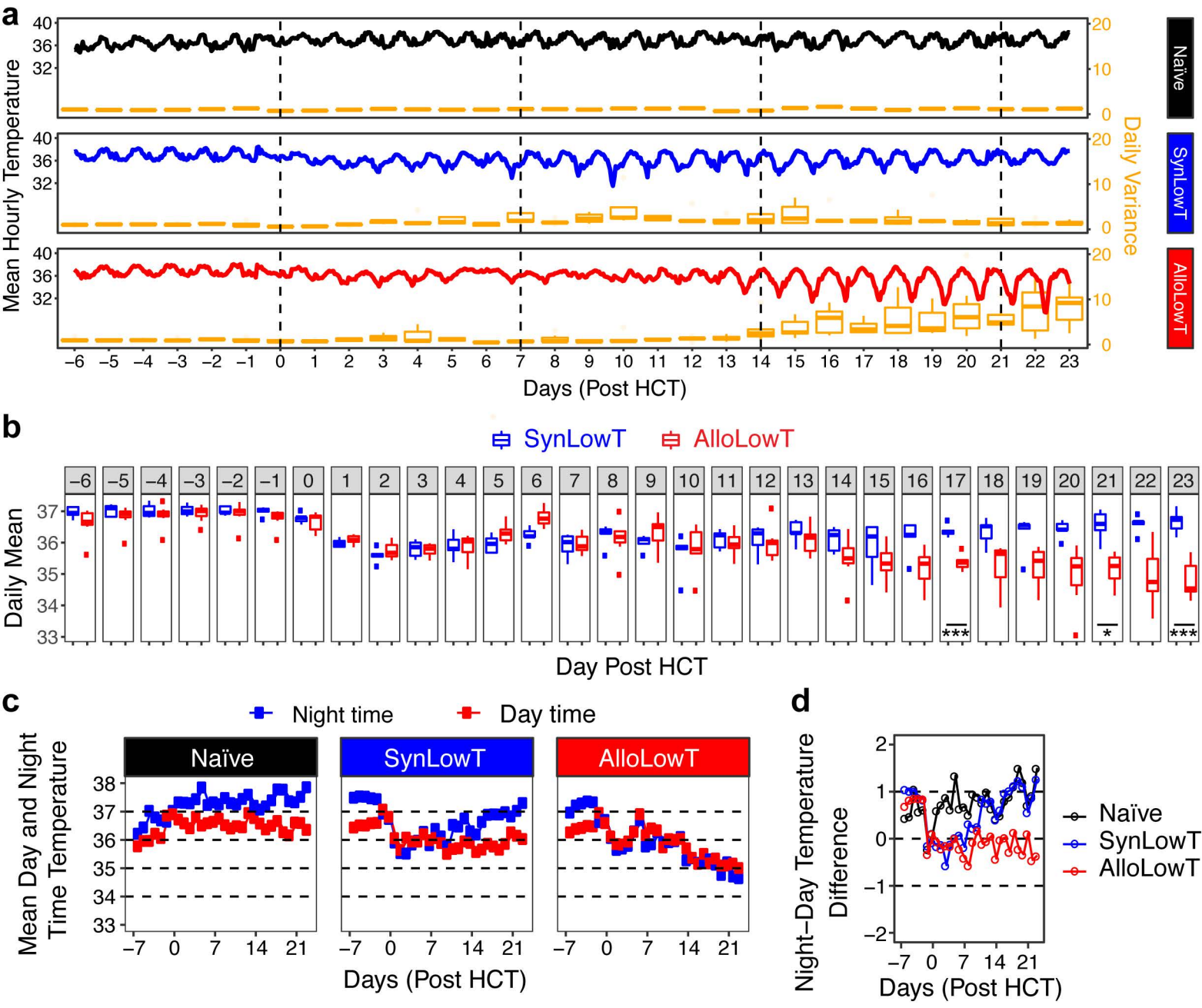

a
Mean Hourly Temperature
Daily Variance
Naïve
SynLowT
AlloLowT
Days (Post HCT)
b
SynLowT
AlloLowT
Daily Mean
Day Post HCT
c
Night time
Day time
Naïve
SynLowT
AlloLowT
Mean Day and Night Time Temperature
Days (Post HCT)
d
Night–Day Temperature Difference
Naïve
SynLowT
AlloLowT
Days (Post HCT)

Figure S5

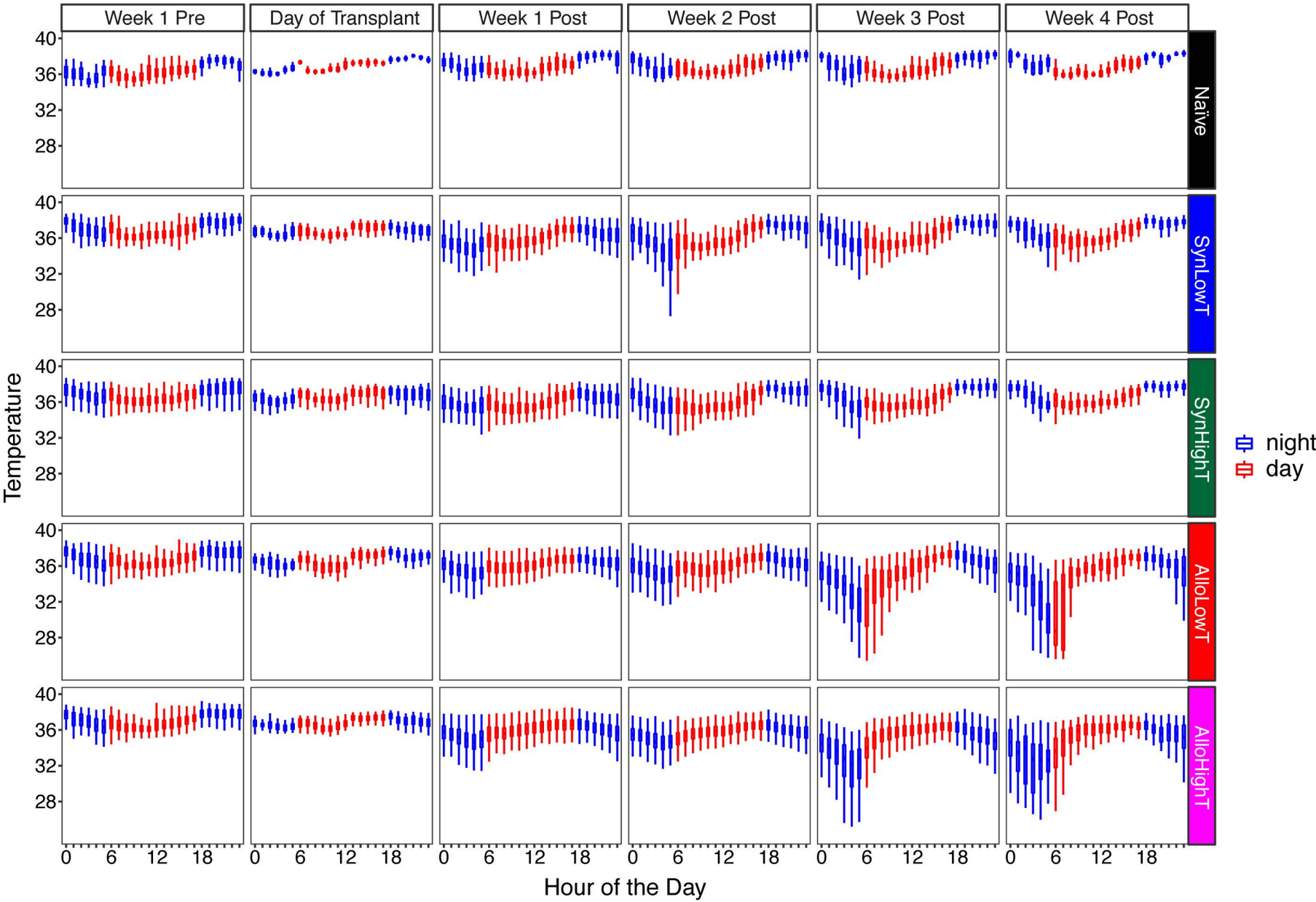
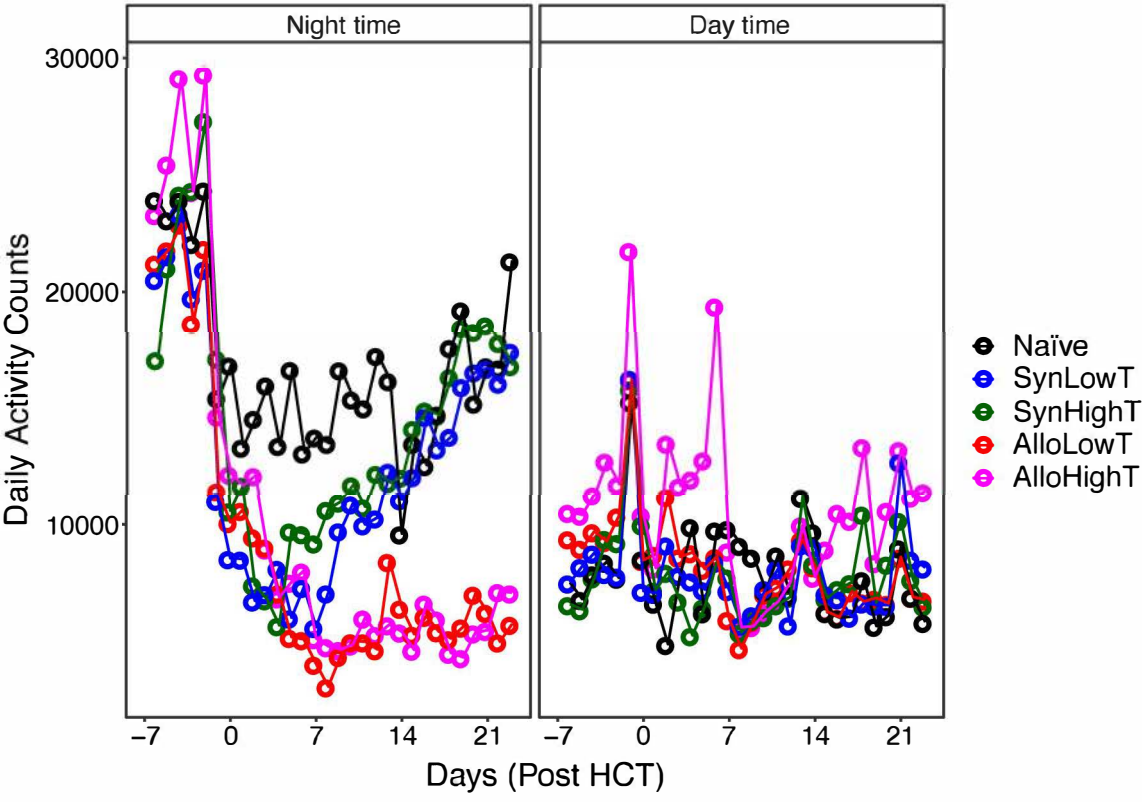

Week 1 Pre
Day of Transplant
Week 1 Post
Week 2 Post
Week 3 Post
Week 4 Post
Naïve
SynLowT
SynHighT
AlloLowT
AlloHighT
Temperature
40
36
32
28
0
6
12
18
Hour of the Day
night
day

Figure S6


Night time
Day time
30000
20000
10000
Daily Activity Counts
−7
0
7
14
21
Days (Post HCT)
Naïve
SynLowT
SynHighT
AlloLowT
AlloHighT