

1 **Supplementary Material for: “Mobile Footprinting: Linking Individual Distinctiveness in**
 2 **Mobility Patterns to Mood, Sleep, and Brain Functional Connectivity”**

3
 4
 5
 6

Supplementary Table 1 | Psychiatric diagnoses of participants

Axis I Diagnosis		Axis II Diagnosis	
Total	n=41		n=41
No Diagnosis	n=3	No Diagnosis	n=7
Diagnosis	n=38	Deferred diagnosis	n= 18
Major depressive disorder	24	Borderline personality disorder	n=12
Bipolar disorder	4	Personality disorder NOS	n=4
Depressive disorder NOS	1		
Mood disorder NOS	1		
Generalized anxiety disorder	14		
Post-traumatic stress disorder	14		
Social phobia	12		
Obsessive-compulsive disorder	11		
Panic disorder	5		
Anxiety disorder NOS	2		
Attention-Deficit Hyperactivity Disorder	6		
Schizoaffective Disorder	1		
Substance Related Disorders	18		

7
 8

1 **Supplementary Table 2 | Footprint mobility metrics**

2

GPS Mobility Metric ¹	Definition	Comments
1	Home Time Amount of time (minutes) spent within a 200-meter radius of home	Home location is determined by the location with the largest total amount of time between 9PM and 6AM over the course of the study period.
2	Distance Travelled Sum of the flight lengths (meters) over all flights that occurred on that day	Flights are quantified as segments of linear movements.
3	Radius of Gyration Root mean square distance (meters) from a central location	As radius of gyration quantifies the distribution around a central location, it acts as a proxy for the area covered in a day.
4	Maximum Diameter Maximum pairwise distance (meters) between any two pause locations that both occur on the same day	
5	Maximum Distance from Home Maximum of the distances (meters) between home and every other pause location.	
6	Significant Locations Visited Number of significant locations a subject visited in a day	The set of significant locations is determined by a K-means procedure on all pause locations with greater than 10 minutes
7	Average Flight Length Average length (meters) of all flights that occur on that day	
8	Standard Deviation of Flight Length Standard deviation of flight length (meters) over all flights that occur on that day	
9	Average Flight Duration Average duration (seconds) of all flights that occur on that day	
10	Standard Deviation of Flight Duration Standard deviation of flight duration (seconds) over all flights that occur on that day	
11	Probability of Pauses Fraction of time stationary relative to the amount of time spent moving	Pauses are defined as periods of no movement
12	Entropy of Significant Locations Visited Randomness of the locations a subject visited in a day	Large values indicate a subject spreading their time out across many different locations fairly evenly for that day, whereas small values indicate a concentration at few significant locations
13	Minutes Missing Number of minutes of missing data for that day	
14	Circadian Routine Degree to which a participant followed their daily routine on a given day	Low values indicate a break from routine on a given day, whereas high values indicate that the subject followed their daily routine
15	Weekend/Weekday Circadian Routine Degree to which a participant followed their weekend or weekday routine	

	Accelerometer Mobility Metric²	Definition	Comments
1	Maximum Magnitude	The maximum magnitude of acceleration, defined by $\ acceleration\ = \sqrt{x^2 + y^2 + z^2}$.	
2	Minimum Magnitude	The minimum magnitude of acceleration.	
3	Average Magnitude	The average magnitude of acceleration.	
4	Median Magnitude	The median magnitude of acceleration.	
5	Standard Deviation Magnitude	The standard deviation of acceleration.	
6	Absolute Duration of Exertional Activity	Total duration of daily exertional activity based on acceleration	The threshold for defining exertional vs. non-exertional activity was set at $0.15g^2$, as described in Panda et al ³ .
7	Relative Duration of Exertional Activity	Duration of daily exertional activity relative to non-exertional activity	

1

2

1 **Supplementary Table 3 | Ecological Momentary Assessment of Mood**

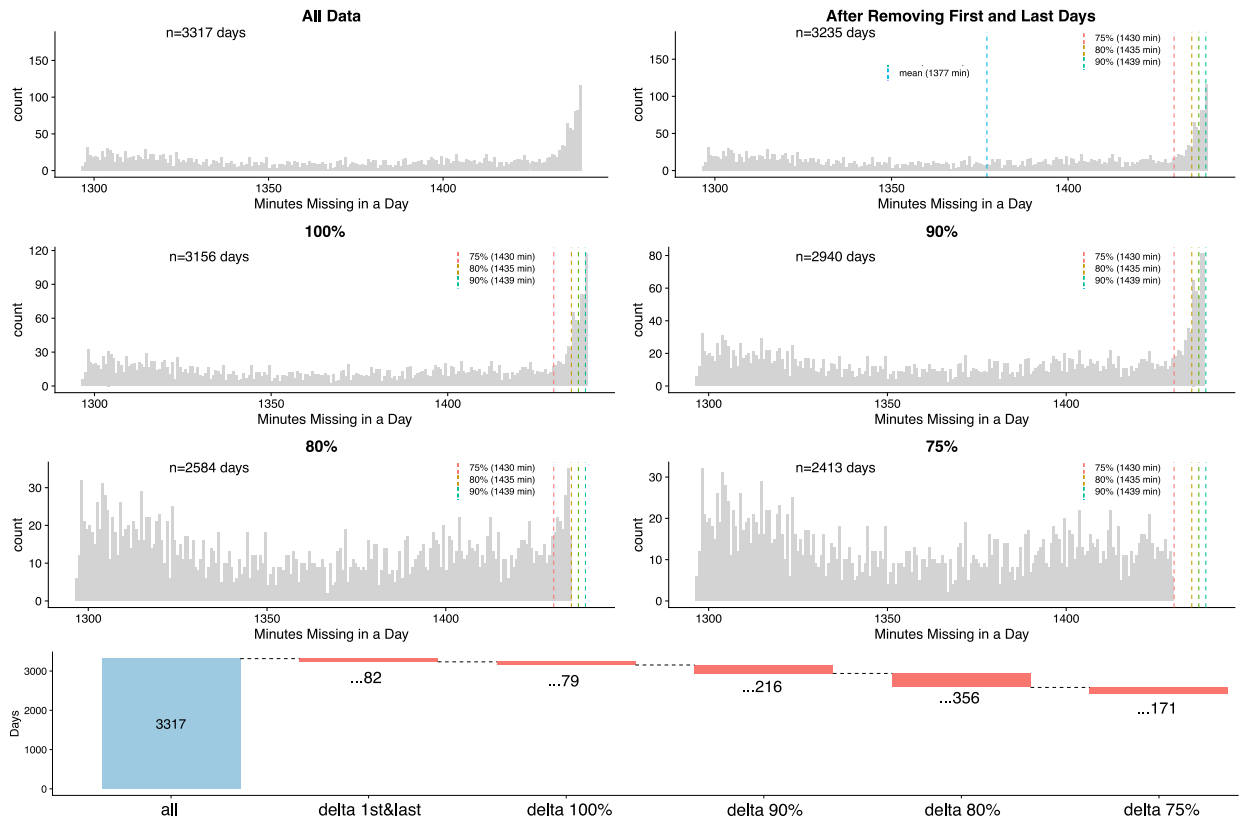
2

Since last beep...

1	I have felt really, really angry or out of control. 1 = not at all; 7 = extremely
2	I thought I am no good, a bad person, or that nobody loves me. 1 = not at all; 7 = extremely
3	I have been social versus isolated myself from others. 1 = extremely isolated; 7 = very social
4	I had an argument. 1 = yes; 2 = no
5	my social interactions have been overall positive versus negative. 1 = very positive; 7 = very negative
6	I have done something impulsive or risky. 1 = yes; 2 = no

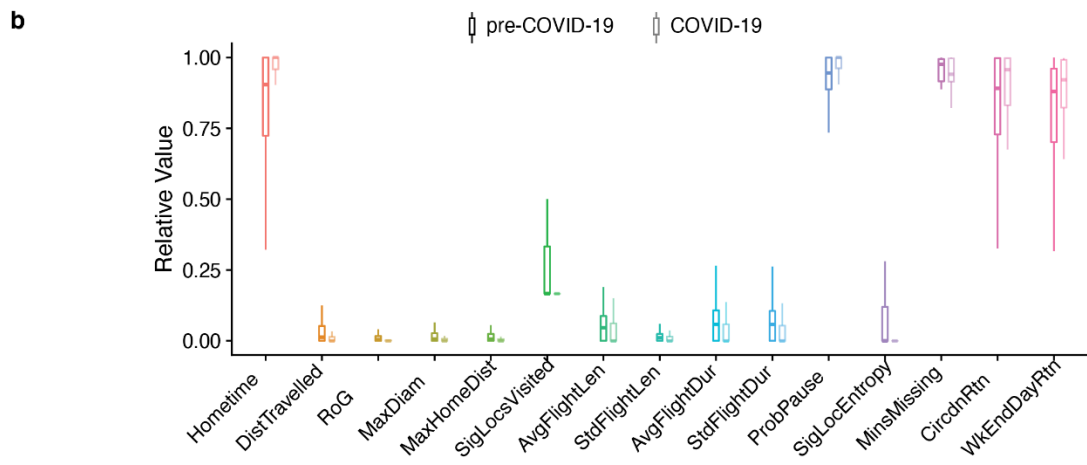
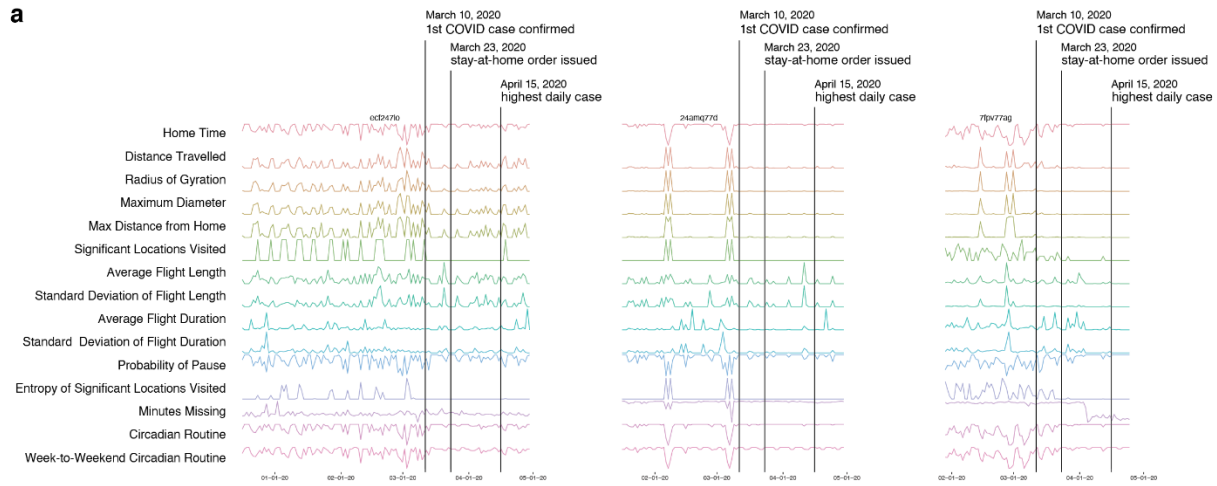
3

1
2



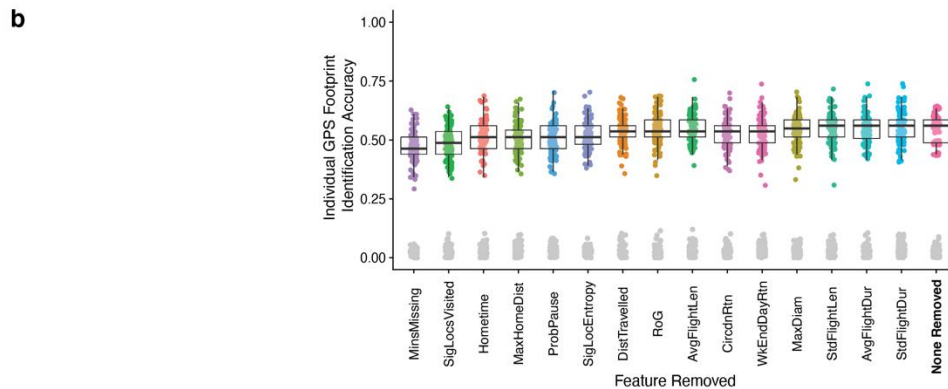
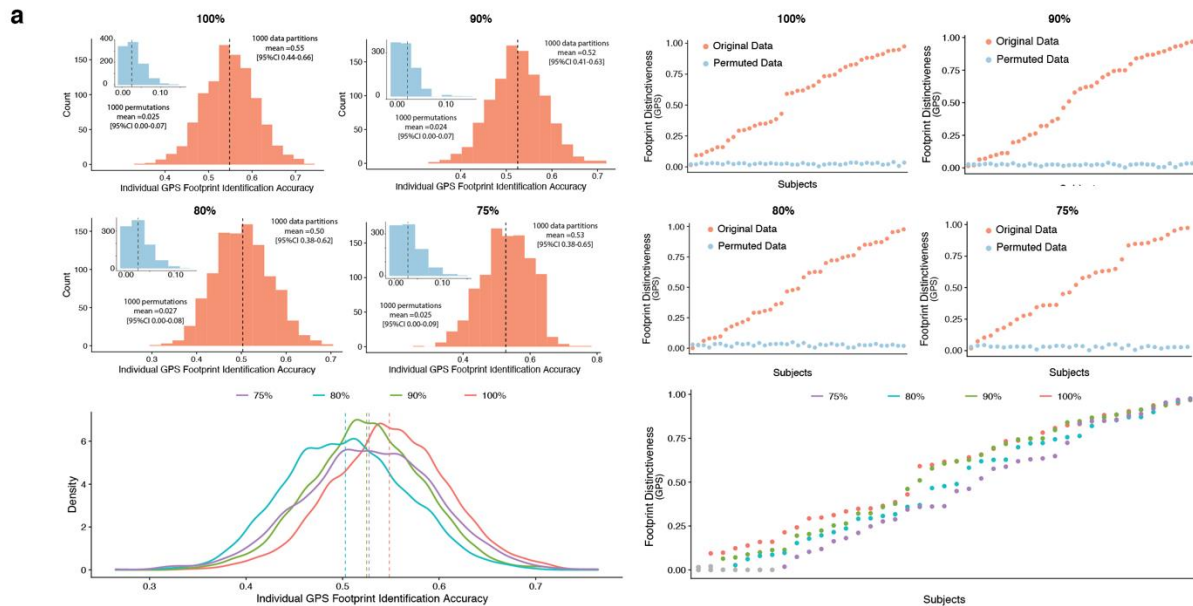
3
4
5
6
7
8
9
10
11
12
13
14

Supplementary Figure 1 | Mobility data collection and thresholds for missing data. In total, 3,317 days of recording across all participants were obtained (mean (s.d.) = 77 (26) days, range 14–132 days). After removing the first and last days of each participant’s study period when only partial GPS data were recorded (n=82) and days that contained no data (n=79), the remaining data available for analysis contained 3,156 days. To assess the results’ sensitivity to data quality, data were then thresholded at three levels 90, 80, and 75 percentiles of missingness. In this case, 100% indicates that a given day was included if any data from that day was available (e.g., missing up to 1,440 minutes, the total number of minutes exist in a day), whereas 75% indicates that days with data missing at the 75th percentile level (missing up to 1,430 minutes).



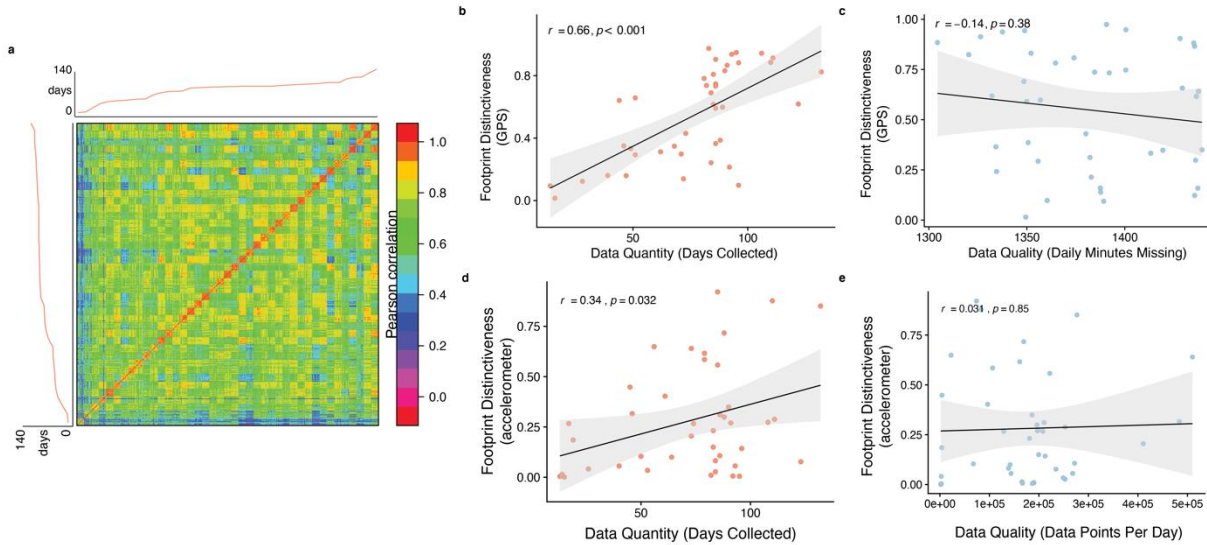
1
2
3
4
5
6
7
8
9
10
11
12
13

Supplementary Figure 2 | Dramatic shift in mobility metrics during COVID-19. a) Three participants still had active mobile phone data collection when COVID-19 hit the Philadelphia area during the end of the study period. Vertical lines indicate milestone dates: 1st confirmed case in Philadelphia (March 10, 2020), stay-at-home order issued by the mayor (March 23, 2020), and the highest daily case in the first wave (April 15, 2020). Visually, mobility metrics across the board exhibited shift since the onset of the pandemic, which we quantified in the bottom panel. b) As expected, participants stayed at home longer, travelled less distance, visited fewer locations, took on a more stable circadian routine. However dramatic, data points collected during COVID-19 represented only 1.1% of all data in the current study. Exclusion of these data points did not significantly change the results.

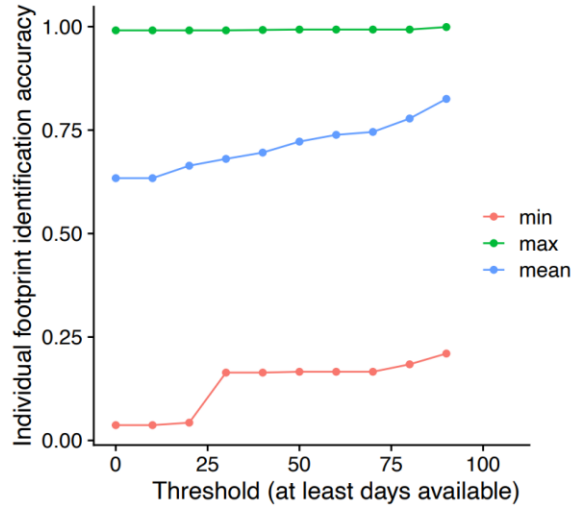


1
2 **Supplementary Figure 3 | Convergent results when different thresholds for data**
3 **missingness are applied or individual mobility features are removed. a)** Histograms of GPS-
4 based identification accuracy as well as subject-level footprint distinctiveness were stable across
5 four inclusion thresholds for data missingness (also see **Supplementary Figure 1**). **b)** Individual
6 identification accuracy was not significantly changed when any of the GPS mobility features was
7 removed from the covariance matrix (“footprint”). Permutation results were shown in grey dots,
8 which demonstrated individual identification were statistically significant regardless which
9 feature was removed.

10



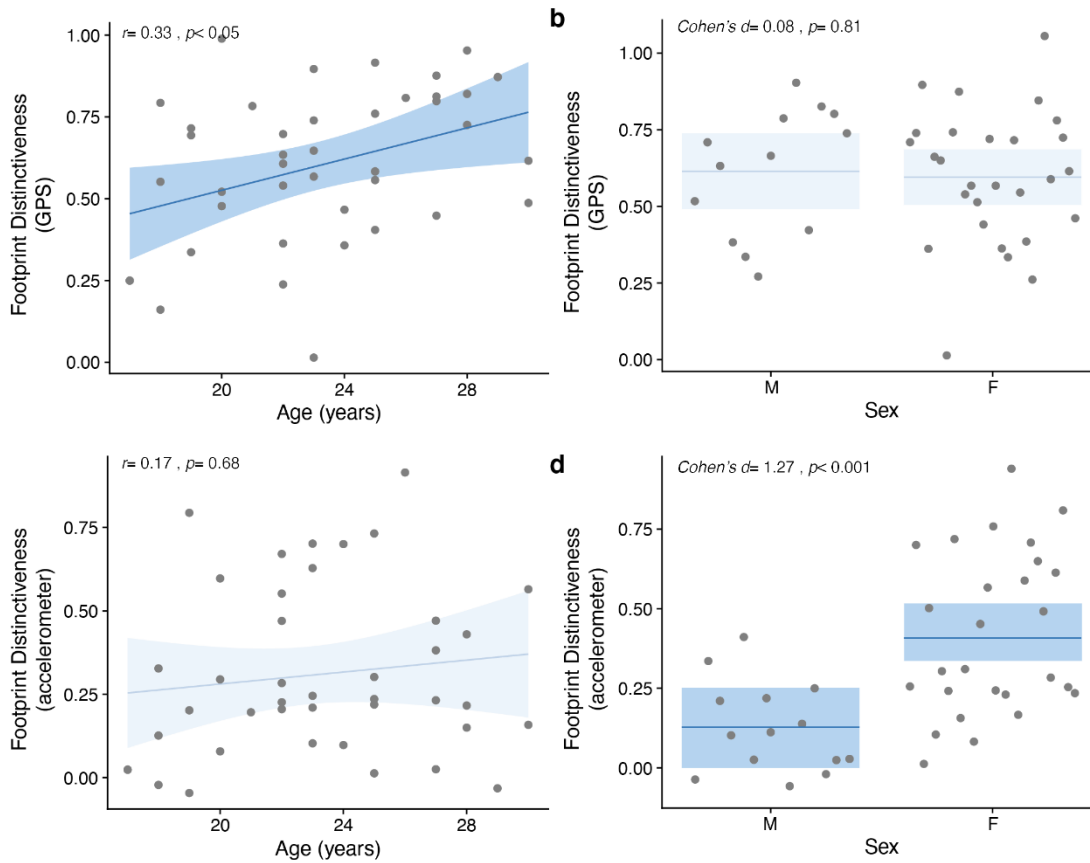
1
2 **Supplementary Figure 4 | Individual footprint distinctiveness is driven by data quantity**
3 **but not data quality.** **a)** We constructed a subject footprint similarity matrix by grouping
4 together multiple random reference and target partitions from the same subject. The visual
5 contrast between the on and off diagonal block structures represented higher within-subject
6 footprint correlations than between-subject. We observed that when subjects were ranked in
7 order of number of days of geolocation data recorded (red linear plots next to the x and y axes),
8 the prominence of the on-diagonal block structures, or the within-subject similarity, also
9 increased accordingly. Note, grouping was only for visualization and not used in any statistical
10 analysis or individual identification. **b)** We found that data quantity was correlated with footprint
11 distinctiveness ($r = 0.66, p < 0.001$). **c)** On the other hand, we found no evidence that data quality
12 (operationalized as the number of minutes missing in daily GPS data) was related to footprint
13 distinctiveness ($p = 0.38$). **d, e)** Similar results were found for accelerometer data. Specifically,
14 data quantity was significantly associated with accelerometer-based footprint distinctiveness ($r =$
15 $0.34, p < 0.05$), but data quality was not ($p = 0.85$).
16



1
2
3
4
5
6
7
8

Supplementary Figure 5 | Individual identification accuracy after thresholding subjects at varying levels of data quantity. As subjects' minimum data quantity increased, the range of footprint distinctiveness narrowed, driven by the increase of lower bound while the upper bound remained the same. Specifically, we observed a range of 4-99% when data quantity was thresholded at 20 days or lower, 17%-99% when thresholded between 30 to 90 days.

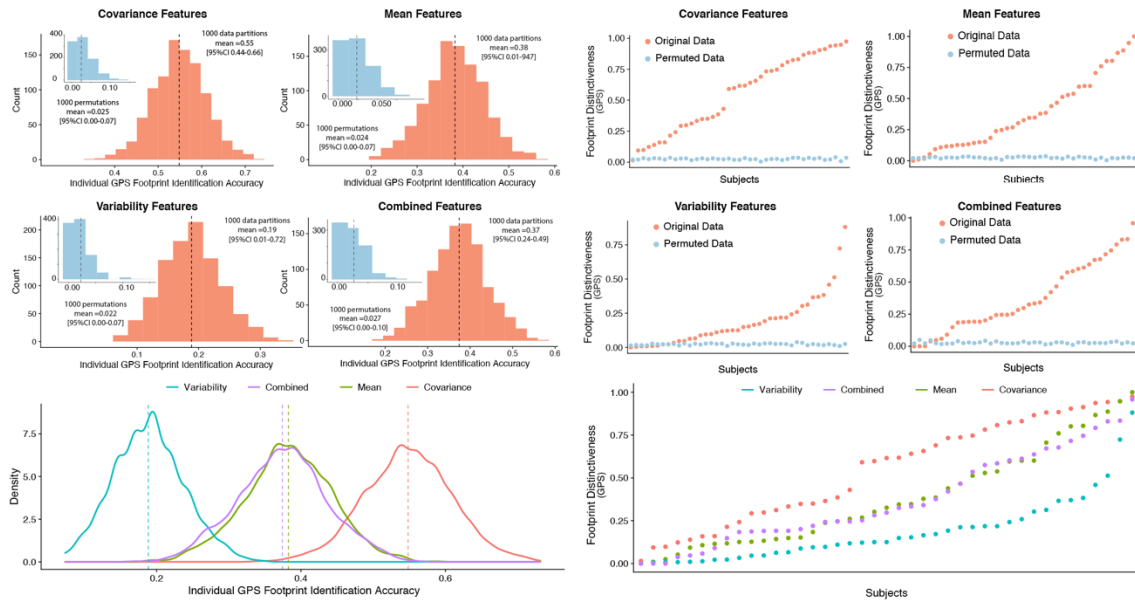
1



2
3
4
5
6
7
8
9

Supplementary Figure 6 | Developmental effects and sex differences in footprint

distinctiveness. **a)** While controlling for data quantity, geolocation-based footprint became more distinct with age (partial $r = 0.33$, $p < 0.05$), **b)** but it did not differ between sexes ($Cohen's\ d = 0.08$, $p = 0.81$). **c)** In contrast, accelerometer data did not exhibit any age effects (partial $r = 0.17$, $p = 0.68$), but females had significantly more distinct accelerometer-based footprint than males ($Cohen's\ d = 1.27$, $p < 0.001$).



1
 2 **Supplementary Figure 7 | Identifying individuals using alternative footprint features.** While
 3 our main analyses used the covariance of mobility features to identify individuals, we also
 4 evaluated alternative feature types. Specifically, we repeated the geolocation-based footprinting
 5 analyses using the daily mean value of each feature, the daily variability of each feature
 6 (operationalized as the root mean square of the successive differences), and a combined feature
 7 set including mean, variability, and covariance features. While all feature sets yielded significant
 8 individual identification accuracy compared to the null distribution, covariance features out-
 9 performed the alternative features by a wide margin (covariance: 54.8%, mean: 38.2%,
 10 variability: 18.9%, combined: 37.4%). This data suggests that covariance features best encoded
 11 individual-specific mobility footprints.
 12

1 **Reference**

2
3
4
5
6
7
8
9

1. Barnett, I. & Onnela, J.-P. Inferring mobility measures from GPS traces with missing data. *Biostatistics* **21**, e98–e112 (2018).
2. Vega, J. *et al.* RAPIDS: Reproducible Analysis Pipeline for Data Streams Collected with Mobile Devices. doi:10.2196/preprints.23246.
3. Panda, N. *et al.* Using Smartphones to Capture Novel Recovery Metrics After Cancer Surgery. *JAMA Surg.* **155**, 123–129 (2020).