

Mobile Footprinting: Linking Individual Distinctiveness in Mobility Patterns to Mood, Sleep, and Brain Functional Connectivity

Cedric Huchuan Xia

University of Pennsylvania

Ian Barnett

University of Pennsylvania

Tinashe Tapera

University of Pennsylvania

Zaixu Cui

State Key Laboratory of Cognitive Neuroscience and Learning-Beijing Normal University

Tyler Moore

Perelman School of Medicine <https://orcid.org/0000-0002-1384-0151>

Azeez Adebimpe

Annenberg Public Policy Center, University of Pennsylvania, Philadelphia <https://orcid.org/0000-0001-9049-0135>

Sage Rush-Goebel

University of Pennsylvania

Kayla Piiwaa

University of Pennsylvania

Kristin Murtha

University of Pennsylvania

Sophia Linguiti

University of Pennsylvania

Ellen Leibenluft

National Institute of Mental Health

Melissa Brotman

National Institute of Mental Health

Melissa Martin

University of Pennsylvania

Monica Calkins

University of Pennsylvania

David Roalf

University of Pennsylvania <https://orcid.org/0000-0002-1728-9782>

Daniel Wolf

University of Pennsylvania

Danielle Bassett

University of Pennsylvania <https://orcid.org/0000-0002-6183-4493>

David Lydon-Staley

University of Pennsylvania

Justin Baker

Harvard Medical School <https://orcid.org/0000-0002-6629-2195>

Lyle Ungar

University of Pennsylvania

Theodore Satterthwaite (✉ sattertt@penncmedicine.upenn.edu)

University of Pennsylvania <https://orcid.org/0000-0001-7072-9399>

Article

Keywords: smartphone, GPS, mobile phenotyping, fMRI, mood, affective instability

Posted Date: May 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-465623/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

3
4 Cedric Huchuan Xia^{1,2}, Ian Barnett³, Tinashe M. Tapera^{1,2}, Zaixu Cui^{1,2}, Tyler M. Moore^{1,2}, Azeez
5 Adebimpe^{1,2}, Sage Rush-Goebel^{1,2}, Kayla Piiwaa^{1,2}, Kristin Murtha^{1,2}, Sophia Linguiti^{1,2}, Ellen Leibenluft⁴,
6 Melissa A. Brotman⁴, Melissa Lynne Martin⁵, Monica E. Calkins^{1,2}, David R. Roalf^{1,2}, Daniel H. Wolf^{1,2,18},
7 Danielle S. Bassett^{1,5,6,7,8,9}, David M. Lydon-Staley^{5,10,20}, Justin T. Baker^{11,12}, Lyle Ungar^{13,14,15,16,17}, Theodore D.
8 Satterthwaite^{1,2,18,19}

- 9
10 1. Penn Lifespan Informatics and Neuroimaging Center, Department of Psychiatry,
11 Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
12 2. Penn/CHOP Lifespan Brain Institute, University of Pennsylvania, Philadelphia, PA
13 19104, USA
14 3. Department of Biostatistics, Epidemiology and Informatics, Perelman School of
15 Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
16 4. National Institute of Mental Health, Intramural Research Program, Bethesda, MD 20892
17 5. Department of Bioengineering, School of Engineering and Applied Science, University
18 of Pennsylvania, PA 19104, USA
19 6. Department of Physics and Astronomy, University of Pennsylvania, PA 19104, USA
20 7. Department of Electrical & Systems Engineering, University of Pennsylvania, PA 19104,
21 USA
22 8. Department of Neurology, Perelman School of Medicine, University of Pennsylvania,
23 Philadelphia, PA 19104, USA
24 9. Santa Fe Institute, Santa Fe, NM 87501, USA
25 10. Annenberg School of Communication, University of Pennsylvania, Philadelphia, PA
26 19104, USA
27 11. Department of Psychiatry, Harvard Medical School, Boston, MA 02115
28 12. McLean Institute for Technology in Psychiatry, McLean Hospital, Belmont, MA 02478
29 13. Department of Computer and Information Science, School of Engineering and Applied
30 Science, University of Pennsylvania, PA 19104, USA
31 14. Department of Bioengineering, School of Engineering and Applied Science, University
32 of Pennsylvania, PA 19104, USA
33 15. Department of Genomics and Computational Biology, Perelman School of Medicine,
34 University of Pennsylvania, Philadelphia, PA 19104, USA
35 16. Department of Operations, Information and Decisions, Wharton School, Philadelphia, PA
36 19104, USA
37 17. Department of Psychology, School of Arts and Sciences, Philadelphia, PA 19104, USA
38 18. Center for Biomedical Image Computation and Analytics, University of Pennsylvania,
39 Philadelphia, PA 19104, USA
40 19. Penn Statistics in Imaging and Visualization Center, Department of Biostatistics,
41 Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA 19104,
42 USA
43 20. Leonard Davis Institute for Health Economics, University of Pennsylvania, PA 19104,
44 USA
45
46

1 **ABSTRACT**

2

3 Mapping individual differences in behavior is fundamental to personalized neuroscience. Here,
4 we establish that statistical patterns of smartphone-based mobility features represent unique
5 “footprints” that allow individual identification. Critically, mobility footprints exhibit varying
6 levels of person-specific distinctiveness and are associated with individual differences in
7 affective instability, circadian irregularity, and brain functional connectivity. Together, this work
8 suggests that real-world mobility patterns may provide an individual-specific signature linking
9 brain, behavior, and mood.

10

11 **KEYWORDS:** smartphone, GPS, mobile phenotyping, fMRI, mood, affective instability

12

13

14

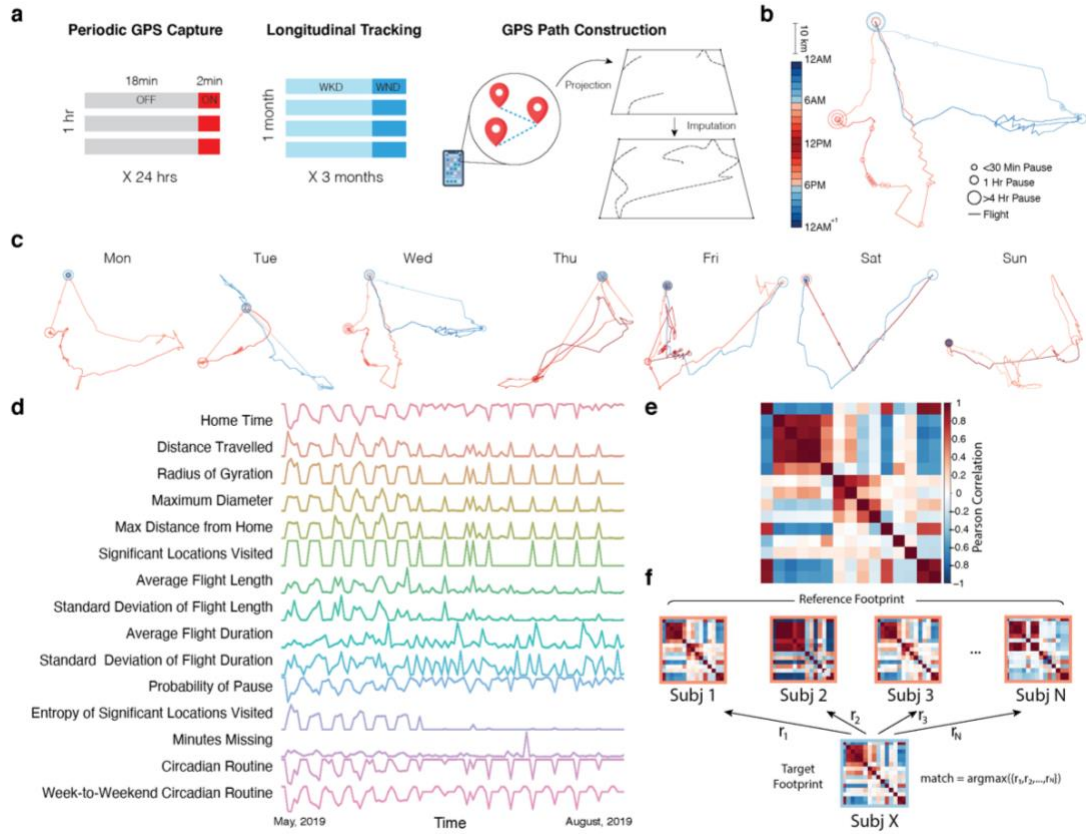
1 INTRODUCTION

2
3 Linking individual differences in behavior to brain function is a central task of behavioral
4 neuroscience¹. However, quantifying complex human behavior in real world settings remains a
5 challenge. One alternative to standard behavioral assessment is *digital phenotyping*, which uses
6 mobility data from personal smartphones to quantify moment-by-moment human behavior².
7 Prior work has associated geolocation features to important clinical outcomes in psychiatric
8 disorders such as bipolar disorder and schizophrenia³, and has linked accelerometer metrics to
9 post-surgical recovery^{4,5}. Furthermore, researchers have recently begun to capitalize on the
10 substantial variability of behavior assessed with digital phenotyping to link individual
11 differences in brain and behavior. For example, lower prefrontal activity during processing
12 negative emotions has been associated with individual exposure to urban green space⁶, while
13 greater functional coupling of the hippocampus and striatum has been linked to location
14 variability⁷.

15
16 While these studies suggest that digital phenotyping can be a powerful tool for studying
17 individual differences, it remains unknown whether mobility patterns are in fact *person-specific*.
18 Recent high-impact work has established that individual humans have unique patterns of
19 functional brain connectivity^{8,9}. The uniqueness of such brain-based “*fingerprints*” (also called
20 “connectotypes”¹⁰) have been associated with development, cognition, and psychiatric
21 conditions¹¹. Establishing analogous person-specific mobility patterns – or mobility “*footprints*”
22 – would constitute an important advance in behavioral neuroscience, and provide the foundation
23 for targeted, individual-specific interventions. Accordingly, here we test the hypothesis that
24 mobility patterns derived from personal smartphones can be used to create person-specific
25 behavioral footprints. Furthermore, we evaluate whether the distinctiveness of these footprints
26 was related to individual differences in mood, sleep, and brain functional connectivity.

27 28 29 RESULTS & DISCUSSION

30
31 As part of a study of trans-diagnostic affective instability in youth, we tracked 3,317
32 person-days of geolocation and 2,972 person-days of accelerometer data from 41 adolescents and
33 young adults (28 females; mean [s.d.] age = 23.4 [3.5] years, range 17–30 years) – approximately
34 3 months per individual (**Fig. 1a, Supplementary Fig. 1**). In this sample, 93% of participants
35 reported clinically significant affective instability in the context of psychiatric disorders
36 (especially borderline personality disorder; see **Supplementary Table 1**). After applying hot-
37 deck imputation to missing GPS data as implemented in the Smartphone Sensor Pipeline¹³, we
38 constructed the daily mobility trajectory for each participant (**Fig. 1b & c**; see Online Methods).
39 Instead of using raw coordinates that would allow trivial individual identification (and raise
40 privacy concerns) given a participant’s exact location, we extracted high-level summary statistics
41 of mobility features. These features (15 geolocation-based and 7 accelerometer-based) included
42 time spent at home, number of locations visited, and many others (**Fig. 1d, Supplementary**
43 **Table 2**).



1
2 **Figure 1 | Constructing personal mobility “footprints”.** a) We collected 3,317 person-days of mobility
3 sensing data via personal smartphones from 41 adolescents and young adults. Geolocation data were
4 recorded in cycles of 2min on and 18min off. Raw geolocation coordinates were de-identified via sphere-
5 to-2D standard space projection and were further imputed for missing data. b) For each individual, we
6 constructed daily personal mobility trajectories, which consist of flights (movement) and pauses
7 (stationary segments). Length of linear lines represents the duration of flights and size of circles
8 represents the duration of pauses. Warm and cold colors indicate daytime and nighttime, respectively. c)
9 A representative week of trajectories is shown, which demonstrates rich characteristics of personal
10 mobility patterns formed over time. d) We extracted timeseries of mobility statistics (e.g. daily time spent
11 at home) from geolocation and accelerometer data that parameterize movement characteristics over weeks
12 to months. The example represented all 110 days of participants’ geolocation metrics recorded. e) For
13 each individual, we constructed a covariance matrix from the mobility metric timeseries. Each cell of the
14 matrix was populated by the Pearson correlation between a given pair of mobility metrics. Warm and cold
15 colors indicate positive and negative correlations, respectively. f) We randomly divided data into two
16 equally sized parts, called the reference and target set. *Subj X* from the target set was matched to the
17 subject in the reference that had the highest correlations between their footprints ($\operatorname{argmax}(r_1, r_2, \dots, r_N)$). The
18 identification was considered correct when underlying data came from the same subject; otherwise, the
19 identification was considered incorrect. We quantified individual identification accuracy as the proportion
20 of correct identifications across the entire sample; this procedure was repeated 1,000 times across
21 different random partitions of the data.

22
23 When tracked over weeks to months, these timeseries of mobility statistics captured rich
24 characteristics of individual mobility patterns. One illustrative example of the sensitivity of the
25 timeseries to track mobility patterns is when COVID-19 pandemic hit the Philadelphia area

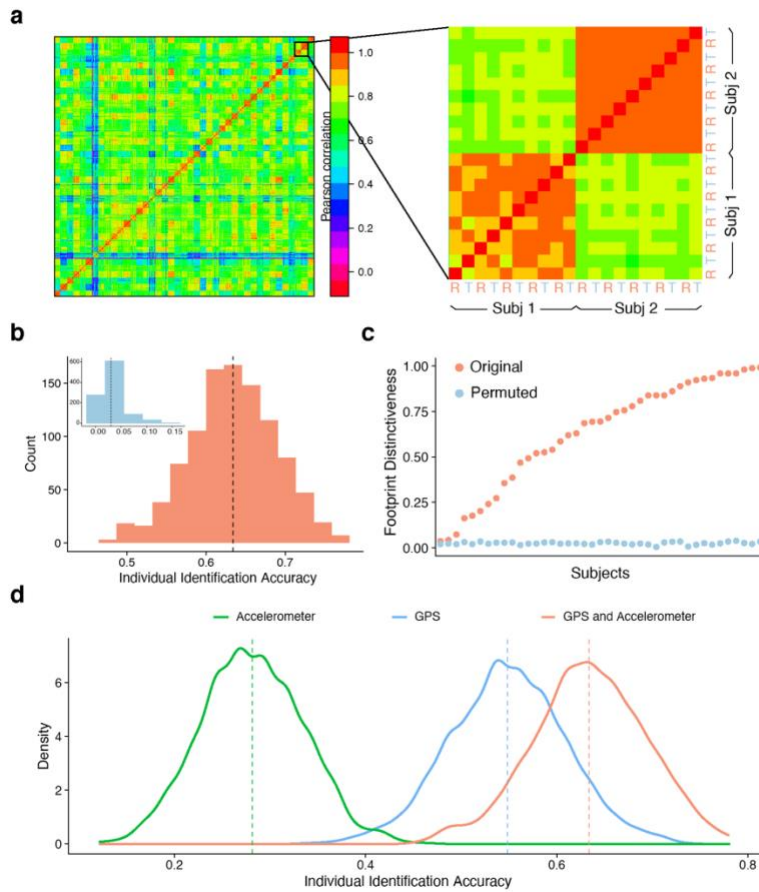
1 towards the end of the study period. Participants who were still engaged in active data collection
2 (n=3) exhibited dramatic shifts in mobility features (**Supplementary Fig. 2**). Of note, as the data
3 points during COVID-19 represented merely 1.1% of all data, the findings reported below did
4 not change significantly when these data were removed.

5
6 Drawing on prior work of brain connectome “fingerprinting,”^{8,10} we created a covariance
7 matrix of each participant’s fifteen geolocation-based and seven accelerometer-based mobility
8 features timeseries to identify individuals (**Fig. 1e**), akin to a person-specific mobility
9 “footprint.” Data from each individual is was partitioned into two groups: the target partition and
10 the reference partition. For each individual, the data in the target partition was separately
11 correlated with every individual’s data in the reference partition; this procedure yielded 41
12 correlation values. A correct identification was declared only when the maximum correlation
13 was from the data belonging to the same individual across the target and reference partitions
14 (**Fig. 1f**). In order to ensure that the random partitioning of the data did not impact results, this
15 matching procedure was then repeated for each individual 1000 times (Online Methods).

16
17 Initial inspection across random partitions of the data revealed that it was visually
18 apparent that there was substantially greater correlation between mobility footprints *within*
19 participants rather than *between* participants (**Fig. 2a**). Permutation testing on the entire sample
20 revealed that individuals could be successfully identified using their mobility footprints (p
21 < 0.001 ; **Fig. 2b**). Across 1,000 random data partitions, the mean individual identification
22 accuracy was 63%. Critically, this accuracy was far better than chance performance determined
23 by a permuted null distribution (mean: 3% accuracy; see **Fig 2b** inset).

24
25 Moving beyond aggregate measures of accuracy across the group, we next investigated
26 whether certain individuals could be consistently identified more accurately than others. Similar
27 to prior studies of brain connectome fingerprinting^{8,10,11}, we refer to this measure as an
28 individual’s “*footprint distinctiveness*”. Notably, individuals exhibited a wide distribution of
29 footprint distinctiveness, ranging from 4% to 99% (**Fig. 2c**). In other words, certain participants
30 had such distinct mobility patterns that it enabled correct identification nearly every single time;
31 other participants were difficult to identify. Nonetheless, permutation testing showed that all
32 participants had significant footprint distinctiveness compared to the null distribution.

33
34 As the group and individual level accuracy results reported thus far were based on the
35 combination of geolocation and accelerometer features, we next examined each feature set
36 separately. Individual footprint distinctiveness derived by geolocation was not correlated with
37 that of accelerometer ($r = 0.18$; $p = 0.26$). Interestingly, while accelerometer data alone yielded
38 lower identification accuracy (28%) than geolocation data (55%), combining these features
39 resulted in higher identification accuracy, suggesting that they encode complementary
40 information (**Fig. 2d**). Importantly, individual identification accuracy was stable across different
41 inclusion thresholds for data missingness and was robust to removal of individual mobility
42 features (**Supplementary Fig. 3**).

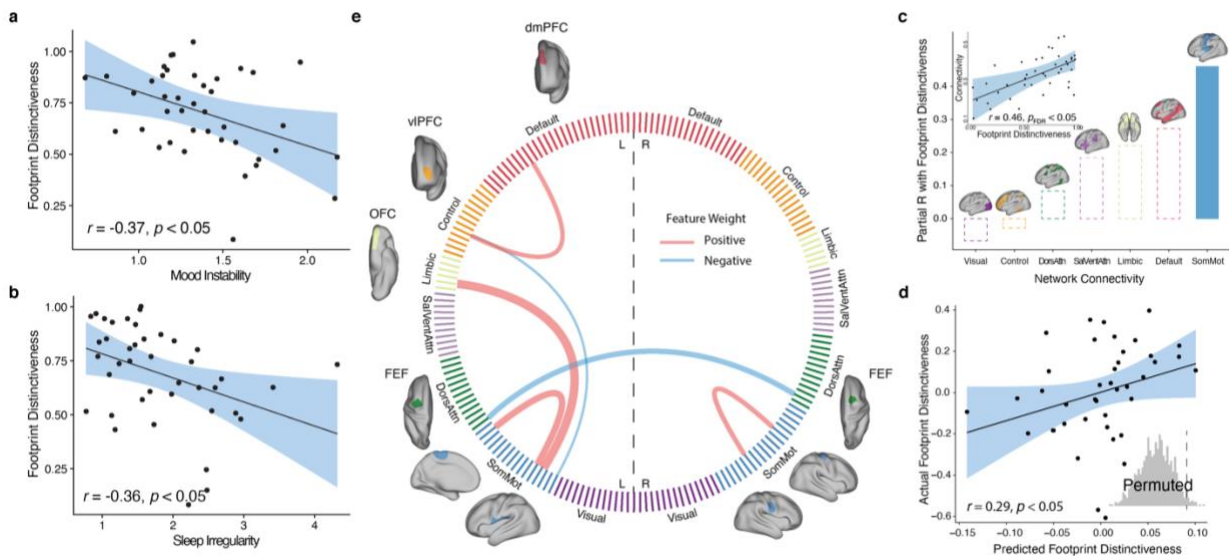


1
2 **Figure 2 | Identifying individuals using personal footprints.** a) As an initial step, we visualized the
3 similarity of mobility features across multiple random reference and target partitions (R & T in inset). It
4 was readily apparent that mobility features were more highly correlated within participants (on diagonal)
5 across data partitions than between participants (off diagonal). Note that this visualization was not used in
6 statistical analysis or individual identification. b) Across 1,000 random partitions, mobility footprinting
7 enabled successful individual identification (mean: 63%, S.D.: 6%). In contrast, the mean chance
8 accuracy from 1,000 permutation was 3% (inset, $p < 0.001$). c) For each individual, we calculated the
9 footprint distinctiveness, or the percentage of correct identification across the 1,000 random partitions of
10 the data. Ranked in ascending order, participants' footprint distinctiveness exhibited a wide range, from
11 4% to 99%. However, even the participant with the lowest identification distinctiveness was significantly
12 higher than the null distribution. d) Individual identification based on geolocation alone had higher
13 accuracy than accelerometer alone. However, they appeared to encode complementary features, as
14 performance was maximal when both measures were used in footprinting.

15
16
17 We next investigated participant factors that influenced footprint distinctiveness. We
18 found that data quantity (i.e. number of days recorded) was associated with footprint
19 distinctiveness (**Supplementary Fig. 4**). In contrast, the amount of missing data *within* a given
20 day was unrelated to footprint distinctiveness. Based on this result, all subsequent analyses of
21 individual differences related of footprint distinctiveness controlled for number of days of data

1 available. As a next step, we evaluated whether footprint distinctiveness was related to age or sex
 2 in our sample of adolescents and young adults. We found that geolocation-based footprints
 3 became more distinct with age across the transition from adolescence to adulthood (partial $r =$
 4 $0.33, p < 0.05$, **Supplementary Fig. 5**). Furthermore, female sex was associated with higher
 5 accelerometer-based footprint distinctiveness (*Cohen's d* = 1.27, $p < 0.001$, **Supplementary Fig.**
 6 **5**).

8 We next evaluated how footprint distinctiveness was related to a key domain of
 9 psychopathology: affective instability. Affective instability is a major feature of many
 10 psychiatric disorders¹⁴, including borderline personality disorder. Affective instability is
 11 particularly prominent in youth¹⁵, and is an important predictor of suicide¹⁶. However, affective
 12 instability is often challenging to quantify using standard tools as it is fundamentally a dynamic
 13 measure¹⁷. We capitalized on participant ratings of multiple mood features collected three times
 14 a day for two weeks using ecological momentary assessment in order to quantify affective
 15 instability. We hypothesized that individuals who had less predictable patterns of mobility (i.e.,
 16 reduced footprint distinctiveness) would have higher levels of affective instability. While
 17 controlling for data quantity, age, sex, and the mean of mood ratings, we found that affective
 18 instability (measured by root mean square of successive differences¹⁸) was associated with
 19 reduced footprint distinctiveness (partial $r = -0.37, p < 0.05$, **Fig. 3a**). Furthermore, given well-
 20 established links between sleep disturbance and mood disorders¹⁹, we also evaluated whether
 21 variability in sleep duration was also associated with footprint distinctiveness. While controlling
 22 for covariates as above, we found that variability in sleep duration was similarly associated with
 23 reduced footprint distinctiveness (partial $r = -0.36, p < 0.05$, **Fig. 3b**).



26
 27
 28 **Figure 3 | Individual footprint distinctiveness is associated with affective instability, sleep**
 29 **irregularity, and patterns of brain functional connectivity. a)** Greater affective instability, measured
 30 by root mean square of successive differences in mood measures from ecological momentary assessment
 31 items acquired three times a day, was associated with reduced footprint distinctiveness ($r = -0.37, p <$
 32 0.05), after controlling for data quantity, age, sex, and mean level of mood ratings. **b)** Similarly, we found
 33 that increased variability in sleep duration was associated with reduced footprint distinctiveness ($r = -$

1 0.36, $p < 0.05$), after controlling for covariates. **c**) Across functional brain networks, only greater
2 connectivity within the somatomotor network had a significant association with footprint distinctiveness
3 ($r = 0.46$, $p < 0.05$, corrected for multiple comparisons with the false discovery rate). **d**) Patterns of brain
4 functional connectivity significantly predicted individual footprint distinctiveness using leave-one-out
5 cross-validation ($r = 0.29$, inset: permutation-based $p = 0.025$). **e**) Six network edges consistently
6 contributed to the sparse regression model. These edges included greater connectivity within
7 somatomotor network, reduced connectivity between left and right frontal eye fields (FEF), increased
8 connectivity between the somatomotor network and the left orbital frontal cortex (OFC) in the limbic
9 network, as well as increased connectivity between the vIPFC (ventrolateral prefrontal cortex in the
10 frontoparietal network) and the dmPFC (dorsolateral prefrontal cortex in the default mode network). Cord
11 thickness reflects the weights in the model, reflecting each edge's contribution to the prediction; cord
12 color indicates the sign of the weights.
13
14

15 As a final step, we investigated whether footprint distinctiveness was related to patterns
16 of functional connectivity. Initially, we examined associations with a simple summary measure
17 of high-dimensional functional connectivity data: the mean connectivity within each of seven
18 canonical large-scale functional networks²⁰. While controlling for covariates as above (as well as
19 in-scanner motion) and correcting for multiple comparisons with the false discovery rate, we
20 found that footprint distinctiveness was associated with greater connectivity within the
21 somatomotor network ($r = 0.46$, $p_{\text{fdr}} = 0.03$, **Fig. 3c**). Previous work has demonstrated that
22 somatomotor network connectivity develops over the lifespan (years)²¹ and is altered acutely
23 (days) during limb disuse²²; our results further suggest that mobility patterns over weeks-months
24 can be predicted by somatomotor network connectivity.
25

26 Lastly, we moved beyond the simple summary measure of mean network connectivity
27 and investigated whether complex multivariate patterns of functional connectivity could predict
28 footprint distinctiveness in unseen data. Given that there were far larger number of features than
29 participants, we used regularized regression with leave-one-out cross-validation and nested
30 parameter tuning, followed by permutation testing to determine significance (Online Methods).
31 We found that multivariate patterns of functional connectivity could predict footprint
32 distinctiveness in unseen data ($r = 0.29$, $p = 0.025$; **Fig. 3d**). The predictive model yielded results
33 that aligned with the mass-univariate analyses (**Fig. 3c**), suggesting that the multivariate model
34 was driven in part by features linked to somatomotor network (67% of edges selected by the
35 model). Moreover, this model also revealed important features beyond the motor system,
36 including increased connectivity between the frontoparietal and default mode system (**Fig. 3e**).
37

38 Taken together, these results establish that mobility patterns collected from smartphones
39 can be used to create a person-specific footprint. Notably, the distinctiveness of this footprint
40 increased with age, was reduced in association with both affective instability and circadian
41 irregularity, and was related to patterns of functional brain connectivity. These results align with
42 impactful prior work on “connectome fingerprinting”⁸ (or “connectotyping”¹⁰), which have
43 shown that individuals can be identified based on their pattern of functional connectivity.
44 Interestingly, result from these prior studies have shown that – like the footprint distinctiveness
45 examined here – connectome distinctiveness increases with age and is reduced in association
46 with psychiatric symptoms¹¹.
47

1 Our finding that footprint distinctiveness is related to data quantity recalls recent work
2 demonstrating that the ability to delineate person-specific functional brain networks is dependent
3 in large part on the quantity of data available^{23,24}. However, while accruing large amounts of
4 functional imaging data is often difficult and expensive, passive collection of long timeseries of
5 mobility data is both tolerable for participants and inexpensive. The high degree of scalability
6 enabled by ubiquitous usage of personal smartphones will allow future studies to test the
7 generalizability of these findings across different age groups and clinical samples. Moving
8 forward, mobility-based digital biomarkers that combine objective measurement and individual-
9 specific analysis of behavior may accelerate the advances in personalized diagnostics for diverse
10 psychiatric illnesses.

11 12 13 REFERENCES

- 14
15 1. Insel, T. R. Digital phenotyping: Technology for a new science of behavior. *JAMA* **318**,
16 1215–1216 (2017).
- 17 2. Onnela, J. P. Opportunities and challenges in the collection and analysis of digital
18 phenotyping data. *Neuropsychopharmacology* **46**, 45–54 (2021).
- 19 3. Fraccaro, P. *et al.* Digital biomarkers from geolocation data in bipolar disorder and
20 schizophrenia: a systematic review. *J. Am. Med. Informatics Assoc.* **26**, 1412–1420
21 (2019).
- 22 4. Cote, D. J., Barnett, I., Onnela, J.-P. & Smith, T. R. Digital Phenotyping in Patients with
23 Spine Disease: A Novel Approach to Quantifying Mobility and Quality of Life. *World*
24 *Neurosurg.* **126**, e241–e249 (2019).
- 25 5. Panda, N. *et al.* Using Smartphones to Capture Novel Recovery Metrics After Cancer
26 Surgery. *JAMA Surg.* **155**, 123–129 (2020).
- 27 6. Tost, H. *et al.* Neural correlates of individual differences in affective benefit of real-life
28 urban green space exposure. *Nat. Neurosci.* **22**, 1389–1393 (2019).
- 29 7. Heller, A. S. *et al.* Association between real-world experiential diversity and positive
30 affect relates to hippocampal–striatal functional connectivity. *Nat. Neurosci.* **23**, 800–804
31 (2020).
- 32 8. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using
33 patterns of brain connectivity. *Nat. Neurosci.* **18**, 1–11 (2015).
- 34 9. Horien, C., Shen, X., Scheinost, D. & Constable, R. T. The individual functional
35 connectome is unique and stable over months to years. *Neuroimage* **189**, 676–687 (2019).
- 36 10. Miranda-Dominguez, O. *et al.* Connectotyping: Model Based Fingerprinting of the
37 Functional Connectome. *PLoS One* **9**, e111048 (2014).
- 38 11. Kaufmann, T. *et al.* Delayed stabilization and individualization in connectome
39 development are related to psychiatric disorders. *Nat. Neurosci.* **20**, 513–515 (2017).
- 40 12. González, M. C., Hidalgo, C. A. & Barabási, A. L. Understanding individual human
41 mobility patterns. *Nature* **453**, 779–782 (2008).
- 42 13. Barnett, I. & Onnela, J.-P. Inferring mobility measures from GPS traces with missing data.
43 *Biostatistics* **21**, e98–e112 (2018).
- 44 14. Koenigsberg, H. W. Affective Instability: Toward an Integration of Neuroscience and
45 Psychological Perspectives. *J. Pers. Disord.* **24**, 60–82 (2010).
- 46 15. Tragesser, S. L., Solhan, M., Schwartz-Mette, R. & Trull, T. J. The Role of Affective

1 Instability and Impulsivity in Predicting Future BPD Features. *J. Pers. Disord.* **21**, 603–
2 614 (2007).

3 16. Yen, S. *et al.* Borderline Personality Disorder Criteria Associated With Prospectively
4 Observed Suicidal Behavior. *Am. J. Psychiatry* **161**, 1296–1298 (2004).

5 17. Trull, T. J. *et al.* Affective instability: Measuring a core feature of borderline personality
6 disorder with ecological momentary assessment. *J. Abnorm. Psychol.* **117**, 647–661
7 (2008).

8 18. Jahng, S., Wood, P. K. & Trull, T. J. Analysis of affective instability in ecological
9 momentary assessment: Indices using successive difference and group comparison via
10 multilevel modeling. *Psychol. Methods* **13**, 354–375 (2008).

11 19. Harvey, A. G. Sleep and Circadian Rhythms in Bipolar Disorder: Seeking Synchrony,
12 Harmony, and Regulation. *Am. J. Psychiatry* **165**, 820–829 (2008).

13 20. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic
14 Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).

15 21. Power, J. D., Fair, D. A., Schlaggar, B. L. & Petersen, S. E. The Development of Human
16 Functional Brain Networks. *Neuron* **67**, 735–748 (2010).

17 22. Newbold, D. J. *et al.* Plasticity and Spontaneous Activity Pulses in Disused Human Brain
18 Circuits. *Neuron* **107**, 580-589.e6 (2020).

19 23. Gratton, C. *et al.* Functional Brain Networks Are Dominated by Stable Group and
20 Individual Factors, Not Cognitive or Daily Variation. *Neuron* **98**, 439-452.e5 (2018).

21 24. Elliott, M. L. *et al.* General functional connectivity: Shared features of resting-state and
22 task fMRI drive reliable and heritable individual differences in functional brain networks.
23 *Neuroimage* **189**, 516–532 (2019).

24
25

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

4 **ONLINE METHODS**

6 **Participants**

8 A sample of 41 adolescents and young adults (28 females; mean (s.d.) age = 23.4 (3.5) years,
9 range 17–30 years) were enrolled as part of a study of affective instability in youth. Participants
10 were recruited via the Penn/CHOP Lifespan Brain Institute or through the Outpatient Psychiatry
11 Clinic at the University of Pennsylvania. Of these 41 participants, 38 participants met criteria for
12 Axis I psychiatric diagnosis based on a semi-structured clinical interview¹; 33 met criteria for
13 more than one disorder (**Supplementary Table 1**). Additionally, 16 of the 41 participants met
14 criteria for a personality disorder (mainly borderline personality disorder) based on assessment
15 with the SCID-II¹. All participants provided informed consent to all study procedures; for
16 minors, the parent or guardians provided informed consent and the minor assented as well. This
17 study was approved by the University of Pennsylvania Institutional Review Board.

20 **Mobility data acquisition**

22 Global Positioning System (GPS) geolocation data were acquired via the Beiwe platform².
23 Participants were asked to download the Beiwe application on their personal smartphone. The
24 application recorded the location of the participant's phone in latitude, longitude, and altitude, as
25 well as the precision of that measure. To conserve battery and minimize degradation of the phone
26 performance, Beiwe was designed to track participant's geolocation in a periodic fashion.
27 Specifically, Beiwe tracked GPS for 2 minutes every 20 minutes, resulting in 144 minutes of
28 data recording and 1296 minutes of dormancy in a 24-hour cycle. Due to user and device related
29 factors in the naturalistic setting, such as phone powered off, no cell signal, or airplane mode,
30 longer periods of recording dormancy were possible. Mobility data were automatically uploaded
31 via WiFi to a cloud-based data management system daily.

33 In total, 3,317 days of GPS tracking across all participants were obtained (mean (s.d.) = 77 (26)
34 days, range 14–132 days, see **Supplementary Figure 1**). After removing the first and last days
35 of each participant's study period when only partial data were recorded and days containing no
36 data, the remaining data available for analysis had 3,156 days.

38 Accelerometer data were also acquired via the Beiwe platform. The application recorded the
39 participants' acceleration in three cardinal axes (x, y, and z) in m/s². In total, 2,972 days of
40 accelerometer data were obtained across all participants (mean (s.d.) = 74 (32) days, range 15-
41 134 days). After removing the first and last days of each participant's study period when only
42 partial data were recorded, the remaining data available for analysis had 2972 days.

44 **Mobility data processing**

1 *GPS data preprocessing*

2
3 Raw GPS data were processed using the Smartphone Sensor Pipeline³, a validated pipeline
4 specifically designed to handle GPS data while accounting for data missingness. First, each
5 subject's GPS longitude and latitude coordinates on the spherical Earth's surface were
6 transformed to a standardized two-dimensional Cartesian plane, thus deidentifying subject's real-
7 world locations. Second, the data were converted to a sequence of flights and pauses, where
8 flights were defined as segments of linear movements and pauses were defined as periods of no
9 movement. Finally, missing flights and pauses were then imputed by the hot-deck method⁴,
10 which resamples from observed events over each missing interval.

11 12 *Mobility metrics calculation*

13
14 Using the constructed subject mobility traces and the Smartphone Sensor Pipeline, 15 GPS-based
15 mobility metrics were calculated for each day of recording, defined as midnight to midnight. See
16 Barnett et al. for details³. An additional seven accelerometer-based mobility metrics were
17 calculated for each day of recording. These were implemented according to methods described in
18 the RAPIDS pipeline⁵. See **Supplementary Table 2** for definitions of each metric.

19 20 *Mobility footprint construction*

21
22 Inspired by person-specific connectome fingerprints^{6,7}, we constructed a mobility footprint for
23 each participant using the covariance matrix of mobility metrics. First, we extracted the mobility
24 metric time series by concatenating the daily mobile metric output from the Smartphone Sensor
25 Pipeline. Then we computed the pairwise Pearson correlation for all the mobility metrics to
26 construct a covariance matrix. The nodes of the network were the mobility metrics, and the edges
27 of the network were the Pearson correlation coefficients between metrics. We refer to the
28 resulting covariance matrix as the "Mobility Footprint." This procedure was carried out
29 separately for GPS- and accelerometer-based mobility data. For the main analysis, the upper
30 triangle of the resulting covariance matrices from GPS and accelerometer metrics were
31 concatenated and were used as input features for the individual identification procedure. We also
32 repeated the identification procedure using GPS or accelerometer features alone.

33
34 As a sensitivity analysis to test performance of alternative features for individual identification,
35 we also computed the mean and the stability of each measure and used these features to identify
36 participants. Stability was defined as the root mean square of the successive differences
37 (RMSSD)⁸ of each measure (**Supplementary Figure 6**).

38 39 *Individual identification procedure ("footprinting")*

40
41 We randomly partitioned each subject's data into two equally sized parts, named the "reference"
42 and the "target", respectively⁶. The objective of the individual identification procedure was to
43 match the subject from the target group to the same one in the reference group. For a given
44 subject, S , we computed the Pearson correlation (r) between that subject's features in the target
45 group, S_T , and everyone's features in the reference group, $S_R 1, S_R 2, S_R N$, where N is the total
46 number of participants.

1
2 Individual identification was operationalized as the maximum of the resulting r_1, r_2, \dots, r_N . In
3 other words, when the subject in the reference group having the mobility features that maximally
4 correlated with that of the target subject, these two participants were declared correctly matched:

$$5 \quad 6 \quad 7 \quad S_M = \operatorname{argmax}(r_1, r_2, \dots, r_N).$$

8 The individual identification accuracy was the number of correct identifications divided by the
9 total number of random data partitions P :

$$10 \quad 11 \quad \text{individual identification accuracy} = \frac{\sum_{i=1}^N \begin{cases} 1, & \text{if } S_T i = S_M i \\ 0, & \text{if } S_T i \neq S_M i \end{cases}}{N}.$$

12 The above individual identification procedure was repeated 1,000 times, each time with a new
13 random data partition (P). We calculated the average individual identification accuracy across
14 the 1,000 runs, which yielded a distribution of sample-wise identification accuracy. Furthermore,
15 we also calculated the accuracy for each participant, defined as the number of correct
16 identifications for that specific participant divided by the number of data partitions (B). We refer
17 to this participant-specific identification accuracy as the individual footprint distinctiveness:

$$18 \quad 19 \quad 20 \quad \text{individual footprint distinctiveness} = \frac{\sum_{i=1}^B \begin{cases} 1, & \text{if } P_T i = P_M i \\ 0, & \text{if } P_T i \neq P_M i \end{cases}}{B},$$

21 where $P_T i$ is target in a partition for subject i , and $P_M i$ is matched subject. We conducted the
22 individual identification procedure using the covariance matrix of the GPS data, accelerometer
23 data, as well as the combined feature set. Sensitivity analyses examined the mean and variance
24 of each feature.

25 26 27 *Similarity matrix construction*

28 To visualize the individual footprint distinctiveness, we constructed similarity matrices between
29 participants' mobility covariance features⁹. First, we concatenated the daily mobility metrics for
30 a participant from multiple random data partitions. Next, a similarity matrix was constructed by
31 computing the Pearson correlation coefficients between every pair of participants. The resulting
32 matrix was a symmetric matrix, where the nodes were each participant and the edges were the
33 correlation coefficients between any two participant's mobility metrics. This grouping procedure
34 was performed solely for visualization, highlighting the within-individual, across-partition block
35 structures on the diagonal of the matrix. This grouping was not used in any statistical analysis.

36 37 38 *Permutation testing*

39 To assess the statistical significance of individual identification accuracy, we used a permutation
40 testing procedure to create a null distribution of accuracy. Specifically, we randomly scrambled
41 the identity of the daily mobility metrics, thus disrupting the linkage between the mobility data
42 and the corresponding participant. We repeated the individual identification procedure for each
43

1 random permutation. The empirical p -value was then calculated as the proportion of times when
2 the permuted data yielded higher accuracy than the original data:

$$P_{\text{permutation}} = \frac{\sum_{i=1}^M \begin{cases} 1, & \text{if } A_i \geq A_{\text{original}} \\ 0, & \text{if } A_i < A_{\text{original}} \end{cases}}{M},$$

3
4
5 where A is the individual identification accuracy, and M is the total permutations.
6
7
8

9 *Sensitivity analysis of data missingness*

10
11 To understand the effect of data missingness on our ability to identify participants' mobility
12 footprint, we conducted sensitivity analyses that used four sets of data constructed using different
13 thresholds for data missingness^{3,10}. Specifically, we applied four thresholds with diminishing
14 tolerance for the number of missing samples (i.e., minutes recorded) in a day's worth of data to
15 be included in analysis (**Supplementary Figure 1**). At the 100th percentile level, which
16 corresponded to retaining all available days except for those with all data missing (or 1,440
17 minutes), 79 recording days were removed, which resulted 3,156 days remaining for analysis. At
18 the 90th percentile, a further 216 days were removed, yielding 2,940 days for analysis. At 80th
19 percentile, a further 356 days were removed, resulting in 2,584 days for analysis. Finally, at 75th
20 percentile, a further 171 days were removed, resulting in 2,413 days remaining for analysis.
21 Using these four sub-samples constructed with different inclusion criteria, we then repeated the
22 individual identification procedure and permutation testing as described above.
23

24 *Feature lesion analysis*

25
26 To further investigate the influence of any single feature's influence on the individual
27 identification accuracy, we conducted a feature lesion analysis. We sequentially removed one
28 metric (out of the total 15 geolocation mobility metrics available) and constructed a new
29 covariance matrix which had one node (and 14 edges) less than the original feature covariance
30 matrix. Using this reduced feature set, we repeated the individual identification and permutation
31 testing procedures as described above (**Supplementary Figure 3**).
32

33 *Ecological momentary assessment*

34
35 Using the Beiwe platform application on personal smartphones, participants completed daily
36 questionnaires specifically designed to assess mood variability at three timepoints throughout the
37 day¹¹. In each survey, participants rated on a scale from 1 ("not at all") to 7 ("extremely") of
38 their endorsement of seven statements assessing mood variability, aggression, impulsivity, and
39 self-esteem since the last time they had answered the survey to capture their mood
40 (**Supplementary Table 3**). All seven items were concatenated to create an overall mood scale.
41 Additionally, every morning, participants were also asked about their sleep patterns and quality
42 from the night before. To quantify the variability of answers to the mood survey, we calculated
43 the root mean square of successive differences (RMSSD) between concatenated answers.
44 Similarly, we also calculated the RMSSD of sleep duration as a measurement of its stability.
45

1 We built a generalized additive model (GAM) to investigate the association between mood and
2 sleep duration stability while accounting for covariate effects including data quantity, sex, age,
3 and mean levels of the measure. Age was modeled using penalized splines within GAM using
4 restricted maximum likelihood (REML) to estimate linear and nonlinear developmental effects
5 without over-fitting the data^{12,13}.

7 **Functional Connectivity Analysis**

9 *Imaging Acquisition*

11 As previously described¹⁴, structural and functional MRI scans were acquired using in a single
12 session on a clinically-approved 3 Tesla Siemens Prisma (Erlangen, Germany) quadrature body-
13 coil scanner and a Siemens receive-only 64-channel head coil at the Hospital of the University of
14 Pennsylvania. Prior to functional MRI acquisitions, a 5-min magnetization-prepared, rapid
15 acquisition gradient-echo T1-weighted (MPRAGE) image (TR = 1810 ms; TE = 3.45 ms; TI =
16 1100 ms, FOV = 180 × 240 mm², matrix = 192 × 256, 160 slices, effective voxel resolution =
17 0.9375 × 0.9375 × 1 mm³) was acquired. We used one resting-state (1200 volumes) scan as part
18 of this study. All fMRI images were acquired with the same multi-band, interleaved multi-slice,
19 gradient-echo, echo planar imaging (GE-EPI) sequence sensitive to BOLD contrast with the
20 following parameters: TR = 500 ms; TE = 25 ms; multiband acceleration factor = 6, flip angle =
21 30°; FOV = 192 × 192 mm²; matrix = 64 × 64; 48 slices; slice thickness/gap = 3/0 mm, effective
22 voxel resolution = 3.0 × 3.0 × 3.0 mm³.

23 *Image Processing*

25 All preprocessing was performed using fMRIPrep 20.0.7¹⁵, which is based on Nipype 1.4.2¹⁶,
26 and XCP Engine^{17,18} (PennBBL/xcEngine: atlas in MNI2009 Version 1.2.3; Zenodo:
27 <http://doi.org/10.5281/zenodo.4010846>). The T1-weighted (T1w) image was corrected for
28 intensity non-uniformity (INU) with N4BiasFieldCorrection¹⁹, distributed with ANTs 2.2.0²⁰,
29 and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped
30 with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using
31 OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF),
32 white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using
33 FAST in FSL 5.0.9²¹. Volume-based spatial normalization to MNI2009c standard space was
34 performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-
35 extracted versions of both the T1w reference and the T1w template.

36 BOLD runs were first slice-time corrected using 3dTshift from AFNI 20160207²² and then
37 motion corrected using mcflirt (FSL 5.0.9)²¹. A fieldmap was estimated based on a phase-
38 difference map calculated with a dual-echo GRE sequence, processed with a custom workflow of
39 SDCFlows inspired by the epidewarp.fsl script and further improvements in HCP Pipelines²³.
40 The fieldmap was then co-registered to the target EPI reference run and converted to a
41 displacement field map with FSL's fugue and other SDCflows tools. Based on the estimated
42 susceptibility distortion, a corrected BOLD reference was calculated for a more accurate co-
43 registration with the anatomical reference. The BOLD reference was then co-registered to the

1 T1w reference using bregister (FreeSurfer) which implements boundary-based registration²⁴.
 2 Co-registration was configured with nine degrees of freedom to account for distortions remaining
 3 in the BOLD reference. Six head-motion parameters (corresponding rotation and translation
 4 parameters) were estimated before any spatiotemporal filtering using mcflirt. Finally, the motion
 5 correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and
 6 T1w-to-template (MNI) warp were concatenated and applied to the BOLD timeseries in a single
 7 step using antsApplyTransforms (ANTs) with Lanczos interpolation.

8
 9 After pre-processing with fMRIPrep, confound regression was carried out in XCP Engine.
 10 Preprocessed timeseries were despiked and then de-noised using a 36-parameter confound
 11 regression model that has been shown to minimize the impact of motion artifact²⁵. Specifically,
 12 the confound regression model included the six framewise estimates of motion, the mean signal
 13 extracted from eroded white matter and cerebrospinal fluid compartments, the global signal, the
 14 derivatives of each of these nine parameters, and quadratic terms of each of the nine parameters
 15 as well as their derivatives. Both the BOLD-weighted time series and the confound regressor
 16 timeseries were temporally filtered simultaneously using a first-order Butterworth filter with a
 17 passband between 0.01 and 0.08 Hz to avoid mismatch in the temporal domain²⁶. Confound
 18 regression was performed using AFNI's 3dTproject. Note that in-scanner head motion was also
 19 included as a covariate in all regression models (see below).

20
 21 *Functional network and community connectivity*

22
 23 Functional connectivity between each pair of brain regions was quantified as the Fisher-
 24 transformed Pearson correlation coefficient between the mean regional BOLD time series. For
 25 each participant, a 200×200 weighted adjacency matrix encoding the connectome was
 26 constructed²⁷. Each node was assigned to one of seven canonical functional brain modules or
 27 communities defined by Yeo et al²⁸.

28
 29 The within-community connectivity is defined as

30
$$\frac{\sum_{j,j' \in C_k} A_{jj'}^i}{|C_k| \times (|C_k| - 1)}$$

31
 32 where $A_{jj'}^i$ is the weighted edge strength between the node j and node j' , both of which belong to
 33 the same community C_k , for the i -th subject. The cardinality of the community assignment
 34 vector, C_k , represents the number of nodes in the k -th community²⁹.

35
 36 *Mass-univariate analysis*

37
 38 For each of the seven canonical networks, we fit generalized additive model (GAM) to
 39 investigate the relationship between within-network connectivity and footprint distinctiveness,
 40 while controlling for in-scanner motion, mobility data quantity, sex, and age. Specifically, we
 41 used penalized splines using restricted maximum likelihood (REML) within GAM to estimate
 42 linear and nonlinear age-related changes^{12,13}. We controlled for multiple comparisons using the
 43 False Discovery Rate ($Q < 0.05$).

Predicting footprint distinctiveness using functional connectivity

We fit a penalized regression model to predict footprint distinctiveness using brain functional connectivity⁶. In each iteration of leave-one-out cross-validation, one subject was left out as the testing set and the rest the training set. Using the training set, we computed residualized footprint distinctiveness from a GAM model with covariates as above (linear terms for in-scanner motion, data quantity, sex; age was modeled with as a penalized spline). Then we fit a lasso regression model to predict the residualized footprint distinctiveness using a sparse collection of functional connectivity edges. L1 lasso hyperparameter was tuned in a nested leave-one-out fashion. Next, we calculated the predicted footprint distinctiveness for the unseen subject in the testing set. After all iterations, we obtained predicted footprint distinctiveness for all participants and then calculated the Pearson correlation between the actual footprint distinctiveness and predicted values.

Code availability and data access

The code for GPS data preprocessing, mobility metric extraction, individual identification, additional analysis, and data visualization is available in R on github:

<https://github.com/PennLINC/footprinting>

Code notebook is available at: <https://rpubs.com/hxia/footprinting>

Data is available upon request.

References

1. Calkins, M. E. *et al.* Persistence of psychosis spectrum symptoms in the Philadelphia Neurodevelopmental Cohort: a prospective two-year follow-up. *World Psychiatry* **16**, 62–76 (2017).
2. Torous, J., Kiang, M. V, Lorme, J. & Onnela, J.-P. New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research. *JMIR Ment. Heal.* **3**, e16 (2016).
3. Barnett, I. & Onnela, J.-P. Inferring mobility measures from GPS traces with missing data. *Biostatistics* **21**, e98–e112 (2018).
4. Andridge, R. R. & Little, R. J. A. A Review of Hot Deck Imputation for Survey Non-response. *Int. Stat. Rev.* **78**, 40–64 (2010).
5. Vega, J. *et al.* RAPIDS: Reproducible Analysis Pipeline for Data Streams Collected with Mobile Devices. doi:10.2196/preprints.23246.
6. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1–11 (2015).
7. Kaufmann, T. *et al.* Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nat. Neurosci.* **20**, 513–515 (2017).
8. Shaffer, F. & Ginsberg, J. P. An Overview of Heart Rate Variability Metrics and Norms.

- 1 *Front. Public Heal.* **5**, 258 (2017).
- 2 9. Gratton, C. *et al.* Functional Brain Networks Are Dominated by Stable Group and
3 Individual Factors, Not Cognitive or Daily Variation. *Neuron* **98**, 439–452.e5 (2018).
- 4 10. Kiang, M. V *et al.* Sociodemographic Characteristics of Missing Data in Digital
5 Phenotyping. *medRxiv* 2020.12.29.20249002 (2021) doi:10.1101/2020.12.29.20249002.
- 6 11. Shiffman, S., Stone, A. A. & Hufford, M. R. Ecological Momentary Assessment. *Annu.*
7 *Rev. Clin. Psychol.* **4**, 1–32 (2008).
- 8 12. Wood, S. N. Fast stable restricted maximum likelihood and marginal likelihood estimation
9 of semiparametric generalized linear models. *J. R. Stat. Soc. Ser. B (Statistical Methodol.*
10 **73**, 3–36 (2011).
- 11 13. Wood, S. N. Stable and Efficient Multiple Smoothing Parameter Estimation for
12 Generalized Additive Models. *J. Am. Stat. Assoc.* **99**, 673–686 (2004).
- 13 14. Jirsaraie, R. J. *et al.* Accelerated cortical thinning within structural brain networks is
14 associated with irritability in youth. *Neuropsychopharmacology* **44**, 2254–2262 (2019).
- 15 15. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat.*
16 *Methods* **16**, 111–116 (2019).
- 17 16. Gorgolewski, K. *et al.* Nipype: A Flexible, Lightweight and Extensible Neuroimaging
18 Data Processing Framework in Python. *Front. Neuroinform.* **5**, 13 (2011).
- 19 17. Ćirić, R. *et al.* PennBBL/xcpcEngine: atlas in MNI2009. (2020)
20 doi:10.5281/ZENODO.4010846.
- 21 18. Ciric, R. *et al.* Mitigating head motion artifact in functional connectivity MRI. *Nat.*
22 *Protoc.* **13**, 2801–2826 (2018).
- 23 19. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**,
24 1310–20 (2010).
- 25 20. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image
26 registration with cross-correlation: evaluating automated labeling of elderly and
27 neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
- 28 21. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the
29 Robust and Accurate Linear Registration and Motion Correction of Brain Images.
30 *Neuroimage* **17**, 825–841 (2002).
- 31 22. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic
32 resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–73 (1996).
- 33 23. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome
34 Project. *Neuroimage* **80**, 105–24 (2013).
- 35 24. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-
36 based registration. *Neuroimage* **48**, 63–72 (2009).
- 37 25. Ciric, R. *et al.* Benchmarking of participant-level confound regression strategies for the
38 control of motion artifact in studies of functional connectivity. *Neuroimage* **154**, 174–187
39 (2017).
- 40 26. Hallquist, M. N., Hwang, K. & Luna, B. The nuisance of nuisance regression: Spectral
41 misspecification in a common approach to resting-state fMRI preprocessing reintroduces
42 noise and obscures functional connectivity. *Neuroimage* **82**, 208–225 (2013).
- 43 27. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic
44 Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
- 45 28. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic
46 functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

- 1 29. Xia, C. H. *et al.* Multi-scale network regression for brain-phenotype associations. *Hum.*
- 2 *Brain Mapp.* **41**, 2553–2566 (2020).
- 3

Figures

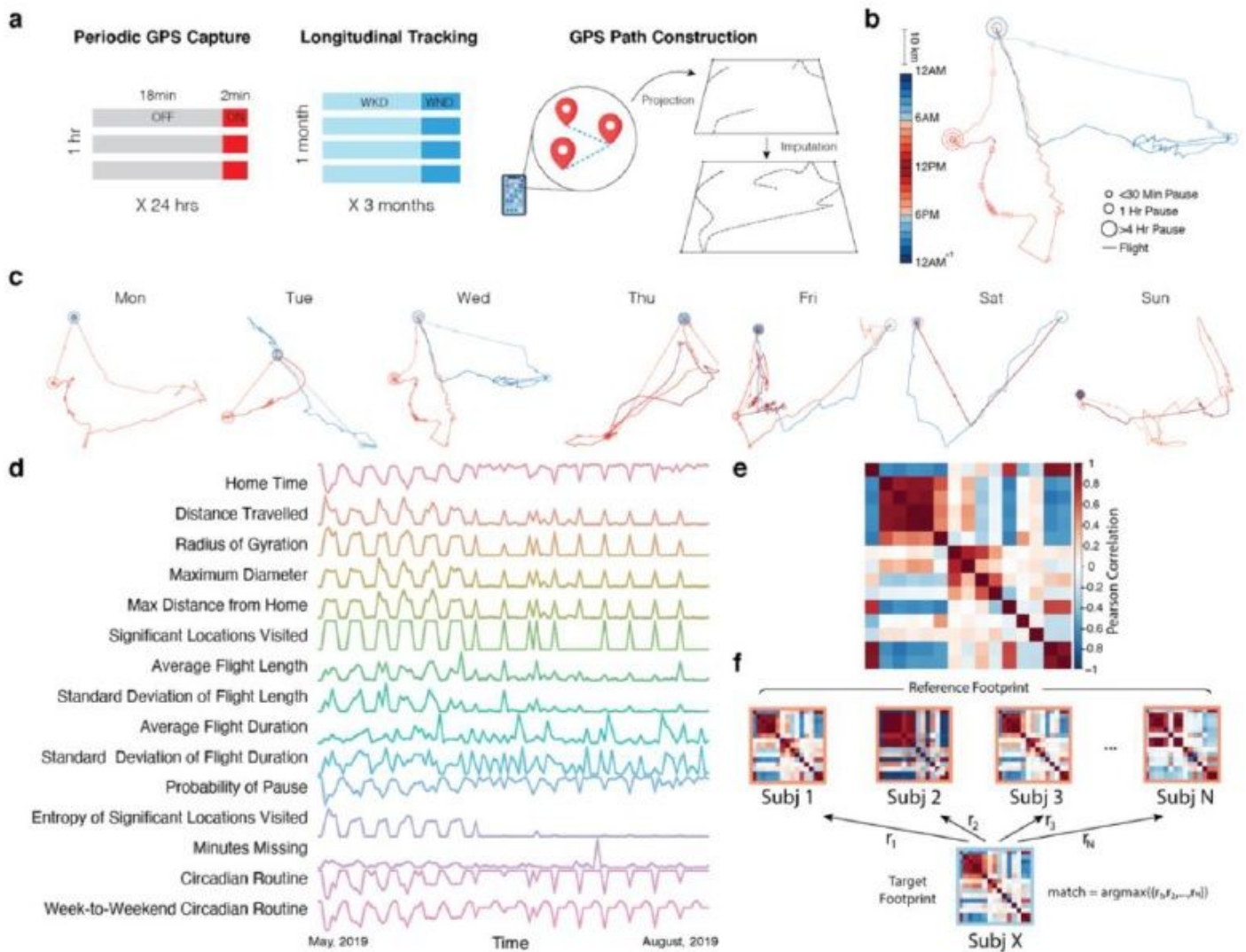


Figure 1

Constructing personal mobility “footprints”. a) We collected 3,317 person-days of mobility sensing data via personal smartphones from 41 adolescents and young adults. Geolocation data were recorded in cycles of 2min on and 18min off. Raw geolocation coordinates were de-identified via sphere to-2D standard space projection and were further imputed for missing data. b) For each individual, we constructed daily personal mobility trajectories, which consist of flights (movement) and pauses (stationary segments). Length of linear lines represents the duration of flights and size of circles represents the duration of pauses. Warm and cold colors indicate daytime and nighttime, respectively. c) A representative week of trajectories is shown, which demonstrates rich characteristics of personal mobility patterns formed over time. d) We extracted timeseries of mobility statistics (e.g. daily time spent at home) from geolocation and accelerometer data that parameterize movement characteristics over weeks to months. The example represented all 110 days of participants’ geolocation metrics recorded. e) For each individual, we constructed a covariance matrix from the mobility metric timeseries. Each cell of

the matrix was populated by the Pearson correlation between a given pair of mobility metrics. Warm and cold colors indicate positive and negative correlations, respectively. f) We randomly divided data into two equally sized parts, called the reference and target set. Subj X from the target set was matched to the subject in the reference that had the highest correlations between their footprints ($\text{argmax}(r_1, r_2, \dots, r_N)$). The identification was considered correct when underlying data came from the same subject; otherwise, the identification was considered incorrect. We quantified individual identification accuracy as the proportion of correct identifications across the entire sample; this procedure was repeated 1,000 times across different random partitions of the data.

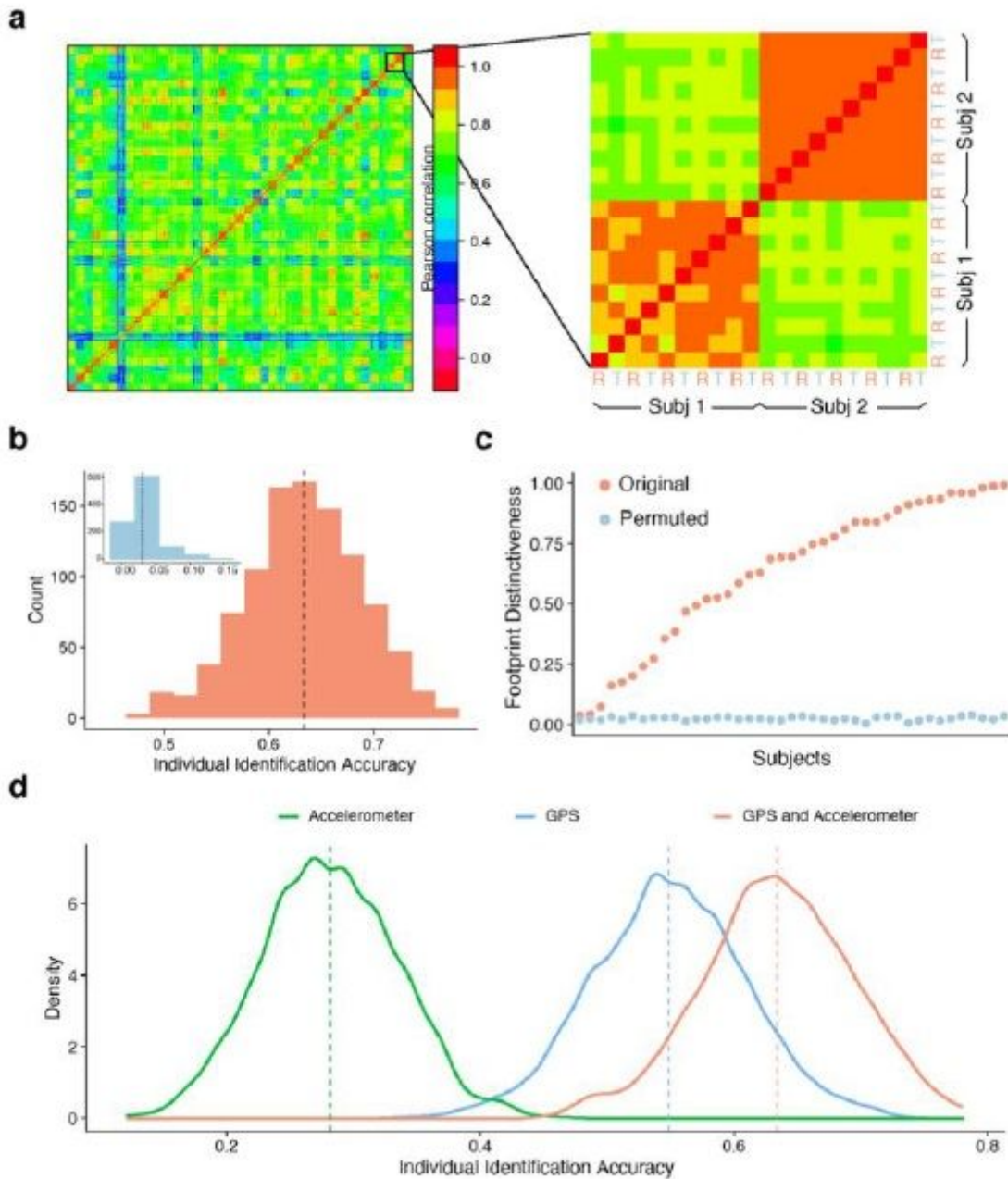


Figure 2

Identifying individuals using personal footprints. a) As an initial step, we visualized the similarity of mobility features across multiple random reference and target partitions (R & T in inset). It was readily apparent that mobility features were more highly correlated within participants (on diagonal) across data partitions than between participants (off diagonal). b) Across 1,000 random partitions, mobility footprinting enabled successful individual identification (mean: 63%, S.D.: 6%). In contrast, the mean chance accuracy from 1,000 permutation was 3% (inset, $p < 0.001$). c) For each individual, we calculated the footprint distinctiveness, or the percentage of correct identification across the 1,000 random partitions of the data. Ranked in ascending order, participants' footprint distinctiveness exhibited a wide range, from 4% to 99%. However, even the participant with the lowest identification distinctiveness was significantly higher than the null distribution. d) Individual identification based on geolocation alone had higher accuracy than accelerometer alone. However, they appeared to encode complementary features, as performance was maximal when both measures were used in footprinting.

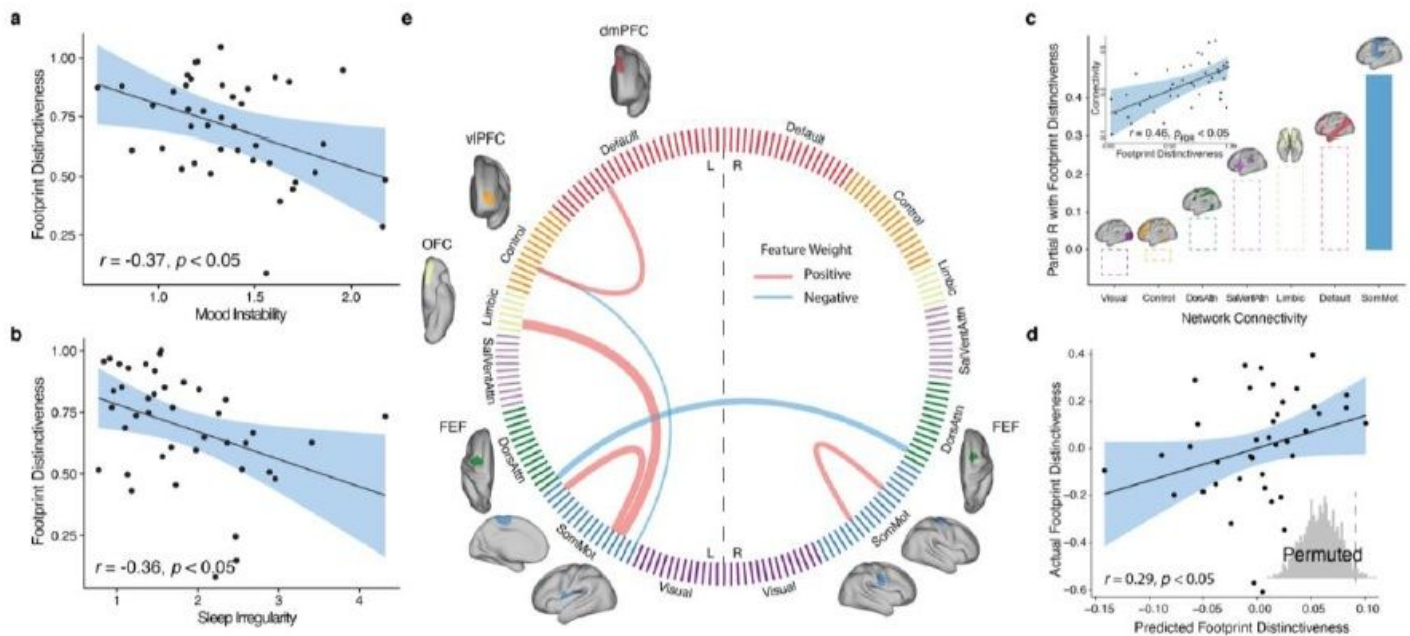


Figure 3

Individual footprint distinctiveness is associated with affective instability, sleep irregularity, and patterns of brain functional connectivity. a) Greater affective instability, measured by root mean square of successive differences in mood measures from ecological momentary assessment items acquired three times a day, was associated with reduced footprint distinctiveness ($r = -0.37$, $p < 0.05$), after controlling for data quantity, age, sex, and mean level of mood ratings. b) Similarly, we found that increased variability in sleep duration was associated with reduced footprint distinctiveness ($r = -0.36$, $p < 0.05$), after controlling for covariates. c) Across functional brain networks, only greater connectivity within the somatomotor network had a significant association with footprint distinctiveness ($r = 0.46$, $p < 0.05$, corrected for multiple comparisons with the false discovery rate). d) Patterns of brain functional connectivity significantly predicted individual footprint distinctiveness using leave-one-out cross-

validation ($r = 0.29$, inset: permutation-based $p = 0.025$). e) Six network edges consistently contributed to the sparse regression model. These edges included greater connectivity within somatomotor network, reduced connectivity between left and right frontal eye fields (FEF), increased connectivity between the somatomotor network and the left orbital frontal cortex (OFC) in the limbic network, as well as increased connectivity between the vlPFC (ventrolateral prefrontal cortex in the frontoparietal network) and the dmPFC (dorsolateral prefrontal cortex in the default mode network). Cord thickness reflects the weights in the model, reflecting each edge's contribution to the prediction; cord color indicates the sign of the weights.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MobileFootprintingSupplementaryMaterial20210412.pdf](#)