LETTER

An Earth–sized planet with an Earth–like density

Francesco Pepe[1], Andrew Collier Cameron[2], David W. Latham[3], Emilio Molinari[4,5], Stéphane Udry[1], Aldo S. Bonomo[6], Lars A. Buchhave[3,7], David Charbonneau[3], Rosario Cosentino[4,8], Courtney D. Dressing[3], Xavier Dumusque[3], Pedro Figueira[9], Aldo F. M. Fiorenzano[4], Sara Gettel[3], Avet Harutyunyan[4], Raphaëlle D. Haywood[2], Keith Horne[2], Mercedes Lopez-Morales[3], Christophe Lovis[1], Luca Malavolta[10,11], Michel Mayor[1], Giusi Micela[12], Fatemeh Motalebi[1], Valerio Nascimbeni[11], David Phillips[3], Giampaolo Piotto[10,11], Don Pollacco[13], Didier Queloz[1,14], Ken Rice[15], Dimitar Sasselov[3], Damien Ségransan[1], Alessandro Sozzetti[6], Andrew Szentgyorgyi[3] & Christopher A. Watson[16]

Recent analyses[1–4] of data from the NASA Kepler spacecraft[5] have established that planets with radii within 25 per cent of the Earth's ($R_\oplus$) are commonplace throughout the Galaxy, orbiting at least 16.5 per cent of Sun-like stars[1]. Because these studies were sensitive to the sizes of the planets but not their masses, the question remains whether these Earth-sized planets are indeed similar to the Earth in bulk composition. The smallest planets for which masses have been accurately determined[6,7] are Kepler-10b ($1.42R_\oplus$) and Kepler-36b ($1.49R_\oplus$), which are both significantly larger than the Earth. Recently, the planet Kepler-78b was discovered[8] and found to have a radius of only $1.16R_\oplus$. Here we report that the mass of this planet is 1.86 Earth masses. The resulting mean density of the planet is $5.57\,\mathrm{g\,cm^{-3}}$, which is similar to that of the Earth and implies a composition of iron and rock.

Every 8.5 h, the star Kepler-78 (first known as TYC 3147-188-1 and later designated KIC 8435766) presents to the Earth a shallow eclipse consistent[8] with the passage of an orbiting planet with a radius of $1.16 \pm 0.19 R_\oplus$. A previous study[8] demonstrated that it was very unlikely that these eclipses were the result of a massive companion either to Kepler-78 itself or to a fainter star near its position on the sky. Judging from the absence of ellipsoidal light variations[8] of the star, the upper limit on the mass of the planet is 8 Earth masses ($M_\oplus$). In addition to its diminutive size, the planet Kepler-78b is interesting because the light curve recorded by the Kepler spacecraft reveals the secondary eclipse of the planet behind the star as well as the variations in the light received from the planet as it orbits the star and presents different hemispheres to the observer. These data enabled constraints[8] to be put on the albedo and temperature of the planet. A direct measurement of the mass of Kepler-78b would permit an evaluation of its mean density and, by inference, its composition, and motivated this study.

The newly commissioned HARPS-N[9] spectrograph is the Northern Hemisphere copy of the HARPS[10] instrument, and, like HARPS, HARPS-N allows scientific observations to be made alongside thorium–argon emission spectra for wavelength calibration[11]. HARPS-N is installed at the 3.57-m Telescopio Nazionale Galileo at the Roque de los Muchachos Observatory, La Palma Island, Spain. The high optical efficiency of the instrument enables a radial-velocity precision of $1.2\,\mathrm{m\,s^{-1}}$ to be achieved in a 1-h exposure on a slowly rotating late-G-type or K-type dwarf star with stellar visible magnitude $m_v = 12$. By observing standard stars of known radial velocity during the first year of operation of HARPS-N, we estimated it to have a precision of at least $1\,\mathrm{m\,s^{-1}}$, a value which is roughly half the semi-amplitude of the signal expected for Kepler-78b should the planet have a rocky composition. We began an intensive

observing campaign (Methods) of Kepler-78 ($m_v = 11.72$) in May 2013, acquiring HARPS-N spectra of 30-min exposure time and an average signal-to-noise ratio of 45 per extracted pixel at 550 nm (wavelength bin of 0.00145 nm). From these high-quality spectra, we estimated[12,13] the stellar parameters of Kepler-78 (Methods and Extended Data Table 1). Our estimate of the stellar radius, $R_* = 0.737^{+0.034}_{-0.042}R_\odot$, is more accurate than any previously known[8] and allows us to refine the estimate of the planetary radius.

In the Supplementary Data, we provide a table of the radial velocities, the Julian dates, the measurement errors, the line bisector of the cross-correlation function, and the Ca II H-line and K-line activity indicator[14], $\log(R'_{HK})$. The radial velocities (Fig. 1) show a scatter of $4.08\,\mathrm{m\,s^{-1}}$ and a peak-to-trough variation of $22\,\mathrm{m\,s^{-1}}$, which exceeds the estimated average internal (photon-noise) precision, of $2.3\,\mathrm{m\,s^{-1}}$. The excess scatter is probably due to star-induced effects including spots and changes in the convective blueshift associated with variations in the stellar activity. These effects may cause an apparent signal at intervals corresponding to the stellar rotational period and its first and second harmonics. To separate this signal from that caused by the planet, we proceeded to estimate the rotation period of the star from the de-trended light curve from Kepler (Methods). Our estimate, of
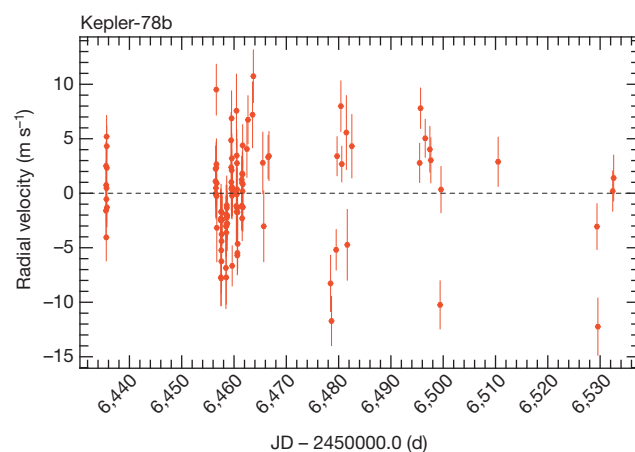

Figure 1 | Radial velocities of Kepler-78 as a function of time. The error bars indicate the estimated internal error (mainly photon-noise-induced error), which was $\sim2.3\,\mathrm{m\,s^{-1}}$ on average. The signal is dominated by stellar effects. The raw radial-velocity dispersion is $4.08\,\mathrm{m\,s^{-1}}$, which is about twice the photon noise. JD, Julian date.

[1]Observatoire Astronomique de l'Université de Genève, 51 chemin des Maillettes, 1290 Versoix, Switzerland. [2]Scottish Universities Physics Alliance, School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK. [3]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [4]INAF - Fundación Galileo Galilei, Rambla José Ana Fernandez Pérez 7, 38712 Breña Baja, Spain. [5]INAF - IASF Milano, via Bassini 15, 20133 Milano, Italy. [6]INAF - Osservatorio Astrofisico di Torino, via Osservatorio 20, 10025 Pino Torinese, Italy. [7]Centre for Star and Planet Formation, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen, Denmark. [8]INAF - Osservatorio Astrofisico di Catania, via Santa Sofia 78, 95125 Catania, Italy. [9]Centro de Astrofísica, Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal. [10]Dipartimento di Fisica e Astronomia ''Galileo Galilei'', Universita' di Padova, Vicolo dell'Osservatorio 3, 35122 Padova, Italy. [11]INAF - Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy. [12]INAF - Osservatorio Astronomico di Palermo, Piazza del Parlamento 1, 90124 Palermo, Italy. [13]Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. [14]Cavendish Laboratory, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. [15]Scottish Universities Physics Alliance, Institute for Astronomy, Royal Observatory, Blackford Hill, Edinburgh EH93HJ, UK. [16]Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK.

~12.6 d, is consistent with a previous estimate[8]. The power spectral density of the de-trended light curve also shows strong harmonics at respective periods of 6.3 and 4.2 d. We note that these timescales are much longer than the orbital period of the planet.

To estimate the radial-velocity semi-amplitude, K_p, due to the planet, we proceeded under the assumption that K_P was much larger than the change in radial velocity arising from stellar activity during a single night. This is a reasonable assumption, because a typical 6-h observing sequence spans only 2% of the stellar rotation cycle. Furthermore, the relative phase between the stellar signal and the planetary signal changes continuously, and so we expect that the contribution from stellar activity should average out over the three-month observing period. Assuming a circular orbit, we modelled the radial velocity of the ith observation (gathered at time t_i on night d) as $v_{i,d} = v_{0,d} - K_p \sin(2\pi(t_i - t_0)/P)$, where t_0 is the epoch of mid-transit and P is the orbital period, each held fixed at the values derived from the photometry, and $v_{0,d}$ is the night-d zero point (an offset value estimated independently for each night). We solved for the values of K_p and $v_{0,d}$ by minimizing the χ^2 function (least-squares minimization), assuming white noise and weighting the data according to the inverse variances derived from the HARPS-N noise model. This is a technique similar to the one used in a recent study[15] of another low-mass transiting planet, CoRoT-7b.

With this technique, we measure a preliminary value of $K_p = 2.08 \pm 0.32\,\mathrm{m\,s^{-1}}$, implying a detection significance of 6.5σ. We confirmed that the radial-velocity signal is consistent with the photometric transit ephemeris by repeating the χ^2 optimization over a grid of orbital frequencies and times of mid-transit (Fig. 2a). This confirms that the values of P and t_0 from the Kepler light curves coincide with the lowest χ^2 minimum in the resulting period–phase diagram.

The Kepler light curve evolves on a timescale of weeks. We therefore expect the stellar-activity-induced radial-velocity signals to remain coherent on the same timescale. To explore this, we fitted the radial velocities with a sum of Keplerian models at different periods, one of which we expect to correspond to the planetary orbit and the others (which we left as free parameters) to represent the effects of stellar activity. Using the Bayesian information criterion as our discriminant (Methods), we found that a model with three Keplerian functions was sufficient to explain the data. The 4.2-d period of the second Keplerian function corresponds to the second harmonic of the photometrically determined stellar rotation period. The period of 10 d for the third Keplerian also appears as a prominent peak in the generalized Lomb–Scargle analysis[16] of the radial-velocity data. Strong peaks at similar periods are also present in periodograms of the Ca II HK activity indicator, the line bisector and the full-width at half-maximum of the cross-correlation function[11] (Extended Data Fig. 1). We conclude that both the 4.2-d and 10-d signals probably have stellar causes.

The best three-Keplerian fit of the data yields an estimate of $K_p = 1.96 \pm 0.32\,\mathrm{m\,s^{-1}}$ and residuals with a dispersion of $2.34\,\mathrm{m\,s^{-1}}$, very close to the internal noise estimates. In Fig. 2b, we show the phase-folded radial velocities after removal of the stellar components, plotted along with the best-fit Keplerian at the planetary orbital period. The orbital parameters are given in Table 1. Having settled on the three-component model, we estimated the planetary mass and density by conducting a Markov chain Monte Carlo (MCMC) analysis (Methods). In this analysis, we adopted previously published[8] values of P and t_0 as Gaussian priors. We replaced the planet radius, R_p, with the published estimate[8] of the ratio $k = R_p/R_*$ and our determination of R_*. The planetary radius then becomes an output of the MCMC analysis. By adopting the mode of the distributions, we find a planet mass of $m_p = 1.86^{+0.38}_{-0.25}\,M_\oplus$, a radius of $R_p = 1.173^{+0.159}_{-0.089}\,R_\oplus$ and a planet density of $\rho_p = 5.57^{+3.02}_{-01.31}\,\mathrm{g\,cm^{-3}}$. These estimates include the contribution from
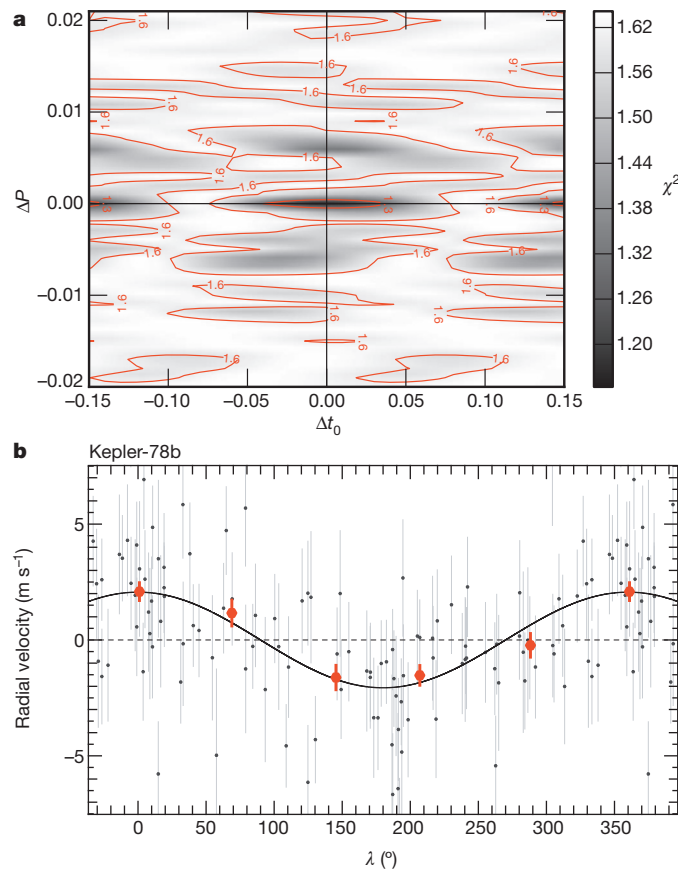

Figure 2 | Model fit of the radial velocities of Kepler-78. a, Greyscale contour plot of the χ^2 surface around the ephemeris for the period and mid-time of transit from ref. 8 (origin of plot). ΔP and Δt_0 represent the departure from the expected ephemeris in units of days. The position of the minimum of the residuals perfectly matches the expected values. **b**, Phase-folded radial velocities and fitted Keplerian orbit of the signal induced by Kepler-78b after removal of the modelled stellar noise components. λ, mean longitude. The transit occurs at $\lambda_T = 90°$. We note that the higher number of data points at maximum and minimum radial velocity is a direct consequence of our strategy of observing at quadrature, where the information on amplitude is highest. The red dots show the radial velocities and the corresponding errors when binned over the orbital phase. The error bars of the individual measurements indicate the estimated internal errors (photon-noise-induced error), of $\sim 2.3\,\mathrm{m\,s^{-1}}$ on average. The error bars on the averaged data points essentially represent the internal error of an individual measurement divided by the square root of the number of averaged measurements.

Table 1 | Planetary system parameters of Kepler-78 b

Planetary system parameter	Kepler-78b
Reference epoch, T_0 (BJD$_\mathrm{UTC}$)	2,456,465.076392
Orbital inclination, I (°)*	79^{+9}_{-14}
Systemic radial velocity, γ (km s^{-1})†	-3.5084 ± 0.0008
Orbital period, P (d)†	0.3550 ± 0.0004
Mean longitude, λ_0, at epoch T_0 (°)†	293 ± 13
Eccentricity, e†	0
Radial-velocity semi-amplitude, K_p (m s^{-1})†	1.96 ± 0.32
Planetary mass, m_p ($M_\oplus$)‡	$1.86^{+0.38}_{-0.25}$
Planetary radius, R_p ($R_\oplus$)‡	$1.173^{+0.159}_{-0.089}$
Planetary mean density, ρ_p (g cm^{-3})‡	$5.57^{+3.02}_{-1.31}$
Semi-major axis, a (AU)†	0.0089
Surface temperature, T (K)*	1500-3000
Number of measurements, N_meas	109
Time span of observations (d)	97.1
Radial-velocity dispersion ($O - C$) (m s^{-1})†	2.34
Reduced χ^2†	1.12 ± 0.07

The parameters are determined from radial velocities. BJD$_\mathrm{UTC}$ indicates barycentric Julian date expressed in coordinated universal time. λ_0 denotes the mean longitude at the mean date of the observing campaign (reference epoch, T_0). This coordinate has the advantage of not being degenerate for low eccentricities. Our choice of T_0 reduces correlations between adjusted parameters. $O - C$ is the standard deviation of the difference between the observed and computed (modelled) data. Stated uncertainties represent the 68.3% confidence interval.
* Taken from the discovery paper[8]. † Fitted orbital parameters (maximum likelihood). ‡ Mode and confidence interval of the real distribution deduced from the MCMC analysis.
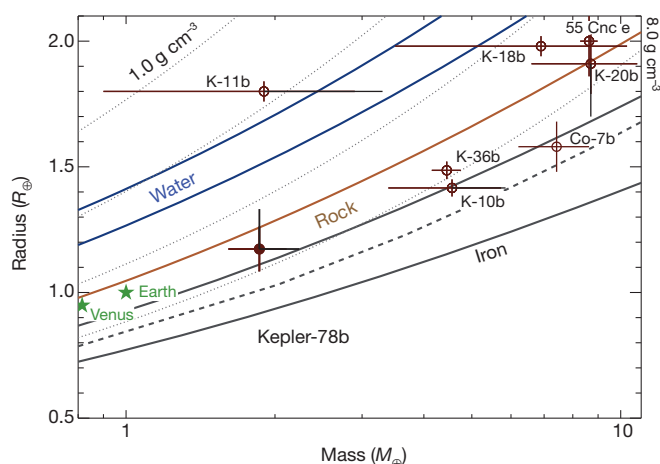
Figure 3 | The hot, rocky planet Kepler-78b placed on a planetary mass–radius diagram. Here masses and radii represent the modes of the corresponding distributions derived from the MCMC analysis, and the error bars represent the 68.3% confidence interval (1σ). For comparison, Earth and Venus are indicated in the same diagram by star symbols. The other exoplanets shown are those for which mass and radius have been estimated (K, Kepler; Co, CoRoT; 55 Cnc, 55 Cancri). From top to bottom, the solid lines show mass and radius for planets consisting of pure water, 50% water, 100% silicates, 50% silicates and 50% iron core, and 100% iron, as computed with the theoretical models of ref. 18. The dashed line shows the maximum mantle-stripping models from ref. 21; that is, Earth-like exoplanets made of pure iron are not expected to form around normal stars. From top to bottom, the dotted lines represent mean densities of 1, 2, 4 and $8 \, \mathrm{g \, cm^{-3}}$.

the uncertainty in the stellar mass. The uncertainty in the density is dominated by the uncertainty in k. Our values for m_p and ρ_P are consistent with those from an independent study[17].

In terms of mass, radius and mean density, Kepler-78b is the most similar to the Earth among the exoplanets for which these quantities have been determined. We plot the mass–radius diagram in Fig. 3. By comparing our estimates of Kepler-78b with theoretical models[18] of internal composition, we find that the planet has a rocky interior and most probably a relatively large iron core (perhaps comprising 40% of the planet by mass). We note that in the part of the mass–radius diagram where Kepler-78b lies, there is a general agreement between models and little or no degeneracy. The extreme proximity of the planet to its star, resulting in a high surface temperature and ultraviolet irradiation, would preclude there being a low-molecular-weight atmosphere: any water or volatile envelope that Kepler-78b might have had at formation should have rapidly evaporated[19]. Kepler-78b is also similar to larger high-density, hot exoplanets (Kepler-10b (ref. 6), Kepler-36b (ref. 7) and CoRoT-7b (ref. 20)), in that in the mass–radius diagram it is not below the lower envelope of mantle-stripping models[21] that tend to enhance the fraction of the planet's iron core. At present, Kepler-78b is the extrasolar planet whose mass, radius and likely composition are most similar to those of the Earth. However, it differs from the Earth notably in its very short orbital period and correspondingly high temperature.

The observations of Kepler-78 have shown the potential of the much-anticipated HARPS-N spectrograph. It will have a crucial role in the characterization of the many Kepler planet candidates with radii similar to that of the Earth. By acquiring and analysing a large number of precise radial-velocity measurements, we can learn whether Earth-sized planets (typically) have Earth-like densities (and, by inference, Earth-like compositions), or whether even small planets have a wide range of compositions, as has recently been established[22,23] for their larger kin.

METHODS SUMMARY

In the case of Kepler-78, the planet-induced radial-velocity variation is small compared with the stellar jitter. If their periodicities are very different, however, it is easy to de-correlate the signals and determine the radial-velocity amplitude of the

planet. We used the Kepler light curve of Kepler-78 to measure the stellar rotational period. After de-trending[24,25] the photometry, we computed its power spectral density, which immediately revealed excess power at a period of 12.6 d, the rotational period of the star.

The HARPS-N observations of Kepler-78 yielded not only radial velocities but also high-resolution spectra, which we combined into a spectrum with a high signal-to-noise ratio. By applying the stellar parameter classification pipeline[12], we derived precise stellar parameters. In particular, we re-determined the stellar radius to be $R_* = 0.737^{+0.034}_{-0.042} R_\odot$, with smaller uncertainties than in the value in the discovery paper[8].

For the purpose of measuring the signal induced by Kepler-78b, several models can be applied to the data, which may all lead to similar results. However, not all of the models represent the data with the same quality. We therefore used the Bayesian information criterion[26–28] to determine which model matches the data best. This analysis led us to select a three-Keplerian model with two sinusoids (zero-eccentricity Keplerians) for the planet and the 4.2-d stellar signal, respectively, and one Keplerian with non-zero eccentricity for the stellar signal at 10 d.

Once a model has been selected, it is adjusted by a least-squares fit to the data. This approach leads to the maximum-likelihood solution but does not provide all statistically relevant solutions and the distributions of their parameters. We used an MCMC analysis to determine the distribution of all orbital and planetary parameters, in particular the planetary mass and density, and to determine their respective errors.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Fressin, F. *et al.* The false positive rate of Kepler and the occurrence of planets. *Astrophys. J.* **766**, 81–100 (2013).
2. Petigura, E. A., Marcy, G. W. & Howard, A. W. A plateau in the planet population below twice the size of Earth. *Astrophys. J.* **770**, 69–89 (2013).
3. Swift, J. J. *et al.* Characterizing the cool KOIs. IV. Kepler-32 as a prototype for the formation of compact planetary systems throughout the galaxy. *Astrophys. J.* **764**, 105–118 (2013).
4. Dressing, C. D. & Charbonneau, D. The occurrence rate of small planets around small stars. *Astrophys. J.* **767**, 95–114 (2013).
5. Batalha, N. M. *et al.* Planetary candidates observed by Kepler. III. Analysis of the first 16 months of data. *Astrophys. J.* **204** (Suppl.), 24–44 (2011).
6. Batalha, N. M. *et al.* Kepler's first rocky planet: Kepler-10b. *Astrophys. J.* **729**, 27–47 (2011).
7. Carter, J. A. *et al.* Kepler-36: a pair of planets with neighboring orbits and dissimilar densities. *Science* **337**, 556–559 (2012).
8. Sanchis-Ojeda, R. *et al.* Transits and occultations of an Earth-sized planet in an 8.5-hour orbit. *Astrophys. J.* **774**, 54–62 (2013).
9. Cosentino, R. *et al.* Harps-N: the new planet hunter at TNG. *Proc. SPIE* **8446**, 84461V (2012).
10. Mayor, M. *et al.* Setting new standards with HARPS. *Messenger* **114**, 20–24 (2003).
11. Baranne, A. *et al.* ELODIE: a spectrograph for accurate radial velocity measurements. *Astron. Astrophys.* **119** (Suppl.), 373–390 (1996).
12. Buchhave, L. A. *et al.* An abundance of small exoplanets around stars with a wide range of metallicities. *Nature* **486**, 375–377 (2012).
13. Yi, S. *et al.* Toward better age estimates for stellar populations: the Y^2 isochrones for solar mixture. *Astrophys. J. Suppl. Ser.* **136**, 417–437 (2001).
14. Lovis, Ch. *et al.* The HARPS search for southern extra-solar planets. XXXI. Magnetic activity cycles in solar-type stars: statistics and impact on precise radial velocities. Preprint at http://arxiv.org/abs/1107.5325 (2011).
15. Hatzes, A. P. *et al.* An investigation into the radial velocity variations of CoRoT-7. *Astron. Astrophys.* **520**, A93–A108 (2010).
16. Zechmeister, M. & Kuerster, M. The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *Astron. Astrophys.* **496**, 577–584 (2009).
17. Howard, A. W. *et al.* A rocky composition for an Earth-sized exoplanet. *Nature* http://dx.doi.org/10.1038/nature12767 (this issue).
18. Zeng, L. & Sasselov, D. A detailed model grid for solid planets from 0.1 through 100 Earth masses. *Publ. Astron. Soc. Pacif.* **125**, 227–239 (2013).
19. Rogers, L. A., Bodenheimer, P., Lissauer, J., Seager, S., The low density limit of the mass-radius relation for exo-Neptunes. *Bull. Am. Astron. Soc.* **43**, 402.04 (2011).
20. Léger, A., Rouan, D. & Schneider, J. Transiting exoplanets from the CoRoT space mission. VIII. CoRoT-7b: the first super-Earth with measured radius. *Astron. Astrophys.* **506**, 287–302 (2009).
21. Marcus, R. A., Sasselov, D., Hernquist, L. & Stewart, S. T. Minimum radii of super-Earths: constraints from giant impacts. *Astrophys. J.* **712**, L73–L76 (2010).
22. Lissauer, J. J. *et al.* All six planets known to orbit Kepler-11 have low densities. *Astrophys. J.* **770**, 131–145 (2013).
23. Charbonneau, D. *et al.* A super-Earth transiting a nearby low-mass star. *Nature* **462**, 891–894 (2009).
24. Smith, J. C. *et al.* Kepler presearch data conditioning II: a Bayesian approach to systematic error correction. *Publ. Astron. Soc. Pacif.* **124**, 1000–1014 (2012).

25. Stumpe, M. C. *et al.* Kepler presearch data conditioning I: architecture and algorithms for error correction in Kepler light curves. *Publ. Astron. Soc. Pacif.* **124,** 985–999 (2012).
26. Schwarz, G. E. Estimating the dimension of a model. *Ann. Stat.* **6,** 461–464 (1978).
27. Liddle, A. R. Information criteria for astrophysical model selection. *Mon. Not. R. Astron. Soc.* **377,** L74–L78 (2007).
28. Burnham, K. P. Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res.* **33,** 261–304 (2004).

Supplementary Information is available in the online version of the paper.

Author Contributions The underlying observation programme was conceived and organized by F.P., A.C.C., D.W.L., C.L., D. Ségransan, S.U. and E.M. Observations with HARPS-N were carried out by A.C.C., A.S.B., D.C., R.C., C.D.D., X.D., P.F., A.F.M.F., S.G., A.H., R.D.H., M.L.-M., V.N., D. Pollacco, D.Q., K.R., A. Sozzetti, A. Szentgyorgyi and C.A.W. The data-reduction pipeline was adapted and updated by C.L., who also implemented the correction for charge-transfer-efficiency errors and the automatic computation of the activity indicator $\log(R'_{HK})$. M.L.-M. and S.G. independently computed the S-index and $\log(R'_{HK})$ values. A.C.C., D. Ségransan, A.S.B. and X.D. analysed the data using the offset-correction method. A.C.C. and D. Ségransan re-analysed the data for the determination of the stellar rotational period based on the Kepler light curve. P.F. investigated for possible correlations between the radial velocities and the line bisector. L.A.B. conducted the stellar parameter classification analysis for the re-determination of the stellar parameters based on HARPS-N spectra. An independent determination of the stellar parameters was conducted by L.M. by analysis of the cross-correlation function. D. Ségransan compared many different models to fit the observed data and selected the most appropriate by using the Bayesian information criterion. D. Ségransan performed a detailed MCMC analysis for the determination of the planetary parameters, with contributions also from A.S.B. and A. Sozzetti. F.P. was the primary author of the manuscript, with important contributions by D.C., D. Ségransan, D. Sasselov, C.A.W., K.R. and C.L. All authors are members of the HARPS-N Science Team and have contributed to the interpretation of the data and the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.P. (francesco.pepe@unige.ch).

METHODS

Photometric determination of stellar rotational period of Kepler-78. In ref. 8, the Kepler light curve of Kepler-78 was analysed and was de-trended using the PDC-MAP algorithm (Extended Data Fig. 2), which preserves stellar variability[24,25]. The light curve displays clear rotational modulation with a peak-to-valley amplitude that varies between 0.5% and 1.5%, and a period of 12.6 ± 0.3 d. We confirmed the rotational period by computing the autocorrelation function of the PDC-MAP light curve: Using a fast Fourier transform we compute the power spectral density from which the autocorrelation function (ACF) is obtained using the inverse transform. We immediately derive a rotational period of 12.6 d (Extended Data Fig. 3a). The amplitudes of successive peaks decay on an e-folding timescale of about 50 d, which we attribute to the finite lifetimes of individual active regions. The power density distribution in Extended Data Fig. 3b finally shows a peak at the stellar rotational period as well as at its first and second harmonics. The main signal at period $P = 0.355$ d of Kepler-78b, as well as its harmonics, are easily identified at shorter periods.

HARPS-N observations and stellar parameters. To explore the feasibility of the programme, we performed five hours of continuous observations during a first test night in May 2013. This test night allowed us to determine the optimum strategy and to verify whether the measurement precision was consistent with expectations. Indeed, 12 exposures, each of 30 min, led to an observed dispersion of the order of 2.5 m s^{-1}, close to the expected photon noise. We therefore decided to dedicate six full HARPS-N nights to the observation of Kepler-78b in June 2013. Given the excellent stability of the instrument (typically less than 1 m s^{-1} during the night) and the faintness of the star, we observed without the simultaneous reference source[10,11] that usually serves to track potential instrumental drifts. Instead, the second fibre of the spectrograph was placed on the sky to record possible background contamination during cloudy moonlit nights. Owing to excellent astroclimatic conditions, we gathered a total of 81 exposures, each of 30 min, free of moonlight contamination and with an average signal-to-noise ratio (SNR) of 45 per extracted pixel at a wavelength of $\lambda = 550$ nm. An extracted pixel covers a wavelength bin of 0.000145 nm.

A first analysis of these observations confirmed the presence of the planetary signal. However, it also confirmed that, as suggested in ref. 8, the stellar variability induces radial-velocity variations much larger than the planetary signal, although on very different timescales. To consolidate our results and improve the precision of our planetary-mass measurement, we decided to perform additional observations during the months of July and August 2013. We preferred, however, to observe Kepler-78 only twice per night, around quadrature (at maximum and minimum expected radial velocity), to minimize observing time and to maximize the information on the amplitude. This strategy allowed us to determine the (low-frequency) stellar contribution as the sum of the two nightly measurements and the (high-frequency) planetary signal as the difference between them. We finally obtained a total of 109 high-quality observations over three months, with an average photon-noise-limited precision of 2.3 m s^{-1}.

The large number of high-SNR spectra gathered by HARPS-N allowed us to re-determine the stellar parameters using the stellar parameter classification (SPC) pipeline[12]. Each high-resolution spectrum ($R = 115,000$) yields an average SNR per resolution element of 91 in the MgB region. The weighted average of the individual spectroscopic analyses resulted in final stellar parameters of $T_{eff} = 5058 \pm 50$ K, $\log(g) = 4.55 \pm 0.1$, $[m/H] = -0.18 \pm 0.08$ and $v\sin(i) = 2 \pm 1$ km s^{-1}, in agreement, within the uncertainties, with the discovery paper. The value for $v\sin(i)$ is, however, poorly determined by SPC. Therefore, we adopted an internal calibration based on the full-width at half-maximum of the cross-correlation functions to compute the projected rotational velocity, which yielded $v\sin(i) = 2.8 \pm 0.5$ km s^{-1}. We note that, assuming spin–orbit alignment, the rotational velocity and our estimate of the stellar radius yield a rotational period of 13 d. This value is in agreement with the stellar rotational period determined from photometry.

The stellar parameters from SPC[12] have been input to the Yonsei–Yale stellar evolutionary models[13] to estimate the mass and radius of the host star. We obtain $M_* = (0.758 \pm 0.046)M_\odot$ for the stellar mass and $R_* = 0.737^{+0.034}_{-0.042}R_\odot$ for the radius, in agreement, within the uncertainties, with the discovery paper. The Ca II HK activity indicator is computed by the online and automatic data-reduction pipeline, which gives an average value of $\log(R'_{HK}) = -4.52$ when using a colour index of $B–V = 0.91$ for Kepler-78. The stellar parameters are summarized in Extended Data Table 1.

Radial-velocity model selection. It is interesting to note that the signature of Kepler-78b can be retrieved despite the large stellar signals superimposed on the radial velocity induced by the planet. To demonstrate this, we adjusted the data with a simple model consisting of a cosine and the star's systemic velocity, while fixing the period and time of transit to the published values[8]. We compared the results of this simple model with a simple constant using the Bayesian information criterion[26–28] (BIC). We derived the relative likelihood of the two models, also called the evidence ratio, to be e$^{-(1/2)\Delta BIC_i} = 4 \times 10^{-10}$. This first estimate tells us that our

cosine model is much superior to the simple constant. In other words, we can say that we have a clear detection of a signal of semi-amplitude $K_p = 1.88 \pm 0.47$ m s^{-1}. Although certainly biased owing to the lack of stellar activity de-trending, the result provides a confirmation of the existence of Kepler-78b and a first estimation of its mass.

To model the stellar signature, we followed two different approaches. The first one consists of removing any stellar effect occurring on a timescale longer than 2 d by adjusting nightly offsets to the data. This method has the main advantage of not relying on any analytical model and it overcomes the difficulty of modelling non-stationary processes that often characterize stellar activity. The approach is also well suited to our problem because the period of the planet is very short. Its only drawback comes from the large number of additional parameters (21 offsets, one per night), which is a direct consequence of our observing strategy. The second approach consists of modelling the stellar activity as a set of sinusoids or Keplerian functions. This approach makes sense provided that spot groups and plages are coherent on a timescale similar to the radial-velocity observation time span. For Kepler-78, the ACF of the light curve shows a 1/e de-correlation of ~50 d (Extended Data Fig. 3a), which compares well with the 97-d time span of the HARPS-N observations.

In total, we studied a series of more than 30 different models of different complexity. We have compared these models using the BIC[28] evidence ratio, ER, and the BIC weight, w, to find the best few models:

$$w_i = \frac{e^{-(1/2)\Delta BIC_i}}{\sum_i e^{-(1/2)\Delta BIC_i}}$$

Of all the models we considered, two are statistically much more significant. They consist of modelling the stellar activity as a sum of periodic signals. The best model, with a BIC weight of 0.71, predicts three Keplerians including the planetary signal ($P_1 = 0.355$ d, $P_2 = 4.2$ d and $P_3 = 10.04$ d; $e_1, e_2 = 0$). The second best contains four Keplerians ($P_1 = 0.355$ d, $P_2 = 4.2$ d, $P_3 = 6.5$ d and $P_4 = 23–58$ d; all eccentricities $e_i = 0$). De-trending of the stellar activity shows much weaker evidence ratios. We therefore retained model 5 (Extended Data Table 2), which consists of one Keplerian describing the planet Kepler-78b and two additional 'signals', a sinusoid of period $P_2 = 4.2$ d and a slightly eccentric Keplerian with $P_3 = 10.04$ d. In Extended Data Table 3, we present the distribution of the parameters of the best model that results from our MCMC analysis (see below). Furthermore, Extended Data Fig. 4 shows the periodogram of the radial-velocity residuals after subtracting the stellar components. The planetary signal is now detected with a false-alarm probability significantly lower than 1%.

MCMC analysis. To retrieve the marginal distribution of the true mass of the planet and its density, we carry out an MCMC analysis based on the model selection process described in the previous section. We sample the posterior distributions using an MCMC with the Metropolis–Hastings algorithm. Because the model is very well constrained by the data, the MCMC starts from the solution corresponding to the maximum likelihood, and the MCMC parameter steps correspond to the standard deviation of the adjusted parameters. An acceptance rate of 25% is chosen. To obtain the best possible end result, we take as priors the transit parameters of Kepler-78b (ref. 8). Symmetric distributions are considered to be Gaussian, whereas asymmetric ones, such as that of the orbital inclination, are modelled by split-normal distributions using the published value of the mode of the distribution. We re-derive the radius of the planet using our improved stellar radius estimation and the planet-to-star radius ratio from the Kepler photometry[8]. All other parameters have uniform priors except for the period P, for which a modified Jeffrey's prior is preferred[29]. We use $\sqrt{e}\cos(\omega)$ and $\sqrt{e}\sin(\omega)$ as free parameters, which translate into a uniform prior in eccentricity[30]. The mean longitude, λ_0, computed at the mean date of the observing campaign, is also preferred as a free parameter. It has the advantage of not being degenerate for low eccentricities, whereas our choice for the reference epoch, T_0, reduces correlations between adjusted parameters. In this analysis, the MCMC has 2,000,000 iterations and converges after a few hundred iterations. The ACF of each parameter is computed to estimate the typical correlation length of our chains and to estimate a sampling interval to build the final statistical sample. All ACFs have a very short decay (1/e decay after 100 iterations and 1/100 decay after 300 iterations) and present no correlations on a larger iteration lag. We build our final sample using the 1/e-decay iteration lag, which is a good compromise between the size of the statistical sample and its de-correlation value. The final statistical samples consist of 20,000 elements, from which orbital elements and confidence intervals are derived. The resulting orbital elements for Kepler-78b are listed in Extended Data Table 3. The results for the mass, radius and density of the planet are given in Extended Data Table 4, and the distributions for the mass and the density are plotted in Extended Data Fig. 5. These distributions are smoothed for better rendering.

29. Gregory, P. C. *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge Univ. Press, 2005).
30. Anderson, D. R. *et al.* WASP-30b: a 61 M_{Jup} brown dwarf transiting a $V = 12$, F8 star. *Astrophys J.* **726,** L19–L23 (2011).
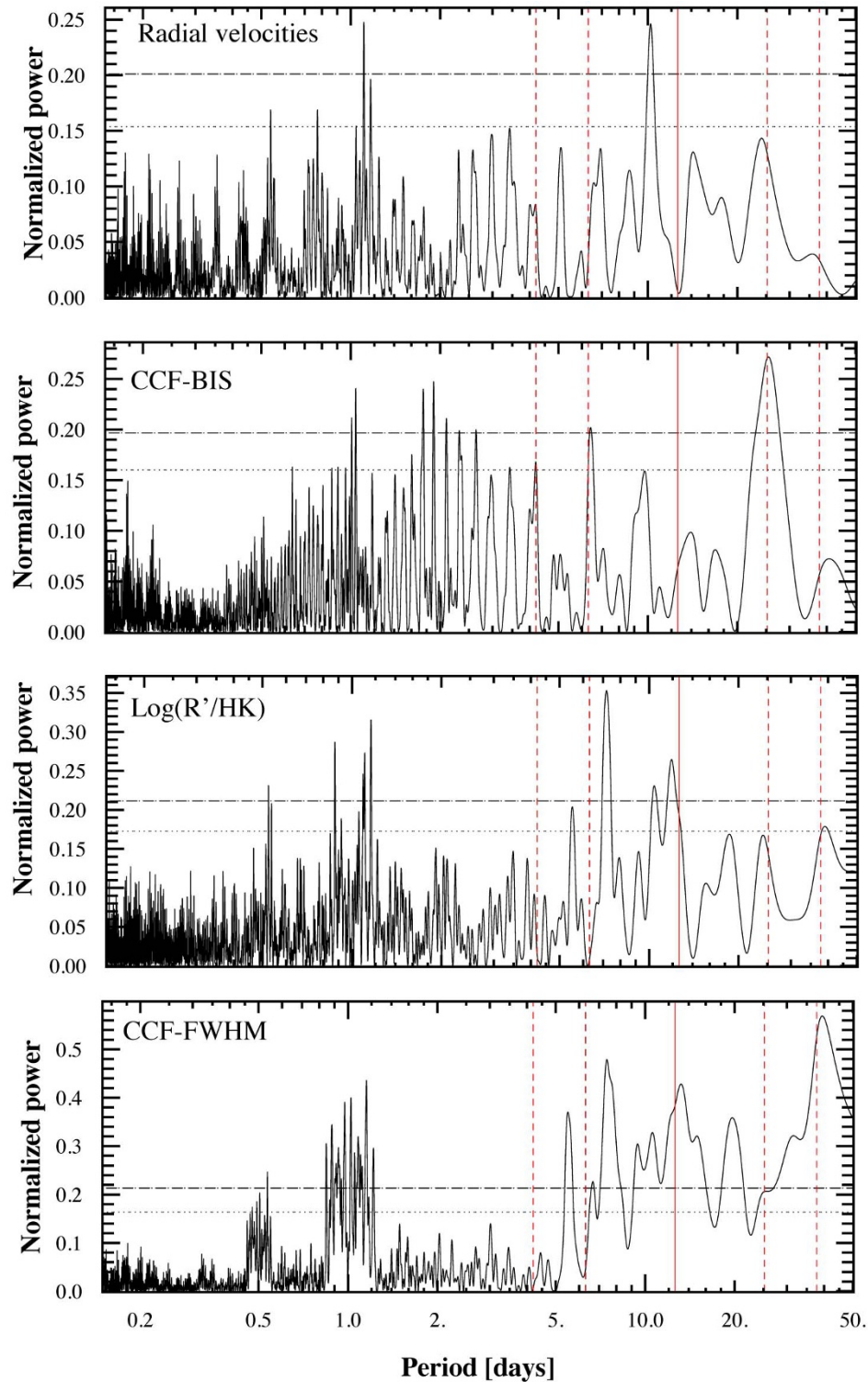
Extended Data Figure 1 | Generalized Lomb–Scargle periodogram of several parameters measured by HARPS-N. The panels show, from top to bottom, the periodogram of the radial velocities (RV) of Kepler-78, the line bisector (CCF-BIS), the activity indicator ($\log(R'_{HK})$) and the full-width at half maximum (CCF-FWHM) of Kepler-78. The dotted and dashed horizontal lines represent the 10% and 1% false-alarm probabilities, respectively. The vertical lines show the stellar rotational period (solid) and its two first harmonics (dashed). All the indicators show excess energy at periods of around 6 d and above, indicating that the peak observed in the radial-velocity data at a period of about 10 d is most likely to have a stellar origin. The additional power in the line bisector periodogram at periods longer than 1 d is most probably induced by stellar spots.
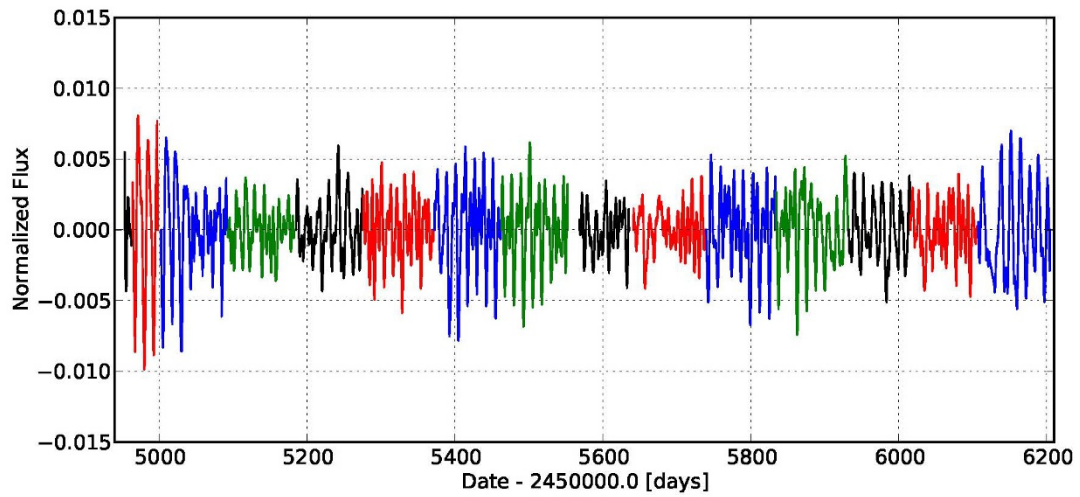
Extended Data Figure 2 | Kepler light curve of Kepler-78d. The data have been de-trended using the PDC-MAP algorithm. Different colours represent different quarters of observation.
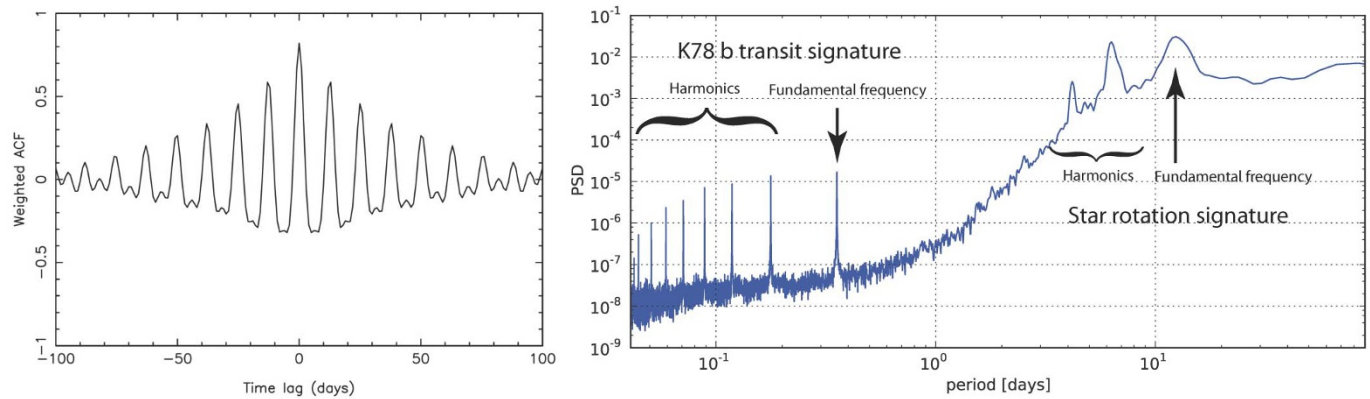
Extended Data Figure 3 | Spectral analysis of the Kepler light curve. Left panel, ACF of the Kepler light curve showing correlation peaks every 12.6 d and a decay on an e-folding timescale of ~50 d. Right panel, the power spectral distribution of the Kepler light curve. Peaks are well identified at the stellar rotational period of 12.6 d and its two first harmonics. At shorter periods, the signal and several harmonics of the transiting planet Kepler-78b can be identified.
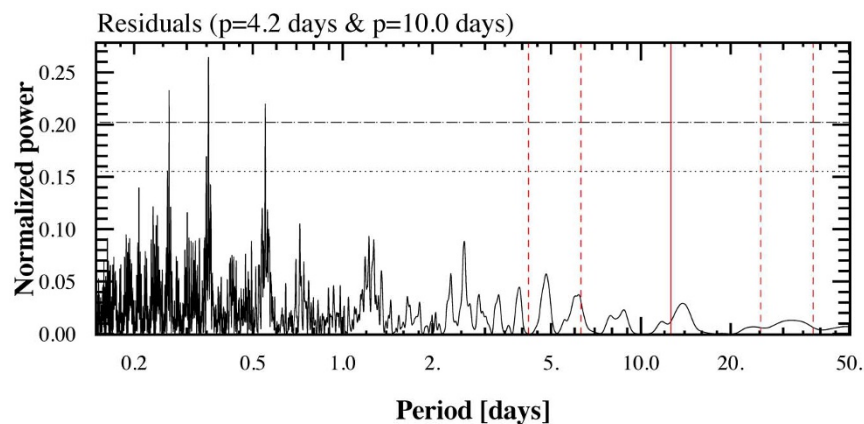
Extended Data Figure 4 | Periodogram of the radial-velocity residuals after subtraction of the 4.2-d and 10.0-d stellar components. The dotted and dashed horizontal lines represent the 10% and 1% false-alarm probabilities, respectively. The signature of Kepler-78b (and its aliases) can now clearly be identified with a false-alarm probability significantly lower than 1%.
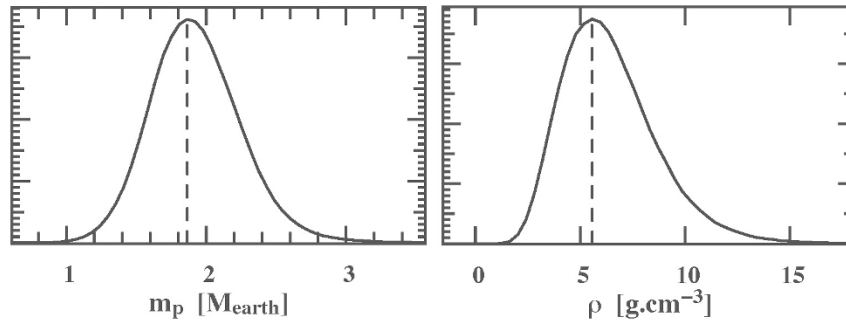
Extended Data Figure 5 | Probability density functions derived from the MCMC analysis. Probability density function of the planetary mass (left) and probability density function of the planetary density (right).

Extended Data Table 1 | Stellar parameters of Kepler-78 computed from HARPS-N spectra

Stellar parameter	Unit	Value	68.3% C.I.
*Effective temperature, T_{eff}	[K]	5058	±50
*Surface gravity, log g	[g in cm s^{-2}]	4.55	±0.10
*Metallicity, [m/H]		-0.18	±0.08
†Mass of the star, M_*	[M$_\odot$]	0.758	±0.046
†Radius of the star, R_*	[R$_\odot$]	0.737	+0.034 -0.042
‡Ca II H&K activity indicator, log(R'$_{\mathrm{HK}}$)		-4.52	±0.015§
‡Projected rotational velocity, v sin$i$	[km s^{-1}]	2.8	±0.5

* Obtained from an SPC analysis of the HARPS-N spectra. †Based on the Yonsei–Yale stellar evolutionary models[13]. ‡Output of the HARPS-N data-reduction pipeline using $B-V = 0.91$. §The error indicates the standard deviation of the individual values.

Extended Data Table 2 | Comparison of the statistical 'quality' of all the considered models

Model	Description	Stellar Jitter	Planet	$X2$	$\chi2$	N_{par}	ER	ΔER	w
1	D0	No	None	374.38	3.47	1	379.07	209.2	0.00
2	D0 + K1	No	K78 (fixed)	326.43	3.05	2	335.81	166.0	0.00
3	D0 + nightly offsets	21 offsets	K78 (fixed)	89.90	1.03	22	193.11	23.3	0.00
4	D0 + K2	2 Keplerians (P2=4.2; e2=0; P3=10.2; e3 adj)	None	166.03	1.69	11	217.63	47.8	0.00
5	D0 + K3	2 Keplerians (P2=4.2; e2=0; P3=10.2; e3 adj)	K78 (fixed)	122.94	1.24	10	**169.85**	0.00	**0.710**
6	D0 + K4	3 Keplerians (P2=4.2; e2=0; P3=6.5; e3 adj; P4=23.58; e4=0)	K78 (fixed)	120.40	1.23	11	172.00	2.2	0.242

Model 5 with the three Keplerians is clearly the one best representing the data, given the fact that its BIC evidence ratio, ER, is lowest and its BIC weight, w, highest. For comparison, the χ^2 and the reduced χ^2 of the residuals are also given. It is interesting to note that the most likely model does not necessarily have the lowest reduced χ^2. N_{par} indicates the number of free parameters in the model. D0 represents a constant term, Kn is the number of Keplerians in the model. e_i and P_i respectively indicate the eccentricity and the period of the ith planet.

Extended Data Table 3 | Orbital parameters (distributions) of the planet and parameters of the two additional Keplerians describing the star-induced signal as determined from the MCMC analysis

Parameters	Unit	Distribution mode	99% Confidence Interval
P_p	[days]	0.35500743	0.35500729 - 0.35500760
K_p	[m s^{-1}]	1.986	1.243 - 2.679
e_p		0	fixed
λ_{0p}	[deg]	297.53	296.82 - 298.23
T_{0p}	[days]	1465.165144271	1465.165144235 - 1465.165144313
$m_p \sin(i)$	[M]	1.82	1.12 - 2.51
a_p	[10^{-3} AU]	8.95323	8.448 - 9.393
P_2	[days]	4.210	4.193 - 4.237
K_2	[m s-1]	3.98	2.66 - 5.29
e_2		0	fixed
λ_{02}	[deg]	146	118 -166
P_3	[days]	10.037	9.977 - 10.272
K_3	[m s-1]	5.33	4.14 - 6.78
e_3		0.39	0.03 - 0.65
λ_{03}	[deg]	-320.0	-329 - -310
ω_3	[deg]	154	91 - 262

P_i represents the (orbital) period, K the semi-amplitude, e_i the eccentricity and λ_{0i} the mean longitude of the ith signal at the reference epoch, T_0. The index i designates the planet (p) and the two additional (stellar) signals (2 and 3). $m_p \sin(i)$ is the minimum mass of the planet and a_p is the semi-major axis of the orbit.

Extended Data Table 4 | Planetary parameters derived from the MCMC analysis

Statistical parameter	Mass [$M_\oplus$]	Radius [$R_\oplus$]	Density [g cm^{-3}]
Mode	1.86	1.173	5.57
Median	1.91	1.194	6.13
68.3% confidence interval	1.61 – 2.24	1.084 – 1.332	4.26 – 8.59
99% confidence interval	1.17 – 3.00	0.942 – 1.584	2.34 – 14.29

The table gives the distributions of the mass, radius and density of Kepler-78b.