JWST's little red dots: an emerging population of young, low-mass AGN cocooned in dense ionized gas

V. Rusakov[1,2,3], D. Watson[2,3], G. P. Nikopoulos[2,3], G. Brammer[2,3], R. Gottumukkala[2,3], T. Harvey[1], K. E. Heintz[2,3,4], R. D. Nielsen[2,3], S. A. Sim[5,2,3], A. Sneppen[2,3], A. P. Vijayan[6], N. Adams[1], D. Austin[1], C. J. Conselice[1], C.M. Goolsby[1], S. Toft[2,3]

***JWST* has uncovered large numbers of compact galaxies at high redshift with broad hydrogen/helium lines. These include the enigmatic population known as "little red dots" (LRDs)[1,2]. Their nature is debated, but they are thought to be powered by supermassive black holes (SMBHs)[3,4] or intense star formation[5]. They exhibit unusual properties for SMBHs, such as black holes that are overmassive for their host galaxies[4] and extremely weak X-ray[6–9] and radio[10–12] emission. Using the highest-quality *JWST* spectra, we show here that the lines are broadened by electron scattering with a narrow intrinsic line core. The data require high electron column densities and compact sizes (light days), which, when coupled with their high luminosities can only be explained by SMBH accretion. The narrow intrinsic cores of the lines imply upper limits on the black hole masses of $10^{5-7}\,M_\odot$, two orders of magnitude lower than previous estimates. These are among the lowest mass SMBHs known at high redshift and suggest that this is a population of young, rapidly growing SMBHs. They are enshrouded in a dense cocoon of ionized gas, probably related to their youth, from which they are accreting close to the Eddington limit. Reprocessed nebular emission from the dense cocoon dominates the optical spectrum, explaining most LRD spectral characteristics and helping to suppress radio and X-ray emission.[13,14]**

Two early clues suggested that these galaxies may be affected by Compton-thick ionized gas: Balmer absorption features are often observed in the broad lines[2,15,16], and if they are active galactic nuclei (AGN), their X-rays might be suppressed by photoelectric absorption in Compton-thick ionized gas in the broad-line region (BLR) near the SMBH[6]. These ideas inspired us to investigate the Hα line profiles to look for electron scattering signatures produced by the large ionized gas column densities.

To test different broad line shapes, we constructed a sample of all broad line galaxies (Hα $\gtrsim$ 1000 km s^{-1}) with high signal-to-noise ratio (SNR) *JWST*/NIRSpec medium-resolution ($R \sim 1000$) spectra [17] in the DAWN *JWST* Archive (DJA)[18]. Details of the sample and selection criteria are presented in the Methods and Extended Data Table 2. Our search yielded 12 objects at $z = 3.4-6.7$ and 18 additional objects for a combined 'stacked' spectrum at $z = 2.32-6.76$. High-resolution ($R \sim 2700$) data were used where available (Objects A and C) and gave consistent results but better constrained narrow component widths.

Our object selection is strongly linked to LRDs, many of which show broad lines. LRDs are spatially compact with characteristic 'v'-shaped optical-UV spectra[1,19], due to a change of slope close to the Balmer limit wavelength, ~ 365 nm[20]. Despite our selection being based only on Hα

linewidth and SNR criteria, our systems are all spatially very compact (Fig. 2), and often have a clear slope change at the Balmer limit in the restframe spectra (see Extended Data Fig. 1).

Electron scattering in a dense ionized gas produces lines with exponential profiles[21–23], while Gaussian or centrally broad profiles are expected from the typical Doppler broadening due to bulk gas motion around SMBHs[24]. We therefore compare two basic models for the broad Hα component: a Gaussian and a double-sided exponential. Raman scattering[9] or turbulence[24] give rise to broad Lorentzian lines. However, since we find the exponential to be a better or equivalent fit in all cases, for clarity of presentation, we discuss the Lorentzian model in the Methods section. In addition to the broad component and a local continuum in the Hα region, we also model the narrow emission lines that cannot easily be masked out: Hα and the [N II] lines $\lambda\lambda$6549,6585. Although rare in star-forming galaxies or AGN at lower redshift, relatively narrow Balmer absorption features are also a common feature in these objects[2] and can be seen on top of the Hα line of many of our sample objects (Fig. 3). They are accounted for in our fits as an absorption line in object E and as P Cygni absorption/emission features in A and D. For the Gaussian/exponential comparison, we examine the quality of these simpler fits and find that while the Gaussian fit leaves systematic residuals with a characteristic 'W'-shape around the central region (Fig. 1), the exponential model is superior in almost every case. The significance of the fit improvement increases with the SNR of the spectra (Extended Data Fig. 6), an indication that the exponential model is the better fit in the general case, even where the spectra do not have high enough SNR to discriminate. The comparison fits to all spectra and the stacked spectrum are shown in Extended Data Fig. 4. Fit statistics are given in Extended Data Table 3. To illustrate more explicitly the exponential nature of the lines, we plot the data with a logarithmic ordinate scale (Fig. 1, *right* and Extended Data Fig. 5). The lines are symmetric and trace a straight line on a semi-logarithmic plot over several orders of magnitude, consistent with moderate optical depth electron-scattering[22]. These results imply that the primary line-broadening mechanism is electron-scattering through a Compton-thick medium and not primarily bulk Doppler motions, immediately excluding high mass SMBHs. The symmetry of the lines suggests that any outflow of the scattering medium is less than a few hundred $\mathrm{km\,s^{-1}}$ (ref. [23]).

After establishing that the basic broad line shape is dominated by electron-scattering effects, we allow that they also have an intrinsic Doppler line core. For our fiducial model we therefore use a Gaussian convolved with an exponential, with the widths free to vary, to measure the intrinsic width of the Doppler lines. Fits for all objects are shown in (Fig. 3). In most cases the intrinsic line width is small; we find average intrinsic velocities of approximately $300\,\mathrm{km\,s^{-1}}$ for ten of the twelve objects in the sample. For objects B and G we recover relatively large Gaussian intrinsic widths, about $2000\,\mathrm{km\,s^{-1}}$. However, we caution that specifically these spectra have the most complex line shapes in our sample, with three apparent peaks. It is probable that our simple model is not adequately reproducing complex radiative transfer effects through a (highly) optically thick medium. We therefore reserve judgement on whether objects B and G really have such large Doppler velocities.

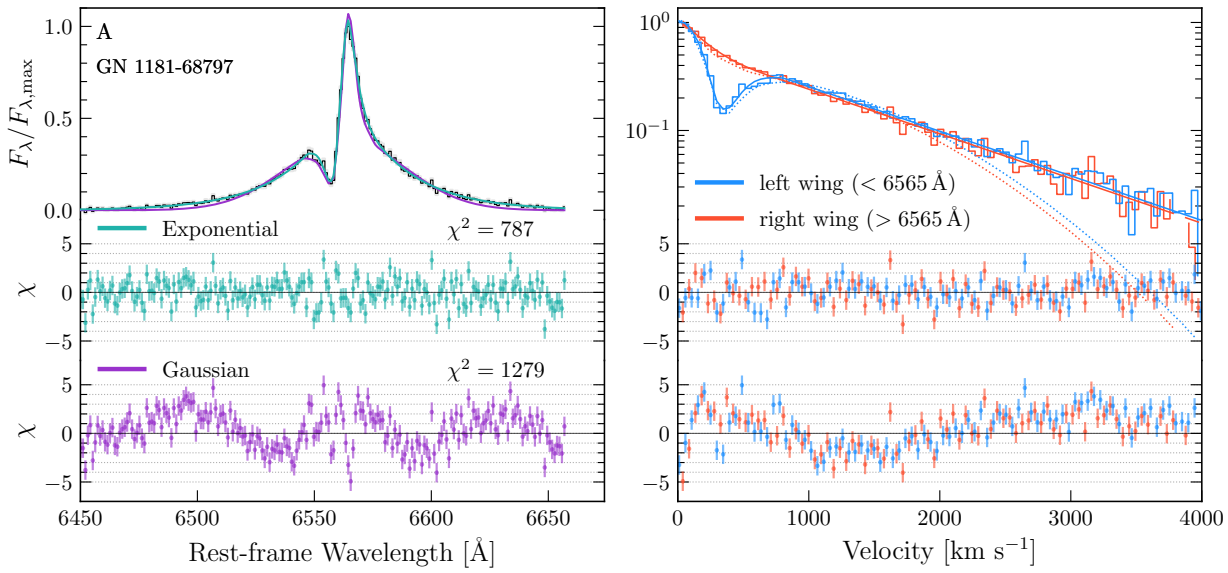


Figure 1: The Hα line profile for GN 1181-68797 (object A) showing the exponential shape of the line. The models in the figure compare the fiducial best-fit model that includes a scattered exponential component with an identical model that instead includes a double Gaussian component. Object A has the highest SNR(Hα) in our sample and was observed with the high-resolution grating on *JWST*/NIRSpec. The absorption feature is modelled with a P Cygni component. *Left:* Gaussian and Exponential-component models compared in linear flux density space. *Right:* Hα line profile plotted in semi-logarithmic space, reflected about the line centre, showing the linearity of the line profile in this space. The wings of the lines are also clearly symmetric. The exponential line shape is significantly preferred by the data both around the line core and the tails, while the Gaussian model falls below the line core and the tails.

The characteristic properties of the gas can be inferred from the line widths. First, we infer the electron column density, N_{e^-}, from the width of the lines[22,23], assuming an electron temperature between 10,000 and 30,000 K. We use our own Monte Carlo electron scattering code with a simple shell geometry to calculate the relationship between the electron-scattering optical depth, τ_e, the temperature, and the characteristic width of the exponential (see Methods). We find $\tau_e = 0.5-2.8$, implying $N_e \simeq 0.7-4.2 \times 10^{24}\,\mathrm{cm}^{-2}$ for the sample (see Table 1 and Extended Data Fig. 8). The absence of strong broadening of the [O III] $\lambda\lambda 4959, 5007$ doublet, while the Hβ line next to it is broadened, suggests that the electron volume density, n_e, in the scattering region, must be at least several times the critical density for these lines. The characteristic spherical size of the scattering region, R_c, can be inferred by taking the ratio of the column and volume densities in a simple constant density sphere, i.e. $R_c \sim 3N_e/n_e$. Assuming $\log(n_e/\mathrm{cm}^{-3}) \gtrsim 6.5$, we find sizes of at most a few hundred light days and ionized gas masses of order $\lesssim 10^5\,\mathrm{M}_\odot$. If we assume densities more in line with estimates in the literature for these objects (i.e. $\log(n_e/\mathrm{cm}^{-3}) \gtrsim 8$[26,27], the sizes would be equivalently smaller, about a few light days.

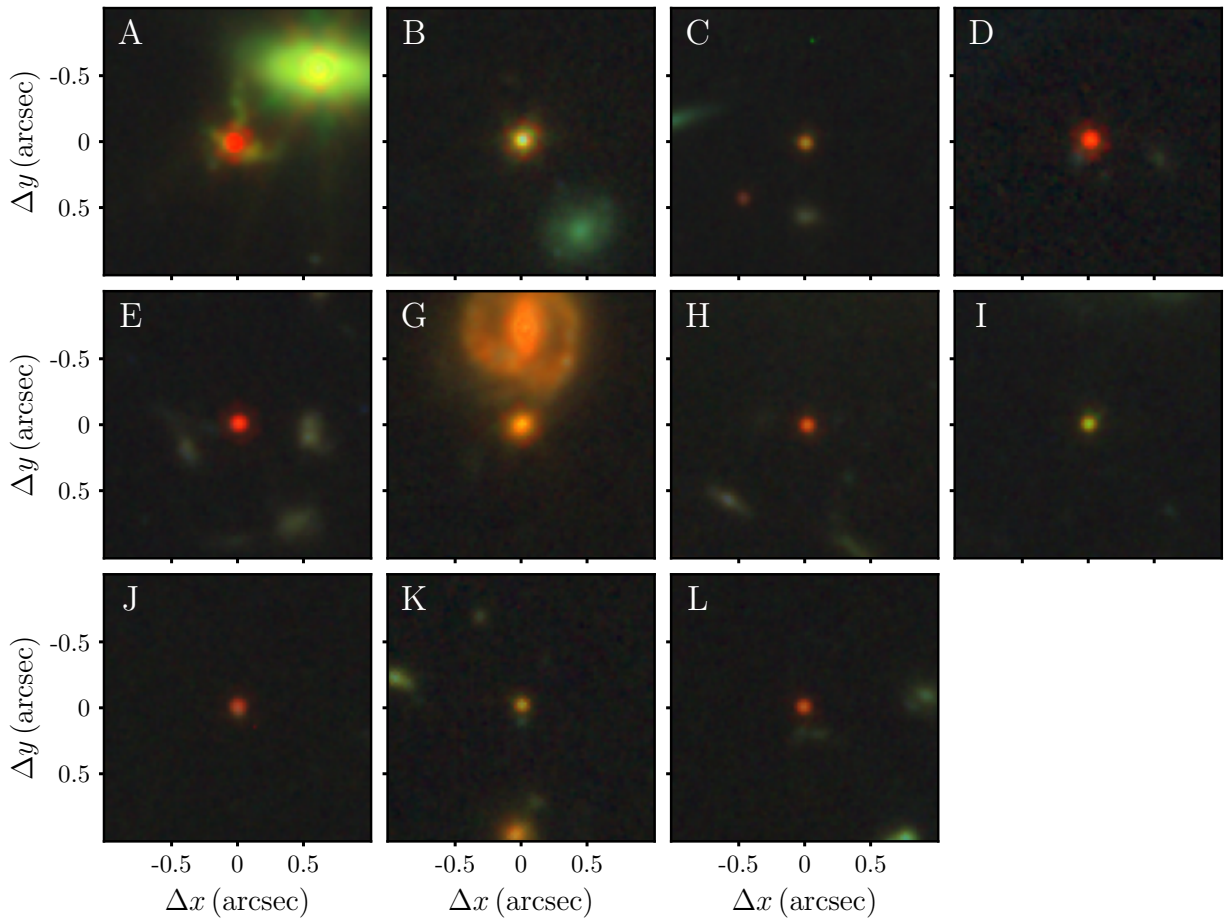


Figure 2: *JWST*/NIRCam images of the sample objects. Despite being selected based on the presence of broad Hα components, all objects appear to be point-like or very compact, occasionally with some nebulosity. Restframe optical colours are mostly red. These properties are similar to LRDs (see also Extended Data Figs. 1 and 2). Object A has an unusual extended cross-like structure. We note that this object also shows rotation in the 2D high-resolution spectra of the narrow nebular lines. Object F was not observed by NIRCam. The RGB colouring is scaled using `Grizli`[25] as a wavelength-dependent power-law (with exponent of 2) based on the fluxes in the F150W (or F200W), F277W, and F444W bands.

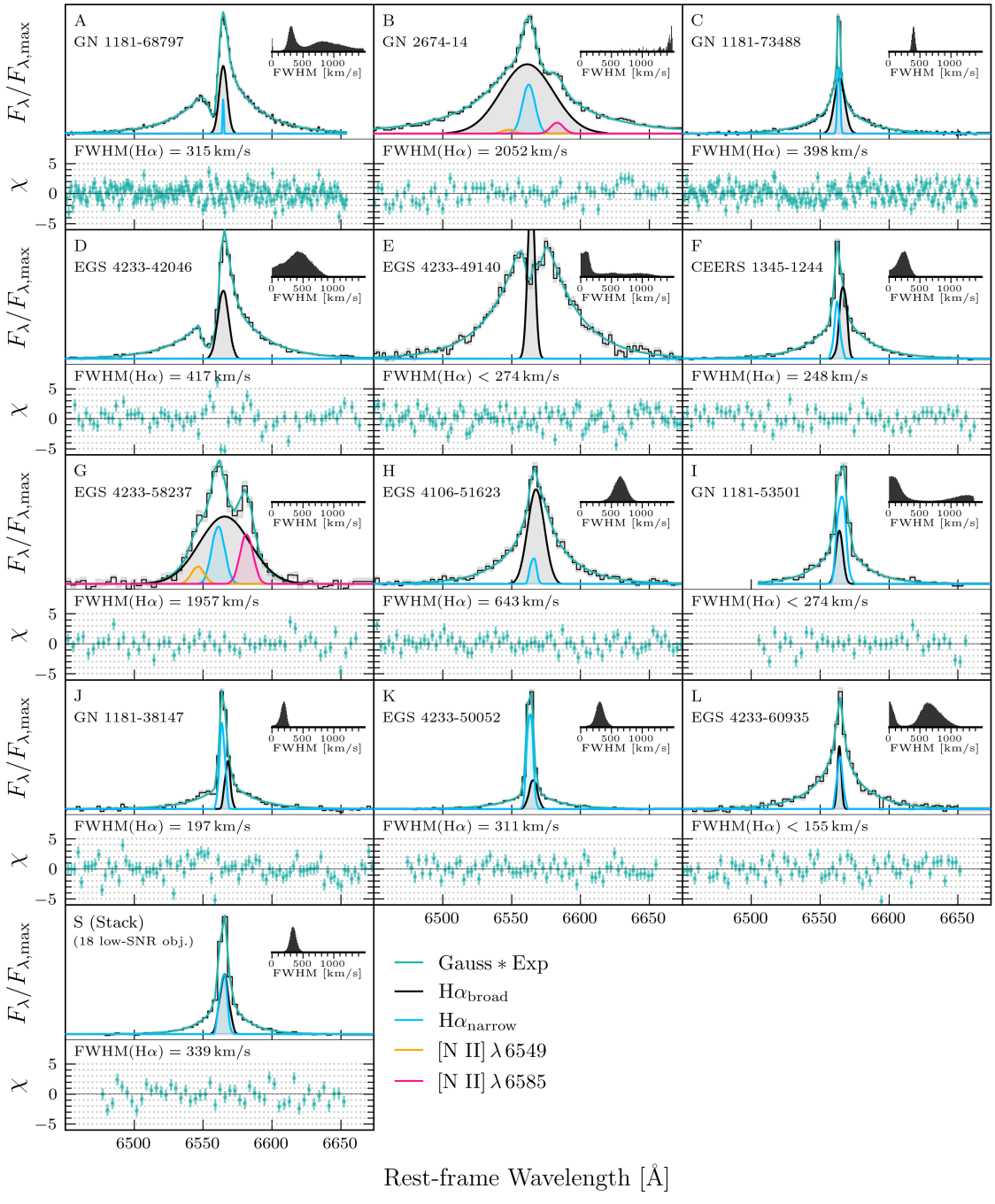


Figure 3: Hα line profiles for the full sample fit with the fiducial scattering model. The total model (cyan line) is the broad scattered and non-scattered components of the intrinsic gaussian line (black line and grey fill), and a narrow Gaussian Hα component from the host galaxy (blue line), as shown in the legend. In several cases, a P Cygni profile or a Gaussian absorption feature are included. The inset panels show the posterior distribution of the FWHM of the intrinsic Doppler component of the broad Hα. In all cases except for objects B and G, which are complex and may not be well-modelled by our simple approach, the widths of the broad Hα lines are dominated by the electron-scattering mechanism (1000–2000 km s^{-1}), while the Doppler motions are either unresolved or on the order of several hundred km s^{-1}.

The dense Compton-thick gas inferred from the line wings offers a natural explanation for the Balmer absorption lines frequently seen in the broad-line systems at high redshift[2,27]. Given the extreme column and volume densities of the ionised gas inferred here, the fraction of the gas in the $n = 2$ level should provide sufficient optical depth to produce these features. We model the features with P Cygni profiles in the spectra of A and D. The P Cygni model gives a much better fit than a pure absorption line due to the sharpness of the transition between the emission peak and the blueshifted absorption feature, characteristic of the P Cygni profile[28]. The properties of the absorption features indicate a largely spherical gas distribution, with mild outflow velocities below a few hundred $\mathrm{km\,s^{-1}}$. In cases of high optical depth and lower velocities, radiative transfer effects may instead induce a complex self-absorption feature in the line centre. The presence of narrow emission lines from the rest of the galaxy may make the profile even more complicated or conceal these features, making it difficult to assess how common such Balmer absorption is. A full investigation of the Balmer absorption lines is beyond the scope of this paper and we defer their analysis to future work.

We now turn to the origin of the lines in these systems. The line widths in most of our sample could be explained by very mild outflows, consistent with the photospheric velocities we find for the P Cygni lines in objects A and D (about 200–300 $\mathrm{km\,s^{-1}}$). Such a mild outflow could be due to feedback from a burst of star formation. The largest maximum radii we infer are roughly consistent with the smaller star clusters observed in R136 in the LMC and NGC 253[29] in the local universe, or in the Sunrise Arc at $z \sim 6$ or the Cosmic Gems system at redshift $z = 10.0$ [30]. However, the luminosity densities of those star clusters are typically about $10^{41}\,\mathrm{erg\,s^{-1}}$ per pc^2 or per pc^3. The more luminous systems in our sample, on the other hand, have ionizing luminosities of $\gtrsim 10^{45}\,\mathrm{erg\,s^{-1}}$, based on observed H$\alpha$ luminosities around $10^{43}\,\mathrm{erg\,s^{-1}}$ [31,32]. This is nearly four orders of magnitude higher than the highest luminosity density star-formation known. Distributing the inferred luminosity across multiple sources over a larger area or volume (e.g. many accreting extremely massive stars[33]) doesn't resolve this problem because the inferred gas masses require luminosities $\gtrsim 10^9 \mathrm{L}_\odot$ to keep them ionized. Because these inferred luminosity densities are based on the profiles of the lines, which represent the average conditions of the emitting regions, this requires each of the putative individual sources to be this luminous. These considerations essentially rule out a star-forming origin for the broad lines. Production of such large ionising luminosities in regions significantly smaller than a few hundred light days can only realistically be powered by accretion onto a SMBH, i.e. an active galactic nucleus (AGN). The AGN radius-luminosity relation implies a Hα broad-line region radius[34] of approximately five light days for a Hα luminosity of $10^{43}\,\mathrm{erg\,s^{-1}}$. This is consistent with the upper bounds inferred above. If this radius is accurate, it suggests that the volume density in the scattering regions may be somewhat lower than typical BLR cloud densities.

As mentioned above, the AGN interpretation of the broad line objects and LRDs discovered by *JWST* faces several challenges: the large inferred black hole masses, both compared to their host galaxies, and as a function of the age of the universe and their cosmic number density, and the lack of X-ray and radio detections. First, as the electron scattering dominates the line widths

in these systems, we need to use the intrinsic line widths prior to scattering, to infer the black hole masses, assuming these intrinsic line widths are entirely due to Doppler orbital motions around the SMBH. These are an order of magnitude lower than inferred with a simple Gaussian fit to the full line profile. The inferred black hole masses, using standard single-epoch black hole mass scaling relations[35], are therefore lower by about a factor of a hundred (Table 1). Our estimates of the maximum SMBH masses are consistent with the typical host galaxy stellar mass found at lower redshifts[35], as shown in Fig. 4, alleviating much of this problem. The much lower black hole masses also alleviate the number density and early formation and growth problems. Second, Compton-thick ionized gas will hinder X-ray[26] and suppress radio[27] emission, by photoelectric and free-free absorption[36] respectively. This is true in X-rays even at low metallicity due to He absorption[37]. The ionized column densities we infer here will attenuate the X-rays by only a factor of a few (see Methods)–less than the ~ 1 dex required to satisfy the observations[6–8]. Steep hard X-ray spectral slopes and/or relatively low energy power-law cut-offs are also required to suppress the X-rays[6], both of which are found in narrow-line Seyfert 1 objects[38]. Narrow-line Seyfert 1s are believed to be low-mass, high accretion rate SMBHs[39].

The high density gas reprocesses essentially all of the Lyman continuum radiation from the central source, resulting in extremely high Lyman optical depths that cause the self-reversal seen as absorption in the line centres of some of the Balmer lines. The bulk of the ionizing flux from the AGN is therefore emitted in the Balmer and Paschen regions of the spectrum, implying that nebular gas emission must be a major component of the optical and NIR spectra of these sources, giving rise to Balmer, Paschen and He I emission lines and continua and the two-photon continua that would dictate the spectral shape[13,40]. In addition, while many of these systems appear in blue star-forming galaxies based on the spatially-extended nebulosity surrounding the point source[41], many may be entirely dominated by the nebular emission at optical and even near UV wavelengths. We also note that our inferred electron-scattering optical depths are all $\tau_e \sim 1$ (see Extended Data Fig. 8). This suggests three things. First, that there may be a population that is even more heavily obscured, suppressing line emission due to self-absorption, and giving rise to spectra that appear in absorption rather than emission, including a Balmer break instead of a jump. Second, it also suggests that the gas distribution is close to spherical without a large opening, because scattered light will be preferentially emitted along the lowest column density sightlines. The fact that we are not detecting large numbers of low-mass, rapidly accreting AGN without these signatures indicates that LRDs are the primary population and that the gas cocoon in these SMBHs is a near-universal, young phase of their accretion-powered growth. Third, that the line ratios of the Balmer lines should be very different from the standard case B values due to very high Lyman optical depths [42], so that high Hα/Hβ ratios do not imply dust extinction.

Another conundrum of these sources was that they are found extremely early in the history of the universe with (previously inferred) very large masses and low accretion rates[16]. If they really represent the early stages of black hole growth and require high accretion rates over very long times to grow rapidly[43], they should be found in high accretion states when they are brighter. The lower black hole masses we infer here solve this puzzle. We calculate a mean Eddington

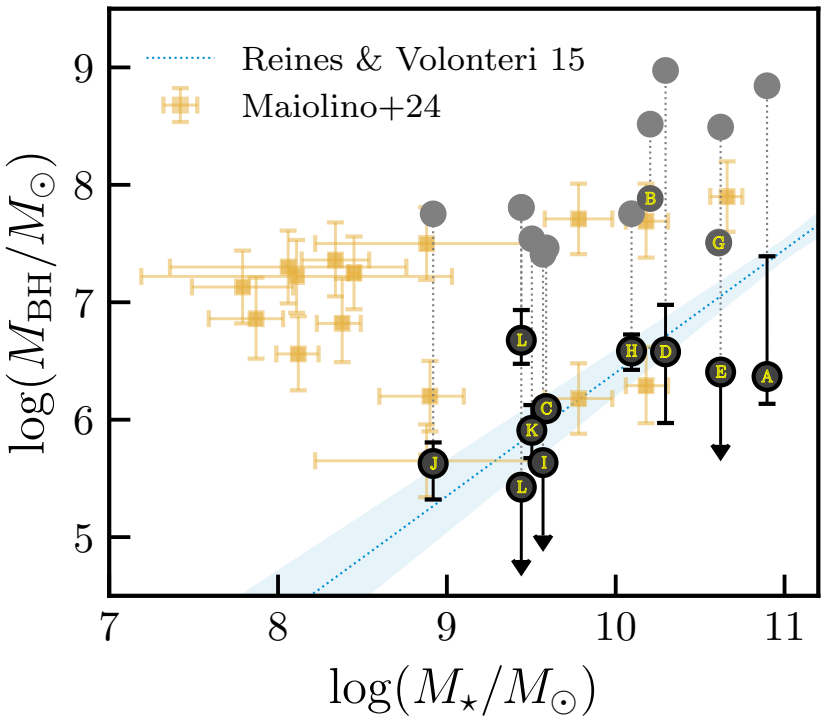


Figure 4: Black hole masses compared to the stellar mass of the host galaxy inferred from SED fitting. The SMBH masses are determined from the Doppler components of our fiducial model (letter-labelled points). We also show the black hole masses inferred with a Gaussian model without broadening by scattering accounted for (non-outlined grey points) to illustrate the difference the scattering model makes to the inferred black hole mass. Objects B and G may have complex absorption features that could substantially change their $M_{\rm BH}$—they are therefore plotted without a black outline. The posterior probability distribution for L is strongly bimodal, so $\log(M_{\rm BH})$ is split into two points according to the posterior solutions for $\rm FWHM_{Doppler}$ in Table 4. Previous estimates from the literature are also shown (yellow points), most of which are included in our low-SNR stack, which has median inferred $\log(M_{\rm BH}) \lesssim 5.6$. Taking account of the broadening due to scattering reduces the inferred SMBH masses by about two orders of magnitude and make the masses consistent with the observed relation of AGN at lower redshift[35]. It is important to note, however, that the stellar mass inferred from SED fitting is an upper limit on the real stellar mass, since a large fraction of the emission in the optical is radiation from the AGN, reprocessed through the ionized gas, rather than emission from stars.

Table 1: Derived properties of the broad Hα systems based on the best-fit fiducial model. We report the estimates of the optical depth, τ, and associated gas column densities, N_{e^-}, based on our electron-scattering gas model and the exponential line widths. We estimate SMBH masses using the relation of ref [35]. Stellar mass estimates from best-fit photometric SED models, are denoted as upper limits because the likely dominant contribution from the AGN's nebular emission to the restframe optical continuum is neglected.

ID	Field PID-SRCID	Redshift	τ	N_{e^-} (10^{24} cm^{-2})	$\mathrm{M_{BH}}$	$\mathrm{M_{BH_{Gauss}}}$	$\mathrm{M_\star}$
					——— ($\log \mathrm{M_\odot}$) ———		
A	GN 1181-68797	5.0405	$1.2^{+0.2}_{-0.1}$	$1.8^{+0.3}_{-0.1}$	$6.4^{+1.0}_{-0.2}$	8.8	< 10.9
B	GN 2674-14	5.1826	$2.8^{+0.3}_{-0.3}$	$4.2^{+0.5}_{-0.4}$	$7.9^{+0.1}_{-0.1}$	8.5	< 10.2
C	GN 1181-73488	4.1327	$0.9^{+0.1}_{-0.1}$	$1.3^{+0.2}_{-0.1}$	$6.1^{+0.1}_{-0.1}$	7.5	< 9.6
D	EGS 4233-42046	5.2757	$1.2^{+0.2}_{-0.1}$	$1.8^{+0.3}_{-0.1}$	$6.6^{+0.4}_{-0.6}$	9.0	< 10.3
E	EGS 4233-49140	6.6847	$1.3^{+0.2}_{-0.1}$	$1.9^{+0.2}_{-0.2}$	< 6.4	8.5	< 10.6
F	EGS 1345-1244	4.4771	$1.0^{+0.1}_{-0.1}$	$1.5^{+0.2}_{-0.1}$	$5.9^{+0.3}_{-0.6}$	7.9	—
G	EGS 4233-58237	3.6505	< 0.8	< 1.2	$7.5^{+0.1}_{-0.1}$	7.5	< 10.6
H	EGS 4106-51623	4.9511	$1.0^{+0.1}_{-0.1}$	$1.5^{+0.2}_{-0.2}$	$6.6^{+0.1}_{-0.2}$	7.8	< 10.1
I	GN 1181-53501	3.4294	$0.7^{+0.2}_{-0.1}$	$1.1^{+0.2}_{-0.2}$	< 5.6	7.4	< 9.6
J	GN 1181-38147	5.8694	$0.9^{+0.1}_{-0.1}$	$1.4^{+0.2}_{-0.2}$	$5.6^{+0.2}_{-0.3}$	7.8	< 8.9
K	EGS 4233-50052	5.2392	$0.9^{+0.1}_{-0.1}$	$1.3^{+0.2}_{-0.2}$	$5.9^{+0.2}_{-0.2}$	7.5	< 9.5
L[a]	EGS 4233-60935	5.2877	$0.9^{+0.2}_{-0.2}$	$1.4^{+0.3}_{-0.2}$	< 5.4 $6.7^{+0.3}_{-0.2}$	7.8	< 9.4
S[b]	—	—	$0.5^{+0.1}_{-0.1}$	$0.7^{+0.1}_{-0.1}$	—	—	< 8.9

[a]Object L has two estimates of M_{BH} corresponding to two distinct FWHM solutions.
[b]$\mathrm{M_\star}$ for the stack is a median of the individual masses.

ratio around unity for our sources A–L, excluding objects B and G, assuming the Hα line is a few percent of the bolometric luminosity. These sources are indeed accreting close to, or in some cases, slightly above, the Eddington limit, as would be expected if we are discovering a population of early, fast-growing SMBHs. As our sample is likely to be biased in favour of more luminous systems, the bulk of the population may have even lower black hole masses, lower accretion rates, and/or greater dust extinction. This suggests that SMBH masses at the $10^5\,M_\odot$ level could be a substantial part of the population.

All of this points to a solution to *JWST*'s riddle of LRDs/compact broad-line objects. They are intrinsically narrow-line AGN, i.e. young[44], low-mass (10^5–$10^7\,M_\odot$), Eddington-accreting SMBHs[39], buried in a thick cocoon of gas, presumably related to their early stage of formation[45]. Their high accretion rates produce copious UV emission that ionizes their gas cocoon and at the same time efficiently cool and weaken the corona, suppressing their hard X-rays[39]. The ionized gas hinders the escape of radio and X-rays, while reprocessing almost all the Lyman radiation into nebular optical emission, producing the broadened Balmer lines and continuum breaks that characterise the classic 'v'-shape spectrum[20]. Many galaxies have been found with stellar masses and ages that are so massive and old that they are difficult to explain in standard scenarios[46,47]. Our results and other recent work[13,14] indicate that galaxies with red colours and even those with Balmer breaks, may, instead of being massive, quiescent or post-starburst galaxies, be low-mass AGN systems with extreme optical depth ionized gas. The distribution of electron column densities and the lack of a similar or larger population of low-mass, high accretion AGN without the nebular cocoon hints that we may be observing the bulk growth of SMBHs in the phase when they are surrounded by a quasi-spherical, dense gas shell, before angular momentum transport has flattened their accretion into a well-defined disk.

References

1. Greene, J. E., Labbe, I., Goulding, A. D., Furtak, L. J. et al., "UNCOVER Spectroscopy Confirms the Surprising Ubiquity of Active Galactic Nuclei in Red Sources at $z > 5$". *ApJ*, **964**(1), 39, March (2024).

2. Matthee, J., Naidu, R. P., Brammer, G., Chisholm, J. et al., "Little Red Dots: An Abundant Population of Faint Active Galactic Nuclei at z $\sim$ 5 Revealed by the EIGER and FRESCO JWST Surveys". *ApJ*, **963**(2), 129, March (2024).

3. Harikane, Y., Zhang, Y., Nakajima, K., Ouchi, M. et al., "A JWST/NIRSpec First Census of Broad-line AGNs at z = 4-7: Detection of 10 Faint AGNs with $M_{\rm BH} \sim 10^6 - 10^8\,M_\odot$ and Their Host Galaxy Properties". *ApJ*, **959**(1), 39, December (2023).

4. Maiolino, R., Scholtz, J., Curtis-Lake, E., Carniani, S. et al., "JADES: The diverse population of infant black holes at $4 < z < 11$: Merging, tiny, poor, but mighty". *A&A*, **691**, A145, November (2024).

5. Baggen, J. F. W., van Dokkum, P., Brammer, G., de Graaff, A. et al., "The Small Sizes and High Implied Densities of "Little Red Dots" with Balmer Breaks Could Explain Their Broad Emission Lines without an Active Galactic Nucleus". *ApJ*, **977**(1), L13, December (2024).

6. Maiolino, R., Risaliti, G., Signorini, M., Trefoloni, B. et al., "JWST meets Chandra: a large population of Compton thick, feedback-free, and intrinsically X-ray weak AGN, with a sprinkle of SNe". *MNRAS*, , February (2025).

7. Yue, M., Eilers, A.-C., Ananna, T. T., Panagiotou, C. et al., "Stacking X-Ray Observations of "Little Red Dots": Implications for Their Active Galactic Nucleus Properties". *ApJ*, **974**(2), L26, October (2024).

8. Ananna, T. T., Bogdán, Á., Kovács, O. E., Natarajan, P. et al., "X-Ray View of Little Red Dots: Do They Host Supermassive Black Holes?". *ApJ*, **969**(1), L18, July (2024).

9. Kokubo, M. and Harikane, Y., "Challenging the AGN scenario for JWST/NIRSpec broad Hα emitters/Little Red Dots in light of non-detection of NIRCam photometric variability and X-ray". *arXiv e-prints*, , arXiv:2407.04777, July (2024).

10. Mazzolari, G., Gilli, R., Maiolino, R., Prandoni, I. et al., "The radio properties of the JWST-discovered AGN". *arXiv e-prints*, , arXiv:2412.04224, December (2024).

11. Gloudemans, A. J., Duncan, K. J., Eilers, A.-C., Farina, E. P. et al., "Another piece to the puzzle: radio detection of a JWST discovered AGN candidate". *arXiv e-prints*, , arXiv:2501.04912, January (2025).

12. Setton, D. J., Greene, J. E., Spilker, J. S., Williams, C. C. et al., "A confirmed deficit of hot and cold dust emission in the most luminous Little Red Dots". *arXiv e-prints*, , arXiv:2503.02059, March (2025).

13. Inayoshi, K. and Maiolino, R., "Extremely Dense Gas around Little Red Dots and High-redshift Active Galactic Nuclei: A Nonstellar Origin of the Balmer Break and Absorption Features". *ApJ*, **980**(2), L27, February (2025).

14. Ji, X., Maiolino, R., Übler, H., Scholtz, J. et al., "BlackTHUNDER – A non-stellar Balmer break in a black hole-dominated little red dot at $z = 7.04$". *arXiv e-prints*, , arXiv:2501.13082, January (2025).

15. Kocevski, D. D., Finkelstein, S. L., Barro, G., Taylor, A. J. et al., "The Rise of Faint, Red AGN at $z > 4$: A Sample of Little Red Dots in the JWST Extragalactic Legacy Fields". *arXiv e-prints*, , arXiv:2404.03576, April (2024).

16. Juodžbalis, I., Maiolino, R., Baker, W. M., Tacchella, S. et al., "A dormant overmassive black hole in the early Universe". *Nature*, **636**(8043), 594–597, December (2024).

17. Jakobsen, P., Ferruit, P., Alves de Oliveira, C., Arribas, S. et al., "The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. I. Overview of the instrument and its capabilities". *A&A*, **661**, A80, May (2022).

18. Heintz, K. E., Brammer, G. B., Watson, D., Oesch, P. A. et al., "The JWST-PRIMAL archival survey: A JWST/NIRSpec reference sample for the physical properties and Lyman-α absorption and emission of $\sim$600 galaxies at z = 5.0 - 13.4". *A&A*, **693**, A60, January (2025).

19. Furtak, L. J., Zitrin, A., Plat, A., Fujimoto, S. et al., "JWST UNCOVER: Extremely Red and Compact Object at $z_{\rm phot} \simeq 7.6$ Triply Imaged by A2744". *ApJ*, **952**(2), 142, August (2023).

20. Setton, D. J., Greene, J. E., de Graaff, A., Ma, Y. et al., "Little Red Dots at an Inflection Point: Ubiquitous "V-Shaped" Turnover Consistently Occurs at the Balmer Limit". *arXiv e-prints*, , arXiv:2411.03424, November (2024).

21. Weymann, R. J., "Electron-Scattering Line Profiles in Nuclei of Seyfert Galaxies". *ApJ*, **160**, 31, April (1970).

22. Laor, A., "Evidence for Line Broadening by Electron Scattering in the Broad-Line Region of NGC 4395". *ApJ*, **643**(1), 112–119, May (2006).

23. Huang, C. and Chevalier, R. A., "Electron scattering wings on lines in interacting supernovae". *MNRAS*, **475**(1), 1261–1273, March (2018).

24. Kollatschny, W. and Zetzl, M., "The shape of broad-line profiles in active galactic nuclei". *A&A*, **549**, A100, January (2013).

25. Brammer, G. "grizli", September (2023).

26. Inayoshi, K. and Maiolino, R., "Extremely Dense Gas around Little Red Dots and High-redshift Active Galactic Nuclei: A Nonstellar Origin of the Balmer Break and Absorption Features". *ApJ*, **980**(2), L27, February (2025).

27. Juodžbalis, I., Ji, X., Maiolino, R., D'Eugenio, F. et al., "JADES - the Rosetta stone of JWST-discovered AGN: deciphering the intriguing nature of early AGN". *MNRAS*, **535**(1), 853–873, November (2024).

28. Castor, J. I. and Lamers, H. J. G. L. M., "An atlas of theoretical P Cygni profiles.". *ApJS*, **39**, 481–511, April (1979).

29. Leroy, A. K., Bolatto, A. D., Ostriker, E. C., Walter, F. et al., "Forming Super Star Clusters in the Central Starburst of NGC 253". *ApJ*, **869**(2), 126, December (2018).

30. Adamo, A., Bradley, L. D., Vanzella, E., Claeyssens, A. et al., "Bound star clusters observed in a lensed galaxy 460 Myr after the Big Bang". *Nature*, **632**(8025), 513–516, August (2024).

31. Greene, J. E. and Ho, L. C., "Estimating Black Hole Masses in Active Galaxies Using the Hα Emission Line". *ApJ*, **630**(1), 122–129, September (2005).

32. Calzetti, D. "Star Formation Rate Indicators". In *Secular Evolution of Galaxies,* Falcón-Barroso, J. and Knapen, J. H., editors, 419. (2013).

33. Gieles, M., Padoan, P., Charbonnel, C., Vink, J. S. et al., "Globular cluster formation from inertial inflows: accreting extremely massive stars as the origin of abundance anomalies". *arXiv e-prints*, , arXiv:2501.12138, January (2025).

34. Cho, H., Woo, J.-H., Wang, S., Son, D. et al., "The Seoul National University AGN Monitoring Project. IV. Hα Reverberation Mapping of Six AGNs and the Hα Size-Luminosity Relation". *ApJ*, **953**(2), 142, August (2023).

35. Reines, A. E. and Volonteri, M., "Relations between Central Black Hole Mass and Total Galaxy Stellar Mass in the Local Universe". *ApJ*, **813**(2), 82, November (2015).

36. Condon, J. J., "Radio emission from normal galaxies.". *ARA&A*, **30**, 575–611, January (1992).

37. Watson, D., Zafar, T., Andersen, A. C., Fynbo, J. P. U. et al., "Helium in Natal H II Regions: The Origin of the X-Ray Absorption in Gamma-Ray Burst Afterglows". *ApJ*, **768**(1), 23, May (2013).

38. Malizia, A., Bassani, L., Bird, A. J., Landi, R. et al., "First high-energy observations of narrow-line Seyfert 1s with INTEGRAL/IBIS". *MNRAS*, **389**(3), 1360–1366, September (2008).

39. Pounds, K. A., Done, C. and Osborne, J. P., "RE 1034+39: a high-state Seyfert galaxy?". *MNRAS*, **277**(1), L5–L10, November (1995).

40. Katz, H., Cameron, A. J., Saxena, A., Barrufet, L. et al., "21 Balmer Jump Street: The Nebular Continuum at High Redshift and Implications for the Bright Galaxy Problem, UV Continuum Slopes, and Early Stellar Populations". *arXiv e-prints*, , arXiv:2408.03189, August (2024).

41. Killi, M., Watson, D., Brammer, G., McPartland, C. et al., "Deciphering the JWST spectrum of a 'little red dot' at z $\sim$ 4.53: An obscured AGN and its star-forming host". *A&A*, **691**, A52, November (2024).

42. Netzer, H., "Physical conditions in active nuclei-I. The Balmer decrement". *MNRAS*, **171**, 395–406, May (1975).

43. Kokorev, V., Fujimoto, S., Labbe, I., Greene, J. E. et al., "UNCOVER: A NIRSpec Identification of a Broad-line AGN at z = 8.50". *ApJ*, **957**(1), L7, November (2023).

44. Mathur, S., "Narrow-line Seyfert 1 galaxies and the evolution of galaxies and active galaxies". *MNRAS*, **314**(4), L17–L20, June (2000).

45. Fujimoto, S., Brammer, G. B., Watson, D., Magdis, G. E. et al., "A dusty compact object bridging galaxies and quasars at cosmic dawn". *Nature*, **604**(7905), 261–265, April (2022).

46. Labbé, I., van Dokkum, P., Nelson, E., Bezanson, R. et al., "A population of red candidate massive galaxies 600 Myr after the Big Bang". *Nature*, **616**(7956), 266–269, April (2023).

47. Weibel, A., de Graaff, A., Setton, D. J., Miller, T. B. et al., "RUBIES Reveals a Massive Quiescent Galaxy at z=7.3". *arXiv e-prints*, , arXiv:2409.03829, September (2024).

Acknowledgements We acknowledge support from the Danish National Research Foundation under grant DNRF140. VR, TH, NA, DA, CC, and CG are funded by the ERC Advanced Investigator Grant EPOCHS (788113). DW, RDN, SAS, and AS are co-funded by the European Union (ERC, HEAVYMETAL, 101071865). RDN is co-funded by Villum Foundation. KEH acknowledges funding from the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00072. SAS acknowledges funding from the UK Science and Technology Facilities Council (STFC, grant number ST/X00094X/1). Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The data products presented herein were retrieved from the Dawn *JWST* Archive (DJA). DJA is an initiative of the Cosmic Dawn Center (DAWN), which is funded by the Danish National Research Foundation under grant DNRF140.

Author contributions VR and DW wrote the manuscript and produced the figures. VR, DW, and GPN analysed the spectroscopic data and tested the models. GB and KEH reduced the spectroscopic and photometric data. RG and TH analysed the photometric and low-resolution spectral data and inferred the stellar masses. DW, SAS, RDN, and AS produced the electron-scattering and P Cygni numerical models. APV, VR, and DA produced stellar population and photoionization models and made comparisons with the data. NA, DA, CJC, CMG and ST helped to improve the manuscript and provided useful comments on the hypothesis and interpretation of the results. All authors reviewed and edited the manuscript.

Competing interests The authors declare no competing interests.

Corresponding Author Vadim Rusakov

Affiliations
[1] *Jodrell Bank Centre for Astrophysics, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[2] *Cosmic Dawn Center (DAWN), Denmark*
[3] *Niels Bohr Institute, University of Copenhagen, Jagtvej 155A, DK-2200, Copenhagen N, Denmark*
[4] *Department of Astronomy, University of Geneva, Chemin Pegasi 51, 1290 Versoix, Switzerland*
[5] *Astrophysics Research Centre, The Queen's University of Belfast, University Road, Belfast, BT71NN, United Kingdom*
[6] *Astronomy Centre, University of Sussex, Falmer, Brighton, BN1 9QH, UK*

Methods

In this paper, uncertainties are given as 1σ or 68% confidence intervals. Upper limits are indicated at the 2σ level unless otherwise stated. We adopt cosmological parameters measured by *Planck*[48], i.e. a Λ cold dark matter (ΛCDM) model with total matter density in units of the critical density, $\Omega_{\rm m} = 0.310$, and Hubble constant, $H_0 = 67.7\,{\rm km\,s^{-1}\,Mpc^{-1}}$.

Spectroscopic Sample This study makes use of the public *JWST* data collected as part of several observational programs with the NIRSpec spectrograph[17] with PIDs: 1345 (CEERS)[50], 1181 (JADES)[51], 1210 (JADES)[52,53], 2674 (PI Arrabal-Haro)[54], 3215 (JADES Origins Field)[55], 4106 (PI Nelson)[56], 4233 (RUBIES)[57], 2565 (PI Glazebrook)[58], 2750 (PI Arrabal-Haro)[59], and 6541 (PI Egami)[60]. These observations have been uniformly reduced and published as part of the Dawn *JWST* Archive[1] (DJA)[18]. Using the DJA, we selected all galaxies observed in the medium resolution grating spectra with a broad Hα component and a spectroscopic redshift produced by `msaexp`[2] [62]. To this, we added *JWST* broad-line objects reported in the literature with publicly available data and processed in DJA. We selected objects with a full width at half-maximum (FWHM) linewidth greater than $\sim$1000 km s^{-1} from the objects in the archive. We then selected spectra with high SNR (median SNR > 5 per 10 Å for the continuum-subtracted region $\pm$2000 km s^{-1} around the Hα line) and also included objects for the stacked spectrum using broad-line objects with lower SNR ($5 >$ SNR/10Å > 1) to ensure that we are not biased by our SNR selection.

Emission Line Models All best-fit results in this paper were produced using the Monte Carlo Markov Chain (MCMC) NUTS sampler in `PyMC` v. 5.17.0 [63], except Objects A and D, which were fitted using the Ensemble sampler in `emcee` v. 3 [64] (due to an incompatibility of the P Cygni model used here with the tensor formalism in `PyMC`). We sampled the posterior distributions with $4k$ walkers (where k is the number of free parameters) and 10^5 samples per walker. We use the mode values of the posterior parameter distributions as the best values and the 68%-highest density interval as the range of uncertainty. Finally, we find that the resolution of NIRSpec gratings is higher than the nominal value. Using the resolved widths of narrow [O III] lines in the high-resolution G235H grating of Object A, we estimate that the medium resolution grating G235M has ~ 1.7 times higher R than the nominal value. This scaling factor on R has been assumed for G395M and also for the G395H grating, as supported by recent instrument modelling[65] for a point source.

In modelling the broad Hα profile, we assumed a broadening mechanism: either a Doppler velocity broadening or a Compton scattering broadening. The former is modelled using a Gaussian function $f(\lambda; A, \mu, \sigma)$ with amplitude A, line centre μ, and velocity dispersion σ. For the Compton-scattered profile we use a symmetric exponential $g(\lambda; B, W) = Be^{-|\lambda-\lambda_0|/W}$ with amplitude B, line centre λ_0, and e-folding scale W.

First, we tested both broadening mechanisms by fitting two sets of models. Figure 4 shows

[1] https://dawn-cph.github.io/dja

[2] `https://github.com/gBrammer/msaexp`

the comparison between the broadening models. To fit the data reasonably well, the models also included (narrow) Gaussians for the host galaxy Hα and [N II] doublet with fixed centroids and velocity dispersions tied to the same value and limited to < 1000 km s^{-1}. The ratio of the two [N II] lines is set to 0.33 [66]. In some cases, an additional Gaussian absorption component (E) or a P-Cygni profile (A, D) are required to accurately model the broad Hα component. We note here that the [N II] lines are only required in the fits in objects B and G, the most complex systems. Whether this is an artefact related to the complex spectral shape due to possible self-reversal in the line, or whether [N II] really is observed in these cases, is unclear and would require higher resolution spectra and more sophisticated models to establish. Finally, wavelength regions around the emission lines [O I] λ6302, He I λ6678, and [S II] $\lambda\lambda$6717,6731, are excluded from the fit.

Although most Hα lines in the sample are predominantly exponential with very high statistical significance (see Extended Fig. 4), to reconstruct the intrinsic Doppler widths, we model the lines with a Gaussian convolved with an exponential, instead of a pure exponential. These models are convolved with the instrumental resolution of the relevant gratings (which were taken from the *JWST* JDox website for NIRSpec) at the Hα peak (we assume the actual resolution is ~ 1.7 times better than the nominal – see Methods above). In order to alleviate the complexity of some of our models and more accurately constrain the narrow Hα components, we use the velocity widths of the optical [O III] lines as a Gaussian prior on Hα and [N II] widths, where relevant spectral coverage is available. The best-fit profiles and intrinsic Doppler components with their widths are presented in Figure 3 and Table 4, while the posterior distributions of the sampled parameters are shown as corner plots in Extended Data Figures 9–21.

We also test whether for example gas turbulence or Raman scattering could be responsible for line broadening by comparing a basic Lorentzian [9,24,68] and exponential line shapes. The former is defined as a symmetric profile with FWHM 2γ centred at λ_0: $h(\lambda; C, \gamma) = C\frac{\gamma}{(\lambda-\lambda_0)^2-\gamma^2}$. The exponential is a significantly better fit in most objects or an equivalent fit in Objects H, L (see Extended Data Fig. 4). As this indicates that any potential contribution from turbulence broadening is not significant, we do not assume a more physically-motivated profile of a Gaussian convolved with Lorentzian (i.e. a Voigt profile). Another reason to exclude this model is that the ratios of the line widths between Hα, Hβ, and Paβ for example, are expected to differ by about a factor of 2–3 in velocity[24], something which is not generally observed in these types of objects. Furthermore, turbulent broadening may sometimes result in enhanced red-wing profiles[68], while all objects here have symmetric Hα wings (see Extended Data Figure 5). However, while it has been argued that the lines cannot be Lorentzian on this basis[27], optical depth effects, which would be quite different for the different lines, could affect the relative line widths and more careful non-LTE radiative transfer analysis would help elucidate this issue. Finally, the comparison between the fit statistics (χ^2 and BIC) for all tested models (Gaussian, Exponential, Lorentzian and Fiducial) are shown in Extended Data Figs. 6, and 7 and Extended Data Table 3.

We also note that some studies reported deviation of the broad Hα lines from single-component Gaussian profiles in similar *JWST* objects[4,72] and explained them with the double-component Gaussian models assuming binary AGN. Although such models may provide a close fit to the

exponential line wings and the Gaussian line cores, they are more complex with more free parameters, and it seems unlikely that eleven in twelve of our objects as well as the stack all represent cases of dual AGN, which are predicted to be only a few percent of cases at intermediate redshifts in simulations[73]. Finally, our objects have broad line profiles that are symmetric (see Extended Data Fig. 5) and they do not exhibit apparent disturbances or irregularities in their spatial light profiles (see Fig. 2) that could be expected from merging galaxies. Therefore, we do not consider such models further.

Optical depth measurement To estimate the approximate optical depth of the scatter from the exponential linewidth measurement (W), we use Monte Carlo simulations of electron scattering at 10,000 K and various optical depths in a spherical shell geometry. Specifically, we simulate Compton scattering in the low energy-limit ($h\nu \ll m_e c^2$) in a thin, uniform density/temperature spherical shell ($\Delta r/r = 0.1$) that is characterised by its radial optical depth to electron scattering. We assume photons enter the scattering region at normal incidence and follow their scattering until they emerge from the simulation domain. We simulate only the Compton scattering process in the domain (no true absorption or line scattering is included). The relation between scattering optical depth and the width of the scattered exponential scales almost linearly (Extended Data Fig. 3). We model the relationship as

$$W = a\tau + b,$$

where a = 428 km s^{-1}, b = 370 km s^{-1}.

The relationship scales roughly as the square root of the temperature[22], so that at temperatures of 20,000 K the inferred optical depths would be 30% lower, for example.

Spectral stacking Individual noisy spectra are combined to obtain a median spectral stack with a greater SNR in the 4000 km s^{-1} region centered on the Hα line. Initially, we fit each Hα line profile with a basic exponential broad Hα component with its amplitude B and e-folding width W. Then we subtract the best-fit continuum around Hα and normalize the individual widths of the line profiles to the median W of the spectral stack. Similarly, individual exponential amplitudes B are normalized to the median amplitude. Next, we resample the spectra to the same restframe wavelength grid oversampled by a factor of ten with respect to the highest resolution Hα spectra in the stack to avoid aliasing effects. Finally, the stacked spectrum is produced by taking the median of the stack and estimating uncertainties by drawing 10000 times from Gaussian uncertainties of individual spectra. In the end, this stacked spectrum is resampled to the wavelength grid with the resolution equal to the median of the spectral stack resolutions.

Spectral energy distribution modelling To estimate stellar masses of host galaxies of our objects (used in Table 1 and Fig. 4), we model their spectral energy distributions (SED) from 0.9 to 13 μm using *JWST* and *HST* photometry from JADES[51], CEERS[50], PRIMER[75] and GO-3577 (PI Egami)[76] surveys, reduced with `Grizli` [25] and available on DJA[18].

We model the rest UV-to-optical SED with the code Bayesian Analysis of Galaxies for Physical Inference and Parameter EStimation (BAGPIPES)[78]. In this work, we model the sources in

our sample with a double power-law star-formation history (SFH) which is characterised by distinct falling and rising slopes (see Equation 10 in ref [78]). We impose the following priors on the SFH: $\tau \in (0, 15)$ Gyr with a uniform prior, where τ is a timescale related to the turnover between the falling and rising components of the SFH, $\alpha \in (0.01, 1000)$ and $\beta \in (0.01, 1000)$ with logarithmic priors, where α and β are the falling and rising exponents of the SFH respectively, $\log(M_{\rm formed}/M_\odot) \in (6, 13)$ with a uniform prior, where $M_{\rm formed}$ is the total mass formed, and $Z_\star/Z_\odot \in (0.01, 1.2)$, where $Z_\star$ is the stellar metallicity. Additionally, we choose a 'Calzetti' dust attenuation law[79] with $A_{\rm V} \in (0, 6)$ mag with a uniform prior, a nebular ionisation parameter of $\log U \in (-4, -0.01)$ with a uniform prior and an intrinsic line velocity dispersion of $v \in (50, 500)$ km s^{-1} with a uniform prior. We fix the redshift of each object to the spectroscopic redshift derived using `msaexp` [62].

The effect of photoelectric absorption on the X-ray luminosity We model the effect of photoelectric absorption by a 10% solar metallicity, ionized gas column at $z = 5$ on the observed photon flux using the 'absori' model in `Xspec`[80] (version: 12.14.0h) with a column density of $N_{\rm H} = 5 \times 10^{24}$ cm^{-2}. This absorber reduces the photon flux by a factor of about two (three) for a Milky Way-absorbed power-law model with photon index $\Gamma = 2.0$ ($\Gamma = 3.0$) and a local equivalent hydrogen column density of $N_{\rm H} = 5 \times 10^{20}$ cm^{-2}. The exact ionisation state of the gas for an absorber in the 10,000–30,000 K range is not very important in this respect—the absorption increases by about 20% for a fully neutral gas—since these temperatures are not hot enough to liberate the L- and K-shell electrons that provide most of the X-ray opacity. Photoelectric absorption from this gas therefore seems insufficient to provide more than half an order of magnitude flux deficit at X-ray wavelengths[6,7].

Table 2: The sample of objects used in this work. Objects ABFGIJ are reported in this work, as extracted from DJA. References to other sources where the objects have been previously reported are as follows: DEHL[15], C[4], K[3]. The stack contains spectra from all of these studies. The typical uncertainties in the redshifts are about 3×10^{-4}. Right ascensions and declinations are relative to J2000.0.

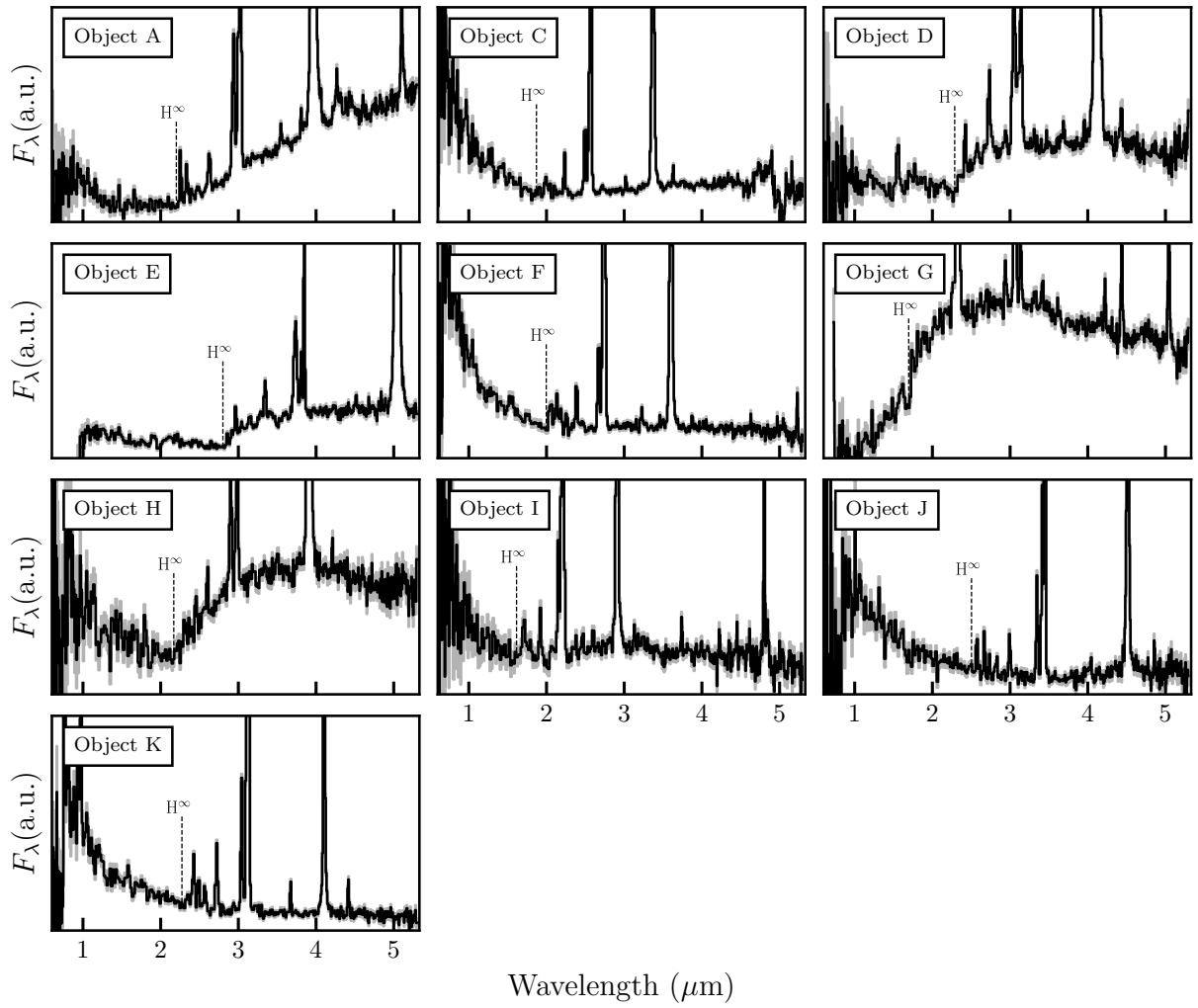
ID	Survey	Field PID-SRCID	α ($^{\rm h}:^{\rm m}:^{\rm s}$)	δ ($^\circ:':''$)	Redshift
A	JADES	GN 1181-68797	12:36:54.993	62:08:46.283	5.0405
B	—	GN 2674-14	12:36:47.955	62:09:41.307	5.1826
C	JADES	GN 1181-73488	12:36:47.375	62:10:38.039	4.1327
D	RUBIES	EGS 4233-42046	14:19:10.888	52:47:19.848	5.2757
E	RUBIES	EGS 4233-49140	14:19:34.139	52:52:38.675	6.6847
F	CEERS	EGS 1345-1244	14:20:57.757	53:02:09.748	4.4771
G	RUBIES	EGS 4233-58237	14:19:24.137	52:51:57.707	3.6505
H	—	EGS 4106-51623	14:19:32.836	52:51:19.401	4.9511
I	JADES	GN 1181-53501	12:37:10.814	62:11:36.860	3.4294
J	JADES	GN 1181-38147	12:37:04.962	62:08:54.307	5.8694
K	RUBIES	EGS 4233-50052	14:19:17.629	52:49:48.996	5.2392
L	RUBIES	EGS 4233-60935	14:19:41.610	52:55:32.134	5.2877
S	Stack	—	—	—	

Code availability. The spectra downloaded from the DJA were processed from the original telescope data with `msaexp`[62]. Our Monte Carlo scattering code is available upon request. Finally, this study has made use of the following publicly available packages: `arviz`[87], `astropy`[88,89], `Bagpipes`[78], `emcee`[64], `Grizli`[25], `matplotlib`[90], `numpy`[91], `pandas`[92],`PyMC`[63], `scipy`[93], `Xspec`[80] and a P Cygni module adapted from Ulrich Noebauer's code `https://github.com/unoebauer/public-astro-tools` [94,95] and presented in [86,96].

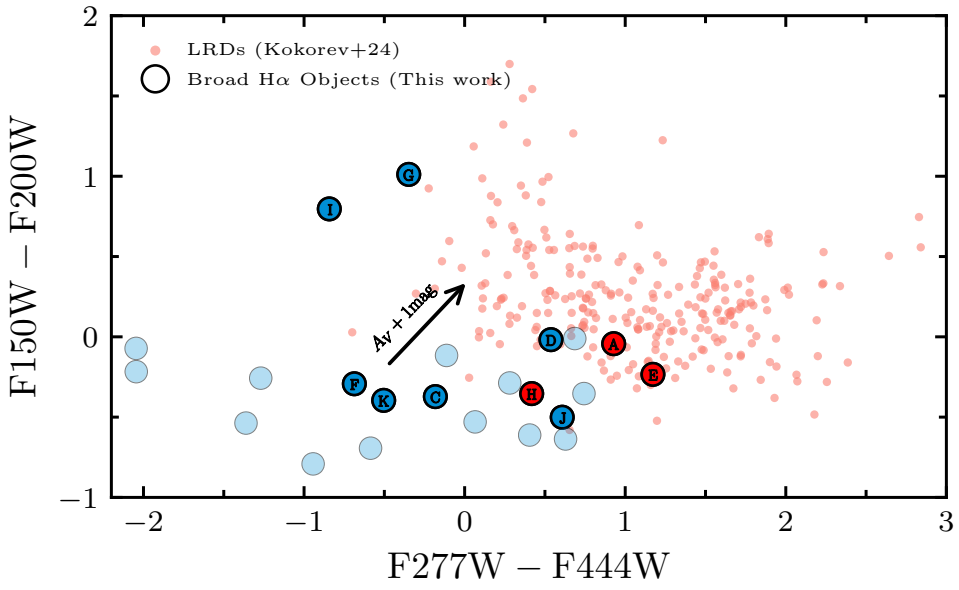
Data availability. This paper makes use of the *JWST* data downloaded from the DAWN JWST Archive (DJA) (`https://dawn-cph.github.io/dja/`). We are grateful for the availability of the raw data from the respective observational programs listed in the Methods Sections "Spectroscopic Sample" and "Spectral energy distribution modelling". Individual data products can be identified in their respective programs or on DJA using PID-SRCID number combinations in Table 2.

References for Methods

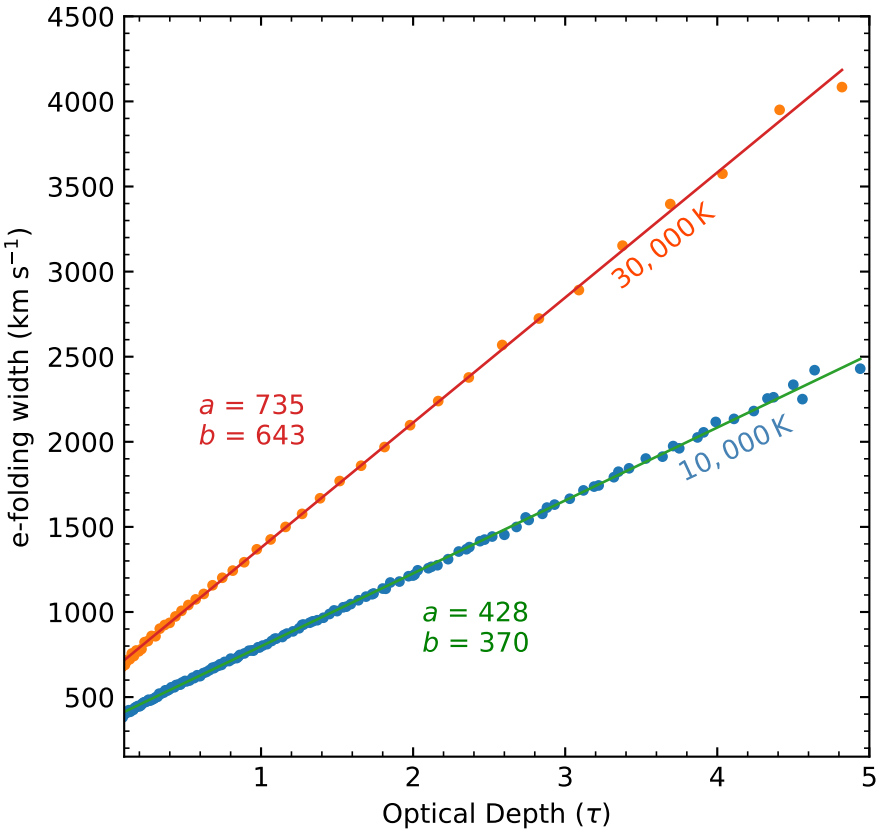
48. Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M. et al., "Planck 2018 results. VI. Cosmological parameters". *A&A*, **641**, A6, September (2020).



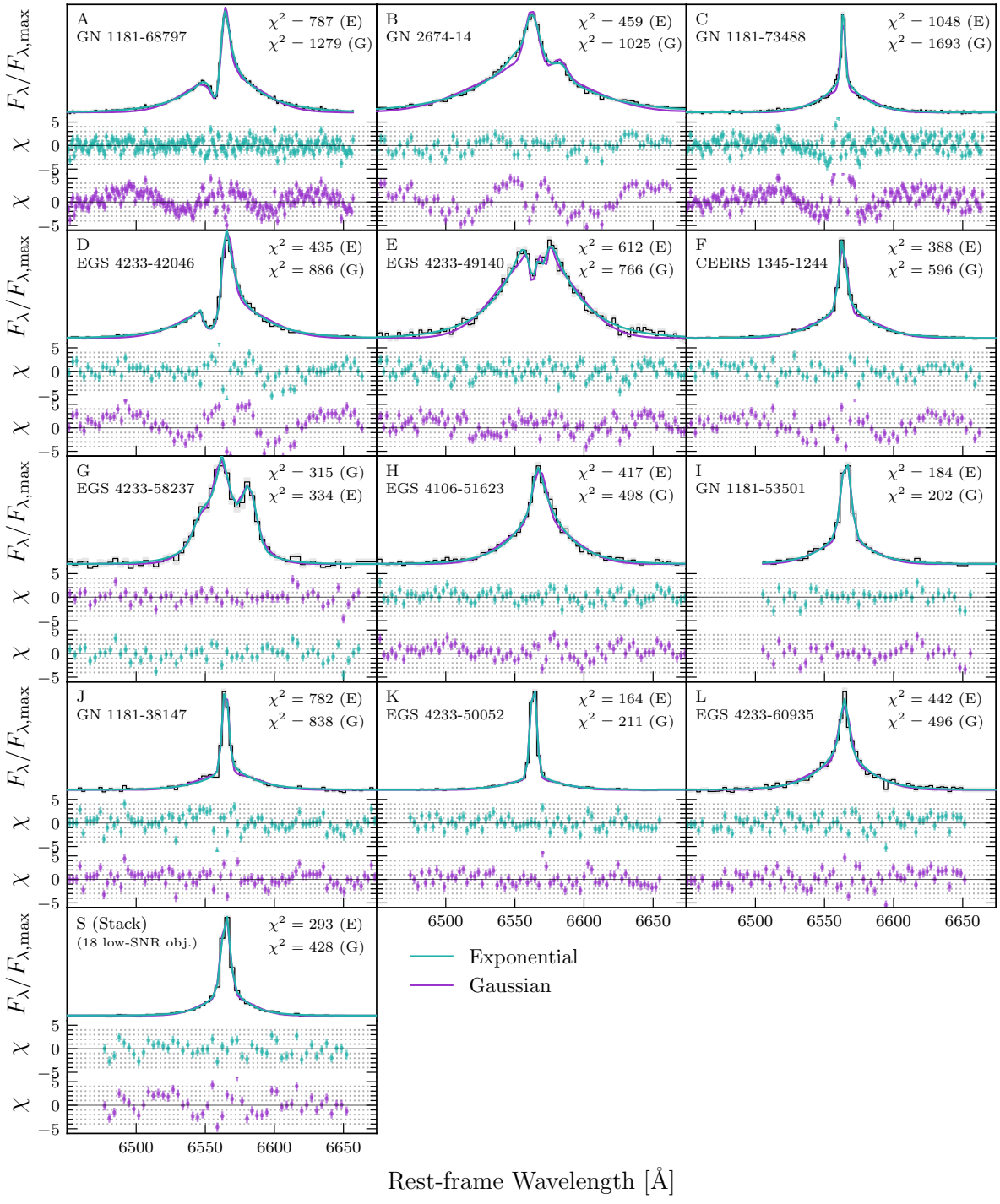
Extended Data Figure 1: A gallery of NIRSpec/PRISM spectra for ten out of twelve high-SNR objects (objects B and L have not been observed with PRISM), with the flux density scaled arbitrarily to emphasize the continuum shapes. The position of the Balmer series limit at ~ 365 nm is labelled on the panels as H^{∞}. Interestingly, objects D, E, G, and H show clear Balmer break features, while objects A, C, F, and I have a turnover in the continuum slope coincident with the location of H^{∞}. These SEDs are also similar to those typical of LRDs.



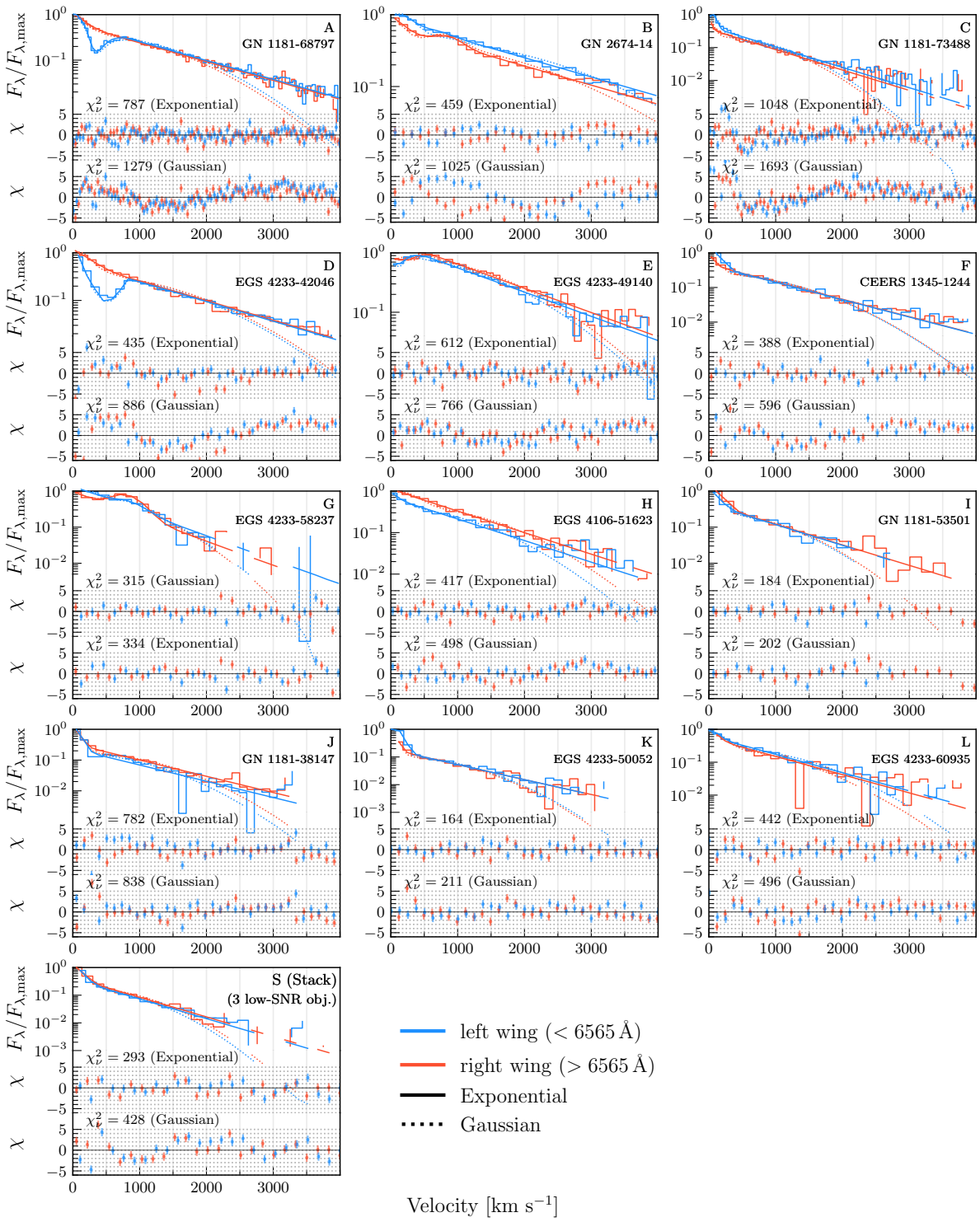
Extended Data Figure 2: Colour space (rest optical vs. rest UV) of the broad Hα sample in this work (large circles) plotted next to the photometrically-identified population of LRDs from [85]. The objects from our sample that satisfy their colour classification criteria of LRDs are shown in red – the rest are in blue (the objects used in the stack are shown as faint blue circles). The high-SNR object B is missing a PRISM spectrum, while object L is missing due to lack of coverage in the rest UV, but has $\mathrm{F277W} - \mathrm{F444W} = 0.24$. Objects G and I are the lowest redshift objects in our sample, which may explain their redness in $\mathrm{F150W} - \mathrm{F200W}$. Some of our selection is somewhat bluer in the rest UV and optical than the population of LRDs. The difference may be explained by host galaxy contributions and dust extinction. The black arrow shows the mean effect of an extinction of $A_V = +1$ mag on the PRISM spectra of our sample.



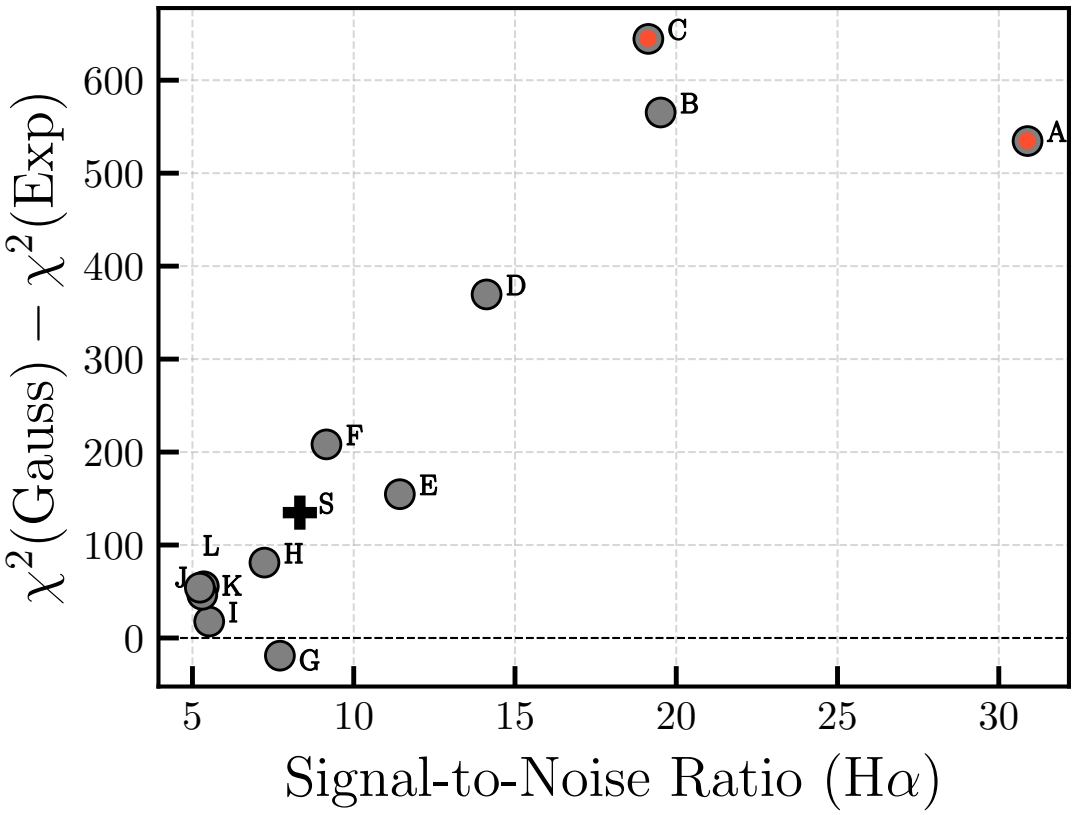
Extended Data Figure 3: Relation between the optical depth of the scatterer and the resulting line broadening. We plot here the results of Monte Carlo slab scattering simulations at electron temperatures of 10,000 K and 30,000 K.



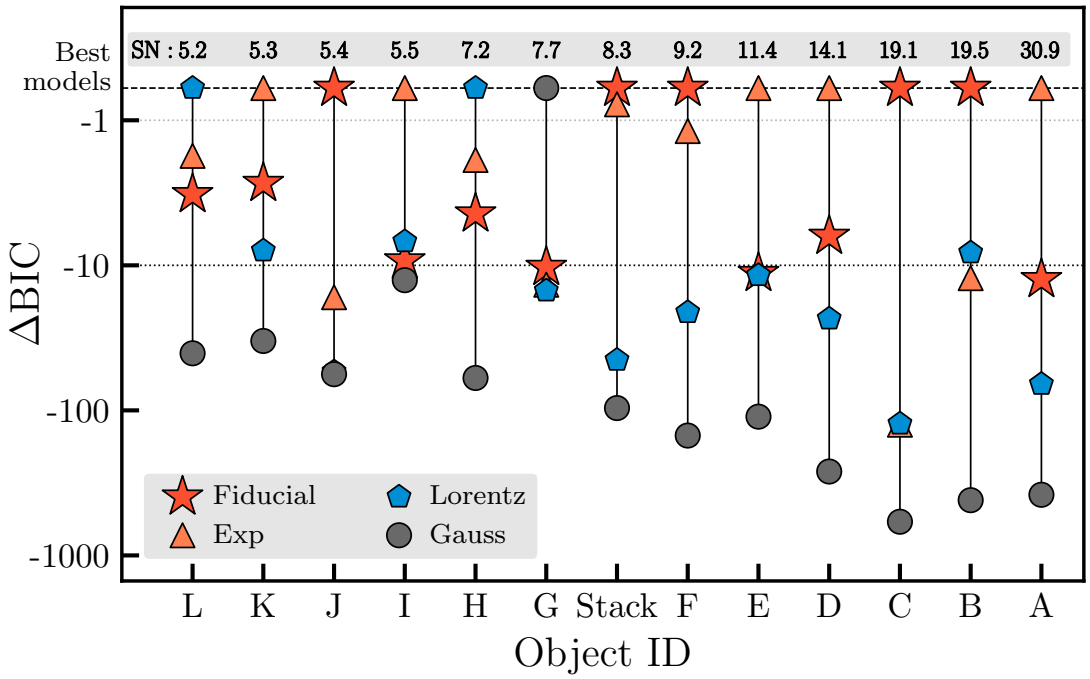
Extended Data Figure 4: Comparison between the best-fitting Hα profiles with either a Gaussian or an Exponential broad-line component. The numbers of degrees of freedom for both model fits are the same. The objects are ordered by decreasing median SNR(Hα) and their peaks are normalized in flux density for ease of comparison. In all cases, except G, and including the stack, the exponential shape is strongly preferred over the typically assumed Gaussian profile. This preference is driven by the narrow Hα cores and widely extended tails.



Extended Data Figure 5: Similar to Figure 4, but with the wings of the Hα profiles mirrored around the central wavelength, demonstrating that the linearity of the data in log-space is best described by the exponential tails (solid lines) and under-fitted by the Gaussian tails (dotted lines).



Extended Data Figure 6: Comparison between the Hα line models that include basic exponential and Gaussian components of the broad Hα. This shows an improvement in the goodness-of-fit of the basic exponential line shape over the Gaussian in terms of χ^2 differences with increasing SNR. The plus sign indicates the stacked spectrum, and two red circles represent the high-resolution spectra.

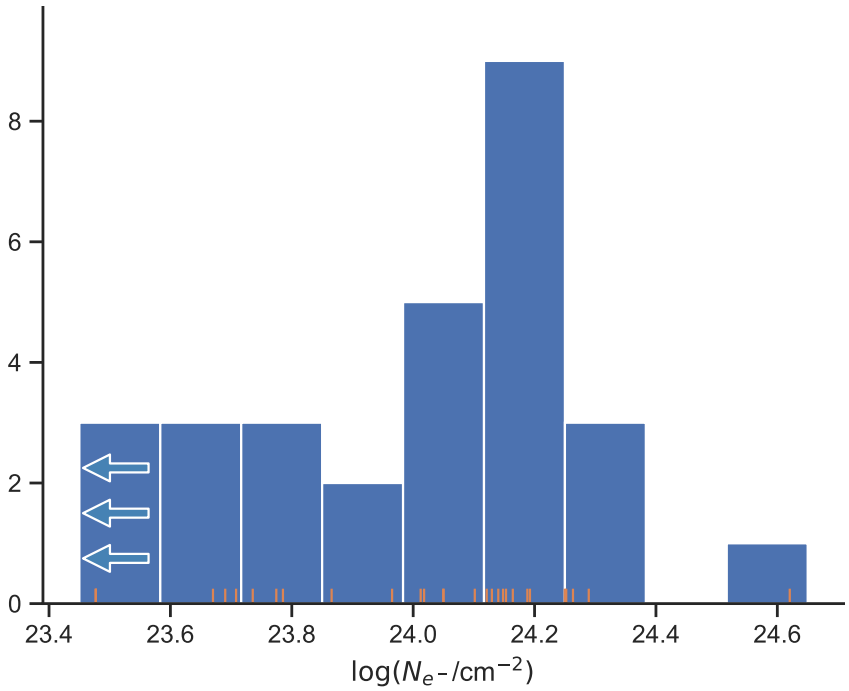


Extended Data Figure 7: Differences between the Bayesian Information Criterion (BIC) values of all broad Hα models with respect to the best one (horizontal dashed line) where the models include Fiducial, and basic Lorentzian, Gaussian, and Exponential. The objects are sorted from left to right in the order of increasing median SN(Hα) (the values are shown at the top). All models have equal degrees of freedom per object, except Fiducial, which has one or two additional parameters. This difference translates to $\Delta\mathrm{BIC} \sim 10$ for the typical number of data points–therefore, for all objects the Fiducial model is either the best-fit model or statistically equivalent to the best one. Where the pure Exponential model is more favoured by ΔBIC, this indicates that the additional components of the fiducial model are not statistically required in those cases. Exceptions are object G, where the Gaussian may be marginally preferred. In objects H and L, the Lorentzian is very marginally preferred, but this is not statistically significant compared to the exponential models. In case of objects A and D, the global minimum is challenging to recover due to the P Cygni feature. We notice that the uncertainties in the spectra may be underestimated and therefore scale all chi-squared values used for calculating BIC in this figure to the single lowest reduced chi-square of $\chi^2_\nu = 1.4$ to correct for that, assuming $\chi^2_\nu = 1.4$ corresponds to an ideal best solution.

Table 3: Goodness of fit of several models using the Bayesian Information Criterion (BIC) and χ^2 statistics. Each model uses n data points and k free parameters.

ID	Hα SNR (/10 Å)	n	k		BIC				χ^2			
			Other	Fiduc.	Gauss	Lorentz	Expo.	Fiduc.	Gauss	Lorentz	Expo.	Fiduc.
A*	30.9	491	12	14	1358.5	916.2	824.0	836.5	1284.2	841.8	749.6	749.8
B	19.5	276	9	11	1075.3	504.3	510.0	488.4	1024.7	453.7	459.5	426.6
C*	19.1	483	8	10	1742.1	1095.1	1097.7	917.1	1692.7	1045.7	1048.3	855.3
D	14.1	198	12	13‡	787.6	450.7	418.0	424.7	724.1	387.3	354.6	356.0
E	11.4	313	11	13	829.6	691.3	674.9	686.2	766.4	628.1	611.7	611.5
F	9.2	224	8	10	639.2	459.9	430.9	426.1	595.9	416.6	387.6	371.9
G	7.7	155	9	11	360.0	380.9	379.0	370.3	314.6	335.5	333.6	314.9
H	7.2	280	8	10	543.3	459.6	462.2	461.3	498.2	414.5	417.2	404.9
I	5.5	103	8	10	238.7	230.5	221.0	230.2	201.6	193.5	183.9	183.9
J	5.4	403	8	10	885.7	882.6	830.1	802.0	837.7	834.6	782.1	742.0
K	5.3	120	8	10	248.9	213.5	202.4	202.4	210.6	175.2	164.1	154.5
L	5.2	222	8	10	539.4	482.8	485.3	483.0	496.2	439.6	442.1	429.0
S†	8.3	154	7	9	463.5	391.4	328.6	324.5	428.2	356.2	293.4	279.1

*Analysis of objects A and C is based on spectra taken with the high-resolution G395H grating, while all others use the medium-resolution G395M. SNR (Hα) is the median SNR within 2000 km s^{-1} of the Hα $\lambda 6564.6$ wavelength. †Stack. ‡The unscattered component is omitted in the fit for object D as it is unconstrained.



Extended Data Figure 8: The distribution of free electron column densities inferred from the exponential fits to the Hα line for our full sample. The first column of the histogram represents upper limits. Our sample selection of FWHM$\gtrsim 1000\,\mathrm{km\,s^{-1}}$ limits the lowest column densities we can detect. Objects with column densities above $\log(N_{e^-}/\mathrm{cm}^2) \sim 24.5$ are less likely to be discovered as they correspond to optical depths $\tau_e > 2$. The intrinsic distribution of the column densities of all sources is therefore likely to be broader than shown here.

Table 4: Properties of the broad component of Hα based on the best-fit fiducial model. In the fiducial model the broad Hα component is represented as a combination of an intrinsic Gaussian convolved with an exponential kernel and a pure Gaussian (electron-scattered and unscattered components), where the best-fit widths of the intrinsic Gaussian and convolved exponential components are reported as $\mathrm{FWHM_{Doppler}}$ and $\mathrm{FWHM_{Exp}}$. The Gaussian line widths ($\mathrm{FWHM_{Doppler}}$) in this table are corrected for instrumental broadening ($\mathrm{FWHM_{instrum}}$). In the case of object L, the table shows two distinct posterior solutions for $\mathrm{FWHM_{Doppler}}$.

Broad Hα Component						
ID	EW [Å]	$F \times 10^{-19}$ [erg s^{-1} cm^{-2}]	$\log L$(Hα)	$\mathrm{FWHM_{Exp}}$ [km s^{-1}]	$\mathrm{FWHM_{Doppler}}$ [km s^{-1}]	$\mathrm{FWHM_{instrum}}$ [km s^{-1}]
A	1106^{+49}_{-60}	1814^{+74}_{-23}	$43.70^{+0.02}_{-0.01}$	1451^{+26}_{-28}	315^{+667}_{-73}	66
B	410^{+6}_{-6}	774^{+9}_{-10}	$43.36^{+0.01}_{-0.01}$	2551^{+109}_{-101}	2052^{+128}_{-148}	171
C	1124^{+42}_{-40}	273^{+3}_{-3}	$42.68^{+0.0}_{-0.0}$	1162^{+29}_{-26}	398^{+24}_{-25}	77
D	1607^{+53}_{-56}	1344^{+16}_{-22}	$43.62^{+0.01}_{-0.01}$	1460^{+22}_{-27}	417^{+233}_{-205}	168
E	1533^{+103}_{-141}	2094^{+245}_{-30}	$44.05^{+0.05}_{-0.01}$	1474^{+34}_{-33}	< 274	137
F	1238^{+51}_{-45}	583^{+12}_{-15}	$43.09^{+0.01}_{-0.01}$	1235^{+28}_{-27}	248^{+82}_{-115}	193
G	201^{+18}_{-21}	340^{+27}_{-34}	$42.64^{+0.03}_{-0.05}$	< 1003	1957^{+82}_{-78}	227
H	733^{+25}_{-24}	240^{+4}_{-5}	$42.80^{+0.01}_{-0.01}$	1266^{+64}_{-61}	643^{+107}_{-103}	177
I	1038^{+335}_{-210}	228^{+58}_{-10}	$42.41^{+0.10}_{-0.02}$	1061^{+85}_{-69}	< 274	238
J	1106^{+68}_{-60}	265^{+8}_{-8}	$43.02^{+0.01}_{-0.01}$	1215^{+40}_{-44}	197^{+42}_{-56}	154
K	996^{+186}_{-141}	186^{+10}_{-11}	$42.75^{+0.02}_{-0.03}$	1171^{+60}_{-66}	311^{+80}_{-70}	169
L	1257^{+123}_{-112}	361^{+13}_{-14}	$43.05^{+0.02}_{-0.02}$	1208^{+84}_{-117}	< 155 627^{+203}_{-122}	168
$\mathrm{S^{(Stack)}}$	—	—	—	871^{+28}_{-24}	339^{+55}_{-51}	246

49. Jakobsen, P., Ferruit, P., Alves de Oliveira, C., Arribas, S. et al., "The Near-Infrared Spectrograph (NIRSpec) on the James Webb Space Telescope. I. Overview of the instrument and its capabilities". *A&A*, **661**, A80, May (2022).

50. Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., Wilkins, S. M. et al., "CEERS Key Paper. I. An Early Look into the First 500 Myr of Galaxy Formation with JWST". *ApJ*, **946**(1), L13, March (2023).

51. Eisenstein, D. J., Willott, C., Alberts, S., Arribas, S. et al., "Overview of the JWST Advanced Deep Extragalactic Survey (JADES)". *arXiv e-prints*, , arXiv:2306.02465, June (2023).

52. Bunker, A. J., Cameron, A. J., Curtis-Lake, E., Jakobsen, P. et al., "JADES NIRSpec initial data release for the Hubble Ultra Deep Field: Redshifts and line fluxes of distant galaxies from the deepest JWST Cycle 1 NIRSpec multi-object spectroscopy". *A&A*, **690**, A288, October (2024).

53. D'Eugenio, F., Cameron, A. J., Scholtz, J., Carniani, S. et al., "JADES Data Release 3: NIRSpec/Microshutter Assembly Spectroscopy for 4000 Galaxies in the GOODS Fields". *ApJS*, **277**(1), 4, March (2025).

54. Arrabal Haro, P., Calvi, R., Daddi, E., Dannerbauer, H. et al. "Environmental effects on galaxy evolution in a z=5.2 proto-cluster". JWST Proposal. Cycle 1, ID. #2674, March (2021).

55. Eisenstein, D. J., Johnson, B. D., Robertson, B., Tacchella, S. et al., "The JADES Origins Field: A New JWST Deep Field in the JADES Second NIRCam Data Release". *arXiv e-prints*, , arXiv:2310.12340, October (2023).

56. Nelson, E., Labbe, I., Benton, C., Bezanson, R. et al. "Extremely massive galaxies in the early universe: a challenge to Lambda-CDM?". JWST Proposal. Cycle 2, ID. #4106, May (2023).

57. de Graaff, A., Brammer, G., Weibel, A., Lewis, Z. et al., "RUBIES: a complete census of the bright and red distant Universe with JWST/NIRSpec". *arXiv e-prints*, , arXiv:2409.05948, September (2024).

58. Glazebrook, K., Nanayakkara, T., Esdaile, J., Espejo, J. M. et al. "How Many Quiescent Galaxies are There at 3". JWST Proposal. Cycle 1, ID. #2565, March (2021).

59. Arrabal Haro, P., Dickinson, M., Bagley, M., Burgarella, D. et al. "Spectroscopic follow-up of ultra-high-z candidates in CEERS: Characterizing true z ¿ 12 galaxies and z 4-7 interlopers in preparation for JWST Cycle 2". JWST Proposal. Cycle 1, ID. #2750, October (2022).

60. Egami, E., Bonaventura, N., Charlot, S., Curti, M. et al. "JWST NIRSpec/NIRCam Follow-Up of the High-Redshift Transients Discovered in the GOODS-S JADES-Deep Field". JWST Proposal. Cycle 2, ID. #6541, November (2023).

61. Heintz, K. E., Brammer, G. B., Watson, D., Oesch, P. A. et al., "The JWST-PRIMAL archival survey: A JWST/NIRSpec reference sample for the physical properties and Lyman-α absorption and emission of $\sim$600 galaxies at z = 5.0 - 13.4". *A&A*, **693**, A60, January (2025).

62. Brammer, G. "`msaexp`: Nirspec analyis tools". `https://doi.org/10.5281/zenodo.8319596`, September (2023). version 0.6.17.

63. Salvatier, J., Wiecki, T. and Fonnesbeck, C. "Probabilistic programming in python using pymc", (2015).

64. Foreman-Mackey, D., Hogg, D. W., Lang, D. and Goodman, J., "emcee: The MCMC Hammer". *PASP*, **125**(925), 306, March (2013).

65. de Graaff, A., Rix, H.-W., Carniani, S., Suess, K. A. et al., "Ionised gas kinematics and dynamical masses of z $\gtrsim$ 6 galaxies from JADES/NIRSpec high-resolution spectroscopy". *A&A*, **684**, A87, April (2024).

66. Dojčinović, I., Kovačević-Dojčinović, J. and Popović, L. Č., "The flux ratio of the [N II] $\lambda\lambda$ 6548, 6583 Å lines in sample of Active Galactic Nuclei Type 2". *Advances in Space Research*, **71**(2), 1219–1226, January (2023).

67. Kollatschny, W. and Zetzl, M., "The shape of broad-line profiles in active galactic nuclei". *A&A*, **549**, A100, January (2013).

68. Goad, M. R., Korista, K. T. and Ruff, A. J., "The broad emission-line region: the confluence of the outer accretion disc with the inner edge of the dusty torus". *MNRAS*, **426**(4), 3086–3111, November (2012).

69. Kokubo, M. and Harikane, Y., "Challenging the AGN scenario for JWST/NIRSpec broad Hα emitters/Little Red Dots in light of non-detection of NIRCam photometric variability and X-ray". *arXiv e-prints*, , arXiv:2407.04777, July (2024).

70. Juodžbalis, I., Ji, X., Maiolino, R., D'Eugenio, F. et al., "JADES - the Rosetta stone of JWST-discovered AGN: deciphering the intriguing nature of early AGN". *MNRAS*, **535**(1), 853–873, November (2024).

71. Maiolino, R., Scholtz, J., Curtis-Lake, E., Carniani, S. et al., "JADES: The diverse population of infant black holes at $4 < z < 11$: Merging, tiny, poor, but mighty". *A&A*, **691**, A145, November (2024).

72. D'Eugenio, F., Maiolino, R., Perna, M., Uebler, H. et al., "BlackTHUNDER strikes twice: rest-frame Balmer-line absorption and high Eddington accretion rate in a Little Red Dot at $z = 7.04$". *arXiv e-prints*, , arXiv:2503.11752, March (2025).

73. Puerto-Sánchez, C., Habouzit, M., Volonteri, M., Ni, Y. et al., "Large-scale dual AGN in large-scale cosmological hydrodynamical simulations". *MNRAS*, **536**(3), 3016–3040, January (2025).

74. Laor, A., "Evidence for Line Broadening by Electron Scattering in the Broad-Line Region of NGC 4395". *ApJ*, **643**(1), 112–119, May (2006).

75. Dunlop, J. S., Abraham, R. G., Ashby, M. L. N., Bagley, M. et al. "PRIMER: Public Release IMaging for Extragalactic Research". JWST Proposal. Cycle 1, ID. #1837, March (2021).

76. Egami, E., Sun, F., Alberts, S., Baum, S. A. et al. "Complete NIRCam Grism Redshift Survey (CONGRESS)". JWST Proposal. Cycle 2, ID. #3577, August (2023).

77. Brammer, G. "grizli", September (2023).

78. Carnall, A. C., McLure, R. J., Dunlop, J. S. and Davé, R., "Inferring the star formation histories of massive quiescent galaxies with BAGPIPES: evidence for multiple quenching mechanisms". *MNRAS*, **480**(4), 4379–4401, November (2018).

79. Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L. et al., "The Dust Content and Opacity of Actively Star-forming Galaxies". *ApJ*, **533**(2), 682–695, April (2000).

80. Arnaud, K. A. "XSPEC: The First Ten Years". In *Astronomical Data Analysis Software and Systems V,* Jacoby, G. H. and Barnes, J., editors, volume 101 of *Astronomical Society of the Pacific Conference Series*, 17, January (1996).

81. Maiolino, R., Risaliti, G., Signorini, M., Trefoloni, B. et al., "JWST meets Chandra: a large population of Compton thick, feedback-free, and intrinsically X-ray weak AGN, with a sprinkle of SNe". *MNRAS*, , February (2025).

82. Yue, M., Eilers, A.-C., Ananna, T. T., Panagiotou, C. et al., "Stacking X-Ray Observations of "Little Red Dots": Implications for Their Active Galactic Nucleus Properties". *ApJ*, **974**(2), L26, October (2024).

83. Kocevski, D. D., Finkelstein, S. L., Barro, G., Taylor, A. J. et al., "The Rise of Faint, Red AGN at $z > 4$: A Sample of Little Red Dots in the JWST Extragalactic Legacy Fields". *arXiv e-prints*, , arXiv:2404.03576, April (2024).

84. Harikane, Y., Zhang, Y., Nakajima, K., Ouchi, M. et al., "A JWST/NIRSpec First Census of Broad-line AGNs at z = 4-7: Detection of 10 Faint AGNs with $M_{BH} \sim 10^6 - 10^8\,M_\odot$ and Their Host Galaxy Properties". *ApJ*, **959**(1), 39, December (2023).

85. Kokorev, V., Caputi, K. I., Greene, J. E., Dayal, P. et al., "A Census of Photometrically Selected Little Red Dots at $4 < z < 9$ in JWST Blank Fields". *ApJ*, **968**(1), 38, June (2024).

86. Sneppen, A., Watson, D., Bauswein, A., Just, O. et al., "Spherical symmetry in the kilonova AT2017gfo/GW170817". *Nature*, **614**(7948), 436–439, February (2023).

87. Kumar, R., Carroll, C., Hartikainen, A. and Martin, O., "Arviz a unified library for exploratory analysis of bayesian models in python". *Journal of Open Source Software*, **4**(33), 1143, (2019).

88. Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M. et al., "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package". *AJ*, **156**(3), 123, September (2018).

89. Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P. et al., "Astropy: A community Python package for astronomy". *A&A*, **558**, A33, October (2013).

90. Hunter, J. D., "Matplotlib: A 2d graphics environment". *Computing in Science & Engineering*, **9**(3), 90–95, (2007).

91. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R. et al., "Array programming with numpy". *Nature*, **585**(7825), 357–362, (2020).

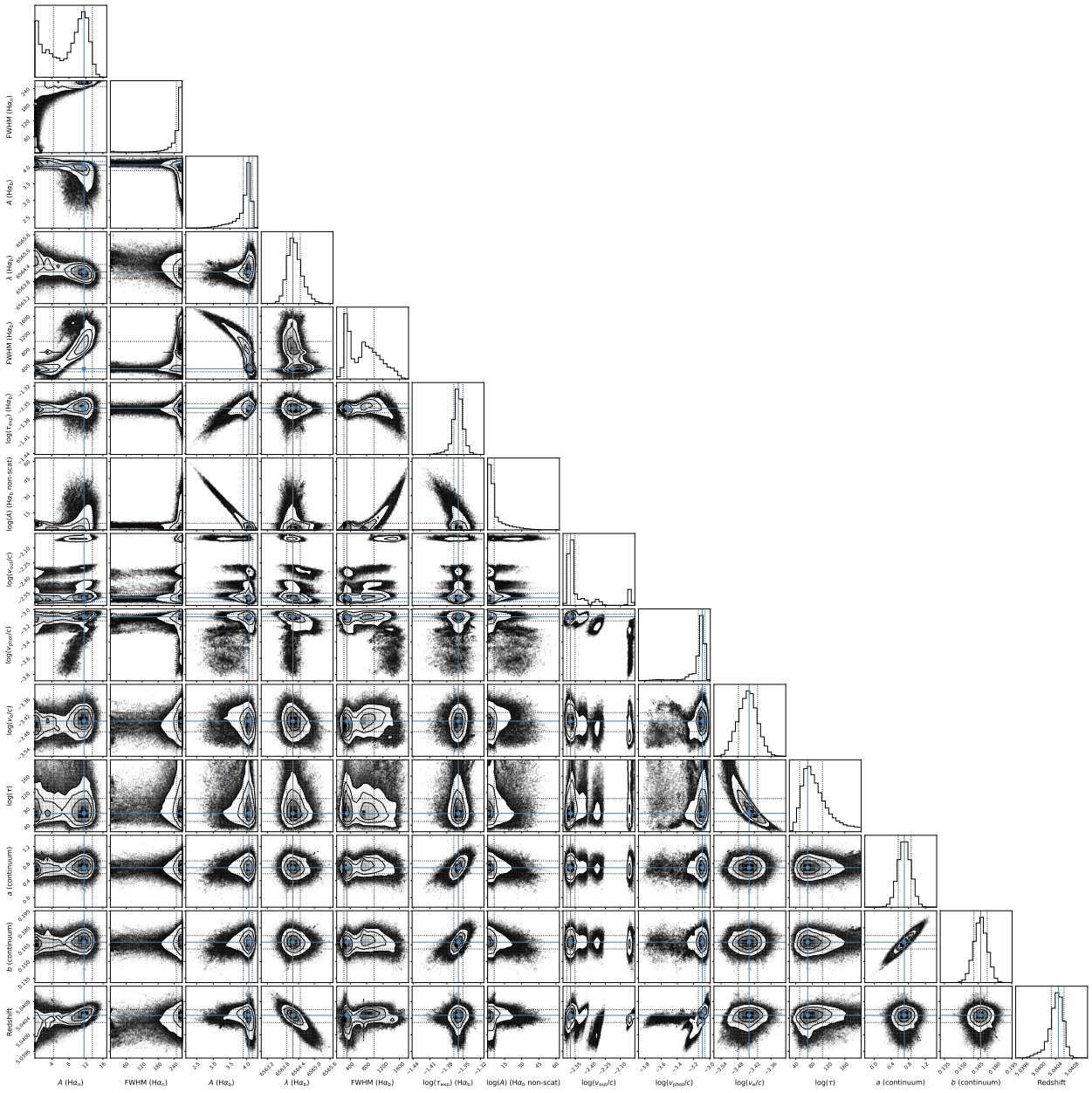
92. pandas development team, T. "pandas-dev/pandas: Pandas", February (2020).

93. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M. et al., "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods*, **17**, 261–272, (2020).

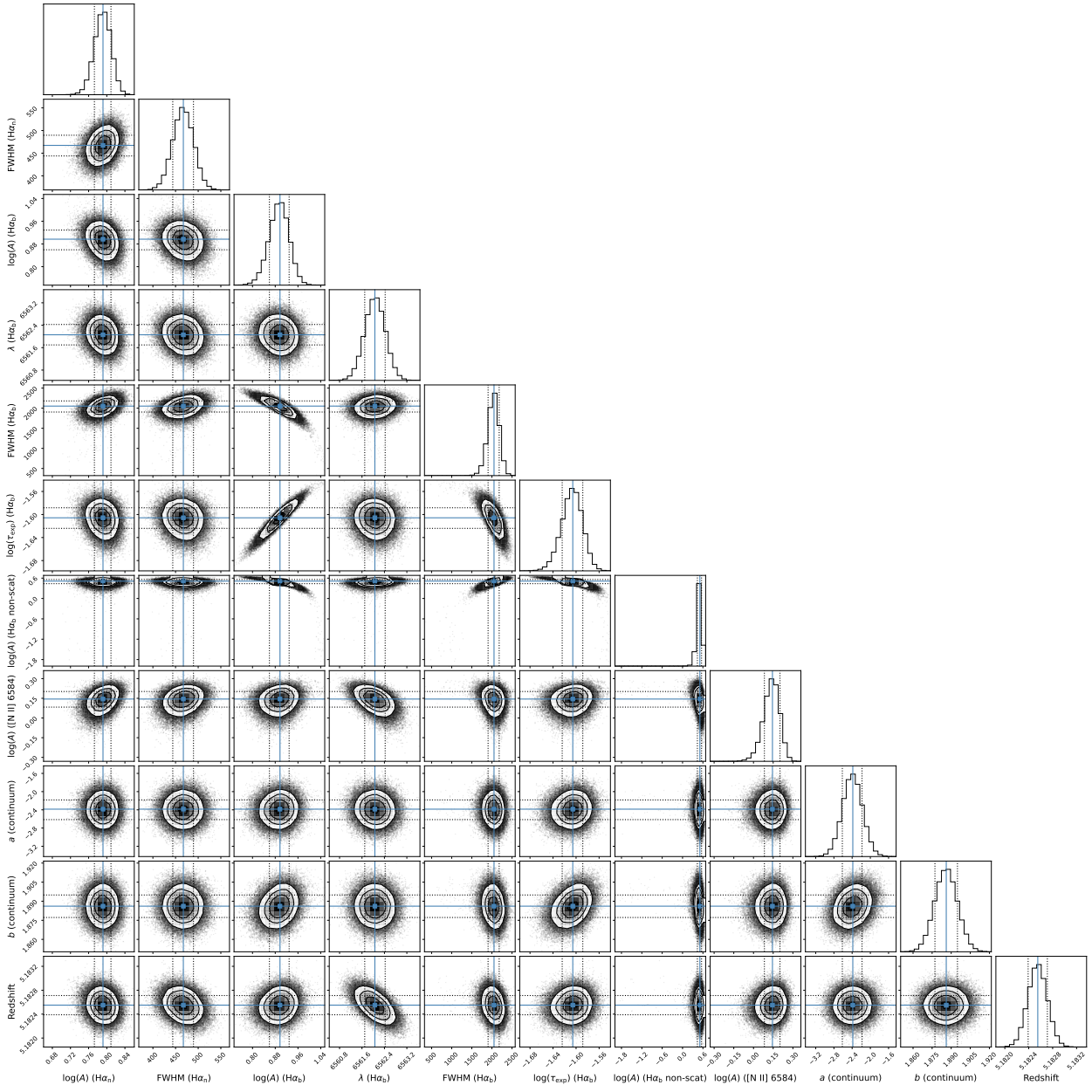
94. Jeffery, D. J. and Branch, D. "Analysis of Supernova Spectra". In *Supernovae, Jerusalem Winter School for Theoretical Physics,* Wheeler, J. C., Piran, T. and Weinberg, S., editors, volume 6, 149, January (1990).

95. Thomas, R. C., Nugent, P. E. and Meza, J. C., "SYNAPPS: Data-Driven Analysis for Supernova Spectroscopy". *PASP*, **123**(900), 237, February (2011).

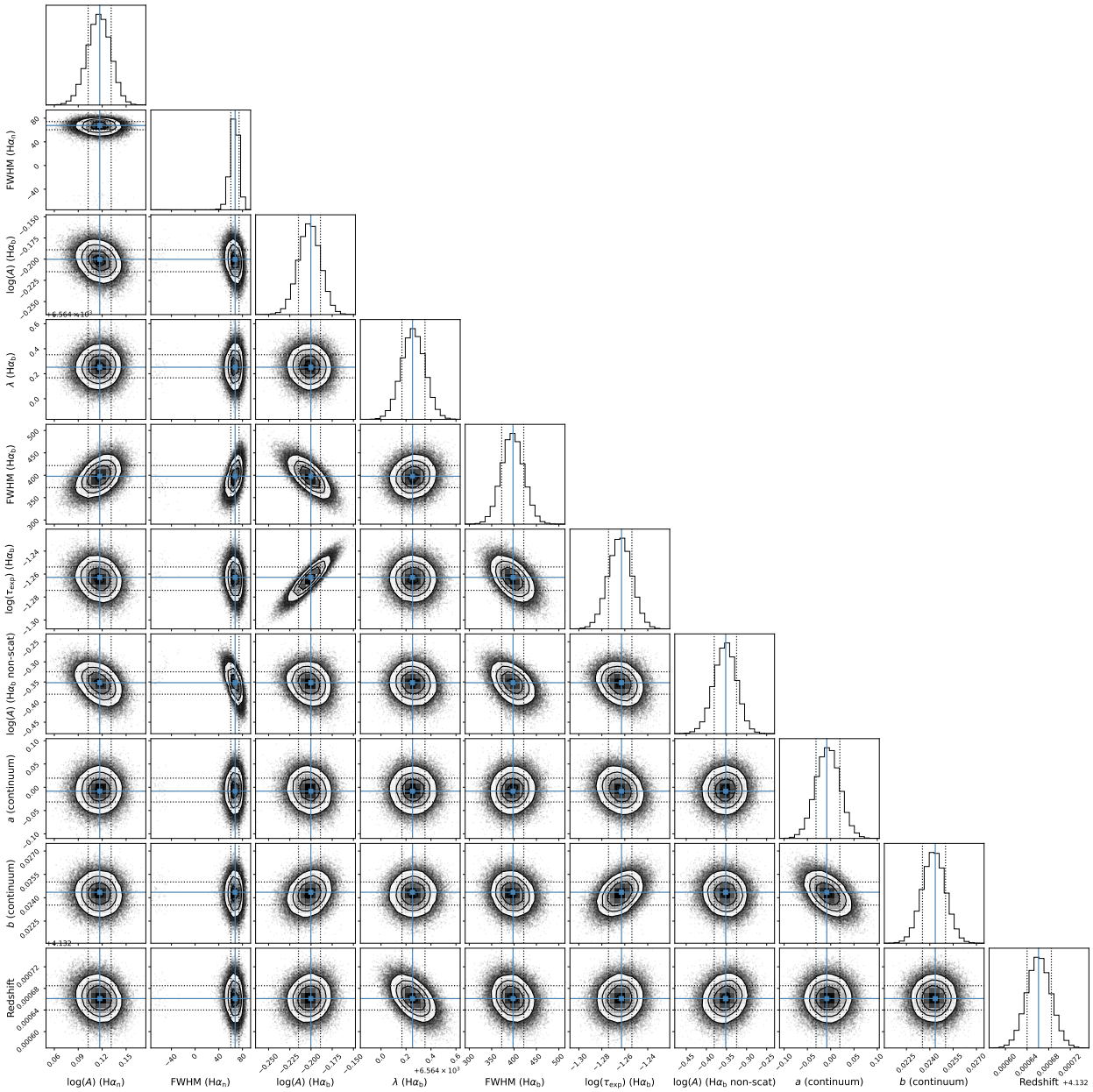
96. Sneppen, A., Watson, D., Poznanski, D., Just, O. et al., "Measuring the Hubble constant with kilonovae using the expanding photosphere method". *A&A*, **678**, A14, October (2023).



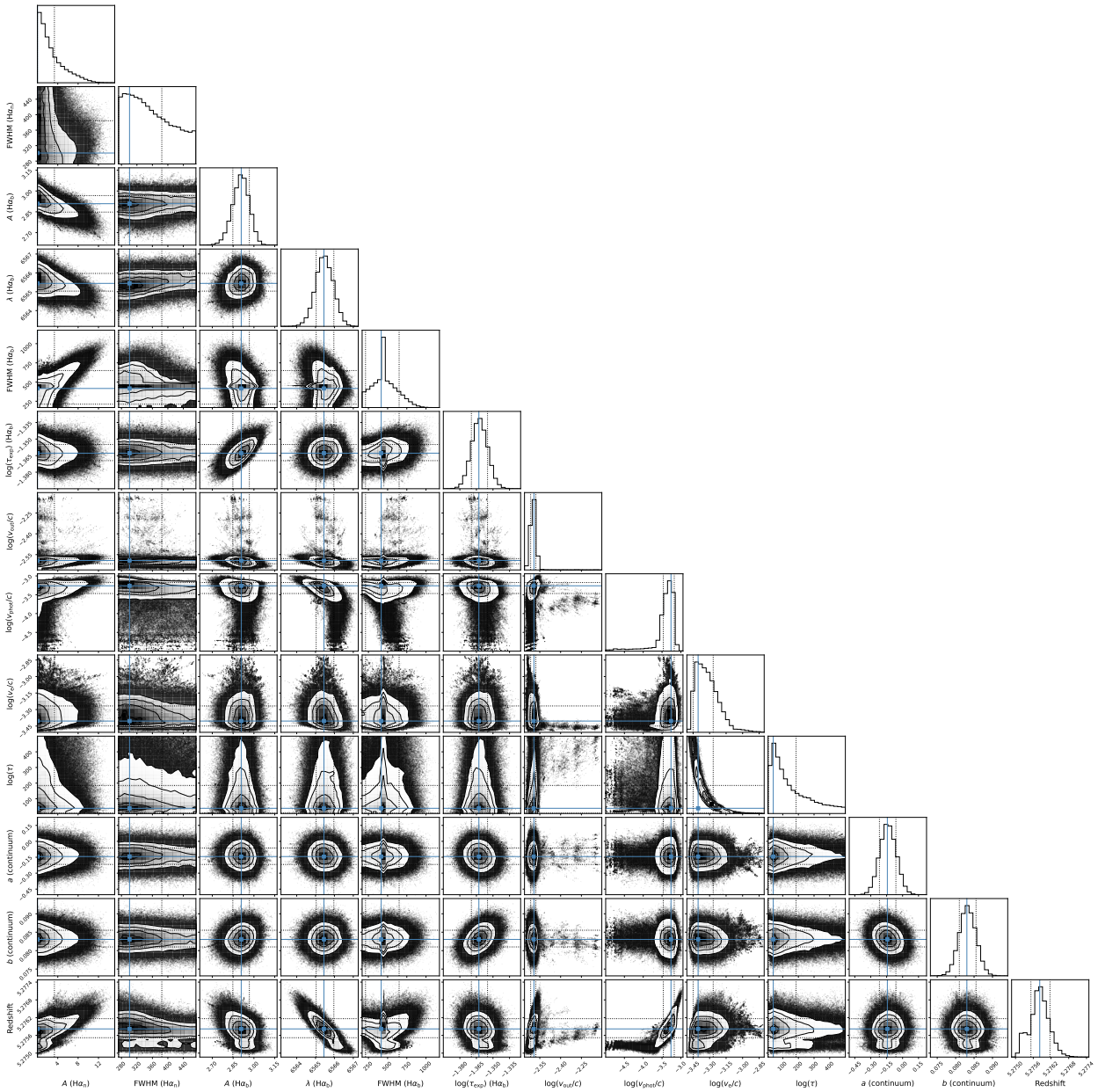
Extended Data Figure 9: Object A. Distributions of the posterior parameter values for the Fiducial model for all components of the Hα line and the continuum. Mode values represent the selected best parameter values (blue solid lines) and the 68% highest density intervals around the modes represent the uncertainty (black dotted lines). The parameters for the narrow H$\alpha_{\rm n}$, broad H$\alpha_{\rm b}$ (either Doppler or electron scattering-broadened) and broad non-scattered H$\alpha_{\rm non-scat}$ line components at some $\mathrm{Redshift}$ include: amplitudes A, FWHM in km s^{-1}, line centroids λ, exponential line widths $\tau_{\rm exp} = 1/W$ in Å^{-1}. The continuum is modelled as a straight line, $ax + b$. The P Cygni absorption/emission feature is modelled using the maximum ejecta velocity $V_{\rm out}$, photospheric velocity $V_{\rm phot}$, reference ejecta velocity $V_{\rm e}$ as fractions of the speed of light c and optical depth of ejecta τ using code presented in [86], originally adapted from Ulrich Noebauer's code `https://github.com/unoebauer/public-astro-tools`.



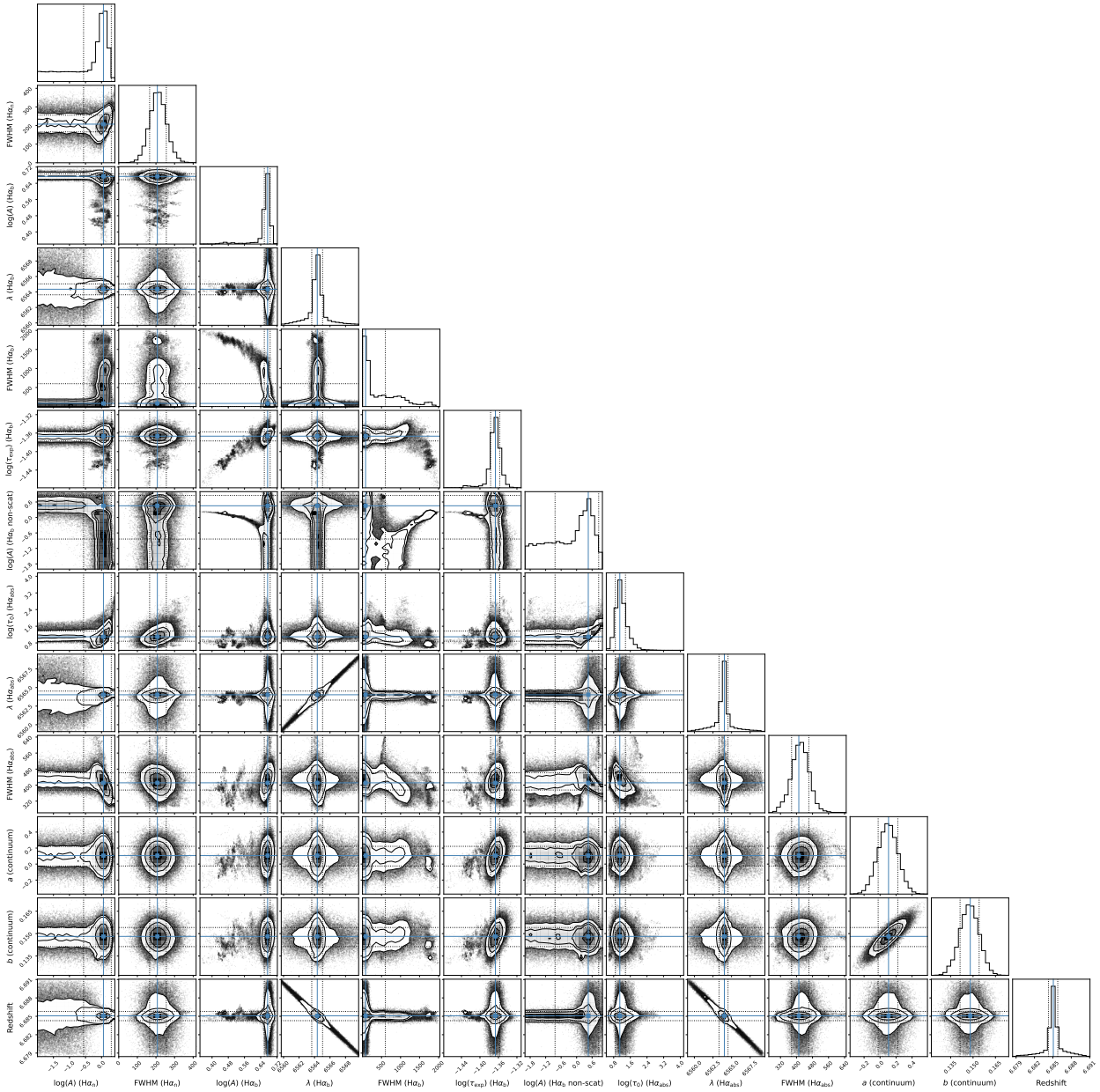
Extended Data Figure 10: Object B. The parameters here are a subset of those described in Extended Data Fig. 9. This object includes the components for [N II] $\lambda\lambda$ 6549,6585 lines fitted using a single amplitude A.



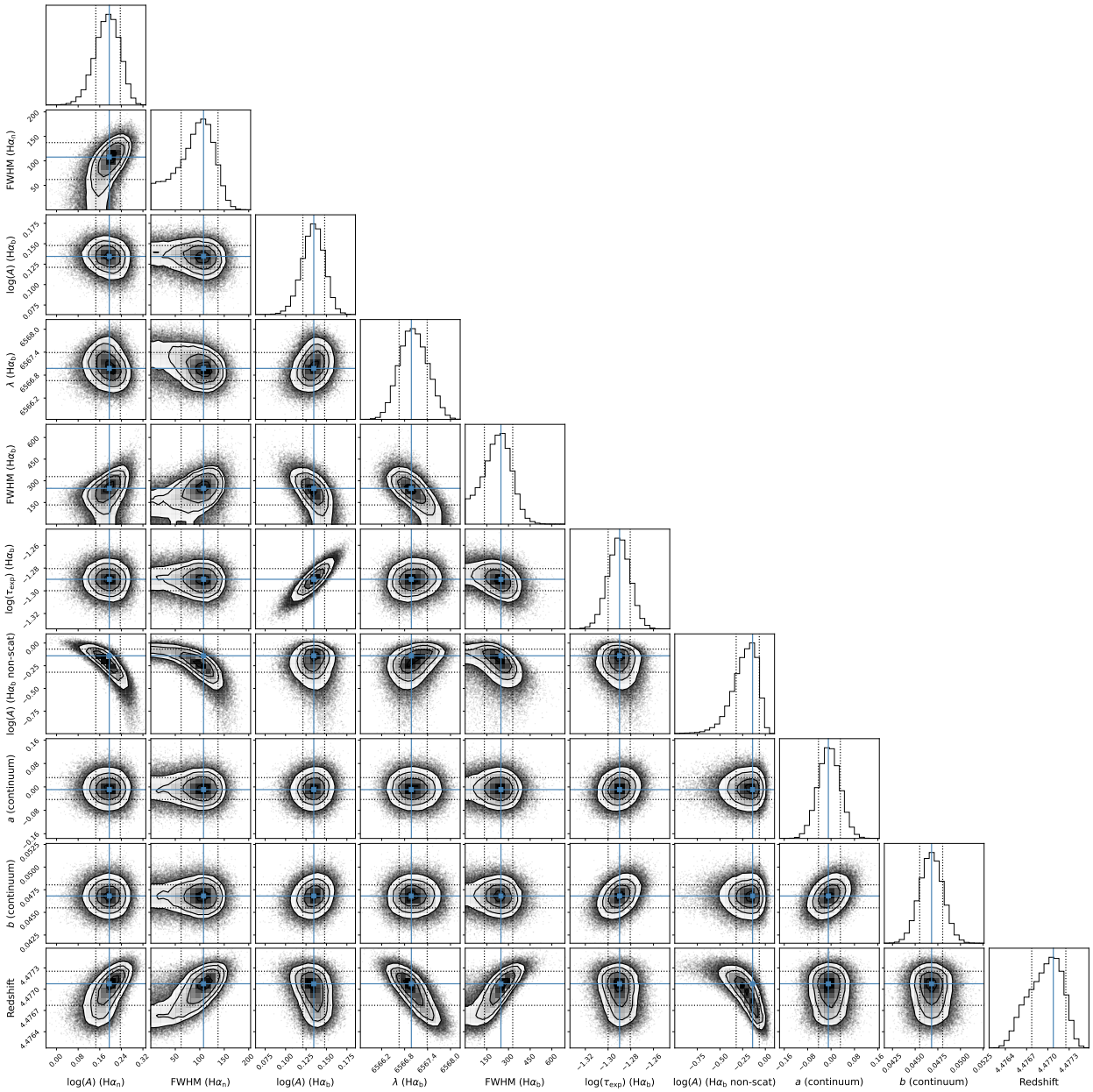
Extended Data Figure 11: Object C. The parameters here are a subset of those described in Extended Data Fig. 9.



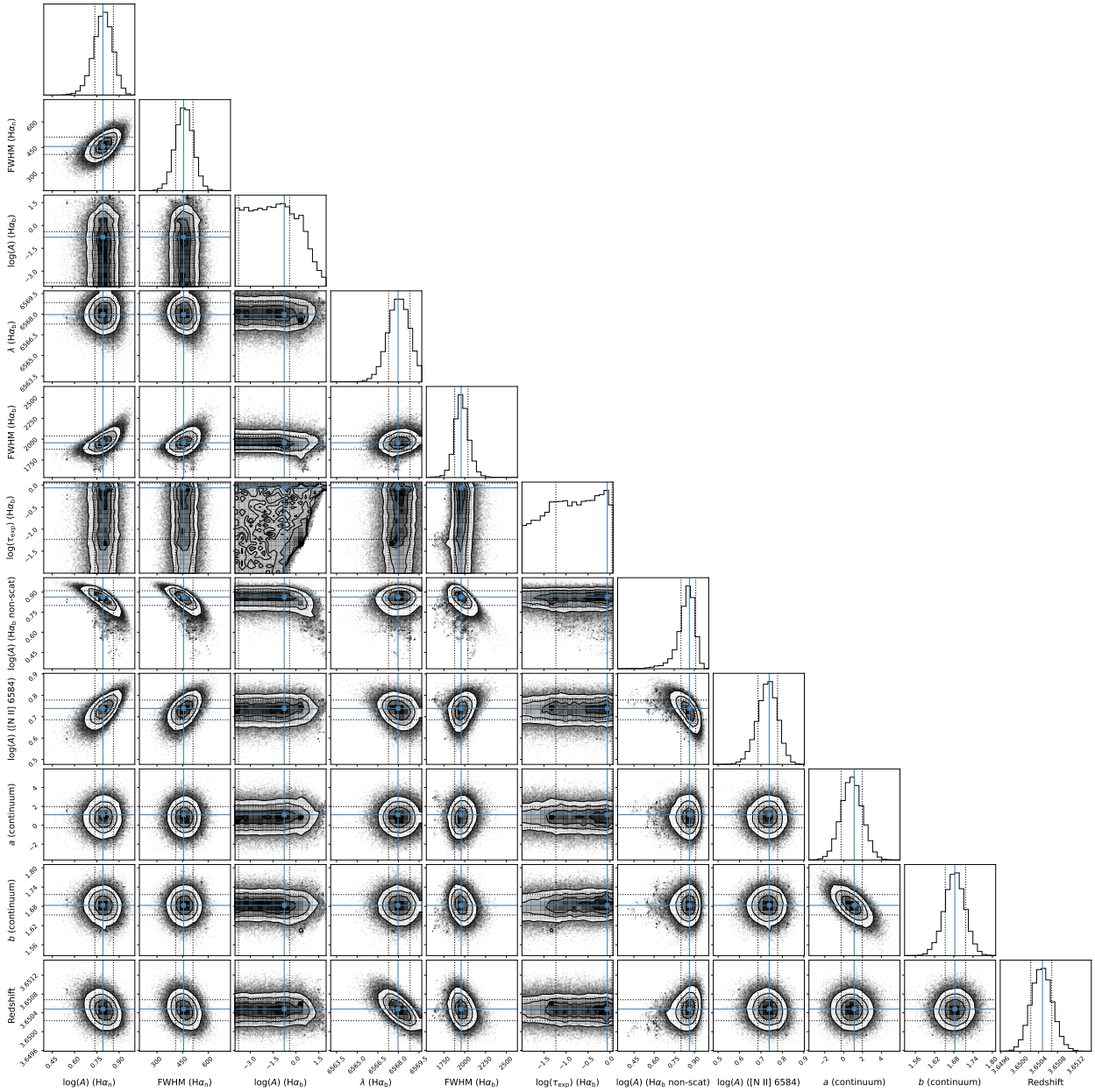
Extended Data Figure 12: Object D. The parameters here are a subset of those described in Extended Data Fig. 9.



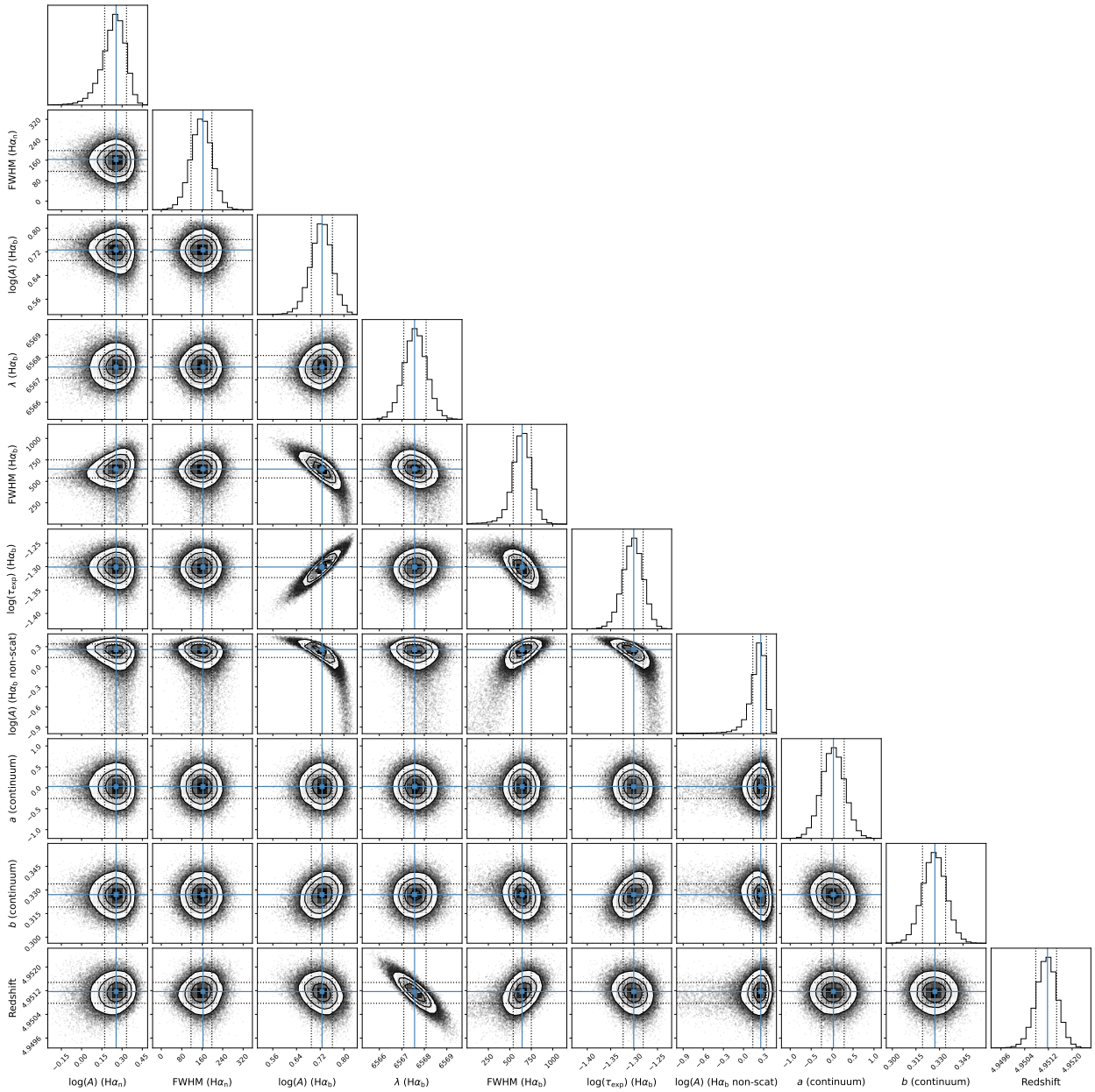
Extended Data Figure 13: Object E. The parameters here are a subset of those described in Extended Data Fig. 9.



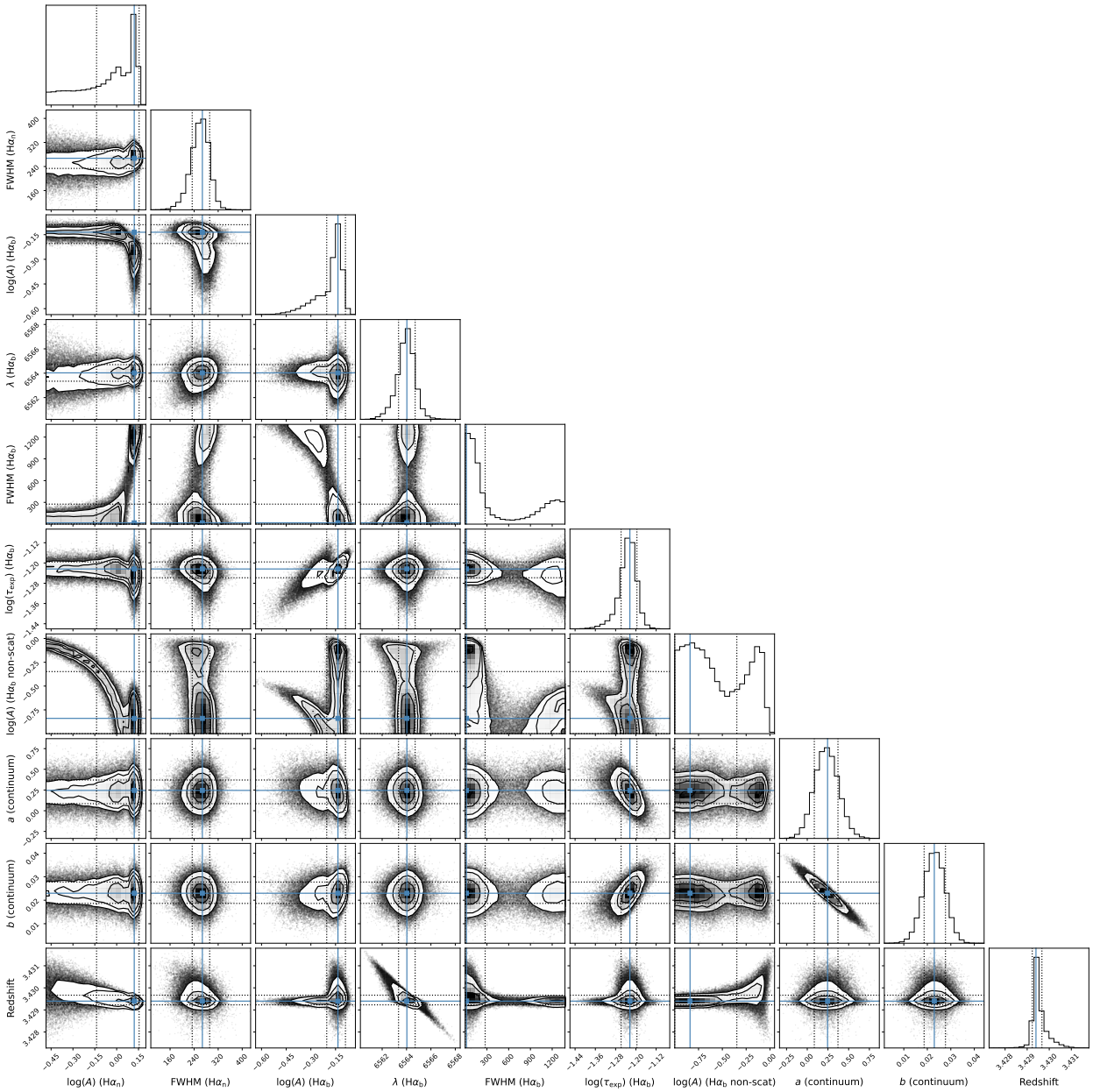
Extended Data Figure 14: Object F. The parameters here are a subset of those described in Extended Data Fig. 9.



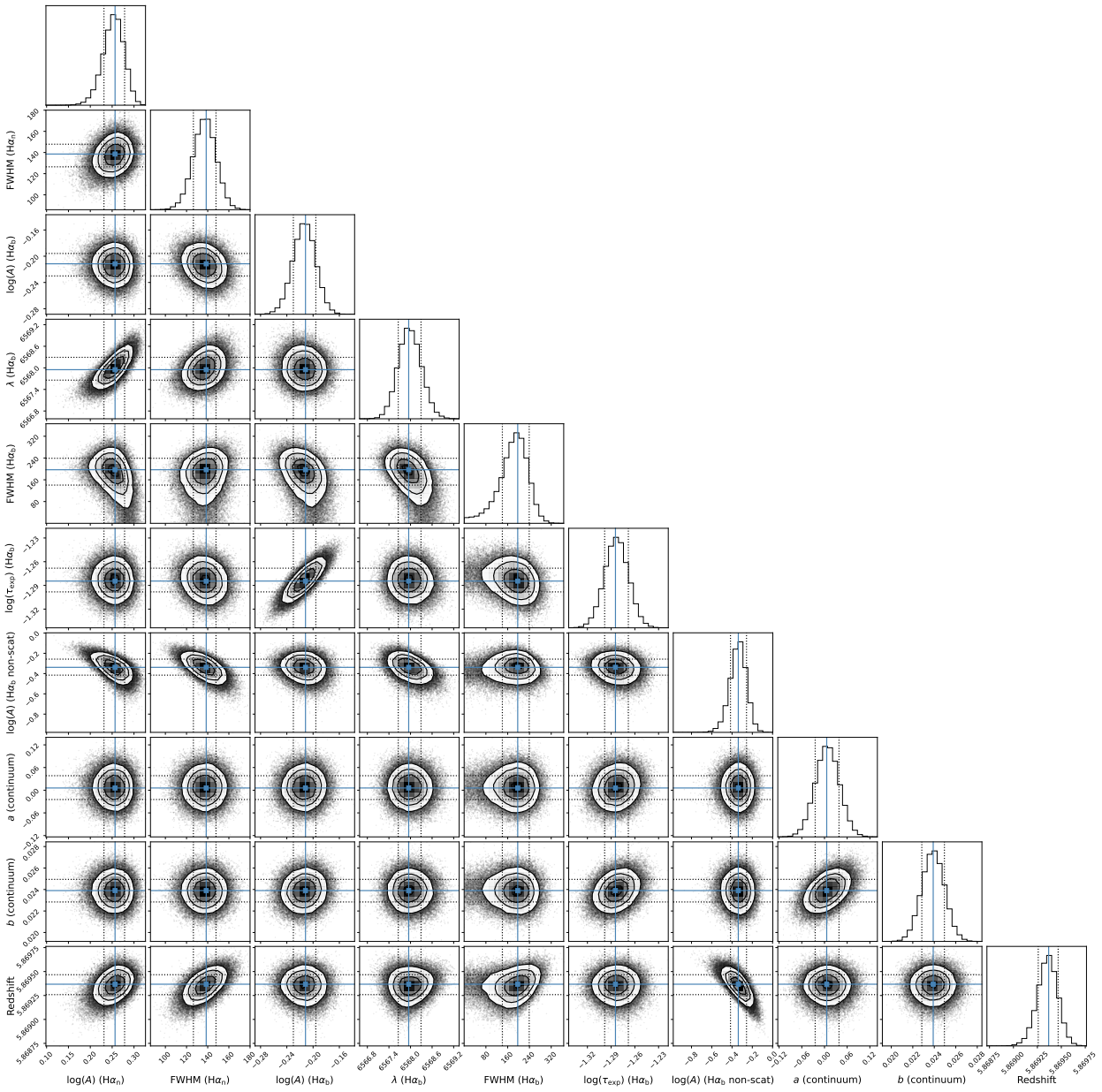
Extended Data Figure 15: Object G. The parameters here are a subset of those described in Extended Data Fig. 9. This object includes the components for [N II] $\lambda\lambda$ 6549,6585 lines fitted using a single amplitude A.



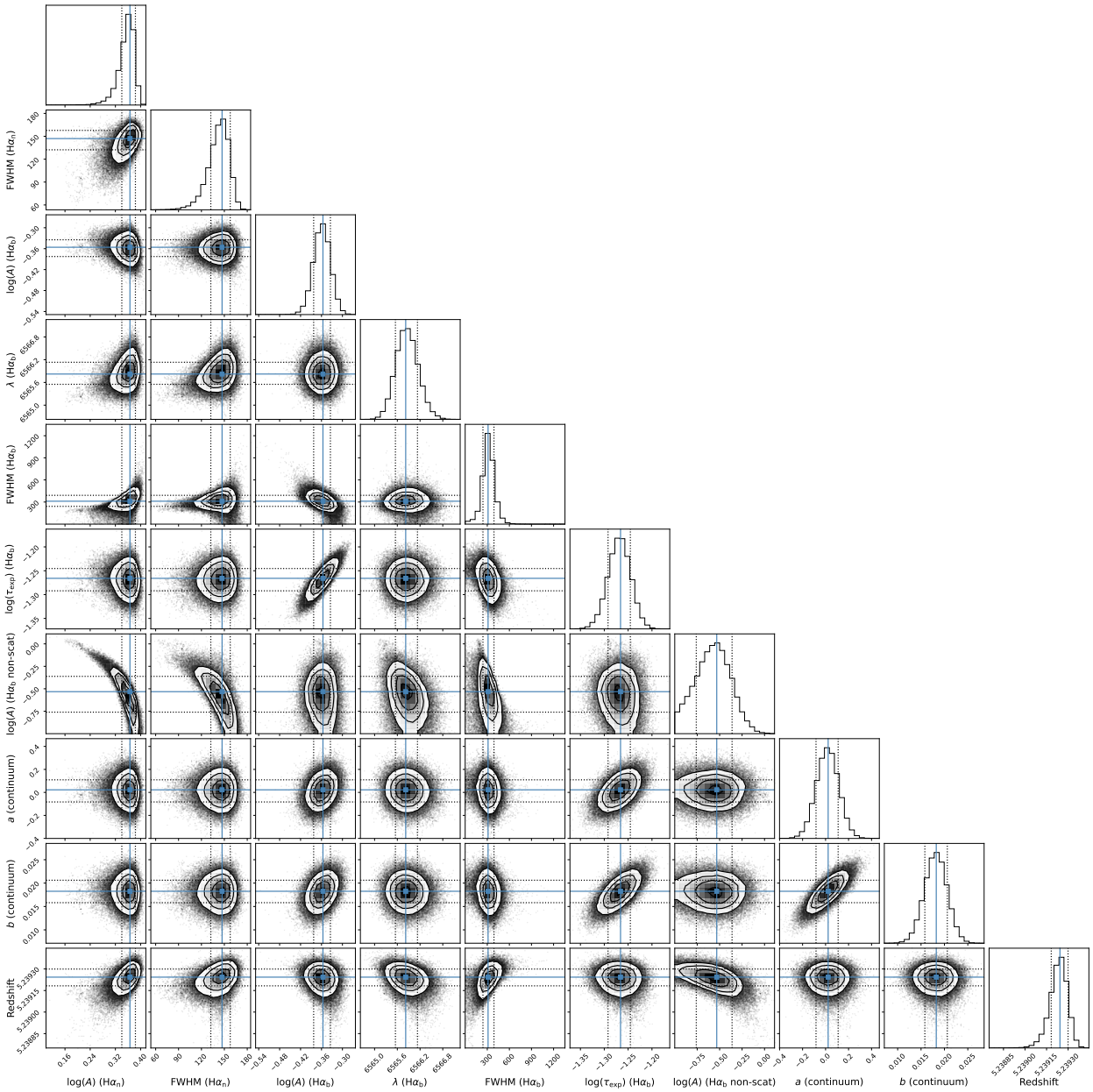
Extended Data Figure 16: Object H. The parameters here are a subset of those described in Extended Data Fig. 9.



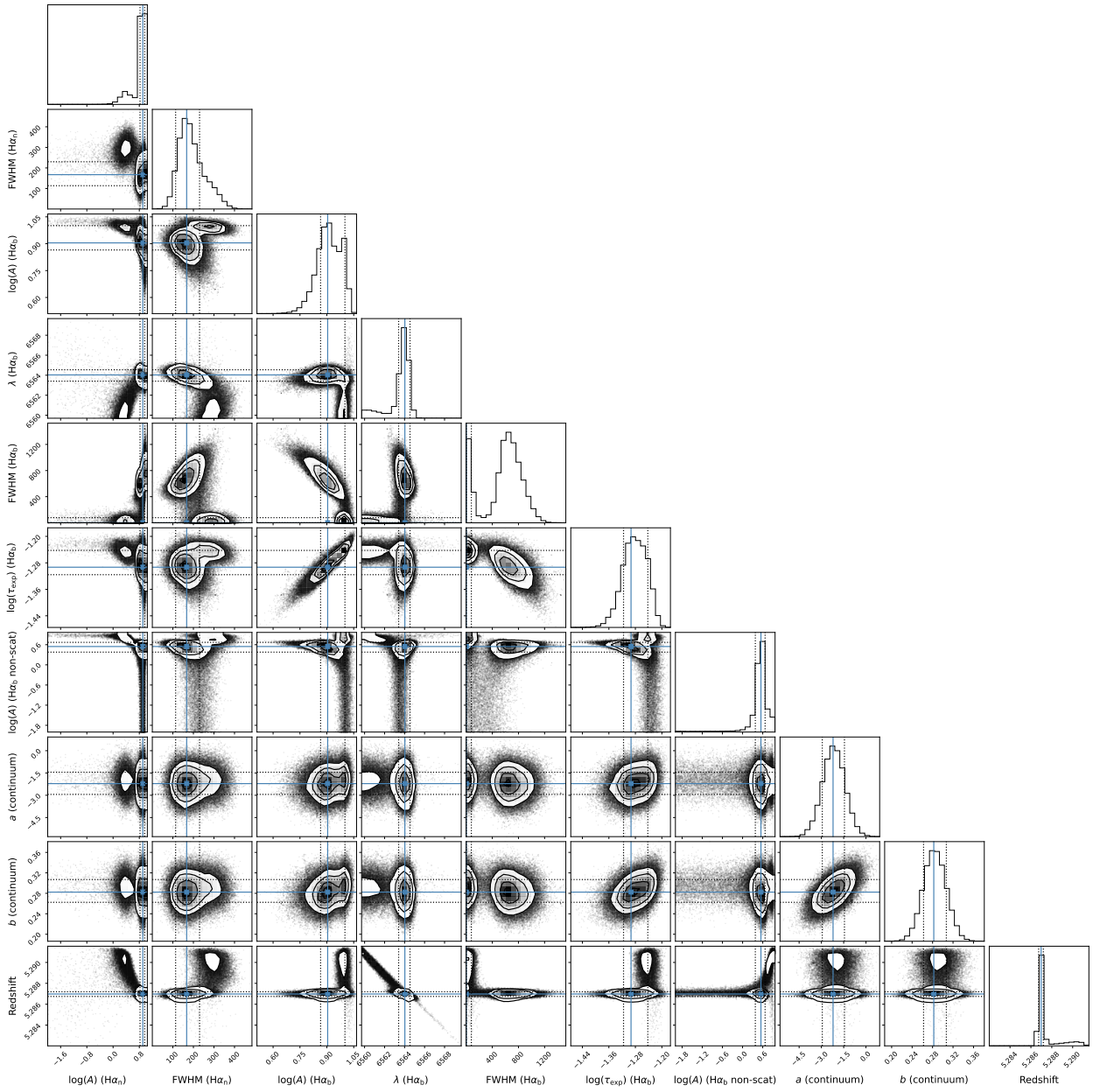
Extended Data Figure 17: Object I. The parameters here are a subset of those described in Extended Data Fig. 9.



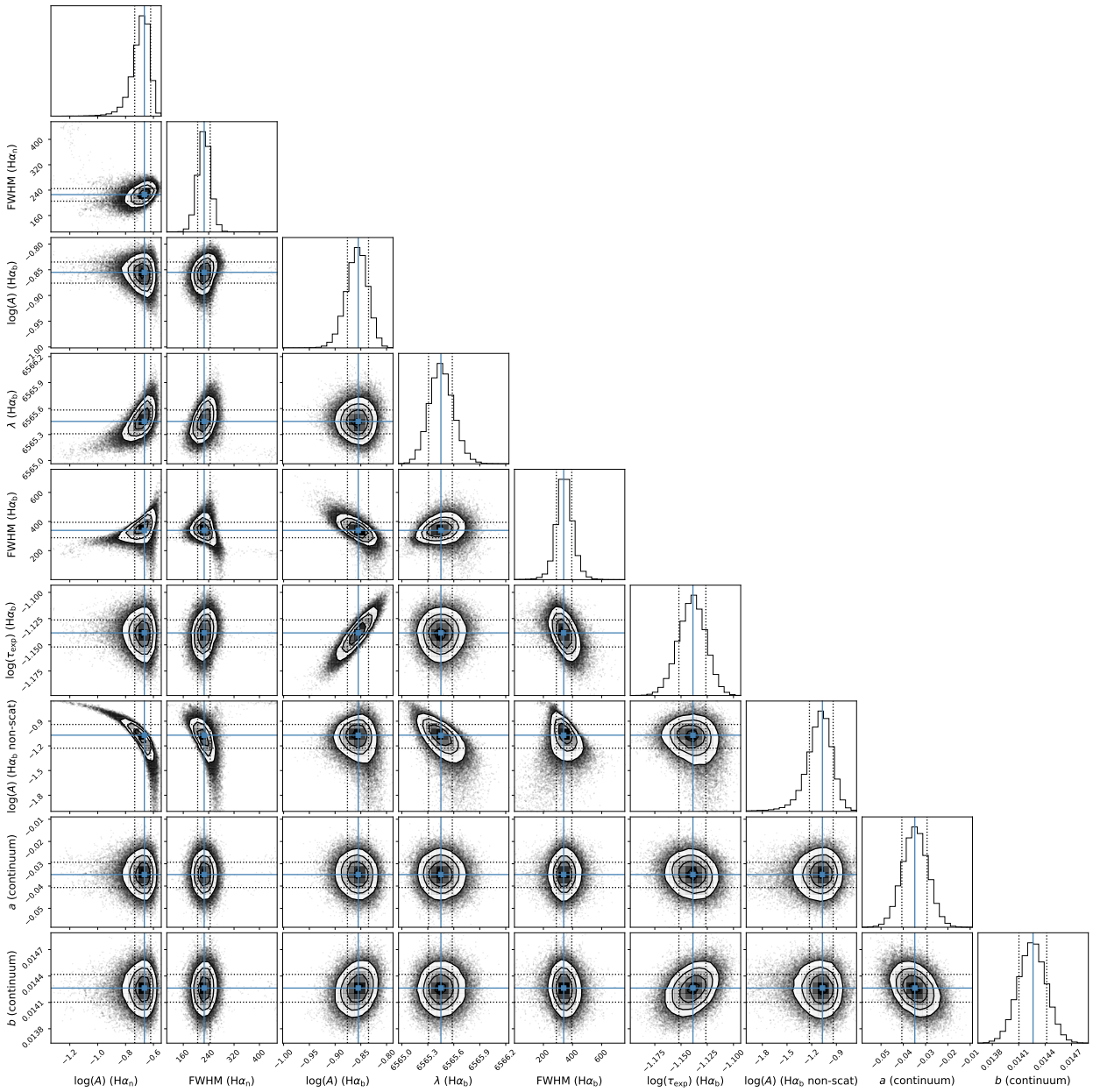
Extended Data Figure 18: Object J. The parameters here are a subset of those described in Extended Data Fig. 9.



Extended Data Figure 19: Object K. The parameters here are a subset of those described in Extended Data Fig. 9.



Extended Data Figure 20: Object L. The parameters here are a subset of those described in Extended Data Fig. 9. This object has two significant peaks in the posterior distribution of the Doppler component width FWHM($\mathrm{H}\alpha_{\mathrm{b}}$), each of which are reported in Table 1, Extended Data Table 4 and Extended Data Figure 4.



Extended Data Figure 21: Stack of low-SNR spectra. The parameters here are a subset of those described in Extended Data Fig. 9.