Constraining the LyC escape fraction from LEGUS star clusters with SIGNALS H II region observations: a pilot study of NGC 628

Jia Wei Teh [1,2,3]★ Kathryn Grasha [2,3,4]† Mark R. Krumholz [2,3] Andrew J. Battisti [2,3]
Daniela Calzetti,[5] Laurie Rousseau-Nepton,[6,7] Carter Rhea,[8,9] Angela Adamo,[10] Robert C. Kennicutt,[11,12]
Eva K. Grebel,[13] David O. Cook [14] Francoise Combes,[15] Matteo Messa,[10,16] Sean T. Linden [5]
Ralf S. Klessen [1,17] José M. Vilchez [18] Michele Fumagalli [19,20] Anna McLeod [21,22]
Linda J. Smith,[23] Laurent Chemin,[24] Junfeng Wang,[25] Elena Sabbi,[23] Elena Sacchi,[26] Andreea Petric,[23]
Lorenza Della Bruna[10] and Alessandro Boselli[27]

Affiliations are listed at the end of the paper

ABSTRACT

The ionizing radiation of young and massive stars is a crucial form of stellar feedback. Most ionizing (Lyman-continuum; LyC, $\lambda < 912$Å) photons are absorbed close to the stars that produce them, forming compact H II regions, but some escape into the wider galaxy. Quantifying the fraction of LyC photons that escape is an open problem. In this work, we present a seminovel method to estimate the escape fraction by combining broadband photometry of star clusters from the Legacy ExtraGalactic UV Survey (LEGUS) with H II regions observed by the Star formation, Ionized gas, and Nebular Abundances Legacy Survey (SIGNALS) in the nearby spiral galaxy NGC 628. We first assess the completeness of the combined catalogue, and find that 49 per cent of H II regions lack corresponding star clusters as a result of a difference in the sensitivities of the LEGUS and SIGNALS surveys. For H II regions that do have matching clusters, we infer the escape fraction from the difference between the ionizing power required to produce the observed H II luminosity and the predicted ionizing photon output of their host star clusters; the latter is computed using a combination of LEGUS photometric observations and a stochastic stellar population synthesis code SLUG (Stochastically Lighting Up Galaxies). Overall, we find an escape fraction of $f_{esc} = 0.09^{+0.06}_{-0.06}$ across our sample of 42 H II regions; in particular, we find H II regions with high f_{esc} are predominantly regions with low Hα-luminosity. We also report possible correlation between f_{esc} and the emission lines [O II]/[N II] and [O II]/Hβ.

Key words: H II regions – ISM: structure – galaxies: individual: NGC 628 – galaxies: star clusters: general – galaxies: star formation – galaxies: structure.

1 INTRODUCTION

Stars are formed primarily in groupings called clusters (e.g. see Portegies Zwart, McMillan & Gieles 2010; Krumholz, McKee & Bland-Hawthorn 2019). Observations of star clusters provide information not just on the properties of clusters themselves, such as the cluster mass and age function (Larsen 2009; Anders et al. 2021), but also allow us to study the hierarchical structure of star-forming galaxies (e.g. Zhang, Fall & Whitmore 2001; Bastian et al. 2005; Grasha et al. 2017a) and to understand the history of star formation (e.g. Pfuhl et al. 2011; Baumgardt et al. 2013). In particular, young star clusters (YSCs; $t \lesssim 10$ Myr; see Mengel et al. 2008; Portegies Zwart et al. 2010; Fujii & Portegies Zwart 2016) are very luminous, and can easily be detected and used to trace active star-forming regions in galaxies that are too distant for individual stars to be resolved (e.g. Whitmore et al. 1999; Bastian et al. 2005; Konstantopoulos

et al. 2009; Adamo et al. 2010; Fedotov et al. 2011; Grasha et al. 2017b; Barnes et al. 2022). One particularly important application of observations of YSCs is to constrain how massive ($\gtrsim 8$ M$_\odot$) stars drive feedback into the surrounding interstellar medium (ISM; for a review see Krumholz et al. 2019), a process that remains only partially understood at best. In this paper, we focus on one aspect of that feedback: ionizing radiation. Lyman-continuum (LyC; $\lambda < 912$Å) emission from O- and B-type stars ionizes the ISM around them, forming H II regions. However, some fraction of the LyC flux may escape the H II region and contribute to ionization elsewhere in the diffuse medium of the galaxy (Niederhofer et al. 2016). Quantifying the LyC escape fraction, f_{esc}, thus allows us to probe the origin of ionization of the diffuse ionized gas (DIG) in galaxies (e.g. see Hoopes & Walterbos 2000; Zurita et al. 2002; Oey et al. 2007; Seon 2009; Weilbacher et al. 2018; Della Bruna et al. 2021; Belfiore et al. 2022). Some fraction of the LyC flux that escapes H II regions may also escape the galaxy entirely and propagate into the intergalactic medium; constraining the fraction that does so is important to the study of cosmological reionization

★ E-mail: jiawei.teh@stud.uni-heidelberg.de
† ARC DECRA Fellow.

(e.g. see Paardekooper et al. 2011; Mitra, Ferrara & Choudhury 2013; Japelj et al. 2017; Ramambason et al. 2020; Chisholm et al. 2022; Saldana-Lopez et al. 2022; Flury et al. 2022a, b). Thus, the measurements of local f_{esc} from star-forming regions impose an upper limit on the latter process.

Astronomers have used a variety of techniques to compute the LyC escape fraction, though there has not been a strong constraint on the range of f_{esc}. One common method is to use surveys of individual massive stars to evaluate the LyC flux, and compare directly to the observed Hα luminosity of their H II regions (e.g. Oey & Kennicutt 1997; Doran et al. 2013; McLeod et al. 2019, 2020; Geist et al. 2022). Recently, McLeod et al. (2019) combined MUSE (Multi-Unit Spectroscopic Explorer) observations of 11 H II regions in the LMC with spectroscopy of massive O stars to directly link the massive stellar population and feedback-related quantities of the regions, yielding a mean escape fraction of $f_{esc} \sim 0.45$. In subsequent work, McLeod et al. (2020) combined observations from MUSE with spectra of massive stars to study two star-forming complexes of the nearby dwarf spiral galaxy NGC 300 and found $f_{esc} \sim 0.28$ and $f_{esc} \sim 0.51$. More recently, Choi et al. (2020) modelled the spectral energy distribution of resolved stars in NGC 4214 and find substantial variation in the LyC escape fraction (0 per cent − 40 per cent). While these surveys represent a significant advance, further progress can be made in two directions: first, with the exception of nearby galaxies, traditional methods of determining the LyC escape fraction are ineffective in extragalactic surveys where spectra for individual massive stars are difficult to obtain. Second, the observations are limited by the small field-of-view (FoV) of these instruments; a larger and more statistically significant sample is necessary to ensure that these results are robust.

The difficulty of obtaining spectra in extragalactic surveys has led to methods to estimate stellar LyC fluxes, and subsequently escape fractions, using broadband photometry rather than spectroscopy. Niederhofer et al. (2016) evaluated the spectral types of the most massive stars in NGC 300 using broadband data from the *HST* and stellar atmosphere models, then compared with surrounding Hα observations. However, they found that the resulting escape fraction is completely dominated by uncertainties in determining stellar parameters from photometric data. Weilbacher et al. (2018) computed the LyC escape fraction of H II regions in the Antennas galaxy by comparing H II region luminosities from MUSE observations to LyC emission from star clusters derived from broadband photometry from *HST*. They estimated the LyC flux by comparing the measured photometry to the GALEV evolutionary synthesis models (Kotulla et al. 2009) and STARBURST99 models (Leitherer et al. 1999). However, they lacked photometric data in the UV band, potentially causing small uncertainty in the age estimates of the youngest stellar populations, which often have the strongest LyC flux. Furthermore, the stochasticity of the stellar initial mass function (IMF) sampling was not taken into account when deriving the LyC flux from star clusters, which may affect the derived ionizing luminosity of clusters at the low-mass end.

In this paper, we revisit the question of the ionizing escape fraction from star clusters using an improved method. First, to ensure that our analysis includes a large enough sample, we use H II region data in NGC 628 drawn from the survey SIGNALS (The Star formation, Ionized gas, and Nebular Abundances Legacy Survey; Rousseau-Nepton et al. 2019). NGC 628 is the first galaxy observed by SIGNALS, which uses the SITELLE Imaging Fourier Transform Spectrograph (IFTS) to achieve high spectral resolution over a FoV large enough ($11' \times 11'$) to enable mapping of full Galactic discs. The SIGNALS catalogue for NGC 628 includes 4285 H II regions imaged

at a spatial resolution of 35 pc (Rousseau-Nepton et al. 2018). We combine this creatalogue with the LEGUS (Legacy ExtraGalactic UV Survey; Calzetti et al. 2015) survey, which provides a total of 1648 candidate star clusters in NGC 628 (see Grasha et al. 2015; Adamo et al. 2017). Combining these catalogues allows us to study the physical connections between H II regions and star clusters for an unprecedented sample size and spatial resolution in a spiral galaxy at a distance of $\sim$ 10 Mpc. Second, we adopt a fully stochastic treatment of star cluster LyC fluxes, following the approach of Della Bruna et al. (2020, 2021, 2022a, b). We use a combination of five-colour photometric data from *HST* (including the crucial UV band) and a stochastic stellar population code (SLUG; da Silva, Fumagalli & Krumholz 2012, 2014; Krumholz et al. 2015a) to infer the ionizing luminosities $Q(H^0)$ from star clusters in NGC 628, and compare this to the observed Hα luminosity $L_{H\alpha}$ of their H II regions.

Compared to previous works, our study has the advantage of both providing a much larger sample size, and including a treatment of the effects of stochastic IMF sampling. Our approach is similar to that of Della Bruna et al. (2022b): they combined MUSE observations of ionized gas with YSC observations from *HST*, identifying $\sim$ 4700 H II region samples in the nearby spiral galaxy M83 to study the link between the feedback of young clusters and their surrounding gas. The consideration of the stochastic sampling of IMF is essential for NGC 628, which has low mass ionizing star clusters ($\gtrsim 10^{2.5}\,M_\odot$; see Grasha et al. 2015). If this method is shown to be feasible, then it can be applied to other galaxies in future LEGUS–SIGNALS observations, allowing us to study star formation across a wide range of galactic environments. It can also be implemented in other surveys, such as the PHANGS MUSE nebular (Emsellem et al. 2022) and *HST* star cluster catalogues (Thilker et al. 2022; Lee et al. 2022b).

In addition, the study of correlations between emission line ratios and the escape fraction of LyC photons has been widely explored at *galactic* scales (e.g. see Jaskot & Oey 2013; Nakajima & Ouchi 2014; Izotov et al. 2018; Jaskot et al. 2019; Nakajima et al. 2020; Ramambason et al. 2020; Flury et al. 2022a, b); at sub-kpc scales, such relations remain partially understood at best. The SIGNALS catalogue, which includes a large number of spatially resolved optical line luminosity measurements, offers a golden opportunity to investigate this question for all rest-frame optical strong line-ratios: [N II]/Hα, [S II]/Hα, [S II]/[N II], [O III]/Hβ, [O II]/Hβ, [O III]/[O II], [O III]/[N II], [S II]λ6716[S II]λ6731, and [O II]/[N II].

The remainder of this paper is as follows. We first describe the observational catalogues we use, and our method for combining them, in Section 2. We assess the completeness of our catalogue and evaluate potential biases in Section 3. We analyse the data in Section 4, discuss the implications in Section 5, and we summarize our conclusions in Section 6.

2 DATA SELECTION

2.1 NGC 628 overview

NGC 628 is a star-forming, grand-design SA(s)c spiral galaxy located at a distance of 9.9 Mpc (Olivares E. et al. 2010; Anand et al. 2021). We select this galaxy for two main reasons. First, NGC 628 is face-on and has a large apparent radius (5.23 ± 0.24 arcmin; Gusev, Egorov & Sakhibov 2014). These features allow detailed study of star-forming regions (e.g. Elmegreen et al. 2006; Berg et al. 2015; Kreckel et al. 2018; Whitmore et al. 2021; Thilker et al. 2022; Lee et al. 2022b; Scheuermann et al., in preparation), dust and metal content (Vílchez et al. 2019), and the temporal evolution of cluster sizes (Grasha et al. 2015; Ryon et al. 2017). Second, NGC 628 is the first galaxy to be

observed by both the SIGNALS and LEGUS surveys, providing a large sample of star clusters and H II regions for analysis.

2.2 Star cluster data

The catalogue of star clusters we use in this work comes from the LEGUS survey (see Grasha et al. 2015; Adamo et al. 2017). LEGUS (Calzetti et al. 2015) is a Cycle 21 Treasury program on the *HST* targeting 50 local ($\lesssim$ 12 Mpc) galaxies. The observations provide broad-band imaging in the NUV (F275W), U (F336W), B (F438W), V (F555W), and I (F814W) filters, covered either by the Wide Field Camera 3 (WFC3) or the Advanced Camera for Surveys (ACS). The star cluster catalogues of LEGUS have enabled a broad range of star formation studies, including testing bar-driven spiral density wave theory (Shabani et al. 2018), the nature and the spatial and temporal evolution of the hierarchical structures of star clusters (Grasha et al. 2015, 2017a, b; Menon et al. 2021; Whitmore et al. 2021; Lee et al. 2022b), IMF studies (Krumholz et al. 2015b; Ashworth et al. 2017), time-scales for star clusters to disassociate with their parent GMCs (Grasha et al. 2018, 2019), and H II region evolution time-scales (Hannon et al. 2019; Whitmore et al. 2020).

LEGUS observations of NGC 628 consist of an east pointing (NGC 628e) and a centre pointing (NGC 628c). At the distance of NGC 628, 9.9 Mpc, the pixel size of *HST*, 0''04 corresponds to a spatial resolution of 1.9 pc. Detailed descriptions of the survey and the standard data reduction of LEGUS imaging data sets are available in Calzetti et al. (2015), whereas the custom pipelines developed to create the initial cluster candidate catalogues, and the analysis of cluster population (i.e. cluster mass function, disruption time-scales) are presented in Adamo et al. (2017). Here, we provide a brief description of steps taken in the automatic pipelines to produce the catalogue of star clusters used in this work. First, the SExtractor (Bertin & Arnouts 1996) algorithm is used on white-light (combination of all five filters) images to extract objects with at least a 3σ detection in a minimum of 5 contiguous pixels. Next, candidates must have a *V* band concentration index (CI; as the difference in magnitudes between radii of 1 pixel and 3 pixels) > 1.4 mag for the centre pointing, and > 1.3 mag for the east pointing, and must be detected in the V band and either B or I band, and a total of at least four bands, with photometric error $\sigma_\lambda \leq 0.3$ mag (see Adamo et al. 2017, for justifications of cluster selection criteria). Second, each cluster candidate with an absolute V band magnitude < −6 mag in the automatic catalogue is then visually inspected and assigned one of four classes: (1) centrally concentrated clusters with spherically symmetric radial profiles; (2) clusters with asymmetric radial profiles; (3) clusters with multiple peaks; and (4) non-clusters such as background galaxies, stars, bad pixels, or chip edge artefacts.

In most of this work, we limit our analysis to Class 1, 2, and 3 clusters, since these are verified to be genuine clusters (or compact associations) by the LEGUS team, using a combination of visual inspection and machine learning techniques (Grasha et al. 2019). We make an exception to this in Section 3.4 and Section 4.1 for reasons we discuss in Section 3, but we set aside the Class 4 sources for the moment. Overall, the LEGUS catalogue contains 1253 Class 1, 2, and 3 clusters. The effective FoV of the survey consists of the regions where there is enough overlapping coverage of the various LEGUS filters such that it is possible to categorize sources into Classes 1–4, which means that there must be overlapping coverage of: (1) V; (2) B or I; and (3) a total at least four bands. We show the footprint of the region satisfying this constraint in Fig. 1.

The LEGUS catalogue also assigns an age, mass, and extinction to every cluster using Yggdrasil (Zackrisson et al. 2011) determinis-

tic stellar population models, using a χ^2-fitting procedure described in Adamo et al. (2017), including propagation of photometric uncertainties through the fit (see Adamo et al. 2010). For NGC 628, the fits assume Solar metallicity in both gas and stars, and include nebular emission computed assuming a 50 per cent covering fraction, a Milky Way extinction curve (Cardelli, Clayton & Mathis 1989), and a foreground extinction of $E(B-V) = 0.06$.

2.3 H II region data

The H II region catalogue comes from the SIGNALS survey (Rousseau-Nepton et al. 2019). SIGNALS is ongoing, and will eventually observe more than 50 000 H II regions in $\sim$ 40 nearby galaxies, using three spectral filters, SN1, SN2, and SN3, covering important optical emission lines for detailed chemical studies of the ISM (see Rousseau-Nepton et al. 2019). SIGNALS has a spatial resolution of $\approx$08, corresponding to 35 pc at the distance of NGC 628 (e.g. see Rhea et al. 2020, 2021a, b). Detailed descriptions of data reduction are available in Rousseau-Nepton et al. (2018). The development of artificial neural network techniques will allow for robust and efficient estimates of the kinematic parameters for the optical emission-line spectra of all the H II regions within SIGNALS.

In SIGNALS, an H II region is defined via a watershed algorithm: one first identifies an emission peak, then determines a zone of influence around it, and lastly outlines an outer limit where the region merges into the DIG background. Once defined, the intensity profile of each H II region is fit to a pseudo-Voigt function, and based on this fit it is assigned one of four categories: (1) symmetric profile, with central concentrated ionizing sources; (2) asymmetric profile, with dispersed ionizing sources; (3) transient region; and (4) diffuse region (see Rousseau-Nepton et al. 2019). We note that the regions in the SIGNALS catalogue are reported as circular regions, by transposing the pseudo-Voigt fitted profile half-width σ into radius (see Rousseau-Nepton et al. 2018, for more details); we discuss the impact of this approximation in Section 3.1.

Additional measures were taken to minimize the inclusion of non-H II regions (e.g. DIG regions, supernova remnants, and planetary nebulae) in the SIGNALS catalogue (Rousseau-Nepton et al. 2018). Nevertheless, we choose a subset of H II regions by applying two additional constraints to further strengthen the fidelity of our sample regions. First, we only select concentrated symmetrical and asymmetrical H II regions in the FoV for the analysis, as transient and diffuse H II regions have intensity profiles that are highly dispersed and are likely dominated by DIG. Second, we check the location of our selected regions within a BPT diagram (Baldwin, Phillips & Terlevich 1981; Kewley et al. 2006) of [O III]/Hβ versus [N II]/ Hα and [O III]/Hβ versus [S II]/ Hα, and retain only regions that lie within the star-forming regime. These constraints reduce the number of H II region candidates from 4067 to 334; we further discuss the possible effect of this selection criteria on our analysis in Section 5.3.1.

2.4 Constructing the combined catalogue

Given our set of 334 H II regions with a specified radius R, and 1253 star clusters, each with a specified central position, we refer to a cluster and H II region as associated if the projected distance D between the central positions is $\leq R$. While this geometrical approach provides a reasonable definition of overlapping regions, we pause to point out three caveats. First, we are treating star clusters as point sources. The approximation here is small, since most clusters have radii $\sim$ 3 pc (Ryon et al. 2017; Brown & Gnedin 2021),
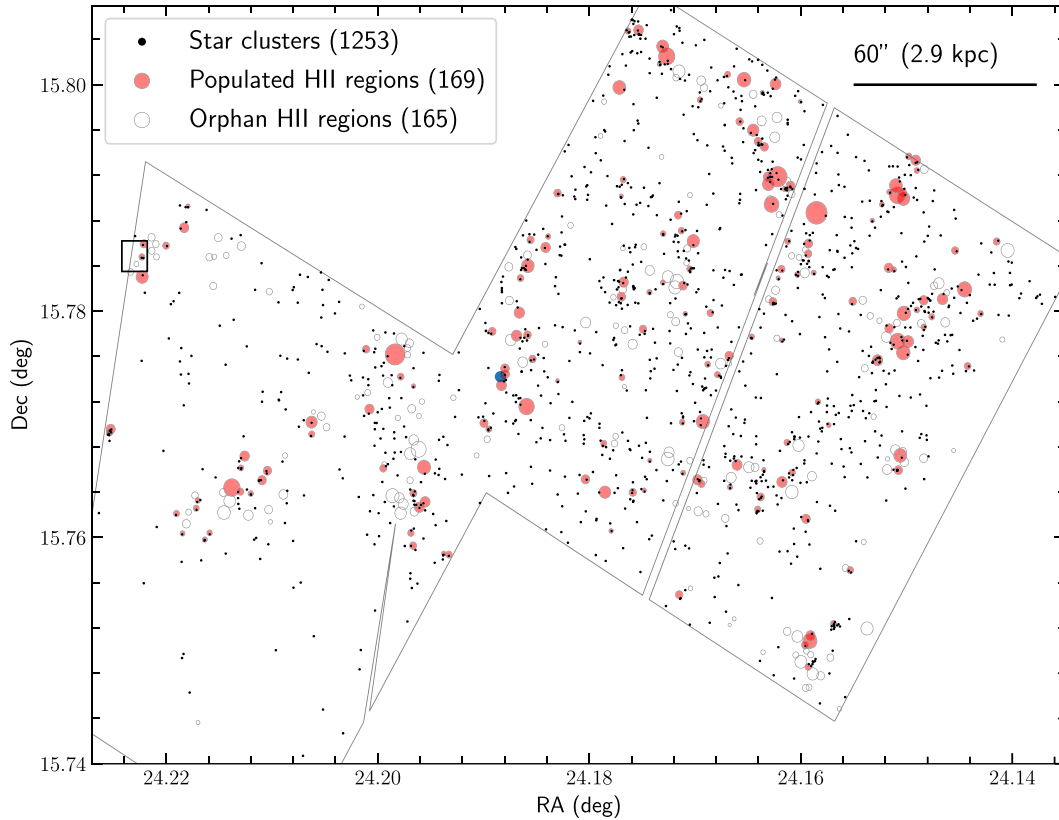
Figure 1. Spatial distribution of LEGUS classified star clusters (Classes 1, 2, and 3; in black dots) and symmetrical + asymmetrical H II regions (in circles, showing true physical sizes with respect to the axes ticks). The filled (red) and empty circles show associated and orphan H II regions respectively. The number of samples in each category is given in the parenthesis. The black outline represents the *HST* FoV defined in Section 2.2. The black horizontal bar indicates a 60″ length, corresponding to 2.9 kpc at distance of NGC 628. A cutout box is randomly selected (black box) for visual inspection (see Fig. 2).

$\sim$ 1 order of magnitude smaller than the SIGNALS resolution of 35 pc. Second, we can only assess overlap in projection, as we do not have information for the depth dimension. This means that false associations due to chance alignments in the line-of-sight are inevitable; we discuss the impact of these in Section 4.1. Third, we do not impose an age cutoff on star clusters in the process of association, for reasons we will also discuss in Section 4.1; this too almost certainly contributes to false associations due to chance overlaps

We present the combined census of 1253 star clusters, and 334 H II regions in Fig. 1. In Fig. 2, we zoom in on a small randomly selected portion of the field, showing the LEGUS NUV and SIGNALS Hα maps directly. It is clear that in cases where there are cluster-H II region overlaps, the association is relatively unambiguous: many of the blue circles marking SIGNALS-catalogued H II regions have one or two obvious slightly extended NUV sources near their centres. Conversely, in the Hα image we clearly see that many of the Hα maxima have clusters sitting on top of them. However, we also observe that many of the H II regions do not seem to host star clusters — 165 out of 334 (49 per cent) H II regions are 'orphans' (refer to Table 1 for terminology). We investigate this topic next.

3 CATALOGUE COMPLETENESS

The phenomenon of H II regions without detected star clusters is particularly intriguing, as there must be an ionizing source that is injecting photons into these regions. How could so many ionizing sources of these H II regions have been missed? In this section, we

present six possible explanations for the high fraction of orphan H II regions. Such an analysis is necessary because, unlike in prior work using targeted observations of H II regions around particular, known star clusters, we are working with catalogues of clusters and H II regions that were made independently of one another. If we are to use the combined catalogue to assess escape fractions, we must understand if there are systematic errors or biases in the two source catalogues we are combining.

3.1 Astrometric uncertainties

We begin by considering the effect of astrometric uncertainties. If the astrometric registration between the LEGUS and SIGNALS images is different, causing an offset in the images, then some H II regions that actually contain star clusters might be misclassified as orphans.

For the SIGNALS survey, a detailed description of the astrometric calibration for all filters in SITELLE is presented in Rousseau-Nepton et al. (2019). They used bright stars in the FoV to calculate the astrometric solution and showed that the resulting calibration has less than 0″.324 uncertainty ($\sim$ 1 SITELLE pixel) almost everywhere throughout the FoV. An additional calibration for the SN3 filter in SITELLE is presented in Martin, Drissen & Melchior (2018), showing that the results are accurate to < 1 pixel. Martin et al. (2018) also compared SITELLE's catalogue of emission-line point-like sources in M31 to other catalogues (Halliday et al. 2006; Merrett et al. 2006) and found an upper-limit uncertainty of 0″.21.

In the LEGUS survey, the WFC3/UVIS data were first aligned using TWEAKREG routine in drizzlepac (Gonzaga et al. 2012).
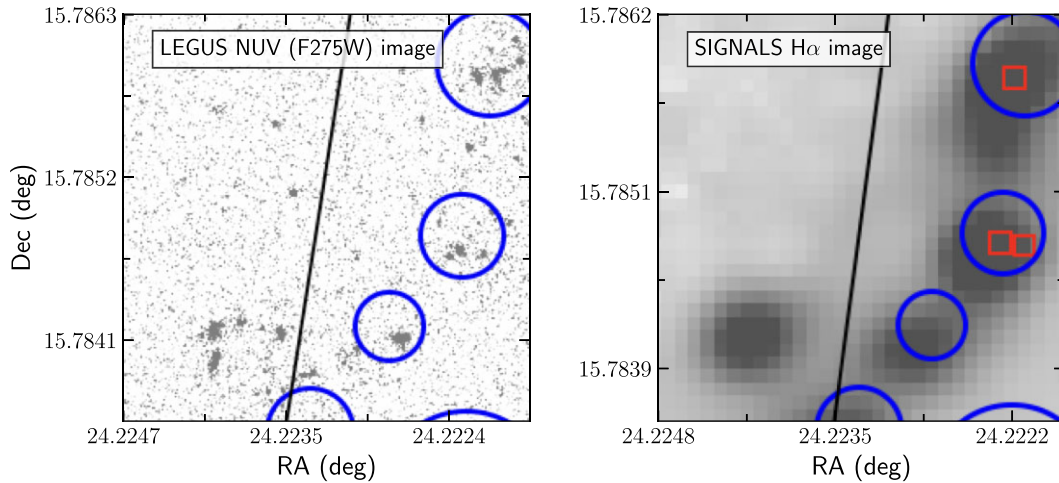
Figure 2. Spatial distribution of H II regions (blue circles) and Class 1, 2, and 3 sources (red boxes) in the cutout defined in Fig. 1, overlaid on the NUV (F275W) image from the LEGUS survey (left-hand panel) and the Hα image from the SIGNALS survey (right-hand panel). We show the red cluster markers only in the right-hand panel to avoid covering up the bright points to which they correspond in the NUV image. Black lines indicate the edge of the LEGUS FoV shown in Section 2.2.

Table 1. A list of the terminology and symbols used throughout this work.

Terminology/symbols	Description
Orphan H II regions	H II regions that lack ionizing star clusters.
Populated H II regions	H II regions that contain ionizing star clusters.
$Q(H^0)$ (photons s^{-1})	The computed LyC flux from star clusters using LEGUS photometric observation.
$L_{H\alpha}$ (erg s^{-1})	The Hα luminosity of H II regions from SIGNALS observation.
$Q(H\alpha)$ (photons s^{-1})	The LyC flux corresponding to $L_{H\alpha}$, obtained with $Q(H\alpha) = 7.31 \times 10^{11} L_{H\alpha}$ (Kennicutt 1998).

The shifts, scale and rotation of individual exposures were then solved and used for matching with catalogues typically containing few hundred bright sources for alignment purposes. The accuracy is better than 0.″004.[1]

These upper limits on the astrometric error are sufficient to rule out astrometric offsets as a significant contributor to the orphan population. Of our 165 orphan H II regions, only 5 have a star cluster within even 0.″1(∼ 4.8 pc) of their outer edge (as defined by their assigned radius in SIGNALS). Thus even if we were to shift the positions of the images by the maximum amount allowed by the possible registration errors, the number of orphan H II regions that could be shifted to the associated category thereby is negligible.

3.2 Single-star H II regions

A second possible explanation for the orphan H II regions is that they are ionized by single, isolated O- or B-type star. In LEGUS, single stars are separated from clusters based on their CI (see Section 2.2), so a single isolated massive star would not be classified as a cluster, even if it had ionizing luminosity large enough to create detectable H II region. However, this too seems unlikely to be a significant contributor to orphans. While 10 per cent − 30 per cent of massive stars are found in isolation from star-forming regions (Gies 1987; Oey, King & Parker 2004), most, perhaps all, of these are runaways (e.g. de Wit et al. 2005; Parker & Goodwin 2007; Pflamm-Altenburg & Kroupa

2010; Gvaramadze et al. 2012). Such stars will be found far from their natal gas, in regions of low gas density (e.g. Hoogerwerf, de Bruijne & de Zeeuw 2001; Drew et al. 2018), where the H II regions they create are likely to be undetectable at extragalactic distances due to low Hα surface brightness. Even in cases where these massive stars travel with low velocities (walkaways; e.g. see de Mink et al. 2012; Renzo et al. 2019), or that they travel along the Galactic plane and end up in high-density regions, the H II regions they produce are typically compact and have sizes ≲ 10 pc (e.g. see Simón-Díaz et al. 2011; Mackey, Langer & Gvaramadze 2013). Given that the orphan regions in SIGNALS have a median radius of ∼ 56 pc, this suggests that the isolated stars may not be a major contributor to the observed phenomenon.

3.3 Oversegmentation

Another possible explanation for the orphan H II regions comes from the possibility of oversegmentation of H II regions from the SIGNALS survey. The idea is fairly straightforward: an area consisting of multiple bright regions of Hα emission, which SIGNALS decomposes into several H II regions, could in reality be a single, giant H II region that is ionized by a central YSC.

Rousseau-Nepton et al. (2018) describes the steps taken to minimize the possibility of false H II region emission peaks. They imposed an intensity threshold on neighbouring pixels of an emission peak, preserving only the brightest peak if two emission peaks are at a distance smaller than the image quality of the observation. They also varied radii of apertures centred on the emission peak to check if the mean flux decreases as the radii increases. However, these checks

[1]We note that the quoted accuracy is limited by the absolute astrometric accuracy of Guide Star II catalogue (< 0.″4; Lasker et al. 2008). Regardless, this is at most ≲ 2 SITELLE pixels, and will not affect our analysis.
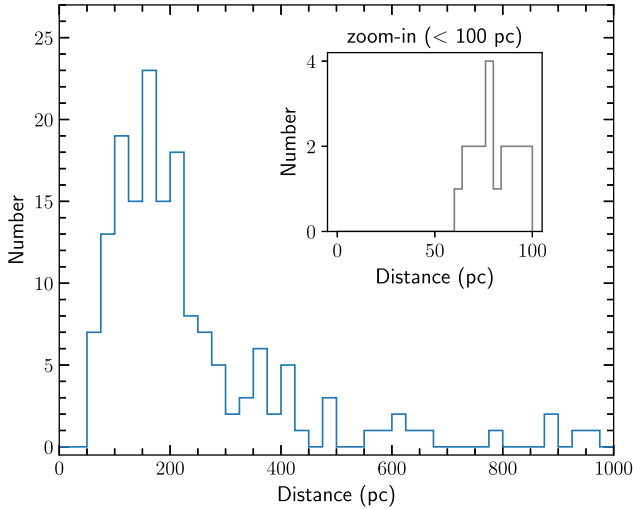
Figure 3. Distance distribution for each orphan H II region to its nearest populated H II region. The inset shows a zoom-in on the part of the histogram in smaller bin (~ 4 pc), showing the number of populated H II regions within distances of <100 pc. We find 20 orphan H II regions out of 165 have a populated neighbour within 100 pc.

only protect against oversegmentation due to surface brightness fluctuations; they are of little help if the Hα surface brightness really is multiply peaked, but is none the less powered by a single source. This could occur in two possible scenarios. First, multiple nearby emission peaks could correspond to spatially separate regions of high-density gas ionized by the same star cluster. Second, a foreground absorber such as a dust lane across an H II region (e.g. Bally 1981) could artificially create surface brightness variations in the H II region large enough for it to be decomposed into multiple emission peaks by the algorithm in SIGNALS. Either of these scenarios could cause an oversegmentation of H II regions, creating the illusion that these regions are ionized by different individual star clusters. If this phenomenon occurs, then we must not simply assume a one-to-one relation for H II regions and star clusters, but must also consider orphan H II regions that are adjacent to a populated H II region that might share the same ionizing source.

To assess this possibility, in Fig. 3 we show the distribution of distances from each orphan H II region to its nearest populated H II region. The plot shows that only 20 orphan H II regions out of 165 have a populated neighbour within 100 pc. Moreover, 100 pc is a generous estimate for the maximum zone of influence of a star cluster; the median radius of all H II regions in the SIGNALS catalogue is ~ 74 pc, and for the Strömgren radius of an H II region to reach 100 pc for a typical ionizing luminosity of 10^{50} erg s^{-1} would require the ionized gas density to be ≲ 3×10^{-3} cm^{-3}, well below the ≈ 10 − 100 cm^{-3} typically inferred from ionized gas density diagnostics in nearby galaxies (e.g. Kewley, Nicholls & Sutherland 2019). This connotes that oversegmentation of H II regions, if present, will not be a major cause of orphan H II regions.

3.4 Potential true clusters in Class 4 LEGUS Sources

Here, we consider the possibility that the orphan H II regions are due to star clusters being falsely identified as non-clusters (Class 4). Identification errors of this type are plausible: Adamo et al. (2017) performed a comparison between clusters of NGC 628 identified from LEGUS and Whitmore et al. (2014), and found only ~ 75 per cent of clusters are common between the studies, as a result of
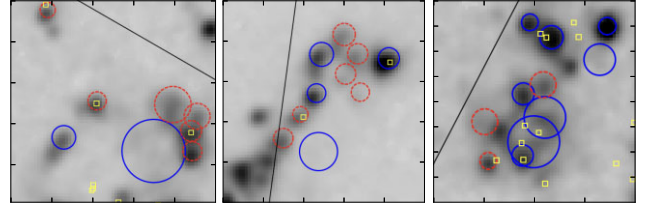
Figure 4. Spatial distribution of Class 4 sources (yellow boxes), orphan H II regions (red dashed circles), and populated H II regions (blue circles), overlaid on the Hα map obtained from the SIGNALS survey. The black outline indicates the FoV defined in Section 2.2. Class 4 sources are seen associated with both orphan and populated H II regions. These Class 4 sources and H II regions are, based on their locations relative to orphan H II regions, possibly associated with them.

a mix of human classification and automated identification. Additionally, Grasha et al. (2019) compared classifications in the LEGUS NGC 5194 cluster catalogue and found 70 per cent − 75 per cent agreement between classifiers for the classes. More recently, Wei et al. (2020) found that different experts agree among themselves 61 per cent of the time as to whether a particular candidate in the PHANGS-*HST* NGC 4656 cluster catalogue should be classified as Class 4.

To assess whether this could contribute to the orphan H II regions, we overlay Class 4 sources, populated H II regions and orphan HII regions on an Hα image of NGC 628 from the SIGNALS data. We find that 36 out of 165 orphan H II regions contain Class 4 sources. Some of these are likely false associations due to chance alignments, but the probability that all 36 associations are chance alignments is very low on both statistical and visual grounds. Statistically, if all Class 4 sources are really background objects or other artefacts, their position should be uniformly distributed across the FoV, and the expected number of chance alignments should then be equal to the fractional area of the FoV covered by orphan H II regions multiplied by the number of Class 4 sources, which is ≈ 13; given this expectation value, the probability of finding 36 overlaps is ~ 10^{-9}. Visually, at least some of the Class 4 sources are almost perfectly aligned with H II regions, as we show in Fig. 4, which makes chance alignment implausible. Therefore, both visual inspection and statistical analysis suggest at least some Class 4 objects are likely true clusters, and can contribute to the orphan H II regions; this motivates us to include Class 4 sources in the analysis of f_{esc} (Section 4). Conversely, however, the fact that we find only 36 Class 4 objects within orphan H II regions, and that we expect 13 of these to be chance alignments, also suggests that misclassification of true clusters as Class 4 objects in the LEGUS catalogue cannot be a major contributor to the phenomenon of orphan H II regions.

3.5 Unclassified star clusters

Next, we investigate whether the orphan H II regions could be powered by star clusters too faint for LEGUS to classify. To do so, we create a set of 10^5 synthetic star clusters using the stochastic stellar population synthesis code SLUG (Stochastically Lighting Up Galaxies; da Silva et al. 2012, 2014; Krumholz et al. 2015a). We then check what fraction of these star clusters would have ionizing luminosities high enough to produce an H II region visible in SIGNALS, while the clusters themselves remain below the LEGUS detection threshold. SLUG creates samples of stellar populations using a Monte Carlo technique that takes into account the effect of stochasticity in the IMF, which is crucial for understanding stellar

populations in the low-mass regime. The parameters we use to create the library of synthetic clusters are a slight modification of those used for the fiducial library described in Krumholz et al. (2015b): we use Padova stellar evolution tracks including TP-AGB stars (Vázquez & Leitherer 2005), STARBURST99 stellar atmosphere models (Leitherer et al. 1999; Vázquez & Leitherer 2005), a Milky Way extinction curve (Cardelli et al. 1989), a Kroupa IMF (Kroupa 2001), Solar metallicity, and 73 per cent ionizing photons converted to nebular emission, with the remaining 27 per cent assumed to be absorbed by dust grains, following McKee & Williams (1997). The ages of the synthesized clusters are randomly drawn from a flat probability density function in age from 10^5 to 10^7 yr, the masses from a $dN/dM \propto M^{-2}$ distribution from 20 to 10^7 M$_\odot$, and the visual extinctions from flat distribution from $A_V = 0 - 3$ (see Krumholz et al. 2015b, and references therein for justifications of these choices of distribution).

Once the library is constructed, the next step is to determine if each synthetic star cluster would be classified as a cluster by LEGUS, and whether it would generate an H II region detectable by SIGNALS. For LEGUS, we use the selection criteria described by Grasha et al. (2015) and summarized in Section 2.2, assuming that a star cluster would not be detected in a given band if it has an apparent magnitude fainter than the 90 per cent completeness limit at the detection thresholds listed in table 1 of Adamo et al. (2017); we use their stated limits for the centre pointing, as this includes the majority of detected cluster candidates. These detectable synthetic clusters are then considered classifiable by LEGUS if they have an absolute V band magnitude < -6 mag. For SIGNALS, we treat a synthetic cluster as detectable if its predicted Hα luminosity, $L_{\mathrm{H}\alpha}$, exceeds the SIGNALS detection threshold for NGC 628 (log $L_{\mathrm{H}\alpha} = 35.65$). We estimate the H$\alpha$ luminosity of the H II regions created by synthetic star clusters assuming

$$L_{\mathrm{H}\alpha} = \alpha_{\mathrm{H}\alpha} f_g Q \left(\mathrm{H}^0\right), \tag{1}$$

where $\alpha_{\mathrm{H}\alpha} = 1.37 \times 10^{-12}$ erg/photon is the conversion efficiency from ionizing photons to Hα emission expected for case B recombination (Hummer & Storey 1987; Kennicutt 1998), $Q(\mathrm{H}^0)$ is the ionizing luminosity of the synthetic star cluster, and $f_g = 0.73$ is the fraction of ionizing photons absorbed by gas rather than dust (McKee & Williams 1997). We further reduce the Hα luminosity by applying a nebular extinction $A_{V,\mathrm{neb}}$ assuming that the ratio of nebular to stellar extinction $A_{V,\mathrm{neb}}/A_{V,\star} = 2.27$ (Calzetti et al. 2000), where $A_{V,\star}$ is the stellar extinction drawn for that cluster.

Fig. 5 shows the age and theoretical Hα luminosity distribution of 10^5 synthesized star clusters. The expected fraction of orphan H II regions is calculated as the ratio of synthetic star clusters that are not detectable or classified by LEGUS but still produce detectable H II regions (grey points above the black horizontal line) to the total number of detectable H II regions (all points above the horizontal line). We find 48 per cent of the H II regions detectable by SIGNALS in our synthetic library are powered by star clusters that would not be classified by LEGUS. For the observed sample size of 334 observed H II regions, this translates to a theoretical number of ~ 160 orphan H II regions. In comparison, we observe 165 orphan H II regions in the combined catalogue. We therefore conclude that the largest factor contributing to orphan H II regions are regions with ionizing star clusters that were not classified by LEGUS.

3.6 OB associations

Finally, we discuss the possibility that the orphan H II regions are ionized not by compact star clusters or by single stars, but are
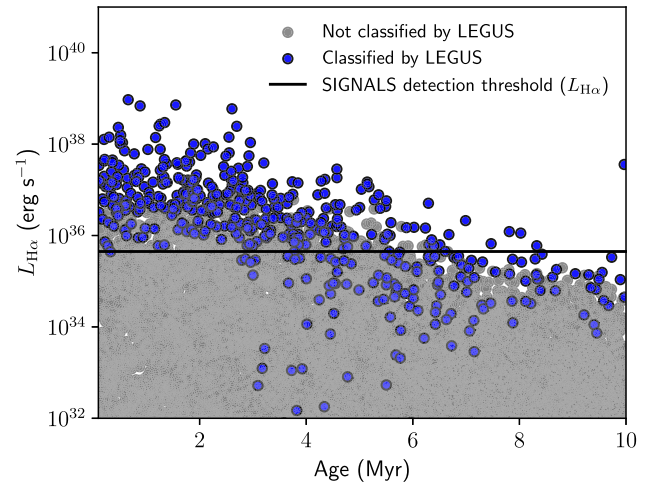
Figure 5. The theoretical Hα luminosity versus age distribution of 10^5 synthesized YSCs from SLUG, showing star clusters that would be detected and classified in LEGUS (in blue), and star clusters that would not (in grey). Data presented in this plot are randomly chosen subsets of the complete analysis to minimize crowding. The horizontal dashed line in black shows the Hα detection limit of SIGNALS for NGC 628 (log $L_{\mathrm{H}\alpha} = 35.65$). We find 48 per cent orphan H II regions (i.e. grey data points above the detection limit of SIGNALS) in our synthetic clusters.

instead powered by loose, gravitationally unbound stellar complexes known as OB associations (e.g. see Lucke & Hodge 1970; McKee & Williams 1997). These OB associations span tens to hundreds of parsecs (Mel'Nik & Efremov 1995; Borissova et al. 2004) and have low spatial densities ($\lesssim 0.1$ M$_\odot$ pc^{-3}), but none the less may be surrounded by dense gas and thus may create detectable H II regions (see Chu & Kennicutt 1988). Most of the OB associations themselves, however, would not be included in the LEGUS catalogue, as LEGUS is optimized to extract only compact sources (see Calzetti et al. 2015; Grasha et al. 2015; Adamo et al. 2017). Indeed, Ryon et al. (2017) and Brown & Gnedin (2021) studied the radius distribution of LEGUS clusters in NGC 628 and showed that LEGUS did not contain sources with radii more extended than ~ 10 pc. These diffuse ionizing sources may contribute not only to the orphan regions, but may also help explain H II regions that are populated by star clusters with seemingly underestimated ionizing luminosity (see Appendix C).

We can estimate the potential contribution of ionizing photons from these missed OB associations by constructing an ionization budget for the entire FoV and checking whether the region as a whole satisfies balance between the injection of ionizing photons and the output of Hα photons. To this end, we include all H II regions (symmetrical, asymmetrical, diffuse and transient), and LEGUS sources (Class 1, 2, 3, and Class 4 sources found in H II regions; see Sections 3.4 and 4.1.1 for justifications). For star clusters, we compute their ionizing luminosity $Q(\mathrm{H}^0)$ using cluster_slug (see Section 4.1.1 for more details); for H II regions, we estimate $Q(\mathrm{H}\alpha)$ using the equation given in Table 1. Overall, we find $f_{\mathrm{esc}} \sim 0.98^{+0.02}_{-0.10}$, where $f_{\mathrm{esc}} = 1 - Q(\mathrm{H}\alpha)/Q(\mathrm{H}^0)$. A value of $f_{\mathrm{esc}} > 0$ indicates that the total ionizing flux required to create H II regions is sufficiently supplied from the photons produced from ionizing sources. Thus, while diffuse ionizing sources missed by LEGUS can potentially explain some orphans, they cannot dominate the total ionization budget of the galaxy.

Table 2. An overview of the analysis on possible causes of orphan H II regions, ordered from highest potential contribution to lowest. Details and justifications for their numbers are available in their respective sections. We note that the results derived in this section should *not* be taken at face value as the representation of the true underlying processes that are creating these orphan H II regions.

Potential Explanation	Potential contribution to orphan H II regions
Unclassified star clusters	160 orphans
Class 4 sources	36 orphans
Oversegmentation of H II regions	< 20 orphans
Astrometric uncertainties	< 5 orphans
Total (in theory)	< 221 orphans
Total (observed)	165 orphans

3.7 Summary of catalogue completeness analysis

In summary, we find that astrometric uncertainties in the catalogues, oversegmentation of H II regions, isolated massive stars, and OB associations will unlikely contribute substantially to the orphan H II regions (see Table 2 for an overview). In addition, the possibility of orphan regions arising from supernovae remnants, planetary nebulae, and DIG regions — if present — represents a minor contribution and will not drive the results we see, as we have discussed in Section 2.3. Instead, we conclude that most orphans are H II regions whose star cluster companions are either not bright enough to be visually inspected in LEGUS, or that they were inspected but were then misclassified as non-clusters (Class 4). The analysis in this section shows that the orphan H II regions are not due to mistaken associations.

4 THE LYMAN-CONTINUUM ESCAPE FRACTION

In this section, we evaluate the LyC escape fraction f_{esc} for H II regions that do have matching LEGUS sources, using two complementary approaches; one where we analyse each H II region individually (Section 4.1), and one where we analyse the collective properties of the cluster population as a whole (Section 4.2).

### 4.1 Individual H II region analysis

4.1.1 Cluster $Q(H^0)$ distributions

The simplest approach to study f_{esc} is to directly compare the ionizing luminosity of star clusters to the ionizing luminosity required to power the Hα emission of associated H II regions. We therefore begin by computing $Q(H^0)$ for star clusters in populated H II regions. To account for stochasticity in the computation of cluster properties, we use cluster_slug, a module that is part of the SLUG software package. cluster_slug uses a Bayesian method to compute the posterior probability density function (PDF) of age, mass, and extinction of star clusters, by taking the photometry of a set of data as input and comparing them to a library of synthesized star clusters. Then, we use the modified version of cluster_slug described in Della Bruna et al. (2021, 2022b) to estimate the cluster ionizing luminosities.

For the library of synthetic clusters, we use the same physical parameters as described in Section 3.5 (i.e. atmosphere models, extinction curve, IMF, and metallicity), with the exception of adopting the Geneva evolutionary tracks (Ekström et al. 2012; Georgy et al. 2012). Both Padova and Geneva models provide a relatively good agreement for models of stars close to the main-sequence, with the Padova models providing superior treatment for stars with ages > 10 Myr, where the contribution of red supergiants and asymptotic giant branch stars become important (e.g. see Leitherer 2005). For the purpose of our study, however, since the majority of targeted clusters are younger than ∼ 10 Myr, we use the Geneva tracks, which are more tuned to reproduce observations of young, massive stars. However, we also present results derived using Padova tracks in Appendix A, and show that they lead to similar qualitative conclusions.

We again assume a flat prior on the extinction from $A_V = 0-3$. However, for the purposes of this analysis we adopt slightly different priors on mass and age than Krumholz et al. (2015b): we take as priors $dN/dM \propto M^{-1}$ in mass from 20 to 10^7 M$_\odot$, $dN/dT \propto T^0$ (i.e. flat) for $T < 10^{6.5}$ yr, and $dN/dT \sim T^{-1}$ for $T > 10^{6.5}$ yr. The mass prior is shallower and the age prior steeper than Krumholz et al. (2015b) used for their analysis of the full LEGUS catalogue because their data set includes all clusters in the galaxy, and therefore contains a larger fraction of older clusters; by contrast, we are selecting not all clusters, but only those associated with bright H II regions, which motivates us to choose priors somewhat more weighted to younger, more massive clusters. Indeed, we compared between our choice of priors and those implemented in Krumholz et al. (2015b), and find that ours can, in general, provide estimations of cluster ionizing luminosity with smaller uncertainty, thereby allowing more data points for further analysis (see Fig. 7). Finally, for each star cluster we use cluster_slug to evaluate the median (50th percentile) and 1-σ uncertainty (16th - 84th percentile) for its ionizing luminosity, $Q(H^0)$, given the measured photometry. We compare the output to the Hα luminosity of H II regions with which these clusters are associated.

Before diving into the analysis, we make four additional constraints regarding the sample of populated H II regions. First, while observations indicate that essentially all clusters older than 10 Myr have cleared their environments (e.g. Whitmore et al. 2011; Hollyhead et al. 2015; Grasha et al. 2018; Hannon et al. 2019, 2022; Kim et al. 2021; Messa et al. 2021), and that the typical time-scale for gas clearing due to feedback from massive stars in clusters is ≲ 5 Myr (e.g. Bromm, Yoshida & Hernquist 2003; Dale, Ercolano & Bonnell 2013; Sukhbold et al. 2016; Chevance et al. 2022; Kim et al. 2022), we do not impose as a prior that the age must be smaller than some value. The reason for this is that we expect at least some of our cluster-H II region associations to be chance alignments along the line of sight, or even clusters that are physically located within H II regions but are so located by chance, not because they are responsible for creating that HII region. We wish to allow our fits to determine older ages in such cases, and these cases are, as we see below, immediately apparent because they lie far from the physically plausible part of parameter space. Second, in this portion of our analysis, we do include Class 4 sources present within H II regions, for the reasons discussed in Section 3.4: at least some Class 4 sources are true clusters, and contribute ionizing fluxes to their associated H II regions. We err on the side of inclusion rather than exclusion because almost all potential interlopers that contribute to the Class 4 population (e.g. background galaxies) are relatively faint in the UV, and this translates to our fitting method assigning them quite low ionizing luminosities. As with older clusters, the extent that these are false associations becomes obvious when we compare them to the observed luminosity. We also present results from an analysis excluding Class 4 clusters in Appendix A, and show that doing so does not yield qualitatively different results. Third, we discard data points where a star cluster is potentially
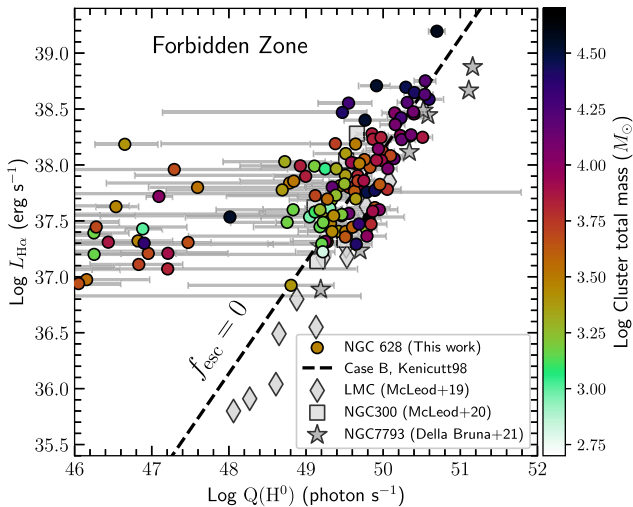
Figure 6. The observed Hα luminosity of H II regions ($L_{H\alpha}$) versus the predicted photon flux of their ionizing sources, $Q(H^0)$, before removing sources with large uncertainties (see Section 4.1). Coloured circles are H II regions in NGC 628 (this study) where the photon flux of corresponding star clusters is computed with modified cluster_slug. The colourbar indicates the sum of the 50th percentile estimates for the masses of star clusters associated with the corresponding H II regions. The $Q(H^0)$ values shown by the points correspond to the median of the PDF, with horizontal errorbars indicating the 15.9th–84.1th percentile ranges. The errorbars are only displayed for a subset of the data to minimize crowding. Diamond, square, and star symbols without error bars are literature values for H II regions in the LMC (McLeod et al. 2019), NGC 300 (McLeod et al. 2020), and NGC 7793 (Della Bruna et al. 2021) respectively. The black dashed line indicates the luminosity–photon flux conversion expected for Case B recombination, i.e. where $f_{esc} = 0$ (see Hummer & Storey 1987; Kennicutt 1998). The plot shown here is contaminated by various sources of uncertainty; we remove them before further analysis (see discussion in Section 4.1).

associated with multiple H II regions. Such multiple associations are possible because, while the SIGNALS H II region decomposition algorithm assigns each pixel of the image to a unique H II region, the SIGNALS catalogue approximates irregular-shaped H II regions as circular regions, and there is no guarantee that these circles do not overlap. Overlapping regions that also contain a LEGUS cluster are therefore excluded, since we cannot make unique cluster-H II region associations for them. Finally, we exclude cluster-H II region associations whose distance from the centre of the H II region to the nearest FoV boundary is $\leq 1.5R$, where R is the radius of H II region. We impose a generous threshold here to avoid the inclusion of H II regions whose ionizing sources are potentially not observed by the LEGUS survey. These constraints further reduce the number of H II region candidates from 169 to 139.

We present the initial result in Fig. 6, showing the Hα luminosity of all 139 H II region versus the ionizing luminosities of its associated sources; where more than one cluster falls within an H II region, we sum their ionizing luminosities. This plot can be understood as follows: if every LyC photon is absorbed by hydrogen atoms in the envelope of an H II region, and the H II region reaches ionization-recombination balance, then all H II regions should lie on the Kennicutt (1998) line. H II regions that lie below this line have lower Hα luminosities than expected from the LyC flux of their source stellar population. This indicates that ionizing photons are not completely absorbed by gas in the region. Conversely, H II regions that lie above the line substantially should be forbidden, as the

inferred ionizing luminosity is insufficient to produce the observed Hα luminosity.[2]

Two noticeable features can be inferred from our initial result in Fig. 6. First, we remark that most data points have large 1-σ uncertainties in ionizing luminosity. This is because cluster_slug 's 68 per cent confidence range properly accounts for the uncertainties from IMF stochasticity and from degeneracies in age, mass, and extinction due to the cluster's location in colour space (see Krumholz et al. 2015b). Moreover, when the PDF is multipeaked, which is a relatively common outcome of a full stochastic analysis (e.g. see Fouesneau et al. 2012; Krumholz et al. 2015a, and Appendix B), the 16th–84th percentile range becomes very broad, since it must expand to 2015a encompass both peaks. Additionally, unless the star clusters are massive enough to fully sample the IMF, the ionizing luminosity diagnosed from optical/UV photometry will come with large uncertainty. This is due to the photometry of low-mass clusters being dominated by a few massive stars, and colours of individual stars in optical/NUV bands are very insensitive to mass once the stellar effective temperature is high enough to shift its spectral energy distribution peak out of the observed bands. A 25 M$_\odot$ star has nearly the same colour in available broadband optical/UV bands as a 75 M$_\odot$ star, but the 25 M$_\odot$ star will have an ionizing luminosity many orders of magnitude lower than the 75 M$_\odot$ star.

Second, we find that data points where the entire 1-σ confidence interval lies above the Kennicutt (1998) line are mostly clusters that are at the low mass end of the observed mass distribution, and with ages $t > 5$ Myr where the inferred ionizing luminosity has large variance. This result is expected and can be interpreted in two ways: one implication is that these clusters have multipeaked or skewed PDFs, due to the stochasticity of IMF sampling in the low-mass regime. Thus, although the median of the ionizing luminosity lies to the left of the Kennicutt (1998) line, there is none the less a reasonable chance that the true ionizing luminosity lies to the right of the line. Another possible explanation is that these are not real associations, i.e. the detected star clusters are chance alignments with their H II region pairs. This is likely to be the case for most of the data points that are located far above ($\gtrsim 0.5$ dex) the Kennicutt (1998) line. In these cases, it is likely that the star clusters that are responsible for the ionization of these H II regions are unclassified because, while bright at ionizing wavelengths, they fall below the LEGUS catalogue's detection limits in the optical bands, while another cluster that is detectable by LEGUS (i.e. that is bright in optical but faint in ionizing photons) is associated with the H II region due to chance alignment.

To mitigate this uncertainty, we truncate the sample by retaining only H II regions for which the 1-σ uncertainty on the ionizing luminosities of the associated star clusters is < 0.5 dex. We show this reduced sample in Fig. 7. We note that, by imposing this error constraint we are effectively retaining the young ($\lesssim 5$ Myr) and high-mass end ($\gtrsim 10^4$ M$_\odot$) of the sample, which are the clusters with the least stochastic variation. Even with this cut, some data points remain well within the forbidden zone (highlighted with green and blue polygons in Fig. 7); we discard these as well, and defer a discussion of them to Appendix C. This reduces the number of data points from 139 to 42. These represent a sample of H II regions for which we can, based solely on the photometry, estimate the ionizing luminosity of the associated star clusters with relatively high confidence.

---

[2] Assuming no systemic errors, H II regions may still lie above the line due to ionization sources other than star clusters.
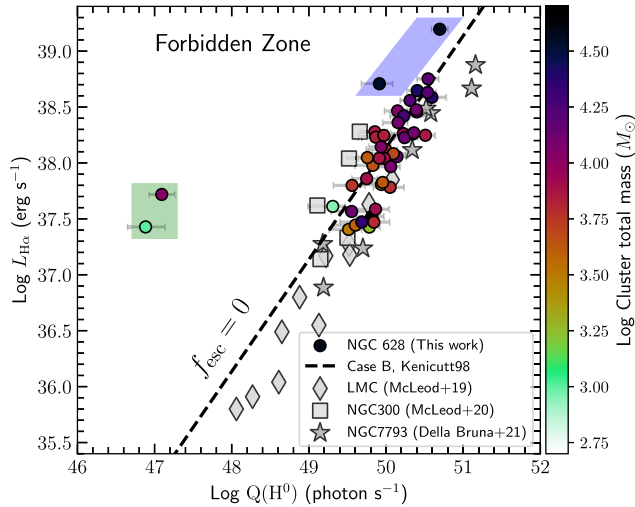
Figure 7. Same as Fig. 6, but only retaining H II regions for which the 68 per cent confidence interval on the ionizing luminosity of the associated star clusters is < 0.5 dex (see Section 4.1). The green and blue polygons, coloured differently to indicate different origins of peculiarity (see discussion in the main text, and in Appendix C), highlight data points which will not be included in our calculations of $f_{\rm esc}$. The remaining data points are in agreement with theoretical expectations (i.e. whose $Q(H^0)$ distribution lie on or below the dashed line).

4.1.2 Probability distribution and caveats for f_{esc}

For our remaining sample, we evaluate the LyC escape fraction by drawing 10^6 Monte Carlo samples from the converged Monte Carlo Markov Chains computed by `cluster_slug` to represent the probability distribution for $Q(H^0)$; where multiple clusters fall within a single H II region, we draw samples for each cluster and sum to produce 10^5 realizations for the *total* ionizing luminosity. For each realization, we compute the escape fraction as

$$f_{\rm esc} = 1 - \frac{L_{\rm H\alpha}}{\alpha_{\rm H\alpha}\, Q(H^0)}, \qquad (2)$$

where $Q(H^0)$ is the ionizing luminosity for that sample, $L_{\rm H\alpha}$ is the observed Hα luminosity, and $\alpha_{\rm H}$ is the conversion from ionizing photons to Hα luminosity; intuitively, this equation simply measures how far below the dashed line in Fig. 7 a given data point lies. The result is a PDF of $f_{\rm esc}$ for each cluster.

We pause here to point out two features of $f_{\rm esc}$. The first is that our definition of $f_{\rm esc}$ includes any mechanism that causes a given ionizing photon not to ionize gas within the H II region. While this could mean that the photon physically escapes the H II region, it could also mean that the photon was absorbed by a dust grain within the H II region; our analysis is not capable of distinguishing between these two possibilities. Second, there is nothing in equation (2) that enforces $f_{\rm esc} \geq 0$. Negative values of $f_{\rm esc}$ correspond to estimates of the ionizing luminosity $Q(H^0)$ small enough that there would be insufficient photons to power the observed Hα emission even for $f_{\rm esc} = 0$. Given the rather broad PDFs of $Q(H^0)$ that we obtain from photometry, it is inevitable that in some cases parts of the PDF will extend into the $f_{\rm esc} < 0$ region. We could avoid this by imposing as a prior that $f_{\rm esc} \geq 0$, but doing so would artificially suppress the width of our confidence intervals, and might conceal problems in our analysis. For this reason, we choose to adopt a purely flat prior of $f_{\rm esc}$, allowing our analysis to produce $f_{\rm esc} < 0$ even though we know such values are unphysical. However, for completeness we also repeat our
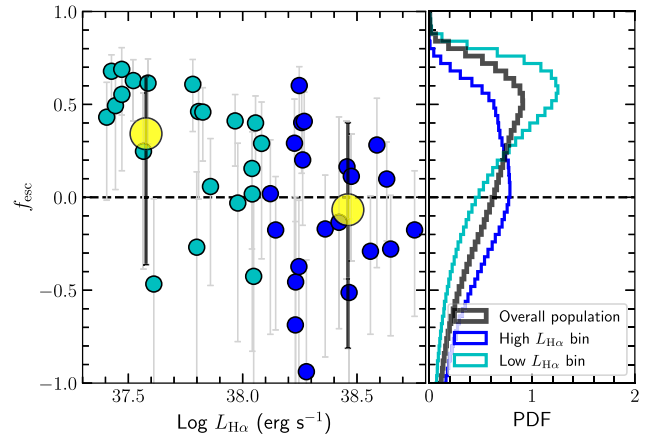
Figure 8. The LyC escape fraction, $f_{\rm esc}$, of individual H II regions after truncating the sample (see Fig. 7), using the technique describeed in Section 4.1.1. In the left-hand panel, circles correspond to the median of the PDF of $f_{\rm esc}$, coloured in cyan and blue to indicate low and high $L_{\rm H\alpha}$ bins, with grey errorbars indicating the 68 per cent confidence interval. The dashed line indicates $f_{\rm esc} = 0$. Negative values of $f_{\rm esc}$ correspond to estimates of the ionizing luminosity $Q(H^0)$ small enough that there would be insufficient photons to power the observed Hα emission even for $f_{\rm esc} = 0$. For the histogram plot, the combined PDF is shown for each bin, along with an overall PDF of $f_{\rm esc}$ (in black) across the truncated sample. Yellow data points illustrate median values of $f_{\rm esc}$ for low- and high-$L_{\rm H\alpha}$ bins, where $f_{\rm esc} = 0.34^{+0.31}_{-0.70}$ and $-0.07^{+0.47}_{-0.75}$, respectively.

analysis with an $f_{\rm esc} \geq 0$ prior in Appendix A, and show there that our qualitative conclusions are similar to those derived in the main paper.

In order to study the possible relationship between $f_{\rm esc}$ and $L_{\rm H\alpha}$, we divide the samples into low and high $L_{\rm H\alpha}$ bins, defined by the intervals $37.41 \leq \log L_{\rm H\alpha,low}({\rm erg\,s^{-1}}) \leq 38.06$ and $38.06 \leq \log L_{\rm H\alpha,high} \leq 38.75$, respectively. These bins are chosen so that each contains roughly half of the sample. We present both the distribution of $f_{\rm esc}$ for individual H II regions and the combined PDF of $f_{\rm esc}$ for each luminosity bin in Fig. 8. Our analysis yields an overall escape fraction of $f_{\rm esc} = 0.13^{+0.43}_{-0.76}$, with $f_{\rm esc} = 0.34^{+0.31}_{-0.70}$ and $-0.07^{+0.47}_{-0.75}$ for low- and high-$L_{\rm H\alpha}$ bins respectively; the figures we quote here are the 50th percentile values, with error estimates corresponding to the 68 per cent confidence range.

4.2 Population analysis

The preceding section attempts to derive an escape fraction separately for each H II region, and then constructs a mean escape fraction for the entire population of H II region simply by summing those. However, given the rather broad PDFs of ionizing luminosity that our photometric analysis returns for each individual cluster, it may be more reliable to attempt to constrain the escape fraction of the population as a whole directly. To this end, let us assume that there exists an overall escape fraction $f_{\rm esc}$ across the H II region population; our goal is to determine the probability density function of this overall $f_{\rm esc}$ given the observed $L_{\rm H\alpha}$ and photometry.

Our basic approach comes from Bayes's theorem: the posterior probability density for the escape fraction, $f_{\rm esc}$, is proportional to the product of the prior probability and the likelihood function, which gives the probability of the data given the model. To determine the likelihood, first consider the simplest case of a single H II region with observed Hα luminosity that is associated with a group of clusters, characterized by a set of photometric measurements $\{F_{\rm phot}\}$.

From our `cluster_slug` analysis, for each H II region we can compute a PDF of ionizing luminosity $p_{\mathrm{SLUG}}\left(Q(H^0)\,|\{F_{\mathrm{phot}}\}\right)$ for the associated star clusters from the photometry exactly as in Section 4.1. If the true escape fraction is f_{esc}, then the corresponding PDF of the expected Hα luminosity has the same shape, with each ionizing luminosity $Q(H^0)$ simply mapped to a corresponding Hα luminosity $L_{\mathrm{H}\alpha} = (1 - f_{\mathrm{esc}})\alpha_{\mathrm{H}\alpha}\,Q(H^0)$. Mathematically, this can be written as

$$p\left(L_{\mathrm{H}\alpha}\mid f_{\mathrm{esc}}, \{F_{\mathrm{phot}}\}\right) \propto p_{\mathrm{SLUG}}\left(Q(H^0)\,|\{F_{\mathrm{phot}}\}\right)\big|_{Q(H^0)=\frac{Q(H\alpha)}{1-f_{\mathrm{esc}}}}, \quad (3)$$

where for convenience we have defined $Q(H\alpha) = L_{\mathrm{H}\alpha}\,/\alpha_{\mathrm{H}\alpha}$ as the ionizing luminosity required to drive a given Hα luminosity in the limit of zero escape fraction. This gives the likelihood function for a single H II region. The generalization to a population of H II regions is straightforward, since each one is independent: given a set of N measured H II region Hα luminosities $L_{\mathrm{H}\alpha,i}$, where $i = 1 \ldots N$, and a corresponding set of photometric measurements $\{F_{\mathrm{phot}}\}_i$ for the clusters associated with those H II regions, the likelihood function is

$$p\left(\{L_{\mathrm{H}\alpha}\}\mid f_{\mathrm{esc}}, \{\{F_{\mathrm{phot}}\}\}\right)$$
$$\propto \prod_{i=1}^{N} p_{\mathrm{SLUG}}\left(L_{\mathrm{H}\alpha,i}\,|\{F_{\mathrm{phot}}\}_i\right)\big|_{Q(H^0)=\frac{Q(H\alpha)_i}{1-f_{\mathrm{esc}}}}, \quad (4)$$

where $\{L_{\mathrm{H}\alpha}\}$ is the set of all observed Hα luminosities and $\{\{F_{\mathrm{phot}}\}\}$ is the set of all observed photometric magnitudes for all clusters in all H II regions; note that the double curly braces on $\{\{F_{\mathrm{phot}}\}\}$ indicate a nested list — all clusters observed in all filters.

For simplicity, we adopt a flat prior on f_{esc}, which means that the posterior PDF for the escape fraction is simply proportional to the likelihood function. Thus our final expression for the posterior PDF of the escape fraction is

$$p\left(f_{\mathrm{esc}}\mid \{L_{\mathrm{H}\alpha}\}, \{\{F_{\mathrm{phot}}\}\}\right)$$
$$\propto \prod_{i=1}^{N} p_{\mathrm{SLUG}}\left(L_{\mathrm{H}\alpha,i}\,|\{F_{\mathrm{phot}}\}_i\right)\big|_{Q(H^0)=\frac{Q(H\alpha)_i}{1-f_{\mathrm{esc}}}}, \quad (5)$$

We can evaluate this expression either for the entire population of H II regions, or for subsets thereof binned by $L_{\mathrm{H}\alpha}$ or in any other way. We plot our posterior PDFs of f_{esc} for both the full sample and for the low- and high-$L_{\mathrm{H}\alpha}$ sub-samples in Fig. 9. We find an overall escape fraction of $f_{\mathrm{esc}} = 0.09^{+0.06}_{-0.06}$ for the full sample, with $f_{\mathrm{esc}} = 0.56^{+0.08}_{-0.14}$ and $0.06^{+0.06}_{-0.06}$ for the low- and high-$L_{\mathrm{H}\alpha}$ bins respectively; as in Section 4.1, the numbers we quote are the 50th percentile value with error bars giving the 68 per cent confidence range. We further discuss the implication of this result in Section 5.3.

5 DISCUSSION

5.1 Improving the number of observed populated H II regions

In Section 3, we show that the orphan H II regions are mainly a consequence of H II regions whose associated star clusters are either less luminous than the LEGUS visual inspection threshold ($M_V < -6$), or that they were inspected and classified as non-clusters (Class 4). Before moving on, it is thus necessary to provide a resolution to this problem if we wish to harness the full potential of oncoming LEGUS-SIGNALS observations. What can we do to improve the number of observed populated H II regions?

There are two possible ways around this issue. The first of these is to allow visual inspection of fainter ($M_V \geq -6$) LEGUS sources, which in turn can be approached both via observational and theoretical grounds. On the observational side, we begin by analysing the location of LEGUS sources that do not meet the magnitude
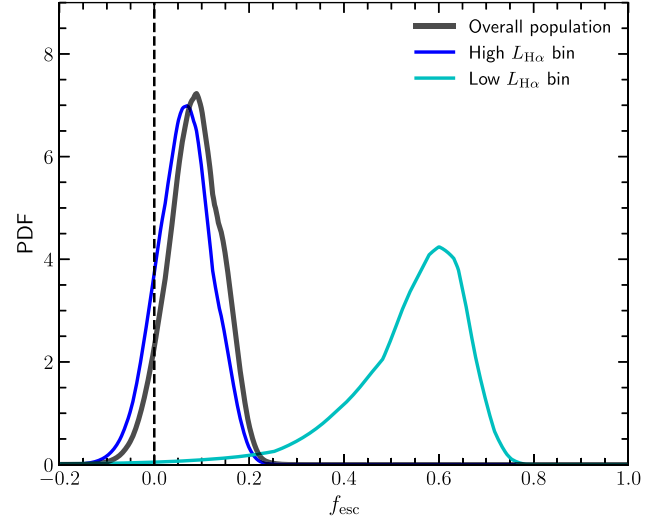

Figure 9. Posterior PDF of f_{esc} derived using the technique described in Section 4.2. Histograms coloured in cyan and blue corresponds to low- and high-$L_{\mathrm{H}\alpha}$ bins. The dashed line indicates $f_{\mathrm{esc}} = 0$. The PDF for the whole population is shown in black. Similar to Fig. 8, we find that the overall escape fraction across the population of low-$L_{\mathrm{H}\alpha}$ H II regions ($f_{\mathrm{esc}} = 0.56^{+0.08}_{-0.14}$) is higher than that of the population of high-$L_{\mathrm{H}\alpha}$ regions ($f_{\mathrm{esc}} = 0.06^{+0.06}_{-0.06}$).

threshold required for visual inspection (Class 0; see Grasha et al. 2015). We overlay these sources on to a map of H II regions that do not contain any of Class 1, 2, 3, or 4 LEGUS sources, and find that 13 out of 129 (~ 10 per cent) of these H II regions contain Class 0 sources. While we note that Class 0 sources may include spurious detections, this never the less suggests that they may contain true clusters.

On the other hand, we can make theoretical predictions of how inspecting fainter clusters would decrease the number of orphan H II regions by repeating the completeness analysis in Section 3.5, with a simple change in the cluster selection criteria: instead of using the LEGUS magnitude cut, $M_V = -6$, we experiment with magnitude cuts ranging from $M_V = -12$ to $M_V = -3$. For each M_V, we calculate the percentage of orphan H II region in the synthesized star clusters and show the result in Fig. 10. We find that changing the magnitude cut from $M_V = -6$ to $M_V = -5$ will result in a decrease in the percentage of orphan H II regions (from 48 per cent to 28 per cent) in the synthesized star clusters. Further allowing clusters that are fainter than $M_V = -4$ seems to have no effect on the orphan H II regions: we suspect that these star clusters are likely unclassified due to the completeness limits in other photometric filters, as $M_V = -6$ is not the only cluster selection criterion. Cluster candidates require detection in at least two filters (the V band and either B or I band), and must be detected in four bands total, implying that the cluster must be detected either in the U or the UV band. This means that clusters with low mass and moderate extinction could, in theory, be detectable in the B, V, and I bands, but not in the bluer bands. Therefore, even if one inspects fainter clusters, some clusters will be excluded from the LEGUS catalogue simply because they will not be detected at all.

Finally, it is worth noting that while varying the magnitude cut seems reasonable on paper, it is impractical. This is because the number of potential clusters that have to go through human visual classification is significantly higher, since the cluster luminosity function of NGC 628 scales as $dN/dL \approx L^{-\beta}$, where $\beta \approx 2$ (see Adamo et al. 2017). Thus, the time-consuming nature of this labour-
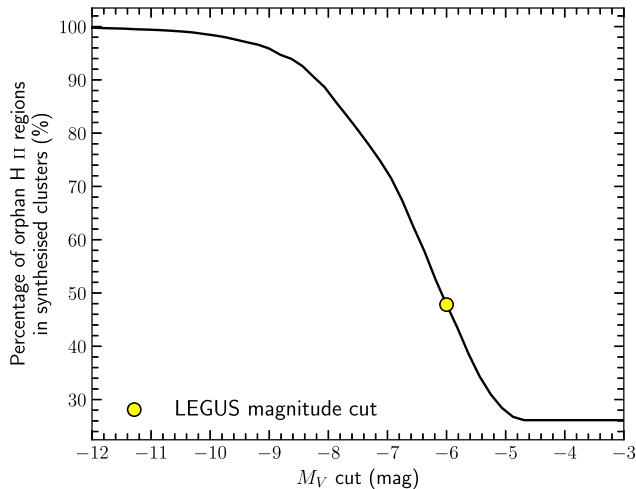
Figure 10. The percentage of orphan H II regions as a function of V-band absolute magnitude cut in the synthesized YSCs using SLUG. Yellow dot indicates the visual inspection magnitude cut in LEGUS. The predicted number of orphan H II regions decreases as M_V cut increases; this is because we are allowing visual inspection on fainter clusters.cut increases; this is because we are allowing visual inspection on fainter clusters. 8−13.6 50−100

intensive task limits the choice of the magnitude cut.[3] Instead, machine learning techniques could be implemented in the LEGUS pipeline to ensure reproducibility of cluster classifications and a significantly higher classification speed (e.g. see Grasha et al. 2019; Wei et al. 2020; Pérez et al. 2021; Linden et al. 2022).

This brings us to the second method. As demonstrated in Section 3.4, we find, via statistical and visual techniques, that a fraction of Class 4 sources are likely true clusters after we combine and inspect spatial information of sources in the LEGUS catalogue with an Hα map from SIGNALS, and look for areas of coincidence. An important corollary of this result is that we can further improve the accuracy of visually identified clusters (especially clusters < 5 Myr) in future surveys by overlaying LEGUS sources on a resolved Hα image.

5.2 Improving measurement of f_{esc}: ways forward

The paper thus far shows that cluster_slug is a powerful tool to infer stellar properties (e.g. age, mass, $Q(H^0)$) from five-band photometry alone, but also suggests some weaknesses and possible ways forward. The first is the lack of far-ultraviolet (FUV; $\approx 8 - 13.6$ eV) data: FUV surveys of stellar populations provide much stronger constraints on the properties of young stellar populations than observations at longer wavelengths alone (e.g. see Knigge et al. 2008; Wofford, Leitherer & Chandar 2011; Smith et al. 2016; Wofford et al. 2016; Sirressi et al. 2022). In our case, as discussed in Section 4.1.1, the dominant contributor to our uncertainties on ionizing luminosities is the fact that all massive stars have similar colours in the optical to NUV bands, meaning that our data offer limited constraints for clusters whose light is dominated by one or a few such stars. The addition of FUV observations would reduce this issue, greatly improving our ionizing flux estimates and

---

[3]However, for those interested to study only clusters that are located in H II regions, it is still worthwhile to inspect all detected sources regardless of their magnitude to allow for more complete coverage of cluster candidates.

therefore our f_{esc} measurements. Existing FUV data from *GALEX* (Galaxy Evolution Explorer; Martin et al. 2005) are not suitable for this purpose due to their limited spatial resolution, which makes it essentially impossible to assign *GALEX*-detected FUV emission uniquely to particular LEGUS clusters. Instead, FUV data with HST-like resolution are required. One possible future source for such data is the *UVEX* survey (Ultraviolet Explorer; Kulkarni et al. 2021), which is designed to perform a multicadence synoptic all-sky survey $50 - 100$ times deeper and at significantly higher resolution than *GALEX* in the FUV/NUV.

A second source of uncertainty in our analysis the loss of LyC photons to dust absorption, and this too could be improved with future data. The inclusion of high-spatial-resolution infrared observations such as from the PHANGS–*JWST* programme (Lee et al. 2022a) and the incoming *JWST*–FEAST programme (Adamo et al. 2021), among several others, will allow us to trace dust emissions around clusters. On the other hand, comparison between our observations and stellar feedback codes such as WARPFIELD (Rahner et al. 2017, 2019), which allows detailed study of clusters and their feedback effect on surrounding natal clouds, will also provide useful insight on the dust content. This in turn will allow better estimates of the fraction of LyC photons that are absorbed by dust grains.

5.3 Linking between f_{esc} and local environments

5.3.1 Notes on the overall f_{esc} and f_{DIG}

In Section 4.2, we report an overall LyC escape fraction of $f_{esc} = 0.09^{+0.06}_{-0.06}$ across our population of H II regions. The escape fraction is lower than what is usually reported for H II regions in nearby dwarf galaxies ($0 \lesssim f_{esc} \lesssim 0.51$; McLeod et al. 2019, 2020; Choi et al. 2020) and spiral galaxies ($f_{esc} \sim 0.67$; Della Bruna et al. 2021, 2022b). We note that these values of f_{esc} should not be compared at face value as they represent a wide range of galactic environments, and that they were evaluated using different methods; however, it is still necessary to address any biases present in our analysis, and their potential effect on f_{esc}. First, in this study we discard H II regions that are deemed diffuse and transient regions by the SIGNALS survey. While this is important for the analysis since we want to minimize contamination from highly dispersed regions that are likely DIG-dominated (Rousseau-Nepton et al. 2018, 2019), it implies that the majority of our regions may be the ones that are compact and ionization-bounded. Thus, one may expect that the overall LyC escape fraction is much closer to zero. There is also a secondary factor, which is that the selected H II regions are limited by the LEGUS FoV: we are essentially observing the inner half of the galaxy disc, thus more towards regions of higher metallicity, extinction and dust-gas mass ratio. This may also point towards a higher likelihood of having a lower overall f_{esc}.

It is also illuminating to compare f_{esc} and the fraction of Hα emission in the DIG (f_{DIG}). Estimates of f_{DIG} in NGC 628 range from ≈ 0.25 to 0.5 (see Crocker et al. 2013; Chevance et al. 2020; Belfiore et al. 2022), much larger than the overall f_{esc} determined in this study. However, given the systematic uncertainties in previous studies we do not conclude that this is inconsistent with the idea that the DIG is powered by leaking H II regions; rather, we posit the possibility that the DIG includes a subdominant contribution from harder ionizing photons from old stars (similar to Belfiore et al. 2022). The reason is that we have only taken into account clusters associated with H II regions, which tend to be very young (< 5 Myr). We do not consider older clusters that are not associated anymore

with compact H II regions, but nevertheless these can still contribute to the DIG.

5.3.2 f_{esc} and the properties of H II regions and clusters

We have found above that f_{esc} varies systematically with Hα luminosity (Fig. 9), with less luminous H II regions having systematically higher f_{esc}. Interestingly, this result differs from the one obtained by Pellegrini et al. (2012) for the Magellanic Clouds; they found that H II regions with higher $L_{H\alpha}$ are leakier. This is not too surprising, given that NGC 628 differs from the Magellanic Clouds in many ways (e.g. morphology, star-formation rate, and metallicity); as such, our result may represent changing ionization conditions in different galactic environments (e.g. see Rahner et al. 2017, for the effects of local environment on f_{esc}). The introduction of both LEGUS and SIGNALS surveys is therefore timely and will offer a crucial opportunity to compare observations over a wide range of galactic environments.

It is also interesting to search for systematic correlations between f_{esc} and other H II region or star cluster properties. For example, we might envision that older clusters, which have had more time to clear their environments, might have systematically leakier H II regions. We might also expect that some of the more luminous regions are less leaky due to a higher proportion of the LyC photons being absorbed by dust (e.g. see Petrosian, Silk & Field 1972; Mathis 1986; Arthur et al. 2004), or that the efficiency of LyC escape might be influenced by metallicity (e.g. see Kimm et al. 2019), which in turns varies with galactocentric radius (e.g. see Shaver et al. 1983; Rolleston et al. 2000; Kewley et al. 2010; Sánchez-Blázquez et al. 2011; Pilkington et al. 2012; Rodríguez-Baras et al. 2018). To investigate all these possibilities, in Fig. 11 we plot f_{esc} as a function of cluster age, cluster mass, extinction, and galactocentric radius.[4] The figure reveals no clear systematic correlations between f_{esc} and any of these properties; though of course, we cannot rule out the presence of such a correlation, especially because our values of both f_{esc} and cluster properties have very large uncertainties at the level of individual clusters. None the less, we find no evidence for f_{esc} correlating with any cluster/H II regions properties other than $L_{H\alpha}$.

5.3.3 f_{esc} and emission line ratios

In Fig. 12, we plot f_{esc} versus these emission line ratios. To quantify possible correlations, we show the Spearman's rank correlation coefficient ρ and the corresponding p-value in Table 3. We find that both [O II]/[N II] and [O II]/Hβ correlate with f_{esc} at high statistical significance ($p < 0.0055$).[5] The physical origin of the

---

[4]For cluster age and mass, we face two complications in assigning single numbers to these quantities: first, SLUG returns PDFs for these quantities, not single numbers. Second, many H II regions are associated with multiple clusters. For this purpose, we first take the age and mass of each cluster to be the 50th percentile value returned by SLUG. We then simply sum the 50th percentile masses to define a total cluster mass, while for age we adopt a 'young' age estimate set equal to the 25th percentile of the ages of the clusters associated with a given H II region. We bias the age estimate young in this way because under the hypothesis that low f_{esc} regions are caused by gas clearing, what matters are the clusters that have had the least time to clear their gas.

[5]Note that, since we are testing nine candidate line ratios for correlation with f_{esc}, Bonferroni correction implies that in order to conclude at 95 per cent confidence that *any* of the line ratios are correlated with f_{esc}, we must obtain a p-value $p < 0.05/9 = 0.0055$ for any particular line ratio; both [O II]/[N II] and [O II]/Hβ reach this significance level.
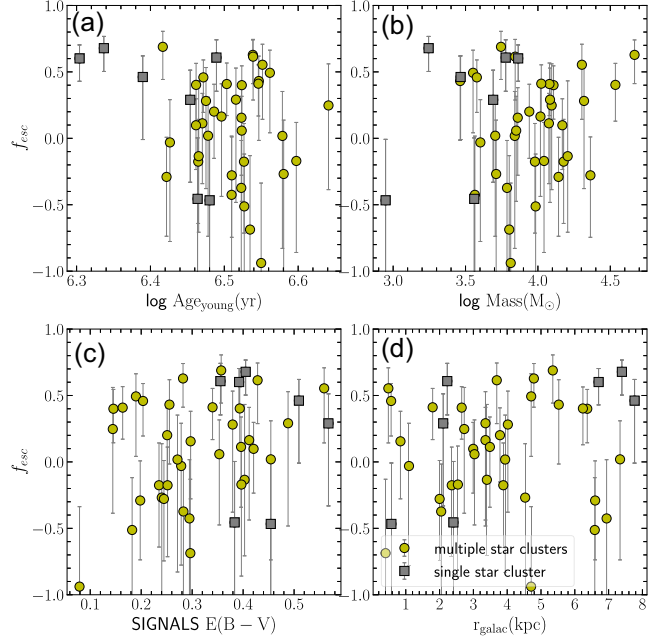

Figure 11. Properties of H II regions and their associated star clusters versus the LyC escape fraction, f_{esc}. We plot H II regions found to be associated with single clusters as grey squares, and those associated with multiple clusters as yellow circles. The errorbars represent 1-σ uncertainties on f_{esc}; we retain distributions of f_{esc} that extend into the $f_{esc} < 0$ region for reasons discussed in Section 4.1.2. The top two panels show respectively the median age across clusters in the younger age bin (age$_{young}$; see Section 5.3.2) and the total mass of star clusters in their corresponding H II region. The two bottom panels show the extinction value, $E(B-V)$, and the galactocentric radius of the H II regions, obtained from the SIGNALS catalogue. We do not find clear evidence of correlation between any of the aforementioned properties and f_{esc}.

correlation is unclear, though we note the possibility that it could relate to either metal abundance or ionization parameter (e.g. see Kreckel et al. 2019; Grasha et al. 2022; Scheuermann et al. 2023). Further progress on this question likely requires improved stellar models in order to obtain better constraints on cluster properties (e.g. Grasha et al. 2021), coupled with theoretical studies and forward-modelling to predict observable correlations between LyC escape and other line diagnostics. For instance, Pellegrini et al. (2020) used a combination of stellar feedback code WARPFIELD (Rahner et al. 2017, 2019), H II region code CLOUDY (Ferland et al. 2017), radiative transfer code POLARIS (Reissl, Wolf & Brauer 2016), and SITELLE observations of H II regions in NGC 628 (Rousseau-Nepton et al. 2018) to directly compare between emission lines from feedback models and observations .

6 CONCLUSION

We present in this work a pilot study of the association between spatially resolved H II regions and star clusters in the nearby spiral galaxy NGC 628. We combine 334 H II regions from the SIGNALS survey (Rousseau-Nepton et al. 2018, 2019) and 1253 star clusters from the LEGUS survey (Calzetti et al. 2015; Grasha et al. 2015), which we link based on their spatial overlap. In the combined catalogue, we find 49 per cent of H II regions lack ionizing star clusters.

We begin in Section 3 by analysing six possible explanations for this phenomenon (see Table 2 for an overview). We consider
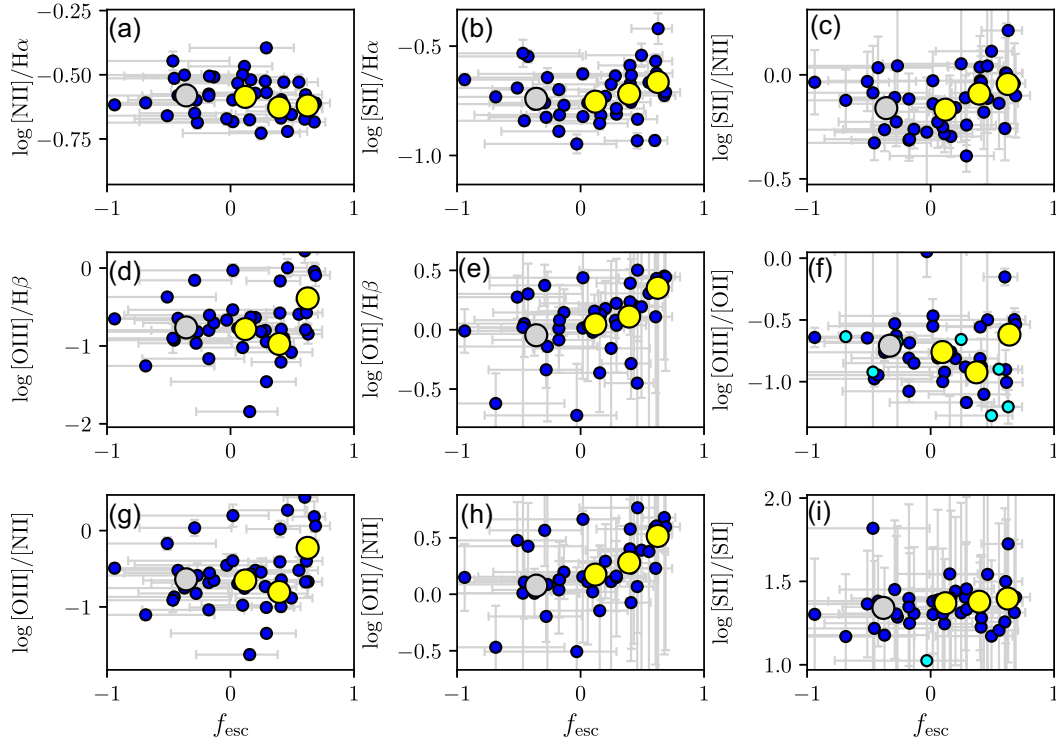
Figure 12. f_{esc} versus extinction-corrected emission line ratios. Blue and cyan points are line ratios with $SNR_{cross} \geq 3$ and $SNR_{cross} < 3$, respectively, where SNR_{cross}, represented by vertical errorbars, is defined as the line ratio's best SNR (see Rousseau-Nepton et al. 2018, for more details). The horizontal errorbars indicate the 68 per cent confidence interval of f_{esc}. Yellow data points represent the mean value obtained by averaging the data in bins of $f_{esc} = 0-0.25$, $0.25-0.5$, and $0.5-0.75$, using the 50th percentile estimate of f_{esc} to assign each H II region to a bin, and excluding data with $SNR_{cross} < 3$. We show the mean and 68 per cent confidence range for each cluster as blue points with horizontal errorbars. In Table 3, we present the detailed description of the line ratios as well as the Spearman's rank correlation coefficient, ρ, and the corresponding p-value.

Table 3. Emission line ratios versus f_{esc} and their corresponding Spearman's rank correlation coefficient ρ and p-value (See Fig. 12).

Emission line ratio	ρ	p-value[a]	Emission line ratio	ρ	p-value
a) [N II]λ6583/ Hα	-0.26	0.0972	f) [O III]λ5007/[O II]λ3727	-0.02	0.9302
b) [S II]λ6716 + 6731/ Hα	0.14	0.3651	g) [O III]λ5007/[N II]λ6583	0.18	0.2571
c) [S II]λ6716 + 6731/[N II]λ6583	0.27	0.0818	h) [O II]λ3727/[N II]λ6583	0.52	0.0004
d) [O III]λ5007/Hβ	0.17	0.2862	i) [S II]λ6716/[S II]λ6731	0.14	0.3717
e) [O II]λ3727/Hβ	0.48	0.0014			

Note. [a]The p-value is quoted for a hypothesis test whose null hypothesis is that two sets of data are uncorrelated. The [O II]λ3727/[N II]λ6583 and [O II]λ3727/Hβ lines show significant correlation (i.e. $p < 0.005$).

and reject explanations that the orphan H II regions are caused by errors in the astrometry, runaway stars, OB associations, or errors in how SIGNALS segments H II regions. Ultimately, we find that the dominant contributor to the orphan H II regions is the incompleteness of the LEGUS catalogue, with the major component of this being clusters that are able to create H II regions detectable by SIGNALS but are too faint to be classified by LEGUS. A subdominant contributor is that the LEGUS visual classification mistakenly identifies some true clusters as Class 4 (non-clusters).

In the second part of this work (Section 4), we attempt to infer the escape fraction of ionizing photons from H II regions by determining the ionizing luminosity of star clusters, $Q(H^0)$ from a combination of photometric data and stochastic stellar population model (SLUG;

da Silva et al. 2012, 2014; Krumholz et al. 2015a), and comparing this to the ionizing luminosity required to power the matched H II regions (see Fig. 7). Overall, we find $f_{esc} = 0.09^{+0.06}_{-0.06}$ across our population, with $f_{esc} = 0.56^{+0.08}_{-0.14}$ and $0.06^{+0.06}_{-0.06}$ for low- and high-$L_{H\alpha}$ bins, respectively (see Fig. 9). We also find that the escape fraction correlates at high statistical significance ($p < 0.0055$) with metallicity tracers, such as [O II]/[N II] and [O II]/Hβ.

Overall, we have demonstrated the use of SLUG as a powerful tool to study the demographics of star cluster populations at extragalactic distances with consideration of the effect of stochastic sampling of the stellar IMF. Combining with the LEGUS and SIGNALS surveys, this allows us to better understand the interaction between stars and their surrounding interstellar medium with high accuracy where most conventional methods cannot.

[6]https://www.astropy.org
[7]https://ui.adsabs.harvard.edu/

DATA AVAILABILITY

The data and the source code underlying this article will be available at https://github.com/JiaWeiTeh/LEGUS-SIGNALS-NGC628. The cluster catalog for NGC 628 underlying this article are publicly available online at MAST https://archive.stsci.edu/prepds/legus/dataproducts-public.html.

REFERENCES

Adamo A. et al., 2017, ApJ, 841, 131
Adamo A. et al., 2021, JWST probes Feedback in Emerging extrAgalactic Star clusTers: JWST-FEAST. JWST Proposal. Cycle 1, ID. #1783, Astrophysics Data System, Cambridge
Adamo A., Zackrisson E., Östlin G., Hayes M., 2010, ApJ, 725, 1620
Anand G. S. et al., 2021, MNRAS, 501, 3621
Anders F., Cantat-Gaudin T., Quadrino-Lodoso I., Gieles M., Jordi C., Castro-Ginard A., Balaguer-Núñez L., 2021, A&A, 645, L2
Arthur S. J., Kurtz S. E., Franco J., Albarrán M. Y., 2004, ApJ, 608, 282
Ashworth G. et al., 2017, MNRAS, 469, 2464
Astropy Collaboration et al., 2013, A&A, 558, A33
Astropy Collaboration et al., 2018, AJ, 156, 123
Astropy Collaboration et al., 2022, ApJ, 935, 167
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Bally J., 1981, PhD thesis, Univ. Massachusetts Amherst
Barnes A. T. et al., 2022, A&A, 662, L6
Bastian N., Gieles M., Efremov Y. N., Lamers H. J. G. L. M., 2005, A&A, 443, 79
Baumgardt H., Parmentier G., Anders P., Grebel E. K., 2013, MNRAS, 430, 676
Belfiore F. et al., 2022, A&A, 659, A26
Berg D. A., Skillman E. D., Croxall K. V., Pogge R. W., Moustakas J., Johnson-Groh M., 2015, ApJ, 806, 16
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Borissova J., Kurtev R., Georgiev L., Rosado M., 2004, A&A, 413, 889
Bromm V., Yoshida N., Hernquist L., 2003, ApJ, 596, L135
Brown G., Gnedin O. Y., 2021, MNRAS, 508, 5935
Calzetti D. et al., 2015, AJ, 149, 51
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Chevance M. et al., 2020, MNRAS, 493, 2872
Chevance M. et al., 2022, MNRAS, 509, 272
Chisholm J. et al., 2022, MNRAS, 517, 5104
Choi Y. et al., 2020, ApJ, 902, 54
Chu Y.-H., Kennicutt Robert C. J., 1988, AJ, 96, 1874
Crocker A. F. et al., 2013, ApJ, 762, 79
da Silva R. L., Fumagalli M., Krumholz M. R., 2014, MNRAS, 444, 3275
da Silva R. L., Fumagalli M., Krumholz M., 2012, ApJ, 745, 145
Dale J. E., Ercolano B., Bonnell I. A., 2013, MNRAS, 430, 234
de Mink S. E., Brott I., Cantiello M., Izzard R. G., Langer N., Sana H., 2012, in Drissen L., Robert C., St-Louis N., Moffat A. F. J., eds, ASP Conf. Ser. Vol. 465, Proc. Scientific Meeting in Honor of Anthony F. J. Moffat. Astron. Soc. Pac., San Francisco, p. 65
de Wit W. J., Testi L., Palla F., Zinnecker H., 2005, A&A, 437, 247
Della Bruna L. et al., 2020, A&A, 635, A134
Della Bruna L. et al., 2021, A&A, 650, A103
Della Bruna L. et al., 2022a, A&A, 660, A77
Della Bruna L. et al., 2022b, A&A, 666, A29
Doran E. I. et al., 2013, A&A, 558, A134
Drew J. E., Herrero A., Mohr-Smith M., Monguió M., Wright N. J., Kupfer T., Napiwotzki R., 2018, MNRAS, 480, 2109
Ekström S. et al., 2012, A&A, 537, A146
Elmegreen B. G., Elmegreen D. M., Chandar R., Whitmore B., Regan M., 2006, ApJ, 644, 879
Emsellem E. et al., 2022, A&A, 659, A191

Fedotov K., Gallagher S. C., Konstantopoulos I. S., Chandar R., Bastian N., Charlton J. C., Whitmore B., Trancho G., 2011, AJ, 142, 42

Ferland G. J. et al., 2017, RMxAA, 53, 385

Flury S. R. et al., 2022a, ApJS, 260, 1

Flury S. R. et al., 2022b, ApJ, 930, 126

Fouesneau M., Lançon A., Chandar R., Whitmore B. C., 2012, ApJ, 750, 60

Fujii M. S., Portegies Zwart S., 2016, ApJ, 817, 4

Geist E. et al., 2022, PASP, 134, 064301

Georgy C., Ekström S., Meynet G., Massey P., Levesque E. M., Hirschi R., Eggenberger P., Maeder A., 2012, A&A, 542, A29

Gies D. R., 1987, ApJS, 64, 545

Gonzaga S., Hack W., Fruchter A., Mack J., 2012, The DrizzlePac Handbook. STScI, Baltimore

Grasha K. et al., 2015, ApJ, 815, 93

Grasha K. et al., 2017a, ApJ, 840, 113

Grasha K. et al., 2017b, ApJ, 842, 25

Grasha K. et al., 2018, MNRAS, 481, 1016

Grasha K. et al., 2019, MNRAS, 483, 4707

Grasha K. et al., 2022, ApJ, 929, 118

Grasha K., Roy A., Sutherland R. S., Kewley L. J., 2021, ApJ, 908, 241

Gusev A. S., Egorov O. V., Sakhibov F., 2014, MNRAS, 437, 1337

Gvaramadze V. V., Weidner C., Kroupa P., Pflamm-Altenburg J., 2012, MNRAS, 424, 3037

Halliday C. et al., 2006, MNRAS, 369, 97

Hannon S. et al., 2019, MNRAS, 490, 4648

Hannon S. et al., 2022, MNRAS, 512, 1294

Harris C. R. et al., 2020, Nature, 585, 357

Hollyhead K., Bastian N., Adamo A., Silva-Villa E., Dale J., Ryon J. E., Gazak Z., 2015, MNRAS, 449, 1106

Hoogerwerf R., de Bruijne H. J., de Zeeuw P. T., 2001, A&A, 365, 49

Hoopes C. G., Walterbos R. A. M., 2000, ApJ, 541, 597

Hummer D. G., Storey P. J., 1987, MNRAS, 224, 801

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Izotov Y. I., Worseck G., Schaerer D., Guseva N. G., Thuan T. X., Fricke Verhamme A., Orlitová I., 2018, MNRAS, 478, 4851

Japelj J. et al., 2017, MNRAS, 468, 389

Jaskot A. E., Dowd T., Oey M. S., Scarlata C., McKinney J., 2019, ApJ, 885, 96

Jaskot A. E., Oey M. S., 2013, ApJ, 766, 91

Kennicutt Robert C. J., 1998, ApJ, 498, 541

Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, MNRAS, 372, 961

Kewley L. J., Nicholls D. C., Sutherland R. S., 2019, ARA&A, 57, 511

Kewley L. J., Rupke D., Zahid H. J., Geller M. J., Barton E. J., 2010, ApJ, 721, L48

Kim J. et al., 2021, MNRAS, 504, 487

Kim J. et al., 2022, MNRAS, 516, 3006

Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215

Knigge C., Dieball A., Maíz Apellániz J., Long K. S., Zurek D. R., Shara M. M., 2008, ApJ, 683, 1006

Konstantopoulos I. S., Bastian N., Smith L. J., Westmoquette M. S., Trancho G., Gallagher J. S. I., 2009, ApJ, 701, 1015

Kotulla R., Fritze U., Weilbacher P., Anders P., 2009, MNRAS, 396, 462

Kreckel K. et al., 2018, ApJ, 863, L21

Kreckel K. et al., 2019, ApJ, 887, 80

Kroupa P., 2001, MNRAS, 322, 231

Krumholz M. R. et al., 2015b, ApJ, 812, 147

Krumholz M. R., Fumagalli M., da Silva R. L., Rendahl T., Parra J., 2015a, MNRAS, 452, 1447

Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, ARA&A, 57, 227

Kulkarni S. R. et al., 2021, preprint (arXiv:2111.15608)

Larsen S. S., 2009, A&A, 494, 539

Lasker B. M. et al., 2008, AJ, 136, 735

Lee J. C. et al., 2022a, ApJ, 944, 23

Lee J. C. et al., 2022b, ApJS, 258, 10

Leitherer C. et al., 1999, ApJS, 123, 3

Leitherer C., 2005, in Popescu C. C., Tuffs R. J.eds, AIP Conf. Ser.Vol. 761, The Spectral Energy Distributions of Gas-Rich Galaxies: Confronting Models with Data. American Institute of Physics, College Park, p. 39

Linden S. T. et al., 2022, ApJ, 935, 166

Lucke P. B., Hodge P. W., 1970, AJ, 75, 171

McKee C. F., Williams J. P., 1997, ApJ, 476, 144

McLeod A. F. et al., 2020, ApJ, 891, 25

McLeod A. F., Dale J. E., Evans C. J., Ginsburg A., Kruijssen J. M. D., Pellegrini E. W., Ramsay S. K., Testi L., 2019, MNRAS, 486, 5263

Mackey J., Langer N., Gvaramadze V. V., 2013, MNRAS, 436, 859

Martin D. C. et al., 2005, ApJ, 619, L1

Martin T. B., Drissen L., Melchior A.-L., 2018, MNRAS, 473, 4130

Mathis J. S., 1986, PASP, 98, 995

Mel'Nik A. M., Efremov Y. N., 1995, Astron. Lett., 21, 10

Mengel S., Lehnert M. D., Thatte N. A., Vacca W. D., Whitmore B., Chandar R., 2008, A&A, 489, 1091

Menon S. H. et al., 2021, MNRAS, 507, 5542

Merrett H. R. et al., 2006, MNRAS, 369, 120

Messa M. et al., 2021, ApJ, 909, 121

Mitra S., Ferrara A., Choudhury T. R., 2013, MNRAS, 428, L1

Nakajima K., Ellis R. S., Robertson B. E., Tang M., Stark D. P., 2020, ApJ, 889, 161

Nakajima K., Ouchi M., 2014, MNRAS, 442, 900

Niederhofer F., Hilker M., Bastian N., Ercolano B., 2016, A&A, 592, A47

Oey M. S. et al., 2007, ApJ, 661, 801

Oey M. S., Kennicutt R. C. J., 1997, MNRAS, 291, 827

Oey M. S., King N. L., Parker J. W., 2004, AJ, 127, 1632

Olivares E. F. et al., 2010, ApJ, 715, 833

Paardekooper J. P., Pelupessy F. I., Altay G., Kruip C. J. H., 2011, A&A, 530, A87

Parker R. J., Goodwin S. P., 2007, MNRAS, 380, 1271

Pellegrini E. W., Oey M. S., Winkler P. F., Points S. D., Smith R. C., Jaskot A. E., Zastrow J., 2012, ApJ, 755, 40

Pellegrini E. W., Rahner D., Reissl S., Glover S. C. O., Klessen R. S., Rousseau-Nepton L., Herrera-Camus R., 2020, MNRAS, 496, 339

Pérez G., Messa M., Calzetti D., Maji S., Jung D. E., Adamo A., Sirressi M., 2021, ApJ, 907, 100

Petrosian V., Silk J., Field G. B., 1972, ApJ, 177, L69

Pflamm-Altenburg J., Kroupa P., 2010, MNRAS, 404, 1564

Pfuhl O. et al., 2011, ApJ, 741, 108

Pilkington K. et al., 2012, A&A, 540, A56

Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, ARA&A, 48, 431

Rahner D., Pellegrini E. W., Glover S. C. O., Klessen R. S., 2017, MNRAS, 470, 4453

Rahner D., Pellegrini E. W., Glover S. C. O., Klessen R. S., 2019, MNRAS, 483, 2547

Ramambason L., Schaerer D., Stasińska G., Izotov Y. I., Guseva N. G., Vílchez J. M., Amorín R., Morisset C., 2020, A&A, 644, A21

Reissl S., Wolf S., Brauer R., 2016, A&A, 593, A87

Renzo M. et al., 2019, A&A, 624, A66

Rhea C. L. et al., 2021b, ApJ, 923, 169

Rhea C., Rousseau-Nepton L., Prunet S., Hlavacek-Larrondo J., Fabbro S., 2020, ApJ, 901, 152

Rhea C., Rousseau-Nepton L., Prunet S., Prasow-Émond M., Hlavacek-Larrondo J., Asari N. V., Grasha K., Perreault-Levasseur L., 2021a, ApJ, 910, 129

Rodríguez-Baras M., Díaz A. I., Rosales-Ortega F. F., Sánchez S. F., 2018, A&A, 609, A102

Rolleston W. R. J., Smartt S. J., Dufton P. L., Ryans R. S. I., 2000, A&A, 363, 537

Rousseau-Nepton L. et al., 2019, MNRAS, 489, 5530

Rousseau-Nepton L., Robert C., Martin R. P., Drissen L., Martin T., 2018, MNRAS, 477, 4152

Ryon J. E. et al., 2017, ApJ, 841, 92

Saldana-Lopez A. et al., 2022, A&A, 663, A59

Sánchez-Blázquez P., Ocvirk P., Gibson B. K., Pérez I., Peletier R. F., 2011, MNRAS, 415, 709

Scheuermann F. et al., 2023, MNRAS, 522, 2369

Seon K.-I., 2009, ApJ, 703, 1159

Shabani F. et al., 2018, MNRAS, 478, 3590

Shaver P. A., McGee R. X., Newton L. M., Danks A. C., Pottasch S. R., 1983, MNRAS, 204, 53

Simón-Díaz S., García-Rojas J., Esteban C., Stasińska G., López-Sánchez A. R., Morisset C., 2011, A&A, 530, A57

Sirressi M. et al., 2022, MNRAS, 510, 4819

Smith L. J., Crowther P. A., Calzetti D., Sidoli F., 2016, ApJ, 823, 38

Sukhbold T., Ertl T., Woosley S. E., Brown J. M., Janka H. T., 2016, ApJ, 821, 38

Thilker D. A. et al., 2022, MNRAS, 509, 4094

Vázquez G. A., Leitherer C., 2005, ApJ, 621, 695

Vílchez J. M., Relaño M., Kennicutt R., De Looze I., Mollá M., Galametz M., 2019, MNRAS, 483, 4968

Wei W. et al., 2020, MNRAS, 493, 3178

Weilbacher P. M. et al., 2018, A&A, 611, A95

Whitmore B. C. et al., 2011, ApJ, 729, 78

Whitmore B. C. et al., 2020, ApJ, 889, 154

Whitmore B. C. et al., 2021, MNRAS, 506, 5294

Whitmore B. C., Chandar R., Bowers A. S., Larsen S., Lindsay K., Ansari A., Evans J., 2014, AJ, 147, 78

Whitmore B. C., Zhang Q., Leitherer C., Fall S. M., Schweizer F., Miller B. W., 1999, AJ, 118, 1551

Wofford A. et al., 2016, MNRAS, 457, 4296

Wofford A., Leitherer C., Chandar R., 2011, ApJ, 727, 100

Zackrisson E., Rydberg C.-E., Schaerer D., Östlin G., Tuli M., 2011, ApJ, 740, 13

Zhang Q. Fall S. M., Whitmore B. C., 2001, ApJ, 561, 727

Zurita A., Beckman J. E., Rozas M., Ryder S., 2002, A&A, 386, 801

APPENDIX A: THE EFFECT OF USING ALTERNATIVE CONDITIONS ON f_{ESC}

In the main text, we make three methodological choices that warrant further exploration. These are the choice of Geneva tracks for stellar evolution, the inclusion of Class 4 objects as potential ionizing sources, and not imposing a prior that f_{esc} be non-negative. In this appendix, we explore the consequences of making different choices. To this end, we repeat the analysis presented in Sections 4.1 and 4.2 with the following combinations of alternative choices: (1) Geneva tracks with no prior on f_{esc}, but only using Class 1–3 sources, (2) Padova tracks with no prior on f_{esc} and using Class 1–4 sources, (3) Padova tracks with no prior on f_{esc} and using only Class 1–3 sources, and (4) Geneva tracks using Class 1–4 sources, but adopting a prior that $f_{esc} \geq 0$.

Figure A1. Same as Fig. 9, except here we impose that $f_{esc} \geq 0$ as a complementary result to that of Section 4.1.2. We observe similar qualitative behaviour in both results.

We report the marginal posterior PDFs for f_{esc} produced by each of these alternative approaches in Table A1. If we focus on the population-level analysis, which is more reliable and has smaller uncertainties, we find that the only methodological choice that produces a difference in the outcome that is larger than the uncertainties is the use of Padova rather than Geneva tracks. We prefer the Geneva tracks in our analysis because doing so produces more data points with smaller uncertainty compared to Padova tracks, which is not surprising as the former are more carefully tuned to model observations of young and massive stars. All other methodological choices produce shifts that are smaller than the uncertainties. For example, if we exclude Class 4 sources, escape fractions shift to somewhat lower values, reaching median $f_{esc} < 0$ in the high luminosity group, which is certainly suggestive that Class 4 sources contain true clusters that contribute a significant amount of ionizing radiation. However, the uncertainty intervals we derive with and without Class 4 none the less slightly overlap.

It is perhaps more surprising that imposing a prior $f_{esc} \geq 0$ does not have larger effects. While there certainly is a modest increase in

Table A1. Alternative model parameters and their effect on our analysis. Column (1) indicates which stellar evolution tracks we use for the SLUG models, column (2) indicates whether we include class 4 LEGUS sources as potential drivers of ionization, and column (3) indicates whether we impose a prior on f_{esc} in our Bayesian analysis.
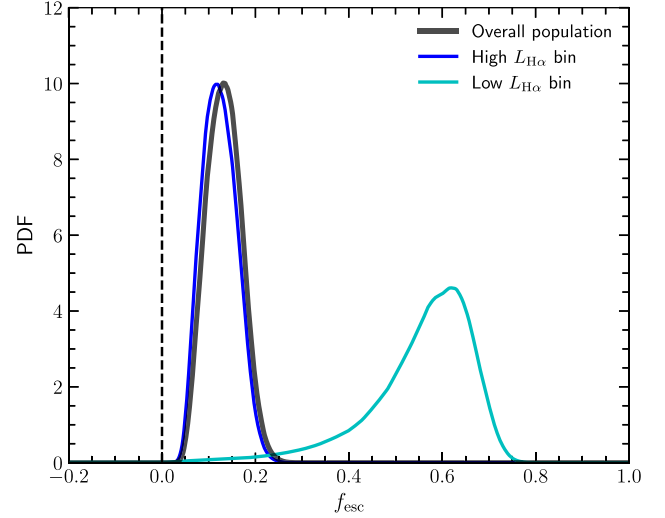
Tracks	Class 4?	f_{esc} prior?	n^a	f_{esc} (Individual, Section 4.1)			f_{esc} (Population, Section 4.2)		
				Low[b]	High	Overall	Low	High	Overall
Geneva[c]	Y	N	42	$0.34^{+0.31}_{-0.70}$	$-0.07^{+0.47}_{-0.75}$	$0.13^{+0.43}_{-0.76}$	$0.56^{+0.08}_{-0.14}$	$0.06^{+0.06}_{-0.06}$	$0.09^{+0.06}_{-0.06}$
Geneva	N	N	33	$0.22^{+0.37}_{-0.96}$	$-0.20^{+0.56}_{-1.13}$	$0.04^{+0.46}_{-1.02}$	$0.55^{+0.10}_{-0.19}$	$-0.07^{+0.07}_{-0.08}$	$-0.04^{+0.07}_{-0.08}$
Padova	Y	N	21	$0.12^{+0.42}_{-0.97}$	$-0.03^{+0.51}_{-1.01}$	$0.04^{+0.47}_{-1.01}$	$0.20^{+0.30}_{-0.51}$	$0.00^{+0.08}_{-0.12}$	$0.02^{+0.08}_{-0.11}$
Padova	N	N	17	$0.23^{+0.39}_{-1.02}$	$-0.37^{+0.65}_{-1.34}$	$-0.06^{+0.65}_{-1.33}$	$0.57^{+0.16}_{-0.50}$	$-0.16^{+0.11}_{-0.13}$	$-0.14^{+0.11}_{-0.12}$
Geneva	Y	Y	42	$0.48^{+0.21}_{-0.27}$	$0.30^{+0.24}_{-0.21}$	$0.41^{+0.23}_{-0.27}$	$0.58^{+0.08}_{-0.13}$		$0.13^{+0.04}_{-0.04}$
Geneva[d]	Y	N	37	$0.39^{+0.30}_{-0.59}$	$0.01^{+0.51}_{-0.67}$	$0.21^{+0.42}_{-0.70}$	$0.55^{+0.11}_{-0.19}$	$0.15^{+0.05}_{-0.05}$	$0.16^{+0.05}_{-0.05}$

Notes. [a]Number of samples after removing sources with large uncertainties and sources in coloured regions (see Section 4.1).
[b]Values quoted for low- and high-$L_{H\alpha}$ bins, and the overall population (see Section 4.1.2).
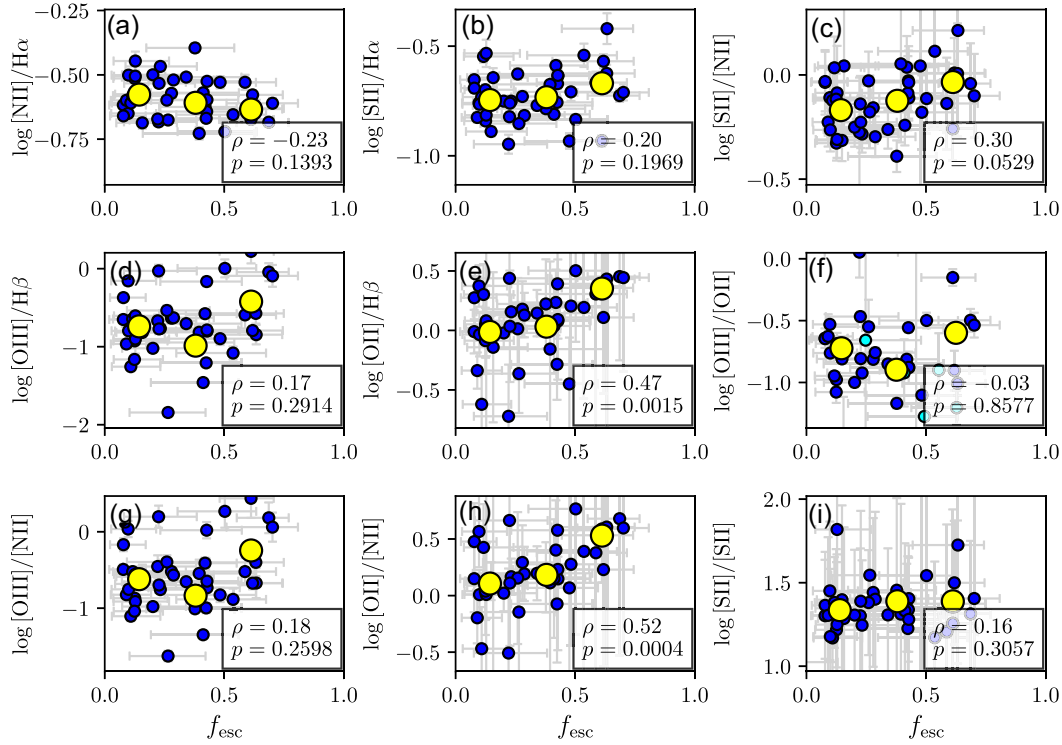[c]Parameters used in the main text.
[d]Same parameter choices as in the main text, but with a slight modification of our approach to associating clusters and H II regions, as described in Appendix C.

Figure A2. Same as Fig. 12, except here we impose that $f_{esc} \geq 0$ as a complementary result to that of Section 4.1.2. We also show the Spearman's rank correlation coefficient ρ and the corresponding p-value. We note the correlation in the line ratios [O II]/[N II] and [O II]/Hβ, as similarly observed in Fig. 12.

median f_{esc} for the high-luminosity part of the sample, this is again comparable in size to the uncertainties. To explore further why this should be, we plot the population-level marginal posterior PDF of f_{esc} derived with a prior $f_{esc} \geq 0$ in Fig. A1, and show the corresponding relationship of f_{esc} with emission line ratios in Fig. A2. Comparing Fig. A1 with Fig. 9, it is clear that the effect of this prior is cut off the low f_{esc} tail of the high $L_{H\alpha}$ bin, but that this does not affect the peak of the distribution, or substantially alter the mean. Similarly, comparing Fig. A2 with Fig. 12, we see that imposing a prior that $f_{esc} \geq 0$ removes the high-uncertainty, low f_{esc} part of the posterior distribution, but that this part of the distribution was not responsible for driving the correlations we find between f_{esc} and [O II]/Hβ or [O II]/[N II]. Thus the significance level of these correlations is not

substantially different than we found without imposing a prior. We therefore conclude that our primary qualitative results are insensitive to whether or not we choose to enforce a positive-only prior of f_{esc}.

## APPENDIX B: VISUALIZING THE DISTRIBUTION OF $Q(\mathrm{H}^0)$

In Section 4.1.1, we discussed the distributions of cluster ionizing luminosity, $Q(\mathrm{H}^0)$, under the effect of stochastic IMF sampling, and how they affect our analysis. To illustrate further, in Fig. B1 we show three common distributions of $Q(\mathrm{H}^0)$ found in this work: (1) sharp, single-peaked profile; (2) multipeaked profile; and (3) broad ($\sigma >$ 0.5 dex), single-peaked profile.
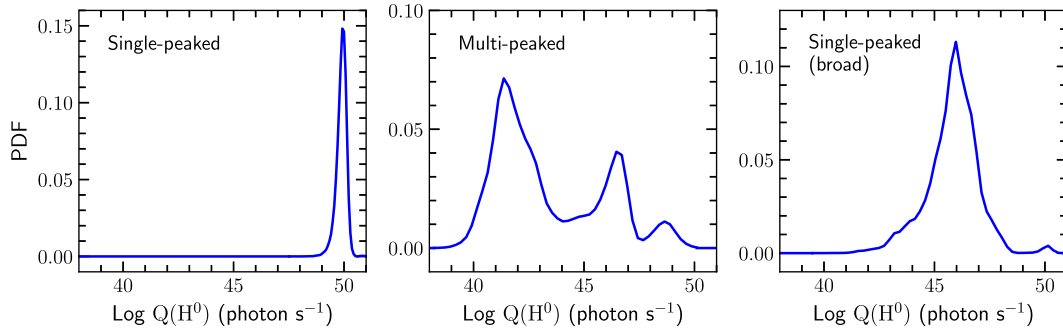
Figure B1. Examples of the posterior PDF of $Q(\mathrm{H}^0)$. From right to left-hand panel: (1) sharp, single-peaked profile; (2) multipeaked profile; and (3) broad (σ > 0.5 dex), single-peaked profile (see Section 4.1.1).
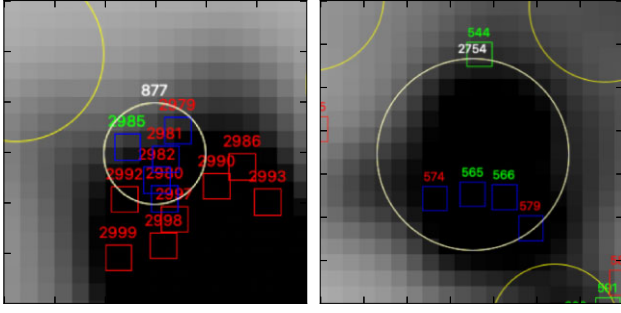
Figure C1. Visual inspection of the two data points in blue regions (see Fig. 7), overlaid on Hα map obtained from the SIGNALS survey. Each figure outlines the target H II region (in white circle) and its corresponding star cluster associations (in blue squares). LEGUS sources that are not associated with the target regions are either outlined with red squares (classes 1, 2, and 3), or with green squares (class 4), whereas the surrounding SIGNALS H II regions are outlined in yellow circles. Additionally, the IDs are shown above every source in white (target H II region), red (classes 1, 2, and 3 LEGUS sources) and yellow (class 4 LEGUS sources).
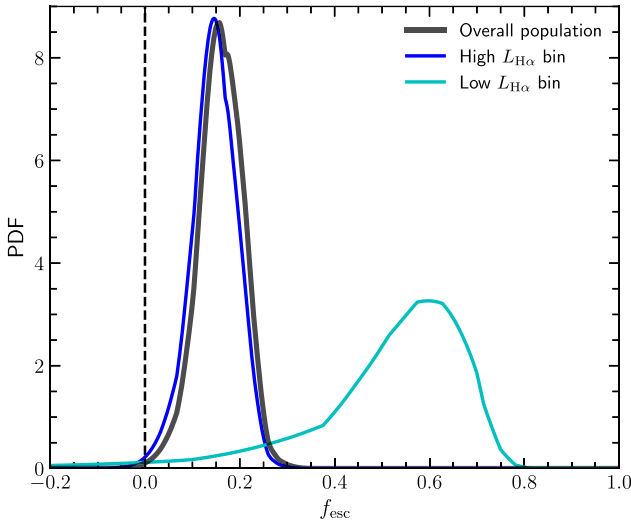
Figure C2. Same as Fig. 9, except here we include clusters within $1.5R$ of H II regions, following our reasoning in Appendix C.

APPENDIX C: POSSIBLE EXPLANATIONS FOR PECULIAR DATA POINTS AND THEIR IMPLICATIONS

Fig. 7 shows two sets of data points (in green and blue regions) that are well within the forbidden zone and are excluded from our analysis. We argue that data points in the green region are either a result of false associations, or that we are missing ionizing clusters due to the completeness of the LEGUS survey. These two arguments, however, cannot fully explain the two data points in the blue region for the following reasons. First, it is highly unlikely that the detected

H II regions and star clusters—both at the high luminosity end—are chance alignments. Second, as illustrated by Fig. 5, the most luminous H II regions are consistently powered by star clusters that are classified by the LEGUS survey. This suggests that the peculiar points could be a consequence of missed diffuse ionizing sources, as we have discussed in Section 3.6.

To further investigate these two data points, we visually examine their location by overlaying LEGUS and SIGNALS data on to the SIGNALS Hα map (see Fig. C1). An interesting feature presented in both sub-figures is that there are LEGUS sources that clearly located within an Hα detection but are not placed in an H II region. Regardless, this analysis shows that it is plausible that the H II region radius reported by SIGNALS can cause an underestimation of the area of influence, thus missing potential ionizing sources. To study this effect, we repeat the analysis in Section 4, except this time we modify our approach as follows: for each H II region in the truncated sample, we consider additional ionizing clusters if the projected distance, D, between the central positions is $\leq 1.5R$, where R is the radius of H II region specified by the SIGNALS catalogue. We show our results of f_{esc} using this approach in Table A1 and Fig. C2, and report that this new approach leads to similar qualitative conclusions.

APPENDIX D: STAR CLUSTER AND H II REGION CATALOGUES

We present in Tables D1 and D2 catalogue entries of star clusters and H II regions used for this paper.

Table D1. The LEGUS star cluster catalogue in the overlapping FoV for NGC 628 (see Grasha et al. 2015; Adamo et al. 2017), showing selected columns and rows. (1) ID_{LEGUS}: identification number, (2) RA: right ascension, (3) DEC: declination, (4) Class: LEGUS assigned class, and (5) $ID_{SIGNALS}$: identification number of associated H II region(s) in the SIGNALS catalogue.

ID_{LEGUS}	RA (deg)	DEC (deg)	Class	$ID_{SIGNALS}$
73	24.1742	15.8082	1	3253
86	24.1751	15.8073	3	3206
...	...	...	...	...
1138	24.1748	15.7859	4	2330
1140	24.1705	15.7858	4	2323, 2359
1141	24.1553	15.7858	3	-
...	...	...	...	...
3671	24.2071	15.7407	2	-

Table D2. The SIGNALS H II region catalogue in the overlapping FoV for NGC 628 (see Rousseau-Nepton et al. 2018), showing selected columns and rows. (1) $ID_{SIGNALS}$: identification number, (2) RA: right ascension, (3) DEC: declination, (4) $L_{H\alpha}$: Hα luminosity of H II region, (5) r: radius of H II region, and (6) ID_{LEGUS}: identification number of associated star cluster(s) in the LEGUS catalogue.

$ID_{SIGNALS}$	RA (deg)	DEC (deg)	$L_{H\alpha}$ (erg s^{-1})	r (pc)	ID_{LEGUS}
397	24.2082	15.7410	1.233×10^{36}	68.0	–
414	24.2068	15.7414	1.882×10^{37}	81.0	–
...	...	...	...	...	...
2013	24.1458	15.7805	1.656×10^{37}	107.0	1669, 1680
2016	24.1777	15.7812	6.339×10^{37}	44.3	1570, 1571, 1605
2017	24.1681	15.7810	3.446×10^{37}	44.3	1619, 1628
...	...	...	...	...	...
3253	24.1740	15.8082	5.577×10^{36}	89.0	73

[1]*Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Straße 2, D-69120 Heidelberg, Germany*

[2]*Research School of Astronomy and Astrophysics, Australian National University, Bart Bok Rd, Stromlo ACT 2611, Australia*

[3]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*

[4]*Visiting Fellow, Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*

[5]*Department of Astronomy, University of Massachusetts, 710 North Pleasant Street, Amherst, MA 01003, USA*

[6]*Canada–France–Hawaii Telescope, Kamuela, HI 96743, USA*

[7]*Department of Physics and Astronomy, University of Hawaii at Hilo, Hilo, HI 96720-4091, USA*

[8]*Département de Physique, Université de Montréal, Succ. Centre-Ville, Montréal, Québec H3C 3J7, Canada*

[9]*Centre de Recherche en Astrophysique du Québec (CRAQ), Québec, QC G1V 0A6, Canada*

[10]*The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova, SE-10691 Stockholm, Sweden*

[11]*Steward Observatory, University of Arizona, Tucson, AZ 85721-0065, USA*

[12]*George P. and Cynthia W. Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA*

[13]*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstr, 12–14, D-69120 Heidelberg, Germany*

[14]*Caltech/IPAC, 1200 E. California Boulevard, Pasadena, CA 91125, USA*

[15]*Observatoire de Paris, LERMA, Collège de France, PSL University, CNRS, Sorbonne University, F-75014 Paris, France*

[16]*Observatoire de Genève, Université de Genève, Chemin Pegasi 51, CH-1290 Versoix, Switzerland*

[17]*Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany*

[18]*Instituto de Astrofisica de Andalucia (CSIC), Gta. de la Astronomía, s/n, E-18008 Granada, Spain*

[19]*Dipartimento di Fisica G. Occhialini, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*

[20]*INAF – Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*

[21]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

[22]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*

[23]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

[24]*Instituto de Astrofisica, Universidad Andres Bello, Fernandez Concha 700, Las Condes, Santiago RM, Chile*

[25]*Department of Astronomy, Xiamen University, Xiamen, Fujian 361005, China*

[26]*Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, D-14482 Potsdam, Germany*

[27]*Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*

This paper has been typeset from a TeX/LaTeX file prepared by the author.