The Astrophysical Journal, 917:23 (44pp), 2021 August 10 https://doi.org/10.3847/1538-4357/ac0251

Stars with Photometrically Young Gaia Luminosities Around the Solar System (SPYGLASS). I. Mapping Young Stellar Structures and Their Star Formation Histories

Ronan M. P. Kerr, Aaron C. Rizzuto[1], Adam L. Kraus, and Stella S. R. Offner
Department of Astronomy, University of Texas at Austin, 2515 Speedway, Stop C1400, Austin, TX 78712-1205, USA; rmpkerr@utexas.edu
Received 2021 March 12; revised 2021 May 11; accepted 2021 May 16; published 2021 August 10

Abstract

Young stellar associations hold a star formation record that can persist for millions of years, revealing the progression of star formation long after the dispersal of the natal cloud. To identify nearby young stellar populations that trace this progression, we have designed a comprehensive framework for the identification of young stars and use it to identify $\sim 3 \times 10^4$ candidate young stars within a distance of 333 pc using Gaia DR2. Applying the HDBSCAN clustering algorithm to this sample, we identify 27 top-level groups, nearly half of which have little to no presence in previous literature. Ten of these groups have visible substructure, including notable young associations such as Orion, Perseus, Taurus, and Sco-Cen. We provide a complete subclustering analysis of all groups with substructure, using age estimates to reveal each region's star formation history. The patterns we reveal include an apparent star formation origin for Sco-Cen along a semicircular arc, as well as clear evidence for sequential star formation moving away from that arc with a propagation speed of $\sim$4 km s^{-1} ($\sim$4 pc Myr^{-1}). We also identify earlier bursts of star formation in Perseus and Taurus that predate current, kinematically identical active star-forming events, suggesting that the mechanisms that collect gas can spark multiple generations of star formation, punctuated by gas dispersal and cloud regrowth. The large spatial scales and long temporal scales on which we observe star formation offer a bridge between the processes within individual molecular clouds and the broad forces guiding star formation at galactic scales.

Unified Astronomy Thesaurus concepts: Stellar associations (1582); Stellar ages (1581); Star formation (1569); Young star clusters (1833); Young stellar objects (1834); Pre-main sequence stars (1290); OB associations (1140)

Supporting material: figure set, machine-readable tables

1. Introduction

Most young stars are not found in isolation, instead residing in comoving star clusters or associations (e.g., Lada & Lada 2003; Krumholz et al. 2019). These stellar overdensities are direct remnants of the molecular cloud that collapsed to create them, and as such, they preserve significant information on the structure and dynamics of those parent clouds (Elmegreen & Lada 1977; Krause et al. 2020). While studies of sites of active star formation are popular due to the presence of both young stars and dense gas for which exquisite dynamical studies can be performed (e.g., Palla & Stahler 1999; Hatchell et al. 2005; Tobin et al. 2009; Kirk et al. 2013, 2017; Kerr et al. 2019), studies of clusters and associations provide a much longer-term view of star formation that has unique value. Rather than providing a snapshot in time, a detailed study of a young association can trace back tens of millions of years of star formation, enabling a complete study of a star-forming event from its onset to the present day and, by extension, the processes that drove the evolution of the population (e.g., de Zeeuw et al. 1999; Pecaut & Mamajek 2016; Wright & Mamajek 2018). Furthermore, as these associations persist long after gas dispersal, the study of stellar populations through associations may also enable the identification of more rapid forms of star formation that do not have known equivalents in active star-forming sites.

Some processes in star formation require a complete record of the star-forming event, not just a snapshot, to be properly studied in nature. One example is sequential star formation, a process in which previous generations of stars compress the cloud beside them, which can then collapse to form stars, producing a self-sustaining cycle of star formation that can slowly propagate across an entire molecular cloud (Elmegreen & Lada 1977). Most cases where a sequential process has been suggested include just two generations of star formation: one recently formed generation powering an H II region and one site of active star formation triggered in a shell that the previous generation compressed (e.g., Lee et al. 2005; Maaskant et al. 2011; Nony et al. 2021). Given that these processes are capable of continuing without limit as long as unused gas remains, the current view of sequential star formation has yet to explore large scales in both time and space.

One of the greatest strengths in using clusters and associations as a record of past star formation lies in the time over which detailed information can be extracted. Having such an extensive star formation record allows for the study of long-lived star-forming processes while also revealing unexpected anomalies in time, such as periods of dormancy in the star formation record. As simulations become increasingly sophisticated and begin to include more physical processes (e.g., Grudić et al. 2021a), having a robust record of star formation complete with currently unexplained features can provide critical comparisons capable of testing new theoretical frameworks.

While more spatially compact open clusters were long ago discovered and cataloged (e.g., Trumpler 1921; Klein Wassink 1927; Mermilliod 1995), much sparser stellar associations are considerably more challenging to identify due to relatively sparse on-sky densities and large spatial extents, often barring identification from on-sky density. For the nearest populations, stars with strong indicators of youth, such as

[1] 51 Pegasi b Fellow.

protoplanetary disks, have typically been used as signposts for the identification of associated populations nearby, such as TW Hydrae and β Pictoris (e.g., Kastner et al. 1997; Zuckerman et al. 2001). Beyond about 50–100 pc, however, recognizing stars with youth indicators and finding potential companions becomes increasingly difficult, leaving the population of low-mass, more distant associations largely unexplored. Due to their unbound nature, small velocities relative to a typical field star, and geometric effects from their wide spatial distributions, the kinematics of associations are often difficult to disentangle from the field. The effective identification of young stellar populations therefore requires both the suppression of older background populations and the use of accurate kinematics and 3D spatial coordinates to properly group the young stars together and distinguish a young association from the field. Improvements in measurements of parallax and proper motions can therefore significantly expand our knowledge of these associations, and such developments frequently result in the addition of new members and the revision of associations' known spatial extents (e.g., de Zeeuw et al. 1999; Preibisch & Mamajek 2008; Rizzuto et al. 2015; Kraus et al. 2017; Luhman 2018; Zari et al. 2018). With its unprecedented breadth of spatial, kinematic, and photometric data, Gaia Data Release 2 (DR2) provides a data set capable of dramatically expanding our knowledge of young stellar populations (Gaia Collaboration et al. 2018b). Gaia photometry can be combined with parallaxes to identify stars based on their high locations on the Hertzsprung–Russell (H-R) diagram, and those stars can then be clustered according to their spatial coordinates and transverse motions.

There have been multiple recent searches for stellar populations in the solar neighborhood enabled by Gaia DR2, including Sim et al. (2019), who reported the discovery of 207 new open clusters, and Kounkel & Covey (2019), who noted the identification of 1901 stellar overdensities in Gaia DR2. However, these surveys included stars of all ages in their clustering, so young populations may not always stand out above the field density of the older stars. The work of Zari et al. (2018) does focus on these younger populations, separating out pre-main sequence stars quite effectively and revealing multiple associations in the form of stellar overdensities. While Zari et al. (2018) revealed substantial young stellar populations, they did not include a clustering analysis, and as such, the structures present within those populations have yet to be rigorously defined. Some recent investigations have focused on young stellar populations while also performing a complete clustering analysis; however, these have all been restricted to individual associations, such as Zari et al. (2019), who studied Orion, and Cantat-Gaudin et al. (2019a), who focused on Vela. No paper to date has performed a spatially unbiased, all-sky search for young stellar populations with both robust youth diagnostics and a broad clustering analysis.

In this paper, we present the deepest comprehensive study of young stellar populations in the solar neighborhood to date. In Section 2, we present the Gaia DR2 data set used for this work. We outline our methods for the identification of young stars, age estimation, and clustering in Section 3, while a detailed analysis into the populations we identify is performed in Section 4. Possible implications of this work with respect to the study of star formation progression in the solar neighborhood are discussed in Section 5, while we conclude in Section 6.

2. Data

2.1. Gaia Astrometry and Photometry

In this paper, we search for nearby young stars among a large sample drawn from Gaia DR2 (Gaia Collaboration et al. 2018b; Lindegren et al. 2018). When querying sources from the Gaia archive, several search restrictions are included to ensure a manageable sample size and guarantee that all stars have high-quality photometric and astrometric measurements. First, we impose a search radius of parallax $\pi > 3$ mas, limiting our sample to stars at distances comparable to Taurus and Sco-Cen, two of the nearest young associations. Both Taurus and Sco-Cen are thought to extend to a maximum distance of approximately 200 pc (Preibisch & Mamajek 2008; Galli et al. 2019), so a minimum parallax of 3 mas ($d < 333$ pc) allows for the exploration of adjoining structures that may exist up to 100 pc beyond their known extents, in addition to numerous other structures, both known and unknown. This radius also includes nearly all of Perseus OB2 and the near edge of Orion (Bally et al. 2008; Zari et al. 2019), allowing for partial coverage in these more distant young associations.

We then impose several restrictions based on quality indicators following Arenou et al. (2018). First, we restrict the sample using the unit weight error (u), which can be interpreted as a goodness-of-fit measurement to the astrometric solution in Gaia DR2. This value is given by $u = \sqrt{\chi^2/\nu}$, where χ^2 represents the χ^2 value between the source and a single-star astrometric solution,[2] and ν is defined as the number of good CCD observations used in the astrometric solution[3] minus 5 (Lindegren et al. 2018). We subsequently impose the following restriction on the unit weight error:

$$u < 1.2 * \max[1, \exp(-0.2(G - 19.5))], \quad (1)$$

as suggested in Arenou et al. (2018). A different condition on the astrometric goodness of fit was proposed by Lindegren (2018), instead using the renormalized unit weight error (RUWE) and the condition RUWE < 1.4. This cut is somewhat more restrictive than our cut on u, producing approximately 5% fewer stars than the cut used above. However, since we find no evidence of spurious astrometric solutions under the looser condition, we opt for the less restrictive cut on u.

In addition, we impose a restriction on the BP/RPs flux excess factor[4] E, which is defined as the sum of the fluxes in the Gaia BP and RP photometric bands divided by the flux in the G band. The configuration of the Gaia photometric bands implies that E should be slightly larger than 1 for stars with good photometric measurements; however, the BP and RP bands are vulnerable to external contamination, since they are based on integrated flux in a small field around the star rather than the profile fitting used for the G band (Evans et al. 2018). This contamination can manifest in the value for E, and contaminated sources can therefore be removed using the following restriction:

$$1.0 + 0.015(G_{BP} - G_{RP})^2 < E < 1.3 + 0.037(G_{BP} - G_{RP})^2. \quad (2)$$

[2] astrometric_chi2_al in the Gaia archive.
[3] astrometric_n_good_obs_al in the Gaia archive.
[4] phot_bp_rp_excess_factor in the Gaia archive.

This nearly matches with the restriction proposed in Arenou et al. (2018), with a slight modification to the factor in the upper bound, where Arenou et al. (2018) used 0.06 rather than 0.037. We found that for known members of Upper Sco (US) and Taurus (Preibisch et al. 1998, 2002; Rizzuto et al. 2015; Luhman 2018), the slightly less restrictive conditions would occasionally show very old photometric ages for these young stars. The more restrictive condition removed nearly all such sources, significantly reducing our rates of false negatives among known young objects.

We also impose restrictions on the number of visibility periods used in the astrometric solution, which are defined as groups of observations separated by at least 4 days. This requires the astrometric solution to be based on a strong baseline of measurements, a condition that was shown in Arenou et al. (2018) to significantly reduce the abundance of apparently spurious astrometric results:

$$\text{visibility\_periods\_used} > 8. \quad (3)$$

Finally, we restrict the sample based on the parallax inverse fractional error[5] (π/σ_π) to exclude stars with a poorly constrained distance, following the condition from Gaia's paper on observational H-R diagrams (Gaia Collaboration et al. 2018a):

$$\pi/\sigma_\pi > 10. \quad (4)$$

Parallaxes are important for generating the absolute magnitudes used in age estimation, and this condition ensures that the absolute magnitude uncertainty remains below ∼0.2 mag, a limit comparable to the uncertainties in the reddening estimates used later in this paper.

While the combination of cuts employed here excludes nearly half of the known young stars in Taurus and US, these restrictions are important to ensure that all absolute photometric measurements can be considered trustworthy, which is critical to the reliable derivation of stellar properties, particularly age. Once stars that pass these rigorous quality cuts are used to identify and define the extents of young groups and associations in space-velocity coordinates, the restrictions can be relaxed in future targeted studies, which can significantly increase the completeness. See Section 3.2 for a more complete and quantitative analysis of the completeness of our results.

3. Methods

The solar neighborhood contains vast quantities of stars, including nearly 5 million within our 333 pc search limit that survive our Gaia quality cuts. However, assuming a constant star formation rate over the last ∼10 Gyr, only about one in 200 stars will have formed in the last 50 Myr. The Sco-Cen association is thought to dominate nearby young populations, and it contains a membership of only ∼10,000 stars covering a massive 80° long swath along the plane of the Milky Way (Preibisch et al. 2002; Rizzuto et al. 2015), so field populations will dominate young structures in most cases. To isolate young populations from the field, photometry can often be used, as many low-mass young stars have yet to leave the Hayashi track (Hayashi 1961) or have not yet settled onto the main sequence. Similarly, O and B stars are readily identified as young, as they do not live long enough to become old. However, in order to reliably identify young stars by photometric means, it is critical that we understand what young stellar populations can and cannot be confidently separated from the contamination of older field stars.

In this section, we describe our approach for identifying and characterizing young stars in the solar neighborhood. In Section 3.1, we describe the implementation of a Bayesian statistical classification approach that uses a model stellar population to identify credibly young stars in the solar neighborhood, and we assess the success of this method in Section 3.2. The rest of this section describes the methods used to further characterize stars and the larger groups they might belong to. Section 3.3 describes our method to obtain age estimates for individual stars, while we identify groups and other young structures in the solar neighborhood using the HDBSCAN clustering algorithm in Section 3.4, lastly computing more precise bulk ages for these larger groups in Section 3.5.

3.1. Generating the Model Population

With the excellent photometric and astrometric data provided by Gaia, we are able to precisely determine the absolute magnitudes and colors of stars in the sample. However, complicating factors such as metallicity, multiplicity, and reddening modify the location of a star in absolute magnitude space, meaning that a star that appears to be photometrically young may be better explained by some combination of other factors (e.g., Sullivan & Kraus 2021). We therefore need to generate posterior distributions in age for each star and marginalize over the other factors that may influence the photometric youth of a star, including reddening and extinction, multiplicity, metallicity, and stellar mass. To do this, we generate a simulated population of 10 million stellar systems, a sample size that balances parameter space coverage with computational limitations. Each system is assigned a multiplicity, and each star within it is assigned an age, mass, and metallicity, as well as an extinction based on external reddening maps (Lallement et al. 2019). We then generate posterior distributions in age, as well as mass, corresponding to the location of real stars from Gaia DR2 in the model population. By integrating the age posterior over all ages less than 50 Myr, we can estimate the probability $P_{\text{age}<50\text{ Myr}}$ that a star is genuinely young, allowing us to isolate young stellar populations in the solar neighborhood.

3.1.1. Age

For the purposes of developing an age distribution, we can approximate the star formation in the solar neighborhood to be constant over its lifetime, for which we take the age of 11.2 Gyr from Binney et al. (2000). This is a simplification of the true star formation history of the solar neighborhood, however, as multiple studies have identified bursts in star formation activity on scales of ∼1 Gyr (e.g., Rowell 2013; Isern 2019). Current literature also indicates the possible presence of larger-scale trends in the star formation rate, with most asserting some form of a decrease in star formation since the galaxy's earliest star formation bursts (e.g., Aumer & Binney 2009; Schönrich & Binney 2009). Most studies that assert a decrease in the star formation rate take into account dynamical heating, a process by which older stars are raised higher in the disk, therefore excluding many of them from consideration. This effect works

[5] parallax_over_error in the Gaia archive.

in opposition to a decreasing star formation rate, implying that despite a possible gradual drop in the star formation rate, the stars that remain within our solar neighborhood are closer to having an even distribution in ages.

We also assume that the age distribution is spatially isotropic, which allows us to generate a single representative population that can be applied to all of our Gaia stars, regardless of location. However, most recent star formation is thought to have occurred within ∼100–200 pc of the galactic plane (Urquhart et al. 2014; Anderson et al. 2019), with stars higher above the plane mainly consisting of older objects potentially raised into their current location by dynamical heating. As such, young stars will represent a significantly higher fraction of the total stars close to the disk rather than at high galactic latitudes. The assumption of an isotropic age distribution may therefore inflate the rate of false-positive young star identifications near the Galactic poles by virtue of the absence of real young stars there, but our rate of recovery among genuinely young stars should be roughly constant across the entire population.

Not all literature agrees with the conclusion of a star formation rate that is either decreasing or steady, on average (e.g., Rocha-Pinto et al. 2000); however, the deviations that these alternative star formation histories might impose would be of order unity and therefore have a limited impact on our posterior distributions. On the main sequence, the photometry of stars is relatively constant, so there is usually little photometric difference between otherwise identical stars at, for example, 1 and 10 Gyr. Our objective in generating the population of older stars should therefore be to ensure that their numbers are approximately in proportion with their abundance in the solar neighborhood, meaning that any star formation bursts can be essentially averaged over. Any larger-scale SFR trends may slightly affect our results by means of changing the relative fraction of young and old stars; however, these effects can be accounted for by modifying the probability cutoff for selecting candidate young stars in Section 3.1.7.

A star formation rate that is to first order uniform over the last ∼11 Gyr implies that stars formed within the last 50 Myr should account for less than 1% of the total stars in the solar neighborhood. Subsequently, a representative sample of ages would overwhelmingly skew toward very old objects, leaving a sparse sample of young stars. Since stars evolve quickly on the pre–main sequence, a sparse sample of model stars at these young ages would be insufficient to effectively assess whether any observed star is genuinely young. We therefore draw model stellar ages from a $\log_{10}$-uniform distribution spanning 1 Myr–11.2 Gyr. This does introduce an exponential bias toward young objects, however, that can be corrected using a prior, as described in Section 3.1.7.

3.1.2. Primary Masses

The initial mass function (IMF), which displays how the abundance of stars varies with mass, is the basis for this project's mass generation. It follows a distribution that has been gradually revised since the first form presented by Salpeter (1955), with the most recent revisions from Chabrier (2003, 2005) using a lognormal distribution at low masses ($M < 1\ M_\odot$) and the Salpeter (1955) power-law distribution for higher masses. The IMF has been consistently shown to be invariant based on environment, so spatial variations to our mass generation are not of concern (Chabrier 2005; Offner et al. 2014). Slightly different

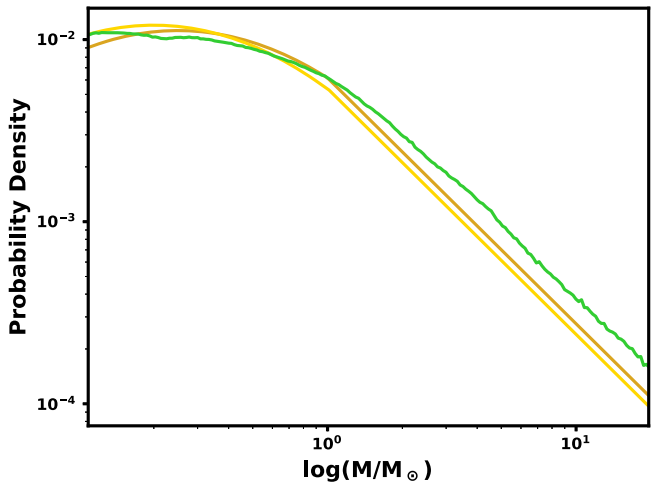

Figure 1. Chabrier (2005) individual object IMF (bright yellow), which was used to set the primary masses in our model systems, plotted against the distribution of system masses in our model after binary and triple companions are added (green). The Chabrier (2005) system IMF is included for reference (dark yellow). Our model systems never deviate from the Chabrier (2005) system IMF by more than 30%, although our model does show a minor deficiency of low-mass binaries with components below the lower mass limit of our model grid (Chen et al. 2015).

lognormal solutions have been presented for individual stars and complete stellar systems, and since we require masses of primaries that may or may not be in multiple systems, we make use of the Chabrier (2005) individual star IMF to generate the masses of primaries. Both the individual and system Chabrier (2005) IMFs are plotted in Figure 1, alongside the distribution of system masses we reach after adding stellar companions in Section 3.1.4.

The minimum mass we consider is 0.11 $M_\odot$, which is roughly the smallest mass that is available in all PARSEC v1.2S isochrones (Chen et al. 2015)). Young stars of this mass also reach Gaia's approximate limiting magnitude of $G \simeq 21$ near our maximum distance of 333 pc (Arenou et al. 2018), so objects in this mass range will be accessible throughout our entire search radius. Our maximum mass was set to 20 $M_\odot$, which excludes a small subset of stars that are both very rare and very bright. Main-sequence stars of more than 20 $M_\odot$ have an absolute magnitude $G < -3$ (Chen et al. 2015), which corresponds to a magnitude of roughly 3 at the distance of Sco-Cen. Gaia photometry and astrometry in this range are known to be poor due to the start of pixel saturation at $G \sim 6$ (Arenou et al. 2018), so the contribution Gaia can make toward furthering the understanding of these very bright stars is relatively minimal.

Due to a combination of the higher average mass of primaries relative to the complete population considered in the individual IMF, as well as the mass limit of 0.11 $M_\odot$ imposed on all model stars, our system mass distribution does not quite replicate the Chabrier (2005) system IMF. For stellar systems, which are discussed in depth in Section 3.1.4, a minimum component mass of 0.11 $M_\odot$ implies that a minimum binary mass is 0.22 $M_\odot$, and as a result, this mass marks a small local minimum in the mass function in Figure 1. As such, the discrepancies that do exist are mostly brought on by limitations in the masses of objects our system can create. However, despite these minor deviations, the final simulated system mass function remains comparable to the system IMF also presented in Chabrier (2005), never differing by more than 30%. We

expect the effects of a slightly discrepant model IMF to be relatively minor in our final results. The overabundance of higher-mass stars relative to lower-mass binaries may overweight the contamination from subgiants; however, particularly in the parameter space occupied by the pre-main sequence, G and F stars follow evolutionary tracks well separated from the early M stars, which have the greatest underabundances in our model (Chen et al. 2015). This suggests that situations where the relative abundances of these different mass model stars have a significant impact on our results will be very rare. Any such effects will be further minimized by our generous 10% probability threshold for the consideration of a star as young and the reliance on kinematics to cull likely erroneously detected field stars, as outlined in Section 3.4.

3.1.3. Metallicity

While stars are thought to be more metal-rich on average at later formation epochs, the age–metallicity relationship in the solar neighborhood appears to be relatively flat beyond an initial early enrichment phase (e.g., Haywood et al. 2019; Lin et al. 2020), inferring that a single non-age-dependent metallicity distribution will be sufficient to represent the solar neighborhood. Regardless of whether metallicity is truly constant with time, it may not influence our detection of young stars, since the nearly age-independent photometry of main-sequence stars means that the age will weigh minimally on the photometry of these stars, regardless of metallicity.

A possibly more consequential bias that a single metallicity distribution may introduce relates specifically to young stars, as most that are known have solar metallicities. Using the full range of possible metallicities to represent these stars may therefore present a wider range of possible photometry compared to what actually exists, particularly for low metallicities. This would result in an underestimation of $P_{\mathrm{age<50\ Myr}}$, as these low-metallicity stars have photometry more consistent with older populations (Chen et al. 2015). However, since we wish to identify structures beyond those already known in the solar neighborhood, we have little cause to tether our populations to an assumption of solar metallicity. Even if the assumption of near-solar metallicity is universally appropriate, placing stars too low on the pre–main sequence due to low metallicity is a conservative error and will therefore not lead to spurious assessments of youth. The resulting lower probability estimations can subsequently be negated using more permissive probability cuts for the identification of young stars. We therefore conclude that a non-age-dependent distribution for metallicity would provide a reasonable approximation to reality. Data Release 2 of the GALAH survey (Buder et al. 2018; Hayden et al. 2019) provides an empirical sample containing over 62,000 stars within 500 pc of the Sun that we use as a representative sample for the metallicities of stars in the solar neighborhood.

Our metallicities were randomly generated from a slightly modified version of the GALAH metallicity distribution. We first restricted the metallicities to between $[\mathrm{Fe/H}] = -1.0$ and 0.5. The lower limit was due to the negligible number of GALAH stars with metallicities in that range ($\sim$0.1%), while the upper limit was chosen to comply with the metallicity range available in PARSEC v1.2S isochrones: $-2.2 < [\mathrm{Fe/H}] < 0.5$ (Chen et al. 2015). There are more stars that exceed this high-metallicity limit in the GALAH data compared to the low-metallicity limit; however, the fraction of total stars occupied by these outliers remains under 0.5% (Hayden et al. 2019). Therefore, these extremely metal-rich stars are not sufficiently numerous for their exclusion to dramatically skew our results. We then binned the GALAH measurements with [Fe/H] between -1.0 and 0.5 into 200 evenly spaced bins over that range and smoothed the resulting distribution using a Savitsky–Golay filter with a 15 bin window length and third-order polynomial. This generated a smooth metallicity distribution consistent with the original histogram from which we can draw samples, which is shown in Figure 2.

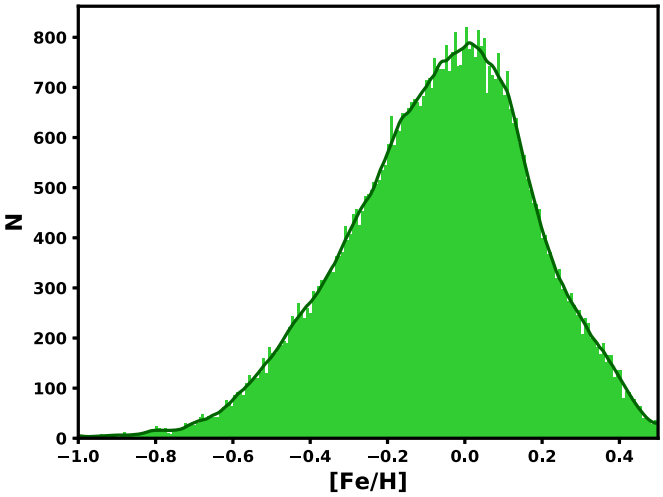

Figure 2. Metallicity histogram for stars observed by the GALAH survey. The dark curve shows the smoothed distribution that we generated metallicities from (prior to normalization).

3.1.4. Multiplicity

Binaries are quite common in the Milky Way, with a mass-dependent abundance fraction that increases in tandem with the mass of the primary (e.g., Duchêne & Kraus 2013). For stars less massive than the Sun, between one-fifth and half of all stars are binaries, and depending on the distance, many of these cannot be resolved by Gaia. While resolved binaries are not of concern for the purposes of this study, as Gaia is able to generate accurate photometry and astrometry for each resolved component, unresolved binaries create combined photometric detections that appear overluminous relative to comparable single stars. Dim companions have a relatively minor impact on the system photometry; however, the presence of an unresolved equal-brightness binary can double the incident flux and subsequently lower the magnitude by up to 0.76 mag, a much more significant contribution. This brightness relative to the main sequence is comparable to what we see on the pre–main sequence, and therefore these older unresolved binaries can easily be mistaken for younger single stars. Therefore, to accurately assess whether a star is young, the abundance of multiple stars, distribution of companion masses, and companion separations must all be well modeled to quantify the probability that stars in any region above the pre–main sequence are multiple systems rather than young stars.

We therefore add multiples to our synthetic population, considering only single, double, and triple systems to avoid complications with the architecture of rare, higher-order systems. We begin with the statistics reported in Duchêne & Kraus (2013), including the multiplicity frequencies (MFs), which measure the occurrence frequency of multiple systems of

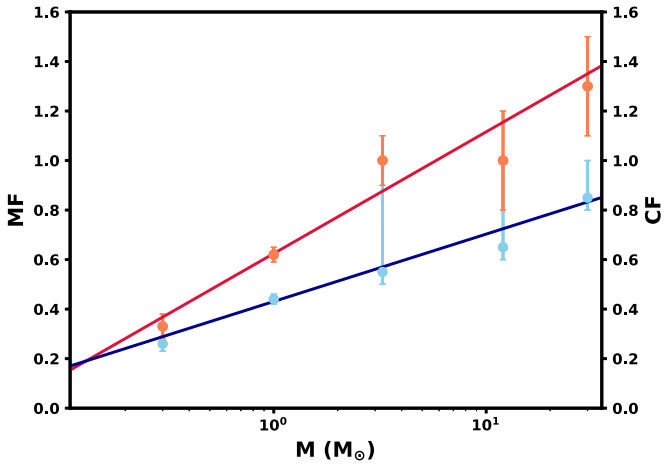

Figure 3. Fits to the MF (blue) and CF (red) distributions as a function of primary mass. Note that the confidence intervals for the two rightmost points for the MF distribution are lower limits, and the dot provided is the chosen intermediate value. These points are heavily deweighted compared to the well-constrained MF values at lower masses and have minimal impact, since our selected sample consists almost entirely of stars of spectral type G or later.

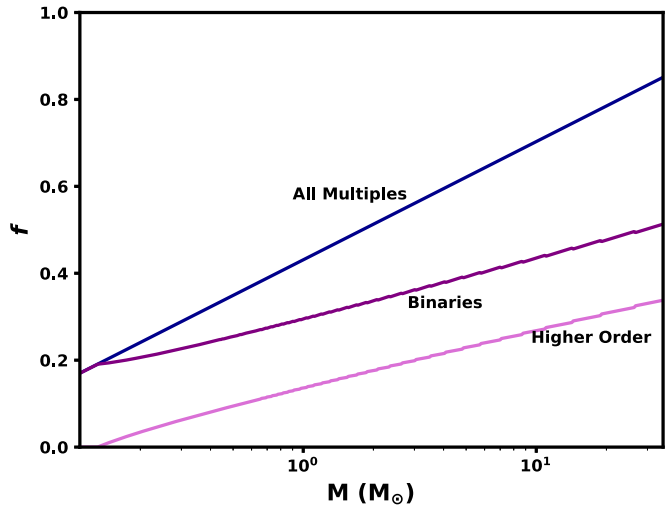

Figure 4. Final distributions used for the binary, higher-order (which are treated as triples), and total multiplicity fractions. The blue line is equivalent to the MF fit in Figure 3.

any kind, and companion frequencies (CFs), which consider the number of companions per primary star (originally sourced from Abt et al. 1990; Preibisch et al. 1999; Delfosse et al. 2004; Kouwenhoven et al. 2005, 2007; Mason et al. 2009; Raghavan et al. 2010; Chini et al. 2012; Dieterich et al. 2012; Sana et al. 2012). While the MF describes the probability of a star being in some form of multiple system, the CF allows for conclusions to be drawn about the occurrence frequencies of both double and triple stars.

To generate a smooth distribution, we fit a logarithmic function as in Equations (5) and (6) to the MF and CF values from Duchêne & Kraus (2013), using reasonable intermediate values for the MF in cases where only upper limits are provided. These fits are shown in Figure 3, matching the linear trends in semi-log mass versus MF and CF that are followed by the Duchêne & Kraus (2013) values for MFs and CFs. The resulting solutions for CFs and MFs as a function of primary mass are as follows:

$$\mathrm{MF} = 0.272 * \log_{10} M + 0.431, \quad (5)$$

$$\mathrm{CF} = 0.491 * \log_{10} M + 0.625. \quad (6)$$

Then, assuming that the probability of each multiplicity follows some geometric series Ck^m (as in Duchêne & Kraus 2013), where m is the multiplicity of the stellar system, we fit for the unique solution for C and k that returns the appropriate values for CF and MF interpolated from our fit. The resulting abundance fraction distributions for multiples, binaries, and higher-order systems are given in Figure 4. The result has the correct edge behavior, as the double and especially triple rates drop dramatically as the primary mass approaches our minimum mass limit, and the multiple rate approaches but does not exceed 1 at the high-mass limit.

The probability distribution of the resulting companion masses is typically described as a function of the mass ratio between the primary and secondary components $m_s/m_p = q$, following a power-law distribution of the form $P(q) = Cq^\gamma$ (Kraus et al. 2011; Rizzuto et al. 2013). We therefore generate secondary and tertiary components based on the mass ratio probability distributions presented in Kraus et al. (2011) and Rizzuto et al. (2013), which cover q values between 0.1 and 1. In order to exclude companions in the brown dwarf regime that are not massive enough to be included in the PARSEC isochrones, we multiply these probabilities by an additional factor, which limits the section of the mass ratio distribution from which companions can be drawn to only that region between the primary mass and the minimum mass accessible through the PARSEC isochrones. The corrective factor for a given primary mass is therefore the integral of the probability distribution over the window from the primary mass to the minimum PARSEC mass, normalized such that the sum over all q is 1. Once this corrective factor to the multiplicity probabilities is applied, the primaries are separated into three populations, each with distributions for q governed by a power law with a different exponent γ: $\gamma = 0.4$ for $M < 0.7\ M_\odot$, $\gamma = 0$ for 0.7–2.5 $M_\odot$, and $\gamma = -0.46$ for masses exceeding 2.5 $M_\odot$. The two lower mass ranges come from Kraus et al. (2011), and the upper mass bracket comes from Rizzuto et al. (2013).

The generation of masses for all primaries and companions represents the final step necessary before computing simulated inherent photometric properties for these stars. The mass distribution of systems after the inclusion of companions is shown in Figure 1, plotted against the Chabrier (2005) IMFs.

The separation of multiple system components must also be considered, as only spatially unresolved systems will display merged photometry. The stellar populations we wish to draw young stars from come from the Gaia mission, and since Gaia reports resolved stars independently, multiples that are resolved in Gaia are seen as a series of two or more singles. To generate approximate separations, we take the distribution of periods presented in Raghavan et al. (2010), randomly select an orbital period from that distribution, and compute an orbital semimajor axis from the mass of the system using Kepler's third law. A corrective factor of $\pi/4$ is applied to account for the projection of these orbits on the plane of the sky, which is modulated by the presence of randomly oriented systems. Recent literature investigating the separations of M dwarf binaries has shown tighter lognormal distributions centered at slightly smaller periods for these stars (e.g., Bergfors et al. 2010); however, these results are not currently well enough constrained for us to justify including a mass-specific period distribution. We may introduce revised separation distributions in a future iteration of

this work, as the current implementation likely underestimates the abundance of unresolved binaries for M dwarfs. However, even if an updated form of the separation distribution were to imply that all M dwarf binaries are unresolved, this would only represent a roughly 30% increase in the unresolved rate for companions to an $\sim$0.2 $M_{\odot}$ star, so we do not expect significant effects from the use of the Raghavan et al. (2010) period distribution.

Most Gaia stars with separations exceeding 1″ are resolved as independent sources (Rizzuto et al. 2018; Ziegler et al. 2018), although this does not appear to be a hard boundary, as there is likely a population of systems at separations below 1″ that are neither unresolved nor reported independently, a factor that may reduce the mean separation at which this transition to independent identification occurs. However, even with that transition separation lowered to 0.″2, the unresolved rate for an average 1 $M_{\odot}$ star only drops by 25%, inferring that the use of 1″ as a hard limit for the transition between unresolved and resolved populations appears to be a reasonable approximation. We take a distance of 234 pc, which is the mean distance to the Gaia stars investigated, and break up test stars with separations exceeding 1″ at that distance. Stars with smaller separations are marked as unresolved systems, which are combined photometrically into a single star. For triple stars, separations are generated for each companion. If both are unresolved, then all three members are photometrically merged. If only one model separation is resolved, then the second component is counted as resolved from the primary, and the third is randomly assigned to either the primary or secondary. The result is that the system is now treated in our model as one unresolved double with a resolved single nearby. The choice to separate binaries at a single fixed distance, rather than make it adaptive to the distance of the target star, was an approximation made to reduce computational costs; however, we do not expect the effects of this choice to be significant. At 10 pc, a companion to a 1 $M_{\odot}$ star would have an approximately 34% chance of being unresolved, compared to about 72% for a star at our field limit of 333 pc and 68% for our chosen distance at 234 pc. As such, the only major inaccuracies will be at very small distances, where we already expect our cluster detection to be less sensitive. Within the nearest 50–100 pc, geometric effects on the 2D Gaia transverse velocities make it increasingly difficult to detect stars with common 3D motions, making structures difficult to identify regardless of the recovery rate of their stars (see Section 3.4).

3.1.5. Generating Photometry of Simulated Stars and Systems

For each mass, metallicity, and age generated, we use isochrones to create corresponding Gaia G, G_{RP}, and G_{BP} magnitudes. We make use of PARSEC version 1.2S[6] isochrones made freely available online (see Bressan et al. 2012; Tang et al. 2014; Chen et al. 2015).

Note that these isochrones do not include the white dwarf cooling sequence; therefore, these objects are treated as "dead," with no photometric contribution. Since the photometry of these objects says little about the white dwarf's age relative to the formation of the system, this exclusion should have a minimal impact on the results. These objects also occupy a very different parameter space on the color–magnitude diagram (CMD), so they are easy to remove and nearly impossible to mistake for young main-sequence stars.

The isochrones cover a metallicity range evenly spaced between $[\mathrm{Fe/H}] = -1.0$ and 0.5 and a less regular age selection, which was designed to optimize the sampling of parameter space in the range of ages $6 < \log_{10}(\mathrm{age/yr}) < 10.049$. The main concern when generating a grid to draw our synthetic population from is that with a sufficiently sparse sample in age, stars of a given mass would not have samples in between, for example, the main sequence and giant branch. This can result in situations where any interpolation on these photometric grids can end up deriving photometry somewhere between the main sequence and giant branch, which may not follow the evolutionary track stars follow between those sequences. To minimize artifacts caused by undersampling, we require that for each slice of the grid in both mass and age, at least two points in age populate the horizontal branch, except for the most extreme massive stars. This requirement ensures that at least the lower red giant branch (RGB) is well covered by our grid, with the top of the RGB and horizontal branch being somewhat less well covered. Further grid resolution beyond covering the lower RGB would only increase the computational cost without improving our calculations, as these more evolved stars do not overlap with the pre-main sequence stars in photometric parameter space.

The PARSEC v1.2S isochrones give a list of masses for each metallicity/age combination, and we regrid these onto a grid with $0.11 < \log(M/M_{\odot}) < 20$ and nonuniform point density. For masses, very sparse sampling is sufficient for stars less massive than the main-sequence turnoff at the end of our age bracket, 11.2 Gyr. All stars below this point do not ever evolve off the main sequence; therefore, the only quick evolutionary process with sensitivity to mass that must be captured is the descent along the Hayashi track, which is significantly slower than evolution along the RGB and much less mass-sensitive (Hayashi 1961; Chen et al. 2015). Not many grid points are required to capture that evolution in the mass dimension, so only 100 linearly spaced grid points are used for masses $M \lesssim 0.8\ M_{\odot}$. We then use 4000 bins between 0.8 and 2.5 $M_{\odot}$, another 4000 between 2.5 and 10 $M_{\odot}$, and 600 more bins between 10 and 20 $M_{\odot}$, for a total of 8700 bins. These binning choices do result in the undersampling of the more evolved stars in the population; however, the pre–main sequence and stars sharing that parameter space are all well covered.

We assigned model stars random ages, metallicities, and masses according to the relevant distributions and linearly interpolated magnitudes for each star from our isochrone grid. When identified as part of an unresolved binary or triple system, we added the magnitudes of the component stars together, creating a merged photometric result for each unresolved system. This produced intrinsic synthetic G, BP, and RP magnitudes for any model star or system, resulting in the CMD presented in Figure 5. While this completes the generation of our model population, these intrinsic magnitudes have yet to be modulated by the conditions in which a real Gaia star might exist, most notably extinction, which we address in the next subsection.

3.1.6. Extinction

Significant interstellar extinction moves the main sequence of a stellar population on the H-R diagram to the right and downward of where it would be without that extinction.

[6] http://stev.oapd.inaf.it/cgi-bin/cmd

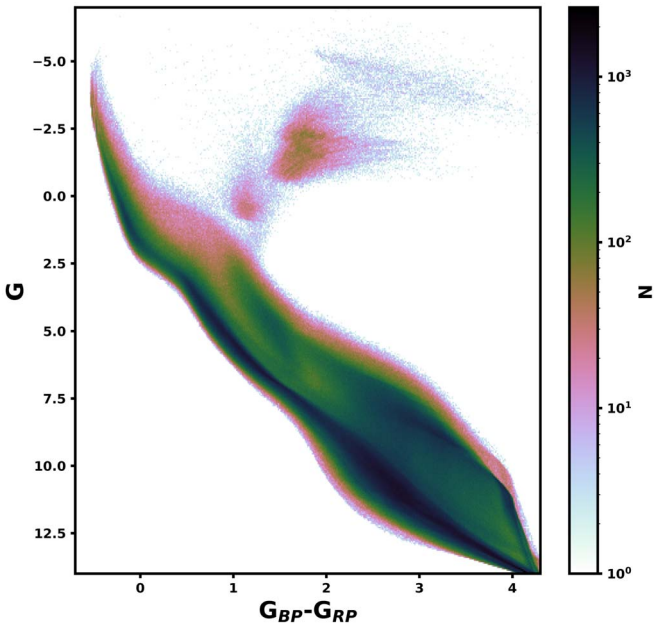

Figure 5. A CMD for the stars generated by simulating a population with our randomly generated ages, masses, metallicities, and stellar multiplicities, then computing observables using the PARSEC isochrones.

Consequently, heavily extincted stars can appear younger than they would without the effects of extinction. Due to the significant regional variations in extinction, the use of a single, directionally independent distribution to modulate our simulated photometry would fail to differentiate between young stars and those that are merely heavily extincted. Instead, we make use of 3D, all-sky extinction maps and generate a unique reddening distribution for each Gaia star based on that map. Extinction values from these distributions are generated and independently applied to the model population for each Gaia star.

We use the STILISM maps of reddening in the solar neighborhood from Lallement et al. (2019), which use Gaia DR2 and Two Micron All Sky Survey photometry and distances to compute reddening for a large volume centered on the solar system. We interpolate reddening parameters for each Gaia star based on its on-sky location using a Monte Carlo framework with distances drawn from the Gaia parallax and parallax uncertainty. This framework captures how Gaia distance uncertainties impact reddening uncertainties.

An additional uncertainty modulation is added to each interpolated reddening result to reflect the inherent uncertainty in the maps themselves. Lallement et al. (2019) reddenings often have very asymmetric uncertainties, so for each generated reddening and corresponding uncertainties, we select a random value from a normal distribution, multiplied by the uncertainty on that side of the reddening value recorded in the STILISM maps. These uncertainty results are added to each randomized reddening, allowing us to generate probability histograms for the reddening of all stars in the Gaia set. The reddenings of the simulated stars are therefore drawn from these histograms separately for each Gaia star. The relationships presented in Wang & Chen (2019) were used to relate $E(B-V)$ to extinction in the Gaia *BP*, *RP*, and *G* filters. While these conversions are expected to differ somewhat by the effective temperature of the target, these effects will be negligible compared to the intrinsic sources of uncertainty in the Lallement et al. (2019) reddening maps, especially given that the Wang & Chen (2019) reddening conversions are based on relatively cool red clump stars that are photometrically similar to many of the pre-main sequence stars we focus on in this paper. The addition of reddening completes a population of simulated stars with a reddening distribution reflecting the reddening expected of each Gaia star based on its location.

3.1.7. Generating Star Statistics

The completed population of sample stars allows us to generate statistics for our population of Gaia DR2 stars. For each real star, we use Bayes' theorem to generate the probability that each sample star is consistent with the Gaia photometry according to the following formulation, as implemented in Huber et al. (2016):

$$p(y|g) \propto p(y)p(g|y) = \prod_i \exp\left(-\frac{(g_i - x_i)^2}{2\sigma_{g,i}^2}\right), \quad (7)$$

where y is a sample star with inherent properties $y = \{\text{age}, [\text{Fe/H}], \text{mass}, E(BP-RP), \text{multiplicity}\}$ and observables $x = \{G, BP, RP\}$. Age, [Fe/H], mass, and multiplicity are all unknowns, while the extinction $E(BP-RP)$ is a prior imposed by the reddening solutions from Lallement et al. (2019). We then compare the observables of our sample stars to our Gaia star g with the same observables, converting from apparent Gaia magnitude to absolute magnitude using the Gaia parallax measurements.

These distances were generated by inverting the parallaxes, and while this is an imperfect estimate for high relative uncertainty parallaxes, the distance estimates from Bailer-Jones et al. (2018) follow the inverted parallax very closely for the stars with $\pi/\sigma_\pi > 10$ that we investigate here. The subscript i multiplies over each observable in x, and $\sigma_{g,i}$ is the uncertainty in the Gaia measurement of the observable i, including both photometric and parallax uncertainties.

In most cases, the prior $p(y) = 1$, owing to the consistency of our metallicity, mass, reddening, and binarity rates with true values. Age, however, does require a nontrivial prior, as we significantly oversample young stars. We therefore add a prior equal to the age of the star, which nullifies the exponential bias cause by the uniform sampling in log space, producing a flat probability distribution in age for the sample stars. To produce any probability distribution, we simply sum the $p(y|g)$ values for each model star into a histogram and normalize over all sample stars. The probability of any given condition on a property, such as $P_{\text{age}<50\text{ Myr}}$, can therefore be found by summing over the relevant bins. The resulting $P_{\text{age}<50\text{ Myr}}$ values for each star in the Gaia sample are presented through the CMD in Figure 6. Most stars have $P_{\text{age}<50\text{ Myr}}$ similar to their neighbors, with some exceptions in cases with larger uncertainties or cases where either the BP or RP filter is discrepant, which may not translate well to this 2D plot from the 3D input magnitude data set.

All Gaia stars with less than 50 simulated stars within 1σ of them in color–magnitude space were culled from the sample, as we identify them as inconsistent with the stars in our simulated population. As we do not consider white dwarfs and white dwarf–main-sequence binaries in our model, most of these stars are excluded, as is evident from the near-absence of a white dwarf cooling sequence in Figure 6. For all Gaia sources that

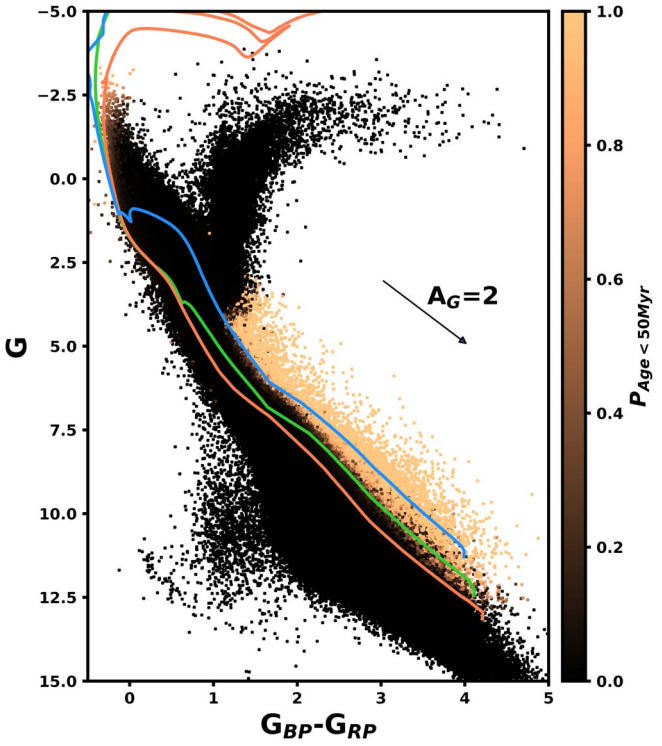

Figure 6. A CMD containing all Gaia stars with more than 50 simulated stars in our sample within 1σ of their locations in the CMD. All magnitudes are absolute and dereddened. The colors represent the probability that the star is young, defined by this paper as younger than 50 Myr. The reddening vector for $A_G = 2$ is labeled, showing the direction along which older stars may move to interfere with the pre–main sequence if not dereddened. Isochrones are also included for ages of 5, 20, and 50 Myr. The orange region with high $P_{\text{age}<50\text{ Myr}}$ values intersects with the 5 Myr isochrone from M dwarfs up to near the subgiant branch, while the 50 Myr isochrone only intersects for extremely low mass M dwarfs, indicating a higher completeness for younger populations (see Section 3.2).

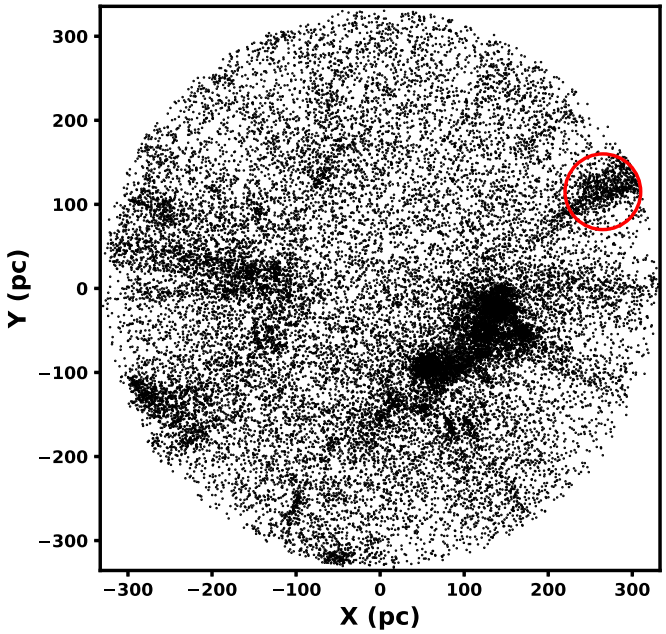

Figure 7. All credible young stars identified by our search to have $P_{\text{age}<50\text{ Myr}} > 0.1$, plotted in X/Y galactic coordinates. The circled region is an example of a reddening anomaly, where the underestimated reddening causes stars to be marked as young. That particular region is in the direction of heavy reddening in Aquila and Serpens. This group is not identified in the clustering stage (see Figure 10), as the stars do not have common motions and are likely dominated by reddened field stars.

are well sampled by our model population, we generated probability distributions for age, mass, metallicity, and multiplicity. To extract a population of young stars from our sample, all objects with $P_{\text{age}<50\text{ Myr}} < 0.1$ were removed, leaving a population of 30,518 credible young star candidates out of the full sample of nearly 5 million Gaia sources. These candidate young stars are presented in XY galactic spatial coordinates in Figure 7 and compiled into a master candidate list in Table 1. As a result of the reddening vector direction shown in Figure 6, the underestimation of the reddening, especially near the subgiant branch, does have the ability to make older stars appear young. Minor anomalous reddening-related clumps do appear. While we do detect minor spatially clustered but kinematically scattered clumps of stars that appear consistent with anomalies from local reddening underestimates, many more young structures are visually identifiable, such as Sco-Cen, Orion, Perseus, and Taurus.

3.2. Recovery

The method we employ in this paper searches the solar neighborhood for stars with a significant probability ($P > 0.1$) of an age less than 50 Myr, and we should therefore have demonstrable capabilities to identify stars in that age range. Here we investigate the efficiency with which those nearby young stars are identified and what kinds of stars are detectable for groups with different ages. There are two main ways in which stars might be missed in this survey: failure of Gaia photometric or astrometric quality cuts and misidentification by our pipeline as old. In this section, we quantify the losses due to each of these factors and provide insight into why some of these losses take place.

The Gaia quality cuts, which remove stars based on factors like goodness of fit to the Gaia photometric and astrometric models, parallax error, and the number of Gaia visibility periods used (see Section 2.1 for a full explanation of these) are necessary to ensure that our Gaia stars have reliable photometry and astrometry. However, it is nonetheless useful to investigate the properties of the stars lost to reveal any potential biases the cuts might impose. By comparing the samples of stars that pass and fail these quality cuts, we can reveal what properties are most likely to lead to the rejection of a star based on quality. We use a large sample of Taurus members drawn from Krolikowski (2021) to make this comparison, which has a wealth of information on ~500 stars in the region, including spectral type, extinction, and binarity. All of these factors may influence the quality of any Gaia astrometric or photometric solutions. We find that our Gaia quality culling appears to correlate very little with stellar spectral type over the late G to M range; however, those cuts do correlate strongly with binarity and extinction. Unresolved binaries are almost twice as likely to be removed by Gaia quality restrictions, with 59% of singles passing these cuts, compared to only 31% of multiples. Binaries are known to both introduce additional astrometric noise and skew the photometric solution (e.g., Arenou et al. 2018), so their more frequent failure of quality cuts is expected. For heavily reddened populations with $A_V > 3$, only 20% of sources pass the quality cuts, compared to 67% for $A_V < 0.5$. This heavy loss rate in highly reddened locations is consistent with our results later in this paper, which include visibly

Table 1
The Full Catalog of All Gaia Objects We Identify with $P(\text{age} < 50\ \text{Myr}) > 0.1$, with Basic Gaia Properties, $P(\text{age} < 50\ \text{Myr})$, and Age as Derived in Section 3.3 for Each

ID	TLC[a]	EOM[a]	LEAF[b]	Gaia ID	R.A. (deg)	Decl. (deg)	π (mas)	g	$bp - rp$	$P(\text{age} < 50\ \text{Myr})$	Age[c] (Myr)		
											Val.	+ Err.	− Err.
0	−1	−1		3239529836638867840	73.8034	4.7175	4.170	17.80	3.50	0.34	28.3	8.4	5.7
1	22	11		6099439604319602560	217.3498	−44.7853	5.960	15.60	3.00	0.94	12.0	2.7	3.0
2	−1	−1		4154795132763964032	278.9114	−10.8489	3.020	14.40	2.40	1.00			
3	21	4		5774202930948040064	266.3687	−82.1980	6.260	16.10	3.10	0.90	18.2	5.6	4.8
4	−1	−1		3442403338420195840	82.6082	28.2391	3.460	11.90	1.40	0.23			
5	17	−1		3442418800302357248	83.0898	28.3245	5.220	16.10	3.00	0.89	15.6	3.3	5.3
6	−1	−1		4154777227001098112	279.5766	−10.6085	3.100	14.30	1.90	0.15			
7	−1	−1		1937498784187322624	351.8766	45.0985	3.830	16.70	2.90	0.24	22.8	9.6	4.3
8	−1	−1		6195524512419762816	204.7577	−21.6912	11.930	13.30	2.60	0.56			
9	−1	−1		1716855559591213824	192.1970	78.9078	7.670	15.10	2.90	0.46	19.9	4.8	4.8

Notes. The TLC, EOM, and LEAF columns mark the top-level, EOM, and leaf clusters, respectively, that each star is a part of, if any. Only a small subset of the 30,518 total objects is shown here.
[a] Objects with $\pi/\sigma_\pi < 25$ were not subjected to clustering and have TLC and EOM markers of zero. Objects not assigned to a cluster are given the marker −1.
[b] Left blank if not related to a leaf cluster.
[c] Left blank if no solution reached.

(This table is available in its entirety in machine-readable form.)

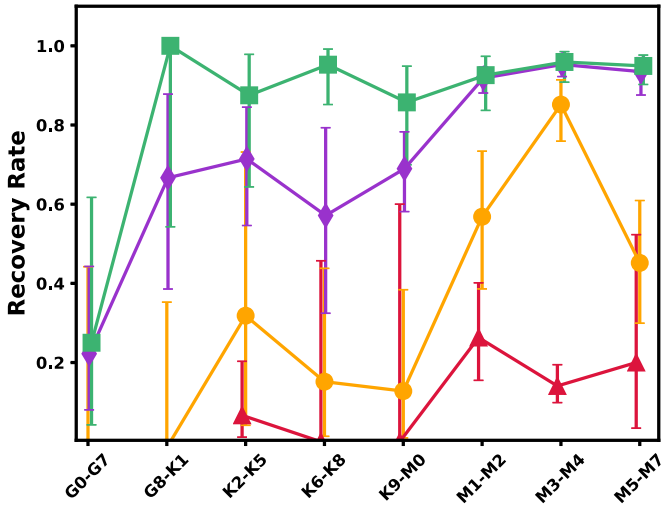

Figure 8. Recovery rates after the imposition of Gaia quality cuts in Taurus (green squares), US (purple diamonds), β Pic (orange circles), and Tucana-Horologium (red triangles), binned by spectral type. In all bins, the order of the groups in recovery rate matches that of their relative ages, with the essentially newborn Taurus association having the most complete recovery. The error bars display the binomial 68% confidence interval.

incomplete samples of young stars in active star-forming sites such as Perseus and Lupus.

Next, we investigate recovery rates from our pipeline for the identification of young stars, which include only objects that pass the Gaia quality cuts. To calculate recovery rates, we gather lists of known members for four well-known associations: US (in Sco-Cen), Taurus, the β Pictoris moving group, and the Tucana-Horologium moving group. These groups cover nearly the full range of detectable cluster ages, spanning from Taurus, which is nearly newborn (<5 Myr; Kraus & Hillenbrand 2009), to Tucana-Horologium, which has an age of approximately 45 Myr (Kraus et al. 2014; Bell et al. 2015). We can therefore use these stars as a strong representative sample of the populations we expect to identify in this paper.

As the youngest population we investigate in this section, stars in the Taurus association should be the easiest to identify reliably as young through purely photometric methods. To verify this, we drew known members from the Esplin & Luhman (2019) Taurus catalog, which compiled verified members from the literature and introduced several new members. These members are focused on the young central subgroups in Taurus (i.e., Greater Taurus groups 8–11 in Section 4.1), making them a more homogeneous sample for comparison relative to the more diverse sample from Krolikowski (2021), which includes both the objects in Esplin & Luhman (2019) and those in more peripheral regions, which may be older. Esplin & Luhman (2019) identified members as candidates using proper motions and photometry and then confirmed them using spectroscopic observations including Li I absorption. We found that 189 of 351 (54%) stars in Taurus have Gaia detections that survive the quality cuts described in Section 2.1. While these restrictions did result in the loss of nearly half of the members, it is important for the reliability of our results that objects with poor Gaia data are excluded. Of the stars that survived the quality cuts, 172 of 189 were identified as young, or 91%. We compile the recovery rates as a function of spectral class for Taurus and the other three test associations in Figure 8, where we show that the recovery rate in Taurus is essentially complete for spectral classes M and K, with a sharp sensitivity decline for G-type stars. This lowered sensitivity is caused by the photometric overlap between pre-main sequence stars earlier than about G8 and the subgiant branch on the H-R diagram, which significantly reduces the probability that a star there is young.

The US association is slightly older than Taurus (age ~5–11 Myr; Preibisch et al. 1999; Pecaut et al. 2012; Rizzuto et al. 2015), and we therefore expect less sensitivity to early-type stars as members begin to merge onto the main sequence. The US sample comes from Rizzuto et al. (2015), Preibisch et al. (1998, 2002), who make use of spectroscopic observations of Li and Hα, as well as complete 3D motions to confirm the membership of the stars. A total of 318 members out of the total list of 477, or 67%, had a high-quality Gaia counterpart that passed our quality cuts. Of those high-quality Gaia detections of known US members, 274 were identified as young by our method, or 86%. As presented in Figure 8, our completeness is best for the lowest-mass stars, which provide essentially complete results for M dwarfs up to about M1. For stars earlier than M1, the sensitivity drops gradually, with the same significant sensitivity drop we observed in Taurus appearing beyond spectral type G8.

To check the nature of the contamination in our sample, we also investigated a subset of the Rizzuto et al. (2015) catalog, which was identified as kinematically and photometrically consistent with US but rejected as young members due to the absence of a lithium feature. Since our method is purely photometric, we identified most of these false positives as young candidates (71%); however, we also found that these stars have disproportionately high RUWEs, with 53% having RUWE > 1.1, compared to 43% for our larger sample of candidate young stars. This is consistent with most of these being spectroscopic binary field interlopers, which appear to be the dominant source of contamination in our sample.

The older β Pictoris moving group (age ~ 23 Myr; Mamajek & Bell 2014) hosts a more elusive population, as these stars sit on a pre–main sequence that is less well removed from the photometry of background stars. To check the recovery rate in β Pic, we used the catalog from Shkolnik et al. (2017), which identified members using both 3D motions and the presence of Li and Hα. Ninety-four of 173 stars in the association, or 54%, survived our Gaia quality cuts. The rate at which we recovered these young stars was considerably lower, however, with 44 out of 94 members being recovered, or 47%. Rather than there being a sensitivity limit around G8, the brightest stars identified in β Pic have spectral types in the early K, caused primarily by the more luminous stars having already settled close to the main sequence. Later K stars showed low recovery rates, suggesting that they are usually only marginally identified (see Figure 8). Significantly better recovery rates were observed for mid-M stars, peaking at approximately 80% for M3–M4, suggesting that our method remains effective at consistently identifying the less massive stars that remain well above the main sequence even at older ages.

The Tucana-Horologium association is near the 50 Myr target age limit with an age of ~45 Myr (Kraus et al. 2014); therefore, we should expect a significantly lower sensitivity to this population. The sample population of Tuc-Hor members that we used comes from Kraus et al. (2014), which, like for our β Pic population, used a combination of 3D kinematic information and stellar youth indicators to verify membership. Of the Tuc-Hor members, 80% (or 115 out of 143) survive the

Gaia quality cuts. As expected, our recovery of Tuc-Hor members showed a significant decay in the recovery rate of stars overall, with no noteworthy sensitivity for stars earlier than a spectral class of about M1 (see Figure 8). The recovery rate for the later M dwarf bins in Tuc-Hor is around 20%, which, while significantly weaker than our recovery for younger stars, does still identify 17 Tuc-Hor members out of the complete sample of 115 across all spectral types, or 15%, enough to make the population potentially identifiable.

The recovery rates among stars with quality Gaia detections in each of the four regions discussed here are compiled in Figure 8, showing clearly how our recovery rates vary with spectral type and the age of the stars being observed. There we show how the effectiveness of our method at a given spectral class negatively correlates with the age of the region, with the nearly newborn Taurus association being essentially complete from G8 to late M excluding the quality cuts and the 45 Myr old Tuc-Hor moving group having extremely limited sensitivity essentially restricted to M dwarfs. For ages in between, sensitivity gradually falls off, with G and then K stars being lost first and M dwarf sensitivity beginning to drop past ∼20 Myr as for our β Pictoris sample. For Tuc-Hor, an association near our target age limit, we maintain a recovery rate of approximately 20% among M stars, while only one earlier star is identified. Despite the restrictions on our sensitivity for these older groups, our ability to detect nonnegligible numbers of members in even the relatively old Tucana-Horologium moving group suggests that large enough groups will remain identifiable up to the upper edge of our target age range at 50 Myr.

Despite the successful stellar recovery among these older populations, the composition of the stars recovered should be treated with caution, as we expect stars with features that artificially inflate a star's photometric youth, such as unresolved binaries, to be detected much more easily than an average star in the samples included in this section. With an overall recovery rate around 15%, as for Tuc-Hor, unresolved binaries may begin to dominate the recovered sample. Particularly in these older populations, demographic studies can be greatly expanded by using the locus of an association in space-velocity coordinates as defined by the limited sample we identify as young to locate and reintroduce probable members that are not recognized as photometrically young by our pipeline. The reintroduction of photometrically older probable members can not only expand the completeness of the original sample but also help to suppress biases produced by the potential overrepresentation of binaries, such as in the derivation of group ages, an application we explore in Section 3.5.

3.3. Age Estimation

The primary objective of generating our model population and corresponding posterior distributions was the identification of young stars in the presence of diverse stellar populations in the solar neighborhood. However, the age distributions that we generate can also be used to derive approximate ages for most of the young stars in our sample. For candidate young stars, the age probability distribution from our pipeline typically consists of a peak at a young age and a significant tail to older ages populated by main-sequence model stars with bright unresolved companions. A Gaussian least-squares fit to that peak would provide a suitable first-order age estimate, emulating the mode of the distribution while avoiding influence from the high-age tail that may dominate estimates from the mean or median (see Figure 9). While these Gaussian fits are effective in representing the age distribution peaks, the wide range of properties considered in these distributions, especially multiplicity, make these peaks unlikely to reflect a star's true isochronal age. Much more accurate ages can be derived by limiting the range of metallicities and multiplicities that we consider and using a population of model stars with known properties to create a map connecting Gaussian least-squares fits to the age and mass probability distributions with a corresponding true age from the model.

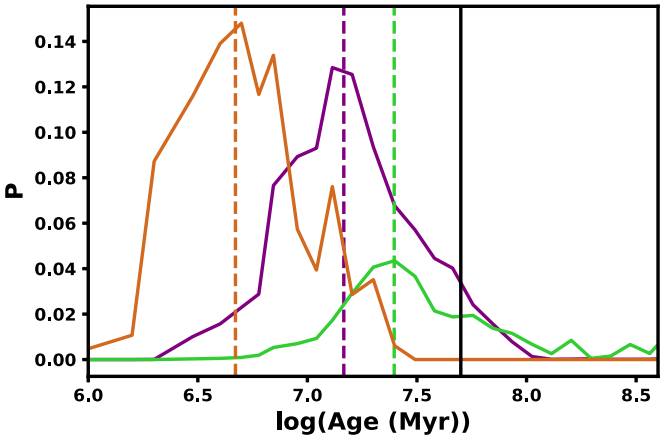

Figure 9. Log(age) probability distributions for three young star candidates. The peaks of the Gaussian fits, which are used in age estimation, are marked with vertical dashed lines, and they closely reflect the mode of the distribution. The reduction in the peak height for the older sources is caused by the presence of a progressively larger high-age tail to older solutions that continues far to the right of the distribution shown. A larger tail reduces $P_{\mathrm{age<50\ Myr}}$, which is defined as the sum of all bins younger than 50 Myr (marked by the vertical solid line).

While binaries are very common, potentially representing a majority of young systems (e.g., Raghavan et al. 2010; Kraus et al. 2011), their companions introduce a very wide range of photometric excesses, depending on the mass of the companion. As such, photometry is a much less consistent age diagnostic for multiple systems compared to stars without a companion. We therefore exclude binaries and higher-order systems from our individual age estimates, subsequently assuming that all Gaia sources have no companion. For single stars, the lack of consideration for binaries will ensure that our results are not influenced by the wide range of photometric solutions possible from the presence of an unresolved companion, making the age estimate of high quality. If a star does have a companion, it will contribute to a background of stars with younger age fits that are not consistent with one another, making them possible to clip out when looking at populations.

We also fix the assumed metallicities of our candidate young stars to the solar value. This assumption has been shown to be appropriate for the vast majority of nearby young clusters and associations (e.g., Viana Almeida et al. 2009; Mamajek 2013), which have a uniformity in their metallicities consistent with the thorough mixing of the interstellar medium (ISM) on 100 pc level scales (e.g., de Avillez & Mac Low 2002). Table 3 in Viana Almeida et al. (2009) presents metallicity values for most nearby young associations, with all solutions falling between $[\mathrm{Fe/H}] = -0.13$ and $+0.04$, mostly within one

standard deviation of solar. The α Persei cluster is one of the most metal-rich known groups accessible through our search, and it has a metallicity of only [Fe/H] = +0.18 (Pöhnl & Paunzen 2010). Due to the mixing of the local ISM, we would not expect any of our groups to depart significantly from the trend of consistently near-solar metallicities for young nearby associations and clusters. Unlike these young groups, older clusters have a more diverse set of metallicities reflective of those in the rest of the local stellar population (Buder et al. 2018; Hayden et al. 2019; see Figure 2). The Hyades and Praesepe have notably supersolar metallicities at [Fe/H] = +0.146 and +0.21, respectively (e.g., Cummings et al. 2017; D'Orazi et al. 2020), while numerous other clusters are much more metal-poor, such as M35 and NGC 2506 at [Fe/H] = −0.21 and −0.52, respectively (Friel & Janes 1993; Bouy et al. 2015). This contrast in the metallicity variation between the youngest stellar populations and various older populations in the solar neighborhood strengthens the decision to restrict the metallicity to solar for our in-depth look at young stars while assuming a much broader prior metallicity distribution for our wider population.

To extract age estimates from our output age probability distributions, we feed model stars with known properties through our pipeline for posterior distributions and generate Gaussian least-squares fits to the output age distributions. We generate these model stars at fixed solar metallicity, chosen at random with uniformly distributed ages between 1 and 50 Myr and masses from 0.1 to 1.3 $M_{\odot}$. This mass limit roughly corresponds to where stars at 1 Myr stop being clearly identifiable on the pre–main sequence as they cross the subgiant branch. The photometry of these test stars is allowed to vary according to the mean Gaia photometric uncertainty and mean reddening uncertainty in the group, providing the expected photometric variation for an individual star in that group. As for the Gaia sources, Gaussian distributions are fit to the resulting output age and mass distributions for each test star, and the input ages are binned according to the output fits for mass and age. For very young stars where the peak is in the youngest age bin, a Gaussian least-squares fit will often view the probability distribution as the tail of an exponential with a peak below zero; therefore, we exclude the 1 Myr bin that often creates these anomalous fits, limiting the ages considered to between 2 and 50 Myr. We compute a median input age and standard deviation for the true ages of stars in each fit age and mass bin, which completes a map that links any given set of age and mass results from the main pipeline to a revised age result excluding any skewing effects related to metallicity and multiplicity.

The vast majority of corrected age solutions are younger than those derived from the peak of the Gaussian age distribution by less than ∼30%–40%, which is expected given that the addition of binaries adds a range of older possible solutions to the photometry of the target. Some of the older (age > 30 Myr) and more massive stars near the limit of our method's sensitivity occasionally require a larger correction to the age of up to a factor of 2. A total of 25,727 of the original 30,518 credible young candidates generate quality age fits, or 84%. Most of the stars without good fits have photometric ages below 2 Myr, a category that includes both genuinely young stars and objects lowered into that photometric age bracket by unresolved binarity or insufficiently corrected reddening. We also lose some objects near the limits for detectability where stars begin to merge into the main sequence. Upon comparing the isochrones associated with these new revised ages to the photometry of these stars, we find that the age results are consistent with our isochronal model. Within Sco-Cen, many of the refined ages within subclusters host uncertainties comparable to internal variation of the age between members, which would be expected assuming a common formation time.

These age estimates assume single stars with solar metallicity, so individual age estimates should be treated with caution when those details are not known. The interference from binaries can, however, be reduced by considering groups of stars rather than individuals. Therefore, when we search for age gradients within a group, rather than considering the age fits of all stars individually, we consider the median of the 10 nearest neighbors in space-velocity coordinates. This is not a perfect solution for a group of stars, as our young star candidates are typically photometrically younger, on average, compared to the population as a whole (see Section 3.5); however, it is nonetheless a useful and efficient way to identify and visualize age patterns in a population.

3.4. Clustering with HDBSCAN

The presence of structures in our young stellar population is clear through the overdensities visible in Figure 7; however, a consistent means of structure identification is necessary for a proper analysis. To this end, we employ Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN; McInnes et al. 2017), a hierarchical density-based clustering algorithm developed as an extension of the frequently used DBSCAN clustering algorithm (Ester et al. 1996). It has already been used quite successfully in the literature for a variety of purposes, including in Kounkel & Covey (2019), where it is used to identify tenuous stellar structures of all ages in their own Gaia DR2 sample. As such, there is a strong precedent for its use in identifying groups and associations within large stellar populations. Unlike other clustering algorithms, which typically presuppose a density threshold, shape, or some other expected cluster property, HDBSCAN uses only a minimum cluster size and a parameter with a smoothing effect on the density distribution. As such, it is an algorithm with next to no built-in assumptions, allowing for young structures to be identified without discrimination based on complicating factors such as unusual shapes or wildly varying scales. Furthermore, because the algorithm searches for structures at different scales, it also provides a unique opportunity to explore different levels of substructure within an association. Lastly, HDBSCAN enables clustering in an arbitrary number of dimensions, allowing for both *XYZ* spatial coordinates and *l* and *b* transverse velocities to be included in clustering simultaneously. In some locations with high localized reddening, the Lallement et al. (2019) maps are insufficient to properly correct for reddening, resulting in anomalous overdensities without common velocities. Through the inclusion of velocities in our clustering, we are able to exclude such anomalies.

The search for overdensities by HDBSCAN makes use of a parameterized density metric, which is based on the distance to the *k*th nearest neighbor (referred to as the core distance, or core_k(object); McInnes et al. 2017), where *k* is the first of two key HDBSCAN input parameters.[7] Increasing *k* effectively

[7] Given by the HDBSCAN clustering parameter min_samples.

smooths the density distribution and causes smaller clumps of stars to be ignored, therefore serving as a proxy for how conservative the clustering will be. The HDBSCAN algorithm generates a new mutual reachability distance metric from these core distances, which for objects a and b is defined as $d_{mreach-k}(a, b) = \max\{\text{core}_k(a), \text{core}_k(b), d(a, b)\}$, where $d(a, b)$ is the distance from object a to object b. Using this metric, HDBSCAN generates a minimal spanning tree, which is essentially a web linking each point in the data set through their nearest neighbor according to their mutual reachability distances. By removing links in the tree of sources with the lowest weights, corresponding to large mutual reachability distances to their neighbors, links connecting clusters in the web are broken, leaving behind a hierarchy of clusters that appear and fragment as the weight threshold is raised.[8]

The methods employed by HDBSCAN are comparable to those of DBSCAN, which defines the cores of clusters by the presence of more than some number of neighbors within a radius ϵ, effectively taking only clusters that exist at some arbitrary chosen scale (Ester et al. 1996). In HDBSCAN, rather than looking for clusters at a single ϵ, clustering is simultaneously extracted at all scales, allowing the most persistent clusters to be identified, rather than those that happen to emerge at any requested scale (McInnes et al. 2017). Any clusters that come out of the tree with less than N members, where N is the second key HDBSCAN input parameter,[9] are thrown out and merged into their parent cluster: a larger-scale node on the clustering tree. The quality of clusters in the tree is set by their persistence, which relates to the range of scales over which the cluster is defined in the clustering tree. The HDBSCAN algorithm also includes the value ϵ as an optional parameter,[10] setting a minimum threshold past which clusters can no longer be fragmented. This is useful in cases where large groups with significant substructure are present, as it allows for the entire region to be identified as a whole, without excessive fragmentation. There are two options for cluster selection: excess of mass (EOM), which selects the best or most persistent clusters in the clustering tree, and leaf, which identifies only the nodes of the clustering tree, effectively selecting clusters from a maximally fragmented clustering solution (McInnes et al. 2017). In most cases, EOM clustering is preferred, as the persistence parameter that defines an EOM cluster is used by HDBSCAN as a proxy for cluster quality, although leaf clustering can be useful when attempting to detect subclusters within a well-defined higher-level cluster. A more complete look at the methodology of HDBSCAN can be found in McInnes et al. (2017), along with excellent visualizations of the tree clustering it implements.

An additional consideration is the incorporation of velocities in clustering. Since transverse velocity in kilometers per second and galactic XYZ coordinates in parsecs do not have matching scales, a corrective factor c is required to reflect the differing standard deviations of groups between distance and velocity coordinates. We therefore apply HDBSCAN to the following 5D data set:

$$(X, Y, Z, c * v_{T,l}, c * v_{T,b}). \quad (8)$$

The known young associations often have wildly different ratios between spatial extent and velocity dispersions; however, ratios of 4–6 pc/km s^{-1} are typical, approximately reflecting the ratios predicted for cloud scales of 25–100 pc in the Larson's law relation (Larson 1981). While Larson's law was designed to relate velocity dispersions to scales in clouds, associations are expected to inherit properties from their parent cloud and maintain those properties for some time after formation, especially in loose associations where dynamical interactions are infrequent (Larson 1979; Wright & Mamajek 2018; Ha et al. 2021). Size–velocity ratios of 4–6 pc/km s^{-1} predict properties consistent with known groups such as US and, to a lesser extent, the more dispersed regions of Upper Centaurus–Lupus (UCL) and Lower Centaurus–Crux (LCC; Wright & Mamajek 2018). This range of factors is also consistent with the relative sizes of the visually identifiable Gaia clumps in spatial coordinates and transverse velocity space. Any variation within this range can be interpreted as a choice to tweak the relative weight of velocity and spatial coordinates in the clustering, where a larger corrective factor weighs velocity more heavily by reducing the relative variability of the spatial component.

For all-sky clustering, a size–velocity corrective factor of $c = 6$ pc/km s^{-1} was chosen, which weights kinematics slightly more heavily, reflecting the larger velocity dispersions expected in the larger groups, such as Sco-Cen, that we wish to identify at this level. A relatively large ϵ value of 25 in units of parsecs (or c km s^{-1} for velocity) supplements this choice, preventing fragmentation of groups below a scale comparable to a moderately sized association like US (Wright & Mamajek 2018). This choice ensures that groups like Sco-Cen can be treated like coherent singular groups while also allowing stellar structures that link subtly separated groups to be detected, which would otherwise be incorporated into the background if those groups were recorded separately. While groups being linked by this condition does not guarantee that they share common formation origins, it does expose spatial and kinematic similarities that can be further explored in subclustering (see Section 5.4). The input parameter N, representing the minimum cluster size, was set to 10, as was k, which is set equal to N by default. This choice avoids the inclusion of negligibly small clusters or high-order multiple systems, producing a set of largely visually convincing clusters. However, some groups with less internally consistent velocities or spatial distributions are also included, so we set a cluster persistence cut at a value of 0.015. This number was chosen to ensure that no groups with velocity distributions comparable to the field were included, most notably the reddening anomaly toward Aquila marked in Figure 7, which is dense enough in space to be marked as a cluster without any persistence-based quality restrictions. Through this choice, our clustering results in this implementation are made to be intentionally conservative, with the objective of highlighting noteworthy structures in the solar neighborhood.

We applied HDBSCAN using the above input parameters to members of our population of probable young stars, with the additional requirement that $\pi/\sigma_\pi > 25$. Clustering analyses in particular benefit from well-constrained distances, which prevent the erroneous merging of groups stretched along a similar line of sight, and this cut restricts the expected distance spreads down to about 13 pc, a scale comparable to those of many of the smaller groups in the solar neighborhood. A total

[8] An excellent visualization of this process is given on HDBSCAN's website at https://hdbscan.readthedocs.io/en/latest/how_hdbscan_works.html.
[9] Given by the HDBCSAN clustering parameter min_cluster_size.
[10] Given by the HDBSCAN clustering parameter cluster_selection_epsilon.

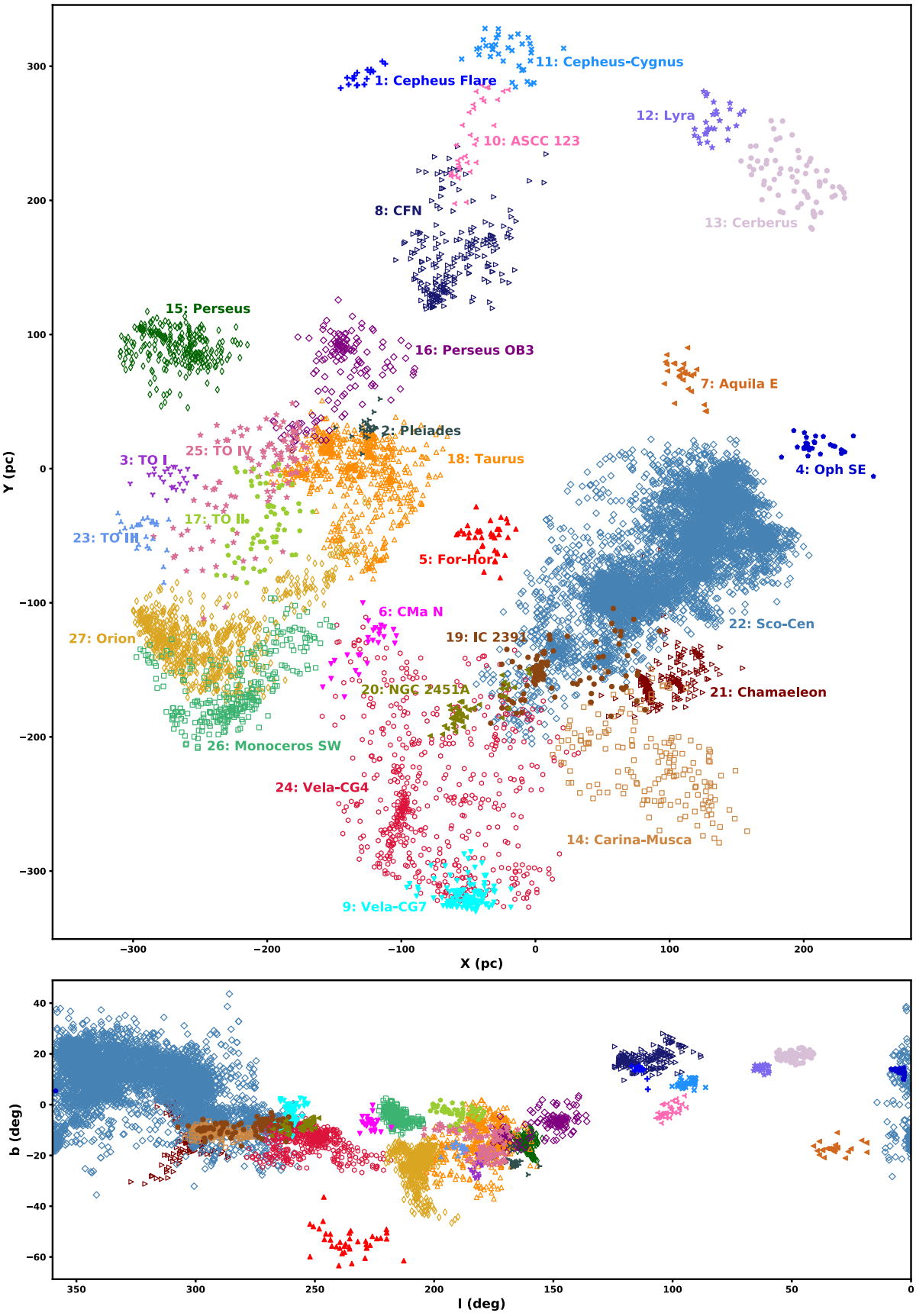

Figure 10. Results of the all-sky clustering analysis from HDBSCAN in X/Y galactic spatial coordinates (upper panel) and l/b sky coordinates (lower panel). Each color corresponds to a separate top-level group identified by HDBSCAN, which are labeled according to the IDs and names in Table 3. Groups with substructure are shown using open symbols, which are given complete subclustering analyses in Sections 4.1 (Taurus), 4.2 (Orion), 4.3 (Perseus), 4.4 (Sco-Cen and Chamaeleon), and 4.5 (all others). See Figure 7 for the X/Y distribution of stars prior to clustering.

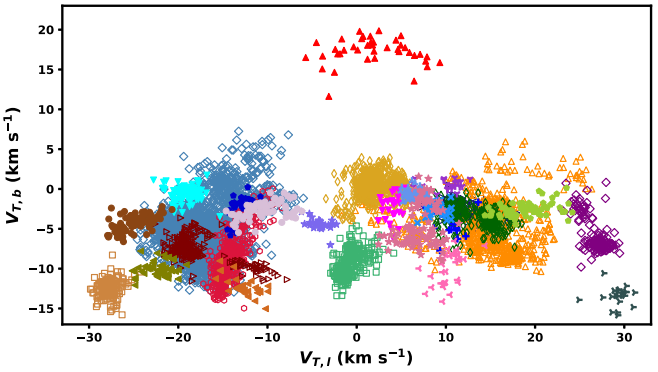

Figure 11. Same as Figure 10 but for transverse velocity.

of 28,340 stars entered into the final 5D data set, a 7% reduction relative to the sample before the additional parallax restriction. Twenty-seven spatially and kinematically distinct young associations from this population, which are shown in *XY* galactic spatial coordinates, l/b sky coordinates, and l/b transverse velocities in Figures 10 and 11, were identified by HDBSCAN. These clustered sources account for 41% of the total young population we identified, with the remaining sources likely being composed of a mixture of young stars ejected from larger associations, young stars in tenuous unrecognized associations, and false positives enabled by reddening anomalies and unresolved multiplicity. Sco-Cen was easily detected, representing by far the largest group with just under 7400 stars, nearly an order of magnitude larger than the next largest group, the near edge of Orion. Sco-Cen appears alongside a series of nearby but spatially distinct structures, such as the dense clusters in Chamaeleon. Perseus, Taurus, and the near edge of Orion are also identified in at least some form, with varying levels of internal division, along with some smaller known groups, such as Perseus OB3 and the near edge of Vela. Many of the remaining groups appear to be previously unknown.

To facilitate a more detailed study of Sco-Cen and other top-level groups, we performed an additional round of clustering on groups that showed visible substructure; however, the use of HDBSCAN there was modified slightly to enable the recognition of hierarchical structures in each region. Rather than identifying only the EOM clusters, as was done for top-level clustering, we also extract the smaller-scale leaf clusters, which might exist within a larger EOM cluster. The approach we employ fuses the two clustering methods, recognizing the EOM clusters containing leaf clusters as intermediate-level groups and the leaf clusters as subclusters within that larger EOM group. For this clustering, N and k were kept the same, while the space-velocity corrective factor c was reduced to 5 to reflect the smaller scales, and ϵ was removed entirely, allowing for subclusters to be identified as separate regardless of the presence of nearby groups. The goal of this lower-level clustering is to identify all potential subgroups in a structure, regardless of scale or membership as part of a larger group. Ten of our 27 top-level groups have substructure, and within those 10 groups, this implementation identifies a total of 60 EOM clusters, with eight of those 60 further subdividing into a total of 27 additional smaller leaf clusters. We present the top-level, EOM, and leaf clustering results for each candidate young star in Table 1, alongside $P_{\mathrm{age<50\ Myr}}$ and the age solutions derived in Section 3.3.

3.5. Group Ages

The ages we derive for entire groups of stars cannot be calculated in the same way as was done for individual stars, as the population we consider during clustering is, by design, skewed toward photometrically younger objects. In groups of fixed age, photometrically younger stars include objects with unresolved companions or, more generally, stars on the high side of the expected photometric variation. This means that group ages calculated using our likely young stars alone will be skewed young, especially in groups older than ∼20 Myr, where only a limited section of the pre–main sequence is reliably detected as young (e.g., see Figure 6). We can reduce this effect by reintroducing stars that our pipeline did not identify as young and fitting an isochrone to a cleaned subset of the expanded population. This can be done by searching for objects that have spatial coordinates and kinematics consistent with the known members of that group, at least as close to the 10th-nearest HDBSCAN-identified member in space-velocity coordinates as the most peripheral HDBSCAN-identified member, mimicking the original HDBSCAN methods for finding members of clusters with $k = 10$. This approach effectively uses the groups in the original clustering analysis as signposts from which more complete young stellar populations can be gathered and used for age estimation, removing influences from our pipeline's selection biases. The stars present in these extended populations are provided in Table 2.

For each candidate young cluster, we perform a single age fit based on the photometry of candidate members, including both the original candidate young stars and the extended population of comoving/cospatial objects mentioned above. Our methods employ a least-squares fit, where we compare the photometry of these populations to a grid of solar-metallicity PARSEC isochrones with ages between 0.25 and 80 Myr (Chen et al. 2015). Only stars within the range of G magnitudes occupied by the isochrone grid are included to enable reliable interpolation, a restriction that also helps to suppress some of the background contamination added upon reintroducing candidate members that are not photometrically young. We also limit the range of $BP - RP$ Gaia color to between 1.2 and 4 to restrict the sample to the pre–main sequence. Most of the groups we recover have pre–main sequences that fall comfortably within the range of photometry given by these 0.25–80 Myr isochrones, with the exception of the Pleiades, which we exclude from age estimation due to its considerably older known age (e.g., Lodieu et al. 2019).

To reduce contamination from external sources, we restrict the sources included in our age fits according to the core distance (as defined in Section 3.4). For the core distance restrictions, we define a weight parameter in which the group member with the smallest core distance in the association has a weight of 1, the member with the largest core distance has a weight of zero, and sources in between have weights spaced linearly according to the core distance between these extremes. For top-level clustering, a default condition of weight >0.3 was required for the inclusion of members in the age fit for the group, which in most cases ensures that a young pre–main sequence dominates background contamination. In more contaminated groups, this was raised as high as 0.5. For subclusters, we often found that no weight cut was required due

Table 2
The Full Catalog of Gaia Stars with Space-velocity Positions Consistent with Our Photometrically Young Populations

ID[a]	TLC	EOM[b]	LEAF[c]	Gaia ID	R.A. (deg)	Decl.	Parallax (mas)	g (mag)	$bp-rp$	Weight
0	1	−1		2297218611804127104	309.8679	79.8920	3.176	18.93	2.87	0.02
1	1	−1		2226243658763815552	336.1287	69.4314	3.035	17.39	3.61	0.23
2	1	−1		2226228579136886144	336.4913	69.3191	3.296	13.51	1.23	0.08
3	1	−1		2286800434190203776	334.4768	80.5854	3.191	17.79	2.69	0.09
4	1	−1		2297464666187525120	324.2971	80.7429	3.391	11.98	0.83	0.01
5	1	−1		2212189293119019264	343.6841	66.3300	3.134	17.04	2.58	0.02
6	1	−1		2226302723153297664	336.9986	69.7687	3.066	18.36	3.76	0.27
7	1	−1		2286846781182031744	339.9039	80.6398	3.300	15.34	1.98	0.08
8	1	−1		2229484297488789248	338.5122	70.9616	3.439	17.85	2.65	0.05
9	1	−1		2229446020739020032	336.3050	70.5487	3.399	17.84	2.56	0.05

Notes. The TLC, EOM, and LEAF columns mark the top-level, EOM, and leaf clusters that each star is a part of, if any. We also include basic Gaia properties and the weight parameter described in Section 3.5, which serves as a proxy for closeness to other members of the top level cluster. Only a small subset of the 94,424 total objects is shown here.

[a] Note that in complex environments, extreme outliers may be shared between two clusters. These cases receive one entry for each possible solution.

[b] Objects not assigned to an EOM cluster are given the marker −1.

[c] Left blank if not related to a leaf cluster.

(This table is available in its entirety in machine-readable form.)

to tighter spatial and kinematic distributions that are less vulnerable to contamination; however, we reimposed the restrictions where necessary. Very few groups needed changes to these conditions, with some examples being the large and badly contaminated subgroups of CHA-1 and ORI-2, as well as Taurus-Orion IV and Cepheus Flare at the top level.

Since unresolved binaries appear photometrically younger compared to single stars, we attempt to remove those likely binary young star candidates from inclusion in our age fitting. We therefore further restricted the population used for age fitting based on the RUWE, an astrometric quality measurement that has been shown to be a strong indicator of stellar multiplicity (e.g., Bryson et al. 2020). In most cases, we removed sources with RUWE > 1.1, roughly corresponding to where the additional contribution to the RUWE curve from binaries becomes negligible in Bryson et al. (2020). This harsh RUWE cut works well to remove binary sequences; however, in regions with consistently high RUWEs, such as groups with embedded sources (e.g., Perseus; see Section 4.3), this condition occasionally removes nearly all stars. Therefore, in cases where more than two-thirds of stars are removed by RUWE < 1.1, we relax the condition to RUWE < 1.2. This value corresponds to approximately where Gaia populations are expected to become fully binary-dominated, and it is the recommended criterion for culling binaries in Bryson et al. (2020). This cut combined with the weight cut significantly reduced the number of new stars used in the fits, from a typically order-of-magnitude increase in populations upon the initial reintroduction of objects to a number more consistent with the expected number, which should be equal to the number of young candidates times the inverse of the expected recovery rate (see Section 3.2).

Once these culling conditions were applied, visually convincing fits were produced for all groups in our sample, which neglects strong background contamination and visible binary sequences. Figure 12 shows two examples of the results of this fitting. We generate fits for all HDBSCAN-identified groups without substructure, as well as a few cases that do show substructure but with a limited enough age spread for a global age to make sense. As we note throughout Section 4, the vast majority of our age fits for known groups are consistent with literature values; however, the uncertainties should be treated with caution, as we report standard errors that cannot capture the dominant systematic errors related to the choice of model and clipping or any inherent age spread. Due to the 80 Myr isochrone limit for inclusion in fitting, some members in associations with ages >50 Myr where these isochrones become close together may be removed, so some additional caution should be applied to the age fits near the upper end of our age range. Groups with a very small number of high-weight members should also be treated with additional caution, as their size makes them more vulnerable to the influence of a small number of field interlopers. These fits are all shown in the Appendix.

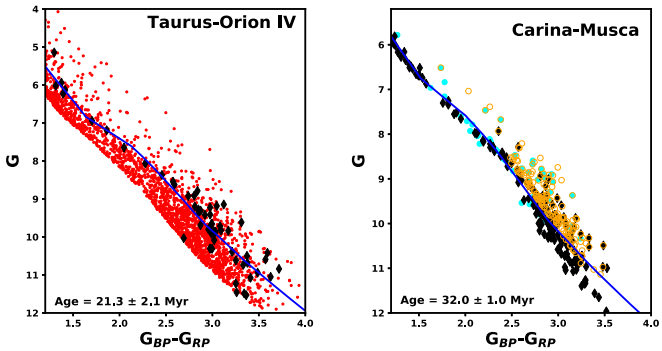

Figure 12. Sample isochrone fits for two of our young associations. In the left panel, we display our handling of severe contamination in Taurus-Orion IV, where a weight cut raised to 0.5 resulted in a very clean pre–main sequence emerging from the noise. Objects that failed the weight cuts are marked with red circles, and the stars that pass are given by black diamonds. In the right panel, we present members of the relatively uncontaminated Carina-Musca association after applying weight cuts. Stars marked by large cyan circles fail the subsequent RUWE cut, while stars marked by black diamonds pass it. Stars that fail the RUWE cut are likely binaries, hence the existence of a major concentration of them along an apparent binary sequence. Only objects marked by open orange circles are identified as young Carina-Musca members using our methods, highlighting the potential bias toward photometrically younger stars in that population that our age-fitting method attempts to rectify. Isochrone fits for all groups and regions are provided in the Appendix.

4. Results

By applying HDBSCAN clustering to our sample of young stars, we have identified extensive young structures in the solar neighborhood. These include 27 top-level regions (see Figure 10), with properties provided in Table 3. Many of these regions have significant internal structure, often with well-separated subclusters or subclusters widely separated in age. Performing a subclustering analysis on each of these regions allows for insight into their star formation histories. We therefore manually inspect each region for visible substructure and apply HDBSCAN subclustering to the 10 regions where substructure is visually evident. While we do not assert an absence of subclustering in the remaining regions, most are too small or sparse for a subclustering analysis to be useful and will therefore likely require a more targeted census before substructure can be ascertained. The details of this HDBSCAN implementation are described in Section 3.4. Further details of individual complex regions will be left to future publications, as will precise characterization of some of the previously unknown or little-known groups that we identify for the first time in this publication. These lesser-known groups are discussed in Section 4.5, along with a discussion of their relations to past literature. Most are assigned a unique name in Table 3 to reflect their distinctiveness as significant nearby young associations.

4.1. Greater Taurus

The general direction toward the Taurus constellation is known to contain multiple nearby stellar populations. The largest of these is associated with sites of active star formation in the Taurus Molecular Cloud (TMC), which form an extensive network of young stars often referred to as the Taurus-Auriga association, or simply "Taurus," as we refer to it here (e.g., see Kenyon et al. 2008). This complex represents one of the nearest and therefore most accessible sites of active star formation to the Sun (∼140 pc; Galli et al. 2018), and as such, it has been subject to many recent studies to expand the known stellar populations, learn more about the structure and history of the region, and study star formation more broadly (Luhman 2018; Galli et al. 2019; Krolikowski 2021; Liu et al. 2021). Through some of that work, kinematic substructure has recently been identified in the region, dividing the stellar populations into four different subgroups, each tied to a different set of clouds within the TMC (Luhman 2018). The lists of known and proposed Taurus members continue to expand, although the existence of older members beyond the well-established young (<10 Myr) populations remains controversial (Kraus et al. 2017; Luhman 2018). Aside from the Taurus populations associated with the TMC, there are also many smaller and less-studied groups in the area, including the 118 Tauri, 32 Orionis, and μ Tauri associations (Mamajek 2016, 2007; Gagné et al. 2020). Since all of these groups have only recently been fully characterized, any relations these associations might have to one another and the other major stellar populations in Taurus remain unsettled.

Using HDBSCAN, we identify an extensive contiguous network of young stars in the direction of Taurus, forming a top-level group that we refer to as "Greater Taurus." Significant substructure is found within the region, breaking it into a total of 11 EOM groups, which are displayed in Figure 13 and listed in Table 4. The velocity dispersions of Taurus subgroups are typically of order 1 km s^{-1} or smaller, with a larger dispersion for the group as a whole at about 4 km s^{-1}. However, most groups closely associated with the TMC are in a somewhat tighter distribution, with even the most extreme exemplars differing in transverse velocity by less than 10 km s^{-1}. Outlying groups typically differ from the core TMC groups in transverse velocities by 5–15 km s^{-1}. However, their distributions remain connected to each other and the core Taurus groups by lower-density distributions of stars in the intervening parameter space; hence, HDBSCAN merges them into a single large-scale Greater Taurus group (see Figure 13). Since the area Taurus covers is quite large, projection effects account for a significant fraction of the velocity changes. Assuming a constant velocity equal to the Taurus *UVW* vector from Luhman et al. (2009) would generate a 3 km s^{-1} velocity dispersion from projection alone. There are, however, subgroups with notably different velocities compared to the projected Taurus *UVW* vector, so internal velocity spreads do exist. Known members in Taurus are marked using filled symbols in Figure 13 as a way to compare our sample to the currently known extent of the region (Kraus et al. 2017; Luhman 2018; Esplin & Luhman 2019; Galli et al. 2019; Krolikowski 2021). The subgroups we identify as part of Greater Taurus extend far beyond these known Taurus members. Higher-order subclusters are also found at the leaf level, with four of 11 EOM groups dividing further into two leaf groups each. Close-up views of the groups containing leaf subclusters are provided in Figure 14.

Four of the groups that HDBSCAN identifies appear to be associated with the ongoing and recent star formation in the TMC (e.g., Rebull et al. 2011; Luhman 2018). These groups contain the majority of currently known Taurus association members (see Figure 14), and they all correspond to known populations of young stellar objects (YSOs) in and around active dense clouds in the TMC. Group GT-8 contains YSOs associated with L1495, B209, L1527, and L1529; GT-9 contains L1517 and L1544; GT-10 contains L1551; and GT-11 contains B213 and L1536 (Onishi et al. 2002; Luhman 2018). These four groups are similar to those proposed in Luhman (2018), with the main differences being our separation between GT-10 and GT-11, which Luhman (2018) grouped together into their "blue" cluster, and our exclusion of their "green" cluster as a distinct subgroup due to its very small membership. All four of our core TMC groups consist almost entirely of known Taurus member stars, with the notable exception of a southeastern extension to GT-9, which forms its own leaf cluster, GT-9B (recognized as Theia 68 in Kounkel & Covey 2019). The other GT-9 subgroup, GT-9A, consists of the known YSOs in the region around L1517. Group GT-8 also fragments into leaf clusters, with GT-8A consisting mainly of YSOs associated with L1529 and GT-8B consisting of B209 and L1495. All of these core TMC groups (aside from GT-9B) have derived ages between 3 and 7 Myr, consistent with our knowledge that active star formation is happening in these regions.

Distributed populations around Taurus have long been proposed, particularly in the direction of GT-1 (e.g., Neuhaeuser et al. 1995; Wichmann et al. 1996; Magazzu et al. 1997); however, these early papers have reached mixed results on potential association with the TMC (e.g., Briceno et al. 1997). We also identify distributed populations in our work, forming an extensive network of peripheral groups that we associate with

Table 3
Properties of the Nearby Clusters Identified Using HDBSCAN

TLC	Name	N	R.A. (deg)	Decl. (deg)	l (deg)	b (deg)	$D_{\rm sky}$[a] (deg)	d (pc)	$\mu_{\rm R.A.}$ (mas yr^{-1})	$\mu_{\rm decl.}$ (mas yr^{-1})	$V_{T,l}$ (km s^{-1})	$V_{T,b}$ (km s^{-1})	σ_{V_T}[b] (km s^{-1})	Age (Myr)
1	Cepheus Flare	17	339.5	73.9	113.9	13.4	5.3×2.2	329.3 ± 3.2[c]	6.8	0.9	9.9	−4.0	2.0×1.4	11.2 ± 3.3
2	Pleiades[d]	23	57.4	24.9	166.4	−22.4	7.6×4.4	139.9 ± 6.7	18.9	−44.7	29.2	−13.4	1.3×1.2	>80
3	Taurus-Orion I[e]	24	65.9	13.5	181.6	−24.4	5.3×2.5	299.2 ± 10.7	4.5	−5.2	9.7	0.4	2.0×0.6	10.6 ± 2.2
4	Ophiuchus Southeast[e]	31	257.5	−18.3	4.5	12.7	4.7×1.7	217.5 ± 14.2	−5.4	−11.6	−13.0	−2.2	1.6×0.7	17.5 ± 1.9
5	Fornax-Horologium[e]	40	54.2	−34.5	235.1	−53.7	20.1×9.6	107.8 ± 10.7	34.7	−3.9	1.9	17.3	3.9×1.7	38.5 ± 3.6
6	CMa North[e]	28	102.3	−15.3	226.4	−7.4	5.5×4.8	184.9 ± 19.7	−0.2	−5.4	4.1	−2.3	1.7×0.8	30.2 ± 2.7
7	Aquila East[e]	30	297.8	−9.0	31.6	−17.3	12.2×3.8	136.6 ± 6.5	8.5	−25.9	−12.7	−12.2	1.9×0.8	20.2 ± 1.5
8	Cepheus Far North[e]	219	329.6	74.7	111.7	17.8	16.9×5.7	179.1 ± 26.3	14.8	6.1	14.1	−2.9	2.4×0.9	23.8 ± 1.4[f]
9	Vela-CG7	111	124.9	−43.7	260.8	−4.2	7.3×4.6	323.7 ± 9.0[c]	−7.3	10.1	−19.2	−0.5	1.2×0.9	14.4 ± 1.2
10	ASCC 123	34	334.0	53.8	101.3	−2.2	8.0×3.5	251.1 ± 24.4	12.0	−2.0	10.3	−9.8	2.2×1.1	49.5 ± 2.2
11	Cepheus-Cygnus[e]	36	311.7	56.6	93.8	8.4	5.9×2.6	314.2 ± 11.7[c]	5.1	3.6	8.9	−2.7	1.0×0.7	32.6 ± 1.2
12	Lyra[e]	29	282.7	32.8	62.7	14.4	3.5×1.9	300.1 ± 11.9	1.7	−3.9	−4.1	−4.4	1.0×0.6	31.0 ± 1.8
13	Cerberus[e]	66	271.4	22.1	48.3	19.5	9.3×2.9	308.6 ± 13.3[c]	−1.0	−7.3	−10.4	−2.7	2.2×0.9	30.0 ± 0.8
14	Carina-Musca[e]	168	158.8	−69.4	291.0	−10.8	15.0×3.3	241.8 ± 33.4	−25.8	3.1	−27.7	−12.5	1.1×0.9	32.0 ± 1.0[f]
15	Perseus	264	58.4	32.9	161.4	−15.9	6.4×4.1	294.8 ± 22.4[c]	5.3	−8.1	13.0	−3.6	2.5×1.2	NC
16	Perseus OB3	145	56.1	44.1	153.2	−7.7	21.2×6.7	167.5 ± 18.6	22.5	−26.6	26.7	−5.8	2.4×0.9	NC
17	Taurus-Orion II[e]	64	89.5	18.3	190.6	−2.8	13.1×4.1	207.4 ± 17.2	7.6	−18.3	19.1	−2.6	2.5×0.8	41.5 ± 1.7
18	Greater Taurus	674	73.3	19.1	181.6	−14.9	23.9×13.5	142.8 ± 24.9	7.6	−22.1	14.5	−4.7	5.0×3.4	NC
19	IC 2391	101	139.1	−58.0	276.7	−8.3	23.2×4.2	157.7 ± 15.8	−26.7	18.9	−24.8	−4.3	1.8×0.7	45.0 ± 1.7
20	NGC 2451A	51	116.7	−40.9	255.1	−8.0	10.0×2.4	189.0 ± 14.0	−22.2	16.4	−22.3	−10.1	1.4×0.9	48.5 ± 1.3
21	Chamaeleon	259	188.0	−74.3	301.8	−13.0	14.9×11.6	190.1 ± 14.5	−18.2	−5.0	−17.7	−7.7	3.3×1.2	NC
22	Sco-Cen	7394	223.0	−40.5	329.1	9.4	50.4×20.0	137.8 ± 24.3	−18.5	−19.3	−18.0	−6.0	2.6×2.0	NC
23	Taurus-Orion III[e]	27	76.4	11.8	189.4	−17.2	5.0×2.5	309.6 ± 10.1	1.9	−3.5	5.9	−0.5	0.7×0.6	15.2 ± 2.8
24	Vela-CG4	551	108.7	−41.3	252.8	−13.9	21.5×8.8	265.7 ± 45.0[c]	−11.0	7.5	−14.1	−8.7	2.9×0.9	33.7 ± 0.9[f]
25	Taurus-Orion IV[e]	108	73.1	19.4	181.0	−15.0	18.1×8.7	218.8 ± 27.4	1.4	−7.6	7.1	−3.9	2.8×1.6	21.3 ± 2.1
26	Monoceros Southwest[e]	292	100.3	−4.6	215.9	−4.4	8.5×4.1	279.0 ± 35.2[c]	−6.6	−2.7	−0.7	−9.3	1.7×0.9	25.5 ± 0.6[f]
27	Greater Orion	757	79.7	−5.5	207.1	−23.0	12.5×9.1	299.1 ± 34.0[c]	1.0	−1.3	2.1	0.4	1.8×1.3	NC

Notes. In the age column, NC indicates that we find the group to be non-coeval; see the age column in the relevant subclustering table for the ages of the subgroups (e.g., Tables 4–7).
[a] On-sky spatial extent in galactic l/b, in the form of the rms in major axis × minor axis, when fit with a bivariate Gaussian.
[b] The radial extent of the velocity distribution, in the form of the rms in semimajor axis × semiminor axis, when fit with a bivariate Gaussian.
[c] Has a member within 1 pc of the search horizon, and therefore members are likely present beyond that limit.
[d] Older cluster: binary sequence dominant in detection.
[e] Name is new: limited to no coverage in the literature or left unnamed in the literature that does exist.
[f] Has substructure but approximately coeval.

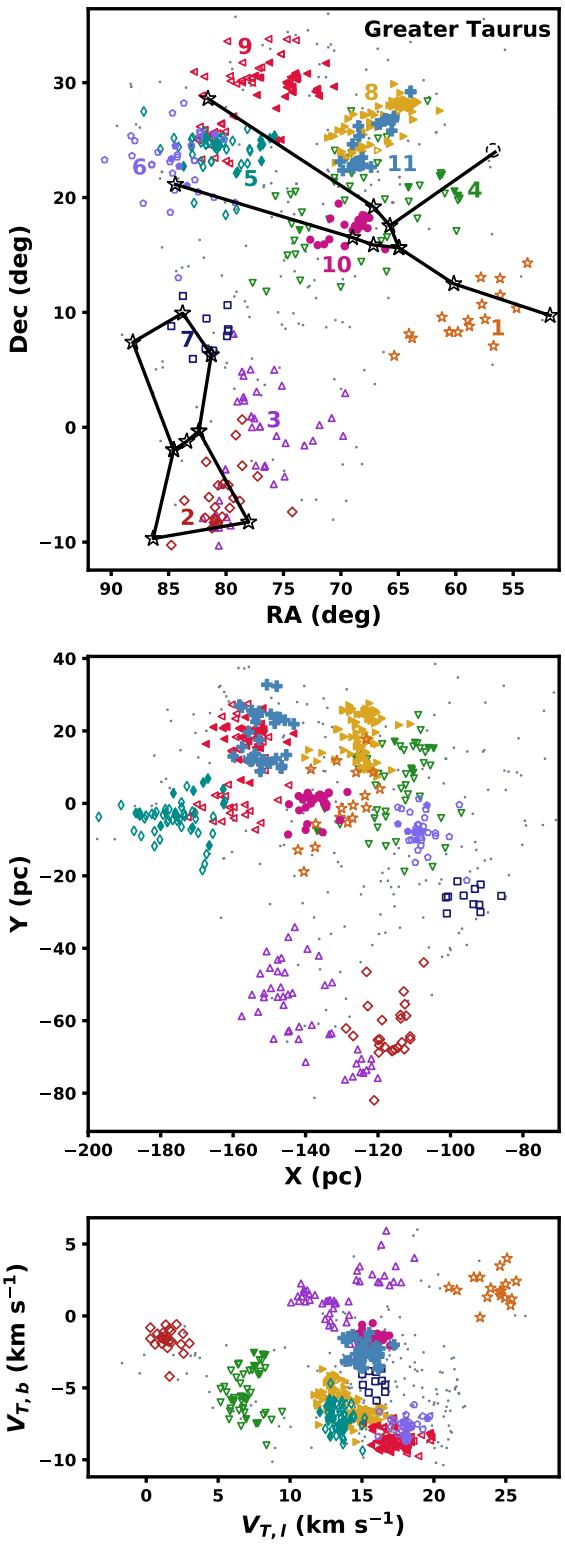

Figure 13. Members of the EOM groups that HDBSCAN identifies in Greater Taurus in R.A./decl., galactic X/Y, and transverse velocity coordinates. Filled symbols indicate known Taurus members, and open symbols indicate objects not previously recognized as part of Taurus. Small gray dots mark unclustered Taurus members. Each group is labeled according to its Greater Taurus ID from Table 4 in the top panel. Our identification of Greater Taurus includes significant populations beyond the known membership. The main stars of the Taurus and Orion constellations are included for reference in the top panel, with the dotted circle attached to Taurus representing the Pleiades.

Greater Taurus, many of which are significantly older than the core TMC groups. Some of these groups, including GT-5, GT-6, and both subgroups of GT-4, contain proposed Taurus members, but as indicated in Figure 13, their coverage in surveys of Taurus to date is far from complete (Krolikowski 2021). The extended Taurus population and clustering given in Liu et al. (2021) has, however, recently broadened coverage of populations around these groups, such as in GT-4, which loosely corresponds to Liu et al. (2021) groups 12–14. The youngest of these groups is GT-5 (which has considerable shared membership with Theia 93 from Kounkel & Covey 2019), a subgroup with TMC-consistent kinematics that appears to represent a slight continuation of Taurus to farther distances. Others have ties to the TMC that are considerably less clear. Group GT-6 has proposed Taurus members in it; however, it was previously identified independently as the 118 Tauri moving group by Mamajek (2016), which has membership falling entirely within GT-6 and kinematically consistent objects. This group represents one of three we identify that have previously been identified independently from the core populations in Taurus, with the other two being GT-1 and GT-7. In GT-7, over half of our membership overlaps with that of the 32 Orionis association (Mamajek 2007; Bell et al. 2017), a somewhat older group where our derived age of 27.2 ± 3.8 Myr is in agreement with the 22 ± 4 Myr isochronal age solution from Bell et al. (2015). The remainder of our membership matches spatially and kinematically with known 32 Orionis members, and while we only identify 11 members of this group, its established membership from Bell et al. (2017) contains only 47 objects, making the sample we gather smaller but consistent with recovery rates for groups in this age range (see Section 3.2).

Our member stars for GT-1 match with those of the much older μ Tauri association in both spatial coordinates and kinematics, where our derived age of $\sim 52.2 \pm 3.4$ Myr is marginally within the uncertainties of the 62 ± 7 Myr solution derived by Gagné et al. (2020). The Gagné et al. (2020) candidate member list contains all of our GT-1 members. With only 17 GT-1 members compared to $\sim$500 μ Tau candidates from Gagné et al. (2020), our detection of this group is limited, although this low recovery rate is likely primarily due to the very old age of the region (see Section 3.2). A few GT-3 members in its far western extension are also included in the Gagné et al. (2020) catalog, an overlap indicative of a possible link between GT-1 (μ Tauri) and GT-3. The cores of these two groups are separated by $\sim$60 pc in space, $\sim$10 km s^{-1} in transverse velocity, and $>$30 Myr in age; therefore, their affiliation with one another is far from certain. However, the separation of these subgroups in velocity is consistent with projection effects, and a low-density, near-contiguous linear structure in the plane of the sky does appear to connect them, lending plausibility to the connection between these subgroups.

Aside from this slight overlap with μ Tau, the remainder of the stars in GT-3 and GT-2 do not match with any current Taurus catalogs related to the TMC, as expected given that these groups stretch up to 25° south of the core TMC groups to the crowded fields in front of the Orion association. Group GT-2 (along with GT-4) is included in Kounkel & Covey (2019) group Theia 44, but these groups are otherwise absent in the literature. Group GT-3 divides into two subgroups, GT-3A to the north and GT-3B to the south, with ages of $\sim$19 and 14 Myr, respectively. Subgroup GT-3B is projected directly

Table 4
Clustering within Greater Taurus

GT	LEAF	Name	N	R.A. (deg)	Decl. (deg)	l (deg)	b (deg)	D_{sky}[a] (deg)	d (pc)	$\mu_{R.A.}$ (mas yr^{-1})	$\mu_{decl.}$ (mas yr^{-1})	$V_{T,l}$ (km s^{-1})	$V_{T,b}$ (km s^{-1})	σ_{V_T}[b] (km s^{-1})	Age (Myr)
1		μ Tau	17	59.0	9.6	180.0	−32.0	8.5 × 2.8	153.5 ± 6.1	23.7	−23.3	24.1	1.9	1.3 × 1.0	52.2 ± 3.4
2			25	80.3	−6.2	208.1	−22.9	6.0 × 3.5	144.2 ± 6.2	−1.1	−3.0	1.5	−1.6	0.8 × 0.6	31.0 ± 2.5
3			46	77.6	−2.0	202.6	−23.2	11.1 × 5.2	165.5 ± 8.2	10.0	−14.2	13.6	1.5	2.4 × 1.3	NC
3	A		13	78.1	2.0	199.1	−20.9	5.2 × 1.6	166.9 ± 3.6	8.6	−14.1	13.1	0.3	0.7 × 0.6	19.3 ± 3.9
3	B		12	80.8	−8.4	210.5	−23.4	2.0 × 1.2	157.2 ± 2.3	8.3	−12.4	11.0	1.5	0.4 × 0.4	13.8 ± 2.2
4			51	68.0	19.7	177.6	−18.8	11.2 × 6.6	121.6 ± 7.8	0.7	−15.1	7.0	−5.3	1.6 × 1.0	NC
4	A		16	70.6	20.4	178.7	−16.6	7.0 × 2.9	119.8 ± 4.4	−1.1	−16.1	6.7	−6.2	0.7 × 0.5	37.8 ± 3.2
4	B		11	61.1	20.1	172.7	−23.6	3.3 × 2.8	120.5 ± 3.3	4.8	−14.0	7.7	−3.5	0.5 × 0.3	29.8 ± 8.3
5		Theia 93	55	80.7	24.2	181.3	−6.8	4.3 × 3.1	177.3 ± 7.4	1.8	−18.1	13.5	−7.2	0.9 × 0.6	8.5 ± 1.3
6		118 Tau	33	84.2	23.0	184.0	−4.8	5.8 × 4.5	108.8 ± 3.6	6.1	−37.6	18.1	−7.7	0.8 × 0.6	16.8 ± 2.0
7		32 Ori	11	81.5	8.3	195.3	−14.7	3.6 × 3.3	102.0 ± 5.1	8.8	−33.2	15.9	−4.6	0.7 × 0.6	27.2 ± 3.8
8		L1527[c]	87	66.6	26.4	171.5	−15.5	5.5 × 1.8	130.7 ± 4.5	7.9	−23.8	14.2	−6.2	1.7 × 0.7	4.5 ± 0.9[d]
8	A	L1529	30	68.3	24.5	174.0	−15.6	3.1 × 1.5	129.4 ± 2.7	7.3	−21.5	12.9	−5.1	0.7 × 0.6	3.4 ± 0.5
8	B	B209/ L1495	38	65.0	28.0	169.3	−15.6	2.6 × 0.7	131.2 ± 3.3	8.6	−25.6	15.2	−7.0	0.7 × 0.4	3.8 ± 0.8
9		L1544[c]	77	77.1	29.0	175.4	−6.7	7.1 × 3.7	159.2 ± 5.6	4.4	−25.3	17.4	−8.6	1.0 × 0.6	NC
9	A	L1517	40	75.5	30.1	173.7	−7.1	4.1 × 1.8	158.0 ± 3.3	4.3	−25.1	16.9	−8.8	0.7 × 0.4	5.6 ± 0.8
9	B		10	81.4	25.8	180.2	−5.4	1.6 × 0.7	162.1 ± 3.6	3.6	−25.5	17.8	−8.5	0.4 × 0.2	11.5 ± 1.3
10		L1551	30	69.0	17.5	180.2	−19.7	3.1 × 1.8	145.6 ± 3.8	12.3	−19.1	15.6	−1.5	0.7 × 0.4	7.0 ± 2.0
11		B213/ L1536	34	67.4	24.9	173.2	−16.0	5.3 × 1.1	159.2 ± 4.2	10.7	−17.4	15.2	−2.5	0.8 × 0.7	3.3 ± 0.9

Notes. Here GT is the Greater Taurus ID, a number from EOM clustering used for identification. Leaf clusters within an EOM cluster are given a subgroup ID, which is given in the LEAF column. In the age column, NC marks subclustered groups with non-coeval populations.
[a] On-sky spatial extent in galactic l/b, in the form of the rms in major axis × minor axis, when fit with a bivariate Gaussian.
[b] The radial extent of the velocity distribution, in the form of the rms in semimajor axis × semiminor axis, when fit with a bivariate Gaussian.
[c] This named structure is not included in any HDBSCAN-defined leaf subclusters. Names associated with subgroups are also contained in the higher-level EOM group.
[d] Has substructure but approximately coeval.

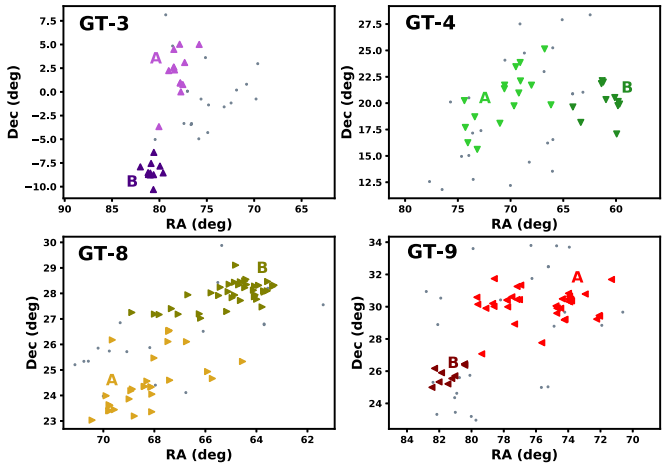

Figure 14. Subclustering in the four Greater Taurus EOM clusters where it is present, displayed in R.A./decl. sky coordinates. In all cases, unclustered members of the parent group are marked as small gray dots, and the subgroups are given as triangles. The ID of the parent EOM cluster is indicated on each panel.

behind the somewhat older (∼21 Myr) GT-2 group in the plane of the sky; however, despite this close physical proximity between GT-3B and GT-2, they are separated by more than 10 km s^{-1} in transverse velocity space, making a close association in formation origin unlikely. Together with GT-4 and GT-1, GT-3 represents a visible outlier from the relatively compact transverse velocity distribution occupied by the core TMC groups, which is less than 10 km s^{-1} across. While GT-2 also qualifies as an outlier to the TMC in transverse velocities, it is actually not inconsistent with a common origin there, as its velocities are consistent with the projection of the Taurus *UVW* velocity vector at its location (Luhman et al. 2009). Group GT-4 is unique in that it has an outlying kinematic signature despite physically overlapping with GT-8 and GT-10.

These distributed extensions to the Greater Taurus association demonstrate that an older population is present in the region. The youngest group we find outside of the well-characterized TMC regions is GT-5 at ∼8.5 Myr old, while all others have ages in excess of 10 Myr. These older regions range from the subgroups of GT-3, at ∼14–19 Myr, to μ Tau (GT-1), at ∼52 Myr old. While their inclusion in Greater Taurus is not necessarily certain for some of the peripheral subgroups (see Section 5.4), the case of GT-9 is quite notable as an older, distributed population that is directly connected to the known elements of the Taurus association. This group consists of two leaf subgroups, GT-9A and GT-9B, which are contiguous with one another in spatial coordinates and have overlapping velocity distributions. Subgroup GT-9A has a near-newborn age solution and contains known Taurus YSOs around L1517 (e.g., Luhman 2018), while GT-9B contains a dispersed population with an older age solution at ∼11.5 Myr old, forming a clear link between an older (age > 10 Myr) distributed group and the current star formation events in Taurus. Group GT-6 and both subgroups of GT-4 also all contain members with known signs of youth and kinematics previously linked to Taurus (Kraus et al. 2017; Krolikowski 2021), and with ages of ∼17, 38, and 30 Myr,

respectively, these groups further enforce Taurus's connection to an older population.

Our work supports recent studies that have also demonstrated the presence of this older distributed population, including Kraus et al. (2017), Zhang et al. (2018), Krolikowski (2021), and Liu et al. (2021). Not all current literature promotes this conclusion, however, such as Luhman (2018), who concluded that the proposed older members with an age >10 Myr are likely not true Taurus members due to differing velocity distributions. Some of our proposed members have kinematics that make them potentially subject to this argument, likely requiring a full kinematic analysis and traceback, as discussed in Section 5.4, to achieve a definitive conclusion on their association with the current star formation in the TMC. Other old subgroups, however, have a clear kinematic overlap with known Taurus members near the TMC, and those with more discrepant velocities are linked by previously unknown low-density stellar distributions to the known Taurus groups, strengthening those connections. The presence of nine groups and subgroups in the 10–40 Myr age range, as well as a much older group in μ Tauri with an age of ∼52 Myr, demonstrates that not only is there an older stellar population surrounding the currently known, essentially newborn TMC-associated groups, but Greater Taurus may represent the oldest association in the solar neighborhood that still hosts active star formation.

4.2. Greater Orion

The Orion OB1 association contains an enormous population of stars in terms of both numbers and extent, while also hosting the Orion Nebula Complex, the nearest site of active high-mass star formation to the Sun (Hillenbrand & Carpenter 2000; Bally 2008). Substructure has long been recognized in the region, with Blaauw (1964) dividing the region into four distinct substructures based on location and photometry, labeled Orion OB1a, b, c, and d. All of these Orion OB1 subgroups are centered beyond this project's horizon at 333 pc; however, both Orion OB1a and OB1b have components that are nearer (Brown et al. 1994; Briceño et al. 2019). The known populations accessible through this project therefore consist of Orion OB1b, a young (∼5 Myr) population immediately surrounding Orion's Belt, and Orion OB1a, a somewhat older (∼11 Myr) population containing groups to the northwest, such as the 25 Orionis and HD 35762 clusters (Bally 2008; Briceño et al. 2019). All recognized subgroups of Orion OB1a have significant components within our search horizon; however, Orion OB1b has a notably bimodal parallax distribution, resulting in only the nearest component being visible, a subgroup given the name Orion OB1b I by Briceño et al. (2019).

Like in Taurus, we also identify a distributed population beyond the known subgroups in Orion, albeit with much less variety in age. The high-level region we identify as Greater Orion hosts three large HDBSCAN-identified groups, with the largest (ORI-1) subdividing into three small subgroups. Their extents in spatial and velocity coordinates are shown in Figure 15 and their properties are listed in Table 5. The mean transverse velocities in these groups never differ from one another by more than 4 km s^{-1}, and all groups have tight velocity dispersions of order 1 km s^{-1} or below. Particularly on the far edge of ORI-1, these velocity distributions are found to be nearly identical to those of known young associations near and beyond our search radius cutoff, such as 25 Orionis and the

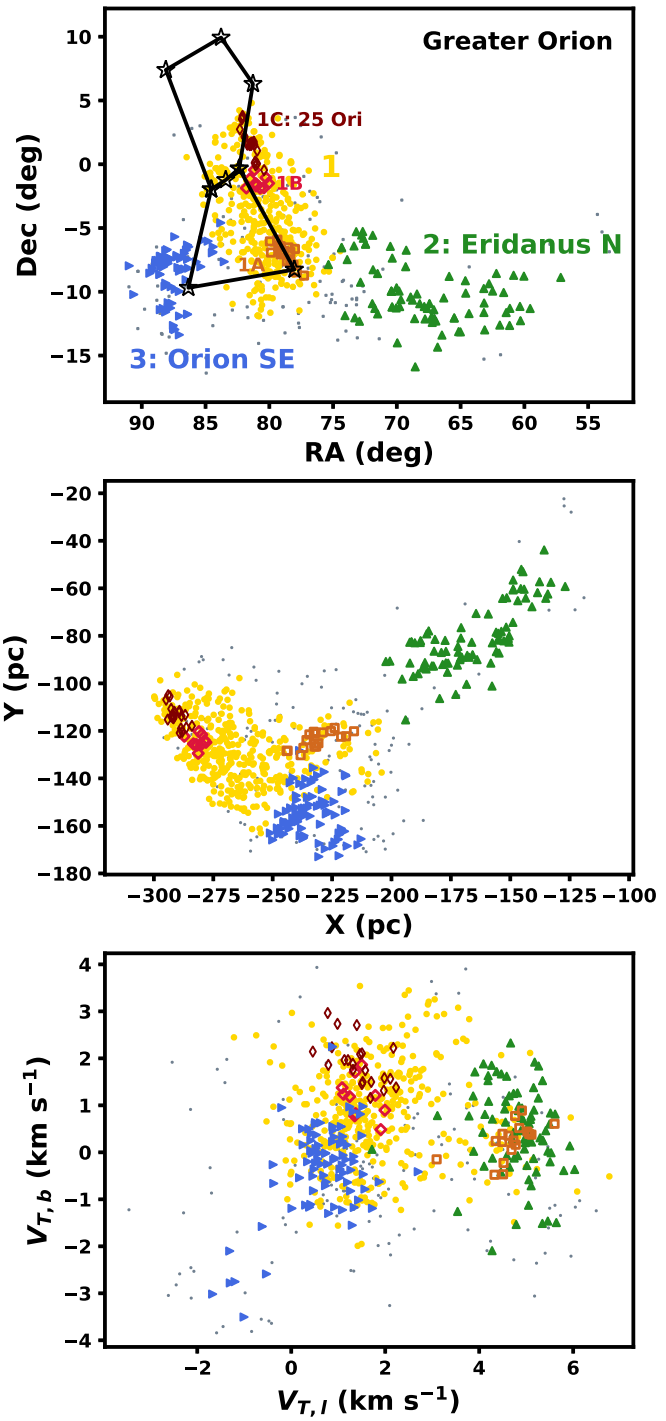

Figure 15. The HDBSCAN-identified groups and subgroups in Orion, labeled according to their Orion ID, and shown in R.A./decl., galactic X/Y, and transverse velocity coordinates. The open diamonds and squares correspond to leaf subclusters of the larger ORI-1 group (yellow circles), while filled symbols are top-level EOM clusters. Small gray dots represent unclustered Greater Orion members. The main stars of the Orion constellation are included for reference.

related subgroups Zari-B_0 and Zari-B_6 defined by Zari et al. (2019). Some of the more distant members of ORI-1 we identify, including those in subgroups ORI-1B and ORI-1C, overlap with the near edge of Orion OB1a (Briceño et al. 2019), with ORI-1C members overlapping directly with known members of the 25 Orionis and HD 35762 subgroups in spatial coordinates. Other sections of ORI-1 overlap with the

Table 5
Clustering within Greater Orion

ORI	LEAF	Name	N	R.A. (deg)	Decl. (deg)	l (deg)	b (deg)	$D_{\rm sky}$[a] (deg)	d (pc)	$\mu_{\rm R.A.}$ (mas yr^{-1})	$\mu_{\rm decl.}$ (mas yr^{-1})	$V_{T,l}$ (km s^{-1})	$V_{T,b}$ (km s^{-1})	σ_{V_T}[b] (km s^{-1})	Age (Myr)
1			458	80.6	−3.6	205.8	−21.4	8.3 × 3.6	316.8 ± 15.7[c]	1.1	−0.9	1.9	0.8	1.3 × 1.0	13.3 ± 0.3[d]
1	A		18	78.7	−7.1	208.2	−24.7	2.1 × 0.8	288.6 ± 7.4	1.7	−3.0	4.7	0.2	0.6 × 0.3	14.0 ± 1.2
1	B		10	80.9	−1.5	203.9	−20.1	1.1 × 1.0	327.7 ± 1.7[c]	1.1	−0.5	1.5	1.2	0.5 × 0.3	12.3 ± 1.6
1	C	25 Ori	22	81.5	1.6	201.3	−18.2	2.5 × 0.4	328.4 ± 2.4[c]	1.5	−0.2	1.4	1.9	0.6 × 0.3	13.2 ± 0.8
2		Eridanus North[e]	82	67.8	−10.2	205.9	−35.7	8.8 × 5.1	227.8 ± 18.2	2.3	−3.8	4.8	0.4	0.9 × 0.6	20.0 ± 1.8
3		Orion Southeast[e]	81	87.5	−8.4	213.6	−17.4	4.0 × 2.6	294.1 ± 8.5	−0.0	−0.6	0.7	−0.4	1.1 × 0.6	19.0 ± 0.9

Notes. Here ORI is the Greater Orion ID, a number from EOM clustering used for identification. Leaf clusters within an EOM cluster are given a subcluster ID, which is given in the LEAF column.
[a] On-sky spatial extent in galactic l/b, in the form of the rms in major axis × minor axis, when fit with a bivariate Gaussian.
[b] The radial extent of the velocity distribution, in the form of the rms in semimajor axis × semiminor axis, when fit with a bivariate Gaussian.
[c] Has a member within 1 pc of the search horizon, and therefore members are likely present beyond that limit.
[d] Substructure present but mostly coeval.
[e] Name is newly assigned by this paper.

membership of Orion OB1b I; however, most of this group is beyond our search radius limit, leaving the remaining members too sparsely distributed for HDBSCAN to identify a subgroup. Both ORI-1B and ORI-1C, along with the rest of the far edge of ORI-1's main body, also appear cut off by our search horizon, as expected given the distance distributions of known Orion subgroups (e.g., Briceño et al. 2019). This indicates that our coverage of known associations in Orion is consistently incomplete.

The lack of overlap between our extended population in Orion and the known Orion subgroups is demonstrated by the DBSCAN-based subclustering analysis of Orion conducted in Zari et al. (2019), which investigated the full range of distances that Orion occupies. The only overlap between our members and those identified with a subgroup by Zari et al. (2019) is at the far edge of our ORI-1 group (including ORI-1C), which overlaps with the membership for Zari-B in 3D positions and proper motions. Zari-B_6 in particular has a distribution in position/velocity parameter space that overlaps fully with the membership of our subgroup ORI-1C, while the near edge of Zari-B_0 also overlaps with part of ORI-1's far edge (Zari et al. 2019). However, most of Zari-B, as well as the rest of the groups they identified, are well beyond our search radius, and despite their search window extending to within 200 pc of the Sun, all populations identified by Zari et al. (2019) are almost entirely located beyond 300 pc.

The more nearby populations accessible through this research therefore largely consist of stars well outside of well-known groups, residing in lower-density extensions that methods like the DBSCAN clustering implementation used in Zari et al. (2019) are not able to recover due to the inflexibility of DBSCAN in simultaneously handling groups with very different densities and scales. Many of the stars in these extensions are linked directly to the known Orion OB1a and OB1b I populations through membership in ORI-1, forming significant extensions to the region that stretch over 50 pc away from those known sections in the approximate direction of the Sun and toward Rigel in the plane of the sky. Some of these nearer populations in ORI-1, as well as ORI-3, appear as part of the Orion D group in Kounkel et al. (2018); however, these populations remain obscure in the literature, and those in ORI-2 remain completely unknown. As ORI-2 and ORI-3 are both separated from known populations and clearly defined, we assign them new names: Eridanus North for ORI-2 and Orion Southeast for ORI-3.

Eridanus North forms a major low-density extension to Greater Orion, spanning from the near edge of ORI-1 to within ∼200 pc of the Sun, with some members residing more than 150 pc from the known membership in Orion OB1 (Briceño et al. 2019; Zari et al. 2019). While there is a subtle separation between the mean transverse velocities in Eridanus North and the majority of ORI-1 of approximately ∼3 km s^{-1}, a small subset of ORI-1 stars at the near edge have transverse velocities that overlap with those in Eridanus North. Stars in ORI-1 that have these unique motions form the ORI-1A subgroup, which has spatial positioning consistent with ORI-1 but transverse motions more similar to Eridanus North. As such, ORI-1A appears to represent a bridge between the main Orion association and the more dispersed extensions into Eridanus North. The stars in Eridanus North appear notably older compared to the known populations in Orion, with an age estimated at ∼20 Myr.

The other new major group we find in the Orion Complex, Orion Southeast, is identified as UPK 422 in Sim et al. (2019) and grouped under Orion D in Kounkel et al. (2018) but otherwise undiscussed in the literature. The group has limited separation from the main body of ORI-1 in transverse velocities but much more visible separation in spatial coordinates. This separation is most evident in the sky coordinates, where the group appears to the southeast of Orion's Belt, with a significant central concentration of stars near the border with Monoceros (see Figure 15). Like in Eridanus North, we find relatively old ages for stars in this group, with an age estimate of ∼19 Myr.

Despite the large extensions we find to known populations, ORI-1 is found to have very consistent ages across its three subgroups, all of which are at ∼12–14 Myr and within the uncertainties of each other. Consequently, ORI-1 appears to be mainly coeval, and no significant hints of age gradients are detectable in the region. One minor exception is a small unclustered group of stars near Orion's Belt associated with Orion OB1b I, where the members have individual age estimates consistent with the 5 Myr photometric age estimate for the group from Briceño et al. (2019). Our ∼13 Myr age solution for ORI-1C is approximately consistent with the age estimates given for the HD 35762 and 25 Orionis subgroups, which vary from ∼8 Myr in Briceño et al. (2019) to 11 and 13 Myr, respectively, for the similarly defined ASCC 21 and ASCC 16 subgroups (Kos et al. 2019). The 11 Myr age for the broader extended populations in Orion OB1a given by Briceño et al. (2019) is also roughly consistent with our bulk age solution for ORI-1, at ∼13.3 Myr, as is the 12.3 Myr estimate from Siess et al. (2000).

4.3. Perseus

Perseus OB2, which we abbreviate to "Perseus" in this paper, is a young association originally identified as an ∼6° wide overabundance of O and B stars near the open cluster IC 348 (Blaauw 1952). While IC 348 and the neighboring cluster NGC 1333 have long been suggested as concentrations within the Perseus region (e.g., Blaauw 1952; Strom et al. 1974), the populations in these clusters have since been significantly expanded, revealing an extensive network of recent and continuing star formation centered at a distance of ∼300 pc (e.g., Cernis 1993; de Zeeuw et al. 1999; Bally et al. 2008). The stellar populations in these clusters are contained within the Perseus Molecular Cloud, an extensive region containing numerous sites of active star formation, as evidenced by the presence of dense cores (e.g., Sargent 1979; Bally et al. 2008; Rosolowsky et al. 2008; Kerr et al. 2019). As such, most of the young stars we see in the region can be tied to previous formation in that cloud (Bally et al. 2008). The presence of a more distributed population reflective of the broader spatial distribution suggested by early works is still accepted in Perseus, having recently been the subject of a substructure analysis from Pavlidou et al. (2021) and new WISE observations from Azimlu et al. (2015). However, most populations currently known in Perseus are thought to be related to earlier generations of formation in the current Perseus cloud (Bally et al. 2008). The affiliation of these populations with the current star formation activity is reflected in the age estimates, which are consistently at or below 6 Myr (de Zeeuw et al. 1999; Bally et al. 2008; Pavlidou et al. 2021), albeit with a few spectral energy distribution fits from Azimlu et al. (2015)

suggesting ages closer to 10 Myr. Astrometric measurements have recently established a kinematic division between stars associated with IC 348 and NGC 1333, with the two populations separated by about 5 km s^{-1} in velocity space (Ortiz-León et al. 2018)

Using HDBSCAN, we identify two spatially overlapping subregions in Perseus, lying along nearly the same line of sight but at different average distances. The nearer of the two at 283 pc is PER-1, while PER-2 has a mean distance of 314 pc. Both of these subgroups in some capacity relate to the known clusters in Perseus, with PER-1 containing NGC 1333 and PER-2 containing IC 348 (Ortiz-León et al. 2018). Both Perseus subgroups, however, contain significant structure beyond the two known young clusters, extending to the east of each cluster's center. These extensions are fully contiguous with the known elements of each subcluster, and therefore they do not form distinct leaf clusters within PER-1 or PER-2. However, the individual age estimates there are distinctly older than those in the known young clusters. The ages in these Perseus subgroups do not appear to form any gradient, instead forming two distinct epochs of star formation in each subgroup, one for the current generation of star formation and another one in the eastern extensions preceding it. To separately investigate each of these generations, we use simple cuts in the plane of the sky to separate the eastern extensions from the western young clusters. We list the properties of each Perseus subgroup in Table 7, and display their extents in spatial and velocity coordinates in Figure 16, marking the eastern extensions with open symbols. For the sake of identification, we refer to the area around NGC 1333 and IC 348 as PER-1A and PER-2A, respectively, while the corresponding eastern extensions for each are referred to as PER-1B and PER-2B.

Both PER-1A and PER-2A have properties consistent with expectations for groups anchored in NGC 1333 and IC 348, respectively. They separate cleanly from one another in velocity space, reflecting the kinematic divisions identified by Ortiz-León et al. (2018) between stars associated with each of the young clusters. Despite these kinematic differences, all of the stars we include in these sections are spatially consistent with the known segments of the Perseus Molecular Cloud, with PER-2 being limited almost entirely to IC 348 and its immediate surroundings, while PER-1 is considerably less tethered to NGC 1333 alone, as it includes some members more spatially consistent with adjacent clouds such as B1 and L1455 and excludes some apparent NGC 1333 members from subclustering due to their location at the outskirts of the broader Perseus group. The inclusion of stars from clouds outside NGC 1333 in PER-1 is consistent with the results from Ortiz-León et al. (2018), where stars associated with B1 and L1455 are found to have kinematics consistent with NGC 1333. Both of the young subgroups we find in Perseus have ages that reflect their essentially newborn status at 4–6 Myr, results consistent with current age estimates in the region (e.g., Bally et al. 2008). Due to these very young ages, many of the stars in Perseus are embedded (e.g., Bally et al. 2008), and as such, we see a significant downselect of members in these young clumps due to Gaia quality cuts, as was demonstrated in Section 3.2. A crossmatch between Gaia sources and members identified in Luhman et al. (2016) by Ortiz-León et al. (2018) found 90 and 351 members in those clusters, respectively, significantly more than the 19 and 43 that we find. As such, while both eastern extensions contain populations larger than

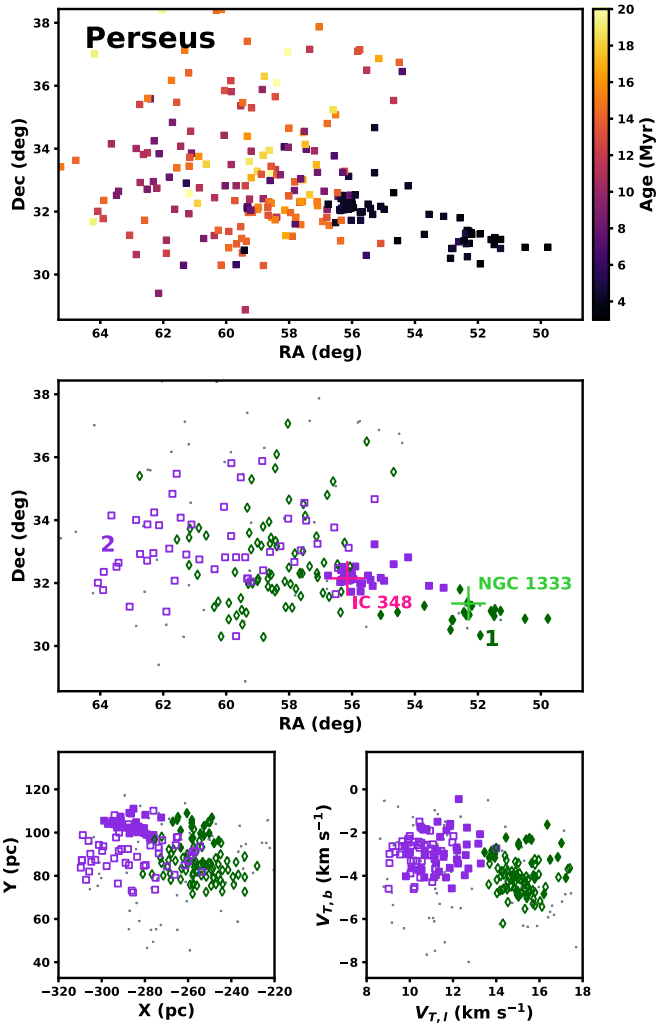

Figure 16. Age distribution and subclustering in Perseus. Ages in R.A./decl. space are shown in the top panel, displaying the clear divide between the young populations to the west near the Perseus Molecular Cloud and older populations to the east. The remaining panels show subclustering in R.A./decl., galactic X/Y, and galactic transverse velocity. The two subclusters are labeled in the middle panel, and objects in the older eastern extensions are marked with open symbols to distinguish them from the young western populations. The centers of IC 348 and NGC 1333 are also marked for reference (Ortiz-León et al. 2018).

we identify in their respective young clusters, upon corrections for recovery rates, both generations appear to be comparable in size.

The old eastern extensions to Perseus are much more dispersed compared to the more concentrated clusters that we identify near the Perseus Molecular Cloud, with both spanning approximately 7° in the plane of the sky, or ∼35 pc. Both PER-1B and PER-2B have very similar velocities to their younger counterparts, with standard deviations of order 1 km s^{-1} and differences in velocity between old and young components not exceeding 1.5 km^{-1}. Due to these kinematic similarities between the young and old populations, PER-1B and PER-2B have clearly separated velocities that reflect the kinematic divisions between their corresponding younger clusters. Despite the kinematic differences seen in each subgroup, PER-1B and PER-2B have remarkably similar age solutions to each other, both at ∼17 Myr. The older ages of these eastern extensions may also be reflected in the broad spatial

distributions they occupy, as dispersal is expected as time advances (e.g., Brown et al. 1997). The near-identical ages of these two populations, combined with the kinematic differences between them, suggests that star formation in those subgroups has progressed essentially in parallel, suggesting that they were assembled by the same initial processes despite their kinematic differences.

The kinematic similarities between the current star-forming regions and their older counterparts suggest that the events that gave rise to the older generations in PER-1 and PER-2 likely originate in the same star-forming processes that led to the current star formation in NGC 1333 and IC 348, respectively. However, due to the significant time lag between the older and younger generations in each subgroup, continuous formation within a static cloud seems unlikely. We present one possible explanation for this lag in Section 5.3, in which the first generation of stars disperses its parent cloud while still being fed by a flow of external material, resulting in another burst of star formation.

4.4. Sco-Cen Association

The Sco-Cen association (also known as Sco-OB2) is an extensive young association covering much of the southern Milky Way and represents the nearest site of recent high-mass star formation (de Zeeuw et al. 1999). The association was first identified as an expansive $40^\circ \times 70^\circ$ overdensity of comoving O and B stars by Kapteyn (1914), with work in the following decades significantly deepening the known membership within this original footprint (e.g., Gutierrez-Moreno & Moreno 1968). The first hints of substructure were proposed by Blaauw (1964), who divided the association into three large subregions that are still frequently referenced in the literature: US, UCL, and LCC. The past age estimates vary from between 5 and 11 Myr for US (e.g., Preibisch & Zinnecker 1999; Pecaut et al. 2012) to a little under 20 Myr for UCL and LCC (e.g., Mamajek et al. 2002). Recent, more complete surveys of Sco-Cen, especially post-Gaia, have identified thousands of new Sco-Cen candidate members, with current membership estimates of around 10,000 (e.g., Pecaut & Mamajek 2016). As the known populations in the region became increasingly complete, Rizzuto et al. (2011) noted the presence of a broad population linking all three of the traditional subgroups, suggesting that these traditional divisions are insufficient to capture the interconnected and continuous nature of star formation in the Sco-Cen association. The Ophiuchus and Lupus clouds are now commonly recognized as part of Sco-Cen, and recent work has proposed the inclusion of many other adjacent groups, including Corona Australis (CrA), the Chamaeleon clusters, the TW Hydrae association (TWA), and IC 2602 (e.g., Mamajek et al. 1999; Mamajek & Feigelson 2001; Damiani et al. 2019).

Sco-Cen is by far the largest association within our search radius, containing nearly 7400 candidate young stars. The region thus contains nearly an order of magnitude more stars than the near edge of Greater Orion, which is the next largest group we identify. Figure 17 plots the populations we identify against the US, UCL, and LCC boundaries as defined by de Zeeuw et al. (1999). Members of our Sco-Cen population clearly reside in all three of the traditional subregions, with significant extensions beyond the traditional boundaries. Our population's lack of adherence to the three traditional subregions and lack of strong divisions on the boundaries between them support the more continuous view of star formation proposed by Rizzuto et al. (2011) while also meriting a reassessment of the way we divide populations in the region. In this section, we present the novel view of Sco-Cen's internal structure that we construct through our subclustering and age distribution analyses, revealing an extensive stellar population with complex substructure and age trends spanning of order 100 pc.

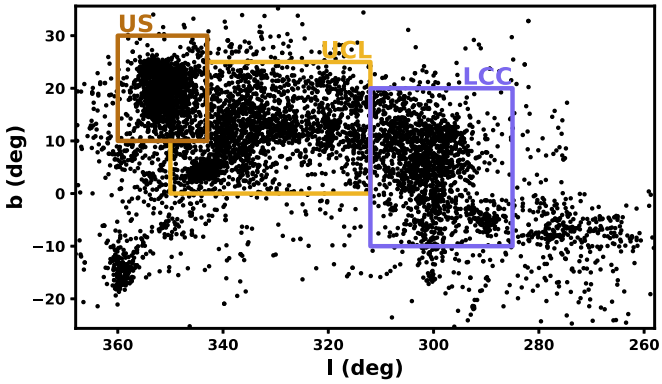

Figure 17. Stars we identify in Sco-Cen, with the de Zeeuw et al. (1999) subregions plotted over it.

4.4.1. Clustering Overview

We find a remarkable diversity of substructures in Sco-Cen with 27 EOM clusters, three of which subdivide into a total of 16 additional leaf clusters (see Table 6). Our top-level clustering results in Sco-Cen are presented in Figure 18, while clustering within the three EOM clusters with substructure occupies Figures 19–21. Compared to the previously discussed Greater Taurus association, most of Sco-Cen has a very compact velocity distribution, especially given its size. Excluding the groups to the far west, all subregions we identify have mean velocities within 7 km s^{-1} of each other, a velocity distribution comparable to the spread in transverse velocities expected from projecting the *UVW* velocity vector for US onto the plane of the sky (Luhman & Esplin 2020). As such, projection effects may dominate the observed velocity spread. The remaining groups can extend up to $\sim$10 km s^{-1} away from the center of the main distribution, although these extended populations remain spatially and kinematically contiguous with the main body of Sco-Cen.

Most of the denser regions of Sco-Cen fall within the de Zeeuw et al. (1999) boundaries; however, 13 of the 27 subgroups we identify in Sco-Cen exist at least partially outside of the de Zeeuw et al. (1999) boundaries displayed in Figure 17. The groups that overflow the traditional Sco-Cen boundaries include the southern tip of SC-27, which is the largest group falling mostly within the traditional boundaries of LCC. Group SC-27 also matches closely with the population referred to as the Crux moving group (CMG) in Goldman et al. (2018), a proposed LCC subgroup. However, given the lack of larger-scale substructure identified in Sco-Cen, we find that SC-27 and, by extension, the CMG are better described as revised extents for LCC, rather than distinct populations within. The US region also has its own EOM cluster with substructure in the form of SC-17; however, the remaining traditional subregion, UCL, does not, instead being loosely comprised of seven clusters with no parent EOM cluster. This lack of identified higher-level structure disputes UCL's existence as a

Table 6
Clustering within the Sco-Cen Association

SC	LEAF	Name	N	R.A. (deg)	Decl.	l (deg)	b	D_{sky}[a] (deg)	d (pc)	$\mu_{R.A.}$ (mas yr^{-1})	$\mu_{decl.}$	$V_{T,l}$ (km s^{-1})	$V_{T,b}$	σ_{V_T}[b] (km s^{-1})	Age (Myr)
1		β Pictoris	23	279.4	−43.6	351.4	−16.6	29.0 × 13.2	57.4 ± 10.4	9.7	−70.3	−16.0	−9.8	2.1 × 0.7	23.3 ± 2.1
2			12	120.1	−59.9	273.2	−15.2	5.7 × 3.9	113.6 ± 7.8	−8.3	21.9	−12.4	1.7	1.3 × 0.5	43.2 ± 2.6
3		IC 2602	101	161.8	−64.6	290.1	−5.0	4.7 × 2.8	152.6 ± 6.7	−18.2	10.5	−15.2	0.7	1.0 × 0.8	40.5 ± 0.8
4		Platais 8	106	138.3	−57.7	277.0	−6.4	5.2 × 4.5	137.5 ± 6.6	−16.3	13.0	−13.5	−1.9	1.0 × 0.8	37.0 ± 0.9
5			14	124.9	−50.8	266.6	−8.2	5.6 × 2.2	162.3 ± 6.6	−8.7	9.8	−10.0	−1.4	1.0 × 0.3	35.8 ± 2.0
6		Pipe	16	257.4	−27.5	356.8	7.4	1.9 × 0.3	160.2 ± 1.9	−1.2	−18.6	−12.0	−7.5	1.1 × 0.5	3.5 ± 0.7
7			12	266.7	−39.7	350.9	−5.8	3.3 × 2.2	141.9 ± 4.9	−7.8	−29.4	−19.7	−5.6	0.6 × 0.4	14.0 ± 1.2
8		CrA	248	281.4	−36.4	359.2	−14.6	4.8 × 2.6	150.4 ± 5.5	1.6	−27.4	−17.5	−8.8	1.1 × 0.7	13.3 ± 0.9
9			15	258.5	−24.2	0.1	8.5	5.6 × 3.0	104.1 ± 3.3	−13.3	−33.8	−17.4	−4.2	0.8 × 0.4	15.6 ± 1.6
10		Theia 67	28	259.4	−24.8	0.0	7.5	3.4 × 1.8	149.3 ± 3.0	−4.8	−21.7	−14.6	−6.0	0.4 × 0.3	20.2 ± 2.0
11		UPK 606	36	216.2	−46.4	319.2	13.5	2.0 × 1.6	169.3 ± 7.5	−20.4	−16.7	−20.1	−6.6	0.6 × 0.4	16.9 ± 1.2
12		Lupus	102	241.5	−39.6	338.9	9.4	3.6 × 2.0	160.8 ± 3.9	−10.4	−23.5	−18.0	−7.8	0.8 × 0.6	5.9 ± 0.6[c]
12	A	Lup IV	46	242.2	−39.1	339.6	9.4	1.0 × 0.4	159.8 ± 2.7	−10.1	−23.5	−17.7	−7.7	0.6 × 0.5	5.4 ± 0.8
12	B	Lup III	14	239.8	−42.1	336.2	8.4	0.8 × 0.4	161.2 ± 2.3	−11.0	−23.4	−18.1	−8.0	0.4 × 0.3	5.9 ± 1.3
13			30	240.1	−24.9	348.3	20.9	3.7 × 3.3	111.9 ± 5.7	−21.2	−32.3	−19.9	−4.4	0.6 × 0.3	25.9 ± 1.5
14			20	242.9	−44.6	336.2	5.0	2.9 × 1.2	185.3 ± 4.2	−13.8	−19.9	−20.8	−4.4	0.4 × 0.2	14.1 ± 1.2
15		Lower Sco	370	250.2	−39.7	343.5	4.5	3.3 × 1.8	176.3 ± 6.3	−12.1	−21.3	−20.0	−4.1	0.6 × 0.4	18.8 ± 0.6
16			12	256.5	−36.7	348.9	2.5	3.6 × 2.2	134.0 ± 3.7	−7.2	−29.1	−17.5	−7.5	0.4 × 0.2	18.5 ± 1.6
17		US	1478	243.7	−23.6	351.8	19.4	6.7 × 4.7	146.4 ± 9.3	−10.5	−23.6	−16.9	−5.8	1.5 × 1.1	11.3 ± 0.3[c]
17	A		11	243.9	−23.9	351.7	19.1	3.2 × 1.4	135.4 ± 3.2	−15.9	−24.7	−18.6	−3.3	0.4 × 0.2	16.2 ± 2.0
17	B		11	247.0	−25.7	352.3	15.8	3.0 × 1.5	140.4 ± 1.5	−11.5	−23.2	−16.7	−4.5	0.4 × 0.4	12.5 ± 1.0
17	C		18	245.1	−24.7	351.9	17.7	1.6 × 0.7	159.3 ± 4.0	−10.4	−20.5	−16.8	−4.6	0.4 × 0.4	12.2 ± 0.6
17	D		22	244.0	−24.9	351.0	18.3	1.5 × 0.9	160.7 ± 2.8	−11.0	−22.4	−18.2	−5.6	0.4 × 0.3	11.2 ± 0.9
17	E		54	242.7	−19.1	354.7	23.1	1.2 × 0.9	140.0 ± 3.3	−8.8	−24.2	−15.8	−6.5	0.4 × 0.3	7.2 ± 0.7
17	F		25	245.3	−22.2	354.0	19.2	1.5 × 1.3	135.3 ± 2.3	−11.6	−25.0	−16.9	−5.1	0.4 × 0.2	12.6 ± 1.1
17	G		29	241.1	−19.6	353.1	23.9	1.1 × 0.7	152.9 ± 2.3	−10.2	−21.4	−16.4	−5.3	0.3 × 0.3	13.2 ± 2.8
17	H		102	240.6	−22.4	350.6	22.3	2.4 × 2.2	143.3 ± 2.6	−11.9	−23.7	−17.1	−5.6	0.6 × 0.5	10.2 ± 0.7
17	I	ρ Oph	110	246.4	−23.9	353.3	17.4	1.5 × 1.0	138.7 ± 2.6	−7.2	−25.9	−15.9	−7.7	0.7 × 0.5	5.7 ± 0.4
18		η Cha	17	132.3	−79.0	292.7	−21.3	2.3 × 1.7	98.3 ± 1.7	−30.7	26.1	−18.1	−5.2	0.3 × 0.2	8.2 ± 1.7
19			10	229.9	−37.8	332.6	16.4	1.7 × 0.5	134.6 ± 2.4	−19.0	−26.9	−19.6	−7.5	0.4 × 0.3	15.3 ± 2.8
20			77	227.7	−44.0	327.8	12.1	3.2 × 1.7	147.4 ± 4.9	−20.8	−21.7	−20.3	−5.4	0.4 × 0.3	22.8 ± 1.0
21			12	212.5	−50.5	315.5	10.5	2.7 × 1.6	134.7 ± 2.5	−25.9	−19.6	−19.6	−6.9	0.9 × 0.3	23.9 ± 2.9
22			102	239.5	−38.9	338.1	10.9	4.9 × 2.9	140.6 ± 4.6	−18.0	−27.1	−20.9	−5.7	0.4 × 0.3	22.7 ± 1.1
23			12	233.6	−35.9	336.3	16.1	3.7 × 1.1	143.6 ± 4.3	−19.1	−24.2	−20.3	−5.3	0.3 × 0.2	21.5 ± 2.1
24			108	206.8	−43.8	313.5	17.9	6.9 × 2.5	134.3 ± 6.6	−25.7	−19.3	−18.7	−8.3	0.5 × 0.3	20.8 ± 1.1
25			11	209.8	−52.5	313.3	9.0	3.5 × 1.5	117.2 ± 2.5	−29.3	−22.2	−18.9	−7.6	0.2 × 0.1	20.7 ± 2.0
26			40	198.1	−49.7	306.4	13.0	3.5 × 2.0	123.0 ± 5.3	−30.3	−17.2	−18.4	−8.5	0.5 × 0.4	21.7 ± 1.1
27		LCC	697	187.3	−60.1	300.1	2.5	13.9 × 6.0	109.5 ± 6.4	−36.1	−12.2	−18.1	−8.0	0.8 × 0.4	NC
27	A	ϵ Cha	17	180.1	−78.5	300.3	−15.9	1.3 × 1.1	101.9 ± 1.2	−41.2	−5.8	−18.9	−6.8	0.3 × 0.2	8.3 ± 1.0
27	B		17	186.0	−72.4	300.8	−9.7	2.7 × 1.4	102.0 ± 2.1	−39.7	−9.9	−18.6	−6.8	0.1 × 0.1	13.0 ± 1.4
27	C	Crux S[d]	100	186.5	−64.4	300.2	−1.6	4.1 × 2.2	106.8 ± 2.3	−37.7	−11.3	−18.4	−7.6	0.3 × 0.2	14.6 ± 0.8
27	D		12	181.9	−51.4	296.0	10.9	1.8 × 1.3	114.8 ± 2.2	−34.4	−10.2	−17.5	−8.7	0.2 × 0.1	23.0 ± 2.3
27	E		48	185.0	−57.1	298.7	5.4	6.0 × 3.4	109.3 ± 2.6	−35.5	−11.7	−17.6	−8.3	0.4 × 0.3	18.5 ± 0.7
…[e]	–	TW Hydrae	19	179.7	−36.2	290.9	25.1	14.7 × 6.2	62.9 ± 10.9	−67.8	−23.0	−17.6	−11.0	1.3 × 0.5	10.8 ± 0.9

Notes. Here SC is the Sco-Cen ID, a number from EOM clustering used for identification. Leaf clusters within an EOM cluster are given a subcluster ID, which is given in the LEAF column. In the age column, NC marks subclustered groups with non-coeval populations.

[a] On-sky spatial extent in galactic l/b, in the form of the rms in major axis × minor axis, when fit with a bivariate Gaussian.

[b] The radial extent of the velocity distribution, in the form of the rms in semimajor axis × semiminor axis, when fit with a bivariate Gaussian.

[c] Has substructure, but most of the region is close enough to coeval that a global age estimation can be meaningful.

[d] Name is newly assigned by this paper.

[e] Defined through crossmatching the Sco-Cen candidate young stars with known TW Hydrae members, not HDBSCAN clustering.

physical structure, although the label remains useful to refer to the area more generally. One of these subgroups contained within the de Zeeuw et al. (1999)–defined extent of UCL is SC-12, or Lupus, which is the third and final group with substructure we identify. The known subregions of Lupus III and Lupus IV correspond to the defined lower-level subgroups of SC-12B and SC-12A, respectively, and some members of Lupus II also appear in SC-12, although these are too sparse for HDBSCAN to identify a subcluster (Galli et al. 2013). The lack of stars detected in Lupus II and near-absence of Lupus I are likely due to heavy extinction in much of that group (Comerón 2008), which leads to dramatically higher rates of rejection based on Gaia quality cuts (see Section 3.2).

The LCC (SC-27) region contains five subgroups, the largest of which is SC-27C, a dense group containing nearly one-seventh of the total population in LCC. This substantial subgroup does not appear to have been previously identified independently from LCC, so we name it the Crux South group, after its parent constellation. Some of the Chamaeleon clusters overlap with LCC and Sco-Cen more generally near LCC's south edge, with SC-27A matching with catalogs for the known ϵ Chamaeleontis cluster (e.g., Murphy et al. 2013). The η Chamaeleontis cluster is located near ϵ Chamaeleontis; however, there is little connecting it to the rest of LCC, hence it being identified as separate under the ID SC-18. Chamaeleon I and II, the other two previously known Chamaeleon-associated clusters, also share kinematics with Sco-Cen, although they are much farther away and apparently non-contiguous with the rest of Sco-Cen, so HDBSCAN does not include them as Sco-Cen subgroups. The relation between these more distant Chamaeleon clusters and Sco-Cen is discussed in Section 4.4.4. Aside from the aforementioned notable features, the rest of LCC contains relatively sparse and small subgroups.

The US (SC-17) subgroup is the largest identified in Sco-Cen, and it also contains the most substructure, with nine subgroups. The largest of these, SC-17I, is centered on ρ Ophiuchi, which is located adjacent to sites of active star formation in and around the L1688 cloud (e.g., Kerr et al. 2019). The rest of US, however, appears to be composed of largely new substructure. Subgroups C and D of SC-17 represent populations distributed nearly 20 pc behind the active star-forming central region around ρ Oph, while subgroups E, G, and, to a lesser extent, F form an arc of dense and populous subgroups located to the north of ρ Oph. The remainder of the subregions are very close to SC-17I in spatial coordinates but reside up to $\sim$5 km s^{-1} away from SC-17I in transverse velocity space, hence their division by HDBSCAN.

In the remainder of Sco-Cen within the de Zeeuw et al. (1999) boundaries, we identify a wealth of diverse subgroups stretching from LCC through to the traditionally defined extent of UCL. The largest of the groups in this region by far is Lower Sco (SC-15), a known clustered group to the area's southeast (Mamajek 2013; also presented as the V1062 Sco moving group in Röser et al. 2018). However, Lower Sco is a relative outlier in spatial coordinates, representing a significant protrusion behind most other groups in the area. In a similar way to Lower Sco, SC-14 also protrudes behind the rest of Sco-Cen and holds an overlapping velocity distribution, possibly indicating a link between the two. Identified as UPK 606 in the Sim et al. (2019) catalog, SC-11 represents another clear protrusion behind Sco-Cen, although this one is much further west. The remainder of the groups in this region follow a fairly contiguous arc of young substructures, with a slight lower-density gap between the groups more closely connected to LCC (SC-24, 21, 25, and 26) and those that fall within the traditional UCL region (SC-12, 22, 23, 19, and 20). Group SC-19 is uniquely dense, with most stars within 15′, making it worth revisiting as a small bound cluster.

The subgroups we identify outside of the de Zeeuw et al. (1999) boundaries for Sco-Cen almost entirely fall into two regions: the southwest, which contains the far south of SC-27, as well as some more western groups, including IC 2602; and the southeast, which contains structure leading to and including CrA. We refer to the subset of the southwestern extension that extends to IC 2602 and beyond as the IC 2602 branch, which includes IC 2602 (SC-3), along with SC-2, SC-4, and SC-5. This is a region that has a somewhat weaker connection to the rest of Sco-Cen, with velocities that are visibly distinct from but still contiguous with the rest of Sco-Cen (see Figure 18). The southeastern extensions are mostly much closer to the known regions of Sco-Cen in velocity space and are anchored by the large CrA association. This is a well-known group; however, it has not typically been linked to Sco-Cen due to its apparent lack of any structure connecting it to the rest of the association. We identify this previously unknown linking structure in the form of four subgroups in the space between US, Lower Sco, and CrA (SC-7, SC-16, SC-6, and SC-10), along with further low-density populations in between subgroups. Group SC-9 is in the same area in the plane of the sky but actually projected in front of these other groups. The inclusion of CrA as a new subregion of Sco-Cen is therefore well supported by both the kinematics from Gaia and the structural layout of the region.

The last of these outlying groups, SC-1, appears to correspond to the β Pictoris moving group. Known β Pic member η Telescopii is the brightest member we identify (Zuckerman et al. 2001), while 10 of the remaining 22 SC-1 members are labeled as β Pic members by Shkolnik et al. (2017), found distributed fairly evenly throughout our population. Our recognition of β Pic members is, however, limited to stars more distant that about $\sim$40 pc, as geometric effects badly skew the transverse velocity vector for clustering nearer populations. While a link between β Pic and Sco-Cen has been proposed in the past (Preibisch & Mamajek 2008), this claim was made on the basis of common motion, rather than the presence of stellar populations linking them. Through our clustering analysis, we show that low-density connections between β Pictoris and the rest of Sco-Cen do exist, representing the first direct evidence that β Pic is fully contiguous with Sco-Cen.

4.4.2. TW Hydrae

TWA is a very nearby young stellar population with an average distance of $\sim$60 pc (Reid 2003). Due to its proximity, TWA is very well studied, and detailed investigations of its membership and kinematics have strongly suggested a connection to the Sco-Cen association, particularly through LCC, the nearest traditional subregion (e.g., Mamajek et al. 1999). While we do not directly identify a subgroup consistent with TWA due to the region's low relative density, many TWA members appear in our Sco-Cen population. In Figure 22, we crossmatch currently known TWA members (Reid 2003; Song et al. 2003; Gagné et al. 2017) with our

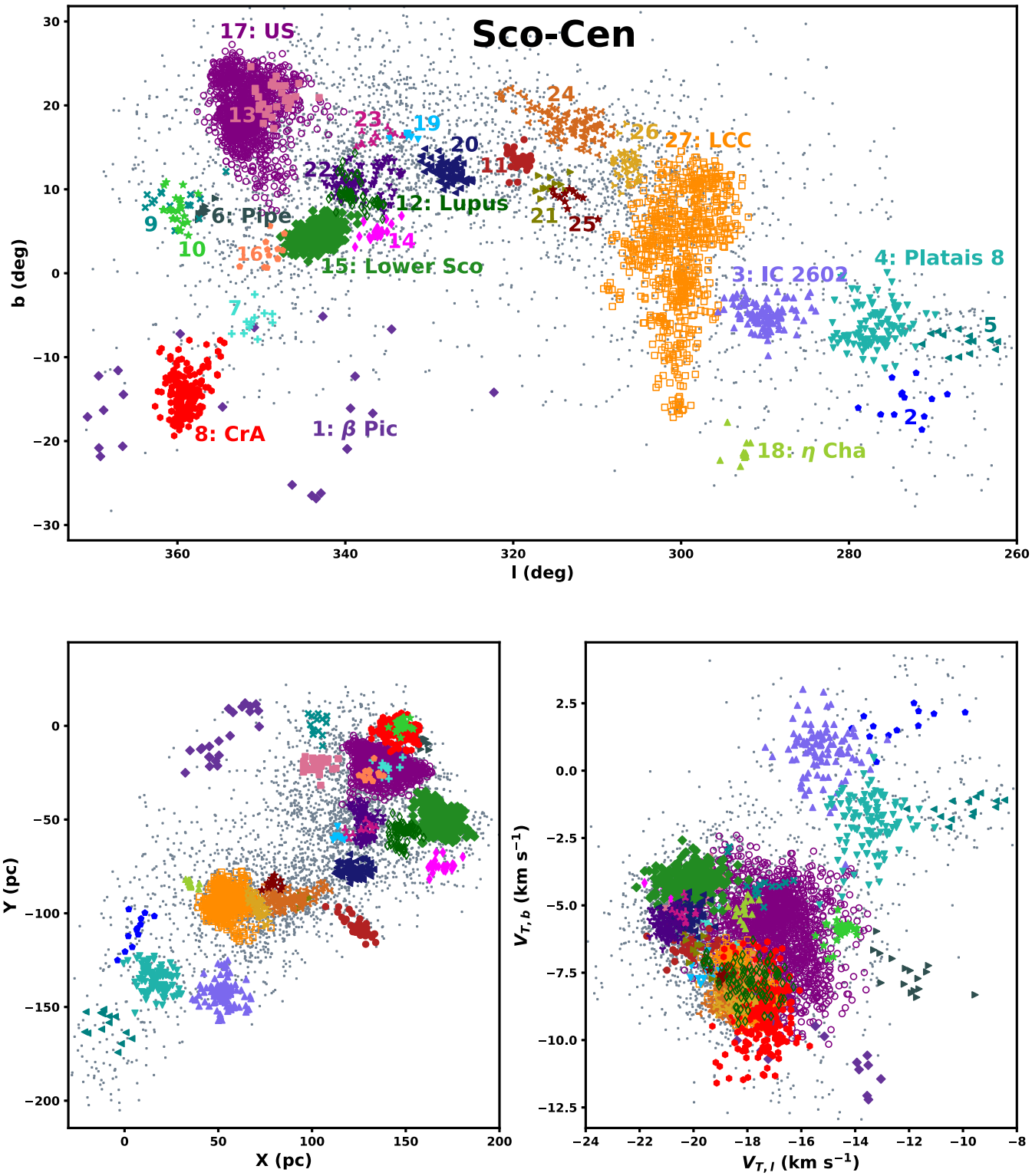

Figure 18. Sco-Cen EOM clustering from HDBSCAN. The top panel displays l/b sky coordinates, the bottom left panel shows galactic X/Y, and the bottom right panel shows l/b transverse velocities. Subgroups are labeled in the top panel, and the corresponding symbols and colors are consistent across the three panels. Three regions (US, LCC, and Lupus) have internal subclustering, and those are marked with open symbols. Gray points represent Sco-Cen members not assigned a subcluster.

complete Sco-Cen population, revealing that nearly half of that known sample is contained within our population of photometrically young Sco-Cen members. All appear along an extended spur of stars that reaches from the near side of LCC in the direction of the Sun, following a distribution fully contiguous with the rest of Sco-Cen. The identification rates for the known TWA members within our Sco-Cen sample appear to taper off for objects on the region's western edge with a distance of less than 40–50 pc from the Sun; however, this is likely due to the transverse velocities, which are increasingly

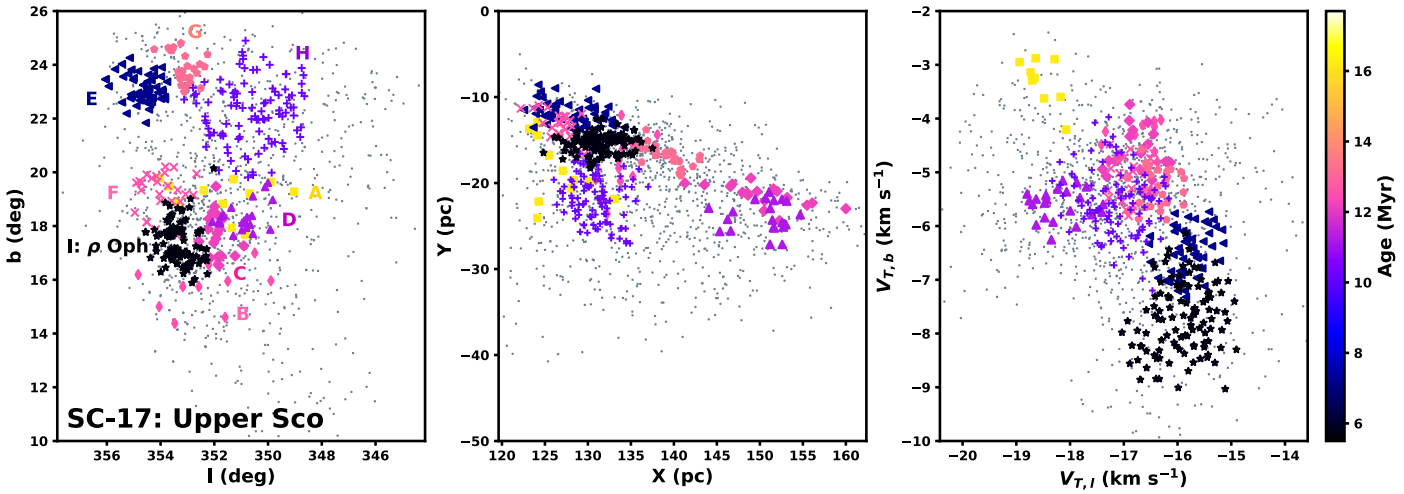

Figure 19. Subclustering in Sco-Cen's Upper Sco subgroup, presented in l/b sky coordinates, X/Y galactic coordinates, and l/b transverse velocity. The US subgroups are marked according to their Sco-Cen subgroup IDs from Table 6, and unclustered US members are shown as small gray dots.

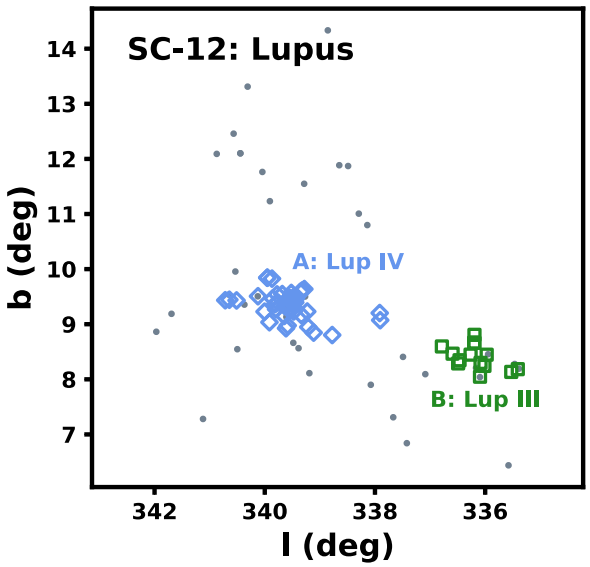

Figure 20. Clustering within Sco-Cen's Lupus subgroup in l/b sky coordinates. Subclusters are marked, and unclustered Lupus members are presented as gray dots. Both subgroups are essentially newborn.

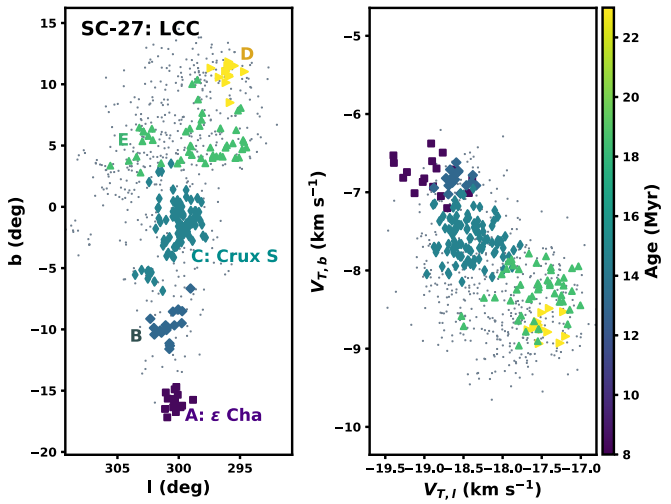

Figure 21. Subclustering in Sco-Cen's Lower Centaurus–Crux subgroup, presented in l/b sky coordinates and l/b transverse velocity. Groups are labeled using the Sco-Cen subgroup IDs from Table 6, and the gray points are unclustered LCC members.

sensitive to geometric effects from their position in space as they get closer to the Sun, therefore making clustering more difficult, especially for objects farther from the line of sight with the rest of the Sco-Cen members. Regardless, the overlap in these populations that we uncover represents the most direct evidence to date of a physical link between TWA and the LCC subregion of Sco-Cen, a link that has been suggested by multiple recent papers (e.g., Mamajek & Feigelson 2001; Murphy et al. 2015).

4.4.3. Age Structure

As shown in Figure 23 and Table 6, ages in the Sco-Cen association range from essentially newborn, such as those in Lupus or the ρ Ophiuchi complex, to over 20 Myr old. Aside from the older populations around IC 2602, all groups with ages older than 20 Myr fall along the northern edge of Sco-Cen, following an arc of old star formation stretching from the north edge of LCC, across the northern and western edge of UCL (including SC-20, SC-21, SC-22, SC-23, SC-24, SC-25, and SC-26), to SC-13 in the foreground of US at the border with Libra. This arc of old star formation will hereafter be referred to as the Libra-Centaurus arc (LCA), in reference to the constellations at the approximate end points of this structure. This arc of old star formation appears to represent a possible starting point for star formation in the Sco-Cen association, as the age distributions in much of the rest of the association can be explained by sequential star formation originating in the LCA.

In most cases, the farther a star is away from the LCA, the younger it is found to be. This trend is most plainly evident in LCC, which has a smooth age gradient from galactic north to south, clearly visible in Figures 21 and 23. Subgroup SC-27D, which we consider part of the LCA, has an age of 23 Myr, while more southerly groups range from SC-27E at ∼18.5 Myr in the north to ϵ and η Chamaeleontis at ∼8 Myr in the south. Other features that protrude from the LCA are similarly young, such as SC-11, which has an age of ∼17 Myr, about 6 Myr younger than the adjacent LCA groups. TWA, which extends

toward the Sun from the northern LCC, is also young compared to adjacent LCA groups, with an age estimated at 8–10 Myr in existing literature (Weinberger et al. 2013; Ducourant et al. 2014; Murphy et al. 2015) and ∼11 Myr in our isochronal age fit. CrA and surrounding groups may also have an age gradient toward the LCA. The southernmost tip of CrA is associated with the CrA dark clouds and is therefore essentially newborn (e.g., see Galli et al. 2020), while the age solution for the broader population centered closer to the LCA is ∼13 Myr, and SC-7 and SC-16, two separate groups even nearer to the LCA, have ages of ∼14 and ∼18.5 Myr, respectively. However, the stellar populations connecting the LCA to CrA are much more sparse, and the distances between subgroups are much greater, making this connection less clear. We investigate Sco-Cen's age trends in the context of propagation rates for sequential star formation in Section 5.2.

These smooth age gradients toward the LCA are not universal, however, particularly for US and groups in the direction of Lupus and Lower Sco. These different age patterns come in the form of both internal age structure, typically with younger stars in the denser regions, such as in US and CrA, and larger gradients that oppose those off the LCA, such as the area between Lupus and Lower Sco. The age gradients in CrA and Lower Sco are given close-ups in Figure 24. The sections of UCL closest to the LCA are older, as expected, showing a subtle age gradient like in LCC, but some of the more distant groups, such as Lupus, which contains essentially newborn stars, break that smooth progression. Lupus represents an anomalously young group for its location in Sco-Cen, and the youth of the region does not continue past Lupus, as Lower Sco and SC-14 have ages of ∼19 and 14 Myr, respectively. There is, however, an age gradient in Lower Sco that is younger in the direction of Lupus, suggesting that the two events may be directly connected.

The other very young region in Sco-Cen, US, is considerably younger than its surroundings, and the transition from the LCA to the adjacent younger populations is quite abrupt. This region is generally younger toward its center, with the large central ρ Ophiuchi group being the youngest. However, most of our US ages outside ρ Oph, including our US at-large solution that assumes common ages across the entire region, converge between 10 and 13 Myr, a result consistent with the US age solutions from Pecaut et al. (2012) and Pecaut & Mamajek (2016) and more recently supported by Sullivan & Kraus (2021). The only subgroup exceeding this age range at ∼16 Myr, SC-17A, is also one of the nearest subgroups to the LCA, so its older age is consistent with our narrative of star formation propagating away from the LCA. The first bursts of star formation in US may have therefore originated relatively early in the history of Sco-Cen, with activity peaking ∼10–13 Myr ago.

The IC 2602 branch appears to contain some of the oldest stars in our sample; however, we treat these groups with caution when considering them in the context of star formation in Sco-Cen more generally. The relatively different velocities in this area may suggest a degree of removal from the star formation processes in the rest of Sco-Cen, and the lack of smooth age gradients leading off the IC 2602 branch supports this. This region may therefore represent an older outbreak of star formation that predates the formation of much of the rest of Sco-Cen, making its influence on the more recent star formation events in Sco-Cen doubtful.

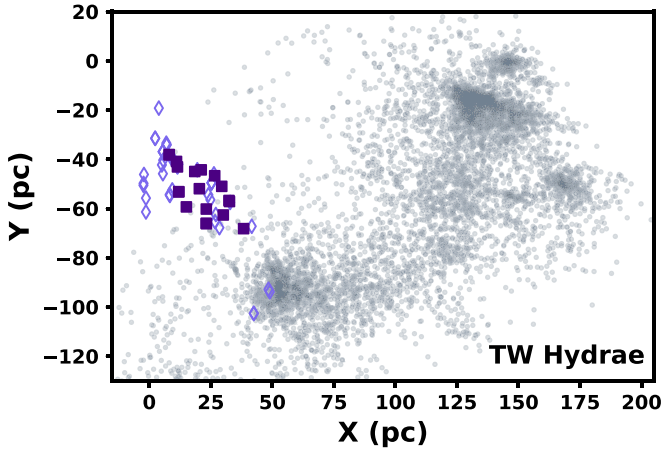

Figure 22. Our members for Sco-Cen (light gray), compared to TWA members (Song et al. 2002; Reid 2003; Gagné et al. 2017). Filled squares are in both our Sco-Cen set and TWA, while open diamonds are other TWA members.

Our overview of the ages in Sco-Cen suggests that smooth, sequential star formation originating along the LCA explains most of the age patterns in Sco-Cen, presenting a much more continuous picture of star formation compared to the set of discrete triggered star-forming events proposed in Preibisch & Mamajek (2008). However, a few notable regions do not fully adhere to these age trends, requiring further investigation to make sense of those outliers. Krause et al. (2018) proposed a star formation scenario that explains the presence of two of these age trend outliers (Lupus and ρ Oph), suggesting that Sco-Cen formed within an extended gas overdensity such that when the earliest generation formed, it preferentially pushed gas out perpendicular to the gas overdensity's long axis, creating superbubbles that later engulfed sparser material along the long axis and guided its collapse. The coeval populations we find along nearly the entire length of Sco-Cen in the form of the LCA are largely inconsistent with this star formation sequence, as our results demand a much broader star formation genesis than the individually coeval star-forming events discussed in Krause et al. (2018). However, this surround-and-squash explanation may contribute to the methods of formation elsewhere in Sco-Cen, particularly in regions like US and Lupus that are somewhat removed from both the LCA and the associated age gradients. Lower Sco may serve an important role in applying the surround-and-collapse concept to our updated view of Sco-Cen, as the group is relatively old and lies opposite to the LCA relative to Lupus and, to a lesser extent, US. As such, a surround-and-collapse pattern guided by superbubbles from the LCA and Lower Sco may provide an explanation for the existence of both of these sites, while the rest of Sco-Cen appears well explained by a sequential star formation pattern, as discussed in Section 5.2.

4.4.4. Chamaeleon

While two of the nearby young clusters in Chamaeleon were identified as Sco-Cen–affiliated by HDBSCAN, the somewhat more distant clusters were not, including the clusters associated with the actively star-forming Chamaeleon I and II clouds (Luhman 2008). We identify these known clusters as CHA-1 and CHA-2, respectively, and they are nestled about 80 pc behind LCC and the near Chamaeleon clusters in the plane of the sky (Lopez Martí et al. 2013). Group CHA-3 represents a third relatively compact and populous group in this region, and,

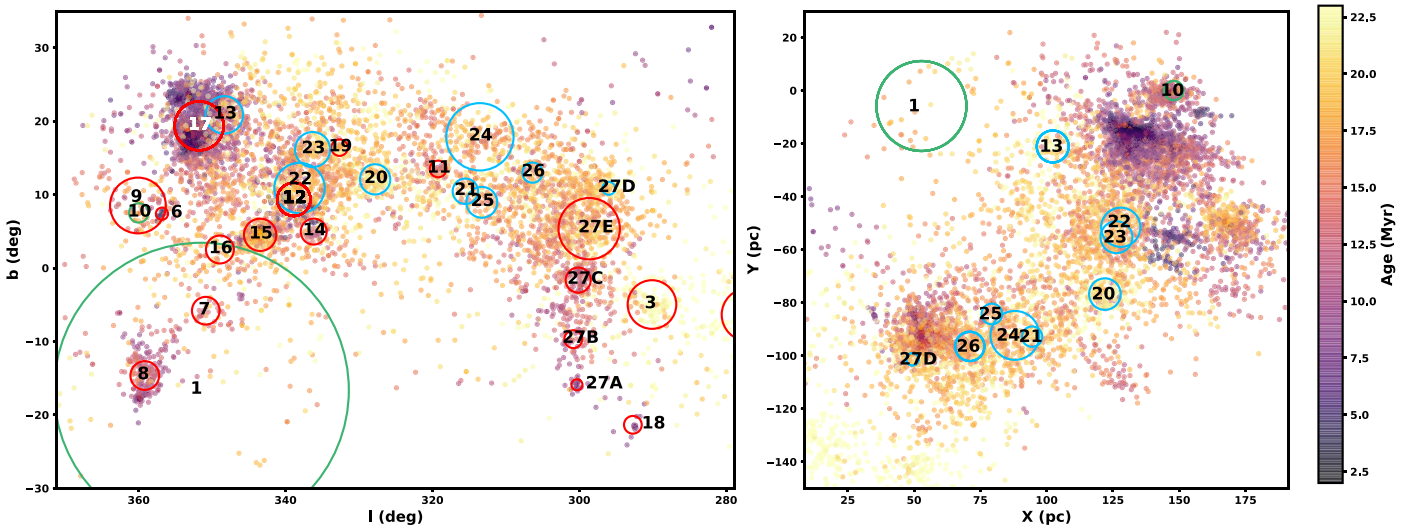

Figure 23. Sco-Cen members plotted in l/b spatial coordinates and X/Y galactic coordinates, color-coded by the median age fit of the 10 nearest neighbors. The groups presented in Figure 18 are labeled and marked with circles that represent the locations and effective radii of each group, with the subclusters in LCC shown as opposed to the top-level group. Most of those outlines are in red; however, groups not on the IC 2602 branch that have age $>$ 20 Myr are marked in blue or green, with the blue outlines marking groups we associate with the LCA.

due to its distinctness as a clustered extension to Chamaeleon, we assign it the name Centaurus South. The last Chamaeleon subgroup, CHA-4, has limited overlap with Theia 94 from Kounkel & Covey (2019) and a much sparser spatial distribution compared to the other, much denser Chamaeleon clusters. Its identification as an independent group can be attributed to its notably different velocity distribution, which, like the other Chamaeleon groups, has a relatively tight $\sim$1 km s^{-1} velocity dispersion, despite its considerably larger extent. The age solutions for the Cha I and II clusters are essentially newborn, as expected, with Centaurus South being older at $\sim$11 Myr. Group CHA-4 is by far the oldest of the groups we identify in Chamaeleon, at $\sim$37 Myr. The properties of all four Chamaeleon subgroups are listed in Table 7.

As shown in Figure 25, despite their physical separation, Cha I, Cha II, and Centaurus South all have similar transverse velocities to both Sco-Cen and each other. The radial velocities for the Centaurus South group cannot be checked due to a lack of past coverage; however, radial velocities for Cha I and II have been reported as 12.9 ± 1.6 and 14.6 ± 1.2 km s^{-1}, respectively (Lopez Martí et al. 2013). This is very similar to the values in LCC, the nearest Sco-Cen subgroup, which has a radial velocity of $\sim$12 km s^{-1} (de Zeeuw et al. 1999). The similarities of the kinematics of these peripheral clusters to the stars in Sco-Cen has previously been used to suggest that the material forming stars in Chamaeleon may be related to that in Sco-Cen (Preibisch & Mamajek 2008); however, we find no evidence for the existence of lower-density stellar populations between the two associations that might suggest a direct structural link.

4.5. Minor Groups

Aside from Taurus, Perseus, Orion, Sco-Cen, and Chamaeleon, 22 other small clusters and groups are also identified in this paper. We assign all of these top-level groups a name in accordance with either past literature or their location in the sky, depending on whether the group has been firmly established in the literature yet. Associations that are known but have not yet been assigned an at-large name in the literature

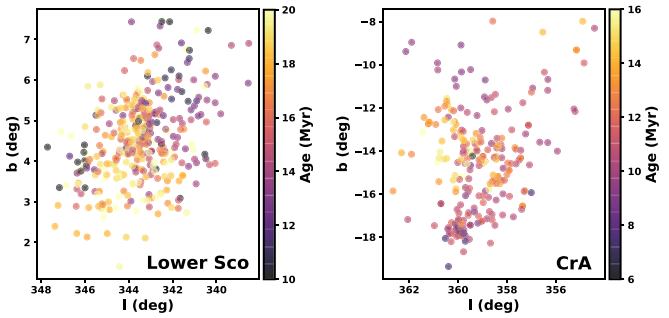

Figure 24. Age distributions in CrA and Lower Sco according to the median age of the 10 nearest neighbors. Both show strong visible gradients, with Lower Sco younger in the direction of Lupus and CrA younger toward a tight clump at the far south, which is associated with recent formation in the CrA molecular cloud.

are assigned one that is descriptive of their location in the sky. While many of these groups have catalog identifications in Kounkel & Covey (2019) or Sim et al. (2019), their youth and proximity make them worth highlighting, hence the inclusion of distinctive noncatalog names. The names used to refer to each of these groups are listed in Table 3, along with basic information on each cluster.

Some of these groups are firmly connected to known objects, including open clusters IC 2391, NGC 2451A, ASCC 123, and the Pleiades, as well as somewhat larger and better-known associations, including Perseus OB3, the Cepheus Flare region, and both Vela subgroups. Other groups, such as Cepheus Far North (CFN) and Monoceros Southwest (MSW), are present in the literature but not yet well established (see the discussion in Sections 4.5.4 and 4.5.2), while the remaining 12 are either absent from the literature completely or only noted as part of very large clustering surveys, such as Kounkel & Covey (2019) and Sim et al. (2019). The Sim et al. (2019) catalog includes Taurus-Orion III as UPK 385, as well as Carina-Musca subgroup CM-2, which is identified as UPK 569. The Theia clusters from Kounkel & Covey (2019) include another five of the 10 remaining clusters: Ophiuchus Southeast (Theia 70), Canis Major North (Theia 69), Aquila East (Theia 53),

Table 7
Subclustering Inside Smaller Regions

TLC	ID	Name	N	R.A. (deg)	Decl. (deg)	l (deg)	b (deg)	D_{sky}[a] (deg)	d (pc)	$\mu_{R.A.}$ (mas yr^{-1})	$\mu_{decl.}$ (mas yr^{-1})	$V_{T,l}$ (km s^{-1})	$V_{T,b}$ (km s^{-1})	σ_{V_T}[b] (km s^{-1})	Age (Myr)
PER	1A	NGC 1333	19	52.3	31.0	158.5	−20.7	2.2 × 0.6	296.9 ± 8.3	7.4	−8.2	15.3	−2.9	1.2 × 0.7	6.0 ± 3.2
PER	1B		77	58.5	32.8	161.7	−16.0	3.1 × 2.4	279.3 ± 12.1	6.4	−10.0	15.1	−4.3	0.8 × 0.6	17.5 ± 0.9
PER	2A	IC 348	42	55.8	32.2	160.2	−18.0	1.3 × 0.5	320.6 ± 7.2[c]	4.7	−6.3	11.6	−2.9	1.0 × 0.9	4.7 ± 0.5
PER	2B		50	60.3	33.2	162.5	−14.7	4.1 × 2.4	308.5 ± 16.3[c]	3.8	−6.3	10.4	−3.0	0.7 × 0.7	17.1 ± 1.1
POB3	1	α Per	89	52.5	48.1	148.4	−6.6	8.5 × 4.8	166.3 ± 15.5	23.5	−27.0	27.2	−6.9	1.0 × 0.7	49.7 ± 0.5
POB3	2		24	68.7	28.7	171.0	−12.6	5.3 × 3.8	181.7 ± 9.1	17.4	−23.4	25.0	−1.9	1.2 × 0.9	30.0 ± 2.5
MSW	1		12	93.5	3.1	205.9	−6.9	1.9 × 1.5	322.4 ± 4.9	−4.2	−3.9	2.3	−8.4	0.9 × 0.7	25.3 ± 2.6
MSW	2	NGC 2232	23	97.0	−4.8	214.5	−7.4	1.6 × 0.9	323.2 ± 7.9[c]	−4.7	−1.9	−0.6	−7.7	0.8 × 0.3	27.8 ± 1.2
MSW	3		16	98.8	−4.2	214.9	−5.6	3.9 × 2.4	223.4 ± 8.0	−7.3	−3.6	−0.1	−8.6	0.7 × 0.4	24.8 ± 2.2
MSW	4	LP 2439	123	103.0	−5.9	218.3	−2.5	3.1 × 2.1	284.2 ± 15.0	−7.3	−2.6	−1.4	−10.3	0.8 × 0.5	27.0 ± 0.7
VCG4	1	UPK 535	21	127.0	−52.4	268.7	−8.1	3.2 × 2.2	310.9 ± 8.9	−13.0	3.6	−15.4	−12.6	1.2 × 0.4	32.2 ± 1.5
VCG4	2		29	105.3	−53.8	264.0	−20.2	9.6 × 2.5	198.7 ± 8.3	−12.7	11.8	−14.4	−7.8	0.9 × 0.7	37.8 ± 1.3
VCG4	3		10	112.3	−46.2	258.4	−13.2	4.2 × 2.1	258.7 ± 2.8	−12.4	7.6	−14.9	−9.7	0.7 × 0.4	39.2 ± 2.0
VCG4	4	Cr 135/UBC 7	145	107.5	−37.2	248.5	−12.6	3.0 × 2.6	285.0 ± 21.9	−10.1	6.8	−14.0	−8.5	1.1 × 0.6	34.8 ± 0.5
VCG4	5		48	119.0	−44.5	259.1	−8.2	4.3 × 1.2	321.6 ± 6.9[c]	−11.6	4.5	−14.9	−11.8	0.9 × 0.6	32.1 ± 1.0
CFN	1		174	334.4	76.1	113.7	17.9	15.3 × 4.5	171.3 ± 15.9	16.2	5.6	14.5	−2.8	2.3 × 0.9	23.8 ± 1.3
CFN	2		14	319.6	70.2	106.7	14.5	3.9 × 2.4	235.1 ± 7.0	10.1	6.4	12.9	−3.5	0.8 × 0.3	25.8 ± 3.7
CHA	1	Cha I	112	167.2	−76.8	297.1	−15.1	1.5 × 0.8	191.2 ± 5.2	−22.5	0.5	−18.9	−7.7	1.0 × 0.7	4.6 ± 0.4
CHA	2	Cha II	33	195.7	−77.4	303.6	−14.5	2.2 × 0.8	199.1 ± 4.4	−20.2	−7.3	−19.3	−6.0	0.9 × 0.5	3.7 ± 0.8
CHA	3	Cen South[d]	39	206.6	−63.4	309.2	−1.3	7.0 × 3.7	186.6 ± 8.2	−20.0	−12.3	−19.6	−7.0	1.5 × 0.4	11.4 ± 1.9
CHA	4		17	195.6	−80.7	302.1	−20.2	14.0 × 3.0	192.7 ± 18.4	−8.9	−8.3	−11.9	−9.7	1.1 × 0.4	36.5 ± 1.7
CM	1		23	133.8	−63.8	280.1	−12.1	6.1 × 1.2	208.6 ± 5.9	−27.4	14.7	−28.3	−12.0	0.7 × 0.5	33.5 ± 2.5
CM	2		75	163.9	−70.7	293.5	−10.3	7.2 × 2.0	254.5 ± 21.3	−25.0	0.9	−27.7	−12.4	1.0 × 0.6	33.9 ± 0.7

Notes. The parent top-level cluster is reported in the TLC column, and the ID column gives the EOM subcluster, with Perseus divided further into its older and younger A and B subregions.
[a] On-sky spatial extent in galactic l/b, in the form of the rms in major axis × minor axis, when fit with a bivariate Gaussian.
[b] The radial extent of the velocity distribution, in the form of the rms in semimajor axis × semiminor axis, when fit with a bivariate Gaussian.
[c] Has a member within 1 pc of the search horizon, and therefore members are likely present beyond that limit.
[d] Name is newly assigned by this paper.

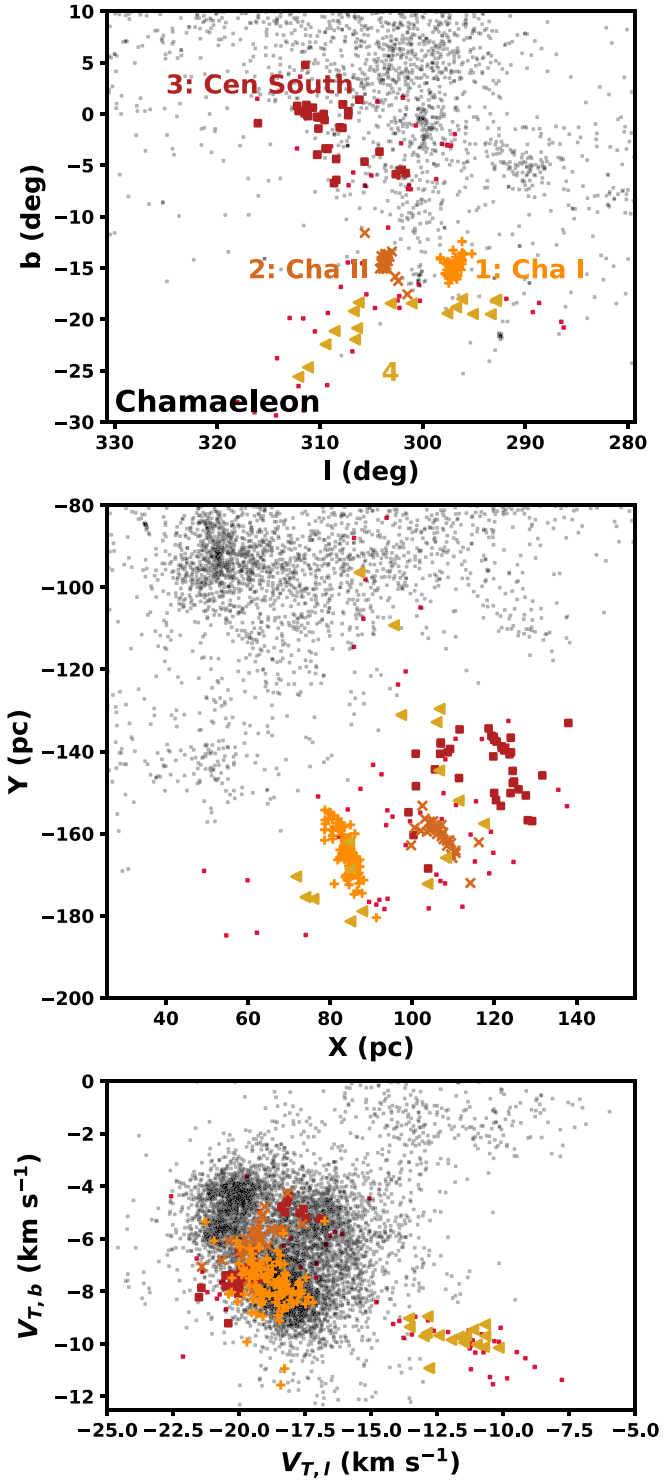

Figure 25. Clusters in the Chamaeleon complex, which hold similar kinematics to Sco-Cen. The panels, from top to bottom, show l/b sky coordinates, XY galactic coordinates, and the galactic transverse velocity. Subgroups are marked in the top panel; small red points represent nonclustered Chamaeleon members, and small black points indicate Sco-Cen members.

Taurus-Orion I (Theia 230), and Taurus-Orion II (Theia 116). Taurus-Orion IV also is partially represented in Kounkel & Covey (2019) as Theia 71, although the Theia group is limited to only the southern half of TO-IV in galactic coordinates. Lyra, Cepheus-Cygnus, and Cerberus (named after the obsolete constellation in eastern Hercules where it resides) are included as part of larger "string" groups, as outlined in Kounkel & Covey (2019), but are not identified as distinct groups, as we do in this work. Finally, the Fornax-Horologium group, which is located at a high southern galactic latitude, appears to be completely absent from the current literature.

Due to the large number of groups we identify, we do not perform a detailed overview of all of them, as we have for Sco-Cen, Orion, Perseus, and Taurus. Instead, we only perform a deep investigation of those that show clear signs of substructure, including Perseus OB3, MSW, Vela-CG4, CFN, and Carina-Musca, which are provided in the subsections below. Properties for each of the subclusters identified are listed in Table 7, while the remaining groups should be adequately described by the statistics in Table 3 and may receive more detailed follow-up in future publications.

4.5.1. Perseus OB3

Perseus OB3, also known as the α Persei cluster or Melotte 20, is a long-recognized nearby association identifiable by its wealth of O- and B-type members, which form an overdensity clearly visible with binoculars or even the naked eye under dark skies (e.g., Boss 1910; Eddington 1910). This region has been well studied over the past century, with the original $<$20 members identified in earlier works expanded to $\sim$300 by the first CCD photometry of the cluster (e.g., Prosser 1992) and later covered by wide-field astrometric studies (e.g., Deacon & Hambly 2004). Hipparcos measurements of Per OB3 members provided distance estimates of $\sim$180 pc (Mermilliod et al. 1997; de Zeeuw et al. 1999), a result that has changed relatively little in recent years (e.g., Lodieu et al. 2019). While most members lie within an $\sim$3° wide core in the plane of the sky, a considerably larger halo has been uncovered in previous works (e.g., Rasmuson 1921; de Zeeuw et al. 1999; Lodieu et al. 2019), which was tied to the much larger (and somewhat controversial; see Crawford 1963) Cas-Tau group by Blaauw (1956). The age estimates for the region have varied considerably, although most appear to fall between $\sim$50 and 80 Myr (e.g., Meynet et al. 1993; Basri & Martín 1999).

As expected, we do identify a significant halo of stars belonging to Perseus OB3, although our extended population contains much more structure than previously identified, with an additional subcluster $\sim$20° to the southeast of α Persei (de Zeeuw et al. 1999). The extensions that we find beyond the core of Perseus OB3 are shown in Figure 26, where both the α Persei cluster (POB3-1) and the new southeastern extension (POB3-2) are shown in spatial and proper-motion coordinates. Group POB3-2 is contiguous with the rest of Perseus OB3 (POB3-1/α Per) through low-density branches of stars connecting the two; however, it is physically closer to sections of Greater Taurus, located within 30 pc of the center of GT-9A in spatial coordinates compared to $\sim$70 pc for α Persei. The connection between POB3-2 and α Persei drawn by HDBSCAN is therefore most strongly rooted in kinematics, as POB3-2 is within only $\sim$5 km s^{-1} of α Persei in transverse velocity, compared to $\sim$10 km s^{-1} for the nearest Taurus groups. While the low-density bridge of young stars connecting the two components of this region supports their connection, the significant distance between the two components in spatial coordinates ($\sim$75 pc) casts doubt on whether this connection is real. Group POB3-2 has near-nonexistent coverage in radial velocity, so we currently lack the 3D velocities necessary to

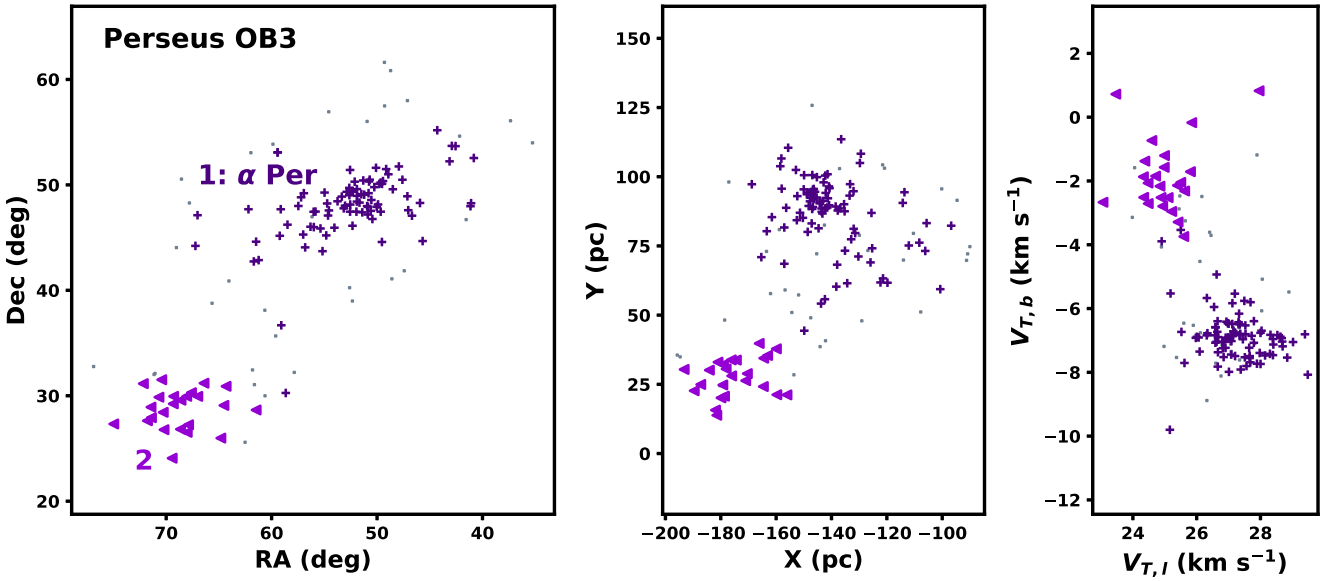

Figure 26. Subgroups in the Perseus OB3 association, plotted in R.A./decl., galactic X/Y, and galactic transverse velocity and labeled according to their POB3 ID in Table 7. Small gray dots mark unclustered Perseus OB3 members. Subgroup POB3-1 corresponds to the known α Persei cluster, while the other subgroup, POB3-2, is a newly discovered, younger extension.

confirm or reject the status of POB3-2 as a genuine subgroup of Perseus OB3.

The age we derive for POB3-2 is considerably younger than α Persei, ∼30 Myr compared to ∼50 Myr for the main cluster. Our ∼50 Myr age solution for α Persei is at the lower edge of the generally accepted age range of ∼50–80 Myr (e.g., Meynet et al. 1993; Basri & Martín 1999); however, large discrepancies would not be surprising in this case, as α Persei has a notably supersolar metallicity ([Fe/H] = +0.18) that may reduce the accuracy of our solar-metallicity model (Pöhnl & Paunzen 2010). Regardless, the fact that the age of POB3-2 is younger is robustly supported by our age solutions.

4.5.2. Monoceros Southwest

Most of the region that we refer to as MSW is a relatively recent discovery, with NGC 2232, a small young open cluster, being its only well-established component. Until recently, relatively little interest has been dedicated to the area around NGC 2232, and as such, the coverage of its parent environment has been quite limited. Claria (1972) performed the first major photometric survey and age estimate for the cluster, while Lyra et al. (2006) significantly expanded on and refined this work, introducing the first CCD survey of the cluster and computing a photometric age of ∼25–32 Myr. Suggestions of a broader population in the area first emerged through Zari et al. (2018), where a small 5°–7° wide clump of stars with a tight proper-motion distribution is identified out of the sample of photometrically young Gaia DR2 stars they investigate. This cluster was later characterized by Liu & Pang (2019), where it is assigned the name LP 2439. Pang et al. (2020) revisited the Monoceros region, identifying objects related to the two known clusters by applying the machine-learning algorithm StarGO (Yuan et al. 2018) to a population of candidates within a 100 pc radius of NGC 2232 and a 2.8 mas yr^{-1} proper-motion radius centered between LP 2439 and NGC 2232. The result is the identification of four subgroups: NGC 2232, LP 2439, and two new, smaller groups, which are referred to by their colors in the Pang et al. (2020) figures, green and purple. Pang et al. (2020) found all four of these groups to be approximately coeval, with an age of approximately 25 Myr, matching the earlier solutions from Lyra et al. (2006)

In our work, we identify a large, substructured group in MSW, which contains just under 300 stars, divided into four subgroups. The results of subclustering in MSW are shown in spatial and velocity coordinates in Figure 27. The most compact of the subgroups we identify, MSW-2, corresponds to NGC 2232, while the larger and more dispersed subgroup MSW-4 corresponds to LP 2439 (Liu & Pang 2019). The group we refer to as MSW-1 matches the green group in Pang et al. (2020), while the purple group, which is slightly more dispersed, shares some membership with our top-level MSW population between NGC 2322 and MSW-1 but is not assigned a subgroup. The methods employed by Pang et al. (2020) for the inclusion of potential members are much more permissive than those we use, only culling objects with inconsistent photometry, rather than our approach, which demands that photometry clearly indicate youth. This approach appears to work well for Pang et al. (2020) in MSW, as the region appears to be relatively uncontaminated and easy to separate from the nearby background, enabling somewhat more detailed subclustering capable of revealing the more tenuous purple group. However, the remaining subgroup that we identify, MSW-3, is not identified in Pang et al. (2020), likely due to it falling partially outside of the 100 pc search radius around NGC 2232 used in that paper. This result highlights the complementary strengths of our survey relative to more targeted studies. While tailoring investigations to an individual group can be very effective at expanding on known populations, assumptions made to isolate that group can result in broader, more dispersed populations being missed, which our more photometrically selective but spatially and kinematically indiscriminate survey excels at finding.

The overall velocity distribution in MSW is similar to that of Orion. All subclusters are within 4 km s^{-1} of each other in transverse velocity, and the velocity dispersions within groups

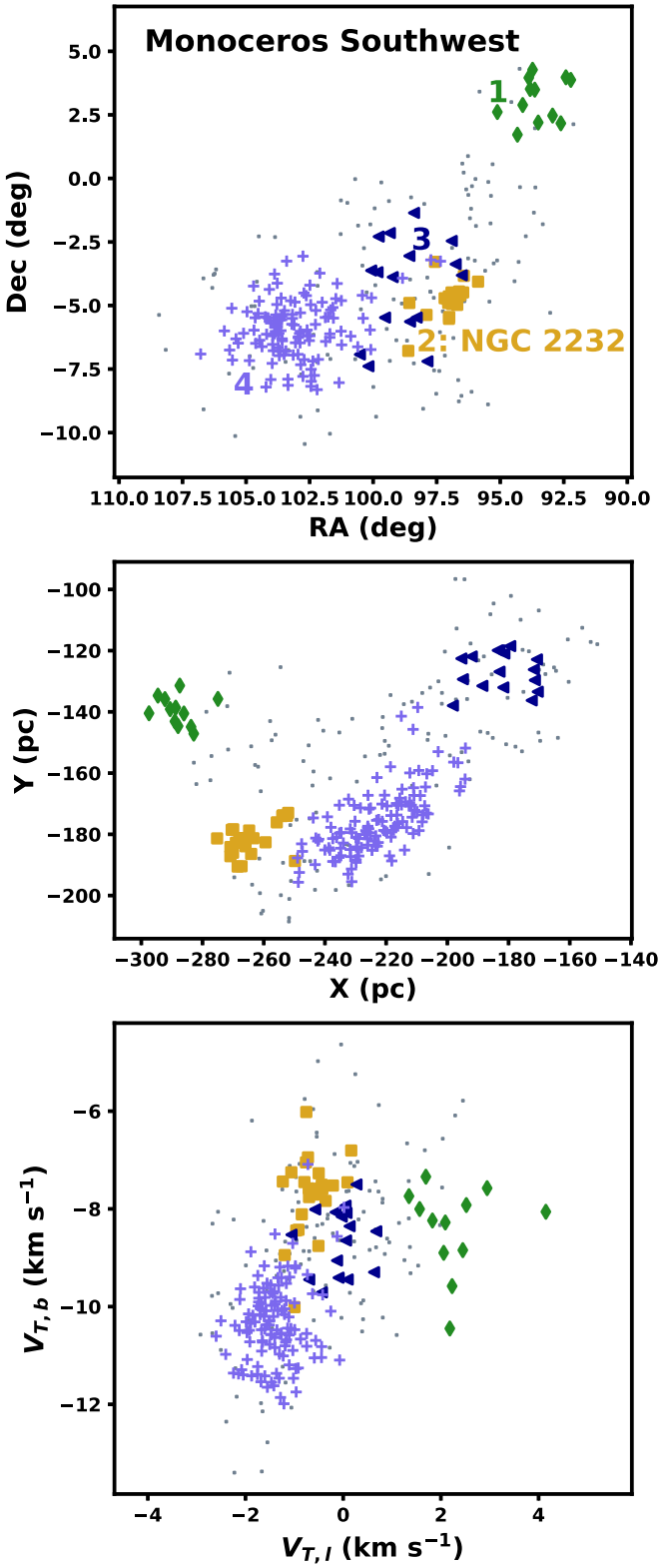

Figure 27. Subgroups in MSW, plotted in R.A./decl., galactic X/Y, and galactic transverse velocity. Each subgroup is labeled according to its MSW ID in Table 7, and small gray dots mark unclustered MSW members.

are consistently below 1 km s^{-1}. Our age fits for each subgroup are all between 24 and 28 Myr, which, combined with our bulk age estimate of 25.5 Myr, is in agreement with the 25 Myr coeval age solution suggested in the literature (e.g., Lyra et al. 2006; Pang et al. 2020).

4.5.3. Vela

The association we now refer to as Vela OB2 was first identified by Kapteyn (1914) and received its first major kinematic analysis from de Zeeuw et al. (1999), with the latter paper confirming both the common motions of the members and the inclusion of the open clusters NGC 2547 and γ Velorum. Later work revealed that the region's kinematic substructure is divided into two distinct populations: one connected to the γ Velorum cluster and another connected to NGC 2547 (Jeffries et al. 2014; Damiani et al. 2017). Two of the most recent major substructure analyses of Vela were performed in Cantat-Gaudin et al. (2019a, 2019b), who identified significantly more substructure. By using the UPMASK unsupervised classification method, Cantat-Gaudin et al. (2019a) identified seven kinematically distinct substructures in the Vela complex. Most of the larger Vela OB2 association they identify extends far beyond the 333 pc search radius used by our survey, limiting the subgroups we can cover, although sections of Cantat-Gaudin et al. (2019a) groups IV and VII (hereafter Vela-CG4 and Vela-CG7) are close enough to be included in this work. Vela-CG4 is the older of the two groups at ∼34 Myr and contains NGC 2547, while Vela-CG7 has an age closer to 14 Myr and contains the γ Velorum cluster.

Vela-CG7 is the smaller of the two Vela subgroups we identify, and our candidate members overlap with the Cantat-Gaudin et al. (2019a) region in proper motions and spatial coordinates. The group's limited size is not inherent, however, as its membership is thought to fall almost entirely outside of our 333 pc search radius. Cantat-Gaudin et al. (2019a) displayed a much more extensive population, which is centered on the γ Velorum cluster at a distance of ∼350 pc. Vela-CG7 does have known substructure outside of our search radius; however, the section we identify consists of only Vela-CG7 component A, as originally defined in Cantat-Gaudin et al. (2019b). As a result, we do not identify substructure, with the entire group having a tight velocity distribution of ∼1 km s^{-1}. Our ∼14 Myr age fit for Vela-CG7 also agrees with the age solution plotted in Cantat-Gaudin et al. (2019a) Figure 7 for the combined Vela-CG7 components A + B. Due to the lack of subclustering within our search radius, we do not include a separate plot for Vela-CG7, although it is visible through the top-level clustering analysis in Figure 10.

Unlike Vela-CG7, Vela-CG4 has considerable substructure within our search radius. The full extent of Vela-CG4 as defined in Cantat-Gaudin et al. (2019a) contains numerous known subclusters both inside and outside of our search radius, including NGC 2451B, NGC 2547, Collinder 140, Collinder 135, and UBC 7. Of these known subgroups, only the adjacent open clusters of Collinder 135 and UBC 7 (Castro-Ginard et al. 2018; Kovaleva et al. 2020) lie entirely within our search radius, together comprising what we call VCG4-4. In total, we identify five subgroups in Vela-CG4, including Collinder 135/UBC 7, which are shown in spatial and velocity coordinates in Figure 28. Like most subgroups identified in other regions, the velocity dispersions of the Vela-CG4 subclusters are all near 1 km s^{-1}, although stars in the region follow a linearly stretched transverse velocity distribution, spanning up to 15 km s^{-1} along the $v_{T,b}$ axis. The mean velocities for all identified subgroups, however, all remain within 5 km s^{-1} of each other, and most of this stretched velocity distribution can likely be attributed to geometric effects from the ∼800 deg^2 extent of the region.

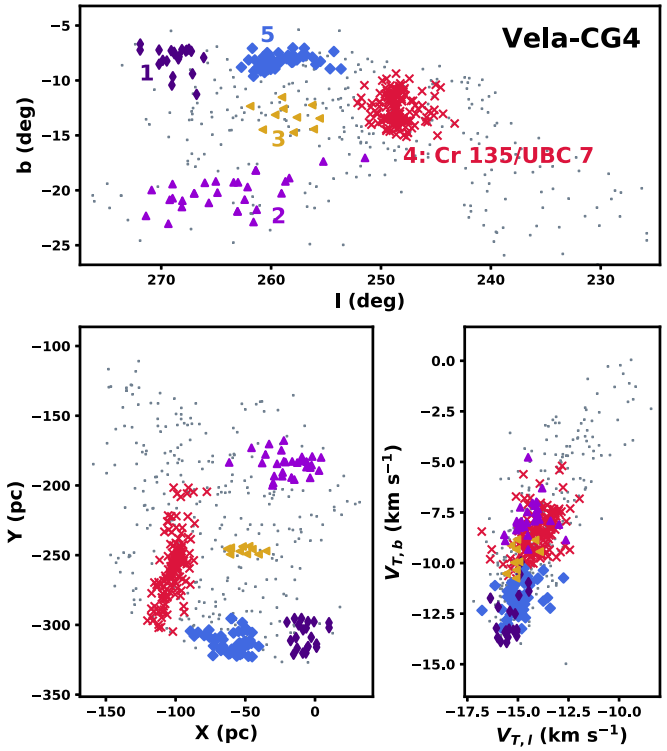

Figure 28. Subgroups in Vela-CG4 (Cantat-Gaudin et al. 2019a), shown in l/b galactic sky coordinates, galactic X/Y, and galactic transverse velocity. Each Vela-CG4 subcluster is labeled according to the VCG4 IDs from Table 7, and small gray dots mark unclustered Vela-CG4 members.

As the top-level HDBSCAN implementation defines it, Vela-CG4 contains a very broad distribution of stars, spanning over 150 pc in both the X and Y galactic coordinates, excluding membership beyond our search radius. Structures that extend far beyond the core Vela-CG4 regions have previously been proposed by Beccari et al. (2020); however, those that we identify are considerably closer and further to the galactic south. The two most distant groups (VCG4-1 and VCG4-5) are located closest to the dense, clustered regions in Vela-CG4 beyond our search radius, such as NGC 2547 (Cantat-Gaudin et al. 2019a), while VCG4-1 also matches spatially and kinematically with Sim et al. (2019) group UPK 535. These distant groups connect through lower-density stellar populations to Collinder 135 and UBC 7 (VCG4-4) and similar low-density populations to VCG4-2 and VCG4-3, overdensities located at ∼200 and ∼260 pc from the Sun, respectively. The ages we derive in Vela-CG4 are consistent with the groups being mostly coeval, with an age spread of just over 7 Myr and a bulk age of ∼34 Myr, similar to the 35 Myr estimate for the populations discussed in Beccari et al. (2020). Cantat-Gaudin et al. (2019a) provided age solutions for UBC 7 and Collinder 135 separately at ∼35 and 40 Myr, respectively, in agreement with our ∼35 Myr age solution.

The largest and closest of the new nearby extensions to Vela-CG4 that we find is VCG4-2, which extends in the direction of the Sco-Cen association and some of the other clusters in Carina, including IC 2391 and NGC 2451A. The individual groups and clusters located in this region host mutually discontiguous transverse velocities, typically separated by >10 km s^{-1}; however, they form a visually identifiable structure stretching between the IC 2602 branch of Sco-Cen (see Section 4.4.1) and Collinder 135/UBC 7 in Vela. The individual components in this structure also host relatively similar ages, spanning between ∼30 and 50 Myr. It is therefore possible that interactions between these individual star formation events may have occurred and contributed to their mutual evolution, without the gravitational influence between them being large enough for the resulting clusters to merge.

4.5.4. Cepheus Far North

A few different papers have previously noted a population of nearby young stars in northern Cepheus (e.g., see Tachihara et al. 2005; Klutsch et al. 2008; Oh et al. 2017; Frasca et al. 2018), although only Klutsch et al. (2020) defined a substantial population with which preliminary analyses have been performed. The extent of these populations as explored by Klutsch et al. (2020) is relatively limited, consisting primarily of a clump centered on $(X, Y) = (-75, 125)$ pc, with most of the 32 confirmed members within 20 pc of that location. They estimated an approximate age of 10–20 Myr for the region (Klutsch et al. 2020).

The population we identify in the region is significantly larger compared to what is presented in Klutsch et al. (2020), with the membership expanded to 219 members spanning more than 100 pc at its longest axis. The clump containing most of the Klutsch et al. (2020) members is included in this population, resting at its near edge, while we identify large numbers of stars extending far beyond this clump, with a mean distance of ∼180 pc. The group is primarily located within the bounds of northern Cepheus but also overflows significantly into Draco, reaching near the north ecliptic pole and overlapping with the TESS continuous viewing zone at its western edge (Ricker et al. 2015). To distinguish it from more distant and better-known features associated with Cepheus (e.g., see Kun et al. 2008), we refer to this region as CFN. The region's extent in spatial and velocity coordinates is displayed in Figure 29, along with a subclustering analysis. This group is divided by HDBSCAN into two subgroups: one for the main body of the region (CFN-1) and one that is more distant and much less populous (CFN-2). We derive age solutions for CFN-1 and CFN-2 of ∼24 and 26 Myr, respectively, suggesting that the star formation event that created CFN-2 may slightly predate star formation in the rest of the region, but the measurements are consistent with the populations being coeval. This age solution for CFN-1 is similar to the estimate of 10–20 Myr given by Klutsch et al. (2020), which is based on a subset of stars in the denser central region of CFN-1.

4.5.5. Carina-Musca

The final minor association with visible substructure that we identify is the Carina-Musca association. At 168 members, the population of identified members in this group is quite large, bigger than any of the individual open clusters, as well as Perseus OB3. It contains two subgroups: CM-1, which is relatively close to the clusters that bridge the gap between Vela and Sco-Cen, and CM-2, which is nearly 100 pc more distant, on average. The association as a whole has very limited recognition in the literature, with the only exception being the densest region of CM-2, which appears as UPK 569 in Sim et al. (2019). The kinematics of the region are very consistent, with a velocity dispersion of only ∼1 km s^{-1} for the entirety of Carina-Musca despite its large size and substructure. The ages, all at ∼33 Myr, enforce the internal consistency of properties

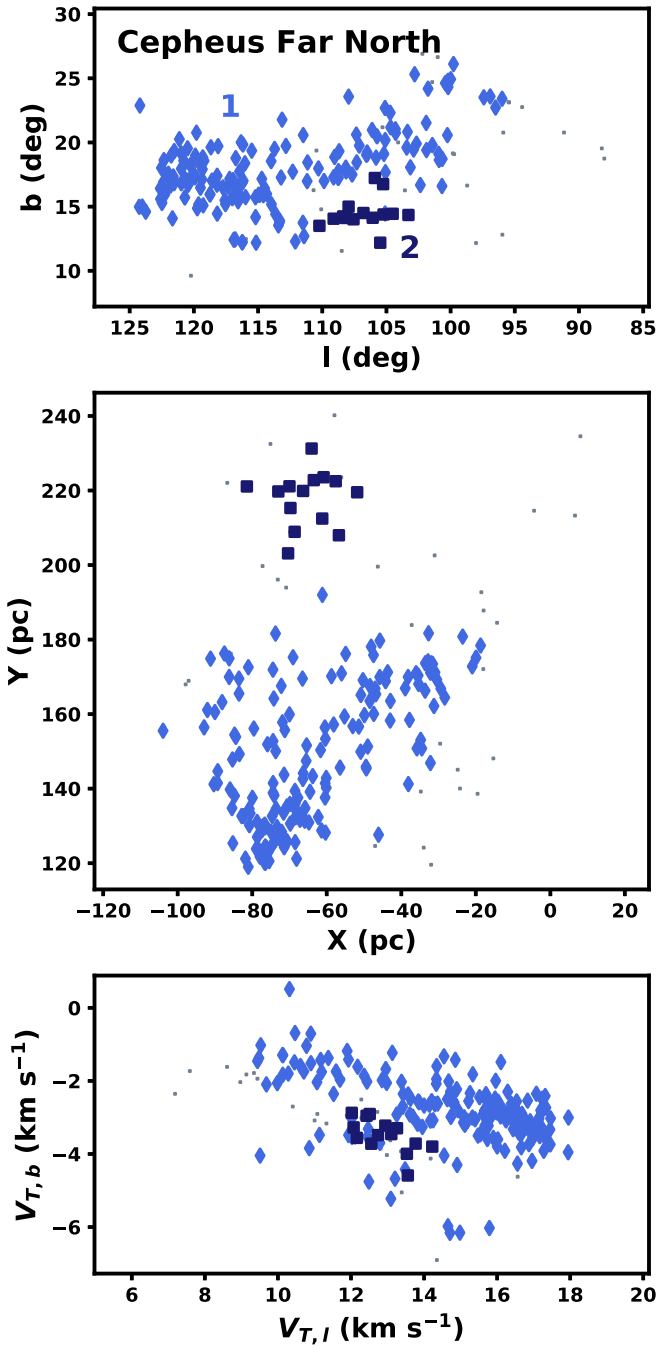

Figure 29. Group Cepheus Far North plotted in l/b galactic sky coordinates, galactic X/Y, and galactic transverse velocity. The two subgroups are labeled according to their CFN IDs from Table 7, and unclustered CFN members are marked with small gray dots. We identify 219 stars in the association, a significant expansion over the sample of 32 centered around $(X, Y) = (-75, 125)$ pc noted in Klutsch et al. (2020).

within Carina-Musca, with an incredibly tight pre–main sequence and less than 1 Myr separating the age fits for the two subgroups.

Carina-Musca also hosts what appear to be filamentary stellar structures crisscrossing the association, especially in CM-2, where the X/Y panel of Figure 30 shows two completely star-free voids centered around $(X, Y) = (84, -221)$ and $(104, -238)$ pc, completely surrounded by thin linear stellar overdensities. This apparent preservation of filamentary structure further enforces the very small internal velocity

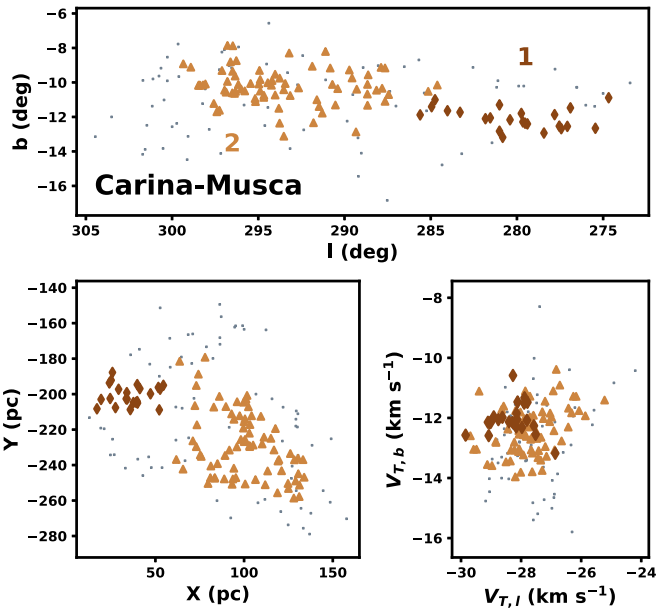

Figure 30. Subgroups in the Carina-Musca association, shown in l/b galactic sky coordinates, galactic X/Y, and galactic transverse velocity. Each Carina-Musca subcluster is labeled according to the CM IDs from Table 7, and unclustered Carina-Musca members are marked by small gray dots. Subgroup CM-1 is not identified in the literature, nor is the top-level structure enclosing the two subgroups.

dispersions, as structures with significant gravitational interactions would not be expected to survive for the ∼33 Myr age of the region. The Carina-Musca association may therefore serve as a useful case study for star formation in a large-scale turbulence-dominated region, a possibility explored in more depth in Section 5.1.

5. Local Patterns in Star Formation

5.1. Large-scale Simultaneous Star Formation

Multiple sites of recent star formation we uncover, including Vela-CG4, MSW, Carina-Musca, and, to a lesser extent, CFN, show remarkably consistent and effectively identical ages for their subgroups despite their immense size, all of which exceed 100 pc across along at least one axis. These groups are also notable for their very low densities, especially for Carina-Musca, which has 168 members over an area more than 100 pc wide, compared to the approximately coeval newborn groups in the TMC, which have an even larger membership despite covering only ∼40 pc from end to end. While some regions we discuss host denser subgroups, such as Collinder 135/UBC 7 in Vela-CG4, all such regions also host subgroups with densities at least comparable to those of the TMC populations. The dearth of gravitational influence inferred by these low densities is supported by the morphology of the regions we observe, all of which show evidence of filamentary structures in the form of extended stellar overdensities, including in MSW, where filamentary structure was independently identified in Pang et al. (2020). Small virialized clusters have been shown to disperse almost entirely over the >20 Myr lifetimes of these regions (Moeckel et al. 2012), so the fact that these tenuous stellar structures remain intact suggests a lack of internal gravitational interactions, effectively allowing the stars to sit in a near-stationary state relative to each other for tens of millions of years. In Carina-Musca, this near-stationary state is

reflected in a very tight, ∼1 km s^{-1} tangential velocity dispersion for the entire association, although the relative motions of the stars are difficult to ascertain without radial velocities.

Due to the apparent weakness of the gravitational forces in these regions, global gravitational collapse is likely a poor explanation for these structures (e.g., Vázquez-Semadeni et al. 2019; Krause et al. 2020). Instead, turbulence-driven collapse seems more likely. In this scenario, rather than collapse being driven directly from the self-gravity of the cloud, the cloud's internal turbulence and associated shocks are able to produce local density fluctuations capable of creating dense cores (Klessen et al. 2000; Mac Low & Klessen 2004; Krause et al. 2020). The cores generated by these turbulent processes tend to form along filaments and sheets, not unlike the stellar structures left behind in Carina-Musca and elsewhere (e.g., Mac Low & Klessen 2004). A turbulent environment is capable of not only driving the generation of overdensities and star-forming cores but also promoting these cores' collapse (e.g., Padoan & Nordlund 2002; Hennebelle & Chabrier 2008). Sustained turbulence in the surrounding environment applies a turbulent pressure to dense cores, driving collapse in locations with weak gravitational binding (McKee & Tan 2002; Field et al. 2011). Recent work on dense cores has shown that in the sparser environments present in Taurus, external cloud turbulence likely has a dominant influence on the collapse of dense cores, relative to more massive environments like Orion A, where self-gravitation and cloud weight pressure are more important (Kirk et al. 2017; Kerr et al. 2019). The sparse groups we identify here may therefore serve as a larger-scale analog to the star formation events visible in Taurus, which could make for useful comparisons with new modeling. It remains unclear whether simultaneous turbulent collapse on 100 pc scales is possible in isolation or whether a significant injection of turbulence, such as from a high-speed supernova shock, is required to produce the approximately coeval formation times we observe (e.g., Padoan et al. 2016). More detailed coverage, including radial velocities and improved ages, will be necessary to better constrain the dynamical histories of groups like Carina-Musca and subsequently assess their origins.

5.2. Evidence for Sequential Star Formation

Age gradients are left behind by multiple different star formation mechanisms, and the patterns can provide important hints as to the star formation history of a region (e.g., Elmegreen & Lada 1977; Krause et al. 2020). While most of our regions have either tenuous or nonexistent age gradients, Sco-Cen is an exception, most notably hosting gradients from older ages along the LCA to progressively younger ages farther away. This gradient appears to apply across all subgroups in and around LCC, as well as all LCA-adjacent groups west of Lupus. We present the age differences between these young populations and the nearest LCA subgroup as a function of distance in Figure 31, and we find a clear linear trend between age and distance. We apply a least-squares fit to these results, only allowing the slope to vary, as the intercept must be at zero to allow the LCA to have ages and distances of zero relative to itself. The result upon inverting that slope is a star formation propagation speed of 4.12 ± 0.19 km s^{-1} (4.12 ± 0.19 km $s^{-1} = 4.21 \pm 0.19$ pc Myr^{-1}) with a reduced χ^2 of 0.55. To our knowledge, this result is the first large-scale empirical measurement of a star formation propagation speed, providing a useful comparison for theoretical models.

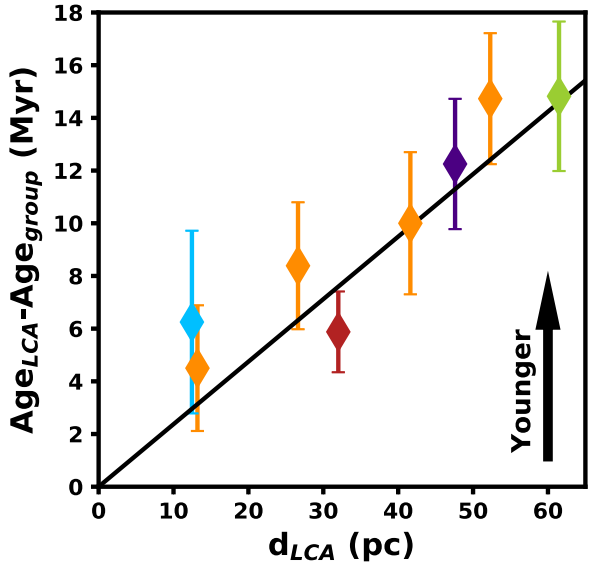

Figure 31. Difference in age relative to the LCA, the oldest arc of subclusters in Sco-Cen (excluding the IC 2602 branch), as a function of distance to the nearest LCA component group. Color schemes match those in Figures 18 and 22, with LCC subgroups in orange, TW Hydrae in indigo, SC-19 in blue, η Cha in green, and SC-11 in red. Younger clusters are consistently located further from the LCA, suggesting a gradual progression of star formation outward from the LCA.

Shocks from supernovae or feedback from energetic young clusters are commonly suggested as triggers of star formation, creating self-gravitating overdensities as they pass through a molecular cloud, forming stars (e.g., Elmegreen et al. 2002; Dale et al. 2007). In many known regions, the most active sites of star formation appear on the periphery of a shell carved out by radiation from the most recent generation of star formation (e.g., Phelps & Lada 1997; Lefloch & Cernicharo 2000; Lee & Chen 2007; Zavagno et al. 2010; Elmegreen et al. 2000). While our stellar populations do not show clear radial trends originating in any one location, we do find a coeval semicircular population in the form of the LCA, possibly consistent with triggered formation from an supernova shell compressing a large gas cloud. However, the populations adjacent to the LCA in Sco-Cen increase in age as a function of distance from the nearest component of the LCA, a result inconsistent with spherical shock propagation or any other form of shock, such as those generated through the collision of molecular clouds (Elmegreen 2011). This close trend between the age and position relative to the LCA suggests that rather than a single external shock creating a propagating overdensity, the triggers for these later generations may originate directly in those earlier generations of stars. Substantial populations of B stars are known to exist in the LCA and LCC, and O stars likely existed before ending their shorter main-sequence lifetimes as supernovae, so they serve as a likely driver for these observed patterns (e.g., de Zeeuw et al. 1999; Rizzuto et al. 2011).

The propagation of star formation prompted by the influence of earlier generations is referred to as sequential star formation, first suggested in Blaauw (1964) and described in detail by Elmegreen & Lada (1977). In this process, radiation pressure from O and B stars powers an ionization shock front, which can generate overdensities capable of supporting star formation. In a large star-forming region, sequential star formation can

continue for millions of years, forming new generations of stars until the cloud is exhausted (Elmegreen & Lada 1977). The precise speed at which sequential star formation is expected to advance has yet to be predicted in the literature, especially on the large scales we observe, due to the lack of simulations covering the range in space and time covered by Sco-Cen. Observational studies of these speeds are also largely absent on the scales we considered, with the existing work only considering gradients on scales of a few parsecs or megayears (e.g., Getman et al. 2018; Lim et al. 2018). While we are currently unable to make comparisons, the star formation progression speed may provide a useful probe of gas properties in the parent cloud, as properties such as magnetic fields and internal turbulence are expected to increase the cloud's resistance to collapse and potentially slow the progression of any sequential star formation event (e.g., Ballesteros-Paredes et al. 2020). Through these inferences of cloud properties, the importance of additional processes that regulate them can also be assessed, such as turbulence excited by stellar feedback, which can inject energy into the cloud ahead of the current generation of star formation (Offner & Liu 2018).

Regardless of the inferences that can be drawn about gas properties and dynamics from the rate of star formation progression, the configuration of star formation near the LCA in Sco-Cen is consistent with sequential star formation and represents the strongest indication to date that this process occurs on timescales of ∼20 Myr. Other features in Sco-Cen also show gradients potentially consistent with these mechanisms, such as those in Lower Sco and toward CrA. However, more complete and precise age mapping is necessary to compare them to our result in the LCA. Expanding the sample of sites with evidence of sequential star formation will be important to ascertain whether the mechanisms that regulate propagation speed are consistent across clouds or whether fundamental changes to the cloud properties can slow or accelerate sequential propagation.

5.3. Repeated Star Formation Bursts

Multiple groups we identify show star-forming events with clear age separation but essentially identical kinematics and similar spatial extents. The clearest example of this is in Perseus, where both subgroups do not have HDBSCAN-defined substructure but do have very clear separation in age, with essentially newborn ages around the Perseus Molecular Cloud in the west and more dispersed stars with ages around 17 Myr in the east. Group GT-9 has subregions separated enough to be identified as independent leaf structures, but there we identify the same age structure: one essentially newborn subgroup in GT-9A and a second, ∼12 Myr old subgroup in GT-9B. Both have no visible gradient in age, ruling out a continuous sequential star formation event. A few other sites show similar patterns, such as Perseus OB3 and a few other subgroups in Taurus, but the larger mutual separations between subgroups in spatial and velocity coordinates make the connections between these different age subgroups weaker relative to those in GT-9 and Perseus.

After star formation takes place in a dense cloud, it is generally expected that the remaining gas present around the young stars will be dispersed by stellar feedback from newly formed stars (e.g., Walch et al. 2012; Dobbs et al. 2014). If the initial collapse only affects part of the cloud, a sequential star-forming event may begin to propagate across the cloud as newly formed stars collect gas at higher densities in a shock that moves across the region, as described in Section 5.2 (Elmegreen & Lada 1977). Otherwise, these feedback mechanisms are expected to disrupt future star formation (Walch et al. 2012; Grudić et al. 2021b). In Perseus (and, to a lesser extent, GT-9), however, we observe time-separated bursts of star formation with no age gradient between them, as well as the near-identical kinematics and a close spatial connection. The internal consistency of these groups makes it unlikely that these star formation bursts originated in fully distinct environments, suggesting that processes from within the same environment were able to force a period of dormancy between the star formation events.

Gas dispersal by stellar feedback is a likely cause of the end of star formation in the initial star-forming events (e.g., Walch et al. 2012). However, each of the groups we discuss here appear to have been able to reaccumulate enough gas afterward to support a new burst of star formation. One possible explanation for these secondary star formation bursts is that gas collection by flows that fed the initial star-forming event continued after the dispersal of the gas at the original collection point. In this scenario, star formation at the first star-forming site ended once stellar feedback dispersed the gas there, but new gas continued to stream in along the original flows. These streams created new collection points capable of supporting a second burst of star formation, like those in Perseus and GT-9. This process may be comparable to the generation of molecular clouds through the collision of convergent flows discussed in Dobbs et al. (2014), with gas from stellar feedback recycling into the next episode of star formation. A similar mechanism was proposed in Kroupa et al. (2018) to explain discrete star-forming events in Orion; however, the dynamical ejection of high-mass stars that they suggest to permit the continuation of star formation after disruption is unlikely to have significant influence in the dynamically cooler Perseus region. The spatial and temporal scales spanned by the stars in our study have not, however, been well sampled in simulations due to limitations in computing power and dynamic range. Our possible explanation of cospatial episodic bursts therefore requires new simulations to explore feasibility and develop a rigorous theoretical model.

5.4. Defining Top-level Regions

In our HDBSCAN implementation, the subgroups that we identify are included as part of the same larger region when they are connected to other subgroups in spatial and velocity coordinates through stellar overdensities. While the presence of apparent physical and kinematic connections is a commonly used and logical standard in clustering (e.g., Kounkel & Covey 2019; Zari et al. 2019), it does not exclude the possibility of false connections emerging. For example, close interactions between two unrelated groups after formation may leave a trail of stars linking them, which our HDBSCAN implementation could group together at the top level. Alternatively, unrelated groups could be made contiguous through the ejection of members from a cluster's core, extending its envelope.

Falsely connected associations would be physically contiguous but have centers that are widely separated in spatial and velocity coordinates. Since we do not include age or traceback in our clustering, we are unable to investigate whether subclusters with wider spacing in spatial and velocity coordinates and therefore apparently weaker mutual connections were likely closer at the time of formation. A few regions we identify have subgroups with these seemingly weak connections to one another. These include Perseus OB3, which

has wide separations between subgroups in space, velocity, and age; Sco-Cen, which has a significant velocity and age division between its main body and the IC 2602 branch; and Taurus, which has many subgroups with significant separations from one another by multiple metrics. All of these sets of subgroups are contiguous with one another through lower-density stellar populations; however, their mutual separations in spatial and velocity coordinates, as well as significant age gaps, make their physical association with one another more uncertain.

Ruling out false connections between subgroups will require the addition of radial velocity measurements to the transverse velocity vector we use for clustering. Transverse velocities are plane-of-sky measurements, and they are therefore vulnerable to projection effects that can hide genuine velocity differences or induce spurious ones. The current implementation suffers greatly from these projection effects, contributing to the velocity scatter in Taurus and elsewhere and preventing our identification of any known groups within ∼50 pc of the Sun. The addition of radial velocity measurements would eliminate all projection effects while also completing the velocity vector, allowing 6D space-velocity clustering, and providing a robust assessment of kinematic similarity between groups, regardless of their distance from one another in the plane of the sky.

Having this complete 3D velocity vector also enables long-term traceback of the stellar motions (e.g., Kraus et al. 2017). Recent developments have been made in improving traceback in young stellar populations (e.g., Crundall et al. 2019), so after the completion of follow-up radial velocity observations, the complete history of these populations can be established, and possibilities for a common origin can be better assessed. Age determinations (such as from the luminosity of the lithium depletion boundary) will also be necessary to exclude the possibility of reddening anomalies causing subgroups to be falsely marked as young while also helping to establish their location at the time of formation in relation to other subgroups. This traceback may also help to establish groups that have differing velocities but may have had interactions with other groups near the time of formation, perhaps as part of opposing colliding gas flows. Therefore, while the top-level regions we identify in this work should not be treated as entirely homogeneous in formation origin, they do connect subgroups where a common origin is likely and mutual association can be robustly evaluated as more accurate ages and complete 3D velocity information becomes available.

6. Conclusion

We have photometrically identified $>3 \times 10^4$ probable young stars within a distance of 333 pc and assessed the distinct groups and clusters among these stars using the HDBSCAN density-based hierarchical clustering algorithm. Through this method, we have significantly expanded the census of known young stars in the solar neighborhood, identifying new structures and revealing new star formation patterns that trace the progression of star formation, particularly in some of the large and sparse groups that are found distributed throughout the solar neighborhood. The key conclusions and discoveries reached through this work are as follows.

1. We identify 27 robust young groups and associations within 333 pc of the Sun at the top level. Of these, 15 have a significant presence in the literature, eight have only been recognized as part of the large catalogs from Kounkel & Covey (2019) and Sim et al. (2019; Carina-Musca, Ophiuchus Southeast, Canis Major North, Aquila East, and Taurus-Orion I, II, III, and IV), and the remaining four (Cerberus, Lyra, Cepheus-Cygnus, and Fornax-Horologium) are identified as distinct groups for the first time.
2. Ten of these groups show significant substructure, which is manifested in the form of HDBSCAN-defined EOM and leaf-level subgroups within each of these larger structures. We provide complete overviews of the structure in each of these cases: Sco-Cen, Greater Taurus, Greater Orion, Perseus, Chamaeleon, Vela-CG4, Perseus OB3, MSW, CFN, and Carina-Musca.
3. Nearly all large known groups contain substructures that we newly identify, including Eridanus North in Greater Orion; Sco-Cen subgroups SC-7, 10, and 16, which connect the region's known extent to CrA; and Centaurus South, a somewhat older clustered extension to the Chamaeleon complex.
4. We identify a complex network of subgroups in Taurus, linking the current sites of active star formation directly to stars considerably older than 10 Myr.
5. The two Perseus subgroups, as well as GT-9 in Taurus, contain an older and a younger stellar population, with near-identical kinematics and close spatial connections between these separate generations. Stellar feedback from the earlier star formation burst may have prompted a period of dormancy in the region during which gas continued to collect, eventually permitting a second burst of star formation.
6. We have identified the LCA, an old semicircular structure along the northern edge of Sco-Cen that appears to represent the first star formation in the region and a possible trigger for later star formation nearby. We also observe a smooth age gradient toward nearby subgroups, which appears to indicate the presence of sequential star formation with a propagation speed of 4.12 ± 0.19 km s^{-1} (4.21 ± 0.19 pc Myr^{-1}).
7. The presence of large, sparse, and approximately coeval populations in MSW, Vela-CG4, CFN, and Carina-Musca may indicate mechanisms for large-scale ($\gtrsim$100 pc) simultaneous star formation in turbulence-dominated environments.

The authors thank the anonymous referee, whose comments greatly improved this publication. A.C.R. was supported as a 51 Pegasi b Fellow though the Heising-Simons Foundation. S.S.R.O. acknowledges support from the National Science Foundation (NSF) Career Grant 1748571 and NSF AAG 1812747. R.M.P.K. was supported in part by NASA grant NSSC18K0405 through the ADAP program. R.M.P.K. acknowledges the use of computational resources at the Texas Advanced Computing Center (TACC) at the University of Texas at Austin, which was used for the more computationally intensive operations in this project.

Facility: Gaia.

Software: Astropy (Astropy Collaboration et al. 2013), HDBSCAN (McInnes et al. 2017), Matplotlib (Hunter 2007), SciPy (Virtanen et al. 2020).

Appendix
Group Isochrone Fits

Here we present the CMDs and corresponding age fits for each group and subgroup we identify in this work, as described

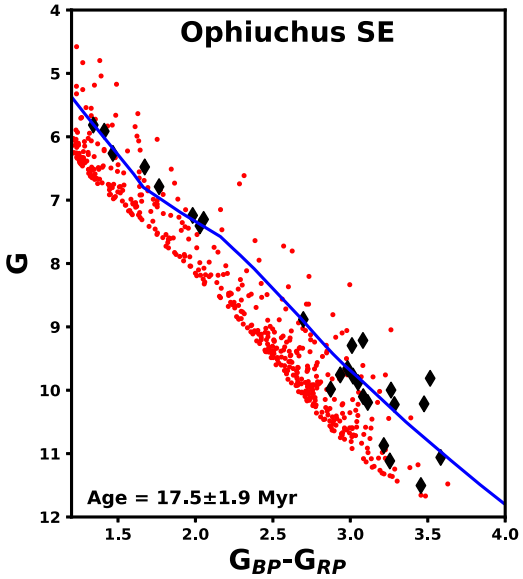

Figure A1. Isochrone fits for each group, with the best-fit isochrone represented by the blue curve. The stars used in the fit are marked with black diamonds, and those that are removed due to RUWE and weight cuts are marked with red dots. The group name and age are both labeled in the figure. The remaining fits for each group are available in the online version of this paper.

(The complete figure set (108 images) is available.)

in Section 3.5. One example is given in Figure A1, and the remaining fits are available in the online version of this paper.

ORCID iDs

Ronan M. P. Kerr https://orcid.org/0000-0002-6549-9792
Aaron C. Rizzuto https://orcid.org/0000-0001-9982-1332
Adam L. Kraus https://orcid.org/0000-0001-9811-568X
Stella S. R. Offner https://orcid.org/0000-0003-1252-9916

References

Abt, H. A., Gomez, A. E., & Levy, S. G. 1990, ApJS, 74, 551
Anderson, L. D., Wenger, T. V., Armentrout, W. P., Balser, D. S., & Bania, T. M. 2019, ApJ, 871, 145
Arenou, F., Luri, X., Babusiaux, C., et al. 2018, A&A, 616, A17
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Aumer, M., & Binney, J. J. 2009, MNRAS, 397, 1286
Azimlu, M., Martínez-Galarza, J. R., & Muench, A. A. 2015, AJ, 150, 95
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58
Ballesteros-Paredes, J., André, P., Hennebelle, P., et al. 2020, SSRv, 216, 76
Bally, J. 2008, in Overview of the Orion Complex, ed. B. Reipurth, Vol. 4 (San Francisco, CA: ASP), 459
Bally, J., Walawender, J., Johnstone, D., Kirk, H., & Goodman, A. 2008, The Perseus Cloud, Vol. 4 (San Francisco, CA: ASP), 308
Basri, G., & Martín, E. L. 1999, ApJ, 510, 266
Beccari, G., Boffin, H. M. J., & Jerabkova, T. 2020, MNRAS, 491, 2205
Bell, C. P. M., Mamajek, E. E., & Naylor, T. 2015, MNRAS, 454, 593
Bell, C. P. M., Murphy, S. J., & Mamajek, E. E. 2017, MNRAS, 468, 1198
Bergfors, C., Brandner, W., Janson, M., et al. 2010, A&A, 520, A54
Binney, J., Dehnen, W., & Bertelli, G. 2000, MNRAS, 318, 658
Blaauw, A. 1952, Bull. Astron. Inst.Netherlands, 11, 405
Blaauw, A. 1956, ApJ, 123, 408
Blaauw, A. 1964, ARA&A, 2, 213
Boss, L. 1910, Preliminary General Catalogue of 6188 Stars for the Eopoch 1900 (Washington, DC: Carnegie Institution)
Bouy, H., Bertin, E., Barrado, D., et al. 2015, A&A, 575, A120
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Briceño, C., Calvet, N., Hernández, J., et al. 2019, AJ, 157, 85
Briceno, C., Hartmann, L. W., Stauffer, J. R., et al. 1997, AJ, 113, 740
Brown, A. G. A., de Geus, E. J., & de Zeeuw, P. T. 1994, A&A, 289, 101
Brown, A. G. A., Dekker, G., & de Zeeuw, P. T. 1997, MNRAS, 285, 479
Bryson, S., Coughlin, J., Batalha, N. M., et al. 2020, AJ, 159, 279
Buder, S., Asplund, M., Duong, L., et al. 2018, MNRAS, 478, 4513
Cantat-Gaudin, T., Jordi, C., Wright, N. J., et al. 2019a, A&A, 626, A17
Cantat-Gaudin, T., Mapelli, M., Balaguer-Núñez, L., et al. 2019b, A&A, 621, A115
Castro-Ginard, A., Jordi, C., Luri, X., et al. 2018, A&A, 618, A59
Cernis, K. 1993, BaltA, 2, 214
Chabrier, G. 2003, PASP, 115, 763
Chabrier, G. 2005, in The Initial Mass Function 50 Years Later, ed. E. Corbelli et al. (Berlin: Springer), 41
Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068
Chini, R., Hoffmeister, V. H., Nasseri, A., Stahl, O., & Zinnecker, H. 2012, MNRAS, 424, 1925
Claria, J. J. 1972, A&A, 19, 303
Comerón, F. 2008, in The Lupus Clouds, ed. B. Reipurth, Vol. 5 (San Francisco, CA: ASP), 295
Crawford, D. L. 1963, ApJ, 137, 523
Crundall, T. D., Ireland, M. J., Krumholz, M. R., et al. 2019, MNRAS, 489, 3625
Cummings, J. D., Deliyannis, C. P., Maderak, R. M., & Steinhauer, A. 2017, AJ, 153, 128
Dale, J. E., Clark, P. C., & Bonnell, I. A. 2007, MNRAS, 377, 535
Damiani, F., Prisinzano, L., Jeffries, R. D., et al. 2017, A&A, 602, L1
Damiani, F., Prisinzano, L., Pillitteri, I., Micela, G., & Sciortino, S. 2019, A&A, 623, A112
de Avillez, M. A., & Mac Low, M.-M. 2002, ApJ, 581, 1047
de Zeeuw, P. T., Hoogerwerf, R., de Bruijne, J. H. J., Brown, A. G. A., & Blaauw, A. 1999, AJ, 117, 354
Deacon, N. R., & Hambly, N. C. 2004, A&A, 416, 125
Delfosse, X., Beuzit, J. L., Marchal, L., et al. 2004, in ASP Conf. Ser. 318, Spectroscopically and Spatially Resolving the Components of the Close Binary Stars, ed. R. W. Hilditch et al. (San Francisco, CA: ASP), 166
Dieterich, S. B., Henry, T. J., Golimowski, D. A., Krist, J. E., & Tanner, A. M. 2012, AJ, 144, 64
Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 3
D'Orazi, V., Oliva, E., Bragaglia, A., et al. 2020, A&A, 633, A38
Duchêne, G., & Kraus, A. 2013, ARA&A, 51, 269
Ducourant, C., Teixeira, R., Galli, P. A. B., et al. 2014, A&A, 563, A121
Eddington, A. S. 1910, MNRAS, 71, 43
Elmegreen, B. G. 2011, EAS, 51, 45
Elmegreen, B. G., Efremov, Y., Pudritz, R. E., & Zinnecker, H. 2000, in Protostars and Planets IV, ed. V. Mannings et al. (Tucson, AZ: Univ. Arizona Press), 179
Elmegreen, B. G., & Lada, C. J. 1977, ApJ, 214, 725
Elmegreen, B. G., Palouš, J., & Ehlerová, S. 2002, MNRAS, 334, 693
Esplin, T. L., & Luhman, K. L. 2019, AJ, 158, 54
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Proc. 2nd Int. Conf. on Knowledge Discovery and Mining, KDD-96, ed. E. Simoudis (Palo Alto, CA: AAAI Press), 226
Evans, D. W., Riello, M., De Angeli, F., et al. 2018, A&A, 616, A4
Field, G. B., Blackman, E. G., & Keto, E. R. 2011, MNRAS, 416, 710
Frasca, A., Guillout, P., Klutsch, A., et al. 2018, A&A, 612, A96
Friel, E. D., & Janes, K. A. 1993, A&A, 267, 75
Gagné, J., David, T. J., Mamajek, E. E., et al. 2020, ApJ, 903, 96
Gagné, J., Faherty, J. K., Mamajek, E. E., et al. 2017, ApJS, 228, 18
Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, A&A, 616, A10
Gaia Collaboration, Brown, A. G. A., & Vallenari, A. 2018b, A&A, 616, A1
Galli, P. A. B., Bertout, C., Teixeira, R., & Ducourant, C. 2013, A&A, 558, A77
Galli, P. A. B., Bouy, H., Olivares, J., et al. 2020, A&A, 634, A98
Galli, P. A. B., Loinard, L., Bouy, H., et al. 2019, A&A, 630, A137
Galli, P. A. B., Loinard, L., Ortiz-Léon, G. N., et al. 2018, ApJ, 859, 33
Getman, K. V., Feigelson, E. D., Kuhn, M. A., et al. 2018, MNRAS, 476, 1213
Goldman, B., Röser, S., Schilbach, E., Moór, A. C., & Henning, T. 2018, ApJ, 868, 32
Grudić, M. Y., Guszejnov, D., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2021a, MNRAS, 506, 2199
Grudić, M. Y., Kruijssen, J. M. D., Faucher-Giguère, C.-A., et al. 2021b, MNRAS, 506, 3239

Gutierrez-Moreno, A., & Moreno, H. 1968, ApJS, 15, 459
Ha, T., Li, Y., Xu, S., Kounkel, M., & Li, H. 2021, ApJL, 907, L40
Hatchell, J., Richer, J. S., Fuller, G. A., et al. 2005, A&A, 440, 151
Hayashi, C. 1961, PASJ, 13, 450
Hayden, M. R., Bland-Hawthorn, J., Sharma, S., et al. 2019, MNRAS, 493, 2952
Haywood, M., Snaith, O., Lehnert, M. D., Di Matteo, P., & Khoperskov, S. 2019, A&A, 625, A105
Hennebelle, P., & Chabrier, G. 2008, ApJ, 684, 395
Hillenbrand, L. A., & Carpenter, J. M. 2000, ApJ, 540, 236
Huber, D., Bryson, S. T., Haas, M. R., et al. 2016, ApJS, 224, 2
Hunter, J. D. 2007, CSE, 9, 90
Isern, J. 2019, ApJL, 878, L11
Jeffries, R. D., Jackson, R. J., Cottaar, M., et al. 2014, A&A, 563, A94
Kapteyn, J. C. 1914, ApJ, 40, 43
Kastner, J. H., Zuckerman, B., Weintraub, D. A., & Forveille, T. 1997, Sci, 277, 67
Kenyon, S. J., Gómez, M., & Whitney, B. A. 2008, in Low Mass Star Formation in the Taurus-Auriga Clouds, ed. B. Reipurth, Vol. 4 (San Francisco, CA: ASP), 405
Kerr, R., Kirk, H., Di Francesco, J., et al. 2019, ApJ, 874, 147
Kirk, H., Friesen, R. K., Pineda, J. E., et al. 2017, ApJ, 846, 144
Kirk, H., Myers, P. C., Bourke, T. L., et al. 2013, ApJ, 766, 115
Klein Wassink, W. J. 1927, PGro, 41, 1
Klessen, R. S., Heitsch, F., & Mac Low, M.-M. 2000, ApJ, 535, 887
Klutsch, A., Frasca, A., Guillout, P., et al. 2008, A&A, 490, 737
Klutsch, A., Frasca, A., Guillout, P., et al. 2020, A&A, 637, A43
Kos, J., Bland-Hawthorn, J., Asplund, M., et al. 2019, A&A, 631, A166
Kounkel, M., & Covey, K. 2019, AJ, 158, 122
Kounkel, M., Covey, K., Suárez, G., et al. 2018, AJ, 156, 84
Kouwenhoven, M. B. N., Brown, A. G. A., Portegies Zwart, S. F., & Kaper, L. 2007, A&A, 474, 77
Kouwenhoven, T., Brown, A., Zinnecker, H., et al. 2005, A&A, 430, 137
Kovaleva, D., Ishchenko, M., Postnikova, E., et al. 2020, A&A, 642, L4
Kraus, A. L., Herczeg, G. J., Rizzuto, A. C., et al. 2017, ApJ, 838, 150
Kraus, A. L., & Hillenbrand, L. A. 2009, ApJ, 704, 531
Kraus, A. L., Ireland, M. J., Martinache, F., & Hillenbrand, L. A. 2011, ApJ, 731, 8
Kraus, A. L., Shkolnik, E. L., Allers, K. N., & Liu, M. C. 2014, AJ, 147, 146
Krause, M. G. H., Burkert, A., Diehl, R., et al. 2018, A&A, 619, A120
Krause, M. G. H., Offner, S. S. R., Charbonnel, C., et al. 2020, SSRv, 216, 64
Krolikowski, D. 2021, AJ, in press, (arXiv:2105.13370)
Kroupa, P., Jeřábková, T., Dinnbier, F., Beccari, G., & Yan, Z. 2018, A&A, 612, A74
Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, ARA&A, 57, 227
Kun, M., Kiss, Z. T., & Balog, Z. 2008, in Star Forming Regions in Cepheus, ed. B. Reipurth, Vol. 4 (San Francisco, CA: ASP), 136
Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57
Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, A&A, 625, A135
Larson, R. B. 1979, MNRAS, 186, 479
Larson, R. B. 1981, MNRAS, 194, 809
Lee, H.-T., & Chen, W. P. 2007, ApJ, 657, 884
Lee, H.-T., Chen, W. P., Zhang, Z.-W., & Hu, J.-Y. 2005, ApJ, 624, 808
Lefloch, B., & Cernicharo, J. 2000, ApJ, 545, 340
Lim, B., Sung, H., Bessell, M. S., et al. 2018, MNRAS, 477, 1993
Lin, J., Asplund, M., Ting, Y.-S., et al. 2020, MNRAS, 491, 2043
Lindegren, L. 2018, Gaia Data Processing and Analysis Consortium, GAIA-C3-TN-LU-LL-124-01, http://www.rssd.esa.int/doc_fetch.php?id=3757412
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
Liu, J., Fang, M., Tian, H., et al. 2021, ApJS, 254, 20
Liu, L., & Pang, X. 2019, ApJS, 245, 32
Lodieu, N., Pérez-Garrido, A., Smart, R. L., & Silvotti, R. 2019, A&A, 628, A66
Lopez Martí, B., Jimenez Esteban, F., Bayo, A., et al. 2013, A&A, 551, A46
Luhman, K. L. 2008, in Chamaeleon, ed. B. Reipurth, Vol. 5 (San Francisco, CA: ASP), 169
Luhman, K. L. 2018, AJ, 156, 271
Luhman, K. L., & Esplin, T. L. 2020, AJ, 160, 44
Luhman, K. L., Esplin, T. L., & Loutrel, N. P. 2016, ApJ, 827, 52
Luhman, K. L., Mamajek, E. E., Allen, P. R., & Cruz, K. L. 2009, ApJ, 703, 399
Lyra, W., Moitinho, A., van der Bliek, N. S., & Alves, J. 2006, A&A, 453, 101
Maaskant, K. M., Bik, A., Waters, L. B. F. M., et al. 2011, A&A, 531, A27
Mac Low, M.-M., & Klessen, R. S. 2004, RvMP, 76, 125
Magazzu, A., Martin, E. L., Sterzik, M. F., et al. 1997, A&AS, 124, 449
Mamajek, E. 2016, doi:10.6084/m9.figshare.3122689.v1, https://figshare.com/articles/figure/A_New_Candidate_Young_Stellar_Group_at_d_121_pc_3655_Associated_with_118_Tauri/3122689
Mamajek, E. E. 2007, in IAU Symp. 237, Triggered Star Formation in a Turbulent ISM, ed. B. G. Elmegreen & J. Palous (Cambridge: Cambridge Univ. Press), 442
Mamajek, E. E. 2013, AAS Meeting, 221, 331.06
Mamajek, E. E., & Bell, C. P. M. 2014, MNRAS, 445, 2169
Mamajek, E. E., & Feigelson, E. D. 2001, in ASP Conf. Ser. 244, Young Stars Near Earth: Progress and Prospects, ed. R. Jayawardhana & T. Greene (San Francisco, CA: ASP), 104
Mamajek, E. E., Lawson, W. A., & Feigelson, E. D. 1999, ApJL, 516, L77
Mamajek, E. E., Meyer, M. R., & Liebert, J. 2002, AJ, 124, 1670
Mason, B. D., Hartkopf, W. I., Gies, D. R., Henry, T. J., & Helsel, J. W. 2009, AJ, 137, 3358
McInnes, L., Healy, J., & Astels, S. 2017, JOSS, 2, 205
McKee, C. F., & Tan, J. C. 2002, Natur, 416, 59
Mermilliod, J.-C. 1995, in Information & On-Line Data in Astronomy, ed. D. Egret & M. A. Albrecht, Vol. 203 (Berlin: Springer), 127
Mermilliod, J. C., Turon, C., Robichon, N., Arenou, F., & Lebreton, Y. 1997, in ESA Spec. Publ. 402, Hipparcos—Venice '97, ESA SP-402, ed. R. M. Bonnet et al. (Noordwijk: ESA), 643
Meynet, G., Mermilliod, J. C., & Maeder, A. 1993, A&AS, 98, 477
Moeckel, N., Holland, C., Clarke, C. J., & Bonnell, I. A. 2012, MNRAS, 425, 450
Murphy, S. J., Lawson, W. A., & Bento, J. 2015, MNRAS, 453, 2220
Murphy, S. J., Lawson, W. A., & Bessell, M. S. 2013, MNRAS, 435, 1325
Neuhaeuser, R., Sterzik, M. F., Torres, G., & Martin, E. L. 1995, A&A, 299, L13
Nony, T., Robitaille, J. F., Motte, F., et al. 2021, A&A, 645, A94
Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 53
Offner, S. S. R., & Liu, Y. 2018, NatAs, 2, 896
Oh, S., Price-Whelan, A. M., Hogg, D. W., Morton, T. D., & Spergel, D. N. 2017, AJ, 153, 257
Onishi, T., Mizuno, A., Kawamura, A., Tachihara, K., & Fukui, Y. 2002, ApJ, 575, 950
Ortiz-León, G. N., Loinard, L., Dzib, S. A., et al. 2018, ApJ, 865, 73
Padoan, P., & Nordlund, Å. 2002, ApJ, 576, 870
Padoan, P., Pan, L., Haugbølle, T., & Nordlund, Å. 2016, ApJ, 822, 11
Palla, F., & Stahler, S. W. 1999, ApJ, 525, 772
Pang, X., Li, Y., Tang, S.-Y., Pasquato, M., & Kouwenhoven, M. B. N. 2020, ApJL, 900, L4
Pavlidou, T., Scholz, A., & Teixeira, P. S. 2021, MNRAS, 503, 3232
Pecaut, M. J., & Mamajek, E. E. 2016, MNRAS, 461, 794
Pecaut, M. J., Mamajek, E. E., & Bubar, E. J. 2012, ApJ, 746, 154
Phelps, R. L., & Lada, E. A. 1997, ApJ, 477, 176
Pöhnl, H., & Paunzen, E. 2010, A&A, 514, A81
Preibisch, T., Balega, Y., Hofmann, K.-H., Weigelt, G., & Zinnecker, H. 1999, NewA, 4, 531
Preibisch, T., Brown, A. G. A., Bridges, T., Guenther, E., & Zinnecker, H. 2002, AJ, 124, 404
Preibisch, T., Guenther, E., Zinnecker, H., et al. 1998, A&A, 333, 619
Preibisch, T., & Mamajek, E. 2008, The Nearest OB Association: Scorpius-Centaurus (Sco OB2), Vol. 5 (San Francisco, CA: ASP), 235
Preibisch, T., & Zinnecker, H. 1999, AJ, 117, 2381
Prosser, C. F. 1992, AJ, 103, 488
Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, ApJS, 190, 1
Rasmuson, N. H. 1921, MeLuS, 26, 3
Rebull, L. M., Koenig, X. P., Padgett, D. L., et al. 2011, ApJS, 196, 4
Reid, N. 2003, MNRAS, 342, 837
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2015, JATIS, 1, 014003
Rizzuto, A. C., Ireland, M. J., & Kraus, A. L. 2015, MNRAS, 448, 2737
Rizzuto, A. C., Ireland, M. J., & Robertson, J. G. 2011, MNRAS, 416, 3108
Rizzuto, A. C., Ireland, M. J., Robertson, J. G., et al. 2013, MNRAS, 436, 1694
Rizzuto, A. C., Vanderburg, A., Mann, A. W., et al. 2018, AJ, 156, 195
Rocha-Pinto, H. J., Scalo, J., Maciel, W. J., & Flynn, C. 2000, A&A, 358, 869
Röser, S., Schilbach, E., Goldman, B., et al. 2018, A&A, 614, A81
Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338
Rowell, N. 2013, MNRAS, 434, 1549
Salpeter, E. E. 1955, ApJ, 121, 161
Sana, H., de Mink, S. E., de Koter, A., et al. 2012, Sci, 337, 444
Sargent, A. I. 1979, ApJ, 233, 163
Schönrich, R., & Binney, J. 2009, MNRAS, 396, 203

Shkolnik, E. L., Allers, K. N., Kraus, A. L., Liu, M. C., & Flagg, L. 2017, AJ, 154, 69
Siess, L., Dufour, E., & Forestini, M. 2000, A&A, 358, 593
Sim, G., Lee, S. H., Ann, H. B., & Kim, S. 2019, JKAS, 52, 145
Song, I., Bessell, M. S., & Zuckerman, B. 2002, A&A, 385, 862
Song, I., Zuckerman, B., & Bessell, M. S. 2003, ApJ, 599, 342
Strom, S. E., Strom, K. M., & Carrasco, L. 1974, PASP, 86, 798
Sullivan, K., & Kraus, A. 2021, ApJ, 912, 137
Tachihara, K., Neuhäuser, R., Kun, M., & Fukui, Y. 2005, A&A, 437, 919
Tang, J., Bressan, A., Rosenfield, P., et al. 2014, MNRAS, 445, 4287
Tobin, J. J., Hartmann, L., Furesz, G., Mateo, M., & Megeath, S. T. 2009, ApJ, 697, 1103
Trumpler, R. J. 1921, LicOB, 333, 110
Urquhart, J. S., Figura, C. C., Moore, T. J. T., et al. 2014, MNRAS, 437, 1791
Vázquez-Semadeni, E., Palau, A., Ballesteros-Paredes, J., Gómez, G. C., & Zamora-Avilés, M. 2019, MNRAS, 490, 3061
Viana Almeida, P., Santos, N. C., Melo, C., et al. 2009, A&A, 501, 965
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Walch, S. K., Whitworth, A. P., Bisbas, T., Wünsch, R., & Hubber, D. 2012, MNRAS, 427, 625
Wang, S., & Chen, X. 2019, ApJ, 877, 116
Weinberger, A. J., Anglada-Escudé, G., & Boss, A. P. 2013, ApJ, 762, 118
Wichmann, R., Krautter, J., Schmitt, J. H. M. M., et al. 1996, A&A, 312, 439
Wright, N. J., & Mamajek, E. E. 2018, MNRAS, 476, 381
Yuan, Z., Chang, J., Banerjee, P., et al. 2018, ApJ, 863, 26
Zari, E., Brown, A. G. A., & de Zeeuw, P. T. 2019, A&A, 628, A123
Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, A&A, 620, A172
Zavagno, A., Russeil, D., Motte, F., et al. 2010, A&A, 518, L81
Zhang, Z., Liu, M. C., Best, W. M. J., et al. 2018, ApJ, 858, 41
Ziegler, C., Law, N. M., Baranec, C., et al. 2018, AJ, 156, 259
Zuckerman, B., Song, I., Bessell, M. S., & Webb, R. A. 2001, ApJL, 562, L87