IDENTIFYING PROTOCLUSTERS IN THE HIGH REDSHIFT UNIVERSE AND MAPPING THEIR EVOLUTION

by

JAY ROBERT FRANCK

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Astronomy

CASE WESTERN RESERVE UNIVERSITY

January, 2018

CASE WESTERN RESERVE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

We hereby approve the dissertation of

JAY ROBERT FRANCK

candidate for the **Doctor of Philosophy** degree.

(signed)

Dissertation Advisor/Committee Chair

Stacy S. McGaugh

Committee Member

Corbin Covault

Committee Member

R.E. Luck

Committee Member

Idit Zehavi

Date of Defense

July 25, 2017

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank $C_8H_{10}N_4O_2$. My advisor, Dr. Stacy McGaugh, deserves a lot of credit for significant encouragement and guidance throughout the CCPC project. The entire Astronomy Department (faculty, staff, students) have been instrumental in my education. Finally, I am grateful to my family and Megan Woo for continuous support and love.

Identifying Protoclusters in the High Redshift Universe and Mapping their Evolution

Abstract

by

JAY FRANCK

To investigate the growth and evolution of the earliest structures in the Universe, we identify more than 200 galaxy overdensities in the Candidate Cluster and Protocluster Catalog (CCPC). This compilation is produced by mining open astronomy data sets for overdensities of high redshift galaxies that are spectroscopically confirmed. At these redshifts, the Universe is only a few billion years old. This data mining approach yields a nearly ten fold increase in the number of known protoclusters in the literature. The CCPC also includes the highest redshift, spectroscopically confirmed protocluster at $z = 6.56$.

For nearly 1500 galaxies contained in the CCPC between redshifts of $2.0 < z < 6.6$, we find archival Spitzer images at 3.6 and 4.5 μm bandpasses. These Spitzer wavelengths serve as a proxy measurement for the stellar mass of the galaxies. The galaxies in protoclusters appear to be consistent with a passively evolving, older stellar population. We find no statistically significant difference between protocluster and field galaxy populations. Galaxy formation models suggest that galaxies in dense environments should be more massive. Comparing the brightness distribution of the data at different epochs provides an evolutionary track for how protocluster galaxies evolve.

We compare the data to the predictions of a large-scale simulation, the Millennium Run. We analyze the simulated data with the same suite of algorithms and metrics as in

the CCPC. The results of this exercise yield a number of significant discrepancies between the theoretical predictions and what is seen. The universe contains a much larger density of bright galaxies than what the model predicts. At $z > 2$, the brightest galaxies are older and more massive than anticipated by the model.

Chapter 1

Introduction

Imprinted on the Cosmic Microwave Background are minute density fluctuations of order $\Delta \sim 10^{-5}$. From this initial distribution, assumed to be Gaussian, the overdense regions of the universe coalesced by gravitational attraction (Zel'dovich, 1970). The largest of these overdensities are thought to correspond to modern day clusters of galaxies. These structures are characterized by 100-1000s of galaxies (Abell, 1958), within a relatively small radius of a few comoving megaparsecs (cMpc), gravitationally bound by a system mass of $\geq 10^{14}\ M_{\odot}$ (Zwicky, 1937). These systems are treated as laboratories for three key astronomical experiments: (1) how the largest structures form within the universe, (2) how galaxies evolve within dense environments, and (3) analyzing large numbers of galaxies throughout cosmic time. This introduction will touch on each of these, while the work presented in this thesis mainly investigates the latter two.

1.0.1 Cluster Formation

Much work has gone into characterizing these clusters theoretically. Understanding when these dense groups of galaxies began to come together, from how large a volume, and from what components are fundamental questions of our Universe. The spherical col-

lapse model of Jeans (1902) was one of the earliest forays into the origin of cluster collapse. With a more developed theoretical framework, Zel'dovich (1970) examined approximations to the collapse of small perturbations in a near-uniform field. Peebles (1970) ran a N-body simulation on the origin of the Coma cluster galaxies, showing that even with crude approximations, a collapsed structure could result. Thus it became mathematically clear that clusters could be the result of collapsing initial overdensities. Press & Schechter (1974) developed an analytical approximation to the evolving halo mass function. Their approximation provides a number density of halos as a function of cosmological parameters.

From here, observations of clusters, namely their masses and number densities, could become a test of various cosmological models (Cen et al., 1994; Eke et al., 1996b,a). In practice, this was used quite successfully to rule out non-flat universes and inform other structure formation parameters (σ_8) by comparing the theoretical results with observations (Davis & Peebles, 1983; Cen & Ostriker, 1994; Bahcall & Fan, 1998; Mortonson et al., 2011). As the halo mass function is steep and evolves with redshift significantly (Eke et al., 1996a), observations of high redshift systems offer the strongest constraints. This was precisely the aim of works like Bahcall & Fan (1998), and more recently (Vikhlinin et al., 2009). If a structure was found to exceed the model expectations, that model could be discarded. Identifying both the highest redshift (and greatest mass) cluster has the potential to test our theoretical understanding of the Universe.

1.0.2 Galaxy Evolution in Dense Environments

Clusters are not only useful as cosmological probes. They have been used extensively to trace galaxy formation and evolution. Galaxies within clusters have a number of characteristics that distinguish them from galaxies that are part of the field. Dressler (1980) noted that there was a strong morphological relation with the density of the galaxies, and not just

near clusters. In general, dense environments show a greater fraction of spheroidal galaxies (elliptical and lenticulars) with older, more massive stellar populations. This strongly suggests that either galaxy-galaxy interactions and/or the initial overdensity for a piece of the universe is a driver of these morphological changes. This nature- versus-nurture debate continues to this day. There are a number of physical processes that can theoretically change spiral, star-forming systems into more spheroidal shapes, like galaxy collisions, gas stripping, and tidal harassment, although the initial density field is also likely to play a role (De Lucia, 2007). It remains to be seen what the dominant mechanism, if any, control the fate of these galaxies.

Relaxed clusters are also characterized by a strong red sequence feature in the color-magnitude relation (CMR), which has been studied for many decades (Baum, 1959; Faber, 1973; Visvanathan & Sandage, 1977). Briefly, the CMR can be described as the brightest galaxies having the reddest colors. It exists primarily in the central regions of clusters (Cooper et al., 2008) which are filled with elliptical and lenticular of galaxies (Bower et al., 1992). The increasingly red color is thought to be metallicity driven, in that greater mass structures appear to retain more of their produced metals [Mass-Metallicity relation; Schombert & Rakos (2009)]. This provides a key clue to unlocking the cluster galaxy formation and evolutionary models. Although a number of mechanisms for this phenomenon have been proposed [wet versus dry mergers; Faber et al. (2007)], no consensus has been reached as of yet (Ferreras et al., 2012; Fassbender et al., 2014; Fritz et al., 2014). De Lucia & Blaizot (2007) suggest that for the Brightest Cluster Galaxies (BCGs), dry mergers are a key growth in the stellar mass at $z < 1$, while the building block galaxies form their stars early ($z_f > 5$).

These cluster galaxies were not always red sequence galaxies. Butcher & Oemler (1978) noticed that in the cores of galaxy clusters, galaxies became progressively bluer as the

redshift increased. Subsequent examinations only strengthened this relation to the 'high' redshift of $z \sim 0.5$ (Butcher & Oemler, 1984). Clearly, cluster galaxies must be evolving over time and were not born as 'red and dead' systems. Therefore, there must be some process(es) that are responsible for this transformation.

Tying the observations to the theory of galaxy formation and evolution has been a goal for nearly a hundred years. It is unsolved to date. A key driver of this thesis has been the investigation of the earliest galaxies in the Universe in the densest environments. When do galaxies first become quiescent? Is there an epoch at which environmental influence becomes pivotal? What can this timescale, if it exists, tell us about the physical process that drives it? To answer these questions, we need a large number of clusters over a range of epochs.

1.0.3 Identifying Galaxy Clusters across Cosmic Time

As the field of high redshift cluster and protocluster identification has matured, so too have the methods. No single technique is necessarily endorsed herein. Perhaps the most traditional approach is to identify volumes of space that contain a larger density of galaxies than neighboring systems. These can be seen visually as 'spikes' in $N(z)$ plots for deep surveys (Franck & McGaugh, 2016b). This has been employed for both clusters and protoclusters (Steidel et al., 1998; Stern et al., 2003; Hayashi et al., 2011; Papovich et al., 2012; Shimakawa et al., 2014; Strazzullo et al., 2016), but requires accurate distance measurements for a large number of galaxies. This necessitates expensive spectroscopic surveys.

Another classic technique is to search for the X-ray emission that accompanies the hot intra-cluster medium (ICM) region of a cluster (Vikhlinin et al., 2006, 2009; Gobat et al., 2011; Willis et al., 2013; Wang et al., 2016). Not only does this technique provide 'proof' of a dense, cluster-like environment, it can also suggest a mass estimate based on the tempera-

ture of the system. The ICM is heated to the virial temperature of the system. It is limited in some respects, as the flux follows the inverse-square law, and is therefore strongly distance dependent. X-ray point sources (AGN) can contaminate the signal and must be removed to get an accurate temperature. The technique also requires the presence of sufficiently hot gas to be detected, which may not be present at high redshift (Wang et al., 2016).

The Sunyaev-Zel'Dovich (SZ) Effect (Zeldovich & Sunyaev, 1969) is the upscattering of Cosmic Microwave Background photons by the hot ICM. Looking towards a region with an ICM at microwave wavelengths, one would see a flux differential with respect to the field at a given wavelength, thus identifying potential clusters. This technique has the benefit of being redshift independent, which theoretically allows the technique to be applied at earlier times in the universe. Many facilities are capitalizing on the use of the SZ effect, such as the South Pole Telescope (Keisler et al., 2011; Bleem et al., 2015; Saliwanchik et al., 2015), the Combined Array for Research in Millimeter-wave Astronomy (CARMA; Muchovej et al., 2007; Mantz et al., 2014), and the Planck Telescope (Planck Collaboration et al., 2013).

Quasars, Quasi-Stellar Objects (QSOs), High Redshift Radio Galaxies (HzRGs), and other AGN have long been used as tracers of structure (Hall et al., 2001; Venemans et al., 2002; Galametz et al., 2010; Wylezalek et al., 2014; Rigby et al., 2014; Franck et al., 2015; Paterno-Mahler et al., 2016). These objects are generally thought to be the most massive systems at any redshift, and are thus assumed to collapse in regions of the largest over-densities in the universe as proto-BCGs. This technique has been successful for clusters and protoclusters to $z \geq 5$ (Venemans et al., 2007). Typically, the regions surrounding a high redshift cluster are photometrically targeted, in narrow and broadband filters, to select potentially high-redshift companion galaxies to the massive system. These candidates can then be spectroscopically confirmed to ascertain their membership along the line of sight. These signposts are generally thought to be more rare than protoclusters themselves, which

may strongly bias this selection as a representative sample of protoclusters (Chiang et al., 2013).

1.0.4 The Realm of Protoclusters

A protocluster is simply the manifestation of a collapsing cluster. From an observers perspective, a protocluster consists of all components that *will* become part of a cluster by $z = 0$, including galaxies, gas, and dark matter. In practice, distinguishing which objects are bound by the system and which are merely interlopers within the volume is difficult (Contini et al., 2016). Throughout this work, we will focus on protoclusters at high redshift ($z \geq 2$), as this is above the epoch where the first clusters can potentially be in a virialized state (Chiang et al., 2013), although there are some possible exceptions (Wang et al., 2016). These systems generally have overdensities of order unity or less, compared to a virialized cluster which can have a density roughly 200 times the critical density. There do exist lower redshift ($z \sim 0.4$) protocluster candidates in the literature (Gonzalez et al., 2005), but these will not be focused on in this work.

These objects exist in an era of incredibly rapid growth and evolution. The redshift range where protoclusters are expected to be observable ($2 < z < 10$ roughly) represents only the first ~ 2.6 Gyrs of our nearly 14 billion year history. However, the halos in which clusters will collapse into can grow in mass by 3 orders of magnitude in this short time (Chiang et al., 2013). This epoch is also associated with rapid galaxy evolution and star formation throughout the universe (Madau & Dickinson, 2014). This coincidental timing of evolution on cluster and galaxy-sized scales grants these systems great leverage in simultaneously testing galaxy and structural growth models. In simulations, it has been shown that the gravitational collapse of protocluster galaxies into cluster systems removes much of their history from the observational record (Chiang et al., 2017). Thus, catching these

nascent galaxies prior to the advent of various environmental quenching processes can provide valuable information as to their origins. More details on their use and comparison to theory will be provided in the following Chapters of this thesis. It is for these reasons that protoclusters have long been the target of observations.

1.0.5 Protocluster Searches

The first observational evidence of a candidate protocluster came from Burbidge et al. (1980). They identified three quasi-stellar objects (QSOs) at $z \sim 2$ in the background of an image featuring M82. Nearly 15 years later, telescopes became powerful enough (8-10 meter class) to discover additional galaxy overdensities at high redshift spectroscopically (Malkan et al., 1996; Francis et al., 1996; Hu & McMahon, 1996; Pentericci et al., 1997; Steidel et al., 1998; Keel et al., 1999; Möller & Fynbo, 2001; Venemans et al., 2002) up to $z < 4$. This became the turning point in the hunt for protoclusters. Some of these systems surrounded HzRGs, while others were found in narrow-band (NB) imaging or Lyman-Dropout catalogs.

Those first groups were merely the vanguard of this new field. Deep spectroscopic catalogs of well-studied fields [e.g., CANDELS; Balestra et al. (2010)] offered great numbers of galaxies with precise redshifts and photometry. These deep fields offer a plethora of potential candidates, and without the need to target biased tracers of mass, like a HzRG (Wylezalek et al., 2014).

To mention each protocluster survey, their technique, and target, goes beyond the scope of this thesis introduction. However, we can sort them broadly by their discovery method. Many modern protocluster searches have looked for 3D overdensities in these deep fields (Diener et al., 2013; Cucciati et al., 2014; Kang & Im, 2015; Toshikawa et al., 2016; Lemaux et al., 2017). The spectroscopic redshifts offer unique insight into the distance of respective

members along the line of sight. There is the added benefit of having a galaxy's spectral lines, which can be used to get a rough idea of their star-formation rate and metallicity in some cases.

In comparison, a rough redshift can be ascertained using a series of images with narrow-band (NB) filters that target emission lines of star forming galaxies (Möller & Fynbo, 2001), typically targeting galaxies around HzRGs. The width of these filters gives little line of sight information for individual galaxies, but is of order 10s of cMpc (Venemans et al., 2002), similar to a photometric redshift. At redshifts of $2 < z < 2.5$, the $H\alpha$ line ($\lambda_{rest} = 6563\ A$) is commonly used (Hatch et al., 2011; Hayashi et al., 2012; Husband et al., 2016). Higher redshift systems are more commonly targeted using the $Ly\alpha$ line at rest-frame $\lambda = 1216$ A (Venemans et al., 2007; Kuiper et al., 2011; Lee et al., 2014; Adams et al., 2015). This technique is much quicker in identifying candidate protoclusters than a blind spectroscopic search. However, there are a few downsides. Low redshift line emitters ([O II], [O III]) can occasionally enter the sample, but this is generally thought to be $< 10\%$ (Venemans et al., 2007). The line of sight structure of these protocluster candidates cannot be known for these NB images until spectroscopically confirmed. Finally, filters need to have bandpasses centered on the redshifted emission lines being targeted, limiting their generalized use.

While the majority of detections belong to the upper two search strategies, there are some exceptions. Targeting a different class of galaxies, some research groups have used far-infrared, sub-millimeter and other long wavelengths to detect dusty galaxies (Hodge et al., 2013; Rigby et al., 2014; Casey et al., 2015) that are clustered in space. It is expected that most galaxies are not spectroscopically detectable at high redshift because of their very red colors and dust obscuration (van Dokkum et al., 2006; Spitler et al., 2014), so these types of observations offer a tantalizing view of protoclusters. Recently, a few groups have used absorption features in spectra to detect gaseous overdensities surrounding HzRGs as

a method to detect protoclusters (Cai et al., 2016; Ogura et al., 2017). Finally, Wang et al. (2016) has made a startlingly discovery of a cluster/protocluster at redshift $z \sim 2.5$ via measurements of this primordial ICM. This is the most distant detection of the ICM known, and the largest mass system at $z > 2$.

1.0.6 Inherent Biases and Survey Challenges

By their nature, protoclusters are diffuse. To build a gravitational bound object of $M_{z=0} \geq 10^{14}\ M_{\odot}$, a significant volume of a near-uniform universe must collapse. For a rich cluster, the value is of order 8 cMpc (Evrard, 1989). For the most massive of protoclusters ($M_{z=0} \geq 10^{15}\ M_{\odot}$) at redshifts $z > 2$, simulations suggest that member galaxies can span 40 cMpc in diameter (Muldrew et al., 2015). This amount of real estate can be unwieldy from an observational perspective; surveys need to be both wide (~0.5 deg) and deep ($z > 2$) to capture an entire protocluster, which is expensive.

It is clear that these past surveys are inherently biased. Many observational studies of protoclusters can only capture the inner regions of the system because of the telescope's small FOV (Venemans et al., 2002; Hatch et al., 2011). As noted earlier, photometric redshifts (or narrowband imaging) can only give line of sight distances to 10s of comoving Mpc, which is insufficient to ascertain cluster membership. Some of these candidate overdensities are found around HzRGs or other cluster signposts (Wylezalek et al., 2014), which are orders of magnitude more rare than protoclusters. Most galaxy searches also only target star-forming galaxies, which are easier to spectroscopically confirm. These probably do not represent the bulk of the population at high redshift (van Dokkum et al., 2006). In a later section of this thesis, an explanation will be provided of the reasons that the approach presented here is less biased than these other searches.

1.0.7 Protocluster Galaxies and their Evolution

How do galaxies in dense environments look compared to their field counterparts at high redshift? Is there a characteristic redshift at which (proto)cluster galaxies are sufficiently distinct to be detectable by observations alone? If the universe grows hierarchically, there should exist an epoch at which the galaxies in a variety of environments look indistinguishable. The red sequence feature can be readily seen to $z \leq 1.5$ within some clusters (Stanford et al., 1995; Rakos & Schombert, 1995; Stanford et al., 1998; Eisenhardt et al., 2008; Mei et al., 2009), suggesting the era of convergence could exist at larger redshifts.

There is some tension within the literature as to the difference (if any) of field and overdense galaxies at $z > 2$. Broadly, there are two camps. In one, mainly dominated by studies of single protoclusters, there appears to be an inherent signature in a protocluster's galaxies. In comparison, studies of multiple systems at high redshift generally show little differentiation of protocluster members with respect to the field.

For example, a number protoclusters at $z \leq 2.5$ may have galaxies that have greater stellar mass with respect to the field (Hatch et al., 2011; Casey et al., 2015), which makes sense on an intuitive scale. There is another study that suggests that *Ly*α emission of these galaxies is greater (Zheng et al., 2016) than their field counterparts. Dey et al. (2016) do not find such an effect, but do note that the overdense galaxies appear younger when compared to field galaxies at $z = 3.8$. Toshikawa et al. (2016) finds smaller Lyman α EW_0 in a protocluster at $z = 3.67$ when compared to the field. Other works suggest that only quiescent galaxies exhibit such differences (Hayashi et al., 2011, 2012). Hatch et al. (2017) found that group environments within a protocluster at high redshift were the only regions in which differentiation could be found, which is similar to what is found in nearby galaxies (Ellison et al., 2009).

Although it is hard to discount any of the previous findings out of hand, it is hard to reconcile the results into a coherent narrative. Some of the findings appear contradictory, while many can be strongly influenced by dust (particularly the star forming galaxies). Small number statistics at discrete redshifts in space can hardly inform galaxy formation models across gigayears of time. The problem is compounded considering the variety of survey types and analyses.

Studies that span a few epochs in time can tell a more complete story of bulk evolution. Ownsworth et al. (2016), Zhao et al. (2017), Diener et al. (2013, 2015) tracked galaxies in a variety of deep fields and compared their evolution to stellar population models. In general, the findings of galaxies in these dense environments were that they were born early ($z \geq 3$) and were mainly driven by passive evolution. Diener et al. (2013) found that the spectroscopic proto-groups and clusters were statistically indistinguishable from field populations.

This is a curious result when compared to the individual protocluster analysis. It could suggest that either the data trends are too noisy over long epochs, too varied from system to system, or that field and protocluster systems are indeed similar. The above examples varied widely in the types of data used. What if a single quantity were measured over many epochs? Stellar mass is a metric that it relatively insensitive to systematics and timescales. It serves as a natural choice to examine the growth of galaxies over time

1.0.8 Stellar Mass of High Redshift Galaxies in the Infrared

The stellar mass of high redshift cluster and protocluster galaxies has been frequently measured (Andreon, 2006; Mancone et al., 2010; Papovich et al., 2012; Andreon, 2013; Wylezalek et al., 2013, 2014), especially with *The Spitzer Space Telescope*. *Spitzer's* IRAC camera (3.6-8μm) can capture the NIR rest-frame emission of a galaxy's stars at high red-

shift, which serves as a stellar mass proxy. In many instances, the cluster galaxies are binned in redshift space, and the magnitude distributions are fit with a Schechter function (Schechter, 1976). The characteristic magnitude $m^*(z)$ can be then estimated, and fit with stellar evolution models.

This provides some sense of the evolution of dense galaxies over time. Apart from a few anomalous instances (Mancone et al., 2010), the evolution of these overdense systems from $0.3 < z < 3.1$ are consistent with passive evolution of an older stellar population (Wylezalek et al., 2014). While individual protocluster galaxies may look distinct from their field counterparts, it appears that *en masse*, these systems have fairly predictable existences. This is similar to the findings above at other wavelengths Diener et al. (2013, 2015); Ownsworth et al. (2016); Zhao et al. (2017) and in stark contrast to the individual protocluster studies. Does theory offer any insight into what is happening with these observations?

1.0.9 Simulations of Protoclusters

The theoretical expectation is that protocluster environments will play host to the most massive galaxies in the universe at $z > 2$ (Muldrew et al., 2015). The greater the initial density field, the greater the halo mass can be. However, there is also the expectation from some models that general properties of interloping galaxies are nearly identical to those bound to a cluster by $z = 0$ (Contini et al., 2016). It can be hard to compare the simulations to observations in an identical fashion. Primarily this is because models, at best, can only hope to emulate the universe at some level. Second, we can 'observe' the distribution of dark matter within a simulation, while the same cannot be said of observations.

The growth stage of protoclusters in simulations is quite rapid. By $z \sim 2$, the most massive halos in the universe can be cluster-sized ($M \sim 10^{14}\ M_{\odot}$) and virialized (Chiang et al., 2013). By the same time, protocluster galaxies are expected to have attained half

of their final, $z = 0$ stellar mass by rapidly forming stars (Chiang et al., 2017). This time period should exhibit a rapid stellar mass growth in the galaxies. However, the high redshift data points of Wylezalek et al. (2014) show no such behavior. It remains to be seen if other high redshift observations will further highlight this discrepancy.

1.0.10 This Thesis

The following four chapters represent four steps in the construction of the Candidate Cluster and Protocluster Catalog, hereafter the CCPC. The project bypasses a number of stumbling blocks of other works by utilizing large open data sets and a unified protocluster finding algorithm.

The first foray into identifying spectroscopically-confirmed protoclusters begins with our pilot project, the CCPC1 (Franck & McGaugh, 2016b). Hereafter it will be referred to as CCPC1. Many previous protocluster surveys had relied on narrow-band imaging (Venemans et al., 2002, 2004, 2007; Hatch et al., 2011), large catalogs of Lyman-Break Galaxies (Steidel et al., 1998, 2003, 2005), or deep spectroscopic catalogs of galaxies (Diener et al., 2013; Chiang et al., 2014; Balestra et al., 2010). Each survey had varying FOVs, selections functions, and targeting criteria. The individual approaches, algorithms, and definitions of what constituted a protocluster made it difficult to compare results. Prior to this work, roughly 20 protoclusters with spectroscopic data were known in the literature, which were mostly contained within Table 5 of Chiang et al. (2013).

To tackle this project, we compiled a catalog of 14,000 galaxies spanning the redshift range of $2.74 < z < 3.7$. This spread approximately corresponds to the traditional Lyman-Break dropout selection (Steidel et al., 1998). Using a relatively simple algorithm, we constructed a catalog of 43 candidate protoclusters that existed as overdensities on Mpc scales. The pilot program will be examined in detail in Chapter 2.

Following the success of the first CCPC catalog, we built an expanded redshift range ($2 \leq z < 7$) spectroscopic galaxy list from which we could identify galaxy overdensities. At the low redshift end ($z \sim 2$), the most massive clusters have the potential to be virialized (Chiang et al., 2013). At the high redshift end of the catalog, there are few galaxies from which an overdensity can be measured. In CCPC2 (Franck & McGaugh, 2016a), we employed our simple algorithm from CCPC1 to again identify candidate protoclusters from 40,000 spectroscopic galaxies. With this expanded data set, we add 173 protocluster candidates to the CCPC, for a total of 216. This is nearly a 10× increase from the literature census at the beginning of this work. In CCPC2, the most distant spectroscopic candidate protocluster is listed at $z = 6.56$. Chapter 3 examines the CCPC2, and delves deeper into the significance of these objects than in Chapter 2.

Chapters 2, 3 provided the data from which greater insight into the nature of the protoclusters and their constituent galaxies form and evolve. Chapter 4 is the attempt to answer a simple question: How does the stellar mass of galaxies in dense environments (protoclusters) evolve as a function of redshift. We obtained Spitzer IRAC data (3.6-8μm) from the Spitzer Heritage Archive for roughly 75% of the 2048 CCPC galaxies in our sample. At the redshifts $2 < z < 6.6$, these wavelengths correspond to rest-frame NIR emission, a proxy for the underlying stellar mass of a galaxy. By fitting the Luminosity Functions of galaxies in discrete redshift bins, we could determine a characteristic brightness (stellar mass) over time via the Schechter Function (Schechter, 1976).

We were also able to compare the protocluster galaxies in dense volumes with isolated, field galaxies as a way of quantifying the role that environment plays in shaping the stellar mass of these objects. Field and protocluster galaxies share an evolution that cannot be distinguished with our data. The results presented in Chapter 4 suggest that the underlying stellar mass of these galaxies changes little over the first few Gyrs after the Big Bang.

The last chapter in the CCPC Saga is the comparison of the CCPC data to theoretical expectations found within large scale simulations. Many previous protocluster surveys have either relied upon these simulations to inform their search techniques (Chiang et al., 2014), or to map their findings into a ΛCDM framework for interpretive purposes (Venemans et al., 2007; Diener et al., 2013; Kang & Im, 2015; Toshikawa et al., 2016). Most of the previous works have utilized the Millennium Run simulation (Springel et al., 2005), for which exists a convenient, queryable database, as well as a variety of value-added products and semi-analytic models (Guo et al., 2011; Henriques et al., 2015). In a consistent way, we 'observe' mock data sets (called lightcones) as we had the CCPC. The same analyses were run, and the results compared to observations. Chapter 5 details the fruits of this exercise, and notes the large discrepancies that exist in comparing the simulation expectations to the data. In general, the mock galaxies within the simulation are significantly fainter (up to 3 magnitudes) at all redshifts. Thus, any ΛCDM comparisons one could make to the data are largely overshadowed by the incompatibility to the simulations.

Chapter 2

The Candidate Cluster and Protocluster Catalog I

Identifying structures over varying epochs offers the possibility of examining the assembly history of the Universe and the environmental evolution of galaxies in structures. Initial density perturbations at the earliest times, aided by gravity, give rise to present day clusters. Although still relatively rare, $z > 1$ clusters are being discovered with increasing regularity based on surveys in the near-infrared (NIR) for galaxy overdensities around radio-loud sources (Hall & Green, 1998; Hall et al., 2001; Franck et al., 2015), in the field (Papovich, 2008; Eisenhardt et al., 2008), by diffuse X-ray emission of the hot intra-cluster medium (Gobat et al., 2011; Willis et al., 2013), and from the Sunyaev-Zel'Dovich (SZ) effect (Mantz et al., 2014; Rettura et al., 2014). A larger sample of cluster candidates, especially at increasing redshift, may offer a clearer picture for structure evolution. Protoclusters, loosely defined here as structures $z \gtrsim 2$, are much rarer, and only a handful of studies containing more than a single object at $z > 2.75$ exist (Venemans et al., 2007; Galametz et al., 2012; Diener et al., 2013; Chiang et al., 2014; Rigby et al., 2014; Wylezalek et al., 2014). With a larger catalog of clusters at various stages of assembly, the general trends of

formation and galaxy evolution can be analyzed over time, instead of being dissected epoch by epoch.

At these higher redshifts, there have been a number of amazing discoveries of groups of galaxies found at $z \sim 4$ (Lee et al., 2014; Venemans et al., 2002), $z > 6$ (Utsumi et al., 2010; Toshikawa et al., 2014), up to $z \sim 8$ (Trenti et al., 2012). The emergence of these protoclusters offer a test of ΛCDM, as simulations provide constraints on the number of clusters and structure mass as a function of redshift and the dark energy equation of state (Mortonson et al., 2011; Vikhlinin et al., 2009), with greater leverage provided at larger redshifts. More fundamentally, these 'pink elephants' offer a glimpse into the unexpected mysteries of this discovery driven science.

By comparing these primordial objects to their assumed, present-day manifestations as rich clusters, the physical processes that foster their evolution can be better understood. To date, a clear evolutionary path has not emerged to tie these high redshift structures to the $z = 0$ clusters observed. For instance, the assembly process for the defining red sequence (RS) feature of nearby clusters is not well established (Ferreras et al., 2012; Fassbender et al., 2014; Fritz et al., 2014). Estimating the formation epoch for the stellar population of clusters has also proven challenging. A range from $z_f = 2$ to $z_f = 30$ is found using a variety of methods at various wavelengths (Rakos & Schombert, 1995; Eisenhardt et al., 2008; Ferreras et al., 2012; van de Sande et al., 2014; Wylezalek et al., 2014).

To this end, we have compiled a list of 43 candidate protoclusters assembled using archival measurements in the redshift range of $2.74 < z < 3.71$. The majority of these candidate structures, at the time of writing, have not been previously identified to the best of our knowledge. This catalog is the largest list of high redshift protoclusters based on spectroscopic redshifts. Section 3.1 explains the search criteria used to identify candidate structures, with the general results explored in Section 5.4. Section 5.5 gives a brief sum-

mary of the findings. The Appendix contains sky position plots and $N(z)$ histograms for each protocluster.

This work assumes a ΛCDM concordance cosmology, with $\Omega_\Lambda = 0.7$, a matter density of $\Omega_m = 0.3$, and $H_0 = 70$ km s^{-1} Mpc^{-1}. At the redshift range of $z = 2.74 \rightarrow 3.7$, the angular size using this cosmology is 0.47 Mpc arcmin^{-1} to 0.43 Mpc arcmin^{-1}, 20 comoving Mpc along the line of sight has a $\Delta z \sim 0.019$ to 0.026, whereas the corresponding age of the Universe is 2.3 to 1.7 Gyrs, respectively (Wright, 2006).

2.1 The Candidate Cluster and Protocluster Catalog (CCPC)

Table 2.1. Candidate Cluster and Protocluster Catalog (CCPC) - Strongest Candidates

Candidate Name	RA	DEC	Redshift	σ_z	Number of Galaxies	Galaxies in $R < 10$ cMpc	Overdensity (δ_{gal})	Cluster Probability (%)	Reference
CCPC-z27-002	02:21:19.92	-04:27:43.20	2.772	0.007	5	5	11.02 ± 6.90[a]	100	
CCPC-z29-001	09:33:35.71	28:44:45.89	2.918	0.005	9	8	11.21 ± 4.76	100	
CCPC-z29-002	09:45:32.76	-24:29:05.28	2.919	0.009	26	26	12.91 ± 4.55	100	1
CCPC-z30-001	12:08:06.67	-30:31:05.16	3.035	0.005	8	8	18.78 ± 10.14	100	2
CCPC-z30-003	22:17:25.92	00:12:37.58	3.096	0.008	54*	25	12.28 ± 2.42	100	3
CCPC-z31-003	03:18:07.58	-25:34:55.56	3.133	0.008	33	33	9.80 ± 2.77	100	1
CCPC-z31-004	01:06:11.53	-25:46:14.52	3.146	0.006	6	6	7.59 ± 4.65	85	
CCPC-z31-005	20:09:54.43	-30:41:19.68	3.152	0.007	12	12	17.77 ± 9.19[c]	100	1
CCPC-z32-002	02:59:04.16	00:12:35.60	3.234	0.003	5	5	13.11 ± 8.63	100	
CCPC-z33-002	02:03:35.13	11:38:06.68	3.372	0.008	5	5	7.44 ± 4.47	85	4
CCPC-z35-001	00:03:25.12	-26:04:52.68	3.597	0.003	4	4	10.18 ± 8.05	100	
CCPC-z36-001	02:03:49.52	11:35:53.48	3.644	0.003	4	4	23.50 ± 14.39	100	

[a]No field galaxies were identified along the line of sight $\Delta z \pm 0.15$. This δ_{gal} should be treated as an upper limit.

[c]Large numbers of diffuse field galaxies within $\Delta z \pm 0.15$ gave a $\delta_{gal} \sim 0$. By limiting the field galaxies to the same surface area (RA/DEC) as the galaxies within the protocluster, these became positive overdensities.

Note. — The names and positions for the strongest CCPC overdensities, along with the mean redshift and dispersion of galaxies. The naming scheme is explained in Section 3.1. For each candidate, the number of spectroscopically confirmed members is listed. If the number of galaxies is followed by a '*', this indicates that the protocluster is an extended system in which 1/3 or more members in the structure are found between $10 < R < 20$ comoving Mpc. The implications of this are discussed in Section 5.4. The galaxy overdensity is listed under δ_{gal}, and the basis of this calculation is explained in Section 2.3. Using each candidates overdensity, we have assigned a conservative probability that the structure will collapse into a massive cluster by $z = 0$ from the values in Figure 8 in Chiang et al. (2013), which plot the fate of overdensities within the Millennium simulation. Structures with greater than 85% probability are included in this Table. If the structure has been previously identified, its reference is included. References: (1) Venemans et al. (2007) (2) Möller & Fynbo (2001) (3) Steidel et al. (1998) (4) Ellison et al. (2001)

Table 2.2. Candidate Cluster and Protocluster Catalog (CCPC) - Additional Candidates

Candidate Name	RA	DEC	Redshift	σ_z	Number of Galaxies	Galaxies in $R < 10$ cMpc	Overdensity (δ_{gal})	Cluster Probability (%)	Reference
CCPC-z27-001	17:01:03.84	64:11:56.40	2.748	0.006	8	8	2.51 ± 1.37	18	1
CCPC-z27-003	03:04:42.00	-00:07:40.80	2.788	0.009	7	7	1.73 ± 0.96[c]	1	
CCPC-z27-004	12:36:51.84	62:11:56.40	2.803	0.007	6*	4	1.98 ± 0.13	10	
CCPC-z27-005	22:32:48.96	-60:31:55.20	2.798	0.006	5	5	6.65 ± 4.38	73	
CCPC-z28-001	14:24:34.08	22:49:04.80	2.814	0.010	9	8	1.97 ± 0.89	10	
CCPC-z28-002	03:32:14.40	-27:52:37.20	2.818	0.009	25*	14	1.27 ± 0.32	10	
CCPC-z28-003	02:59:05.76	00:11:27.60	2.825	0.014	10	10	1.01 ± 0.50	10	
CCPC-z28-004	00:54:28.56	-23:53:49.20	2.857	0.009	31	30	0.75 ± 0.25	1	2
CCPC-z28-005	21:41:59.76	-44:13:26.40	2.856	0.005	23	23	3.98 ± 1.72	48	
CCPC-z28-006	03:32:38.17	-27:47:08.16	2.863	0.009	14*	2[b]	0.31 ± 0.11	1	
CCPC-z28-007	03:04:41.13	-00:10:38.78	2.864	0.010	7	7	1.52 ± 0.83	10	
CCPC-z29-003	14:18:04.13	52:29:54.28	2.924	0.013	30*	18	1.36 ± 0.35	10	
CCPC-z29-004	12:36:47.71	62:12:55.80	2.931	0.010	26	25	2.93 ± 0.83	18	
CCPC-z29-005	14:24:45.91	22:56:42.11	2.974	0.008	13	9	3.26 ± 1.11	48	
CCPC-z29-006	02:58:53.07	00:09:53.60	2.982	0.013	6	6	0.52 ± 0.28	1	
CCPC-z29-007	03:32:19.29	-27:43:03.14	2.979	0.012	14*	7	0.30 ± 0.10	1	
CCPC-z30-002	14:24:30.29	22:52:49.91	3.074	0.012	15	11	1.12 ± 0.41	10	
CCPC-z31-001	03:32:24.31	-27:41:52.44	3.113	0.009	27*	16	3.67 ± 0.93	48	3
CCPC-z31-002	04:22:18.53	-38:45:15.48	3.118	0.009	9	9	0.47 ± 0.28	1	
CCPC-z31-006	13:49:06.70	-03:36:32.00	3.166	0.010	13	11	0.98 ± 0.46	1	
CCPC-z31-007	12:36:47.35	62:15:12.74	3.179	0.014	12	10	0.55 ± 0.24 [c]	1	
CCPC-z32-001	12:05:19.10	-07:42:31.00	3.206	0.010	14	14	0.67 ± 0.35	1	
CCPC-z32-003	03:32:34.54	-27:46:12.36	3.258	0.014	15*	3[b]	1.1 ± 0.43[c]	10	
CCPC-z33-001	12:36:50.74	62:13:49.51	3.363	0.008	9	9	1.48 ± 0.65	10	
CCPC-z33-003	03:32:34.90	-27:47:08.16	3.368	0.013	11*	5	1.70 ± 0.77[c]	10	
CCPC-z33-004	01:06:04.60	-25:46:54	3.388	0.010	10	10	3.63 ± 1.91	48	4
CCPC-z33-005	22:39:38.45	11:54:09.61	3.389	0.014	14	11	2.74 ± 1.19	18	
CCPC-z34-001	12:36:52.42	62:13:37.74	3.423	0.014	7	7	0.61 ± 0.30	1	
CCPC-z34-002	03:32:18.48	-27:52:06.60	3.476	0.010	23*	11	3.75 ± 1.02	48	
CCPC-z36-002	03:32:16.90	-27:48:38.16	3.658	0.012	8*	5	0.32 ± 0.14	1	
CCPC-z37-001	03:32:10.42	-27:40:53.04	3.704	0.013	11*	7	1.02 ± 0.42	10	

[a]For all footnotes (a, b, c), see Franck & McGaugh (2016b).

Note. — Identical to Table 2.1, but for candidate structures with < 85% probability to collapse based on their galaxy overdensity (δ_{gal}). The probabilities are assigned based on overdensities analyzed by Chiang et al. (2013) within the Millennium simulation (their Figure 8). If the structure has been previously identified, its reference is included. References: (1) Prescott et al. (2008) (2) Venemans et al. (2007) (3) Bond et al. (2010) (4) Frank et al. (2003)

Table 2.3. CCPC: Member Redshift Reference List

Candidate Name	Redshift References
CCPC-z27-001	Peter et al. (2007); Shapley et al. (2005); Hewett & Wild (2010); Law (2008)
CCPC-z27-002	Adams et al. (2011)
CCPC-z27-003	Steidel et al. (2003); Cowie et al. (1995)
CCPC-z27-004	Papovich et al. (2001); Moth & Elston (2002); Fernández-Soto et al. (2001); Conselice et al. (2003)
...	Steidel et al. (2003); Reddy et al. (2006)
CCPC-z27-005	Iwata et al. (2005); Cristiani et al. (2000)
CCPC-z28-001	Steidel et al. (2003)
CCPC-z28-002	Balestra et al. (2010); Wuyts et al. (2008); Schreier et al. (2001); Santini et al. (2009)
...	Pirzkal et al. (2013); Conselice et al. (2011)
CCPC-z28-003	Songaila (1998); Steidel et al. (2003)
CCPC-z28-004	Venemans et al. (2007); McCarthy et al. (1996)
CCPC-z28-005	Storrie-Lombardi & Wolfe (2000); Fynbo et al. (2003)
CCPC-z28-006	Balestra et al. (2010); Santini et al. (2009); Grazian et al. (2006)
CCPC-z28-007	Steidel et al. (2003)
CCPC-z29-001	Steidel et al. (2003)
CCPC-z29-002	Venemans et al. (2007); Roettgering et al. (1997); Doherty et al. (2010)
CCPC-z29-003	Steidel et al. (2003); Georgakakis et al. (2006)
CCPC-z29-004	Steidel et al. (2003); Cowie et al. (2004); Pirzkal et al. (2013); Moth & Elston (2002); Adams et al. (2011)
...	Reddy et al. (2006); Dawson et al. (2001); Chapman et al. (2004c); Barger et al. (2002)
CCPC-z29-005	Steidel et al. (2003)
CCPC-z29-006	Steidel et al. (2003)
CCPC-z29-007	Balestra et al. (2010); Le Fèvre et al. (2004); Pirzkal et al. (2013)
...	Wuyts et al. (2008); Bond et al. (2011)
CCPC-z30-001	Möller & Fynbo (2001); Fynbo et al. (2001)
CCPC-z30-002	Steidel et al. (2003); Petry et al. (1998)
CCPC-z30-003	Chapman et al. (2004a); Steidel et al. (2003); Inoue et al. (2011); Steidel et al. (2000)
...	Hayashino et al. (2004); Smail et al. (2004); Nestor et al. (2011); Yamada et al. (2012)
...	Lehmer et al. (2009b); Chapman et al. (2004b); Lehmer et al. (2009a)
CCPC-z31-001	Bond et al. (2010); Wuyts et al. (2008); Santini et al. (2009); Ciardullo et al. (2012)
...	Balestra et al. (2010); Cristiani et al. (2000)
CCPC-z31-002	Cantalupo et al. (2007); Osmer et al. (1994)
CCPC-z31-003	Venemans et al. (2005); Kuiper et al. (2012); Maschietto et al. (2008); Le Fevre et al. (1996)
CCPC-z31-004	Noll et al. (2004)
CCPC-z31-005	Venemans et al. (2007)
CCPC-z31-006	Fynbo et al. (2003)
CCPC-z31-007	Adams et al. (2011); Pirzkal et al. (2013); Steidel et al. (2003); Moth & Elston (2002)

Table 2.3 (cont'd)

Candidate Name	Redshift References
...	Reddy et al. (2006); Barger et al. (2008, 2002)
CCPC-z32-001	Grove et al. (2009)
CCPC-z32-002	Steidel et al. (2003)
CCPC-z32-003	Wuyts et al. (2009); Balestra et al. (2010)
...	Wuyts et al. (2008); Bond et al. (2011)
CCPC-z33-001	Lowenthal et al. (1997); Fernández-Soto et al. (2001); Adams et al. (2011); Lanzetta et al. (1996)
...	Steidel et al. (2003); Pirzkal et al. (2013); Reddy et al. (2006)
CCPC-z33-002	Ellison et al. (2001); Steidel et al. (2003); de Bruyn et al. (1996)
CCPC-z33-003	Balestra et al. (2010); Fiore et al. (2012); Bond et al. (2011)
...	Wuyts et al. (2008)
CCPC-z33-004	Noll et al. (2004); Mehlert et al. (2002)
CCPC-z33-005	Steidel et al. (2003)
CCPC-z34-001	Fernández-Soto et al. (1999); Moth & Elston (2002); Dawson et al. (2001); Adams et al. (2011)
...	Barger et al. (2003); Steidel et al. (2003); Reddy et al. (2006)
CCPC-z34-002	Balestra et al. (2010); Vanzella et al. (2006); Szokoly et al. (2004)
...	Wuyts et al. (2008)
CCPC-z35-001	Steidel et al. (2003); Savaglio et al. (1997)
CCPC-z36-001	Ellison et al. (2001); Steidel et al. (2003)
CCPC-z36-002	Gnerucci et al. (2011); Balestra et al. (2010); Vanzella et al. (2006); Wuyts et al. (2009)
...	Le Fèvre et al. (2004); Grazian et al. (2006)
CCPC-z37-001	Conselice et al. (2011); Balestra et al. (2010); Pirzkal et al. (2013)
...	Vanzella et al. (2006); Fontanot et al. (2007)

Note. — The protocluster candidates matched with references for the spectroscopic measurements of their respective members.

Table 2.4. CCPC Mass Estimates

Candidate Name	δ_m (b = 3)	Overdensity Volume (cMpc^3)	Overdensity Mass Estimate (10^{14} $M_\odot$)	R_{hms} (Mpc)	σ (km/s)	Virial Mass Estimate (10^{14} $M_\odot$)
CCPC-z27-001	0.84	193	0.1	2250	486	2.5
CCPC-z27-002	3.67	45	0.1	1688	541	2.3
CCPC-z27-003	0.58	220	0.1	2063	686	4.5
CCPC-z27-004	0.66[c]	224	0.2	1125	582	1.8
CCPC-z27-005	2.22	19	< 0.1	1375	436	1.2
CCPC-z28-001	0.66	587	0.4	2688	758	7.2
CCPC-z28-002	0.42	11168	6.7	3938	735	9.9
CCPC-z28-003	0.34	1048	0.6	4563	1071	24.4
CCPC-z28-004	0.25	2642	1.4	3188	724	7.8
CCPC-z28-005	1.33	828	0.8	2063	352	1.2
CCPC-z28-006	0.10	6040	2.9	5313	672	11.2
CCPC-z28-007	0.51	1197	0.8	3625	781	10.3
CCPC-z29-001	3.74	601	1.2	2313	357	1.4
CCPC-z29-002	4.30	1640	3.6	2500	699	5.7
CCPC-z29-003	0.45	10147	6.1	4313	988	19.6
CCPC-z29-004	0.98	3742	3.1	2375	736	6.0
CCPC-z29-005	1.09	3393	3.1	3125	614	5.5
CCPC-z29-006	0.17	29	< 0.1	3500	995	16.1
CCPC-z29-007	0.10	6579	3.1	4438	878	15.9
CCPC-z30-001	6.26	71	0.2	1000	408	0.8
CCPC-z30-002	0.37	3033	1.8	4063	917	15.9
CCPC-z30-003	4.09	14192	30.5	3938	616	6.9
CCPC-z31-001	1.22	13398	12.4	3313	629	6.1
CCPC-z31-002	0.16	575	0.3	2438	645	4.7
CCPC-z31-003	3.27	1740	3.1	2250	581	3.5
CCPC-z31-004	2.53	44	0.1	1125	462	1.1
CCPC-z31-005	5.92[c]	424	1.3	2000	482	2.2
CCPC-z31-006	0.33	1393	0.8	3188	716	7.6
CCPC-z31-007	0.18[c]	3465	1.9	4625	1038	23.2
CCPC-z32-001	0.22	860	0.4	2000	739	5.1
CCPC-z32-002	4.37	30	0.1	1063	218	0.2
CCPC-z32-003	0.35[c]	8715	5.1	5000	955	21.2
CCPC-z33-001	0.49	154	0.1	1688	520	2.1
CCPC-z33-002	2.48	25	< 0.1	1063	514	1.3
CCPC-z33-003	0.56[c]	2348	1.6	4250	869	14.9
CCPC-z33-004	1.21	245	0.2	2375	703	5.5

Table 2.4 (cont'd)

Candidate Name	δ_m (b = 3)	Overdensity Volume (cMpc^3)	Overdensity Mass Estimate (10^{14} $M_\odot$)	R_{hms} (Mpc)	σ (km/s)	Virial Mass Estimate (10^{14} $M_\odot$)
CCPC-z33-005	0.91	2595	2.1	3125	974	13.8
CCPC-z34-001	0.20	798	0.4	2563	976	11.3
CCPC-z34-002	1.25	6521	6.2	3875	667	8.0
CCPC-z35-001	3.39	4	< 0.1	813	214	0.2
CCPC-z36-001	7.83	50	0.2	1000	194	0.2
CCPC-z36-002	0.11	1564	0.7	1688	743	4.3
CCPC-z37-001	0.34	3725	2.1	3563	857	12.2

[c]As in Table 2.1, these overdensities were found near strong, extended fields within $\Delta z \pm 0.15$, which gave $\delta_m \sim 0$ values. The field galaxy numbers, when limited to the same aperture as the overdensity, show stronger mass overdensities. In the case of CCPC-z27-004, we merely had to limit Δz to ± 0.05, as it intersected another structure (CCPC-z29-004).

Note. — The estimated mass overdensities ($\delta_m = \delta_{gal}/b$) are listed for the CCPC structures, and a linear bias parameter of $b = 3$ is adopted from previous works (Steidel et al., 1998; Venemans et al., 2007). Names in boldface are the strongest candidates and can be found in Table 2.1. Based on the minimum and maximum RA/DEC values of galaxies within the overdense region, a rectangular volume encompassing the structure is listed in units of cMpc^3. If we assume that the volume containing δ_m will eventually collapse, we can calculate an estimated mass of the collapsed structure as $M = \rho_{crit,z} V(1+\delta_m)$. We can also infer a mass using the virial equation $M_{vir} = \frac{2\sigma^2}{G} R_{hms}$. The radius ($R_{hms}$) used is approximated by $1.25 * R_e$, where the effective radius contains 50% of the galaxy members of a protocluster in physical units. The virial mass estimator is almost universally larger than the expected, collapsed mass of the system we infer from δ_m. This suggests that these objects are not in virial equilibrium. Protoclusters with the highest probabilities (Tables 2.1, 2.2) typically have mass estimates within the expectations of simulated systems, while low probability systems (< 10%) have virial masses much larger than predicted at this epoch (Chiang et al., 2013).

2.1.1 Construction of the High Redshift Galaxy List

To identify structures of galaxies in the high redshift universe, a large list of galaxies was first compiled. Utilizing the NASA Extragalactic Database (NED), we assembled a list of ∼14,000 spectroscopic redshifts for galaxies at $z \gtrsim 2.7$, removing sources that were flagged as gravitationally lensed objects. Occasionally NED will not flag galaxies with redshifts determined photometrically as PHOT, as in the case of Rafelski et al. (2011), which had to be identified and removed manually. The redshift limit was chosen because it coincides with the effective onset of Lyman-Break Galaxies (LBGs), which were followed

up spectroscopically (Steidel et al., 2003). With the expectation that protoclusters would be observed as galaxy overdensities, we developed a simple algorithm that identifies candidate groups of galaxies within this high-redshift object list.

The spectroscopic galaxy list contains objects selected by a large range of criteria. Many galaxies in our list were part of spectroscopic surveys of *Hubble/Chandra/Spitzer* Deep fields. One of these (Balestra et al., 2010) in the GOODS-S field was surveyed with the *Very Large Telescope's* VIMOS grisms. The survey depth for galaxies with measured redshifts was to 24-25 AB magnitudes in *B*,*R*, and out to a redshift of $z < 3.5$ with a redshift uncertainty of $\sigma_z \sim 0.0008$. Another large contributor to our spectroscopic galaxy list were LBGs that were followed up spectroscopically. Steidel et al. (2003) used rest-frame UV photometry to identify more than 2000 photometric, $z \sim 3$ candidates to a limiting magnitude of $R_{AB} = 25.5$ in fields totaling 0.38 square degrees. Almost 1000 redshifts were measured using Keck telescopes. Other sources of redshifts come from narrow band imaging around high redshift AGN that were followed up spectroscopically (Venemans et al., 2007), or targets selected based on combined narrow and broad band images to a typical limiting magnitude of $R_{AB} = 25.5$ (Fynbo et al., 2003; Grove et al., 2009).

It follows naturally that high redshift galaxies that are collapsing to form a nascent structure should exist as a large concentration of objects within a small volume. From the high redshift galaxy list, groups of objects coincident on the sky were identified by searching within an angular search radius of 2 arcminutes and a redshift range $\Delta z < 0.03$ from a search galaxy. At the maximum redshift of our galaxy list ($z = 4.05$), the volume probed is a cylinder with radius 4 and depth of ±20 comoving Mpc from the estimated center of the distribution. This initial redshift depth was chosen in order not to miss candidate galaxy members, assuming that the protocluster dispersions could be $\sigma \gtrsim 0.01$ (600 km s^{-1} at $z \sim 4$).

The number of unique sources (those separated by at least 3″) within the search volume were counted. Prospective groups with three or more galaxies within the aperture were added to an initial group list for further analysis. The 'center' of a group in RA and DEC was determined iteratively by maximizing the number of galaxies within the sky aperture. Many prospective groups of galaxies that are in the final list of the Candidate Clusters and Protoclusters Catalog (CCPC) contained more than the minimum number of galaxies in this small volume. The heterogeneity of the sample (differing sky coverage and depths) does not permit us to make any estimate of group completeness. All we can state is that these candidates appear to be real, physical associations, and provide a lower limit to the number of galaxy members. Of the initial list of 14,000 galaxies, 603 sources are used in this work. Their references can be found in Table 2.3.

Diener et al. (2013) has shown that by linking associations of only $3 \leq N \leq 5$ member galaxies within a few comoving Mpc in the zCOSMOS field at $z > 1.8$, they could effectively identify nascent galaxy groups in the high redshift universe. Comparing the systems they discovered to mock galaxy light cones (Kitzbichler & White, 2007) from the Millennium simulation (Springel et al., 2005), they could track the fate of these primordial systems. Their method, when coupled with the complete spectroscopic sampling of the zCOSMOS field, is expected to identify $\sim 65\%$ of the protoclusters ($M_{z=0} > 10^{14}\ M_{\odot}$) within the survey volume (Diener et al., 2013). We adopt a similar, simple yet effective search criteria to identify initial regions of target protoclusters.

2.1.2 Candidate Protocluster Criteria

Protoclusters are the extended, collapsing manifestations of present day clusters. As the Universe is expanding, the volume surveyed for a fixed area on the sky will be dependent on the redshift of the source. To consistently measure protoclusters iso-volumetrically,

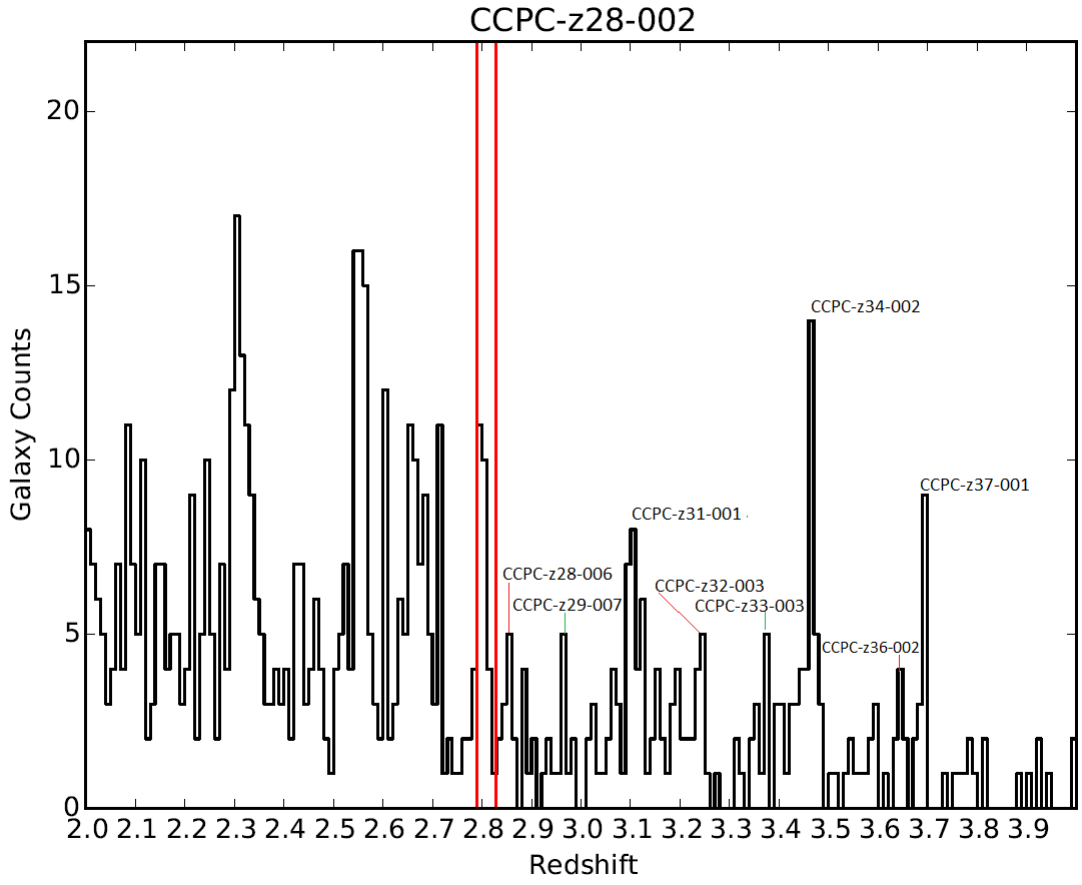

Figure 2.1: The number of galaxies as a function of redshift around CCPC-z28-002 at $z = 2.82$. The red, vertical lines correspond to a Δz that encompasses ± 20 cMpc along the line of sight from the mean redshift of the protocluster. Within this aperture of the CANDELS GOODS-S field, the partial peaks from other CCPC systems can be clearly identified. There are also strong peaks at redshifts $z \sim 2.3, 2.5, 2.6$, but are beyond the scope of this paper.

we have adopted the use of comoving Mpc (cMpc) throughout this study. This allows easy comparison to other observational studies that have also employed comoving volumes, as in Rigby et al. (2014). Analysis of large ΛCDM simulations (e.g. Millennium Simulation) also typically utilize cMpc when tracking the growth and evolution of protoclusters at various epochs (Chiang et al., 2013; Muldrew et al., 2015).

Generally, structures at high redshift have extended galaxy distributions of $R \gtrsim 10$ comoving Mpc from the highest density region. This seemingly large volume has both theoretical and empirical bases. Analysis of the Millennium Simulation by Chiang et al. (2013) found that the effective radii of the most massive protoclusters are typically $R_e \gtrsim 8$ cMpc at $z = 3$, collapsing to $R_e \gtrsim 1$ cMpc at $z = 0$. Muldrew et al. (2015) also analyzed the Millennium data for protoclusters, and found that 90% of the stellar mass of a $\gtrsim 10^{15}$ $h^{-1}\,M_{\odot}$ cluster is contained within a comoving radius of $\sim 20\,h^{-1}\,Mpc$ at $z = 3$.

Observationally, Rigby et al. (2014) investigated protocluster candidates around high redshift radio galaxies using *Herschel* . They found that galaxy overdensities peak at a radius $\sim 6 - 7$ cMpc, with distributions that flattened out at $R \sim 10$ cMpc (see their Fig 7). A protocluster identified by Prescott et al. (2008) at $z = 2.75$ extended 20×50 cMpc, while Lee et al. (2014) found an incredibly large structure of three protoclusters within $72 \times 72 \times 25$ cMpc^3 at $z \sim 3.8$. These examples illustrate the importance of wide search radii in identifying these extended, perhaps filamentary structures.

To refine our initial list of galaxy groups into a catalog of protoclusters, we chose a search radius of 20 comoving Mpc from the approximate center of each initial groups' distribution. This corresponds to $11.3'$ and $9.9'$ on the sky at $z = 2.75, 3.7$, respectively. Many redshift surveys used in this work do not extend to $R = 20$ cMpc, the expected size of the most massive protoclusters (Muldrew et al., 2015), so this should be treated as the maximum volume probed. In redshift space, $\Delta z \sim 0.019$ corresponds to ±20 cMpc

at $z = 2.75$, which increases to $\Delta z \sim 0.026$ at $z = 3.7$. More than 2/3 (435 out of 603 galaxies) of the redshifts have published uncertainties in NED, and the median uncertainty is ~0.0008, which corresponds to an uncertainty of ±0.9 comoving Mpc at $z \sim 3$ (Wright, 2006). This encompasses a 8% uncertainty in the velocity dispersion, assuming a typical $\sigma_z \sim 0.01$.

If within the search radius of 20 comoving Mpc on the sky (and associated length in Δz) there were 4 or more galaxies, which consist of a galaxy overdensity $\delta_{gal} > 0.25$ (see Section 2.3), this candidate structure was assigned to the Candidate Cluster and Protocluster Catalog (hereafter CCPC). Although four objects does not necessarily constitute a protocluster, this is at least a group of bright galaxies within a relatively small space, and might turn out to be a richer group with a more complete redshift census within the search volume.

Chiang et al. (2013) illustrate that a galaxy overdensity of $\delta_{gal} \sim 1 \pm 1$ is representative of low mass, $\gtrsim 10^{14}\ M_{\odot}$ clusters at $z = 0$ within the Millennium simulation (see their Fig 6). This implies that even modest overdensities at high redshift may represent protoclusters. We adopt $\delta_{gal} > 0.25$ as a working definition for candidate protoclusters, as these regions are more dense than the surrounding field while not excluding much of the overdensity distribution for the lowest mass protoclusters predicted by Chiang et al. (2013). Some of these objects may represent the tip of an iceberg, some may prove fictitious, while others may turn out to be filaments or void walls. At this juncture, all appear to be bona fide associations of galaxies at high redshift. Only two candidates have the minimum of 4 members. The median number of total members is 11 galaxies. Ten candidate protoclusters have over 20 members each.

Tables 2.1, 2.2 contain the full list of candidates. The naming convention we have employed is CCPC-z, followed by the first two digits of the redshift (e.g. $z = 2.9$ is indicated by 29-), and ends with a running index of objects in the respective redshift bin. Table 2.1

contains the sources that are at least 85% likely to collapse into a cluster at $z = 0$ based on the strength of their overdensity, while Table 2.2 contains other overdensities that are less strong. The basis of these probabilities will be explained in the following section. The source of the individual galaxy redshift measurements for each candidate cluster can be found in Table 2.3.

The search radius of $R = 20$ comoving Mpc is the effective size of the most massive protoclusters, with smaller systems being much less extended (Chiang et al., 2013; Muldrew et al., 2015). In the few instances in which the survey width encompasses $R \gtrsim 20$ cMpc (e.g. CANDELS GOODS-S), it is possible that we may be identifying two less massive protoclusters instead of a single, large system. However, by reducing our search radius to 10 comoving Mpc, only a few (or zero) galaxies were removed from many candidate protoclusters. Consequently, most objects appear to be more centrally concentrated than required by our search criterion, primarily because the fields-of-view (FOV) of most surveys are smaller than the maximum surface area we are probing. Upwards of 1/3 of a candidate's galaxies are found between $10 < R < 20$ comoving Mpc for 12 CCPC objects, which are marked by a '$*$' next to their galaxy number in Tables 2.1, 2.2. The implications of this are discussed in Section 5.4.

Each candidate protocluster has an associated sky position plot, showing the distribution of members in RA/DEC within 20 cMpc (black points within the outer red circle), as well as field galaxies (green $\times$'s) along the line of sight. The field galaxies serve as an illustration of the data footprint that contains the protocluster, with clear survey boundaries seen in many cases. The sky plots of the candidate structures illustrate the radial distributions of galaxies within a system. Some protoclusters are very extended (e.g. CCPC-z28-002 shown in Fig 4 in the published version), while others are strongly concentrated (CCPC-z29-002) within the inner red ring (corresponding to $R = 10$ cMpc), with the outer red circle

marking the search radius of 20 comoving Mpc for comparison. The differences in these systems, as noted earlier, is primarily dependent on survey width, and has little to do with the protocluster itself.

A $N(z)$ distribution (number of spectroscopic galaxies in redshift bins from $2 < z < 4$ within the search radius) for each protocluster is also located in the Appendix in Franck & McGaugh (2016b). The $N(z)$ plots have a large amount of variation as well, particularly with respect to differences in survey depth in our high redshift galaxy catalog. For example, CCPC-z28-005 shows an unmistakable peak of galaxies at the cluster redshift, with no other sources detected at other distances. In contrast, CCPC-z28-002, located in CANDELS GOODS-S field, contains a multitude of galaxies with peaks corresponding to 8 other CCPC members, as shown in Fig 2.1.

One distinct difference between this CCPC sample and the samples of Venemans et al. (2007); Galametz et al. (2012); Wylezalek et al. (2013); Rigby et al. (2014); Wylezalek et al. (2014) is that our protoclusters candidates were not initially targeted based on the presence of a high redshift radio galaxy (HzRG). Rather, we identify spatial coincidences among spectroscopically confirmed galaxies with $z > 2.74$. Some candidate structures in our sample contain radio galaxies, and our search methodology has recovered a number of previously identified protoclusters that were first identified using HzRGs as signposts (see Tables 2.1, 2.2). That we recover these previously identified structures is encouraging.

2.1.3 Candidate Protocluster Overdensities

To statistically measure the significance of these structures, we have estimated the galaxy overdensities (δ_{gal}) of each CCPC member. We computed the galaxy overdensity for each protocluster as $\delta_{gal} = (n_{proto}/n_{field}) - 1$, where the number density (n_{proto}) is the number of galaxies along the line of sight within $z \pm \sigma$ of the redshift of the structure. The

n_{field} density was determined by taking $z_{proto} \pm 0.15$, sans the region around the overdensity of $\pm\sigma$. A maximum field redshift range of $dz = 0.15$ (excluding the overdensity region) was chosen to adequately sample the field. Increasing the protocluster overdensity redshift range to encompass the full ±20 cMpc (instead of $z \pm \sigma$) typically increases the value of δ_{gal}, but at the expense of decreasing the field galaxies in some cases. When estimating the overdensity, we always assign galaxies to the field in cases of doubt. Some of these are likely members residing in the outskirts, so δ_{gal} is a conservative estimate of the overdensity. All listed uncertainties are Poissonian.

For the relatively modest overdensity of CCPC-z33-003, there are 6 galaxies found within the redshift space of $\Delta z = \sigma = 0.012$, giving a $n_{proto} = \frac{6}{2(0.012)}$ along the length ($dz = 0.024$) of the probed volume. Within the same aperture, there are 24 galaxies along the line of sight of $\Delta z = 0.284$, so the field length is $\Delta - 2(\sigma) = 0.284 - 0.024 = 0.26$. From this, a field density of $n_{field} = \frac{24}{0.26}$ is estimated, and $\delta_{gal} = 1.70$. All values of δ_{gal} are listed in Tables 2.1, 2.2. The field line of sight is limited to the Δz range where field galaxies are identified. For example, CCPC-z28-004 was previously identified using narrow-band filters by Venemans et al. (2007), so understandably has a narrower redshift distribution (as seen in its $N(z)$ plot Figure 4.4 in the published version) than many other systems. We identify 17 galaxies along the short length of Δz=0.028 outside of the overdensity region, which defines our field measurement. This serves as an upper limit on the field in that some of these 'field' galaxies may themselves be infalling future cluster members.

In some instances, $\delta_{gal} \sim 0$ as extended field density estimates roughly matched the candidate structure densities. This is most likely the result of the relatively small number of galaxies in these structures, combined with possible intervening (sub)structure. Limiting the field galaxies to the same sky area as the structure galaxies transforms these relatively small underdensities into overdensities. In the interesting case of CCPC-z27-004, the edge

of the field length intersects CCPC-z29-004. By limiting the field calculation to $\Delta z \pm 0.05$, the other structure is averted and a positive galaxy overdensity becomes apparent.

It should be noted that choosing a different Δz can also change the value of δ_{gal}. This highlights the difficulty in defining a field number density along the line of sight, as intervening (sub)structures can boost the field counts significantly. We have purposefully avoided fine-tuning the galaxy oversensities in a way to produce larger overdensities. Most importantly, we urge caution in drawing conclusions from these galaxy overdensities, as the definition of the field, coupled with the selection effects of the sample, can alter the significance of the structure.

Seven structures had 2 or fewer field counts in calculating their overdensity, with CCPC-z27-002 having none at all. In these instances, we have manually added 7 field galaxies to their number density to provide a lower limit to the actual δ_{gal}. Some of these candidate protoclusters, like CCPC-z30-001 and CCPC-z31-005, have incredibly large calculated galaxy overdensities (δ_{gal} >10). These are the result of an overdensity coupled with a very weak field density measurement along the line of sight as a selection effect of narrowband imaging, which targets galaxies at a specific redshift (Venemans et al., 2007). As Muldrew et al. (2015) point out, this is one of the dangers in calculating galaxy overdensities using narrowband imaging, as field counts can be under-represented.

Chiang et al. (2013) calculated the expected galaxy overdensities for protoclusters identified within the Millennium Simulation, combined with the semi-analytic model of Guo et al. (2011). For the lowest mass protoclusters ($M_{z=0} < 3 \times 10^{14}\ M_{\odot}$) in the redshift range of $2 < z < 5$, in cubes with sides of 25 cMpc, a $\delta_{gal} \sim 1 \pm 1$ is typical for galaxies with a $SFR > 1\ M_{\odot}$. This overdensity increases to $\delta_{gal} \sim 3$ for the most massive ($M_{z=0} > 10^{15}\ M_{\odot}$) cluster progenitors. Table 5 in Chiang et al. (2013) contains a list of more than 20 protoclusters found in the literature, with measured δ_{gal} ranging from $0.7^{0.8}_{-0.6}$ (Venemans et al.,

2007) to 16±7 (Toshikawa et al., 2014). The median value of galaxy overdensity within the CCPC is $\delta_{gal} \sim 2$, which agrees well with values from the literature and theoretical predictions (Chiang et al., 2013).

In Tables 2.1, 2.2 we provide a conservative estimate of the probability that a given candidate structure will collapse into a cluster at $z = 0$ based on its δ_{gal} value. These estimates are taken from the z = 3 case in Figure 8 of Chiang et al. (2013), which plots the fraction of overdensities which will form a $M > 10^{14}\ M_{\odot}$ halo at the present day. We made the conservative assumptions that our protoclusters onservative assumptions that the galaxies identified as cluster members are the most biased tracers of mass (the brightest galaxies), the overdensity volumes are [15 cMpc]^3, and that none of the field galaxies used to calculate our overdensity values will become part of the cluster. If some of our member galaxies are less biased tracers (e.g. $M_* < 10^{10}\ M_{\odot}$), the probability can increase by more than 20% for a given δ_{gal} (Chiang et al., 2013). If the box volume is increased to [25 cMpc]^3, a $\delta_{gal} = 0.86$ has a 50% of collapsing into a cluster, while for the smaller box volume of [15 cMpc]^3, a $\delta_{gal} = 3.43$ is required (Chiang et al., 2013). Systems not fated to be clusters can end up as groups of galaxies, or may simply be false positives.

We list our candidate structures in Tables 2.1 and 2.2. Table 2.1 lists the most probable protoclusters, containing 12 candidates that have at least an 85% chance of collapsing into a cluster by $z = 0$ according to Chiang et al. (2013). Lower probability structures are listed in Table 2. Six out of the twelve protoclusters in Table 2.1 have been previously identified, while only 4 of 31 candidates in Table 2.2 were found in the literature. Typically, overdensities of $\delta_{gal} > 7$ are expected to be protoclusters at this redshift.

It is important to remember that these probabilities are based on analysis of the Millennium simulation by Chiang et al. (2013). There is no guarantee that this represents the Universe in which we reside. Therefore, these probabilities should be treated as a mere

guide.

2.1.4 Mass Estimates

Using the positions of galaxies within each CCPC system and their respective redshift dispersions σ, we have provided a crude mass estimate for each structure in Table 2.4. We estimate the mean harmonic separation as $R_{hms} \approx 1.25R_e$, where R_e is the radius from the approximate center of the galaxy distribution that contains half of all member galaxies in the system. The R_{hms} value for each CCPC protocluster is recorded in Table 2.4. Another variation of an R_{hms} estimator using Equation 2 from Carlberg et al. (1996) with weights set to unity provides values roughly a factor of 2 larger, so the R_e approximation is conservative. The virial mass is then estimated by

$$M = \frac{2\sigma^2}{G} R_{hms}. \tag{2.1}$$

There is no guarantee that these systems are virialized, symmetric structures, and so we stress that these mass estimates are merely indicative.

Although some systems have seemingly reasonable masses for a present-day cluster ($\gtrsim 10^{14}$-10^{15} $M_\odot$), these estimates are much larger than theoretical predictions for a $z = 3$ protocluster. Figure 2 in Chiang et al. (2013) tracks the halo mass of protoclusters as a function of redshift within the Millennium Simulation. At $z = 3$, the main halo of a protocluster of mass $\sim 3 \times 10^{13}$ $M_\odot$ will grow into a 10^{15} $M_\odot$ cluster at $z = 0$. Smaller clusters ($\gtrsim 10^{14}$ $M_\odot$ at $z = 0$) will be in the range $\sim 5 \times 10^{12}$-10^{13} $M_\odot$ at $z = 3$. Almost all of the mass estimates for our protoclusters exceed 10^{14} $M_\odot$, with 13 structures above $M \gtrsim 10^{15}$ $M_\odot$. Either these CCPC systems are orders of magnitude larger than ΛCDM predicts at this epoch, or more probably, the crude mass estimate we have employed is a

poor representation of the physical nature of these early structures. Only 3 systems (CCPC-z32-002, CCPC-z35-001, CCPC-z36-001) have reasonable mass estimates of $< 3 \times 10^{13}$ $M_{\odot}$, and each has 5 or fewer member galaxies. Chiang et al. (2013) find only 2% of almost 3000 clusters have $M_{z=0} > 10^{15}$ $M_{\odot}$ within the Millennium Simulation's comoving box (sides of 500 h^{-1} Mpc). With the significant number of $M^{vir}_{z\sim3} > 10^{15}$ $M_{\odot}$ protoclusters in the CCPC, this suggests that the virial masses are overestimates, or massive clusters emerge earlier than anticipated.

However, comparing the results of simulations to actual observables can be problematic. For instance, Chiang et al. (2013) report that a $z = 0$, $M > 10^{15}$ $M_{\odot}$ protocluster in the Millennium Simulation would have a velocity dispersion along the line of sight of 400 ± 60 km s^{-1} at $z = 3$. Many members in the CCPC are well above this velocity dispersion, with some at $\gtrsim 1000$ km s^{-1}. Converting the radius from Chiang et al. (2013) for the progenitor of a 10^{15} $M_{\odot}$ cluster from cMpc into physical units, and setting $\sigma_v = 400$ km s^{-1}, Equation 2 provides a mass estimate of 4×10^{14} $M_{\odot}$. This 'observable' mass is an order of magnitude larger than the mass computed by the simulation for the same protocluster ($\sim 3 \times 10^{13}$ $M_{\odot}$).

As an illustrative exercise, doubling the velocity dispersion to $\sigma_v = 800$ km s^{-1} (more typical of CCPC systems), we recover a comparable mass estimate ($\sim 10^{15} M_{\odot}$) to those listed in Table 2.4, as expected. This increased velocity dispersion is not unique to our analysis. As pointed out in Chiang et al. (2015), other known protoclusters have line of sight velocities of $\sigma_v = 900$ km s^{-1} at $z \sim 3$. The larger line of sight dispersions of observed σ_z values than the theoretical predictions boosts our structure mass estimates by orders of magnitude over their simulated sibling protoclusters. It is therefore necessary to take great care to measure apples with apples when comparing observations to simulations.

Taken at face value, the masses we compute are problematic for ΛCDM. Mortonson et al. (2011) predict that no clusters should exist with $M > 6 \times 10^{14}$ h^{-1} $M_{\odot}$ at $z = 2$, and

they should be smaller still at higher redshift. The CCPC contains 14 clusters that exceed this virial mass, all at $z > 2.7$. Galaxies at high redshift also appear to have a similar issue, in that the number density of massive halos is $\sim 10^2$ larger than predicted by ΛCDM (Steinhardt et al., 2016). We note in passing that clusters this large are predicted to exist at redshifts of 2-3 by MOND (Sanders, 1998; Nusser, 2002; McGaugh, 2004; Katz et al., 2013; McGaugh, 2015).

Our mass estimates assume that the objects we identify as cluster candidates are virialized systems. This may not yet be the case at these redshifts, as they are still in the process of collapse. We therefore urge great caution in interpreting these results.

Another estimator of protocluster mass is the use of galaxy overdensities (δ_{gal}) as biased tracers of mass overdensity (δ_m), as utilized by Steidel et al. (1998); Venemans et al. (2007). The equation adopted is

$$M = \rho_{crit,z} V (1 + \delta_m) \tag{2.2}$$

where $\delta_m = \delta_{gal}/b$, with b existing as a bias parameter, with values assumed to be $\sim 3 - 6$ (Steidel et al., 1998; Venemans et al., 2007). Take CCPC-z33-003 as an example again, for which we computed a δ_{gal}= 1.59 in Section 2.3. At $z \sim 3$, the comoving critical density is $\rho_{crit} = 4.2 \times 10^{10}\ M_\odot$ cMpc^{-3} within our cosmology. Assuming a bias parameter of $b = 3$ and a computed volume of approximately 3300 cMpc^3, the implied $z = 0$ mass that this structure may collapse to is $2.2 \times 10^{14}\ M_\odot$. The virial mass estimate for this system was $14.9 \times 10^{14}\ M_\odot$. For the entire sample of the CCPC, the median mass estimate is 8×10^{13} $M_\odot$.

There are a number of systematic uncertainties within this calculation. The bias parameter b is typically assumed to be in the range 3-6, although this value has never been directly measured. As discussed previously, the value of δ_{gal} (and therefore δ_m) is strongly dependent on the definition of field galaxies, or lack thereof. One of the largest contributors to

the mass estimate is the volume V that is expected to collapse by $z = 0$, which is calculated using the minimum and maximum RA/DEC values in a rectangular box along the line of sight $\Delta z = \sigma$. If the width of the field of observations is small, the calculated volume of a collapsing protocluster will be a lower limit. For instance, recomputing the overdensity volume of $\Delta z_{proto} = 20$ cMpc and not σ, and a field width of $\Delta z = 0.05$ instead of the ± 0.15 adopted in this work, the median δ_{gal} value is unchanged from this work, although with greater scatter. However, the increase in volume of the candidate structures boost the overdensity mass from $8 \times 10^{13}\ M_{\odot}$ to $4 \times 10^{14}\ M_{\odot}$.

There appears to be no correlation between the virial mass estimates and the masses inferred from the overdensities. The two estimates are physically distinct, as the former ostensibly measures the current system/halo mass, while the latter is an estimated collapsed mass. Yet some link should exist between them, as massive main halos generally correlate with massive $z = 0$ clusters within simulations (Chiang et al., 2013; Muldrew et al., 2015). The fact that there is no connection implies that systematic uncertainties dominate, and that one or both of these methods are inherently flawed. As a result, these mass estimates should be treated with caution.

It should be noted that the systems identified as having the highest probability (85% or larger) of collapsing into a cluster in Table 2.1 based on the strength of their overdensity, typically have reasonable mass estimates (both virial and δ_m) of $\sim 10^{15}\ M_{\odot}$ or less in Table 2.4. A notable exception is CCPC-z30-003, but this is the richest structure in the CCPC and was previously identified by Steidel et al. (1998). Encouragingly, the candidates with lowest probability ($< 10\%$) of collapsing into a cluster often have the most anomalously large virial mass estimates of $> \times 10^{15}\ M_{\odot}$.

2.2 Discussion

2.2.1 Confirmation tests of structure

To confirm that these CCPC targets are indeed structures, we have used our initial list of ~14,000 $z > 2.75$ objects as a benchmark sample. We searched around each galaxy (including galaxies identified within CCPC systems), and plotted any companions that were within the same comoving volume as the candidate protoclusters (search radius of 20 cMpc, Δz corresponding to ±20 cMpc). Many of these galaxies had no spectroscopic companions within the volume. We plotted the mean distribution of these galaxies as a function of comoving radius from the central source (in many cases the sole source) in Fig 3.4 as the red line in the top panel, and the black line in the bottom plot.

For a simple metric of the distribution of galaxies within protoclusters at $z > 2.75$, we have added the distribution of each individual CCPC system in the top panel of Fig 3.4 to the distribution of all galaxies. The distributions can be seen in the striking variability in concentration of these early mass overdensities. Some structures are fully contained within a radius of 5 comoving Mpc, while others do not even have 40% of their total members by $R \gtrsim 10$ cMpc. In our analysis, these variations do not appear to be dependent of the redshift of the system, as there are both extended and compact structures at the redshift extrema of the catalog.

As a further test, we computed the mean fraction of all CCPC members within the circular 1 Mpc annuli from the central galaxy out to 20 comoving Mpc in the bottom panel of Fig 3.4. Separating the candidates into three redshift bins $2.75 < z < 3.00$ (19 objects, blue line in Fig 3.4), $3.00 < z < 3.25$ (12 objects, green line), and $3.25 < z < 3.7$ (12 objects, red line) shows there is no significant distribution difference between the bins, despite the > 0.5 Gyr time difference between $z \sim 3.7 \rightarrow 2.75$. It appears that although the

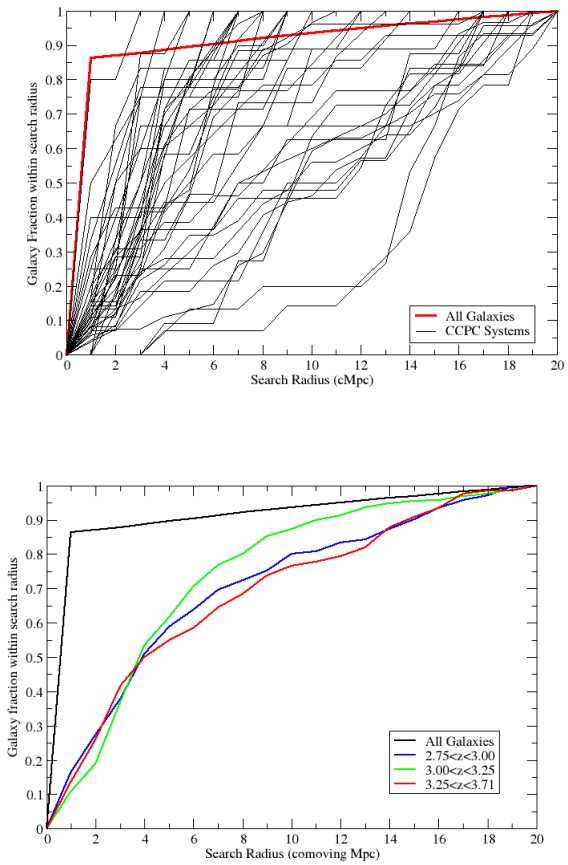

Figure 2.2: Top : The fractional number of galaxies as a function of radius in individual CCPC structures. There is a large spatial variability, as some objects are very compact while others are diffuse, extended systems. The red line represents the mean distribution for 14000 galaxies and any companions found within $R < 20$ comoving Mpc (see Section 5.4). The candidate protoclusters (Black Lines) clearly have more neighbors than is typical in the field, as well as different distributions. *Bottom* : Mean distribution of the fraction of total members in the CCPC as a function of search radius in comoving units. The black line distribution represents the mean All Galaxies sample. The CCPC catalog is divided into three redshift bins, which show no significant difference in the mean distributions, but there is significant scatter between CCPC members within each bin (as seen in the top figure). It may appear that some evolution is taking place from $2.75 < z < 3.00$ (blue line) to $3.00 < z < 3.25$ (green line), in that the more distant systems are more centrally concentrated on average. However, the trend is reversed when looking at the mean distributions of $3.00 < z < 3.25$ to $3.25 < z < 3.71$ (red line).

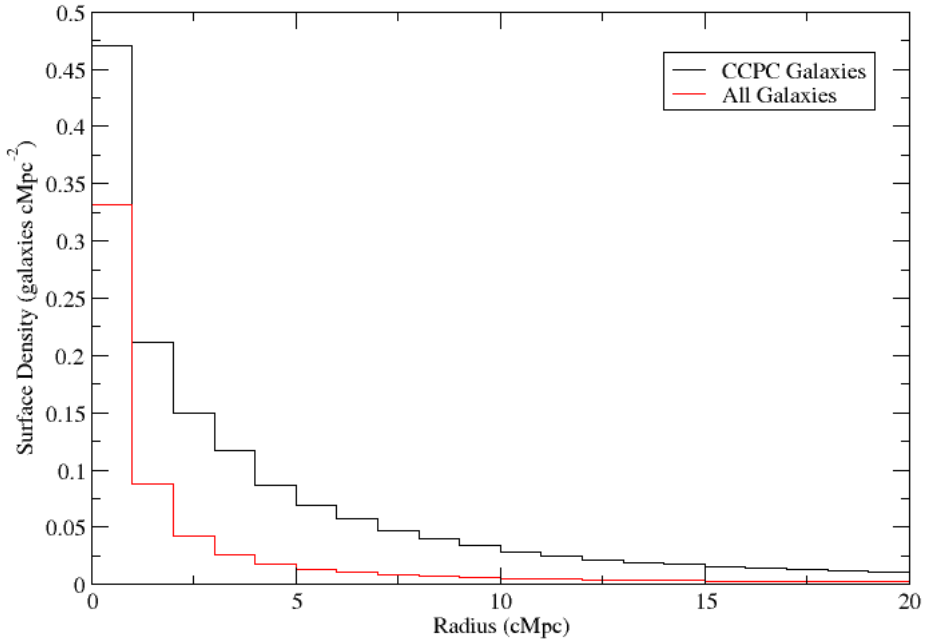

Figure 2.3: The mean surface density profile of All Galaxies (14,000 spectroscopic sources with $z > 2.75$) that were used in the initial list to identify structures is shown in red, with the mean density profile of the CCPC sources shown in black. Most galaxies in the All Galaxies list have no companions within 20 cMpc, although galaxies within the CCPC structures are included in All Galaxies. The CCPC surface density profile has more galaxies within the inner R=1 cMpc bin, and has a more gradual decrease in density than the distribution of galaxies in general, as expected for galaxy overdensities.

mean distribution generally does not vary as a function of redshift, individual protoclusters have a wide range of spatial differences for their members. If the total sample is limited to members within $R \sim 10$ comoving Mpc, the distributions still show no significant difference as a function of redshift, but have considerably less scatter, as expected. This is a possible indication that in general, protoclusters at these epochs have similar assembly histories. There also does not seem to be a significant trend in the number of galaxy members as the redshift decreases in the CCPC, with a similar median galaxy membership (13) occurring in structures between $3.00 < z < 3.25$ when compared to the median over all candidates $2.74 < z < 3.71$ (11 galaxies). Median membership in the bins $2.74 < z < 3.00$ and $3.25 < z < 3.71$ are 10 and 9 galaxies, respectively.

Fig 3.3 illustrates the difference in surface density of the structures in the CCPC (black line) compared with all of the galaxies in the initial list of ~14,000 spectroscopic galaxies (red line). The red distribution of galaxies is consistent with a single object with no other companions found within $R_{search} = 1$ cMpc ($\Sigma = \frac{1\,gal}{\pi(1\,cMpc)^2} = 0.33$ cMpc^{-2} in the first bin), and rapidly decreasing as the radius increases. CCPC galaxies have a larger central concentration (0.47 cMpc^{-2}) within 1 cMpc radius, and a more gradual decrease in surface density. A two-sample Kolmogrov-Smirnov test gives a value of $KS = 0.69$, suggesting that these cumulative densities are distinct distributions. In $log - log$ space, the slope of all galaxies is steeper (-1.65) compared to the -1.5 slope for CCPC galaxies. Interestingly, outside of $R \sim 5$ cMpc, the slopes are both ~ -1.5, suggesting a possible characteristic radius at which the galaxy overdensity and the field begin to merge.

We note that Figures 3.4, 3.3 should be only used as tests of structure in comparison to the field galaxy distributions, and should not interpreted as the expected radial profile of protocluster systems. These data were taken from a number of different surveys of various widths and selection techniques. Not all protoclusters are centered within a survey's FOV,

nor do they necessarily encompass the full search radius of $R = 20$ cMpc, as can be seen in the sky plot of each CCPC member within the Appendix. Thus, the cumulative distribution will be strongly dependent on the survey characteristics and the protoclusters position within the survey's width.

A further test that a candidate protocluster's galaxies are not merely coincident on the sky, but rather constitute physically associated structures can be found by plotting the number of galaxies at each redshift along the same line of sight as the structure in a $N(z)$ plot. Protoclusters, as an overdensity of galaxies, should be visible as a peak in counts within a small redshift bin. The Appendix in Franck & McGaugh (2016b) contains a $N(z)$ plot for each CCPC member in the right panel. For many of these protocluster systems, the overdensity peak is unmistakable from the background distribution (like CCPC-z30-003), or there are no other intervening sources along the line of sight, as in CCPC-z28-005. In deeper surveys, the distinction between structure and field galaxies becomes harder to identify, as peaks can be smaller relative to the continuum. Fig 2.1 shows the $N(z)$ plot for CCPC-z28-002 in the CANDELS GOODS-S survey. The number of galaxies in each redshift bin is larger than in other pointings, presumably because this field has been surveyed more deeply.

As the CANDELS GOODS-S field is essentially a pencil-beam survey between the redshift values $2.74 < z < 3.71$ (roughly $50 \times 50 \times 900$ cMpc in our assumed cosmology), we can compare the number and richness of our candidate structures within this pointing to the expectations of stochastic alignments of galaxies arising from Poissonian fluctuations in a smooth density field. Sheth (2001) computed a toy-model of galaxy clustering with a density distribution of galaxies n, in which the expected number of systems N with M galaxy members and linking length l can be expected for a sample size of N_G objects along

the line of sight via

$$N_M = N_g e^{-2nl}[1 - e^{-nl}]^{M-1}. \tag{2.3}$$

If the Poissonian expectation of structure is lower than the recovered number of structures, this is an indication that physical structures exist in these data, are not mere chance alignments of galaxies.

To utilize this simple model, the linking length l must be computed. Using the auto-correlation function

$$\xi(l) = \frac{N_{DD}}{N_{DR}} \frac{n_R}{n_D} - 1 \tag{2.4}$$

one measures the excess Data-Data pairs of galaxies N_{DD} over the number of Data-Random pairs (N_{DR}) of galaxies in the interval $l \pm \Delta l$ (Davis & Peebles, 1983). Within the GOODS-S field, ξ was found to be $\gtrsim 1$ at $l < 1.78$ cMpc, effectively measuring the separation l at which clustering is strongest.

A one-dimensional, Friends-of-Friends (Huchra & Geller, 1982; Press & Davis, 1982) algorithm was constructed to identify structures of galaxies separated by a linking length $l < 1.78$ cMpc within the CANDELS GOODS-S catalog. We then compared the number of structures with M members (N_M) with the expectation value using the toy model of Sheth (2001). Within the data set, the number of systems with $4 < M < 30$ galaxies identified using the Friends-of-Friends algorithm always exceeded the expected, Poissonian value in our analysis. For instance, there are 7 structures identified with $M \geq 9$ galaxies along the line of sight, where only one such structure should exist by chance. As a check of the Sheth (2001) model, we ran a Monte Carlo simulation of 500 randomized fields, identifying galaxy systems using our Friends-of-Friends algorithm. The Monte Carlo results provides a slightly smaller estimate of the number of structures at almost all values of M, but are generally similar. For example, for $M = 5$, the Monte Carlo predicts 7.6 systems to be

identified, where $N_M = 8.6$ in the toy model. At $M = 9$, where $N_M \sim 1$, the Monte Carlo predicts only 0.1 structures to exist, on average.

It is important to emphasize that this analysis is specific to this single field, as the parameters l, n, and N_G are unique to this survey. To apply the results of the toy model to a another field (or the CCPC as a whole) with a different selection function would not be a physically-motivated comparison.

Based on the combination of (1) positive galaxy overdensities (δ_{gal} listed in Tables 2.1, 2.2) within the protocluster volume, (2) radial distribution profiles of CCPC structures that look significantly different from the stochastic positions of field galaxies (Figs 3.4, 3.3), (3) redshift distributions of galaxies along the line of sight that show up as strong peaks in the $N(z)$ plots in the Appendix of Franck & McGaugh (2016b), (4) the excess number of galaxy groups over Poissonian expectations in the deep CANDELS GOODS-S survey, and (5) that we recover a number of previously identified, rare protoclusters in the literature suggests that the galaxy associations identified in the CCPC are strong candidate protoclusters.

2.2.2 Extended Protoclusters

As mentioned in Section 3.1, there appears to be two populations of structures within the CCPC. While many protocluster candidates add few members outside a 10 comoving Mpc radius, 12 CCPCs have a $1/3$ or more of member galaxies at $10 < R < 20$ comoving Mpc from the center of the distribution. In Tables 2.1, 2.2, these extended sources are indicated by a '$*$' next to their galaxy counts. By examining the galaxy position plots of each candidate protocluster in the Appendix of Franck & McGaugh (2016b), the differences between the extended and centrally concentrated structures are easily distinguishable.

These extended objects (CCPC-z27-004, CCPC-z28-002, CCPC-z28-006, CCPC-z29-003, CCPC-z29-007, CCPC-z30-003, CCPC-z31-001, CCPC-z32-003, CCPC-z33-003, CCPC-

z34-002, CCPC-z36-002, CCPC-z37-001) tend to have larger numbers of members overall, and the richest candidate (CCPC-z30-003) is among their number. They span almost the entire range of redshifts in the sample. This may be an indication that in this epoch of the universe, protoclusters are in various stages of assembly, such that some structures are centrally concentrated while others are condensing more slowly. Protoclusters identified by Muldrew et al. (2015) within the Millennium Simulation are consistent with there being a variety of evolutionary stages that exist for structures at these redshifts, with a minority ($\sim 10\%$) of protoclusters having their mass concentrated in a single, dominant halo at $z > 2$.

Chiang et al. (2013) mapped the growth of protoclusters from $0 < z < 5$ within the Millennium Simulation to estimate observable features in future surveys. In the context of this work, the effective radius (R_e), within which $\sim 65\%$ of the mass of bound halos are found, is particularly interesting. Regardless of redshift, the greater the mass of the system at $z = 0$, the larger R_e was measured to be, with the expected result that at redshifts of $z \sim 5$, the radius was largest (see their Fig 2). For instance, at $z \sim 3$, the effective radius is > 8 comoving Mpc for a Coma-like $M_{(z=0)} = 10^{15}\ M_\odot$ halo, while it is $R_e < 5$ cMpc for a halo $M_{(z=0)} < 3 \times 10^{14}\ M_\odot$. It is conceivable that these extended sources could also be more massive systems. As the extended sources also typically have the largest number of galaxies within our catalog, this is a plausible explanation.

It is important to note that a significant number of these extended objects are in the extensively studied CANDELS (PI: Faber, Ferguson) ~170 arcmin^2 GOODS-S field, and two others were previously identified as protocluster candidates (Bond et al., 2010; Uchimoto et al., 2012). As these fields have greater survey depth and width, these extended objects most likely constitute a selection effect and are not structurally different than the other CCPC objects. With a wider, deeper search around the non-extended sources, it is possible that their protocluster galaxy counts could similarly grow. Of the five CCPC mem-

bers that were found in the Hubble Ultra Deep Field (Beckwith et al., 2006), only one of these structures have significantly extended populations despite the impressive survey depth, while all of the CANDELS protoclusters are extended. As the Ultra Deep Field has only a $3.3' \times 3.3'$, the survey width is insufficient to detect such diffuse structures. The search radius corresponding to 20 cMpc at $z = 3.7$ is $10'$.

2.3 Summary

We present a catalog of 43 candidate protoclusters in the redshift range of $2.74 < z < 3.71$. These structures were identified using published position and spectroscopic measurements within a comoving search radius of 20 cMpc. Prospective structures were initially identified by flagging groups of three or more galaxies with a $\Delta z < 0.03$ and within a 2 *arcminutes* radius of each other. This list was later refined by requiring at least 4 spectroscopic sources within a search radius of 20 comoving Mpc, and a galaxy overdensity of $\delta_{gal} > 0.25$. The median number of galaxies in each candidate protocluster is 11, while the maximum number of galaxies is 54. There appears to be little evolution in galaxy numbers as a function of redshift. As a statistical measure of the significance of these structures, we calculated a galaxy overdensity δ_{gal} for each CCPC member following the examples of Steidel et al. (1998) and Venemans et al. (2002). The median δ_{gal} value is ~ 2, which is comparable to the overdensities of protoclusters within simulations and observationally. Twelve of these have high probability (85% or larger) of collapsing into a cluster at $z = 0$ based on their overdensity compared to analysis of Millennium run data (Chiang et al., 2013).

There are a number of tests that suggest that the CCPC systems identified by our simple algorithm are coherent structures. The protoclusters exist as overdensity peaks, both in

δ_{gal} and visually above the continuum of galaxies in the N(z) plots shown in the Appendix. They follow a distinctly different radial distribution than the mean distribution of the 14000 galaxies in which the CCPC was drawn from, as shown in Figs 3.4, 3.3, many of which do not have any companion galaxies within 20 cMpc of the central source. Furthermore, we have recovered 10 previously identified candidate protoclusters from the literature. We have also found that in the CANDELS GOODS-S field, there is significantly more structure (excess of 6 systems of 9 or more galaxies) than can be expected from Poissonian fluctuations.

The mean radial distribution of members for each CCPC structure follow a similar trend, independent of redshift. However, individual CCPC structures have a wide range of distributions, with some protoclusters completely contained within $R < 10$ comoving Mpc while others add more than a dozen of their constituent galaxies in the range of $10 < R < 20$ comoving Mpc. These variations can be primarily attributed to differences in survey depth and volume for the various regions of our galaxy list. This distribution difference could be also an indication of the various stages of assembly protoclusters are in at this epoch of the Universe, in conjunction with differences in total halo mass, as more massive protoclusters are expected to have larger radii. A combination of all three seems likely.

For each system, we computed a characteristic radius (R_{hms}) and a rough virial mass estimate based on its size and velocity dispersion σ_z. The mass estimates assume a virialized distribution, which may not apply at this epoch. As a result, these crude protocluster mass estimates are up to 10^2 times larger than those predicted in analysis of the Millennium Simulation by Chiang et al. (2013). In addition, by assuming a linear bias parameter of $b = 3$, we can calculate a mass overdensity ($\delta_m = \delta_{gal}/b$) within the volume of the structure. This can then be transformed into a mass estimate using $M = \rho_{crit,z} V(1 + \delta_m)$ (Steidel et al., 1998; Venemans et al., 2002). These are estimates of the collapsed mass of the system, and

have a median value of 8×10^{13} $M_{\odot}$ in the CCPC. These mass estimates are uncorrelated, and caution should be exercised in interpreting these results.

This work significantly increases the number of spectroscopically confirmed, high redshift protoclusters known. To our knowledge, it represents the largest catalog of such sources to date.

Chapter 3

The Candidate Cluster and Protocluster Catalog II

Protoclusters are the diffuse, extended building blocks that will collapse into a galaxy cluster at the current epoch. Structures at high redshift have power as tests of cosmology, as the maximum mass that can collapse and virialize at a given epoch depends on the mass density ($\Omega_{\rm m}$), the power spectrum (σ_8), and the dark energy equation of state (w) (Bahcall & Fan, 1998; Vikhlinin et al., 2009; Mortonson et al., 2011). These primeval systems are also unique laboratories in which we can observe galaxies in the Universe assemble into larger collections and evolve in dense environments. Galaxy formation and hierarchical accretion scenarios can be examined by compiling large numbers of high redshift galaxies at various times (Cooke et al., 2014; Wylezalek et al., 2014). Protocluster galaxies have been found to have enhanced mass assembly (Casey et al., 2015) and evolution (Hatch et al., 2011). Star formation rates of 10^4 $M_\odot$ yr^{-1} for a protocluster (Clements et al., 2014) are in significant excess of hydrodynamic simulation expectations (Granato et al., 2015; Contini et al., 2016). As a result, the search for high redshift clusters and protoclusters within the astrophysical community has been rapidly expanded in the last few years.

Less than two decades ago, protoclusters were relatively unknown until rich, Lyman-Break Galaxy (LBG) overdensities were discovered by Steidel et al. (1998) at $z \sim 3$ and *Ly*α emitters (LAEs) at $z \sim 4$ (Venemans et al., 2002). Until recently, few systematic surveys for $z \gtrsim 2$ systems existed. With the advent of Clusters Around Radio Loud-AGN (CARLA) *Spitzer* survey (Wylezalek et al., 2013), the Search for Protoclusters with *Herschel* (Rigby et al., 2014), data mining the GOODS-N/GOODS-S fields (Kang & Im, 2009; Salimbeni et al., 2009; Kang & Im, 2015), and finding structures within the COSMOS/UltraVISTA field (Diener et al., 2013; Chiang et al., 2014), the number of photometrically/spectroscopically identified candidate protoclusters is now in the hundreds. Catalogs of pure spectroscopically identified protoclusters in the high redshift universe, as in Venemans et al. (2007), typically have fewer than 10 such candidates, however. Most of these previously identified structures are found within $z \sim 3 \pm 1$ (Venemans et al., 2007; Wylezalek et al., 2014; Franck & McGaugh, 2016b), with a few at $z \sim 6$ (Utsumi et al., 2010; Toshikawa et al., 2014). These objects were identified with varying instruments, selection techniques, and fields-of-view (FOV) that rarely encompass the entire system.

It was this understanding that prompted a systematic search of archival data in a simple manner to identify high redshift ($z > 2$) clusters and protoclusters. This work builds on the original harvest of 43 members in the Candidate Cluster and Protocluster Catalog (CCPC) between $2.74 < z < 3.71$ (hereafter known as CCPC1 Franck & McGaugh, 2016b). These structures were identified as galaxy overdensities in fixed comoving volumes. In this second paper (hereafter CCPC2), we have extended the redshift range of our search to $2.00 < z < 7.00$ using a similar search method. We eschew the common search technique of using High Redshift Radio Galaxies (HzRGs) as biased tracers of structure (Venemans et al., 2002; Wylezalek et al., 2013; Rigby et al., 2014), but we have recovered overdensities identified in this way. Presented in this work are 173 additional protocluster candidates, 23 of which

have been found in the literature.

We have organized CCPC2 as follows: Section 3.1 discusses the archival data from which these structures are gleaned, the algorithm that identifies structures, calculation of the overdensity of each system, and the explanation of the two mass estimators. In Section 5.4, we explore the nature of these overdensities, their significance as structures, compare their properties to simulations, and discuss three supercluster candidates.

We assume a standard cosmology, adopting $H_0 = 70$ km s^{-1} Mpc^{-1}, a matter density of $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. The Universe is 3.2 Gyr old at the redshift of $z = 2.0$, and has a comoving angular scale of 1.51 Mpc arcmin^{-1}. At $z = 7$, the corresponding age is 0.75 Gyr, while the angular scale is 2.51 comoving Mpc arcmin^{-1} (Wright, 2006).

3.1 The Candidate Cluster and Protocluster Catalog (CCPC) II

3.1.1 Data

In CCPC1, a NASA Extragalactic Database (NED) search was used to compile a list of ∼14,000 galaxies with spectroscopic redshifts between $2.74 < z < 3.71$. Many of these galaxies were found in Hubble Deep Fields, and spectroscopic follow-up of Lyman Break Galaxies (LBGs) Steidel et al. (2003). In this work, we have expanded the redshift range to $2.00 < z < 10.23$, which constitutes a galaxy list of 47,000 objects. The upper limit of $z = 10.23$ is an extreme example; only 9 galaxies have $z > 7.5$ in our list. The NED database holds published redshift uncertainties for 813 of the ∼ 1400 candidate protocluster galaxies, with a mean value of $\sigma = 0.001$. At a redshift of $z = 3$, this represents a comoving distance uncertainty of 1.0 Mpc (Wright, 2006).

Nearly 200 sources of spectroscopic measurements were used to identify candidate clus-

Table 3.1. Candidate Cluster and Protocluster Catalog (CCPC) II - Best Candidates

Candidate Name	RA (deg)	DEC (deg)	Redshift (z_{avg})	σ_z	N	$N_{R\leq10}$ cMpc	Overdensity (δ_{gal})	Cluster Probability (%)	Q	Recovered Reference
CCPC-z20-002	222.20	8.92	2.002	0.008	11	7	9.38 ± 5.34	100.0	1	8
CCPC-z20-003	29.62	-25.05	2.018	0.004	10	10	19.43 ± 13.06	100.0	1	1,2
CCPC-z20-009	150.04	2.21	2.098	0.005	10	4	13.15 ± 6.54	100.0	1	4,5
CCPC-z21-004	175.15	-26.47	2.155	0.007	24	24	6.34 ± 3.36	100.0	1	9,1
CCPC-z21-005	214.31	52.40	2.160	0.007	5	3	9.07 ± 6.93	100.0	1	
CCPC-z21-006	334.35	0.32	2.172	0.005	4	3	18.85 ± 13.55	100.0	1	
CCPC-z21-007	356.58	12.80	2.174	0.002	7	7	17.27 ± 10.57	100.0	1	
CCPC-z21-008	149.98	2.11	2.179	0.002	5	1	9.41 ± 5.38	100.0	2	4
CCPC-z22-007	255.20	64.26	2.296	0.008	32	30	7.77 ± 2.90	100.0	1	6
CCPC-z23-002	334.46	0.14	2.309	0.009	4	3	11.45 ± 9.05	100.0	1	
CCPC-z23-003	214.39	52.49	2.333	0.008	4	4	11.73 ± 10.66	100.0	1	
CCPC-z23-007	258.53	50.27	2.390	0.005	7	6	16.63 ± 14.52	100.0	1	1,20
CCPC-z24-003	164.20	-3.64	2.426	0.005	7	7	15.06 ± 8.92	100.0	1	
CCPC-z24-005	150.00	2.26	2.442	0.009	14	8	9.27 ± 4.93	100.0	1	4,10
CCPC-z25-002	255.18	64.17	2.537	0.002	4	4	19.86 ± 13.41	100.0	1	
CCPC-z25-003	143.36	28.77	2.548	0.003	5	5	10.89 ± 7.70	100.0	1	
CCPC-z25-007	216.14	22.84	2.581	0.007	5	2	10.90 ± 6.72	100.0	1	
CCPC-z27-012	16.48	-25.81	2.758	0.008	4	3	12.83 ± 10.91	100.0	1	
CCPC-z28-011	36.36	-4.32	2.820	0.006	4	2	9.10 ± 7.53	100.0	1	
CCPC-z28-016	36.27	-4.28	2.866	0.006	5	1	15.46 ± 11.42	100.0	1	
CCPC-z29-009	136.34	34.14	2.905	0.010	5	5	13.33 ± 9.13	100.0	1	
CCPC-z29-011	339.95	11.87	2.925	0.008	13	3	9.62 ± 5.14	100.0	1	
CCPC-z29-013	13.31	12.63	2.934	0.002	5	4	14.67 ± 10.21	100.0	1	
CCPC-z31-015	339.87	11.88	3.148	0.008	9	2	8.15 ± 4.56	100.0	1	
CCPC-z32-007	46.15	-0.19	3.233	0.007	5	4	11.47 ± 9.32	100.0	1	
CCPC-z33-006	150.07	2.28	3.303	0.008	4	4	17.49 ± 13.89	100.0	2	
CCPC-z33-007	334.35	0.07	3.310	0.012	7	5	11.52 ± 9.15	100.0	1	
CCPC-z33-010	216.12	22.83	3.379	0.009	7	3	10.44 ± 7.10	100.0	1	
CCPC-z36-007	34.54	-5.30	3.688	0.010	5	2	15.22 ± 16.25	90.0	2	
CCPC-z44-003	189.15	62.23	4.424	0.010	5	3	13.79 ± 12.03	90.0	1	
CCPC-z26-001	339.84	11.81	2.617	0.003	4	4	7.33 ± 6.16	85.6	1	
CCPC-z26-006	255.14	64.22	2.688	0.005	5	5	7.24 ± 5.18	85.6	1	
CCPC-z27-008	36.39	-4.51	2.729	0.006	6	2	7.27 ± 4.83	85.6	1	
CCPC-z31-008	339.89	11.88	3.104	0.007	8	5	7.70 ± 4.58	85.6	1	
CCPC-z31-017	36.76	-4.56	3.187	0.006	11	3	7.26 ± 3.67	85.6	1	
CCPC-z34-006	36.54	-4.63	3.472	0.013	6	1	7.82 ± 5.40	85.6	2	

Note. — The names and positions (1^{st} through 3^{rd} columns) of the most overdense candidate protoclusters. If we recover an overdensity that was previously identified, the discovery references are listed in the last column. References: (1) Galametz et al. (2012), (2) Galametz et al. (2013), (4) Diener et al. (2013), (5) Yuan et al. (2014), (6) Steidel et al. (2005), (8) Gobat et al. (2013), (9) Pentericci et al. (1997), (10) Chiang et al. (2014), (20) Keel et al. (1999). See text for column descriptions.

Table 3.2. CCPC II: Mass Estimates-Best Candidates

Candidate Name	R_e (Mpc)	σ (km s^{-1})	Virial Mass Estimate (10^{14} $M_\odot$)	δ_m (b = 3)	Overdensity Volume (cMpc^3)	Overdensity Mass Estimate (10^{14} $M_\odot$)
CCPC-z20-002	4.5	749	14.7	3.13	5361	9.8
CCPC-z20-003	2.3	367	1.8	6.48	125	0.4
CCPC-z20-009	4.3	446	4.9	4.38	589	1.4
CCPC-z21-004	3.4	701	9.6	2.11	519	0.7
CCPC-z21-005	5.0	689	13.8	3.02	799	1.5
CCPC-z21-006	2.9	479	3.9	6.28	58	0.2
CCPC-z21-007	1.2	156	0.2	5.76	13	< 0.1
CCPC-z21-008	3.7	211	1.0	3.14	176	0.3
CCPC-z22-007	3.1	709	8.9	2.59	3032	4.7
CCPC-z23-002	2.6	783	9.3	3.82	60	0.1
CCPC-z23-003	2.3	711	6.8	3.91	171	0.4
CCPC-z23-007	1.5	442	1.7	5.54	82	0.2
CCPC-z24-003	2.4	399	2.2	5.02	327	0.8
CCPC-z24-005	4.1	770	14.1	3.09	9110	15.5
CCPC-z25-002	1.3	208	0.3	6.62	28	0.1
CCPC-z25-003	2.2	295	1.1	3.63	116	0.2
CCPC-z25-007	3.1	620	6.9	3.63	215	0.4
CCPC-z27-012	1.3	632	2.9	4.28	217	0.5
CCPC-z28-011	2.6	504	3.8	3.03	280	0.5
CCPC-z28-016	4.0	429	4.2	5.15	458	1.2
CCPC-z29-009	2.2	757	7.2	4.44	27	0.1
CCPC-z29-011	4.1	628	9.3	3.21	4579	8.0
CCPC-z29-013	1.7	189	0.3	4.89	56	0.1
CCPC-z31-015	3.7	558	6.6	2.72	2011	3.0
CCPC-z32-007	2.6	504	3.8	3.82	280	0.6
CCPC-z33-006	1.8	578	3.5	5.83	115	0.3
CCPC-z33-007	3.2	818	12.3	3.84	101	0.2
CCPC-z33-010	3.4	611	7.3	3.48	730	1.4
CCPC-z36-007	3.3	634	7.7	5.07	165	0.4
CCPC-z44-003	1.4	575	2.6	4.60	456	1.0
CCPC-z26-001	0.7	245	0.2	2.44	3	< 0.1
CCPC-z26-006	2.2	366	1.7	2.41	139	0.2
CCPC-z27-008	3.5	463	4.3	2.42	5570	7.9
CCPC-z31-008	2.6	541	4.3	2.57	1038	1.6
CCPC-z31-017	3.7	462	4.5	2.42	10025	14.6
CCPC-z34-006	4.0	878	17.7	2.61	270	0.4

Note. — The mass estimates for the largest overdensities in CCPC2 (same order as Table 3.1). The effective radii (R_e, in physical units) and velocity dispersions (σ) are used to compute a virial mass estimate, with the caveat that these systems are not expected to be in equilibrium at the relevant redshifts. We obtain the mass overdensity δ_m by assuming galaxies are linearly biased tracers of mass with a slope of b = 3. See text for details.

ters (listed in Table 4 of this published article), although many are concentrated in a few catalogs. The single largest source of spectroscopic redshifts that were identified as protocluster galaxies is from Steidel et al. (2003). These galaxies were identified as Lyman Break Galaxies initially (with additional color criteria), and then followed up spectroscopically to a limiting magnitude of $R_{AB} = 25.5$ using the Keck telescopes over a field of view of 0.38 square degrees. Many other spectroscopic surveys utilized very deep, multi-wavelength fields (e.g. CANDELS GOODS, Hubble Deep/Ultra Deep) to identify candidate high redshift galaxies with various color cuts, dropouts, or *Chandra* emission (Noll et al., 2004; Le Fèvre et al., 2005; Reddy et al., 2006; Balestra et al., 2010). These galaxies would then be targeted by VIMOS/FORS/MOIRCS (or a similar instrument) to confirm their redshift.

Another common source of redshifts is from surveys that capitalize on strong line emission from faint star forming galaxies, either with $Ly\alpha$ or $H\alpha$. By using a narrowband (NB) filter which selects out one of these lines at a specific redshift, one can effectively identify a large number of sources efficiently. These candidates can then be targeted using a large telescope to confirm their redshift. This technique has been used to identify a number of high redshift protoclusters (Fynbo et al., 2003; Venemans et al., 2007; Cooke et al., 2014). An interesting observation that was noticed by Koyama et al. (2013) is that $Ly\alpha$ and $H\alpha$ are rarely simultaneously observed in star-forming galaxies within a protocluster at $z = 2.16$, suggesting that they may be entirely different populations. They could also have varying amounts and distribution of dust. Simulations suggest that roughly 90% of protocluster galaxies are actively forming stars at these redshifts (Contini et al., 2016).

3.1.2 Candidate Protocluster Criteria

While the most massive, $z = 0$ clusters have radii on the order of a few Mpc, the components that form these systems (i.e. protoclusters) are much more extended at higher redshift.

Previously identified structures at high redshift have observed sizes larger than 50 comoving Mpc (cMpc) (Shimasaku et al., 2003; Matsuda et al., 2005; Lee et al., 2014). These seemingly large sizes have a theoretical basis within ΛCDM simulations as well. Chiang et al. (2013) and Muldrew et al. (2015) mined the Millennium simulation (Springel et al., 2005) for clusters (collapsed halos with $\gtrsim 10^{14}\ M_{\odot}$) at $z = 0$. Using the semi-analytic model of Guo et al. (2011) to trace the galaxy distributions of the primordial clusters backwards, they analyzed the evolution of these systems. Muldrew et al. (2015) found that at $z = 2$, 90% of the stellar mass of a protocluster can be found spread across 35 h^{-1} cMpc, which grows to more than 40 h^{-1} cMpc at $z = 5$. This example is typical for the largest mass systems (clusters with $M_{z=0} > 10^{15}\ M_{\odot}$), while a cluster with $M_{z=0} = 10^{14}\ M_{\odot}$ might have a modest size of $\sim 20\ h^{-1}$ cMpc. On the sky, the most massive protoclusters can span more than 0.5 degrees, while the smallest mass clusters are roughly half of that (Muldrew et al., 2015).

This can be problematic when searching for such systems, in that many deep, spectroscopic surveys (with the exception of the CANDELS GOODS-S field) would not encompass the full breadth of the most massive protoclusters. In Franck & McGaugh (2016b), we plotted the positions of field and CCPC1 protocluster galaxies within search radii of 20 cMpc to illustrate the distribution of members with respect to the survey widths in the Appendix. Surveys rarely extended beyond the probed search radius. Similarly, Muldrew et al. (2015) make an example of two candidate systems found within the literature (Koyama et al., 2013; Cooke et al., 2014) that have particularly small fields of view which could only target the innermost halo of a protocluster. This imposes a selection effect, in that most ($\sim$ 90%) simulated protoclusters do not have a single, 'main' halo at these redshifts (Muldrew et al., 2015).

We took an initial list of ~14,000 galaxies and identified galaxy groups of 3 or more

within a radius of 2 arcminutes and a $\Delta z < 0.03$ in CCPC1. Individual galaxies are required to be more than 3" from one another to be considered unique. From these groups, we expanded our search to a radius of 20 cMpc on the sky and a redshift depth of ±20 cMpc. Any cylindrical volume that contained 4 or more galaxies and had a galaxy overdensity of $\delta_{gal} > 0.25$ (calculation found in the following subsection) was considered a candidate protocluster and included in CCPC1.

These requirements are based on the need for a simple, adaptable, and effective means of identifying galaxy structures from a variety of surveys. The large search radius enables wide surveys to be adequately probed corresponding to the size of the largest protoclusters. The modest richness requirement of 4 or more galaxies is sensitive to surveys of small volumes with few expected galaxies. These two extremes of survey depths are moderated by the density requirement of $\delta_{gal} > 0.25$. The $N \geq 4$ galaxy requirement serves primarily as a signpost from which to calculate the overdensity. A more detailed discussion of the algorithm will follow.

In this paper, we have removed the intermediate step of finding groups of 3 or more galaxies within a 2 arcminute radius. The overdensity of these candidate protoclusters, which may not be accompanied by a dense, central knot of galaxies, is of greater importance than the possible chance alignment of galaxies on the sky. Therefore, any group of 4 or more galaxies that also exhibits a galaxy overdensity of $\delta_{gal} > 0.25$ is considered to be part of the CCPC. The following subsection will describe the overdensity calculation in detail. In Section 3.2.1 we justify the removal of the intermediate step statistically.

Identifying structure in large, high redshift data sets is not a novel exercise. Diener et al. (2013) found infalling groups of spectroscopic galaxies separated by small physical separations in space in the zCOSMOS field. The GOODS-N and GOODS-S fields have been the subject of many searches for overdensities found with photometric redshifts and

supplementary spectroscopic catalogs at $z > 2$ (Kang & Im, 2009; Salimbeni et al., 2009; Kang & Im, 2015). Indeed, we may have recovered a number of these overdensities, which have $0 \leq N \leq 4$ spectroscopic member galaxies in Kang & Im (2015). In CCPC1 (Franck & McGaugh, 2016b), our candidate protocluster CCPC-z37-001 was identified near a photometric redshift overdensity by Kang & Im (2009), while CCPC-z27-004,CCPC-29-004, and CCPC-z34-001 most likely coincide with overdensities found in Kang & Im (2015). CCPC-z22-001,CCPC-z23-001 and CCPC-z24-001 (all featured in CCPC2) have similar positions and redshifts to the overdensities reported by Salimbeni et al. (2009), while CCPC-z25-004, CCPC-z25-005, CCPC-z40-001, and CCPC-z42-001 may coincide with candidate structures in Kang & Im (2015). CCPC-z28-002, CCPC-z34-002, and CCPC-z37-001 in Franck & McGaugh (2016b) are all overdensities originally found in the VIMOS GOODS-S spectroscopic survey by Balestra et al. (2010).

The CCPC is unique in that it applied a single and simple algorithm to all archival spectroscopic data available. The number of galaxy overdensities identified in this paper (hereafter referred to as CCPC2) is 173, the 36 strongest of which are listed in Table 3.1. The division of these 'Best' candidates is explained in Section 2.3 (Overdensities). The total list of candidates can be found in Table 3 of Franck & McGaugh (2016a). The RA/DEC coordinates in the Table are centered on the galaxy from which the number of members is maximized for a given overdensity. The redshift listed is the mean value of the system members, and may be minimally offset from the search galaxy's redshift in some cases ($\Delta z \sim 0.0005$ on average). We have included the number of galaxies within $R = 10, 20$ cMpc from the central galaxy, which are equivalent in many instances. This is primarily dependent on the limited FOV of the surveys. In combination with the 43 protocluster candidates in CCPC1, the Candidate Cluster and Protocluster Catalog contains 216 systems between the redshifts of $2 < z < 7$. To the best of our knowledge, this represents the

largest collection of spectroscopic, $z \geq 2$ galaxy overdensities in the literature at the time of writing. Each CCPC candidate has an individual list of redshift measurement references in Table 4 of Franck & McGaugh (2016a).

As in CCPC1, many of the protoclusters have more than the minimum number of members, with a median value of 6 galaxies per candidate. There are 40 CCPC2 systems with the minimum 4 galaxies, and 9 candidates with 23 or more members (Table 3 of the online version). There does not appear to be a strong trend in numbers of galaxy members as a function of redshift for the bulk of the CCPC. However, all candidates with more than 20 galaxies are found at $z < 4.5$. The median number of galaxies in protoclusters in the redshift bins of $2 < z < 3$, $3 < z < 4$, $4 < z < 5$, $5 < z < 7$ are 5, 6, 8, and 7, respectively within CCPC2. Although these are only slight variations, one would naively expect that the median numbers would decrease as a function of distance. Perhaps at high redshift we are only identifying the richest overdensities, and thus their median members are larger.

Each candidate structure in CCPC2 has a spectroscopic rating ('Q') associated with it in Table 3.1 and Table 3 of Franck & McGaugh (2016a). In many cases, redshift values cataloged by NED have an accompanying Qualifier flag that distinguishes the quality of a given redshift measurement. If a protocluster has 4 or more member galaxies with no quality flags raised, thus satisfying our criteria as a candidate structure, it is assigned the greatest rating of '1'. If there are 4 or more galaxies that have either no flags or were identified based on a single line, the system is rated slightly lower with a '2'. A rating of '3' is assigned to collections of galaxies with redshift flags of a somewhat uncertain measurement or tentative result. Galaxies with redshifts flagged as photometric redshift, modeled from SEDs, or highly uncertain/questionable values, are not used in this work. In CCPC2, there are 135 (out of 173) systems identified with the highest quality spectroscopic data (RATING=1), and 32 that have a rank '2' rating.

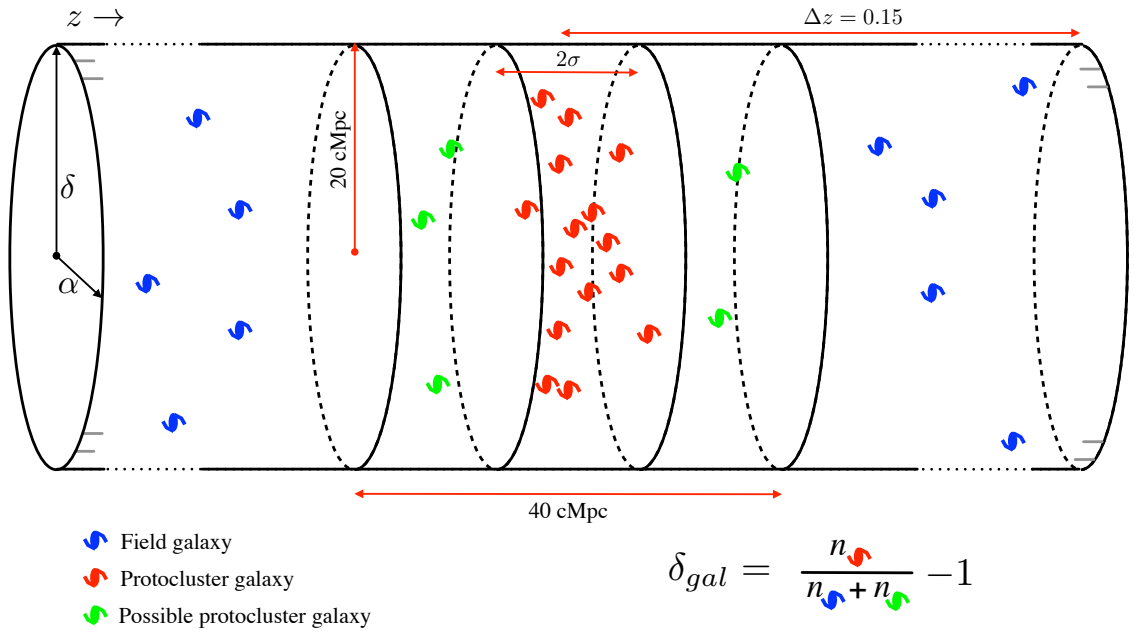

Figure 3.1: A visual representation of the overdensity (δ_{gal}) measurement. Galaxies within ±20 cMpc but outside of $z \pm \sigma$ are considered possible protocluster members, but are treated as field galaxies for the computation of the overdensity so that it is a conservative estimate.

3.1.3 Candidate Overdensities

To measure the strength of these candidate systems quantitatively, we estimate their galaxy overdensity using the simple formula $\delta_{gal} = (n_{proto}/n_{field}) - 1$. The number density of protocluster galaxies (n_{proto}) is based on the density of galaxies along the line of sight (LOS) within the dispersion (σ_z) from the center of the redshift distribution. Limiting the galaxies in the overdensity to those only within the redshift dispersion likely excludes some objects that are on the outskirts of the candidate structure, and so the δ_{gal} values are conservative estimates. These overdensities and their uncertainties can be found in Table 3.1 and Table 3 in Franck & McGaugh (2016a). Fig 3.1 illustrates the calculation of the galaxy overdensity.

To calculate the number density of the field (n_{field}), we use the same aperture on the sky

as our search criteria (R = 20 cMpc) and identify all galaxies along the line of sight (using the same galaxy surveys) within a maximum length of $\Delta z = 0.15$ from the protocluster center. The overdense region is excised from the field counts. The choice of the field length is typically an order of magnitude longer than the overdense region to ensure a fair sample. A visual representation of the calculation is shown in Fig 3.1. It is important to note that in every step of calculating the overdensity (δ_{gal}), we are adopting the most conservative values. The purpose of the CCPC was to investigate protoclusters using methods which did not 'cherry-pick' the largest values of overdensities.

The uncertainties in field counts are estimated using the cosmic variance calculator from Trenti & Stiavelli (2008), while the uncertainty within the overdense region is $\sqrt{N_{proto}}$. For each structure, we input the volume probed by field systems and an assumption of the completeness of the spectroscopic survey. The output is not particularly sensitive to the completeness assumption ($\sigma \pm 0.1$ galaxies between 10-90%), and so we have adopted a 50% completeness. The field and protocluster uncertainties are then added in quadrature. On average, the inclusion of cosmic variance and completeness adds 0.3 galaxies to the uncertainty of the CCPC when compared to the Poissonian treatment of uncertainties.

When measuring n_{field} or n_{proto} for the calculation of δ_{gal}, the length (Δz) that the density is computed over is limited by the extent of the data, and not the volume queried (e.g. $\Delta z \leq 0.15$ for the field). For example, NB filters do not extend the full possible width of the field distribution. In < 15% of cases in the CCPC, low numbers of galaxies clustered in redshift space increased n_{proto} or n_{field} to large levels (e.g. $\delta_{gal} > 50$). We set Δz equal to the protocluster redshift dispersion (σ_z) for these low richness cases. In some overlapping instances, the field galaxy counts were very low ($N_{field} < 3$) compared to other candidate structures. For these systems in CCPC1 we injected seven more galaxies into the counts for the field number density n_{field} to reflect the median field counts in the low-richness

sample. This effectively decreases the overdensities to more reasonable values. In this work we have neglected this rather un-physical method, but instead put brackets around the overdensity values to reflect their low-richness status in Table 3 in Franck & McGaugh (2016a). The resulting overdensities are highly uncertain as a result, and should not be relied on without further observations. Although these overdensity vales are questionable, the average number density of these systems ($n \sim 7 \times 10^{-2}$ cMpc^{-3}) exceeds that of the mean value for CCPC2. This suggests that these systems are likely overdensities, and need not be removed simply because they lack field counts.

The median value of $\delta_{gal} = 2.9$ for CCPC2, only slightly larger than for CCPC1 ($\delta_{gal} \sim 2$). Only 15% of CCPC2 systems have $\delta_{gal} < 1$. Generally, $\delta_{gal} \sim 2$ is typical of many protoclusters in the literature (see Table 5 for a summary in Chiang et al., 2013). These values have considerable breadth, from $0.7^{+0.8}_{-0.6}$ (Venemans et al., 2007) to 16±7 (Toshikawa et al., 2014), which are consistent with the range in the CCPC. The mean number density of the δ_{gal} volume is $n = 5.9 \times 10^{-2}$ cMpc^{-3} in CCPC2. For a general comparison, the density of LAEs at $z \sim 3$ is $n = 1.5 \times 10^{-3}$ cMpc^{-3} (Gronwall et al., 2007), but this value is dependent on the galaxy type. In simulated protoclusters within Millennium, analysis shows that overdensities of $\delta_{gal} = 1 \pm 1$ from redshifts $z = 2 - 5$ are consistent with low mass, $M_{z=0} \approx 10^{14}\ M_{\odot}$ clusters (Chiang et al., 2013). This assumes that galaxies in the overdensity have $SFR > 1\ M_{\odot}\ yr^{-1}$ and are within boxes of [25 cMpc]^3. The most massive systems $M_{z=0} > 10^{15}\ M_{\odot}$ have overdensities roughly a factor of 3 larger, as one would expect (Chiang et al., 2013).

These simulations also show that δ_{gal} expectations for protoclusters are inversely proportional to the volume probed. Steidel et al. (2003) identified LBGs in windows of $R \sim 5$ cMpc on the sky at these redshifts, which is only 25% the size of the maximum CCPC search radius. The line-of-sight distance probed is not similarly hampered. Figure 3 in Chi-

ang et al. (2013) suggest that the stacked overdensity profiles of Millennium protoclusters can be factors of 6 or more larger when evaluated in such small boxes ($R_e \sim 5$, $z = 3$). Indeed, the CCPC candidates in the regions of Steidel et al. (2003) do have a larger typical overdensity with a median value of $\delta_{gal} > 5$, 20 times larger than our minimum required overdensity and 70% larger than that of the CCPC2 median value. This illustrates the inherent uncertainty when evaluating overdense regions.

A relatively compact overdensity region can be dwarfed by the diffuse nature of field galaxies along the line of sight in some instances (like CCPC-z21-003), resulting in $\delta_{gal} \sim 0$. For four these special cases, we limit the sky aperture of our field counts to the surface area of the overdense region. The overdensities estimated in this way are noted in Table 3 in Franck & McGaugh (2016a). Occasionally, the maximum field length of $\Delta z = 0.15$ will intersect other structures in the same field. As the Universe does not contain a simple, smooth distribution of galaxies, it can be quite difficult to measure the 'field' surrounding an overdensity. Therefore, the value of δ_{gal} should be taken as an estimate of the strength of the structure, and should not be treated as an absolute metric. As an example of this, we have CCPC-z23-001 and CCPC-z23-005, which are two structures identified by our algorithm whose volumes are fused together in an unbroken $\sim$ 80 cMpc long galaxy distribution. They have 43 and 23 galaxies, respectively, which is significantly larger than the median number of member galaxies in CCPC2 (6). Despite their richness, they have middling δ_{gal} values of 2.0 and 0.7, as their field counts intersect one another. We will discuss the implications of this 'superstructure' in Section 5.4 with more detail.

3.1.4 Probable Reality of Protoclusters

It is possible that some of the structures we identify are real, and others are chance coincidences. To estimate the probable reality of each overdensity, we refer to simulations

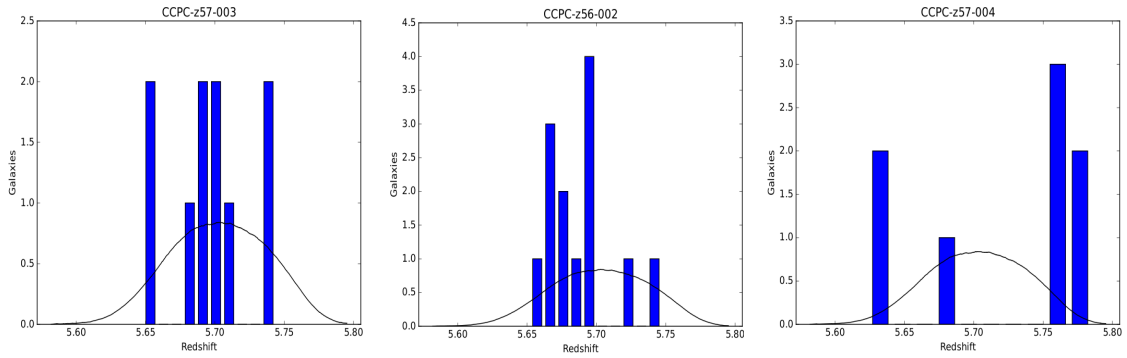

Figure 3.2: To test the possibility that galaxies observed through a narrowband filter could produce a false positive protocluster detection from its transmission function, we constructed a Monte Carlo simulation. Shown here are the distribution of LAE galaxies within the NB filter NB816 for three protocluster candidates. Plotted in black is the transmission curve of NB816 at this redshift, arbitrarily normalized to illustrate its shape. The simulated galaxies were distributed following the transmission function of the filter. The simulated distributions were then compared to the mean redshift ($< z_{obs} > - < z_{MC} >$) and dispersion ($\sigma_{z,obs} - \sigma_{z,MC}$) of the actual data. The left panel shows one extreme case in which the distribution of real galaxies show only a 0.5σ offset from the Monte Carlo simulation. The center panel is a galaxy distribution in the filter that shares the median value difference (2σ) between the simulated LAEs and the observations, while the right plot shows the most extreme difference example ($\sim 5\sigma$). With the possible exception of the leftmost case, the observed distributions do not follow simply from the shape of the filter transmission.

of structure formation. Chiang et al. (2013) computed the probability distribution function for overdensities of galaxies becoming a cluster at $z = 0$ (their Fig 8) using the semi-analytic models of Guo et al. (2011) in the Millennium simulation. For different values of δ_{gal}, they calculated the fraction of volumes that would become halos of $M > 10^{14}$ $M_{\odot}$ (i.e. clusters). They chose to identify galaxy overdensities in the simulation in [15 cMpc]^3 boxes at redshifts $z = 2 - 5$ for different galaxy populations. Using their results for galaxies with $M_* > 10^{10}$ $M_{\odot}$ (the most biased tracers of mass), we attribute a conservative probability that each CCPC candidate will collapse into a cluster at $z = 0$ based on its overdensity. If the galaxies within these structures are less biased tracers (e.g. galaxies with $M_* < 10^{10}$ $M_{\odot}$), these probabilities can increase by as much as 40%. These percentages can be found in Tables 3.1, 3 (online version) next to each CCPC candidate. Table 3.1 contain systems that have $\geq 85\%$ probability of becoming a cluster at $z = 0$.

There are indications that the overdensities and probabilities we calculate for the CCPC are too conservative. In Section 5.4, we estimate the number of structures expected within the volume of the CANDELS GOODS-S field. Compared to the candidate protoclusters we identify (and the probability of collapse we assign), the discrepancy is at a minimum of a factor of two smaller, and may be an order of magnitude too low.

It has been noted that a spurious protocluster signal could be the result of the transmission curve of a NB filter (Venemans et al., 2007). A false-positive overdensity is possible if more galaxies are detected at the central wavelength of the filter because of its greater transmittance, while the less-responsive tails yield fewer detections. For instance, the FWHM of the Subaru Telescope's NB filters used to select LAEs at $z = 3.1, 5.7, 6.5$ are only slightly larger than the expected diameter of the most massive protoclusters at these epochs. A smooth distribution of galaxies could appear as an overdensity at the central wavelength with respect to the edges, where 'field' galaxies might reside. We find that this is not the

case, and the galaxy distributions differ significantly ($> 2\sigma$) from the transmission functions. This serves as a further confirmation that we are detecting actual structure.

To model the possibility of false detections, we adopted a similar approach to Venemans et al. (2007). We ran a Monte Carlo simulation based on the transmission curves of the Suprime-Cam's NB503, NB816, and NB921 filters with 10^4 iterations for all of the 15 protoclusters detected from these observations. For each iteration, the number of galaxies observed within the filter were distributed according to its transmission probability. The difference between the real vs. simulated mean redshift ($< z_{obs} > - < z_{MC} >$) and observed vs. Monte Carlo redshift dispersions ($\sigma_{z,obs} - \sigma_{z,MC}$) were calculated in units of the standard deviation in the simulation. When compared to the actual data of galaxies identified through these filters, the average deviation from the Monte Carlo was 2.4σ (combined $< z >$, dispersion deviations). This suggests that the galaxy distribution (and therefore the overdensities) are not merely the result of the NB filter. In Fig 3.2, the galaxy distributions of three protoclusters (CCPC-z57-004, CCPC-z56-002, CCPC-z57-004) are plotted with respect to the transmittance of NB816. These examples show the full range of the deviations from the Monte Carlo ($0.5 - 4.9\sigma$).

3.1.5 Mass Estimates

The mass of clusters and protoclusters at various redshifts is a stringent probe of cosmological parameters ($\Omega_{\rm m}$, Ω_Λ, σ_8 and w), as the cluster mass function is tied to these values (Press & Schechter, 1974). We attempt to provide two mass estimates in this work. The first is based on the volume and mass overdensity (δ_m) that may collapse by $z = 0$, and the second is a crude virial mass estimate based on the velocity dispersion σ and effective radius of the system. However, these two values do not appear correlated and are highly uncertain.

Steidel et al. (1998) and Venemans et al. (2007) have estimated the expected collapsed masses of protoclusters at high redshift by estimating a galaxy overdensity (δ_{gal}), a volume encompassing the galaxies within the overdensity V, and the critical density of the Universe at that redshift ρ_{crit}. Galaxies are assumed to be mere tracers of the dark matter distribution, so it is required to assume a bias parameter b to convert the observed galaxies into a mass overdensity δ_m. Historically, linear bias parameters of $3 \leq b \leq 6$ have been used (Steidel et al., 1998; Venemans et al., 2007), and the matter overdensity can be found via $\delta_m = \delta_{gal}/b$. In this work, we adopt a bias of $b = 3$. At a redshift of $z = 3$, $\rho_{crit} = 4.2 \times 10^{10}\ M_\odot\ \mathrm{cMpc}^{-3}$ in our assumed cosmology. If the entire volume is assumed to collapse into a single halo by $z = 0$, that cluster will have an estimated mass of

$$M = \rho_{crit,z} V (1 + \delta_m). \tag{3.1}$$

Table 3.2 contains the mass estimates for the protoclusters with the most significant overdensities from Table 3.1. All system masses are located in Table 5 of Franck & McGaugh (2016a).

There are a number of assumptions that go into this calculation, the most critical of which is the volume estimate. Galaxies on the outskirts of the extended protocluster distribution may not collapse into a single halo by $z = 0$, or be bound to the structure at all, as is seen in simulations (Muldrew et al., 2015). In addition, the SAM used by Contini et al. (2016) to investigate protocluster galaxies suggests that these may be indistinguishable observationally from their field counterparts. Therefore, including these as tracers of the volume that will collapse into a cluster can greatly increase the mass estimate of the system, possibly by orders of magnitude. Some previous works that have utilized this mass estimator include a corrective factor for redshift space distortions (Steidel et al., 1998, 2005). This can result in a difference factor of ~ 2 in some instances. As the volume assumptions

can change the mass estimate by orders of magnitude, this space distortion calculation is neglected.

The bias parameter b is an assumed value, and depends on the galaxies used as tracers, in that higher mass galaxies are more biased tracers of mass. If the galaxies that trace the protoclusters do not have $b = 3$, the overdensity mass estimate will also be systematically affected. Bias parameters can also evolve over time, growing larger at higher redshifts. For LAEs at $z = 3.1$, Gawiser et al. (2007) estimate a value of $b = 1.7$, while LAEs at $z = 4.86$ have $b \approx 3$ (Ouchi et al., 2003). Biases of larger mass galaxies (LBGs, for instance) can have $b \gtrsim 4$ at $z > 4$ (Ouchi et al., 2004b). Based on the no-merger model of Fry (1996), strong evolution of the bias ($\Delta b \gtrsim 3$ from $z = 0$) is predicted for high redshift, large bias systems (e.g. $b = 6$ at $z = 5$). For systems of modest bias ($1 < b < 2$), the evolutionary difference is less than unity. If our algorithm is selecting only the most biased sources, the implied mass overdensities may not be sufficient to become clusters. However, if the bias is similar to the roughly constant, unevolving value of Gawiser et al. (2007), our candidates might be more significant than we claim.

The mean estimated collapsed mass of candidates in CCPC2 is $M_{z=0} = 1.8 \times 10^{14}\ M_{\odot}$, consistent with low mass clusters found in the local universe. Systems of this mass comprise roughly 70% of the population of clusters in the Millennium simulation (Chiang et al., 2013), while clusters of $M_{z=0} > 10^{15}\ M_{\odot}$ should represent only 2% of cases. However, our search methodology is not expected to be mass-blind, and preferentially selects the highest overdensity systems.

For clusters at low redshift ($z \sim 0$), a traditional method of estimating a mass was to assume the system was virialized, measure the velocity dispersion (σ) and effective radius

R_e, and compute the system mass via

$$M = \frac{2\sigma^2}{G} R_{hms}, \tag{3.2}$$

where $R_{hms} \sim 1.25 R_e$. However, this assumption is not expected to hold at higher redshifts in ΛCDM . At 3 Gyrs after the Big Bang ($z \sim 2$), analysis of the Millennium simulation shows that the the progenitors of the most massive clusters at $z = 0$ ($M_{z=0} > 10^{15}\ M_\odot$) have a dark matter halo of $\geq 10^{14}\ M_\odot$, while the lowest mass clusters at the present day may not have assembled this mass until $z \approx 0.6$ (Chiang et al., 2013). Prior to this epoch, it is likely that only subhalos have virialized.

For each object in the CCPC, we have calculated the effective radius (R_e) in which 50% of the total protocluster members reside, as well as the velocity dispersion of the system σ, and computed a 'virial' mass estimate. Mass estimates for the entire CCPC2 list are in Table 5 of Franck & McGaugh (2016a), while the most overdense protoclusters can be found in Table 3.2. We again urge caution in interpreting these results, as it is unlikely that such extended, diffuse systems are in virial equilibrium. We use the 'virial' term only because we utilize the virial mass equation. As there are few protocluster members in some CCPC systems, their diffuse nature can imply large values of both R_e and σ, as is the case for CCPC-z24-007. It has only 4 galaxy members (the minimum number) and an effective radius of 4.8 (physical) Mpc, a dispersion of 760 $km\ s^{-1}$. These yield a virial mass estimate of $1.6 \times 10^{15}\ M_\odot$, a factor of two larger than the average estimate in CCPC2 ($\sim 8 \times 10^{14}\ M_\odot$), despite its minimal richness.

In contrast, there are a few systems (CCPC-z20-005, CCPC-z21-011, CCPC-z22-002) that have a large number of candidate galaxy members ($N \geq 18$) with large implied virial masses ($\geq 10^{15}\ M_\odot$). These systems are at an epoch ($z \leq 2.3$) that can theoretically host virialized clusters (Chiang et al., 2013). With the significant increase in richness, these mass

estimates may be at least more physically meaningful than the previous example (CCPC-z24-007), although these would still be an order of magnitude larger than predicted. Further examination and discussion of the mass estimates (via mass overdensities and velocity dispersions) can be found in Section 3.2.3.

3.1.6 Objects of Interest

We have compiled a list of targets that are of potential significance, but violate our strict spectroscopic redshift criteria. These Objects Of Interest (OOI) are generally the result of narrow-band or photometric redshift observations, but otherwise fulfill the requirements of a CCPC target.The last two entries (OOI-z65-001 and -002) are spectroscopic galaxies, but had no field sources with which to calculate a δ_{gal} value. Some of these will likely prove to be fictional if targeted spectroscopically. These are listed in Table 6 of Franck & McGaugh (2016a).

3.2 Discussion

3.2.1 Tests of Structure

The method in which we identify structures is relatively simple. In CCPC1 (Franck & McGaugh, 2016b), we performed a number of tests on the algorithm in which candidate protoclusters were identified. In addition, the significance of the candidates were also evaluated. It is trivial to list positions of galaxy associations, but much more difficult to find physical systems. We have three overarching diagnostics of structures, each with sub-tests for significance: (1) the CCPC systems exist as overdensities, (2) they are statistically distinct spatially when compared to the 'field', and (3) their number densities in deep surveys are comparable to expectations from large simulations. That we recover a number of previ-

ously identified structures with our algorithm is an additional confirmation of its fidelity.

One of the criteria for CCPC protoclusters is that candidate systems must show an overdensity of galaxies ($\delta_{gal} \geq 0.25$) when compared to the local field along the line-of-sight ($\Delta z \pm 0.15$). This criteria is simply a lower limit, and the median overdensity is $\delta_{gal} = 2.9$. Within the literature, previously identified protoclusters (and simulations) at these redshifts show similar overdensities (Table 5 and Section 3.6, respectively, in Chiang et al., 2013).

Furthermore, these overdensities can be observed visually as spikes of galaxies along the line of sight in $N(z)$ plots in Fig 4 of Franck & McGaugh (2016a). These plots were constructed using the full aperture of the sky search radius ($R = 20$ cMpc). Many systems, such as CCPC-z20-003, have clear overdensities along the line of sight. Other overdensities, particularly those with centrally condensed galaxy distributions on the sky (e.g. CCPC-z20-008), may appear as little more than noise. This underlines the importance of having more than a single protocluster identifier (e.g. only $N(z)$ spikes, δ_{gal}, or a minimum number of galaxies).

It is of interest that the δ_{gal} values and CCPC galaxy members show little correlation. It is logical that spectroscopic surveys which have deeper limiting magnitudes would find more galaxies in both the field and in structures, all galaxies being equal. However, if protoclusters are indeed regions of enhanced mass assembly (Casey et al., 2015), rapid star formation or galaxy evolution (Hatch et al., 2011; Clements et al., 2014), it would be reasonable to assume that some manifestation of this behavior would present itself.

In fact, some of the largest overdensities ($\delta_{gal} \geq 10$) have fewer than 10 galaxy members within the search volume. For many of these cases (e.g. CCPC-z21-007 and CCPC-z23-007), the Poissonian uncertainties are large ($\geq \pm 10$) because of the small number of galaxies in the overdensity calculation (especially field galaxies). Some cases of these low N, high δ_{gal} cases are, interestingly, previously identified structures.

CCPC1 and CCPC2 have nearly identical median overdensities of $\delta_{gal} \sim 2.0, 2.9$ over the entirety of the redshift range. This is not a perfectly direct comparison, as the CCPC1 criteria required 3 or more galaxies to be found within 2 arcminutes of the center of the search radius. Potentially, CCPC1 may have been selecting the minority of protoclusters that have dominant main halos (Muldrew et al., 2015), which would represent stronger overdensities. Also, there is an expectation of some δ_{gal} evolution with redshift (Chiang et al., 2013), but with larger variations depending on the galaxies observed as tracers.

Despite these differences in redshift and methodology, the protoclusters at $2.74 < z < 3.71$ in the CCPC2 have a median $\delta_{gal} \sim 1.7$ and are consistent within the uncertainties of $\delta_{gal} \sim 2$ for CCPC1 and CCPC2. This is of two-fold importance: (1) there were 66 overdensities left undiscovered in the redshift range of CCPC1 (which contained 43 structures) by requiring an initial core group of galaxies to be a criterion, and (2) that the slightly greater median δ_{gal} between CCPC1 and CCPC2 suggests we are not admitting poorer candidates by removing this step.

Galaxy overdensities should not only exist along the line-of-sight (as observed in $N(z)$ plots and $\delta_{gal} > 0$), but their spatial distributions should also be distinct from the field. As protoclusters exist as very extended, diffuse systems on the order of tens of cMpc (Figs 1,2 in Muldrew et al., 2015), their profiles will naturally look more similar to isolated galaxies than $z = 0$ clusters. However, in CCPC1 we showed that there were significant differences between the surface density and cumulative distribution of field and protocluster galaxies.

In this work, we present similar findings based on the mean surface density of all CCPC2 systems (Fig 3.3). In order to capture the essence of 'field' galaxies, we used the initial list of $> 47,000$ galaxies from which we identified the candidate protoclusters at $z > 2$ to search for any companion galaxies within a radius of 20 cMpc, and ±20 cMpc along the line of sight. This mirrors the search volume of the CCPC algorithm. Included in this galaxy list

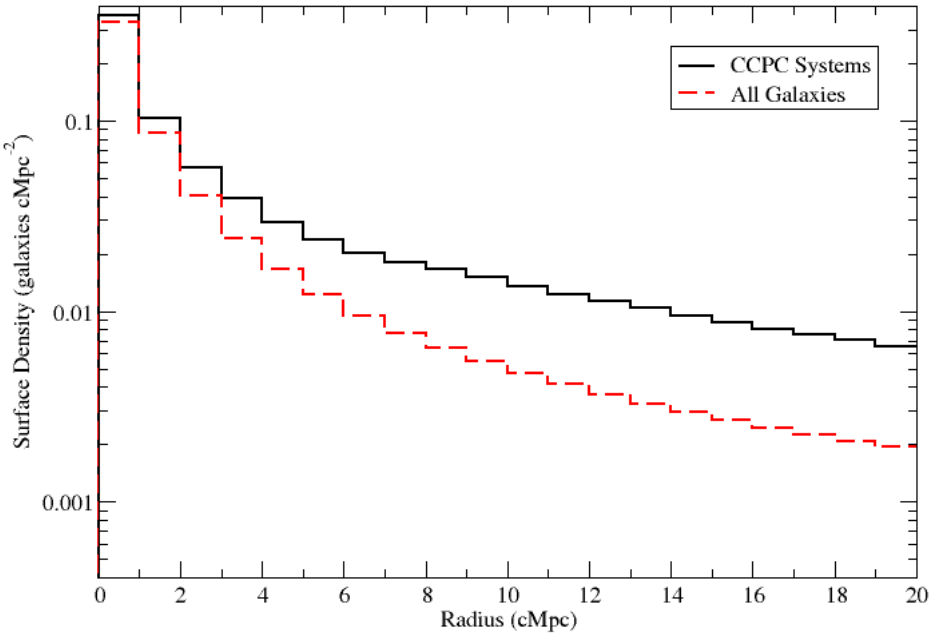

Figure 3.3: The mean galaxy density on the sky of CCPC2 galaxies is shown in black. In contrast is the mean surface density profile of All Galaxies in our initial list (>40,000 spectroscopic sources with $z > 2.00$), which is shown as the red dashed profile. The 'All Galaxies' distribution contains CCPC2 galaxies. This 'field' proxy is consistent with there being a single galaxy in the center of a $R = 20$ cMpc search radius with few companions. A two-sided KS test (0.6) suggests that these are distinct distributions.

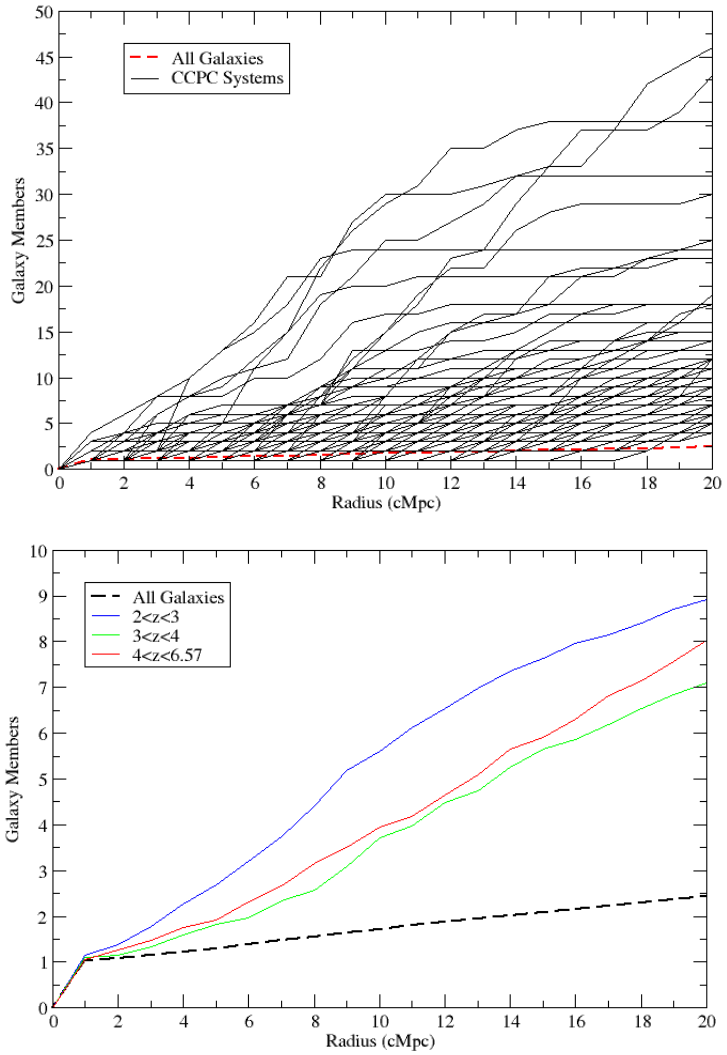

Figure 3.4: *Top* : The distribution of galaxies for each of the 173 CCPC2 systems are shown in black as a function of distance from the center of the search radius in comoving Mpc. The mean distribution of 44,000 spectroscopic galaxies from which we identified structures is shown as a dashed red line. For the CCPC systems, there is a clear difference in both the number and distribution of galaxies. In many cases, the survey edge can be seen as a flat line at $R < 20$ cMpc. *Bottom* : At different bins in redshift, we have plotted the mean number of galaxies as a function of radius, compared to the black dashed line representing All Galaxies at $z \geq 2$. As shown by the top panel, there is considerable variation in the distribution of galaxies. The difference in survey widths also prevents meaningful analysis of the comoving distribution of the galaxies, especially at radii > 10 cMpc. Primarily, this plot is effective at illustrating the difference between field galaxy distributions and those within protoclusters at all redshifts.

are all of the CCPC2 member objects. The mean sky surface density (galaxies cMpc^{-2}) is shown as a red dashed line in Fig 3.3. This is consistent with a single galaxy found within $R \leq 1$ cMpc ($\Sigma = 0.33$ cMpc^{-2}). This is labeled 'All Galaxies' instead of 'Field Galaxies', as we did not separate out isolated galaxy systems from overdense regions. By comparison, there are more mean galaxies in the central regions of CCPC systems (black line), and a shallower slope (-1.3 in $log - log$ space) compared to the field galaxy slope (-1.7). A two-sample Kolmogrov-Smirnov test (KS) value of 0.6 shows that these are distinct distributions at the 99.93% level.

Fig 3.4 illustrates the number of galaxies as a function of search radius $N(r)$ for each individual CCPC2 overdensity in the top panel. As in Fig 3.3, the red dashed line is representative of field galaxies, with clear differences in the distribution and mean number of sources at different comoving radii. The distribution of protocluster galaxies is highly variable, with some overdensities being very concentrated while others have lower central concentrations, but continually add galaxies to large radii. Plotting the number distributions instead of cumulative distributions (as in CCPC1) highlights the differences in survey widths more effectively, as some $N(r)$ profiles flatten out well before the maximum search radius of $R = 20$ cMpc. It is therefore important to not infer a 'characteristic' distribution of galaxies in protoclusters from such a plot, as some spectroscopic galaxy surveys will only target the innermost regions of structure, while others will trace out to larger radii. In addition, simulations suggest that there exists a menagerie of galaxy distributions within these high redshift systems, even among protoclusters that will have the same mass at $z = 0$ (Chiang et al., 2013; Muldrew et al., 2015).

The bottom panel of Fig 3.4 shows the mean 'All Galaxies' distribution as a *black* dashed line (instead of a red dashed line in the top panel), while CCPC2 galaxies in various redshift bins are shown in blue ($2 < z < 3$), green ($3 < z < 4$), and red ($4 < z < 6.57$). On

average, 'All Galaxies' will have roughly 1.5 companion galaxies surrounding it within the fixed CCPC volume, while at a minimum a CCPC system will have 3 other companions, and on average more than 7.

A KS test of the radial distributions of galaxies between 'All Galaxies' and the mean CCPC2 galaxies gives a value of 0.75, more significantly different than the surface density comparison (KS= 0.60). The various redshift bins, within the considerable scatter illustrated by the top panel, show little evolution. Simulations suggest there are large distribution variations of individual systems even at the same mass and redshift (Chiang et al., 2013; Muldrew et al., 2015). When coupled with the heterogeneous spectroscopic data (survey width/depth differences) from which we draw our candidates, the lack of meaningful $N(r)$ variations is not particularly surprising.

The most significant difference in selecting protocluster candidates between CCPC1 and CCPC2 was that the former contained a 'group finding' intermediate step. Initial groups were selected by requiring at least 3 galaxies within 2' of the search center, and within a $\Delta z = 0.03$. Requiring this intermediate step does not appear to affect the quality of sources, as the median overdensities between CCPC1 and CCPC2 are equivalent ($\delta_{gal} \sim 2$). Interestingly, requiring a centrally concentrated group of galaxies does not seem to have a significant effect on the surface density profiles (KS= 0.35 between CCPC1 and 2). Only the inner regions ($R \sim 1$ cMpc) show significant differences, with CCPC1 having a larger surface density (as a criterion) of $\Delta\Sigma = 0.12$ cMpc^{-2}. We conclude that the structures identified using the selection methods of CCPC1 and CCPC2 are not significantly different from one another, and both are selecting plausible protocluster candidates.

The CANDELS GOODS-S field is a deep, multi-wavelength field in which 27 CCPC structures have been identified. Of these, 9 were identified in Franck & McGaugh (2016b) (in the redshift range $2.74 < z < 3.71$). It is the deepest, widest field from which we draw

candidate spectroscopic galaxies to identify protoclusters. The majority of the spectroscopic footprint falls on an area of the sky of roughly 0.4 degrees on a side, which corresponds to a box with sides approximately 35 cMpc wide at $z = 2$. At $z = 5.7$, the box side length increases to $\sim$ 55 cMpc in our assumed cosmology, with a length along the line of sight ($z = 2 \rightarrow 5.7$) of almost 3 cGpc. This length is the expanse of the field in which we identify protoclusters.

From this pencil beam survey, we estimate the volume that we have probed to be $\sim 5.89 \times 10^6$ cMpc^3. The Millennium Simulation has a volume of $[500\,h^{-1}\,cMpc]^3$ (Springel et al., 2005), which is almost 2 orders of magnitude larger than the GOODS-S data. Chiang et al. (2013) identified 2832 clusters at $z = 0$ with masses $M > 10^{14}\ h^{-1}\ M_\odot$ within their analysis of the simulation. This corresponds to a number density of 7.8×10^{-6} cMpc^{-3}. Therefore, one would expect to find 46 clusters at $z = 0$ in the volume of the GOODS-S field. As we have only identified 27 systems in this region, our algorithm is probably not over-identifying structure. Using the number density of low mass clusters ($M_{z=0} < 3 \times 10^{14}$ $M_\odot$) in Millennium, there should be an estimated 32 such systems in this deep field (Chiang et al., 2013), while 14 Virgo-mass or larger ($M_{z=0} \geq 3 \times 10^{14}\ M_\odot$) protoclusters are expected within this volume. It is probable that our algorithm is only identifying some of the richest overdensities that will collapse into Virgo or larger mass systems at $z = 0$, while missing many of the smaller protoclusters in this field.

In all cases of protocluster candidates within the GOODS-S field, the probability that these objects will collapse into clusters by $z = 0$, based on the overdensities of Chiang et al. (2013), is considerably less than 100% (see Table 3 in the online version). If we sum their fractional collapse estimates (e.g. a candidate's 10% chance of collapse can be approximated as $\frac{1}{10}$ of a cluster), for the entire CCPC we have identified only $\sim$ 3.7 clusters from $2 < z < 5.7$ in the CANDELS GOODS-S survey. As we expect more than

40 structures in this probed volume, there exists a serious discrepancy. One explanation could be that our conservative estimates for both δ_{gal} and its application in determining a collapse probability from the work of Chiang et al. (2013) is much too stringent. This is a probable scenario, as explained in the outline of overdensities and probabilities within Section 3.1. Another option could be that the number density estimates are for all clusters, while we are mainly identifying higher mass systems. However, 14 Virgo-like protoclusters are expected in this volume, and our probabilities are an order of magnitude smaller than that. It is also plausible that CANDELS GOODS-S field is not accurately represented within the Millennium simulation. Indeed, a combination of these elements are probable to span the gap of the excess number of simulated systems.

As mentioned earlier, many of these deep, pencil-beam surveys from which we identify protoclusters have sky widths smaller than the expected size ($\sim$ 20 cMpc) of the most massive protoclusters (Muldrew et al., 2015). Therefore, if the overdensity does not significantly fall within the footprint of the spectroscopic data, it may be missed by our search method. Intriguingly, Kang & Im (2015) delved into both the GOODS-S and GOODS-N combined spectroscopic and photometric data and found an excess of structures in the redshift range $0.6 \leq z \leq 4.5$ when compared with the Millennium Simulation protoclusters. While in the CCPC we have limited our comparison to the number density, they included the mass of their structures, finding a factor of 5 more systems with $M > 7 \times 10^{13}\ M_{\odot}$. Kang & Im (2015) provide some plausible explanations for the overabundance (elements of the input physics in the models could be simply incorrect, ΛCDM may not be an accurate representation of the universe, overestimation of masses), but no definitive diagnosis for this complex problem.

That we derive the opposite conclusion from the same data set, similar search methodology, and identical simulation is puzzling. It could be possible that the precision of pho-

tometric redshift measurements produces an increase in false-positive structure detections. This seems unlikely, as we found more overdensities (24 versus 9) in the shared redshift space with Kang & Im (2015) in the GOODS-S field. We find it more likely that the authors' method is sufficiently distinct from our own so as to be difficult to compare results.

3.2.2 Poissonian Expectation Model

A further confirmation that these associations are more than chance groups of galaxies can be found by estimating the number of false positives that should be expected in a smooth density field (i.e. lacking structure) along the line of sight that could arise from Poissonian fluctuations. We can utilize the analytical formula

$$N_M = N_g e^{-2nl}[1 - e^{-nl}]^{M-1} \tag{3.3}$$

from they toy model of Sheth (2001), which approximates the number (N) of Poissonian fluctuations of M galaxies that would manifest as protoclusters. This model uses the number density along the line of sight, n, for N_{gal} total galaxies, separated by a linking length l. The CANDELS GOODS-S field is approximately a 1-D pencil beam survey in the context of this work ($\sim 40 \times 40 \times 3000$ cMpc from $2 < z < 5.7$). The number density of spectroscopic sources varies from $2 < z < 5.7$, which makes it a linearly decreasing density field as a function of redshift, unlike the toy model. At $2 < z < 3$, $n = 0.76$ galaxies cMpc^{-1}, while over the full range ($2 < z < 5.7$), this decreases to $n = 0.40$ galaxies cMpc^{-1}. We adopt the latter value ($n = 0.40$ galaxies cMpc^{-1}) as the estimate of the number density.

We require a suitable estimate of the linking length l between galaxies in this survey. We can measure the strength of clustering at different length scales by using the auto-correlation function (Davis & Peebles, 1983).

$$\xi(l) = \frac{N_{DD}}{N_{DR}} \frac{n_R}{n_D} - 1. \tag{3.4}$$

This compares the number of pairs of actual galaxies (N_{DD}) found within shells separated by $l \pm \Delta l$ in the GOODS-S field, to the number of *D*ata-*R*andom pairs of galaxies (N_{DR}). When the value of $\xi(l)$ crosses $\xi = 1$, the clustering strength at length l is said to transition from strong to weak. At $l = 1.1 - 1.75$ this occurs within the spectroscopic data set. This is similar to the linking length measured between $2.74 < z < 3.71$ found in CCPC1 (Franck & McGaugh, 2016b).

The CCPC algorithm does not utilize linking lengths to identify structure, as l would vary from survey to survey. To similarly match the fluctuations in the toy model of Sheth (2001), we identified structures using a one-dimensional Friends-of-Friends (FoF) algorithm (Huchra & Geller, 1982). We also ran 500 Monte Carlo simulations of the field and computed the mean number of groups found using the FoF algorithm as a check on the toy model. For $l = 1.2 - 1.75$ cMpc, an excess number of groups are identified above the Poissonian expectation for any choice of the minimum member galaxies ($M \geq 2$). There were 15 FoF structures found within the CANDELS data set using separation $l = 1.2$ cMpc with $M \geq 6$ galaxies and 6 associations of $M \geq 8$ members. The toy model predicts 4 and 0.6 such groups of galaxies should exist (respectively) from Poissonian fluctuations, while the Monte Carlo finds a mean of 1.4 and 0.08 systems of such richness. Choosing a larger separation of $l = 1.75$ cMpc reveals an excess of 21 FoF groups over the expectation of Sheth (2001) for associations of 6 or more members, and an excess of 25 compared to the Monte Carlo value. For groups of 10 or more galaxies with separation of $l = 1.75$ cMpc, there should be fewer than one chance fluctuation in the Poissonian and Monte Carlo models, while 5 systems are identified in the GOODS-S volume.

The clear excess of FoF groups using a range of values for l and M compared to what would be expected stochastically illustrates that there are physical associations within these data. Values of N_m and l are unique to this survey and do not apply to CCPC protoclusters

outside of the CANDELS GOODS-S sample. This example is only used to illustrate that an excess of clustering is found over Poissonian fluctuations in a simple toy model.

3.2.3 Mass Estimates

In Sec 3.1, we outlined two distinct methods to estimate the mass of protoclusters at these high redshifts. Unlike the methods utilized for 'nearby' clusters at $z \leq 1$ which use signatures from a massive $M > 10^{14}\ M_{\odot}$, virialized halo (e.g. SZ effect, X-ray emission), protoclusters at $z \geq 2$ are not expected to have main halos of this magnitude (Chiang et al., 2013; Muldrew et al., 2015). More uncertain means must be employed to provide some metric of the mass of these systems.

One method used by Steidel et al. (1998) and Venemans et al. (2002) is to calculate the volume of the overdensity that will collapse into a cluster at $z = 0$. The mass is simply the density of the Universe (ρ_{crit}) multiplied by the mass overdensity ($\delta_m = \delta_{gal}/b$) and volume of the system (Equation 1). The mean overdensity mass is $M_{z=0} \sim 1.8 \times 10^{14}\ M_{\odot}$ for the CCPC2 catalog. As the majority of clusters ($\sim 70\%$) in the Universe are expected to be of this mass (Chiang et al., 2013), this appears reasonable. We have also made the most conservative estimates of the volume and overdensities (δ_m) which would enhance this expectation.

There does not appear to be a significant trend in decreasing mass with increasing redshift. There is considerable scatter in the $\delta_{gal} - z$ distribution, which subsequently persists into the mass estimates. Most of the sources with the highest probability of collapsing into a structure (Tab 3.2) are at low redshift and can have masses in excess of $M_{z=0} > 5 \times 10^{14}\ M_{\odot}$. A large mass estimate for CCPC-z65-005, the highest redshift system, is an example of the lack of mass evolution. Located at a redshift of $z = 6.56$, its large volume (~ 7000 cMpc^3) and overdensity imply a collapsed mass of nearly $M_{z=0} \approx 6 \times 10^{14}\ M_{\odot}$. The CCPC1

has a mean mass estimate of 2.5×10^{14} $M_{\odot}$, only slightly larger than CCPC2.

It should be noted that the δ_{gal} estimator relies on a number of assumptions. Primarily, the tracer of both the volume and matter overdensity are galaxies, which are not a significant contributor to the density of the Universe in ΛCDM . We must assume that all galaxies, especially those at the outskirts that define the volume of the overdensity, will collapse by $z = 0$. Numerous simulations have shown that many galaxies within the comoving volume of the structure may not be bound to the cluster by $z = 0$ (Muldrew et al., 2015). They also lack physical properties (e.g. enhanced SFRs, stellar colors, etc.) that could distinguish them as outliers observationally in the SAMs of Contini et al. (2016). We also must adopt a linear bias parameter b that translates a galaxy overdensity into a mass overdensity ($\delta_m = \delta_{gal}/b$). This bias value is assumed to be in the range of 3-6 at these redshifts (Steidel et al., 1998), and is dependent on the types of galaxies observed. Furthermore, the values of δ_{gal} are highly dependent on how the field is defined and the scales at which the overdensity exists, as we have pointed out here and in CCPC1.

The 'virial' mass estimate also has a number of systematic uncertainties. These systems are the diffuse, primordial manifestations of clusters and are not expected to be in equilibrium in ΛCDM . In the Millennium simulation, Chiang et al. (2013) found that the first $M \geq 10^{14}$ $M_{\odot}$ halos are not present until $z \leq 2.3$, and may not be virialized until a dynamical time later. However, there are some indications that subhalos in protoclusters can be virialized, as observed by Venemans et al. (2007); Shimakawa et al. (2014) and Topping et al. (2016). These subhalos can be seen as bimodal distributions of protocluster galaxies along the LOS. Intriguingly, (Wang et al., 2016) recently identified a cluster core emitting X-rays at $z = 2.5$. The cosmological implications of this discovery are uncertain.

Virial equilibrium is unlikely to hold, but we can query the data to see what mass is implied. Equation 2 is the familiar virial mass estimate that only requires an effective radius

($R_{hms} = 1.25R_e$) and velocity dispersion (σ) for a system in equilibrium. Based on the position of galaxies in the plane of the sky, we have calculated the radius in which 50% of the member galaxies can be found. The mean R_e is 2.8 ± 1.0 (physical) Mpc. The relatively small number of galaxies (minimum of 4) in our criteria is a further source of significant uncertainty, in that values of R_e can be based on only two systems. The dispersion velocity is simply the standard deviation from the mean redshift of the system, and is on average 653 km s^{-1}. This is typical of dispersions found for other protoclusters, as compiled in Table 5 by Chiang et al. (2013). We note that the few galaxy members involved in the calculation of the dispersion of these systems could introduce a large bias. Some candidates with $N = 4$ members have σ_z values based on only two galaxies. Richness alone does not appear to be a significant driver, however. The average velocity dispersion for systems with $N \geq 23$ members is 661 km s^{-1}, the same as the mean value of the entire group.

Interestingly, these large velocity dispersions, like those at $z \geq 3$ as found in Venemans et al. (2007), CCPC1, and here, appear to be larger than those predicted by simulations. Chiang et al. (2013) report that dispersions along the line of sight for overdensities at $z = 3$ in the Millennium Simulation are 400 ± 60 km s^{-1} for the progenitors of $M_{z=0} \geq 10^{15}$ $M_\odot$ clusters. Our typical redshift uncertainty ($\sigma \sim 0.001$) represents a velocity uncertainty of 75 km s^{-1} (at $z = 3$), which cannot account for the +100 km s^{-1} dispersion excess with respect to the simulated systems. Similarly, Venemans (2005) compared the protocluster dispersions in their sample to simulated dark matter halos of clusters within a similar window size. For $\sigma_{z=0} = 1000$ km s^{-1} systems (e.g. the Coma cluster), the dark matter velocity dispersion was found to be systematically lower than what was observed in the protoclusters at nearly all redshifts sampled.

It could be possible that galaxies not bound to the protocluster, or those at the outskirts, are boosting the dispersion significantly above expectations. Contini et al. (2016) plot the

velocity field for simulated protocluster galaxies in their Fig 2 for a system with $M_{z=0} \geq 10^{15}\ M_{\odot}$. At $z = 3$, the members near the central galaxy ($R < 5$ cMpc) have velocities of 200-400 km s^{-1}. Galaxies more than 20 cMpc away can have velocities in excess of 1000 km s^{-1} (Contini et al., 2016). Survey sizes within the CCPC1 and 2 vary in both width and depth, yet this phenomenon persists. In our smallest survey volumes, the inner regions should contain the brightest, most massive galaxies and the velocities are expected to be the smallest (Contini et al., 2016). Greater spectroscopic completeness within some of the richest overdensities might shed light on this mystery. This would allow a more apt comparison to simulated protoclusters, which 'observe' all galaxies and not just the brightest sources.

As an attempt to mitigate the weight of outliers for small numbers of galaxies when calculating a velocity dispersion for protoclusters, Venemans et al. (2007) utilized the biweight location estimator (Beers et al., 1990). The dispersions discussed previously, both here and in CCPC1, were simple standard deviations. We employed this biweight estimator in the AstroPy package (Astropy Collaboration et al., 2013) and iteratively solved for the velocity dispersion for each CCPC system. The median value of these biweight dispersions across the CCPC2 catalog is 717 km s^{-1}, which is $< 1\sigma$ larger than a simple median value of all dispersions (677 km s^{-1}). Outlier galaxies do not appear to be biasing our results significantly.

We note that unlike CCPC1, this work did not require a knot of ≥ 3 galaxies to exist near the central search regions. This had the occasional effect of 'centering' the galaxy distribution. For CCPC2, we sought only to maximize the number of galaxy members within the search radius of $R \leq 20$ cMpc. Therefore, some R_e values could possibly be reduced if a different protocluster center were chosen (as in CCPC-z28-017), whereas others would be unaffected by this method change. However, for the bulk of the CCPC2 population, this is

unimportant. The mean R_e = 2.2 Mpc for CCPC1 is only marginally smaller than CCPC2 and within its standard deviation. The velocity dispersions are essentially equivalent for CCPC1 and 2 as well (668 and 653 km s^{-1}, respectively).

By limiting the distance we probe for structure along the line of sight, the calculated velocity dispersions are decreased, as one would expect. For example, restricting the search redshift range to half its original length ($\Delta z \pm 10$ cMpc), only 32 protocluster candidates exceed the 400±60 value for the expected dispersion of a $M_{z=0} \geq 10^{15}\ M_{\odot}$ cluster at $z = 3$ (Chiang et al., 2013). These anomalous systems also have only ~ 6 galaxies, on average. To establish an expectation of the role window size affects the observed dispersion, we designed a Monte Carlo simulation of a protocluster at $z = 3$. $N = 6$, 10, and 200 galaxies were randomly distributed 10^4 times with varying LOS distances ($1 < d < 20$ cMpc). We measured the mean dispersion and standard deviation of the simulation for each LOS distance. The Monte Carlo produces a linear relationship between LOS window size and velocity dispersion. When compared to the CCPC2 dispersion values in different window sizes, as well as the expectation value from the Millennium simulation (Chiang et al., 2013), the Monte Carlo results were not distinct at a statistically significant level. These results suggest that the velocity dispersion excess could merely be the result of window size rather than a physical characteristic of the system.

The Monte Carlo simulation is only an effective modelling tool if we assume the correct, underlying distribution of galaxies and their velocities. For instance, we can make a simple assumption that all galaxies in the simulation have a true distance corresponding to $z = 3$, and then imprint a Gaussian distribution of velocities of $\sigma = 400$ km s^{-1}. These results can fit the data point of CCPC2 with a window size of ±10 cMpc, as well as the expectation value of Chiang et al. (2013). We can modify the simulation again by allowing galaxies to be normally distributed within a specified radius, and then applying a random velocity

(a pseudo-virial distribution) to each one individually. This can provide a fit to the CCPC2 for the full window size of the redshift distribution, but no other data points. Applying only infalling velocities to spatially dispersed galaxies can also match the observations. These examples are meant to illustrate the underlying degeneracy of the simulation and the complexity of the problem. Fundamentally, the menagerie of simulated protocluster sizes, velocity fields, and evolutionary states (Chiang et al., 2013; Muldrew et al., 2015; Contini et al., 2016) make the prospect of a direct comparison of dispersions to real data a daunting task, even at a specific redshift and structure mass.

The average virial mass estimate is $M = 8 \times 10^{14}\ M_{\odot}$ in CCPC2. Included are 61 structures (plus 13 in CCPC1) that have masses $M > \times 10^{15}\ M_{\odot}$ at $z > 2$. The Millennium simulation analysis by Chiang et al. (2013) suggest that for the single most massive halo in a protocluster, none should have masses in excess of $\sim 10^{14}$ prior to $z = 2.3$. Likewise, Mortonson et al. (2011) predict that at $z \geq 2$, there should not be a single collapsed structure in a ΛCDM universe with $M > 6 \times 10^{14}\ h^{-1}\ M_{\odot}$. If any of these systems are in virial equilibrium, it could pose a serious challenge to the concordance cosmology.

Chiang et al. (2013) also computed the radii (in comoving units) for the main halos in protoclusters at redshifts $z = 2 - 5$ (their Fig 2). As noted earlier, our search algorithm was not designed to minimize the R_e of candidate systems, and so the radii listed in Table 3.2 may be an overestimate. Subsequently, these large radii could bias our anomalously large virial mass estimates. To test this, we limited our entire sample to those with R_e values less than the expected range for the most massive protoclusters (Chiang et al., 2013). There were 66 such sources, with an average radius of 6.5 cMpc and redshift of $z \sim 3.4$ (but spread between $2.00 < z < 6.5$). The average mass of these systems is $M_{vir} = 2.7 \times 10^{14}$ $M_{\odot}$. At $z = 2.0$, this is approximately the mass expected for the largest main halos of the universe at this epoch. Observational mass estimates for structures at $z \sim 2$ from SZ/X-ray

emission are consistent with this result (Gobat et al., 2011; Mantz et al., 2014). However, at larger redshifts, especially those at $z \geq 6$, the discrepancy is on the order of 10^2 from the predictions of Chiang et al. (2013). For at least this subsample, the radii do not appear to be the aberrant property in determining the mass.

It would seem that these discrepancies can be directly attributed to the inherent uncertainties in the velocity dispersions. Gobat et al. (2011) identified a (proto)cluster at $z \sim 2$ (identified here as CCPC-z20-002) and calculated its mass via X-Ray emission to be $M_{X-ray} = 5 - 8 \times 10^{13}$. In a follow-up spectroscopic study of the galaxies in the system, they obtained a velocity dispersion of 1300 km s^{-1} (Gobat et al., 2013), which would suggest a much larger mass than inferred by the emission from the intracluster medium. Considerable work needs to be done to establish reliable methods for estimating protocluster masses at high redshift.

The overdensity and virial mass estimates (Equations 1 and 2, respectively) do not correlate with one another in the CCPC. The two estimators are ostensibly measuring two different properties, so taken at face value this lack of a link is not too surprising. The overdensity method is attempting to quantify the amount of mass that will collapse at $z = 0$, while the virial estimate is a representation of the current mass of the system in dynamical equilibrium. Despite the fact that even cluster progenitors of the same $z = 0$ mass can exist in a variety of evolutionary stages at $z \geq 2$ (Chiang et al., 2013; Muldrew et al., 2015), one would expect some relationship to exist. The conclusion that can be drawn from this comparison, in addition to the previous paragraphs discussing both estimators, is that these mass values are highly uncertain; no reliable mass indicator is available at present. While each may represent some aspect of the physical nature for these systems, we caution the reader not to rely on these values.

3.2.4 Protocluster Groups

CCPC-z23-001 and CCPC-z23-005: These overdensities were identified as two distinct systems by our search algorithm. After inspection, it was realized that these objects appear to be a single, extended system at $z = 2.33$. Fig 3.5 shows that the galaxy footprints overlap one another on the sky. However, the line-of-sight distribution of the galaxies appear to show a continuous distribution of sources of roughly 70 cMpc in length. The $N(z)$ plot exhibits a single, wide spike in galaxy counts (i.e. not a bimodal distribution). This spike was also identified using a combined catalog of primarily photometric with some spectroscopic galaxies by Salimbeni et al. (2009) as well as Kang & Im (2015). On the sky, the system is approximately 22×31 cMpc in RA/DEC, respectively. This is a large structure, but it is not unique. For example, Lee et al. (2014) found three overdensities at $z = 3.78$ within a $75 \times 75 \times 25$ cMpc^3 volume. Dey et al. (2016) provided follow-up spectroscopic coverage to that extended system, which may be part of a filament stretching $\sim$ 170 cMpc. More recently, Zheng et al. (2016) identified four protoclusters of LAEs with volumes in excess of $[15cMpc]^3$ around an overdensity discovered by Balestra et al. (2010) and listed in the CCPC1 as CCPC-z28-002.

The galaxy overdensities of CCPC-z23-001 and CCPC-z23-005 are relatively modest ($\delta_{gal} = 2$, 0.7 respectively). However, this can be partially attributed to the strong 'field' counts from the neighboring system. CCPC-z23-001 is one of the richest systems in the combined CCPC ($N = 43$), while together they boast 66 member galaxies. These galaxies fill a volume of 4.6×10^4 cMpc^3 and have a combined overdensity of $\delta_m = 0.52$. To compute the mass overdensity, we found the number density of galaxies within $z = 2.329 \pm 0.03$ and a field length $\Delta z = 0.15$. The volume is simply the rectangular region that encapsulates all galaxy members multiplied by the length of the box in comoving units. This volume and

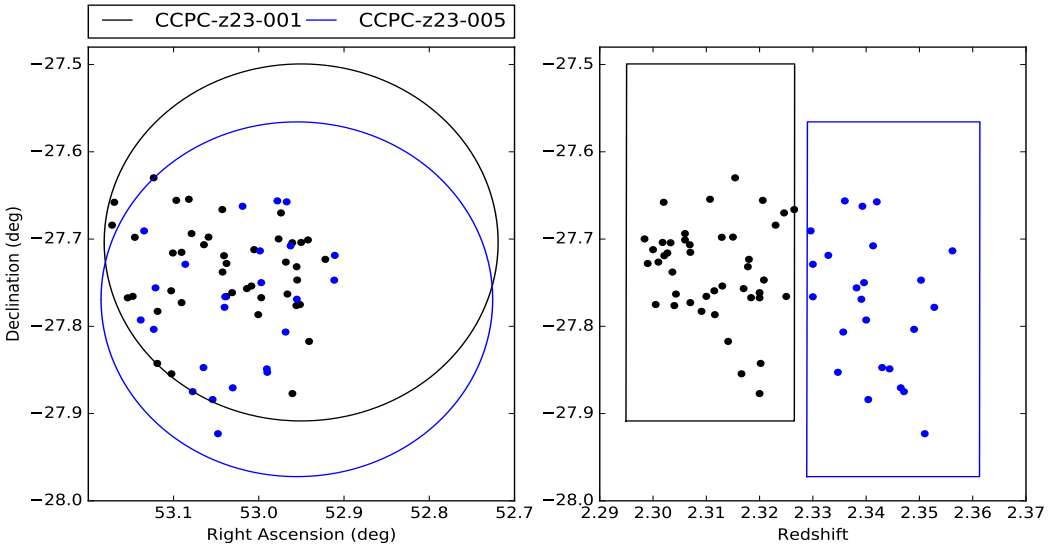

Figure 3.5: At $z \sim 2.33$, our search algorithm selected the two galaxy associations (CCPC-z23-001 and CCPC-z23-005) as being strong protocluster candidates. *Left* : Sky plot showing the overlap of the two structures on the sky, with the ellipses illustrating the boundary of our search volume ($R = 20$ cMpc). *Right* : Line-of-sight distribution of galaxies in the two candidates, which appear as a single, unbroken distribution. The boxes are the boundaries of Δz corresponding to ±20 cMpc. Further analysis shows that these objects appear as one continuous structure of volume 4.6×10^4 cMpc^3. The comoving volume is calculated by assuming a box with a length corresponding to the minimum and maximum redshifts, and width bounded by the galaxy positions in the sky plot (*Left* panel). There is also no break in the $N(z)$ distribution that would indicate two separate structures. A mass estimate based on its overdensity is 3×10^{15} $M_\odot$, which would be one of the largest structures known.

overdensity imply a mass of 3×10^{15} $M_{\odot}$ if the superstructure were to collapse. However, this large of a volume is not expected to become one system by $z = 0$ in a ΛCDM Universe. Within this volume, assuming a number density of clusters to be 7.8×10^{-6} cMpc^{-3} (Chiang et al., 2013), we would expect 0.3, $M \geq 10^{14}$ $M_{\odot}$ clusters to be found. If they are each $\geq 10^{15}$ $M_{\odot}$ systems, as their overdensity and volume implies, the cluster number density decreases to 1.6×10^{-7} cMpc^{-3}. Within the estimated 5.893×10^{6} cMpc^{3} volume in which we identify structure in the GOODS-S field ($2 < z < 5.7$), we would expect to find fewer than one $\geq 10^{15}$ $M_{\odot}$ protocluster. That we find two systems in a volume 8×10^{-3} smaller is intriguing. The nature and fate of this system(s) is not yet understood.

CCPC-z65-004 and CCPC-z65-005: This object is similar to the previous example in that the algorithm detected these two objects as separate sources. However, the redshift at which this object is found makes it even more interesting. Along the line of sight, these two systems essentially overlap (right panel of Fig 3.6), are much shorter than the $\Delta z = 40$ cMpc search length (< 20 cMpc), and are thin. On the sky, their respective search centers are merely 40 cMpc offset from one another. This could mean that this is a single, very massive protocluster ($\delta_{gal} \sim 4$, $M > 2 \times 10^{15}$ $M_{\odot}$) with a geometric center in between the two coordinates listed. Of the two other $z \sim 6$ protoclusters known (Utsumi et al., 2010; Toshikawa et al., 2014), this combined system has an overdensity mass estimate at least 5 times larger.

Regardless of whether CCPC-z65-004 and CCPC-z65-005 are a single system or two separate, high mass protoclusters, this detection represents the highest redshift association of galaxies that has been spectroscopically confirmed to the best of our knowledge. Trenti et al. (2012) and Ishigaki et al. (2016) have both identified protocluster candidates at $z \sim 8$ based on strong overdensities of Y-dropout galaxies, but these have yet to be spectroscopically confirmed.

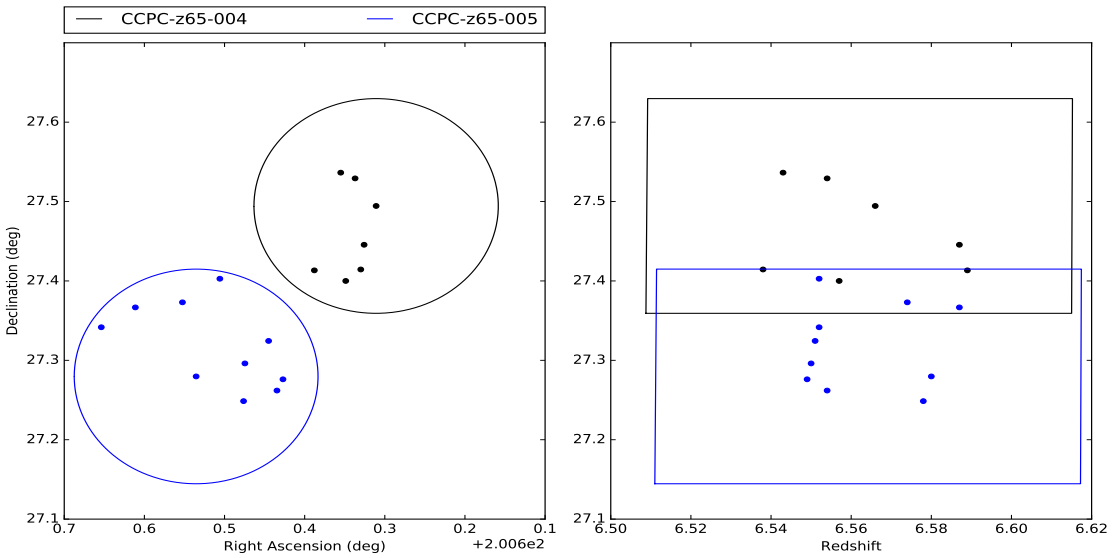

Figure 3.6: CCPC-z65-004 and CCPC-z65-005 are two systems separated by ∼40 cMpc on the sky (*Left* panel), with their search volumes (ellipses) nearly touching. Indeed, a search center located at the midpoint (RA: 201 *deg*, DEC: +27.4 *deg*) would capture the majority of galaxies in the two distributions. In the *Right* panel, the Δz plots corresponding to a length of 40 cMpc, almost perfectly overlap, showing the galaxies along the line of sight. These could be two distinct associations, or one system with a geometric center in between the two groups. The mass of this system is estimated to be $M > 2 \times 10^{15}\ M_{\odot}$. These two (sub)protoclusters are the highest spectroscopically identified protoclusters at the time of writing.

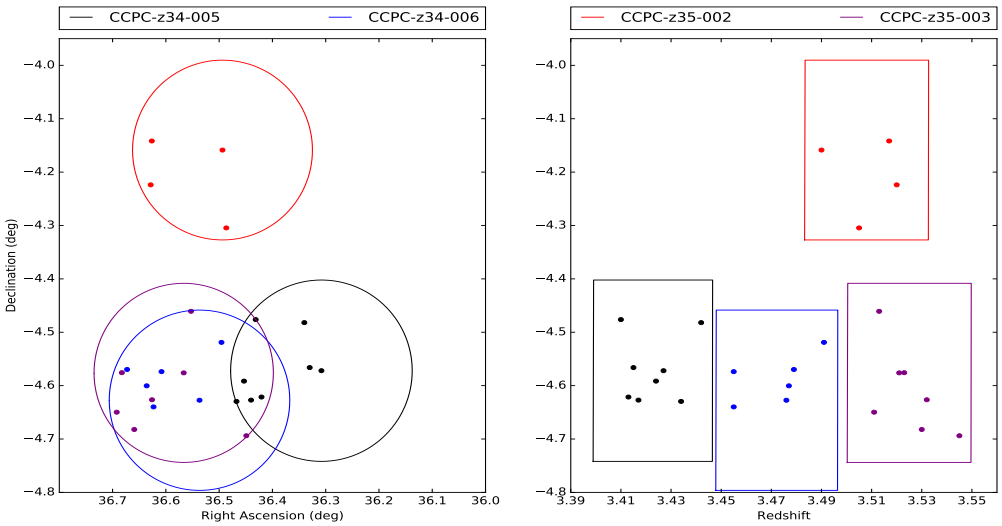

Figure 3.7: A group of 4 protocluster candidates at $z = 3.5$ that are proximate in space. The $Left$ panel is the sky association of the candidates, while the *Right* panel shows their galaxy distribution along the line of sight. The ellipses/boxes show the search volume boundaries ($R_{search} = 20$ cMpc, $\Delta z \pm 20$ cMpc). Three of these (CCPC-z34-005, CCPC-z34-006, CCPC-z35-003) exist in a chain along the LOS stretching ≤ 120 cMpc. This may become a supercluster-sized structure at $z = 0$.

CCPC-z34-005, CCPC-z34-006, CCPC-z35-002, and CCPC-z35-003 Complex: These overdensities, although individually unremarkable, are part of a linked superstructure. These four separate systems almost touch on the sky in Fig 3.7. CCPC-z34-005, CCPC-z34-006, and CCPC-z35-003 all exist along the same line of sight in a chain of ≤ 120 cMpc in length. The center of CCPC-z35-002 is separated from the chain by approximately 40 cMpc. This complex may be a proto-supercluster in the process of assembly.

There are a number of other systems in CCPC2 that may be associated with one another. CCPC-z27-006 and CCPC-z27-010 appear to be nearly touching along the LOS and overlap one another on the sky. They appear similar to CCPC-z23-001 and CCPC-z23-005 (as seen in Fig 3.5), but are not nearly as rich. CCPC-z28-015 and CCPC-z28-016, and CCPC-z45-001-CCPC-z45-002 are more similar in nature to the distribution of galaxies in Fig 3.6 with similar redshifts but separated on the sky by roughly the search radius of the algorithm. However, the associations are not as strong as that of CCPC-z65-004-CCPC-z65-005. Nevertheless, it appears that it is not uncommon for protocluster candidates to reside in very large scale associations.

We emphasize again that the volumes estimated in this manner are highly uncertain. Candidate member galaxies on the outskirts of a distribution, whether projected on the sky or along the line of sight, can significantly enhance the volume implied. These volumetric tracers may not be bound to the structure at all, or may be sufficiently separated that they will not reside within the cluster's halo at $z = 0$.

3.3 Summary

We have extended the Candidate Cluster and Protocluster Catalog to redshifts $2 < z < 6.6$ in CCPC2, adding 173 protocluster candidates to the 43 in CCPC1. In the CCPC2, we

identified galaxy overdensities ($\delta_{gal} > 0.25$) of 4 or more galaxies within a search radius of $R = 20$ cMpc and within a Δz of ± 20 cMpc. In Table 3 of Franck & McGaugh (2016a), all candidate protoclusters are listed. The 36 systems that have the largest collapse probabilities ($\delta_{gal} > 6$) are found in Table 3.1. The median number of galaxy members is 6 in CCPC2, and 9 candidates have $N > 23$ galaxies. Above a redshift of $z > 4$, we have identified 40 structures. Prior to this work, fewer than 10 had been identified. At the time of writing, this list includes the most distant spectroscopically-confirmed protocluster known (CCPC-z65-005). The combined CCPC is the largest known list of high redshift, spectroscopic protoclusters to date.

Following the examples of Steidel et al. (1998) and Venemans et al. (2002), we estimate the strength of these protoclusters by computing their overdensities with respect to field counts of galaxies at similar redshifts. These structures contain a median overdensity of $\delta_{gal} \sim 2.9$, slightly larger than the median value found in CCPC1 and typical of the overdensities found in the literature (summarized in Table 5 of Chiang et al., 2013), as well as expectations from simulations. These overdensities and their visual counterparts ($N(z)$ plots for each CCPC member in the Appendix) are two pieces of evidence suggesting that these galaxy associations are indeed structures and not merely coincident on the sky. We emphasize that these overdensities are conservative estimates (Fig 3.1).

In Section 5.4, we compare the spatial distribution and surface densities of CCPC systems to field galaxies. Using the list of $\geq 40,000$ galaxies from which we identify overdensities, we built the mean surface density of 'All galaxies' as a proxy for the field in Fig 3.3. Compared to the mean distribution of CCPC galaxies out to $R = 20$ cMpc, there are significantly fewer galaxies at all radii in the field. A KS test of these two distributions shows a statistically significant difference (KS= 0.60). Fig 3.4 illustrate a similar trend in the total number of galaxies as a function of the distance from search center. There are fewer field

galaxies, at all redshifts and radii, than in CCPC systems. We estimate the expected number of protoclusters that could be found within the CANDELS GOODS-S pencil beam survey volume based on the number density of clusters in Millennium (Chiang et al., 2013). In CCPC1 and CCPC2, we found 27 candidate structures, whereas 46 clusters of $M \geq 10^{14}$ $M_{\odot}$ were expected in a similar volume of the ΛCDM N-body simulation. This suggests that we are not over-identifying structure using our method, and may be recovering only the most significant overdensities. For systems of Virgo-mass and larger ($M \geq 3 \times 10^{14}$ $M_{\odot}$), 14 are expected in the GOODS-S volume.

We apply an analytic toy model by Sheth (2001) to estimate the expected number of galaxy groups stochastically produced in a smooth density field. We then compare this expectation value to the number of FoF groups identified in CANDELS GOODS-S survey and found an excess of FoF groups compared to what would be expected stochastically from Poissonian fluctuations and Monte Carlo simulations. This test furthers the notion that these are physical structures and not mere chance overdensities.

These tests were to ensure that the structures identified with the heterogeneous data available were legitimate structures. We have shown that: (1) these objects exist as galaxy overdensities (median $\delta_{gal} \sim 2.9$), (2) have 4× larger number densities than LAEs (Section 3.1), (3) differ spatially from field spectroscopic catalogs in Section 5.4, (4) are not the result of Poissonian fluctuations in a smooth density field (Sheth, 2001), and (5) we do not over-identify structures in the deepest survey used (CANDELS GOODS-S). It is statistically unlikely that these tests, when taken together, would fail to distinguish the field from the candidates if a significant number of these systems are not actual structures.

For each CCPC member, we estimated the mass of the structure using two distinct methods. The first technique uses a linear bias parameter b to transform the galaxy overdensity δ_{gal} into a mass overdensity δ_m. If one assumes that the volume (V) traced by galaxies will

collapse to a cluster, the mass of the system at $z = 0$ can be approximated by the product of the volume and mass density (Steidel et al., 1998). The average overdensity mass estimate for the CCPC2 is 1.8×10^{14} $M_{\odot}$. We also computed the virial mass for each system based on its effective radius R_e and velocity dispersion σ, making the crude assumption that these systems are in virial equilibrium. The mean 'virial' mass is $M = 8.4 \times 10^{14}$ $M_{\odot}$ calculated for all CCPC2 members. With the combined CCPC1 and CCPC2 catalog, a total of 74 structures have mass estimates of $M \geq 10^{15}$ $M_{\odot}$. Such large masses are not expected at redshifts $z > 2.3$ in N-body simulations (Chiang et al., 2013). There is little agreement between the volumetric and virial mass estimators which emphasizes their uncertainty and the dubiousness of the necessary assumptions (e.g. virial equilibrium).

There does appear to be a discrepancy between the observed and predicted velocity dispersions (Chiang et al., 2013; Contini et al., 2016). The median value is $\sigma \geq 650$ km s^{-1} in both CCPC1 and CCPC2, with some systems having dispersions as large as 900 km s^{-1}, here and in Venemans et al. (2007). This is more than double the expected value at the relevant redshifts for the most massive protoclusters (400 ± 60 km s^{-1} in Chiang et al., 2013). This is important because no assumption of equilibrium is made. However, it may not be possible to make an apples-to-apples comparison, as objects near the edge of the protocluster have large ($\sim$ 1000 km s^{-1}) simulated velocities (Contini et al., 2016). By reducing the line-of-sight window size to $\pm$10 cMpc generally removes this observed excess to expected values. We note that it is difficult to map 'observations' of simulations to real data. The observed excess of velocity dispersion in the CCPC is most likely not a physically-relevant result.

The CCPC2 also has three groups of protoclusters (Fig 3.5, Fig 3.6, and Fig 3.7) that may be primordial superclusters based on their small ($\leq$ 100 cMpc) separations. CCPC-z23-001 and CCPC-z23-005 appear to be physically connected in a chain of length $\sim$ 70

cMpc. Its nature is not understood currently, as it appears to be a single, extended structure of volume $\geq 4 \times 10^4$ cMpc^3 and mass $\geq 10^{15}$ $M_\odot$. CCPC-z65-004 and CCPC-z65-005 are the two highest redshift spectroscopically confirmed protoclusters known to date, and are separated by a mere 40 cMpc. They may be a single, massive structure with a geometric center offset from their galaxy distributions, or two disparate protoclusters very near to one another. CCPC-z34-005, CCPC-z34-006, CCPC-z35-002, and CCPC-z35-003 form a protocluster complex, with three existing in a chain along the line of sight.

In total, we have identified 173 protocluster candidates that appear to be genuine, spectroscopically confirmed, physical associations of galaxies at high redshift. Some of these systems reside in close proximity, as if part of (proto) super-clusters. Candidate protoclusters of high confidence are found up to $z \approx 6.5$. There appears to be a rich amount of structure still to be revealed in the high redshift universe.

Chapter 4

The Candidate Cluster and Protocluster As Seen by Spitzer

The nascent study of protoclusters is at a juncture of two important evolutionary epochs in the universe: the early growth of large structures and the rapid assembly of galaxy mass at $z \geq 2$. Careful study of these objects can probe both cosmology and galaxy formation and evolution. The initial mass overdensities in the early universe (and their subsequent collapse) are governed by the cosmic matter density ($\Omega_{\rm m}$), σ_8, and the cosmological constant Ω_Λ. Thus, by examining their properties (mass, evolutionary state) and number density of these early structures, they can provide constraints on these cosmological parameters. The tracers of these overdense structures (i.e., galaxies) can be used to investigate the role environment plays in their growth and evolution. Formation models of galaxies must match observations across both cosmic time and throughout space to be considered viable. More simply, void and cluster galaxies must both be reproduced for all $z \geq 0$. The focus of this work will be primarily on the properties of galaxies in dense environments, but this juncture of structure and galaxy evolution are clearly relatable in many ways.

It has been clear for many decades that galaxies at $z = 0$ have varied properties that cor-

relate strongly with density in the well-established morphology-density relation (Dressler, 1980; Baldry et al., 2004). Galaxies in dense environments show clear evolution at higher redshift (Butcher & Oemler, 1984), in that there are more passive galaxies in local clusters, but a greater fraction of star forming objects in higher redshift systems. This can be seen observationally in the fraction of quiescent, 'red-sequence' objects in contrast with the 'blue cloud' of star forming galaxies (SFGs) as a function of redshift (Stanford et al., 1995; Rakos & Schombert, 1995). These two facts provide an initial scaffolding from which galaxy evolutionary models can be wrought: quiescent systems dominate high density regions of the local universe, and this need not hold throughout cosmic time. What is the path that must be taken to satisfy these two simple observations?

One prescription to turn these observational facts into a coherent model is to identify physical processes that effectively turn SFGs into quiescent systems preferentially in dense environments. Commonly invoked interactions are ram-pressure stripping of cold gas by the intracluster medium (ICM; Gunn & Gott, 1972), removal of the hot gas halo of a galaxy to halt the cooling of gas to sustain star formation (e.g. strangulation; Larson et al., 1980), galaxy-galaxy interactions that disrupt the galaxy (Moore et al., 1996), and a variety of others. The overarching umbrella that is used to refer to these proposed mechanisms is called 'quenching' (Peng et al., 2010), loosely defined here as the environmental process(es) that abruptly discontinue star formation in a galaxy.

Each proposed quenching process has a distinct time scale at which it can effectively halt star formation. These are generally related to the crossing/dynamical time of the galaxies interacting with other galaxies and/or the ICM. For this latter case, it also assumes that the hot ICM is in place at the epoch in question so as to produce the desired effect. These quenching timescales have been estimated to operate on the order of a few Gyrs for an effective change to manifest itself (Brodwin et al., 2013). If cluster galaxies are transformed

primarily by one quenching mechanisms, it might be possible to examine the galaxy populations in clusters at various cross-sections in time (e.g., redshifts) to isolate the process responsible.

Another proposed scenario used to explain the evolution of the galaxy population in clusters is from galaxy mergers. Hierarchical accretion suggests that galaxies assemble from the bottom up. In the context of clusters, these mergers have been proposed to be either dry (e.g. gas-free; van Dokkum, 2005) or wet (gas-rich collisions; Faber et al., 2007). In the wet merger scenario, two or more blue cloud galaxies come together in a burst of star formation, turning into a spheroidal configuration (Mihos & Hernquist, 1996), and are then quenched via some mechanism. This is effectively a transition from the blue cloud onto the red sequence. Dry mergers, on the other hand, push galaxies along the red sequence as they grow in stellar mass, but remain with a generally passive stellar population (color). It is possible and even probable that the various quenching mechanisms and merger scenarios could all play a role of varying importance. Do any observations hint at a timescale of rapid development in the history of cluster galaxies?

The red sequence feature of galaxy clusters has been historically used as a tracer of the core galaxy population at redshifts $z < 1.5$ (Stanford et al., 1995; Rakos & Schombert, 1995; Stanford et al., 1998; Eisenhardt et al., 2008; Mei et al., 2009). Using models of galaxy colors a mean stellar age of the individual systems can be estimated. Progenitor bias (van Dokkum & Franx, 2001) can ultimately lead to an increased estimated formation redshift (z_f) for a cluster's galaxy population as a whole. Despite this caveat, the stellar age of the brightest systems on the red sequence can give some indication as to when the most massive galaxies formed their stars. Models of galaxy growth used to match the evolution of the red sequence feature typically rely on pure, passive evolution models that were formed in a single burst at high redshift $z_f > 3$ (Eisenhardt et al., 2008). In fact, some studies

of clusters suggest even larger formation redshifts of $z_f > 5$ (Rakos & Schombert, 1995; Schombert & Rakos, 2009). As the redshift of the cluster sample increases, the formation redshift also grows (with an accompanying increase of scatter).

For instance, for cluster redshifts $z > 1$, it appears that $z_f = 30$ is not ruled out (Fig 19 of Eisenhardt et al., 2008). Not all cluster galaxies need form at $z_f > 5$, but at least some passive systems were born remarkably early in the universe. This could simply be a manifestation of galaxy downsizing (Cowie et al., 1996), in which the largest stellar mass galaxies (presumably formed in the densest regions) formed earlier in the universe. Within the literature, there is no consensus for any single model or mechanism for the redshift evolution of the red sequence feature (Fassbender et al., 2014; Fritz et al., 2014).

By shifting the focus from the highly biased red sequence galaxies to cluster populations in general, there is the hope that the bulk stellar properties of these systems could be investigated as a function of redshift. Mancone et al. (2010) mapped the luminosity functions (LFs) of cluster galaxies spanning $0.3 < z < 2.0$ at *Spitzer* wavelengths. At the highest redshift of this sample, *Spitzer* 3.6 μm coverage measures rest-frame J band, which is a tracer of the stellar population for a range of ages. They mapped the evolution of the characteristic luminosity $m^*(z)$[1] of the clusters by comparing the data to models of simple stellar populations with various formation redshifts z_f. This is similar to the exercise performed for red-sequence fitting. The mean formation age of these systems was $z_f \sim 2.5$ (Mancone et al., 2010), with the same behavior noted previously: higher redshift clusters favor higher formation redshifts. Their two highest redshift bins ($z \geq 1.5$) have $m^*(z)$ values nearly a magnitude fainter than the predicted evolution of their best fitting model (their Fig 7). The conclusion drawn from this observation is that rapid mass assembly (up to 4× growth) must

[1]We designate the characteristic magnitude in lower case (m^*) to emphasize that it is a measure of the apparent magnitude and to distinguish it from the stellar mass ($M_\star$) also found in this text.

occur in cluster galaxies $z \leq 1.5$ (Mancone et al., 2010). Brodwin et al. (2013) investigated the star formation activity of galaxies in these clusters to look for clues as to the nature of this mass assembly. They found that the star formation within the core of these clusters transition from unquenched to quenched at the same epoch ($z \sim 1.4$) as the rapid assembly era within Mancone et al. (2010). This behavior is generally attributed to wet mergers within the cluster core, rapidly growing the mass of these systems and then abruptly turning off the star formation activity.

In the previous examples, all of the structures were considered to be clusters. Generally speaking, the highest redshift at which virialized halos of $M \geq 10^{14}\ M_{\odot}$(e.g. clusters) are expected is at $z \sim 2$ in large ΛCDM simulations (Chiang et al., 2013). The collection of components that will constitute a cluster in the future is referred to as a protocluster. Observationally, galaxy overdensities at $z > 2$ are designated as protoclusters for the sake of simplicity, as it is difficult to confirm these systems to be in virial equilibrium apart from a handful of cases (Gobat et al., 2011; Wang et al., 2016). This unique transition point in the universe represents an epoch at which galaxies could first begin to interact with one another. Mancone et al. (2010) and Brodwin et al. (2013) presented tantalizing evidence that the majority of mass assembly occurred around $z \sim 1.5$, but higher redshift luminosity functions of structure might yield further insight into the galaxy growth within dense environments.

Wylezalek et al. (2013) identified *Spitzer* galaxy overdensities around high redshift ($1.3 < z < 3.1$) radio-loud AGN and built 3.6 and 4.5 μm LFs in Wylezalek et al. (2014). This redshift range overlapped the sample of Mancone et al. (2010) and extended the age probed by more than 1 Gyr. Remarkably, the $m^*_{3.6}$ and $m^*_{4.5}$ evolution over the redshifts probed are well fit by a passive stellar evolution model formed at $z_f = 3$ or larger. They also do not match the results of Mancone et al. (2010), in that they fail to see a burst of mass

assembly at $z \leq 1.5$. This is attributed to a sampling bias, in that the high redshift overdensities are thought to be the most massive, rare systems in the universe, while the lower redshift sample is tracing the growth of less massive clusters. This is analogous to galaxy downsizing, in that the most massive overdensities are fated to assemble into a cluster mass halo more quickly in ΛCDM (Chiang et al., 2013). Therefore, signatures of mass assembly for the progenitor systems of (Wylezalek et al., 2014) could potentially be observable beyond their redshift limit of $z \approx 3$. In Section 4.2, we probe these earlier epochs to possibly identify epochs at which rapid mass growth or quenching might be exhibited.

Thus far, it appears that galaxies within clusters, as traced by both red sequence and LF models, form at high redshift ($z_f \geq 3$) and evolve passively thereafter. This is an interesting result, as the cosmic star formation rate in the universe does not peak until approximately $z \sim 2$ (Madau & Dickinson, 2014). Indeed, if galaxies in dense environments form earlier than their 'field' galaxy counterparts, which follow the mean trend, then evidence of this should be apparent at high redshifts. The number of spectroscopically confirmed protoclusters has evolved considerably after the first few discoveries (Steidel et al., 1998; Venemans et al., 2002), but were still only numbered in the few dozens up until recently. These were also identified by a wide range of selection techniques, from blind spectroscopic surveys (Steidel et al., 1998) to targeted narrowband (NB) imaging around high redshift quasars (Venemans et al., 2007). In the instances in which these galaxy overdensities were compared with field galaxies at a similar redshift, the results of environmental evolution are varied at best. We continue the exploration of these results at higher redshift in Section 5.4.

The majority of cases in the literature where protocluster galaxies were measured with respect to field sources consist of one or two candidate structures that are compared to a 'blank' field-of-view. For instance, Casey et al. (2015) studied a protocluster at $z = 2.5$ within the COSMOS field and found that it had evidence of greater AGN activity, more indi-

cations of merging/interacting galaxies, and a population of Lyman-Break Galaxies (LBGs) with $\sim 1.5\times$ greater stellar mass. These had similar star formation rates (SFRs) when compared to field sources, though. For other protoclusters at $z \leq 2.5$ identified with NB filters centered on the redshifted $H\alpha$ line, the candidate galaxies were also found to be dustier (Cooke et al., 2014), more massive, and not significantly forming more stars than their field counterparts (Hatch et al., 2011).

Similar studies that trace the $Ly\alpha$ emission of protocluster galaxies find that they are generally brighter (Zheng et al., 2016), *less* dusty (although this may be a selection effect), and younger (Dey et al., 2016). $Ly\alpha$ equivalent widths (EWs) have been used extensively as an estimator for the star formation rate of galaxies in the high redshift universe (Dijkstra & Westra, 2010), and Zheng et al. (2016) find evidence that the EWs are stronger at $z = 2.8$. Dey et al. (2016) do not find such EW dependence at $z = 3.8$, while Toshikawa et al. (2016) finds smaller EW_0 in a protocluster at $z = 3.67$ when compared to the field. Hayashi et al. (2011) and Hayashi et al. (2012) find only the reddest galaxies in their structures have statistically significant environmental dependence. It seems clear that when individual high redshift structures are analyzed, usually an environmental influence is found, but the effect is varied. It is also apparent that in some protoclusters, the property in question is enhanced (e.g. dustier galaxies), while in others it is diminished, even at similar redshifts.

This does not seem to be the case when multiple candidate structures are identified in the same manner, or large surveys are systematically analyzed. Ownsworth et al. (2016) used a constant number density selection technique for the UKIDSS survey to measure the evolution of galaxies. This selection technique is thought to be much less-biased when compared with a mass-limited selection in matching progenitor galaxies to their offspring. Their results point to a relatively early formation redshift ($z \geq 3$ and possibly earlier) and subsequent passive evolution with little environmental influence. With a similar method,

Zhao et al. (2017) tracked the growth and evolution of $z = 0$ Brightest Cluster Galaxies (BCGs) from $z \sim 2$. A key result is that most of the mass growth must occur at $z < 2$, as BCGs are not divergent from similar mass systems. Diener et al. (2013, 2015) identified more than 40 spectroscopic galaxy group and larger systems in the COSMOS field. Compared to field sources, their analysis revealed no statistically significant color (stellar population) difference in the group environments with respect to the field.

This begs the question of why overdensities do not show up as significant deviations from their field counterparts when analyzed systematically at a variety of redshifts, while individual systems have found statistically different evolution over similar epochs. It is often difficult to match the results of these protocluster studies coherently. These protoclusters exist at a wide range of redshifts, each within their infancy and characterized by the rapid changes expected in a ΛCDM universe (Chiang et al., 2013; Muldrew et al., 2015). Particularly for the individual case studies (Hatch et al., 2011; Hayashi et al., 2012; Cooke et al., 2015; Dey et al., 2016; Zheng et al., 2016), different instruments and selection techniques were used, which targeted different populations of galaxies. Furthermore, the galaxy properties themselves were not all analyzed in the same manner.

In an attempt at tackling the complex problem of galaxy evolution, it can be helpful to simplify the approach (Abramson et al., 2016), and confront the issue in a new way instead of adding further epicycles. It was this impetus that inspired us to construct the Candidate Cluster and Protocluster Catalog (hereafter CCPC; Franck & McGaugh, 2016b,a). With a straightforward algorithm, we were able to systematically detect galaxy overdensities from disparate spectroscopic catalogs in the high redshift universe. Then, the properties of the galaxies within these candidate structures can be traced through cosmic time in a series of cross-sections. We hope to address the evolution of galaxy stellar mass, as Mancone et al. (2010) and Wylezalek et al. (2014) did at lower redshifts, while simultaneously mapping the

field evolution in a consistent manner over a range of redshifts. Although not a longitudinal study, these snapshots of galaxies in dense environments may provide a powerful glimpse into the behavior governing their evolution.

We present here a detailed analysis of the CCPC sample to date using *Spitzer* IRAC and supplementary *Hubble Space Telescope* near-infrared (NIR). In this back-to-basics approach, we measure the 3.6 and 4.5 μm LFs of the galaxies in the CCPC as a function of redshift. This serves as a tracer of the stellar mass of these objects, and are compared with 'field' galaxies identified in the same spectroscopic surveys used for the CCPC.

Throughout this work we assume a cosmology of $\Omega_{\rm m} = 0.3$ and $\Omega_{\Lambda} = 0.7$, with a Hubble value of $H_0 = 70$ km s^{-1} Mpc^{-1}. All magnitudes quoted are in the AB system, with apparent magnitudes in the four *Spitzer* IRAC channels denoted as [3.6], [4.5], [5.8], and [8]. The accompanying apparent magnitudes from *Hubble Space Telescope* measurements will be referred to by the filter name (e.g. $F160W$).

4.1 Observations

The CCPC identifies structure around galaxies by mining archival spectroscopic redshift catalogs. Any volume within a search radius of $R = 20$ comoving Mpc (cMpc) and distance in redshift space Δz corresponding to ±20 cMpc which contains 4 or more galaxies and display a galaxy overdensity of $\delta_{gal} > 0.25$ is considered a candidate system (Franck & McGaugh, 2016b,a). These are the minimum requirements and, in many cases, are exceeded.

The algorithm was designed to be used on a variety of survey depths and widths, where $N \geq 4$ is used as a signpost from which the volume overdensity can be computed. The average number density of these systems is $n \sim 0.05$ cMpc^{-3}, and a mean galaxy overdensity of

$\delta_{gal} \sim 2.0$. We have shown in Franck & McGaugh (2016b) and Franck & McGaugh (2016a) that these protocluster candidates are statistically distinct both spatially and along the line of sight from non-CCPC galaxies.

4.1.1 Galaxy Selection

The catalog contains a total of 216 structures spanning $2 < z < 6.56$. We include two objects of interest at $z = 6.56$ from Franck & McGaugh (2016a) to bring the total number of candidate systems to 218. These systems lack field galaxies, and so a galaxy overdensity (δ_{gal}) cannot be computed, a requirement for inclusion in the CCPC. In this list there exist 2048 galaxies. The vast majority of these objects are identified either by spectroscopically targeting Lyman Break Galaxies (LBGs Steidel et al., 1998) or follow-up spectroscopy on suspected $H\alpha/Ly\alpha$ emitters from NB imaging (Venemans et al., 2007).

We note here briefly that SFGs are not the dominant massive galaxy population at these epochs. In the high redshift universe, van Dokkum et al. (2006) estimated that only 20% of all galaxies with $M_\star > 10^{11}\ M_\odot$ are LBGs. The majority of systems ($\sim$ 70%) are thought to be Distant Red Galaxies (DRGs), which are often too faint in the observed optical passbands to be spectroscopically targeted. They can also lack the strong emission lines of their unobscured star-forming counterparts. We will discuss further implications of this in Section 5.4.

4.1.2 Data

All of the measurements in this work came from archival data sets. The *Spitzer* Heritage Archive provided nearly 600 processed images in all four IRAC channels covering most of the CCPC fields. The greatest wavelength coverage of these systems came from the first two channels (3.6 and 4.5 μm), with galaxy contributions from 177 and 184 CCPC structures,

respectively. Nearly $\sim$ 75% of the 2048 CCPC galaxies were measured at 4.5 μm. All photometry was performed using 2" radius apertures using the IRAF[2] (Tody, 1986, 1993) QPHOT package, and the magnitudes were computed by $m_{AB} = 23.9 - 2.5\log(f_\nu/1\mu\mathrm{JY})$. Aperture corrections were applied to each galaxy (0.32 mags in [3.6] and 0.36 for [4.5]), as in Papovich et al. (2010). Our photometry is consistent with the reported [3.6] and [4.5] AB magnitude values compiled in the 3D-HST database (Skelton et al., 2014) within the CANDELS fields, apart from the aperture corrections.

We obtained more limited data coverage of the CCPC with the *Hubble Space Telescope* in the $F160W$ bandpass via the *Hubble Legacy Archive*. The scope of these images were primarily concentrated within the CANDELS fields (Grogin et al., 2011). Approximately 25% of the CCPC was measured within these images. Fluxes were measured in apertures of radius 0.4" with a zeropoint magnitude of 25.75, values which were adopted from the WFC3 Handbook. The pixel scale of these images ranged from ~7 to 33.33 pix/".

With *Hubble Space Telescope* and *Spitzer* data, there were cases in which a galaxy was measured in a number of images. All magnitudes listed here are the uncertainty-weighted mean value of the photometry.

4.1.3 Building the Luminosity Function

We map the evolution of the CCPC galaxy luminosity function using the Schechter (1976) form

$$\phi(L)(dL) = \phi^*(\frac{L}{L^*})^\alpha e^{-L/L^*}\frac{dL}{L^*}, \tag{4.1}$$

which relates the characteristic number density ($\phi(L)$) of sources over a range of luminosities. L^* is the characteristic luminosity of the distribution where the number density

[2]IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

decreases rapidly, and α is the slope of the faint end. The scaling factor of the Luminosity Function is ϕ^*. In this work, we will adopt the magnitude functional form of

$$\phi(m) = 0.4\ln(10)\phi^* \frac{10^{0.4(m^*-m)^{\alpha+1}}}{\exp\left[10^{0.4(m^*-m)}\right]}. \tag{4.2}$$

To construct the distribution function, we first compute the density of each candidate structure by finding the minimum rectangular region (in units of arcmin^2) that bounds all galaxies. We then place galaxies in magnitude bins of $\Delta m = 0.4$ for a given redshift range, and finally divide this by the summed surface area of candidates at that redshift. The uncertainty for each magnitude bin is the 1σ photometric uncertainty of the galaxy magnitudes in that bin. The number density uncertainty is Poissonian ($\sqrt{N}$). The redshift bins were designed to offer a balance between similar temporal spacing and number of galaxies at each epoch. This balance is necessary, for if the time period probed by each snapshot is too variable or too large, then any evolutionary inferences that the luminosity function might provide could be lost. In addition, too few galaxies in a redshift bin can be insufficient in fitting the parameters of the Schechter function. We aim for $N \geq 10^2$ galaxies per redshift bin whenever possible, as this provides a robust fit to the data in practice.

We computed the $e-$ and $k-$ corrections of each galaxy by estimating (using EzGal; Mancone & Gonzalez, 2012) the observed magnitude difference of a BaSTI simple stellar population model (Percival et al., 2009) between the model's magnitude at the observed redshift of the galaxy and the center of its redshift bin. Typically these corrections are less than $\Delta m \sim 0.05$ magnitudes, which are smaller than our photometric uncertainties. There is no net change in the values of m^* when the corrections are applied, as these shifts in magnitude are balanced out by the galaxies at either end of the redshift bin.

For fitting the Schechter function to the data, we used SciPy's *curvefit* routine. This routine takes the data table, the Schechter function, the parameters to be fit and a set of initial

guesses for those parameters. These initial input values are insensitive to the outcome. The solution to Equation 4.2 is optimized via a Levenberg-Marquardt algorithm. In practice, if the data have sufficient depth to fit the faint-end slope α, all three parameters can be solved for simultaneously, as Wylezalek et al. (2014) were able to do. However, the variability of our data depth does not permit us to reliably fit α at all redshifts, and we therefore set it as a constant $\alpha = -1$. This is standard procedure for *Spitzer* LFs of insufficient depth (Mancone et al., 2010). Wylezalek et al. (2014) obtained the same results independently of whether α was fixed or not. The selection function of spectroscopic surveys used to construct the CCPC will typically favor bright galaxies out of necessity, which would artificially restrict the faintest regions in magnitude space, regardless.

The main science goal of these LFs is to investigate the temporal evolution of $m^*(z)$ and number density of the largest galaxies in these systems, and so the faint slope of the galaxies is of relatively minor importance. Even at low redshifts, determining α can be problematic, as low surface brightness galaxies are often missed (McGaugh, 1996). When the value of α is allowed to vary, it is generally consistent with $\alpha \approx -1$ within the uncertainties for the lowest redshift sources in our data. At higher redshifts, there are too few galaxies in the faint magnitude bins from incompleteness, and the fitting routine breaks down. In short, the optimization procedure of *curvefit* provides a more robust fit to the data with a constant $\alpha = -1$. There have been no completeness corrections implemented on the data set. The focus of this research is on tracing the brightest, most massive galaxies at a given epoch. Attempting to adjust the number of faint galaxies is (1) not important in achieving the research aim and (2) uncertain at best, as spectroscopic surveys of high redshift sources are inherently biased in this regard.

Once the fitting routine has provided values of $m*$ and ϕ^*, the uncertainties are calculated via bootstrapping. The galaxies in each redshift bin are resampled with replacement

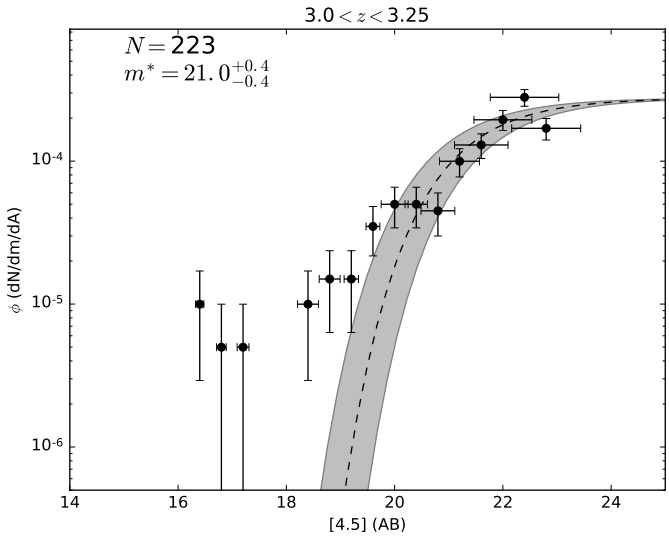

Figure 4.1: The 4.5μm Luminosity Function of CCPC galaxies at redshifts $3.0 \leq z < 3.25$. ϕ is the number of galaxies at a specific magnitude per square arcminute ($dN/dm/dA$). Uncertainties in the number density of sources are Poissonian ($\sqrt{N(dm)}$), while the magnitude uncertainties in each bin are the average uncertainty of the bin's galaxies. SciPy's *curvefit* routine's best fit to the data is the black dashed line. The number of galaxies in the plot and the value of m^* and its 2σ bootstrapped uncertainties are listed in the top left corner. In this work, α is defined to be -1. The gray shaded area represents the 95% bootstrapped confidence region of the fit.

10^4 times. Each instance is fitted to Eq 4.2 in the same manner as the full data set. The 95% confidence region of the data is provided by fitting the $\pm 2\sigma$ values of m^*. Figure 4.1 illustrates this.

4.2 Results

4.2.1 Luminosity Functions

Table 4.1 and Table 4.2 contain the [3.6] and [4.5] *Spitzer* parameters estimated for the CCPC, respectively. In general, the values of m^* are flat as a function of redshift, suggesting

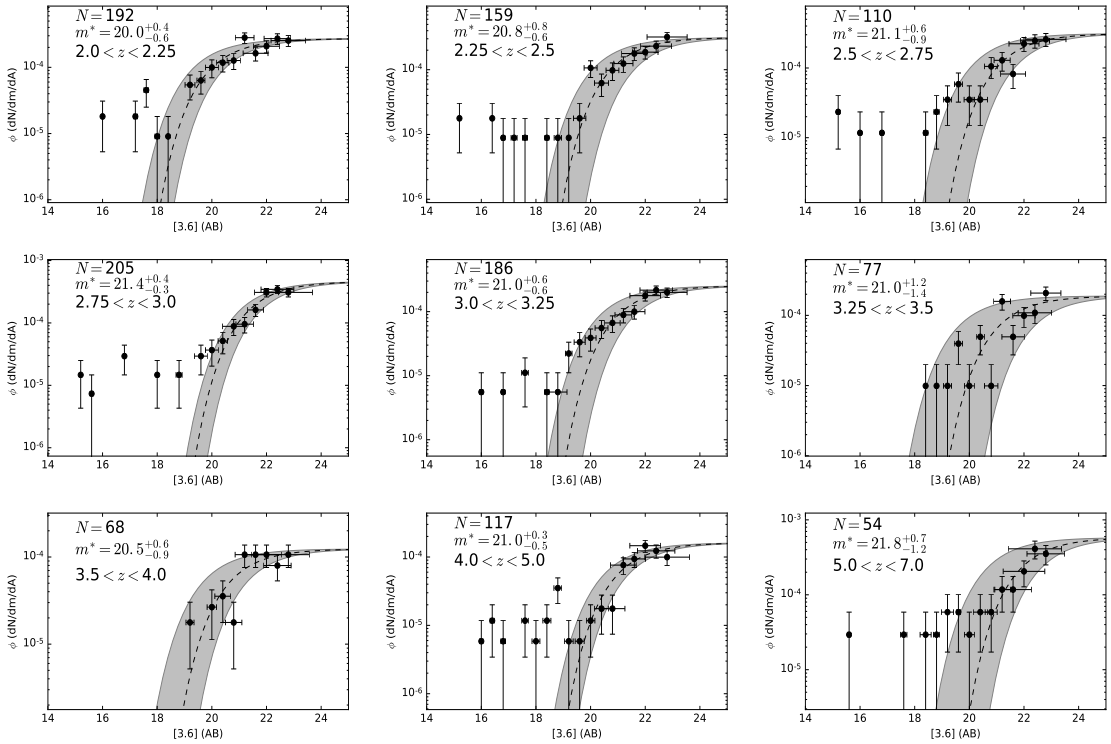

Figure 4.2: The 3.6μm *Spitzer* Luminosity Function at nine epochs across $2 \leq z < 7$. The surface number density (ϕ) of galaxies is measured in square arcminutes. The number density uncertainty is calculated by the root N value in each bin, while the magnitude uncertainties are the average photometric errors in that bin. The gray shaded regions are the 95% confidence intervals calculated by bootstrapping. The Schechter fit (Eq 4.2) to the CCPC data is the black dashed line, with α defined as -1. The number of galaxies and the value of m^* are listed in the upper left hand corner. The m^* uncertainties are the bootstrapped 2σ values. Galaxies much brighter than m^* we refer to as 'hyperluminous' sources and will be discussed in Section 5.4.

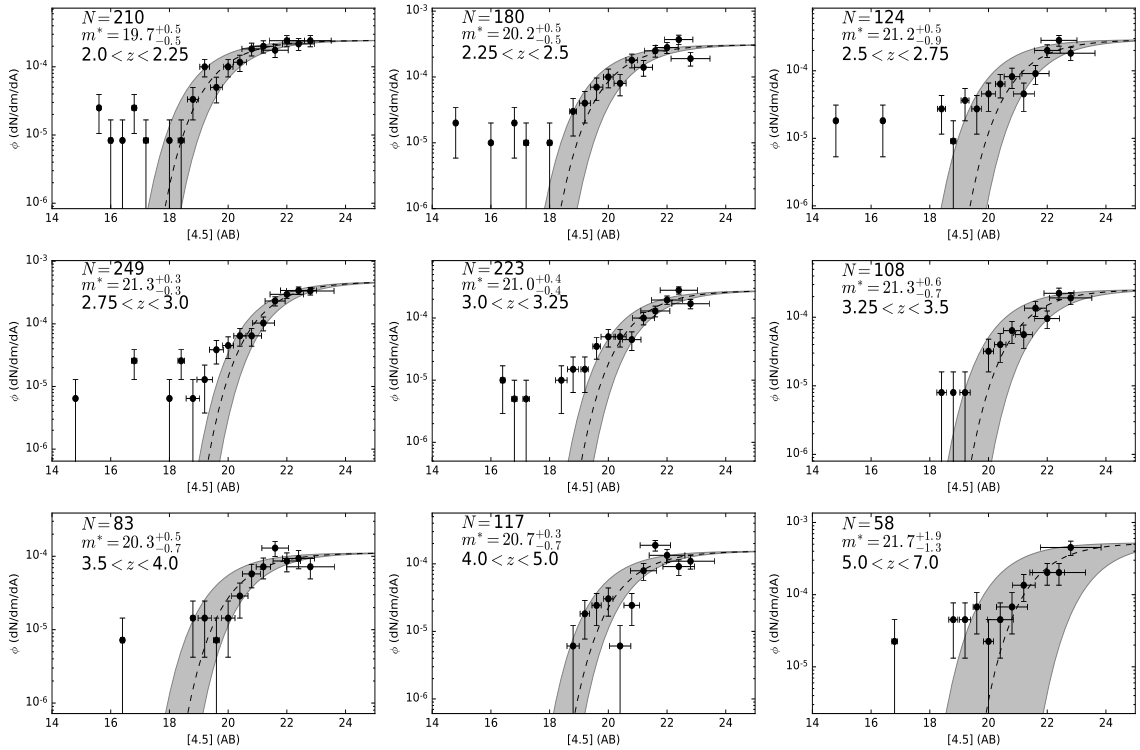

Figure 4.3: The 4.5μm version of Figure 4.2.

Table 4.1. 3.6μm CCPC Luminosity Function

Redshift Range	N Galaxies	m^* (AB)	$2\sigma(m^*)$ (95% CI)	ϕ^* (dN/dm/dA)
$2 \leq z < 2.25$	192	20.01	$^{+0.50}_{-0.65}$	2.94×10^{-4}
$2.25 \leq z < 2.5$	159	20.89	$^{+0.88}_{-0.68}$	3.41×10^{-4}
$2.5 \leq z < 2.75$	110	21.13	$^{+0.67}_{-0.96}$	3.39×10^{-4}
$2.75 \leq z < 3$	205	21.42	$^{+0.46}_{-0.33}$	5.01×10^{-4}
$3 \leq z < 3.25$	186	21.06	$^{+0.60}_{-0.65}$	2.74×10^{-4}
$3.25 \leq z < 3.5$	77	21.07	$^{+1.22}_{-1.47}$	2.05×10^{-4}
$3.5 \leq z < 4$	68	20.51	$^{+0.63}_{-0.99}$	1.34×10^{-4}
$4 \leq z < 5$	117	21.06	$^{+0.38}_{-0.49}$	1.75×10^{-4}
$5 \leq z < 6.6$	54	21.83	$^{+0.74}_{-1.23}$	6.30×10^{-4}

Note. — The results of fitting a Schechter function (Equation 4.2) to the CCPC galaxies in a series of redshift bins (1^{st} column). The number of galaxies (N) in each redshift bin is listed in the 2^{nd} column, followed by the fitted m^* parameter in the 3^{rd} column. The 4^{th} column represents the 95% confidence interval of the m^* value. This uncertainty was computed by bootstrapping with resampling of the data. The characteristic density (ϕ^*) in units of the number of galaxies per magnitude bin per square arcminute is found in the final column. The faint end slope of the LF was fixed to be $\alpha = -1$.

Table 4.2. 4.5μm CCPC Luminosity Function

Redshift Range	N Galaxies	m^* (AB)	$2\sigma(m^*)$ (95% CI)	ϕ^* (dN/dm/dA)
$2 \leq z < 2.25$	210	19.73	$^{+0.56}_{-0.59}$	2.65×10^{-4}
$2.25 \leq z < 2.5$	180	20.27	$^{+0.56}_{-0.53}$	3.35×10^{-4}
$2.5 \leq z < 2.75$	124	21.25	$^{+0.58}_{-0.97}$	3.14×10^{-4}
$2.75 \leq z < 3$	249	21.36	$^{+0.39}_{-0.31}$	5.13×10^{-4}
$3 \leq z < 3.25$	223	21.08	$^{+0.41}_{-0.46}$	3.00×10^{-4}
$3.25 \leq z < 3.5$	108	21.31	$^{+0.71}_{-0.80}$	2.76×10^{-4}
$3.5 \leq z < 4$	83	20.38	$^{+0.51}_{-0.74}$	1.21×10^{-4}
$4 \leq z < 5$	117	20.71	$^{+0.35}_{-0.74}$	1.68×10^{-4}
$5 \leq z < 6.6$	58	21.75	$^{+1.97}_{-1.38}$	5.69×10^{-4}

Note. — Identical to Table 4.1, but at 4.5μm.

little evolution. Figure 4.1 shows a Luminosity Function for a single epoch to illustrate the finer details of the fit. Figure 4.2 shows the full set of Luminosity Functions at each epoch at 3.6μm, while Figure 4.3 does the same for [4.5].

4.2.2 Field Luminosity Functions

In an effort to compare galaxies in overdense environments with their 'field' counterparts, we assembled a list of all spectroscopic galaxies that were in the same surveys as the protoclusters in our sample. This list contains more than 4000 galaxies, double the number of the CCPC (N = 2048 objects). We imposed no richness or density criteria on this sample set, apart from that its members were not within the volume of a candidate protocluster. Although some of these galaxies may exist in a volume with $N \geq 4$ galaxies (a requirement for CCPC candidacy), they did not have the sufficient galaxy density to be flagged as a protocluster candidate. The references to the spectroscopic measurements of the field systems

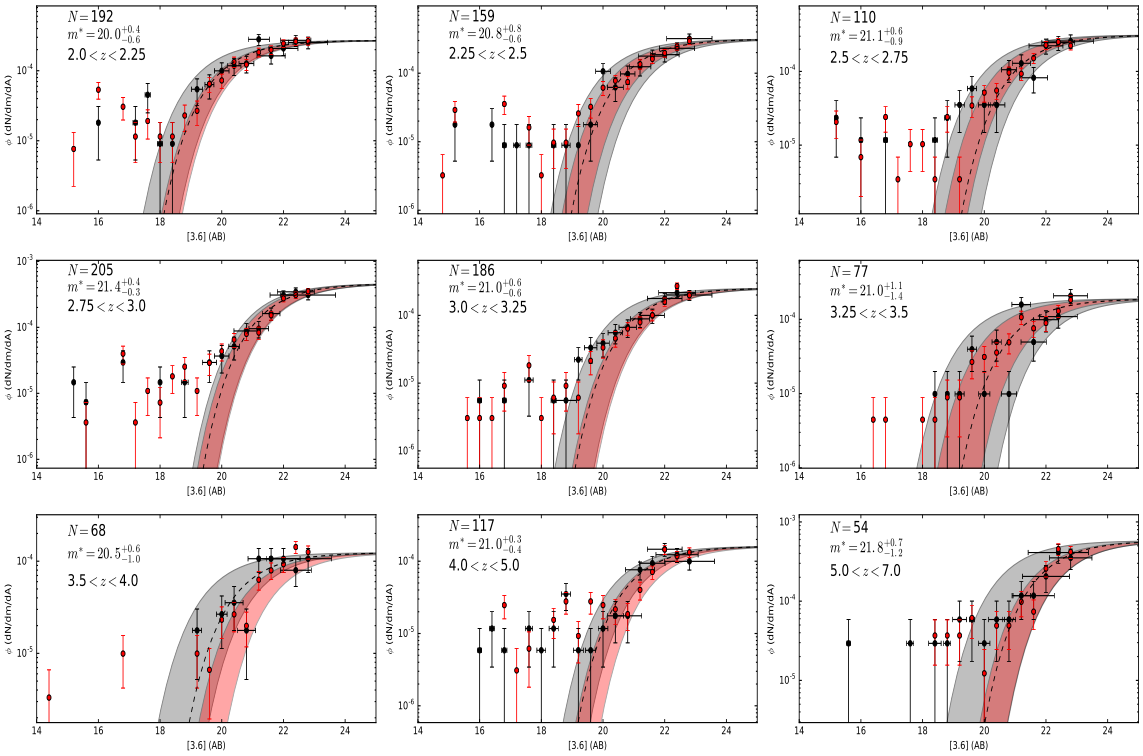

Figure 4.4: Superimposed on the LF of CCPC galaxies (Fig 4.2) are spectroscopic 'field' galaxies (red points), taken from the same redshift source catalogs and fields as the CCPC candidates (see Appendix). These are simply scaled to the CCPC ϕ values by their relative numbers at each redshift bin ($\phi_{field}(z) = (N_{CCPC}/N_{field})\phi_{CCPC}(z)$). Within the uncertainties, these distributions show no statistically significant difference. The red shaded regions are the bootstrapped 95% confidence interval for the field galaxies, and they overlap the CCPC's 2σ range (gray shaded region) at all times.

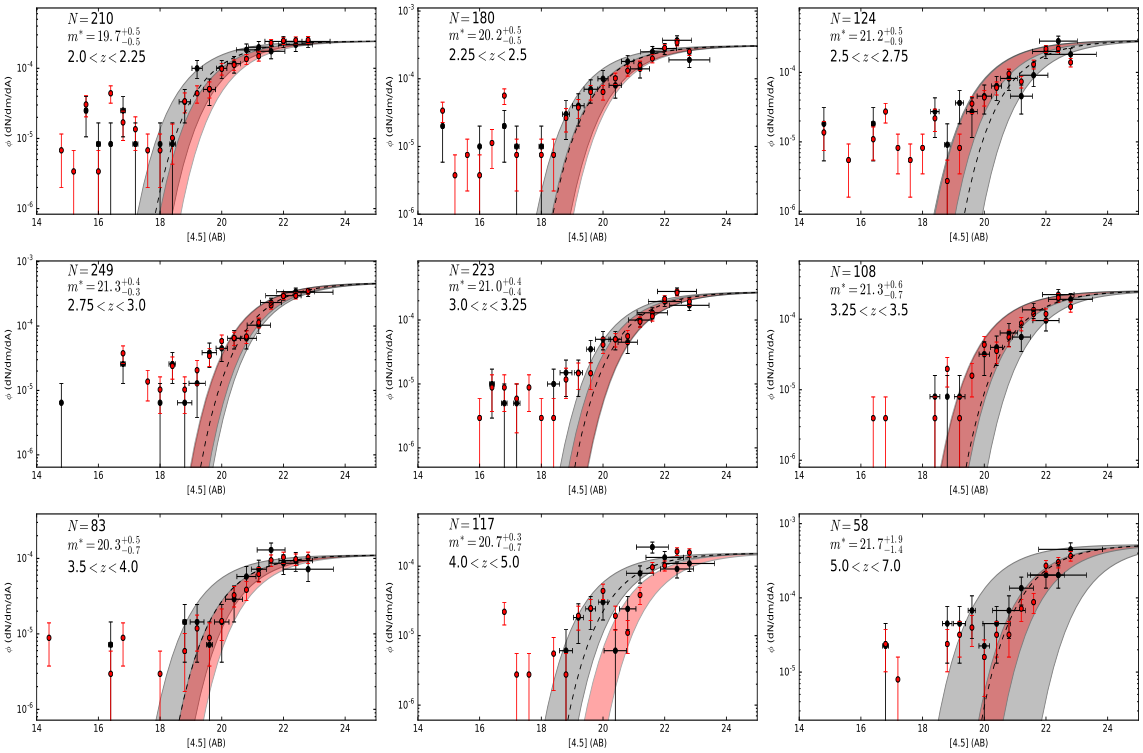

Figure 4.5: The same as Fig 4.4 but with [4.5] magnitudes.

are located in the Appendix, and can also be found in the references of Franck & McGaugh (2016b) and Franck & McGaugh (2016a).

We analyzed these galaxies in the same manner and built an 'All Galaxy' LF (as a proxy for field galaxies) at each redshift bin. The value of ϕ^* was scaled to the CCPC LF at each epoch. As can be seen in Figs 4.4 and 4.5, the number density of field galaxies (red points) in all bins, at all epochs, are consistent with the CCPC $\phi(m)$ (black points). Fitting a Schechter function to 'All Galaxies' produces equivalent values of $m^*(z)$ to that of the CCPC galaxies, as can be seen by the overlapping 95% confidence intervals (red and gray shaded regions, respectively).

The CANDELS GOODS-S field is the deepest, most continuous spectroscopic survey from which dozens of structures are identified in our sample (Franck & McGaugh, 2016b,a). Nearly 25% of galaxies in the CCPC originate in this field. To minimize the effects of varying spectroscopic selection functions from the heterogeneous sample the CCPC is constructed from, we constructed GOODS-S LFs for CCPC and non-CCPC galaxies. The m^* values remain unchanged at all redshifts in this subsample, for both field and overdense galaxies, within the uncertainties. Although spectroscopic selection is not definitively ruled out as a variable for the entirety of the CCPC LFs, it does not appear to be a driving factor in the CANDELS GOODS-S data.

As a further test, we limited our analysis to galaxies in the 'All Galaxy' list that had $N < 4$ galaxies within the CCPC search volume (R = 20 cMpc, Δz±20 cMpc). Although this 'Reduced Field Galaxy' sample limited the number of galaxies to 2299, this list is still larger than the number of CCPC galaxies, and thus remains a fair comparison. Interestingly, this imparted no measurable difference when compared to the shape of the 'All Galaxy' or CCPC LFs. It is possible that at these very early epochs, the galaxies that are spectroscopically selected have not had sufficient time to interact in their modestly dense environments

and thus differentiate themselves. Another plausible explanation is that many of the galaxies selected in the overdense volumes are not future cluster galaxies, but rather are field interlopers that will disperse by $z = 0$ and thus show little/no differentiation from our field sample. These results will be discussed in more detail in Section 5.4.

4.2.3 Galaxy Colors

A selection tool used to identify high-redshift ($z > 1.3$) galaxies with *Spitzer* is the popular $[3.6] - [4.5] > -0.1$ color cut (Papovich et al., 2012; Wylezalek et al., 2014). It is a simple and effective way of removing low redshift galaxies from a sample, regardless of the underlying stellar population. Interestingly, more than 1/3 of the CCPC sample (all with spectroscopic redshifts $z > 2$) failed this color cut. Within the photometric uncertainties, many of these systems emit 'true' colors that would satisfy the criterion. However, this blind cut can remove a significant portion of a sample of high redshift galaxies. This is not unique to our photometry. The galaxies in CCPC structures that are coincident with objects in the 3D-HST database (Skelton et al., 2014) show a similar result, with 33.8% of more than 300 galaxies having $[3.6] - [4.5] < -0.1$, with a median color of $[3.6] - [4.5] = -0.25$. This is a further piece of evidence suggesting that this aberrant fraction is the result of photometric uncertainty that has scattered the colors below the cut. There is no obvious correlation with apparent magnitude, therefore AGN or hyperluminous source contamination is likely not an issue.

Although effective at measuring the underlying stellar mass of high redshift galaxies from their rest-frame NIR emission, the colors of *Spitzer* are not sensitive to different stellar populations (Cooke et al., 2014). Even at the highest redshift of the CCPC ($z = 6.56$), the rest-frame wavelength observed at 4.5μm falls at just 5900Å. As the majority of these galaxies were spectroscopically targeted as UV bright, star forming systems, it stands to

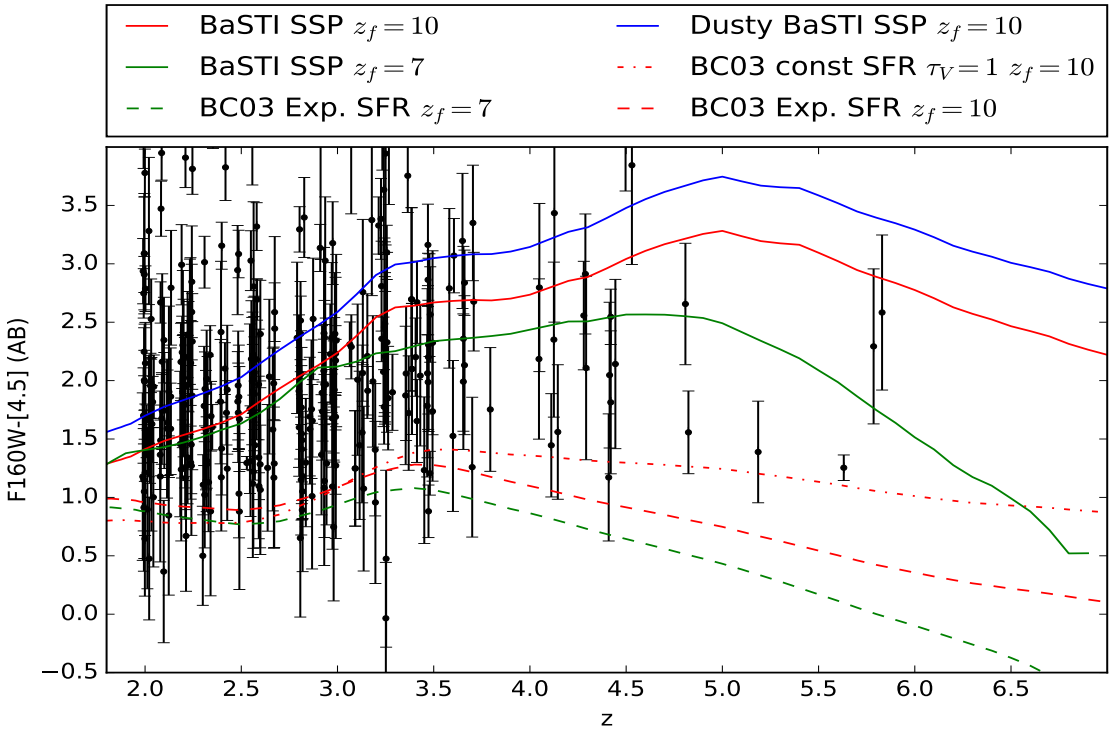

Figure 4.6: We plot the $F160W - [4.5]$ color evolution of six stellar populations: two BaSTI SSPs at $z_f = 7, 10$ (green and red solid lines, respectively), with an added Charlot & Fall (2000) dust component to the $z_f = 10$ SSP (blue solid line), along with two BC03 exponential decaying SFR with $\tau = 1$ Gyr at $z_f = 7, 10$ (green and red dashed lines), and finally a constant star formation model with an extinction of τ_V and $z_f = 10$ (red dot-dashed line). All models have solar metallicity. Plotted as black points are the colors of CCPC galaxies. These span the entire range of model stellar populations. For an individual galaxy, the uncertainties are generally too large to be assigned to one model or another. However, most systems can be separated between SSPs (generally quiescent) and star forming models.

reason that they should have blue colors, which *Spitzer* is not sensitive to. The *Hubble Space Telescope* WFC3 $F160W$ filter probes rest-frame wavelengths of 2100 − 5300Å between the redshift range $2.0 < z < 6.56$. Figure 4.6 illustrates the redshift evolution of the $F160W - [4.5]$ galaxy color as a function of redshift for a variety of stellar population models (Bruzual & Charlot, 2003; Percival et al., 2009), with different initial mass functions (Chabrier, 2003; Kroupa, 2002) and dust extinction (Charlot & Fall, 2000). Plotted on the models are the $F160W - [4.5]$ colors as a function of redshift for the CCPC. We only show galaxies with photometric uncertainties in an individual filter of $\sigma < 0.75$ mags. There is no clear preference for a single galaxy type, with a large scatter of blue and red galaxies throughout. We also cannot ascertain a preferred stellar population, dust content or formation redshift from these colors alone for an individual galaxy, as the uncertainties can be too large. For many systems the difference between a star forming galaxy and a passively evolving SSP can be assigned within the errors.

Noirot et al. (2016) spectroscopically targeted two candidate structures identified by Wylezalek et al. (2014) at $z \sim 2$. Their *Hubble Space Telescope* photometry of the overdensity revealed a wide spread of F140W-[3.6] colors of roughly $-1 < F140W - [3.6] < 3$, with a much smaller range for the galaxies with confirmed redshifts (approximately $0.5 < F140W - [3.6] < 2$).

The 3D-HST data (Skelton et al., 2014) coincident with CCPC galaxies in the CANDELS fields at $2.0 < z < 2.05$ has a mean color of $< F140W - [3.6] >= 0.5 \pm 0.6$, and a range $-0.4 < F140W - [3.6] < 2.3$. The color range is more broad than the spectroscopic sample of Noirot et al. (2016), but does not go beyond the full range of candidate galaxies in their protoclusters. This suggests a general agreement of their protocluster galaxy population and ours. Higher redshift CCPC galaxies show little color evolution in this plane as well.

4.2.4 SED Fitting and Galaxy Stellar Mass

The usual methodology of obtaining stellar masses for galaxies at high redshift is to match a stellar population model with a number of observed magnitudes across a range of filters (e.g. spectral energy distribution (SED) fitting; Magdis et al., 2010). Thus, a galaxy is matched with the stellar age, star formation history, metallicity, initial mass function, and dust content of a given model. Hereafter we will refer to this as 'traditional' SED fitting. The implied stellar mass-to-light ratio of the best fit model is then used to compute the $M_\star$ of the galaxy. There are a number of degeneracies that arise from this method. This can be particularly important with limited wavelength coverage and inconsistent rest-frame colors over a large range in redshift. These factors will be briefly considered in Section 5.4.

To test the relative strength of a fit to different model SEDs, we follow the procedure used in Bolzonella et al. (2000) to estimate photometric redshifts:

$$\chi^2 = \sum_{i=1}^{N_{filt}} \frac{(F_{i,obs} - F_{i,model})^2}{\sigma_i^2} \tag{4.3}$$

which measures the observed flux in a filter ($F_{i,obs}$) relative to the flux of a model stellar population ($F_{i,model}$) and weighted by the squared observational flux uncertainty (σ_i^2). The SED's flux is calibrated to match the observed [4.5] magnitude of the galaxy being measured in our algorithm. We limit our analysis to magnitudes with uncertainties $\sigma_m < 0.75$ and require a F160W measurement. As stated previously, *Spitzer* colors cannot differentiate between varying stellar populations at these redshifts. This maximum allowable uncertainty is effective in limiting severely anomalous photometric measurements from a heterogeneous sample of surveys which do not have a constant depth. We will show that the photometric uncertainties are generally of minor importance when compared to the model degeneracies of our fits. The stellar mass implied by different models for the same redshift and magnitude have greater variance than the uncertainty introduced by photometric errors. This will be

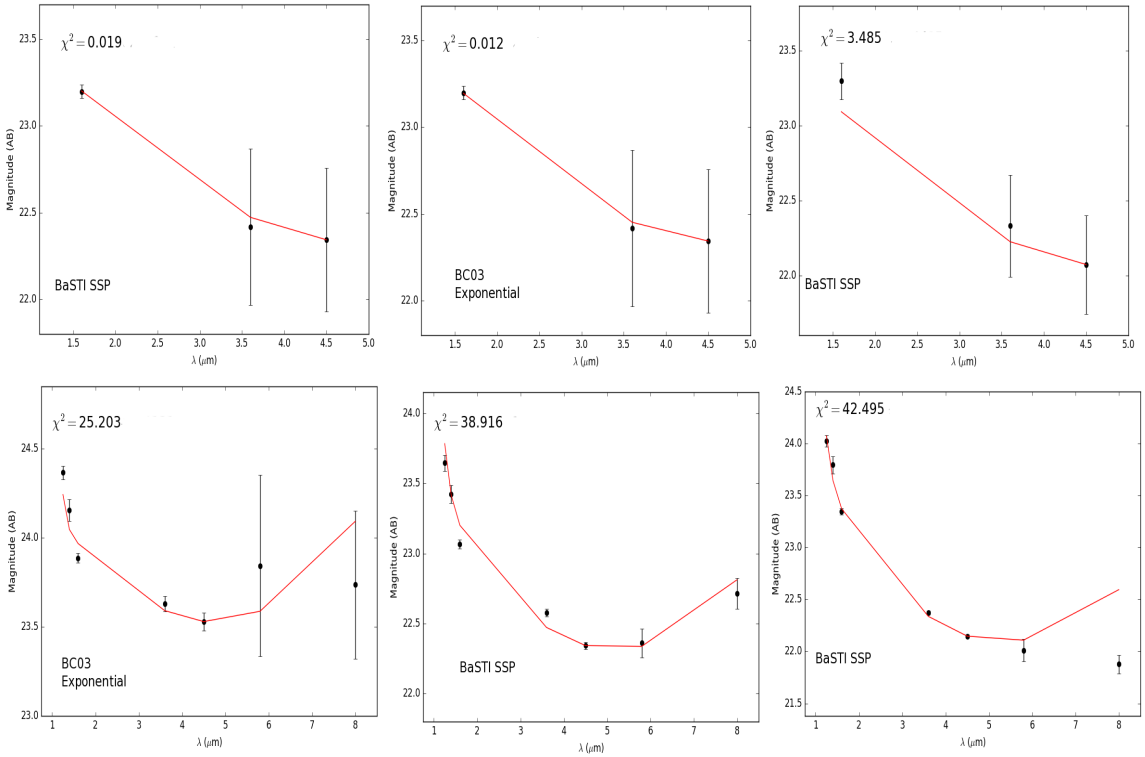

Figure 4.7: *Top:* Three examples of CCPC galaxy SED fits. The black points are the apparent magnitudes at observed wavelengths (F160W, [3.6], and [4.5] filters), while the red lines are the template magnitudes. The formation redshifts, from left to right, are z_f = 3.0, 7.0, and 3.25. The left and right templates are BaSTI SSPs with solar metallicities, while the center template is a BC03, exponentially decaying model with τ = 1 Gyr. The χ^2 values of the fits are listed, which are used to distinguish the relative goodness-of-fit between various models for a single galaxy. *Bottom:* CCPC galaxies found within the 3D-HST database (Skelton et al., 2014) can have greater wavelength coverage (F125W, F140W, F160W, [3.6], [4.5], [5.6], and [8.0]), although these are limited to the GOODS fields. The templates presented are dust-free. The left panel is a BC03 exponential model with τ = 1 Gyr and metallicity of Z = 0.008, while the middle and right panels are BasTI SSP with solar metallicities. They have z_f = 3.75, 3.0, and 3.5 from left to right, respectively. The larger χ^2 values of the models in this panel compared to the *Top* panel are the result of more measurements (with relatively small uncertainties) from which a galaxy model can differ. They are not necessarily a poorer fit. The χ^2 values also depend upon the inherent biases within the models as well, and should therefore not be treated as an absolute metric.

discussed further in Section 5.4.

Using the EzGal code (Mancone & Gonzalez, 2012), we built grids for each filter and stellar population model as a function of formation redshift ($2 < z_f < 10$) and observed redshift (z_{obs}). We consider the following models: BaSTI simple stellar populations (SSPs) with metallicities of $Z = 0.008, 0.0198$ (Percival et al., 2009) and a Kroupa IMF (Kroupa, 2002), Bruzual & Charlot (2003) (hereafter BC03) constant star formation models with extinctions of $\tau_V = 0.2, 1.0$ and a Chabrier (2003) IMF, and two BC03 exponential decaying SF models with $\tau = 1.0$ Gyrs and $Z = 0.008, 0.02$. The model predictions become erratic when the stellar age of the system is low, so we implemented a cut of $z_f - z_{obs} > 0.05$ when fitting the formation redshift.

For the models that did not have a dust component built in, we also explicitly calculated the extinction from a Charlot & Fall (2000) model for the rest-frame wavelengths observed in our filters at a stellar age computed from the z_f value in the grid. The dusty fluxes were recomputed and the fits measured using Equation 4.3 in the same manner as the dust-free systems.

Once the best-fitting model is identified for each galaxy, we calibrate the model to the observed magnitude m_i, for each filter measured (i), and query EzGal for the implied stellar mass at the observed redshift. For an individual filter's mass measurement, $M_{i,\star}$, we estimate its uncertainty by taking the implied mass of the galaxy if the magnitude was changed by its photometric uncertainty ($m_i \pm \sigma_i$) in that filter. Ultimately, the estimated stellar mass of the system is computed from the uncertainty-weighted mean value from each wavelength measured. If the system is found to have a best-fit SED that is dusty, we subtract the dust absorption from the observed magnitudes prior to computing the underlying stellar mass estimates.

In this manner, we were able to fit 414 galaxies that were below a minimum χ^2 threshold,

with a median value of $\chi^2 = 1.6$ and a median of $N_{filters} = 3$ (F160W, [3.6], and [4.5]). Fig 4.7 shows a few examples of SED fits with a variety of χ^2 values. The median mass implied for these galaxies is $3.3{\times}10^{10}\ M_{\odot}$. Nearly half of the systems were best fit by an exponentially decaying, BC03 star forming model (185 objects), with 201 others well fit by a BaSTI simple stellar population. Roughly 50% of galaxies were found to be best fit by a dusty component, and only 18% had less than solar metallicity. The average formation redshift fit by the algorithm was generally old, at $z_f = 7.9$. The 3D-HST data (Skelton et al., 2014) has greater wavelength coverage in *Hubble Space Telescope* filters within the CANDELS fields. The fits to 395 SEDs using this expanded data set did have a marginally smaller median $M_{\star}$ value of $0.9{\times}10^{10}\ M_{\odot}$, a lower percentage of dusty galaxies (27%), and 51% low-metallicity systems (compared to $< 15\%$ in our data set). The 3D-HST catalog's photometry did not use aperture corrections in their *Spitzer* magnitudes, which (if instituted) would systematically increase the stellar mass of these galaxies. Only 7% of the 3D-HST galaxies were best fit by a BaSTI SSP, while the BC03 constant star formation models was applied to roughly 1/3 of CCPC sources. The majority were fit by an exponential model, just like our own photometric set. The mean formation redshift for the 3D-HST photometry was $z_f = 5.9$. The 3D-HST fits have a median value of χ^2 =30.5.

With the recent success of modeling galaxy SFHs as a log-normal distribution (Gladders et al., 2013; Abramson et al., 2016), we attempted to fit the CCPC with a similar analysis. The log-normal distribution adopted is of the form

$$SFR \propto \frac{\exp\left(-\frac{\left(\ln(t)-T_0\right)^2}{2\tau^2}\right)}{t} \tag{4.4}$$

where t is the time since the Big Bang, T_0 is the half-mass time of the galaxy, and τ is the half-mass width of the distribution (Abramson et al., 2016). We took the two BaSTI SSPs (different metallicities) as our base models, and then computed Complex Stellar Populations

(CSPs) using EzGal for a variety of values of T_0, τ ranging from $0.05 \leq (T_0, \tau) \leq 1.0$. This range corresponds to the breadth of values fit to observed SEDs at low and high redshifts (Gladders et al., 2013). We evaluated the goodness-of-fit with Equation 4.3, as before. We included the optional Charlot & Fall (2000) dust extinction by assuming a stellar population age (z_f) coincident with a SFR(t) = 10^{-3}(SFR$_{max}$). The factor 10^{-3} is fairly arbitrary, with an order of magnitude adjustment changing z_f by approximately 0.2.

The log-normal routine was able to adequately describe only 207 galaxies, roughly half of the traditional number of successful SED fits. The fits were also poorer than the traditional fitting, with a median $\chi^2 = 24.6$ (compared to $\chi^2 = 1.6$). This is somewhat surprising, as the volume of parameter space explored was much larger than in traditional fitting. The median mass for these galaxies (1.5×10^{10} $M_\odot$) was similar to the traditional fit, with no preference for low metallicity systems (15%). Nearly 85% of the galaxies were best fit with a dust component included. The mean formation redshift was $z_f = 10.7$, only slightly older than the stellar age from traditional fitting. The median values of T_0, τ are 0.05 and 0.15 (~1.05 and 1.16 Gyrs), respectively. Running the SED fitter on the 3D-HST data set generally confirms the earlier results from our photometry, with a few notable exceptions. Interestingly, more galaxies were successfully fit by using the larger wavelength coverage (262) but with a much lower fraction of dusty systems (0.33), a larger mean T_0 (0.65), an older mean stellar population ($z_f = 14.6$), and a slightly lower median mass than the traditional models (0.7×10^{10} $M_\odot$). The median values of $\tau = 0.15$ was equivalent to the log-normal value derived with our photometry. The 3D-HST fits had a similar value of $\chi^2 = 21.7$ to the log-normal fit of our data, but are in fact more robust, as more filters were used. We will also briefly note that although the median values here may appear to be slightly different if traditional or log-normal SED fitting is used, or the 3D-HST data adopted versus our own photometry, the variance of the properties are much larger than

their differences. This suggests that the underlying properties of these galaxies are still very uncertain. This will be discussed further in Section 5.4.

Many galaxies have no *Hubble Space Telescope* photometry in the CCPC. In order to estimate their $M_\star$ we adopt the simple BC03 exponential decaying SF model with a metallicity of $Z = 0.02$, $\tau = 1$ Gyr, and $z_f = z + 0.25$. This choice is reasonably justified, as the majority of the CCPC galaxy sample was selected as unobscured star forming systems (UV-bright LBGs or $Ly\alpha/H\alpha$ line-emitters), and this exponential model appears to be a consistent fit with the evolution of m^*. It supplies a conservatively low mass estimate for the system.

Figure 4.8 plots the galaxy stellar masses of the objects measured by various SED fitting methods. The scatter is large, even for a single galaxy fit by different algorithms (log-normal versus traditional) or different photometry (3D-HST versus our own). The gray points are the conservative mass estimate for each galaxy, which is simply an exponential model (BC03, $Z = 0.02$, and $z_f = z + 0.25$ described above). These points can be thought of as lower limits of stellar mass for individual galaxies. An example presented in Section 5.4 will illustrate the circumstances in which this is the case.

4.3 Discussion

4.3.1 Hyperluminous Sources

It is clear that the Schechter function fits the data adequately within a few magnitudes around m^*. However, there are some very bright sources ($m \leq m^* - 2$) that are clearly anomalous in Figures 4.2 and 4.3. These points are rare, generally consisting of 1-3 galaxies per bin, but they are apparent at all redshifts and at both wavelengths. In the context of ΛCDM and hierarchical accretion, it is predicted that in the densest regions of the uni-

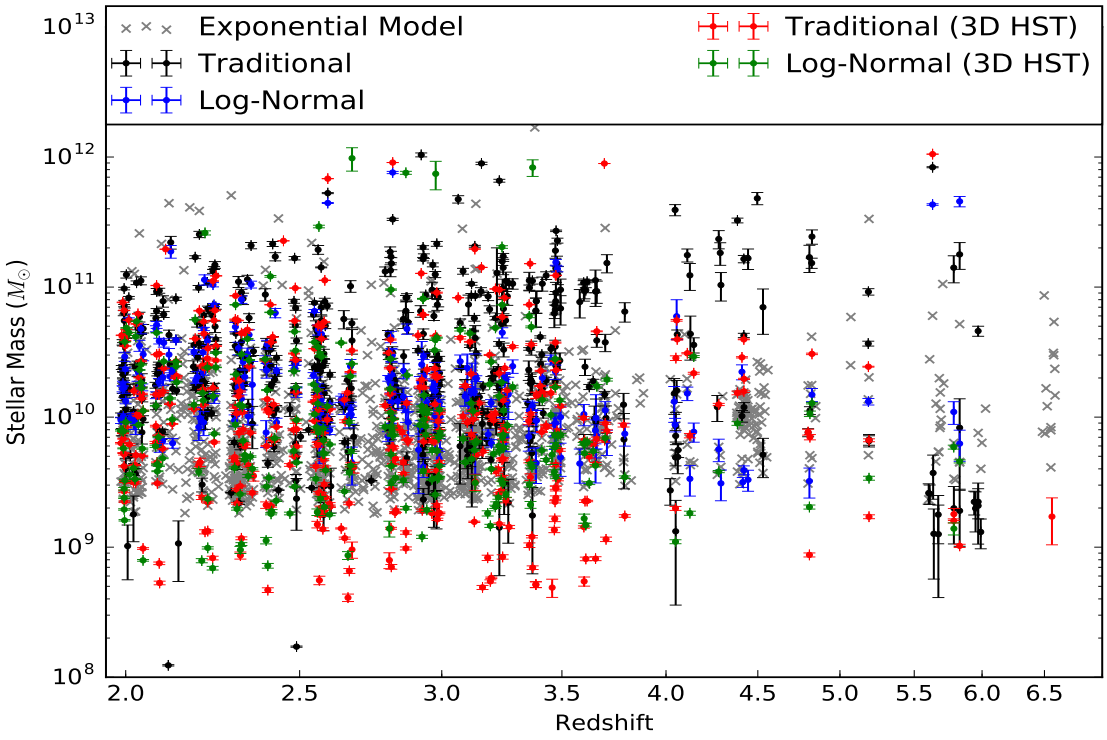

Figure 4.8: The stellar masses (in units of $M_\odot$) as a function of redshift (note the logarithmic scale) for CCPC galaxies estimated via traditional SED fitting (black points), and log-normal SFH fitting (blue points). All galaxies with a [4.5] measurement are shown as gray points, which assume a young, bright stellar population with an exponentially decaying SFR. Also included are galaxies photometrically measured by the 3D-HST project (Skelton et al., 2014) by traditional and log-normal SED fitting (red and green points, respectively). Individual galaxies with more than one stellar mass estimate can have large variances (much larger than the photometric errors) based on the model selected with the minimum χ^2 value. For instance, a bright galaxy with a young stellar population can have an order of magnitude lower stellar mass than the same galaxy fit by an old SSP. We are not able to adequately select a unique model for a given galaxy with our data. The masses are the uncertainty-weighted mean values from the adopted population model. The gray error bars are suppressed for clarity, but are generally larger than the other data points, as they are only measured at 4.5μm. The galaxies with the largest stellar masses are typically the 'hyperluminous' sources seen in in Figs 4.2 and 4.3, as expected. These include a probable mix of AGN, low-redshift interlopers, and simply massive systems.

verse at high redshift ($z > 2$), the most massive galaxies will reside (Muldrew et al., 2015). Therefore, protocluster galaxies might be expected to be in the most massive halo systems. Curiously, field galaxies appear to have the same proportion of hyperluminous galaxies as their overdense counterparts (Figs 4.4, 4.5). We will discuss a few other possibilities for the origin of these sources.

In their Fig 23, Guo et al. (2011) plot their semi-analytic model (SAM) Schechter stellar mass functions at redshifts $2 < z < 4$, overplotted with observational data from Pérez-González et al. (2008) and Marchesini et al. (2009). The data diverge from those models in much the same manner as the data presented here, with a number of bright objects not fit by the exponentially declining number density. At a redshift of $z \approx 7$, Bowler et al. (2014) find a similar trend at rest-frame UV wavelengths. It could be possible that at these epochs, Eq 4.2 is not representative of the stellar mass of galaxies. There is some indication that the most massive galaxies observed at very high redshifts ($z > 4$) had not the time to assemble in a ΛCDM universe (Steinhardt et al., 2016), and their halo mass density is larger than theoretical predictions. Some of the hyperluminous sources may be these galaxies and their descendants.

AGN can have strong, non-stellar emission that dominates the flux of the galaxy. These objects are contributing to the number of hyperluminous sources. Roughly half of the hyperluminous objects were spectroscopically selected as part of quasar and AGN surveys, and additionally some of our data come from targeted overdensities surrounding these types of sources, as in Venemans et al. (2007). There is also some evidence that AGN are found in greater density surrounding protoclusters (Casey et al., 2015). However, AGN are quite rare in LBG studies (Magdis et al., 2010).

Some of these objects were detected using NB filters centered on redshifted $H\alpha$ or $Ly\alpha$ lines, which were then confirmed to be emission lines spectroscopically. However,

emission line galaxies can be incredibly faint and show little or no continuum emission (Fynbo et al., 2003), and thus no other absorption or other emission features are identified. Therefore, some of these objects could be [O II] emitters ($\lambda\lambda 3726, 3729$), or other line-emitting galaxies at lower redshifts, and are therefore less distant than expected. Venemans et al. (2007) discuss various tests than can be used to disqualify candidate *Ly*α systems, and they estimate that interlopers are $\leq 10\%$ in their sample. These low redshift interlopers could account for a few of these hyperluminous sources.

4.3.2 Field Galaxy Comparisons

More massive galaxy halos are systematically found in denser environments. This is an observed effect at high redshift, where the two-point correlation function amplitude appears tied to the UV luminosity of LBGs (Ouchi et al., 2004b). It also has a theoretical basis found within large ΛCDM simulations, where the most massive galaxies reside almost exclusively in the densest environments (Muldrew et al., 2015). However, it is readily apparent in Figs 4.4, 4.5 that the luminosity functions of field galaxies are in no way distinct from their CCPC galaxy counterparts. They contain an equal measure of hyperluminous sources, and their respective Schechter function parameters are equivalent.

Taken at face value, the rest-frame NIR emission of galaxies at all redshifts and densities are essentially equal, and therefore their stellar mass contents should be similar as well. In order to reconcile this fact with the points laid out in the previous paragraph, we will suggest a few possible solutions. We note that our data are not sufficient to endorse any of these over another. The null hypothesis is that the galaxy stellar populations at $z > 2$ are the same, regardless of environment.

If galaxies in protoclusters are inherently brighter, as expected, but also had a greater fraction of dusty galaxies or more dust extinction in general, this could balance the mag-

nitude of field and protocluster galaxies. Assuming a Charlot & Fall (2000) dust model, we can predict the dust extinction for a given wavelength and stellar age. At the redshifts measured, it is a low extinction of median $\Delta m \sim 0.6$ mags for a starbursting galaxy and $\Delta m \sim 0.3$ mag for an older population from $2 < z < 4$ at 4.5μm, with the extinction increasing at higher redshifts. This hypothesis would require an extremely convenient steady increase in dust absorption across the range of rest-frame emission to account for the stellar mass difference between protocluster galaxies and the field. The maximum extinction during a starburst is hardly significant ($\Delta m \approx 0.6$ mags), and might not be detected within the uncertainties of m^*.

Another option is that our selection of 'field' sources actually targets marginally overdense systems, and are therefore not isolated enough to be different from the CCPC systems. As a reminder, the initial 'Field' sample was composed of galaxies within the same survey fields as the CCPC galaxies to minimize bias, but were not found within the same volume as a CCPC candidate. After this sample showed no differentiation, a smaller subsample was crafted which contained $N < 4$ galaxies in the same volume as the CCPC. This also did not show any difference in the LFs when compared to the CCPC. Analysis of galaxies limited to the GOODS-S survey also showed no statistically distinct difference, suggesting that in this particular instance, the myriad of spectroscopic selection functions of the CCPC galaxies did not dilute a potential signal.

Contini et al. (2016) analyzed zoomed-in protocluster galaxies in a SAM, and found that for galaxies in the region of a protocluster, but not bound to it at $z = 0$, the galaxy properties (color, mass, etc.) were indistinguishable. It is possible these field galaxies may be similarly camouflaged. A related plausibility is that a large number of interlopers within the overdensity volumes mask a detectable differentiation. If, however, the rarest, most massive galaxies form only in the densest regions of the universe (Muldrew et al., 2015),

then presumably some evidence of these could be solely evident in the CCPC LFs. Recently, Hatch et al. (2016) found evidence suggesting that dense sub-groups in a protocluster at $z \sim 1.6$ exhibited differentiation with respect to the field, while 2/3 of the member galaxies outside of groups showed no variation.

A further possible, but poor, explanation is that the spectroscopic selection of the surveys used are more incomplete in these field regions than in the CCPC volumes, and therefore may well be overdense themselves. However, spectroscopic completeness is strongly correlated to flux for practical purposes, and we see no difference with galaxy densities brighter than m^*. At the present time, we do not have a satisfactory explanation for this discrepancy.

4.3.3 m^* Evolution

In the context of previous works, our results for the redshift dependence of m^* are puzzling. Mancone et al. (2010) and Brodwin et al. (2013) analyzed the *Spitzer* LFs and SFR (respectively) of the same cluster sample at $z < 2$. They conclude that the epoch at which cluster galaxies are undergoing rapid mergers (and therefore mass assembly) is approximately $z \sim 1.5$. Wylezalek et al. (2014) investigated the *Spitzer* LF of clusters and protoclusters at $z \leq 3$, and find no evidence of such a rapid mass assembly at $z \sim 1.5$. They find that a passively evolving stellar population is consistent with the full $m^*(z)$ range of both Mancone et al. (2010) and Wylezalek et al. (2014). Wylezalek et al. (2014) hypothesize that as they are probing higher redshifts, and thus rarer/denser volumes, they do not observe the mass assembly seen at $z \sim 1.5$ by Mancone et al. (2010), which correspond to more common overdensities. As a result, they speculate that it might be possible to observe a different epoch of cluster assembly at higher redshifts ($z > 3$) than their sample. This is akin to a cluster-scale version of galaxy downsizing, where the densest clusters will form

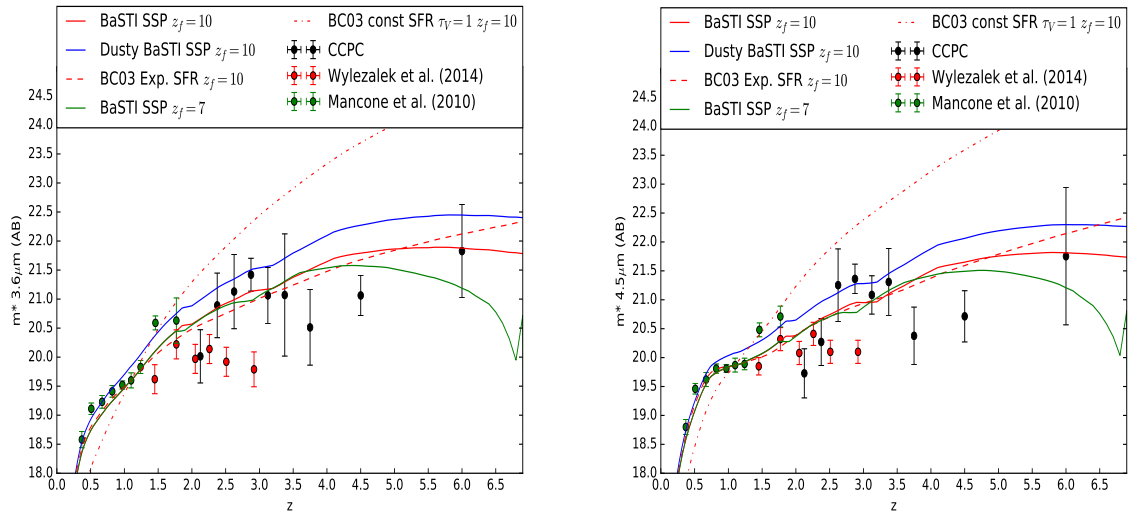

Figure 4.9: We show the evolution of m^* as a function of redshift for 3.6μm (*Left*) and 4.5μm (*Right Panel*). The black points are the CCPC values, while green and red values are from Mancone et al. (2010) and Wylezalek et al. (2014), respectively. The error bars shown for the CCPC are 1σ values computed by bootstrapping. Overplotted are various models computed using EzGal. Briefly, we have included two BaSTI simple stellar populations at formation redshifts $z_f = 7, 10$ (solid green and red lines, respectively), and added a Charlot & Fall (2000) dust prescription to the $z_f = 10$ SSP (blue solid line). From the BC03 models, we included an exponentially declining SFR model with $\tau = 1$ Gyr (dashed red line) and a constant star forming model with an extinction of $\tau_V = 1$ (dash-dot red line). This latter model provides a poor fit to the m^* evolution, while the two SSPs ($z_f = 7, 10$) and exponential model are consistent with the data. Each model is scaled to match the $z = 1.1$ m^* value from Mancone et al. (2010), although this choice is not unique. The scaling simply moves the evolution curves to brighter or fainter magnitudes, but the shape remains constant. Therefore, for any scaling choice, a constant star forming model will provide a poor fit. The dust model was originally calibrated to the same m^* value as the others, but was not re-scaled to show the effect.

the quickest. We do not see any behavior analogous to the high redshift ($1.5 < z < 2.0$) m^* variability of Mancone et al. (2010) in our sample, which would indicate rapid mass assembly.

When our data is analyzed with the two previous works (Fig 4.9), the evolution of the characteristic magnitude (m^*) is fully consistent with a simple stellar population formed at $z_f \leq 10$. Even more puzzling is that the field sample at these redshifts is equivalent in all respects to the overdense sample, apart from the scaling. We gain little insight into the epoch of rapid mass assembly of cluster galaxies, or of the field for that matter. More complex stellar populations, such as an exponential decay model with $\tau \sim 1$ Gyr, can also be consistent with the data. A constant star formation model is not favored, regardless of which $m^*(z)$ value it is calibrated to.

It should be noted that this is not a progenitor-matched study, in that as the redshift increases, the possibility of detecting a weak overdensity will decrease. CCPC candidates at $z > 4$ will likely not evolve into the $z = 2$ candidates from Wylezalek et al. (2014) in a one-to-one fashion. However, some overlap could occur, particularly with their strongest overdensities. From our simple analysis, it appears that *Spitzer* m^* values in dense environments, a tracer of the stellar mass content of these galaxies, is consistent with the passive evolution of a single burst of star formation at $z_f = 10$ over nearly 10 Gyrs ($0.3 < z < 6.6$) of time.

It could be hypothesized that the marginally overdense CCPC candidates could merely be field galaxies. Their inclusion in the luminosity functions (Fig 4.4 and Fig 4.5) could thus mask a weak signal of differentiation in the stellar mass functions of protocluster versus field galaxies. As a simple test of this hypothesis, we re-computed the [4.5] m^* values for sub-samples of galaxies that exist in overdensities of $\delta_{gal} > 2, 3, 4$ and 5. The redshift and magnitude bins remained unchanged from the analysis of the total sample. The m^* redshift

evolution of these subsamples are broadly consistent with the 2σ uncertainties presented Table 4.2. There appears to be no statistically significant correlation between m^* and δ_{gal} within our limited sub-samples. For CCPC galaxies at $z < 3.25$, m^* is brighter by less than 0.3 mags between the subsamples of $\delta_{gal} > 4$ and $\delta_{gal} > 2$, on average. This is consistent with the uncertainties listed in Table 4.2. Significant deviations between the total and sub-samples of $\Delta m^* > 1$ occur only in some high redshift, high overdensity bins that lack significant galaxy numbers ($N \sim 10^1$) to provide a satisfactory solution.

4.3.4 Galaxy Selection Implications

A further curiosity is evident when the evolution of m^* is considered in the context of the galaxy type predominantly represented in our sample. CCPC galaxies generally originate from LBG and line-emitting galaxy ($Ly\alpha$ emitters) spectroscopic surveys, which are selected directly because of their large UV-luminosity/SFR. Fig 4.9 illustrates that a consistent fit to the model of m^* galaxies from $0.3 < z < 6.6$ is a simple stellar population, not a constant star formation model, dusty or otherwise. At low redshifts this is not surprising, as a passive stellar population model has historically been well fit to overdense regions (Stanford et al., 1995). However, the high redshift systems are star forming galaxies *by selection*, and yet the data clearly disfavor the model. The exponential decaying SFR model (Tinsley, 1972) with $\tau = 1$ Gyr is also consistent with the evolution of m^*. However, this type of model is not able to fit roughly $1/4$ of all $z = 0$ galaxy SFHs (Oemler et al., 2013). Exponential decay models with $\tau = 10$ Gyrs look similar to that of the constant star formation model. Fig 4.9 is normalized to match the data of Mancone et al. (2010) at redshift $z < 1.1$.

It could be assumed that by normalizing the star formation model to a different $m^*(z)$ value, the data might be better fit. However, the shape of the model does not change, just its scaling. The constant SFR model will be too bright at lower redshifts and too faint at

higher redshifts, regardless of the scaling. However, this simple observation should not be considered wholly unreasonable. There were a variety of individual sources that were fit with SED templates of exponential decaying SFRs, which are still forming stars at these early epochs of the universe, and the $\tau = 1$ Gyr model in Fig 4.9 is consistent with the data. At these wavelengths, the predictions between a young, SSP and a decaying SFR cannot be disentangled within the uncertainty of the data.

An important factor to consider in the context of this entire work (not just the m^* evolution) is that our spectroscopic galaxy sample does not consist of the majority of galaxies at $z > 2$. In fact, van Dokkum et al. (2006) showed that approximately 80% of galaxies are not LBGs between $2 < z_{phot} < 3$. Many of these other objects are distant red and dusty star-forming galaxies (DRGs and DSFGs). This trend appears to become even more pronounced at higher redshifts. In the range of $3 < z < 4$, only 14% of galaxies with photometric redshifts would be identified as LBGs, with the remainder being nearly split between DSFGs and DRGs (Spitler et al., 2014). It is unclear, without spectroscopic confirmation of protocluster membership, how these DRGs/DSFGs might cluster differently than their LBG counterparts, or if they might have LFs that vary with environment.

These DRGs are forming a not-insignificant amount of stars, on the order of a 20% contribution to the cosmic star formation density at $1.5 < z < 2.5$ (Webb et al., 2006). They therefore may not be completely unlike the spectroscopic galaxies in our sample. Unfortunately, Spitler et al. (2014) noted that most of these DRGs are much fainter (~ 2 mags) than the canonical spectroscopic limit of $R_{AB} \leq 25.5$ for current instrumentation. For the present, it appears that this question will remain unanswered in the context of a spectroscopically-confirmed protocluster sample like the CCPC.

Although we do not expect to have DRGs/DSFGs in the CCPC, which are the dominant galaxy populations at high redshift, we can make comparisons to cluster and protocluster

candidates that do have these systems at lower redshift. In fact, we have already performed such an analysis, in that Mancone et al. (2010) and Wylezalek et al. (2014) do not rely solely on spectroscopic redshifts for cluster membership. Mancone et al. (2010) compute photometric redshift probabilities from deep, multi-wavelength data, while Wylezalek et al. (2014) utilize *Spitzer* color cuts to identify high redshift galaxies in their overdensities. Both of these techniques are sensitive to galaxy populations not characterized by bright UV continuum selection (e.g. LBGs). As the entirety of the m^* evolution is consistent with a simple stellar population, passively evolving, it is plausible that the LFs of DRGs and spectroscopically confirmed LBGs may not be significantly divergent at these redshifts. Indeed, the redshifts at which Wylezalek et al. (2014) and this work overlap are in agreement within the uncertainties, despite the differences in selection.

4.3.5 Inferred Stellar Masses and Galaxy Properties

SED fitting is an incredibly useful tool for estimating redshifts and galaxy properties at high redshift over a range of populations (Bolzonella et al., 2000; van Dokkum et al., 2006). However, it is possible to fit more models and parameters to the data than can actually be constrained. In addition, there are significant degeneracies among model parameters that can match the same data at these high redshifts (Papovich et al., 2001; Magdis et al., 2010), such as the well known age-metallicity-dust degeneracy. The models differ among themselves, with varying treatment of thermally-pulsating asymptotic giant branch stars or the adoption of varying IMFs (Bruzual & Charlot, 2003; Percival et al., 2009). Furthermore, with little *a priori* knowledge of the uncertain galaxy zoo extant at high redshift, the difficulties compound. These issues are indeed true in the case of the CCPC, with our limited wavelength coverage, as well as in numerous other studies. However, some properties can be loosely constrained by our data, and is therefore a useful exercise if one is cognizant of

the limitations of SED fitting.

Primarily, the rest-frame NIR data provided by *Spitzer* provides a proxy for the underlying stellar mass of the CCPC galaxies at high redshift. Unfortunately, for the same reason they are a powerful tool for measuring stellar mass, these colors provide little information in determining any other property of the underlying galaxy (e.g. passive versus star forming, metallicity variations, formation redshift; see Cooke et al., 2014). *Hubble Space Telescope* filters (e.g. F125W, F140W, F160W) that measure rest-frame optical bands at the redshifts of the CCPC are able to generally distinguish between passive and star-forming galaxies. In Fig 4.6, a number of model stellar populations are plotted as a function of redshift and F160W-[4.5] color. A clear bifurcation is shown at $z \sim 2$ which grows more pronounced at larger redshifts. Within the photometric errors of our colors, it is not possible to assign a preferred star formation rate or formation redshift to the CCPC galaxies. Clearly a range of stellar populations may exist. Dust obscuration is also uncertain, as a typical reddening of these colors is ~ 0.5 mags or less for the BaSTI SSPs. This which is the 1σ photometric uncertainty in many cases.

To briefly illustrate the perils of mass estimation among various models, let us take an idealized example of a galaxy with a measured [4.5] = 20 AB magnitude. We can infer the stellar mass of this system from a menagerie of models available using a formation redshift $z_f = 10$ and solar metallicity. At $z = 2$, the observed system can have a range of $1 \times 10^{11} < M_\star < 3 \times 10^{11}$ $M_\odot$ for the extreme cases of a BC03 SFG to a BaSTI SSP, respectively. At a redshift $z = 6$, this gap can widen to roughly $4.5 \times 10^{11} < M_\star < 9.5 \times 10^{11}$ $M_\odot$. Lowering the formation redshift will also decrease the implied stellar mass by a factor of ≤ 3. Notice that this example did not take into account any photometric uncertainties, dust, metallicity variations, or flux measurements at other wavelengths. A change in z_f can have an outsized role if it is close to the redshift of the galaxy (e.g. a young galaxy). An

old, bright galaxy can have 10× the stellar mass than that of a young system of the same luminosity.

In practice, the χ^2 values from successive SED fits were observed to not change significantly. This was in spite of their sometimes drastically different stellar populations (e.g. quiescent versus star forming). A brief examination of Figure 4.6 reveals that various models can exist within the color uncertainty of the CCPC galaxies. We wish to caution readers that SED fitting can have a difficult time *excluding* models, especially with the limited wavelength coverage presented here. Therefore, stellar mass uncertainties are dominated by systematic variations in the models and the subsequent fitting procedure.

Comparing the model fits on a system-by-system basis provides a cautionary tale for determining galaxy properties (mass, age, metallicity) via the SED fitting method. We applied our algorithm, for both log-normal and more traditional SEDs, to the data we measured, in addition to a companion photometric catalog in the CANDELS fields (3D-HST; Skelton et al., 2014). The stellar mass estimates could vary by an order of magnitude or more for a single galaxy, but we found *no* statistically significant trends among the combination of two data sets and the two SFH prescriptions (log-N versus traditional). Although there might appear to be a mean offset of formation redshift between our data and the 3D-HST catalog of $< z_f - z_f(3D) >\sim 2$, for instance, the scatter between the two ($\sigma = 4.5$) is much larger. The mean mass and χ^2 differences follow much the same pattern, where occasionally a galaxy will be better fit or more massive via log-normal fitting, but a subsequent galaxy will have the opposite effect. This appears to be a classic case of overfitting the data, with various models supporting divergent implications (SFG vs. quiescent) being equal fits to the photometry. It does not appear we are able to constrain the stellar populations or masses for the CCPC galaxies with any reliability.

Despite these concerns, our estimated stellar masses are not wholly unreasonable. We

compare our mean $M_\star$ values to the sample investigated by Magdis et al. (2010) of LBGs at $z \sim 3$ with *Spitzer* data. Their mean stellar masses are 2.8×10^{10} $M_\odot$ and 4.2×10^{10} $M_\odot$, depending on which suite of models they use. Our mean values are 3.0×10^{10} $M_\odot$ for log-normal fitting and 6.7×10^{10} $M_\odot$ using our traditional SEDs. Their catalog also contains a few very bright, non-AGN sources that exceed $M_\star \geq 5 \times 10^{11}$ $M_\odot$, much like our own results (the hyperluminous sources). Although individual objects may suffer from systematic uncertainties in the $M_\star$ estimates, the CCPC as a whole is a reasonable match to other stellar mass studies of bright galaxies at high redshift.

4.4 Summary

Although longitudinal data is required to perfectly map the evolution of galaxies (Abramson et al., 2016), astronomy must content itself with studies that contain as minimal inherent bias as possible. This manuscript details the *Spitzer* photometry of protocluster galaxies in the Candidate Cluster and Protocluster Catalog. The catalog probes galaxies between redshifts $2 < z < 6.6$ in dense environments. We built luminosity functions of the galaxies in various redshift bins at 3.6 and 4.5 μm wavelengths. These measure the rest-frame NIR emission of the galaxy populations to trace their stellar mass as a function of redshift.

The galaxies in both the field and CCPC samples contain extremely bright sources up to 5 magnitudes brighter than the characteristic magnitude m^*. These galaxies are divergent from the shape of the Schechter function, and exist at nearly all redshifts. Many of these are expected to be bright AGN and a few ($< 10\%$) low redshift interlopers. Semi-analytic models do not predict that these types of galaxies should exist (Guo et al., 2011), although they have been observed previously at similar redshifts (Pérez-González et al., 2008; Marchesini et al., 2009). Their nature is not yet established.

Field samples of galaxies are also photometrically measured, and remarkably the luminosity functions of the overdense regions are not statistically distinct from their field counterparts. In our current understanding of galaxy formation, the expectation is that the most massive galaxies at any epoch will be found in the densest environments. In Section 5.4 we analyzed a number of possibilities that might explain this phenomenon, but cannot find a satisfactory conclusion. We believe this to be the most fundamental result of this work.

We model the fitted LF parameter m^* as a function of redshift in the context of various stellar population models. By including the measurements at lower redshifts from Mancone et al. (2010) and Wylezalek et al. (2014), we find that a passively evolving stellar population formed in a single burst at high redshift ($z_f = 7 - 10$) is consistent with the data at all redshifts ($0.3 < z < 6.6$). An exponentially decaying star formation model with $\tau = 1$ Gyr is also in agreement with the data. Despite the fact that the majority of CCPC galaxies were spectroscopically selected based on their star forming properties (e.g. LBGs and line emitters) a constant star forming model is a poor fit to the observed $m^*(z)$.

A SED fitting technique has provided stellar mass estimates and some general information about the properties of the CCPC galaxies. We use supplemental *Hubble Space Telescope* data to probe the rest-frame optical emission to measure stellar colors for additional model constraints. However, we are careful to note that even with greater wavelength coverage than that which is presented here, SED fitting can be fraught with degeneracies (dust, age, metallicity) and inter-model uncertainties. Overall, the CCPC appears to be composed of $M \gtrsim 10^{10}\ M_\odot$ galaxies with mean formation redshifts $z_f > 7$. Apart from these broad statements, we cannot provide reliable dust content, metallicity information, or unique model fits to individual CCPC sources.

The spectroscopic redshifts used in the Field samples of galaxies (Section 4.2) came

from the following sources (and references therein), as compiled by NED: Worseck et al. (2008); McIntosh et al. (2004); Noll et al. (2004); Le Fèvre et al. (2005) Bond et al. (2012); Balestra et al. (2010); Wuyts et al. (2008); Santini et al. (2009); Le Fèvre et al. (2004) Bonzini et al. (2012); Luo et al. (2008); Mainieri et al. (2008); Bond et al. (2011); Straughn et al. (2011) Mignoli et al. (2005); Xue et al. (2010); Trump et al. (2011); Treister et al. (2009); Hewett & Wild (2010); Wuyts et al. (2009) Tanaka et al. (2013); Eales et al. (2003); Barger et al. (2002); Menéndez-Delmestre et al. (2009); Reddy et al. (2006) Pope et al. (2006); Conselice et al. (2011); Pirzkal et al. (2013); Erb et al. (2006); Steidel et al. (2004); Papovich et al. (2001) Fernández-Soto et al. (1999); Bertincourt et al. (2009); Fernández-Soto et al. (2001); Hayashi et al. (2009); Ly et al. (2009) Ramos Almeida et al. (2009); Whitaker et al. (2011); Park et al. (2010); Gobat et al. (2012, 2013, 2011) Kocevski et al. (2009); Erb et al. (2003); Shapley et al. (2005); Colbert et al. (2006) Keel et al. (2002); Steidel et al. (2003); Abraham et al. (2004); Vanzella et al. (2002); Wuyts et al. (2007); Kilic et al. (2005) Sealey et al. (1998); Galametz et al. (2013); Gavignaud et al. (2006); Worseck & Wisotzki (2006); Silverman et al. (2010) Szokoly et al. (2004); Elmegreen & Elmegreen (2010); Brammer et al. (2013); Trump et al. (2014); Kriek et al. (2008); Wolf et al. (1999) Doherty et al. (2010); Brusa et al. (2010); Söchting et al. (2012); Song et al. (2014) Pentericci et al. (2002); Kurk et al. (2004a); Pentericci et al. (2000); Croft et al. (2005); Kurk et al. (2004b); Roettgering et al. (1997) Lowenthal et al. (1997); Papovich et al. (2005); Moth & Elston (2002); Wang et al. (2004); Tadaki et al. (2011) Yoshikawa et al. (2010); Matsuda et al. (2011); Weiner et al. (2005); Georgakakis et al. (2006); Smail et al. (2004); Onodera et al. (2010) Shapley et al. (2004); Vogel & Reimers (1995); Peter et al. (2007); Chapman et al. (2004b); Moorwood et al. (2000) Cristiani & D'Odorico (2000); Stalin et al. (2010); Melnyk et al. (2013); Simpson et al. (2014); Donley et al. (2010); Casey et al. (2011) Salimbeni et al. (2009); Grazian et al. (2006); Rangel et al. (2014); Osmer et al.

(1994) Silverman et al. (2005); Storrie-Lombardi et al. (1996); Dannerbauer et al. (2004); Cowie et al. (2004); Erb et al. (2004); Adams et al. (2011) Barger et al. (2000); Lowenthal et al. (2009); Phillips et al. (1997); Trouille et al. (2008); Swinbank et al. (2004) De Breuck et al. (2004); Anderson & Margon (1987); Digby-North et al. (2010); Erb et al. (2011); Harrison et al. (2012) Lehmer et al. (2009b); Petitjean & Srianand (1999); Hamann et al. (1997); York et al. (1991); Cowie et al. (1995) Zheng et al. (2004); Rodney et al. (2014); Yang et al. (2011); Donley et al. (2007); Georgantopoulos et al. (2011); Wirth et al. (2004) Dannerbauer et al. (2006); Kulas et al. (2012); Law et al. (2012); Dobrzycki & Bechtold (1996); Tripp et al. (1997) Tanner et al. (1996); Hainline et al. (2011); Schreier et al. (2001); Kirkpatrick et al. (2012); Brusa et al. (2009a); Bond et al. (2014) Daddi et al. (2005); van Breukelen et al. (2005); Knudsen et al. (2008); Bauer et al. (2002); Dawson et al. (2001) Fasano et al. (1998); Conselice et al. (2003); Riechers et al. (2011); Ivison et al. (2011); Wang et al. (2006); Barger et al. (2001b) Postman et al. (2001); Reimers et al. (1989); Iwata et al. (2005); Castro-Rodríguez & López-Corredoira (2012); Rafelski et al. (2011) Teplitz et al. (2007); Diener et al. (2013); Bothwell et al. (2013); Lanzetta et al. (1996); Campos et al. (1999); Coppin et al. (2008) Wolfe & Prochaska (2000); Bielby et al. (2013); Rafelski et al. (2009); Rigby et al. (2008); Akiyama (2005) Ross et al. (2012); Lilly et al. (2007); Sharp et al. (2002); Barger et al. (2001a); Rodriguez-Pascual et al. (1995); Simcoe et al. (2006) Venemans et al. (2007); McCarthy et al. (1996); Tapken et al. (2007); Santos et al. (2004); Vernet & Cimatti (2001) Songaila (1998); Sarajedini et al. (2011); Chapman et al. (2004c); Magdis et al. (2010); Fynbo et al. (2003); Storrie-Lombardi & Wolfe (2000) Gavignaud et al. (2008); Cristiani et al. (2000); Feruglio et al. (2011); Shapley et al. (2001); Rottgering et al. (1995) Civano et al. (2011); Fynbo et al. (2001); Möller & Fynbo (2001); Stanford et al. (2004); Webb et al. (2003); Savaglio et al. (1997) Srianand (1996); Venemans et al. (2005); Le Fevre et al. (1996); Maschietto et al. (2008); Ciardullo et al. (2012) Bond

et al. (2010); Straughn et al. (2008); Xu et al. (2007); Lehmer et al. (2009a); Cantalupo et al. (2007); Lu et al. (1998) Trump et al. (2009); Grove et al. (2009); Barger et al. (2008); Petry et al. (1998); Tran et al. (2005) Sand et al. (2005); Hawkins & Veron (1993); Veron & Hawkins (1995); Inoue et al. (2011); Nestor et al. (2011); Shapley et al. (2006) Chapman et al. (2004a); Watabe et al. (2009); Hayashino et al. (2004); Steidel et al. (2000); Vanzella et al. (2004) Steidel et al. (1999); Mehlert et al. (2002); Ellison et al. (2001); de Bruyn et al. (1996); Polletta et al. (2008); Lemoine-Busserolle et al. (2010) Krumpe et al. (2008); Fiore et al. (2012); Vanzella et al. (2006); Tang et al. (2014); Gnerucci et al. (2011) Jangren et al. (2005); Boutsia et al. (2011); Hornschemeier et al. (2001); Barger et al. (2003); Hainline et al. (2006) Scott et al. (2000); Schaye et al. (2000); Ouchi et al. (2008); Vanzella et al. (2008) Fontanot et al. (2007); Tozzi et al. (2009); Brusa et al. (2009b); Patnaik et al. (1992); Saito et al. (2008) Shim et al. (2011); Vanzella et al. (2010); Rhoads et al. (2009); Daddi et al. (2009); Ouchi et al. (2004a) Yoshida et al. (2006); Savaglio et al. (1994); Schneider et al. (1991); Iwata et al. (2007); Tapken et al. (2006); Wang et al. (2009) Dawson et al. (2004); McLure et al. (2006); Ouchi et al. (2005); Saito et al. (2006); Maiolino et al. (2009) Hu et al. (2010); Ranalli et al. (2013); Zheng et al. (2013); Finkelstein et al. (2009a,b); Raiter et al. (2010) Stanway et al. (2007); Vanzella et al. (2005); Coppin et al. (2010); Pentericci et al. (2011); Bunker et al. (2004) Malhotra et al. (2005); Rhoads et al. (2005); Buchner et al. (2014); Hathi et al. (2008); Finlator et al. (2007); Stanway et al. (2004b) Dow-Hygelund et al. (2007); Finkelstein et al. (2011); Husband et al. (2013) Péroux et al. (2001); Djorgovski et al. (2003); Casey et al. (2012); Taniguchi et al. (2009); Martin et al. (2008); Henry et al. (2012) Murayama et al. (2007); Davies et al. (2010); Hu et al. (1996); Carilli et al. (2013); Elston et al. (1996) Wampler et al. (1996); Riechers et al. (2006); Kajino et al. (2009); Stark et al. (2013, 2011); Kakazu et al. (2007) Stern & Spinrad (1999); Weymann et al. (1998); Dawson et al. (2002); Neri et al. (2014); Reshetnikov & Vasil'Ev (2002) Spin-

rad et al. (1998); Ando et al. (2004); Tan et al. (2014); Kashikawa et al. (2011); Shimasaku et al. (2006); Toshikawa et al. (2014) Jiang et al. (2013); Nagao et al. (2007); Dawson et al. (2007); Lemaux et al. (2009); Anderson et al. (2001) Hu et al. (2004, 1999); Curtis-Lake et al. (2012); Ouchi et al. (2010); Grazian et al. (2012); Ouchi et al. (2009) Schenker et al. (2012); Ono et al. (2012); Stanway et al. (2004a); Jiang et al. (2011); Taniguchi et al. (2005) Kashikawa et al. (2006); Iye et al. (2006); Nagao et al. (2004, 2005) and Rhoads et al. (2004)

Chapter 5

Comparison of the Millennium Simulation with the Candidate Cluster and Protocluster Catalog

5.1 Introduction

To understand the physical processes involved in shaping clusters and their constituent galaxies, it is important to study their evolutionary history. Astronomical timescales are generally long enough that humans are not afforded the luxury of a longitudinal study of a galaxy. However, we can attempt to piece together the history of these objects by examining galaxies and their environment at ever increasing redshift. High redshift protoclusters, which are simply the un-assembled components of a galaxy cluster, provide a brief glimpse into the rapid formation and evolution of galaxies in their first few Gyrs of existence. At redshifts $2 < z < 7$, the main halos of cluster progenitors are growing in mass by 3 orders of magnitude (Chiang et al., 2013), galaxies are approaching peak star formation density (Madau & Dickinson, 2014), and the first examples of clusters can be observed (Gobat

et al., 2011; Wang et al., 2016). This epoch exists as a crossroads of rapid evolution, and serves as a fundamental test of theoretical predictions.

Protoclusters are identified observationally at high redshift ($z \geq 2$) as galaxy overdensities. Protoclusters exist at even lower redshifts ($z \sim 0.4$ Gonzalez et al., 2005), can be characterized as galaxy groups, and may not be in dynamical equilibrium. It is difficult to know which galaxies in the overdensity will be bound to the structure at $z = 0$, and which are interlopers. To provide some context for the observations of candidate protoclusters, simulations can be used to track the evolution of these structures, provided the theoretical underpinnings of the simulations match reality.

There has been an increased focus on characterizing the evolution of protoclusters using simulations as computational power increases. Structure formation simulations provide broad statistics on the number density, physical size, dynamics, mass, and richness of overdensities. Some of these works have utilized N-body and semi-analytic models (SAMs), such as Suwa et al. (2006) and Contini et al. (2016). Others have focused on mining published databases of simulations, like the Millennium Run (Springel et al., 2005), to run their analysis (Chiang et al., 2013; Muldrew et al., 2015; Chiang et al., 2017). The general finding of these works (Suwa et al., 2006; Chiang et al., 2013; Contini et al., 2016; Muldrew et al., 2015; Chiang et al., 2017) is that protocluster properties, such as their effective radii (R_e), galaxy and mass overdensities (δ_{gal}, δ_m), and primary halo masses, are positively correlated to their $z = 0$ cluster halo mass. The scatter among these correlations can be quite large ($\sim 20\%$ variation in radii in Fig 2 of Chiang et al., 2013), suggesting that these simulated protoclusters exist in a menagerie of possible evolutionary states at a given redshift.

Many basic properties of observed protoclusters and their constituent galaxies to be in agreement with the simulations. This includes the number density of galaxies as a function of redshift (Diener et al., 2013) and overdensities (Table 5 in Chiang et al., 2013), for

example. Exceptions include a color discrepancy for drop-out galaxies (Toshikawa et al., 2016) and an excess of observed massive overdensities over what would be expected in ΛCDM (Kang & Im, 2015).

Numerous large-scale, high redshift spectroscopic surveys have begun in the last decade. The Great Observatories Origins Deep Survey (GOODS) with VIMOS instrument (Balestra et al., 2010), zCOSMOS (Lilly et al., 2007), and HETDEX (Hill et al., 2008; Adams et al., 2011) have all contributed significant numbers of spectroscopically confirmed galaxies at $z > 2$. Even more ambitious surveys are planned that will increase the sky coverage and breadth of objects (a large list is provided in Overzier, 2016). However, spectroscopic surveys are generally biased towards star forming systems such as Lyman Break Galaxies (LBGs) or emission line galaxies[1]. Studies suggest that such galaxies may represent only the tip of the iceberg, with a large fraction (80% in some cases) of very massive galaxies ($M_\star > 4\times10^{10}\ M_\odot$) belonging to Distant Red Galaxy (DRG) or Dusty Star Forming Galaxy (DSFG) population (van Dokkum et al., 2006; Spitler et al., 2014).

To date, there are a few hundred spectroscopically confirmed protoclusters (Chiang et al., 2013; Franck & McGaugh, 2016b,a), spread across $2 < z < 6.6$. Large simulations can provide large comparison samples. In the Millennium Run, there exist an order of magnitude more clusters in the box at $z = 0$ than all observationally known protoclusters. Simulated clusters can be analyzed at many epochs prior to their virialization. As a tool, simulations can provide a zoo of mock protoclusters in a variety of dynamical states, masses, and redshifts that can be compared to observations.

The Candidate Cluster and Protocluster Catalog (CCPC) includes 216 spectroscopically confirmed overdensities of galaxies spanning $2 \leq z < 6.6$ (Franck & McGaugh, 2016b,a). These objects were selected from archival spectroscopic galaxy catalogs with the minimum

[1]Some recent protocluster searches now target absorption features (Cai et al., 2016; Ogura et al., 2017)

requirements of a galaxy overdensity of $\delta_{gal} \geq 0.25$ and $N = 4$ galaxy members within a tube of maximum size $R \leq 20$ cMpc and length $\Delta z \leq 40$ cMpc (the expected size of the largest protoclusters in ΛCDM ; Muldrew et al., 2015). The CCPC contains the largest number of spectroscopically confirmed protoclusters at $z \geq 2$, as well as the most distant protocluster ($z = 6.56$) known to date.

Franck & McGaugh (2017) utilized archival *Spitzer* images in 3.6 and 4.5μm filters to trace out the evolution of the stellar mass in CCPC galaxies in nine epochs from $2 \leq z < 6.6$. The photometric data were fit to a Schechter function (Schechter, 1976) to calculate the characteristic magnitude $m^*(z)$ at each epoch. Intriguingly, the $m^*(z)$ remains approximately constant across the redshift range in question at both wavelengths. Including results from lower redshift (Mancone et al., 2010; Wylezalek et al., 2014), the evolution of $m^*(z)$ is consistent with a simple stellar population (SSP) formed in a single burst at high redshift ($z_f > 7$).

A further curious observation was that galaxies in the same field as the CCPC galaxies, yet not found in the overdensities, had indistinguishable $m^*(z)$ values at each time period and at both wavelengths from their overdense counterparts. Galaxies found in overdense regions might be expected to have greater stellar masses than field galaxies at the same redshift, based upon ΛCDM predictions (Muldrew et al., 2015). For a protocluster at $z \sim 4.5$ identified by Lemaux et al. (2017), there is some observational evidence that the protocluster galaxies have greater mass. At low redshift, there is a clear density correlation between galaxies in different environments (Dressler, 1980). However, this distinction is not always obvious at high redshift.

The focus of this work is to attempt to map the CCPC and its component galaxies in the framework of a large ΛCDM simulation. Provided that the CCPC and the simulation are (relatively) well matched, a context for the significance of the CCPC structures can

be established. How many candidate structures identified will become clusters at $z = 0$? Are the observed overdensities, sizes, and distributions consistent with what the simulation suggest? The growth of the stellar mass function of galaxies in high density environments is of particular focus. The results of Franck & McGaugh (2017) suggest that not only are overdense and field regions statistically equivalent, there is a constant $m^*(z)$ value at 3.6 and 4.5μm, consistent with the passive evolution of an old stellar population. Do simulations provide insight into this observation?

Results in this manuscript are based on the assumption of a ΛCDM universe with values of $\Omega_\Lambda = 0.685$, $\Omega_m = 0.315$, and $H_0 = 67.3$ km s^{-1} Mpc^{-1}, in line with the 2013 values of the Planck Collaboration et al. (2014), unless otherwise noted. At $z = 2$, the age since the Big Bang is 3.27 Gyrs, and the comoving angular scale is 1.547 cMpc arcmin^{-1}. All previous work on the CCPC assumed slightly different parameters ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1}). The resulting differences in age and angular scale are marginal, with values of $\Delta t \sim 0.05$ Gyrs and $\Delta\Theta \sim 0.04$ cMpc arcmin^{-1} at $z = 2$ (Wright, 2006). All magnitudes are based on the AB magnitude system. *Spitzer* IRAC magnitudes at 3.6 and 4.5 μm wavelengths will be abbreviated as [3.6] and [4.5], respectively.

5.2 Simulations

This work makes use of the Millennium Run (MR) simulation (Springel et al., 2005). The MR is a cube of sides 500 h^{-1} cMpc that houses 2160^3 particles of mass 8.6×10^8 h^{-1} $M_\odot$. In its original incarnation, the MR's 64 snapshots extended from $0 < z < 127$. The version used in this work (Henriques et al., 2015), scaled to Planck cosmology (Planck Collaboration et al., 2014), has a range extending into the future, $-0.1 < z < 56.4$. The final snapshot (63) is approximately 1.63 Gyrs in the future of the assumed cosmology,

while snapshot 58 represents $z \sim 0$. All redshifts from the simulations include peculiar motions.

The Millennium Run and its value added products have been widely used to investigate protoclusters (Diener et al., 2013; Chiang et al., 2013, 2014; Muldrew et al., 2015; Kang & Im, 2015; Toshikawa et al., 2016). This makes it the natural choice with which to compare to the CCPC. Chiang et al. (2013) and Muldrew et al. (2015) analyzed the entire MR box for the evolution of protoclusters and clusters at various mass ranges, albeit with slightly different methods and cosmological parameters. In addition to using the entire simulation box, we also employ the use of mock catalogs constructed in lightcones to 'observe' the simulation.

5.2.1 Lightcones

A number of protocluster searches have employed the tactic of using lightcones to test the effectiveness of their search algorithms on real data. Mock catalogs allow one to trace the descendants of structures identified as protoclusters at high redshift. This work adopts a similar approach, but does not presuppose the simulation results are necessarily related to the universe. If the observations match the results, then the underlying physics governing structure formation and galaxy evolution can be further studied.

Henriques et al. (2015) have provided simulated lightcones in addition to their SAM of the MR simulation, sampling the volume with different orientations and at different epochs to construct a continuous distribution of galaxies. Each lightcone has a 2 degree diameter. The stellar population model that is the default for these lightcones is from Maraston (2005). A total of 9 lightcone catalogs were obtained by querying the Millennium Database. Each cone is limited in depth to $R_c \leq 25.5$ mags, which is approximately the ground-based spectroscopic limiting magnitude of many of the archival surveys (Stern et al., 2003) used

in the assembly of the CCPC (see other references in Franck & McGaugh, 2016b,a). This is not an absolute rule, of course, as strong emission line galaxies can be fainter than this limit but still be detected (Fynbo et al., 2003).

It is important to note here that the $R_c \leq 25.5$ selection is unimportant to the results of this paper. In future sections we will perform tests to determine the effect of magnitude cuts on the simulated data set. However, the primary result concerns the bright end of the luminosity function and is insensitive to selection effects.

For a comparison simulation sample, the Theoretical Astrophysical Observatory (TAO) All-Sky Virtual Observatory (ASVO)[2] provided a suite of SAMs (Somerville et al., 2008; Croton et al., 2016) imprinted upon the Millennium simulation with a varying dust model (Devriendt et al., 1999). This group of models adopts a slightly different set of cosmological parameters ($\Omega_\Lambda = 0.75$, $\Omega_\mathrm{m} = 0.25$, $H_0 = 73.0$ km s^{-1} Mpc^{-1}), but as was shown previously, these have little effect on the results presented here. The limiting magnitudes of these lightcone catalogs are equivalent to those used from Henriques et al. (2015).

5.2.2 Tracking Structure

The MR simulation provides a number of methods to trace the merger trees of galaxies across cosmic time. In this paper, we will utilize the *treeRootId*, which matches all halos that will eventually merge, to a single, unchanging value across all snapshots (i.e., simulation time-steps). We assume that the minimum halo mass that defines a cluster to be $M_{crit,200} = 10^{14}\ h^{-1}\ M_\odot$. For each cluster, we identified all galaxy *treeRootIds* associated with the main halo of minimum stellar mass of $M_{\star,z=0} \geq 10^{10}\ h^{-1}\ M_\odot$. This is approximately the median stellar mass estimated for CCPC galaxies in Franck & McGaugh (2017) at $z \geq 2$. The number of halos greater than or equal to the minimum cluster mass at $z \sim 0$

[2] tao.asvo.org.au

Table 5.1. Sample Terminology

Sample Name	Description
CCPC	Observed Galaxies in the Candidate Cluster and Protocluster Catalog (Franck & McGaugh, 201
treeRootId	Simulated galaxies that are/will be part of a (proto)cluster based on their Millennium identi
Overdense	Simulated galaxies from the Millennium Run identified using the CCPC overdensity identifier
Field	Simulated galaxies not contained in the Overdense/*treeRootId* samples

Note. — A concise overview of the sample names that are used within this work.

(*snapnum*=58) is $N = 2293$. This is similar to the number found by Muldrew et al. (2015) with a slightly different cosmology. The cluster mass function is quite steep: only 76 structures exist with $M_{crit,200} > 5 \times 10^{14}\ h^{-1}\ M_{\odot}$, and a mere 8 objects have $M_{crit,200} \geq 10^{15}\ h^{-1}\ M_{\odot}$.

With the *treeRootId* identifier code, one can examine simulated galaxies that belong to a cluster halo at $z = 0$. We have included a summary of the sample names used throughout this work for easy reference in Table 5.1.

5.3 Results

Perhaps the most important check that can be performed when comparing data to simulations is to ensure that their bulk properties agree with one another. If the comparison lacks a robust overlap, then the remaining results may be of questionable importance.

A simple test is the number density of galaxies in a lightcone above a minimum magni-

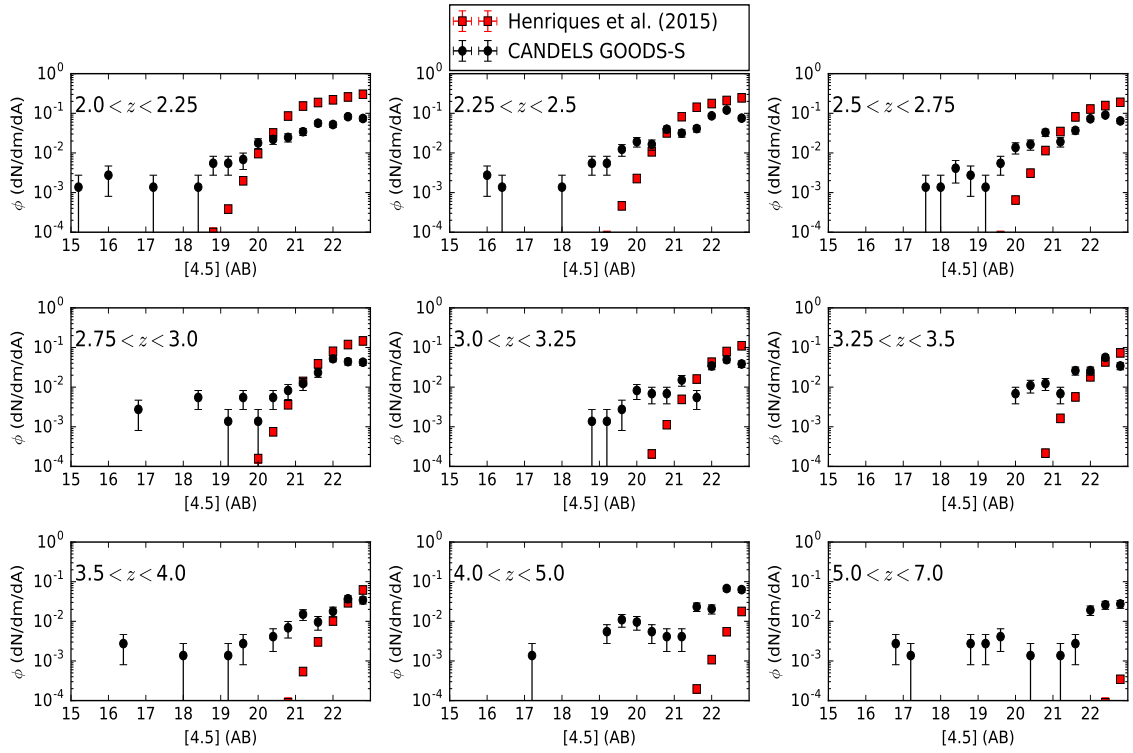

Figure 5.1: The [4.5] magnitude Luminosity Functions of galaxies in the CANDELS GOODS-S field (black circles) (Franck & McGaugh, 2016b,a). This field is the most complete spectroscopic catalog used in the CCPC, and as such offers the closest comparison to a simulated lightcone with 'perfect' completeness. The red squares represent the simulated galaxy luminosity functions from the lightcones of Henriques et al. (2015). At every redshift, the data have greater number densities of bright objects than the simulations predict. Uncertainties shown are $\sqrt{N}$. The number density of bright sources is not affected by the $R_c \leq 25.5$ minimum magnitude cut. Bright galaxies simply do not exist in the Millennium catalogs in sufficient numbers to explain this discrepancy.

tude in the simulated versus real data. We are not including any cuts or selections based on the fate of these systems (protocluster vs field galaxies), but merely querying all available data in a survey volume. In the CANDELS GOODS-S field, the deepest and most complete spectroscopic survey used in the CCPC (for partial survey details see Balestra et al., 2010), the number density of galaxies with [4.5] < 22.8 mag is $n = 6336\ deg^{-2}$.

Using 9 lightcones from Henriques et al. (2015), which have a surface area of $9\pi\ deg^2$ (compared to the $\sim 0.45^2\ deg^2$ of GOODS-S), provide a number density of $n = 9138\ deg^{-2}$ to the same limiting magnitude. If one considers the uncertain completeness function of any high redshift spectroscopic survey, this $< 2\times$ discrepancy factor could be expected. Potential source blending, significant populations of distant red galaxies difficult to target spectroscopically (van Dokkum et al., 2006; Spitler et al., 2014), and edge effects may all contribute to this finding. It could also be possible that this is a real discrepancy, and that the simulations overproduce galaxies.

However, other details may prove to be more challenging. For example, if the redshift range of the previous simple test is restricted to $3 < z < 4$, the CANDELS field has a larger number density of these galaxies ($n = 1348\ deg^{-2}$) with respect to the MR ($n = 922\ deg^{-2}$). This discrepancy is more difficult to explain, especially for relatively bright sources, whereas spectroscopic completeness is anti-correlated with redshift in general. Figure 5.1 illustrates this issue by plotting the Luminosity Functions of spectroscopic galaxies in the CANDELS GOODS-S field with the expectations from Henriques et al. (2015). There exist many more bright galaxies in the actual data than the Millennium Run produce. In Section 5.4 we will examine this result in more detail.

As mentioned earlier, the $R_c < 25.5$ limiting magnitude imposed on the simulated catalogs has little change on this result. The number density of galaxies with $z > 2$ and [4.5] < 22.8 in the simulation is still only a factor of 2 larger ($n \sim 13000\ deg^{-2}$) than in

the incomplete CANDELS field. As we limit the data to higher redshifts again ($3 < z < 4$) there still exist only $n \sim 1100\ deg^{-2}$ mock galaxies in the simulation (an increase of $\sim 200\ deg^{-2}$). If we take $\sqrt{N}$ as our shot noise for the simulation, the difference between the real data and the simulation is 13σ. If we were able to accurately account for spectroscopic incompleteness of the real data, this discrepancy would increase.

5.3.1 Overdensity Algorithm

We developed a number of simple algorithms in Franck & McGaugh (2016b) and Franck & McGaugh (2016a) that sort through archival spectroscopic redshift catalogs to identify candidate structures. In brief, overdensities were found by locating groups of galaxies ($N \geq 4$) within tubular volumes of radius 20 cMpc and lengths $\Delta z = 40$ cMpc. This corresponds to the volume the most massive protoclusters in the universe typically exhibit at $z > 2$ (Chiang et al., 2013; Muldrew et al., 2015). Any such group that contained a galaxy overdensity $\delta_{gal} \geq 0.25$ was considered a potential candidate. The galaxy overdensity is a simple ratio of number densities $\delta_{gal} = (n_{proto}/n_{field}) - 1$, where n_{proto} is bound by the volume within the velocity dispersion ($z_{avg} \pm \sigma$) of galaxies in redshift space. This is done partially to reduce the number of interlopers in the overdensity calculation, which are expected to increase as a function of radius from the geometric center of a protocluster (Muldrew et al., 2015; Contini et al., 2016). The field density is found in the same surface area, and along the line of sight within $\Delta z \pm 0.15$ of the central redshift of the overdensity. Further details can be found in Franck & McGaugh (2016b) and Franck & McGaugh (2016a).

We applied the same methods used in the CCPC to nine galaxy lightcones from Henriques et al. (2015). The simulated data are analyzed without amendment: no noise fluctuations, incompleteness, or other selection effects are initially applied to the mock data, apart from the $R_c < 25.5$ limiting magnitude requirement. Attempting to faithfully repro-

duce the myriad differences that comprise the spectroscopic sample selections used in the CCPC, of which there are hundreds, we run the CCPC algorithms on a perfect data set. Are protocluster and field sources separable in a 'pure' dataset? What is the height of effectiveness of our methodology without uncertainty to degrade the results? Key results, like the Schechter parameter estimates, have bootstrapped uncertainties that highlight sample variation. Even more fundamentally, implementing uncertain adjustments (like completeness effects) would only enhance the divide between what is observed and what the simulations predict, as we have shown in the previous subsection. In Section 5.4, we randomly selected 50% of the lightcone data to examine the role spectroscopic incompleteness might play in the conclusions presented here. In general, this did not change any of the underlying results.

These lightcones represent a much larger volume than that which is probed in the CCPC, and also has perfect spectroscopic redshift 'completeness', in contrast to real data. Thus, some metrics, particularly in regards to richness, may not be directly comparable. If the underlying galaxy population between the mock data and observed galaxies is inherently different, the implications of the galaxy overdensity will also change (Chiang et al., 2013). The bias parameter b relates the galaxy density to the underlying mass density of the volume via $\delta_m = \delta_{gal}/b$. For example, the bias parameter for varying populations of galaxies in the MR has been analyzed by Chiang et al. (2013) for the Guo et al. (2011) SAM at high redshift. For galaxies of $M_\star > 10^9\ M_\odot$ at $z = 2$, a value of $b = 1.7$ was found, while larger galaxies of $M_\star > 10^{10}\ M_\odot$ were found to be more biased, with $b = 2.0$. These biases grow larger at higher redshift, and more disparate, in the simulation. If the underlying bias parameter of the CCPC data is more or less biased than the underlying mass density (δ_m) of the simulations, the comparisons between the two will become less apt.

We find that the number of simulated overdensities per light cone is fairly static, with roughly 300 candidate structures found between $2 < z < 7$. They have a median richness

of 31 galaxies each, but with considerably fewer at high redshift where the number density of galaxies decreases significantly (Fig 5.1). There is not an easy comparison between the richness in the CCPC and the MR overdensities. As discussed earlier, the number density of galaxies in the MR does not neatly coincide with the numbers drawn from the spectroscopic samples used in the CCPC, even for surveys like the GOODS-S field.

The estimated median overdensity for these simulated systems is $\delta_{gal} = 6.43$. This is more than a factor of 2 larger than the CCPC (median of $\delta_{gal} \sim 2.9$: Franck & McGaugh, 2016a). The overdensity value, when limited to $2 < z < 4$, is slightly more modest ($\delta_{gal} \geq 4$). At larger redshifts, the scarcity of simulated galaxies enhances the density in regions that clump together. The implications of this will be discussed in Section 5.4. Briefly, one would naively expect that although the richness or number density of the galaxies in the simulation might change, this would scale both field and overdense systems alike. However, it appears that greater galaxy number densities exist in the overdensities within the simulation than when compared to the observed data.

The mean velocity dispersion of the overdensities in our analysis of the Millennium was $< \sigma >= 544.4$ km s^{-1}. While the mean value of the CCPC is larger on average ($< \sigma >= 650$ km s^{-1}), the distribution found between Millennium expectations and the observations share significant overlap. We ran a KS test on the cumulative distributions of the velocity dispersions, which provide a p-value of 0.48 (KS=0.31). There had been some tension, in Franck & McGaugh (2016a) and elsewhere (Venemans, 2005; Venemans et al., 2007), between the observed and expected velocity dispersions of protocluster galaxies, in which the real data found large spreads, occasionally in excess of 1000 km s^{-1}. The MR overdensities exhibit similarly large dispersions as well. As pointed out in Franck & McGaugh (2016a), this could be a purely observational effect of low N observations combined with observational window size.

Table 5.2. 4.5μm Luminosity Functions of mock protoclusters identified by the CCPC algorithm

Redshift Range	Overdense m^* (AB)	Overdense ϕ^* (dN/dm/dA)	Field m^* (AB)	Field ϕ^* (dN/dm/dA)	CCPC m^* (AB)	CCPC ϕ^* (dN/dm/dA)	Δm^* Over-Field	Δm^* Over-CCPC
$2.0\leq z<2.25$	$21.03^{+0.09}_{-0.08}$	0.0075	$21.24^{+0.04}_{-0.04}$	0.0427	$19.73^{+0.56}_{-0.59}$	0.000265	-0.2±0.1	1.3±0.6
$2.25\leq z<2.5$	$21.59^{+0.11}_{-0.11}$	0.0064	$21.65^{+0.05}_{-0.05}$	0.0333	$20.27^{+0.56}_{-0.53}$	0.000335	0.0±0.1	1.3±0.5
$2.5\leq z<2.75$	$21.77^{+0.11}_{-0.11}$	0.0062	$21.93^{+0.05}_{-0.05}$	0.0320	$21.25^{+0.58}_{-0.97}$	0.000314	-0.1±0.1	0.5±1.0
$2.75\leq z<3.0$	$21.98^{+0.14}_{-0.14}$	0.0050	$22.39^{+0.07}_{-0.06}$	0.0273	$21.36^{+0.39}_{-0.31}$	0.000513	-0.4±0.2	0.6±0.3
$3.0\leq z<3.25$	$22.67^{+0.16}_{-0.15}$	0.0059	$22.70^{+0.08}_{-0.08}$	0.0265	$21.08^{+0.41}_{-0.46}$	0.0003	0.0±0.2	1.6±0.5
$3.25\leq z<3.5$	$22.96^{+0.25}_{-0.23}$	0.0064	$23.08^{+0.12}_{-0.12}$	0.0235	$21.31^{+0.71}_{-0.8}$	0.000276	-0.1±0.3	1.7±0.8
$3.5\leq z<4.0$	$23.13^{+0.31}_{-0.30}$	0.0040	$23.38^{+0.13}_{-0.13}$	0.0307	$20.38^{+0.51}_{-0.74}$	0.000121	-0.2±0.3	2.8±0.8
$4.0\leq z<5.0$	$23.82^{+0.55}_{-0.54}$	0.0074	$24.09^{+0.25}_{-0.24}$	0.0371	$20.71^{+0.35}_{-0.74}$	0.000168	-0.2±0.6	3.1±0.9
$5.0\leq z<7.0$	...	...	...	...	$21.75^{+1.97}_{-1.38}$	0.000569	...	...

Note. — The characteristic magnitude m^* and surface density ϕ^* (units: number per magnitude bin per square arcminute) of simulated galaxies in *Spitzer* [4.5] (AB) magnitudes at a range of redshifts. The 2nd and 3rd columns are for galaxies identified within overdensities in nine Henriques et al. (2015) light cones using the same procedure as in the CCPC (Franck & McGaugh, 2016a). The next two columns are all remaining galaxies in the lightcones, which represent field sources. The 6th and 7th columns are the observed CCPC values (Franck & McGaugh, 2017), which were found to be indistinguishable from their field counterparts. The final two columns show the m^* value difference between the simulated overdensity vs. field galaxies and the simulated overdensity vs. real data, respectively. All uncertainties shown are 2σ values found by bootstrapping the fit 10^4 times. The last redshift bin in the lightcones had too few galaxies to perform a reliable fit to these data to determine m^*. The simulated data shows only a few instances in which the overdense systems can be reliably distinguished from their field counterparts by their brighter m^* values ($2.0\leq z<2.25$ and $2.75\leq z<3.0$). The difference between the simulated characteristic magnitude and the real data is apparent at nearly all redshift bins. The discrepancy can be more than 3 magnitudes different for galaxies $4.0\leq z<5.0$. The observations are universally brighter.

5.3.2 Simulated Luminosity Functions

Schechter (1976) supplied a function that relates the number distribution of objects per magnitude per unit area. This distribution function has been widely used to characterize the luminosity in populations of differing density, ages, and redshifts. Throughout this work, we will use the magnitude functional form of the Schechter equation, where $m^*(z)$ is the characteristic magnitude at redshift z. In this way, we can trace the evolution of the underlying stellar population of galaxies as a function of redshift in the *Spitzer* wavelengths. This

Table 5.3. 4.5μm Luminosity Functions of Protocluster Galaxies that join mock clusters by $z = 0$ (*treeRootId*)

Redshift Range	*treeRootId* m^* (AB)	*treeRootId* ϕ^* (dN/dm/dA)	Field m^* (AB)	Field ϕ^* (dN/dm/dA)	CCPC m^* (AB)	CCPC ϕ^* (dN/dm/dA)	Δm^* Proto-Field	Δm^* Proto-CCPC
$2.0 \leq z < 2.25$	$20.69^{+0.08}_{-0.08}$	0.0061	$21.32^{+0.04}_{-0.04}$	0.0441	$19.73^{+0.56}_{-0.59}$	0.000265	-0.6±0.1	1.0±0.6
$2.25 \leq z < 2.5$	$21.15^{+0.11}_{-0.11}$	0.0046	$21.72^{+0.05}_{-0.05}$	0.0372	$20.27^{+0.56}_{-0.53}$	0.000335	-0.5±0.1	0.9±0.5
$2.5 \leq z < 2.75$	$21.51^{+0.11}_{-0.10}$	0.0049	$21.99^{+0.05}_{-0.05}$	0.0334	$21.25^{+0.58}_{-0.97}$	0.000314	-0.4±0.1	0.3±1.0
$2.75 \leq z < 3.0$	$21.87^{+0.13}_{-0.13}$	0.0040	$22.42^{+0.07}_{-0.07}$	0.0283	$21.36^{+0.39}_{-0.31}$	0.000513	-0.5±0.1	0.5±0.3
$3.0 \leq z < 3.25$	$22.31^{+0.16}_{-0.15}$	0.0047	$22.80^{+0.09}_{-0.08}$	0.0281	$21.08^{+0.41}_{-0.46}$	0.0003	-0.4±0.2	1.2±0.5
$3.25 \leq z < 3.5$	$22.83^{+0.27}_{-0.24}$	0.0049	$23.11^{+0.12}_{-0.12}$	0.0195	$21.31^{+0.71}_{-0.8}$	0.000276	-0.2±0.3	1.5±0.8
$3.5 \leq z < 4.0$	$22.95^{+0.23}_{-0.22}$	0.0055	$23.49^{+0.14}_{-0.14}$	0.0312	$20.38^{+0.51}_{-0.74}$	0.000121	-0.5±0.3	2.6±0.8
$4.0 \leq z < 5.0$	$24.09^{+0.36}_{-0.37}$	0.1392	$24.01^{+0.28}_{-0.29}$	0.0326	$20.71^{+0.35}_{-0.74}$	0.000168	0.1±0.5	3.4±0.8
$5.0 \leq z < 7.0$	...	...	...	...	$21.75^{+1.97}_{-1.38}$	0.000569	...	...

Note. — The equivalent of Table 5.2, but for galaxies identified as protocluster galaxies via their *treeRootId* in the simulation. The field galaxies are all remaining galaxies not identified as progenitors of massive ($M_\star \geq 10^{10}\ h^{-1}\ M_\odot$) cluster galaxies at $z = 0$. The final two columns show the magnitude difference between the characteristic magnitude of simulated protocluster - simulated field galaxies and simulated protocluster - real galaxies in the CCPC (Franck & McGaugh, 2017), respectively. Within the bootstrapped uncertainties of the fits, the brightness difference between protocluster and field galaxies is clearly distinguishable at nearly every redshift epoch (apart from $3.25 \leq z < 3.5$). This is in contrast to the simulated galaxies selected by overdensities with the CCPC algorithm from the previous Table, which have much smaller brightness variations between overdense and field galaxies.

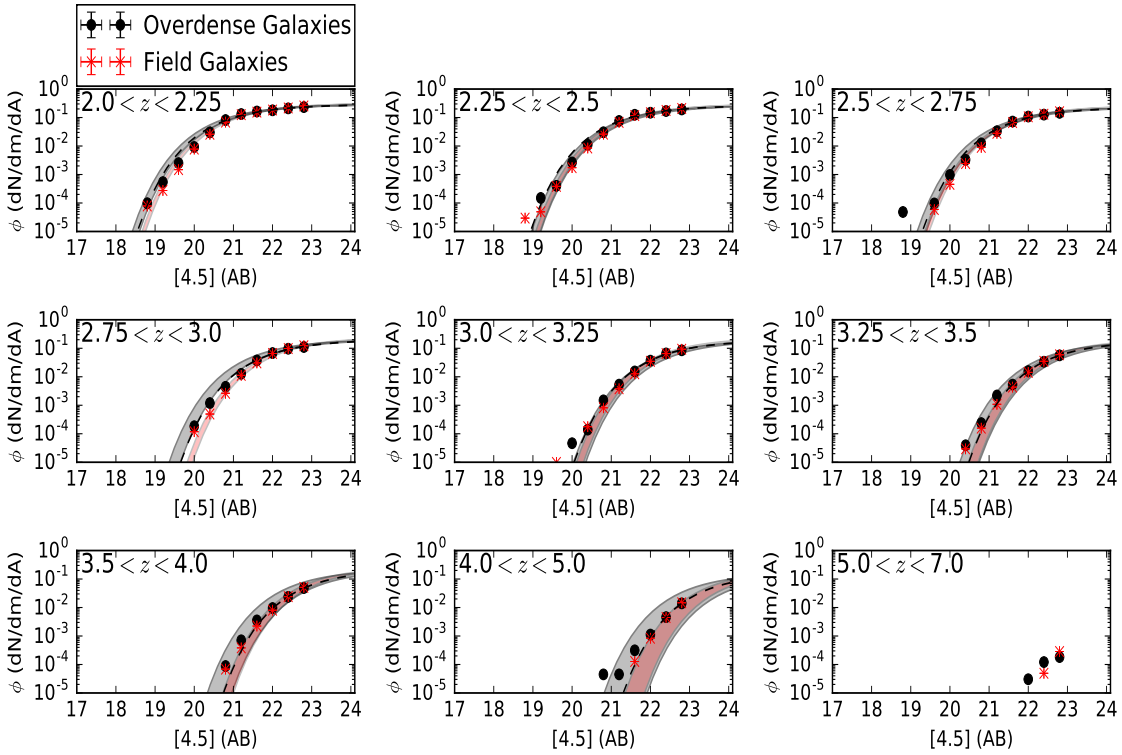

Figure 5.2: The 4.5μm *Spitzer* Luminosity Functions of Millennium Run mock galaxies identified within overdensities found by the CCPC algorithm (black points) alongside the galaxies in the light-cones that were not part of an overdensity (red asterisks). The vertical axis is the number of galaxies (dN) per magnitude (dm) bin per square arcminute (dA). These LFs contain galaxies in 9 mock lightcones provided by Henriques et al. (2015). The overdense galaxies are scaled in density to the field galaxies for easy comparison. The overdense galaxy distributions were fit using a Schechter Function as shown by the dashed line. The grey shaded regions illustrate the 95% confidence interval determined by bootstrapping the fit 10^4 times. The red shaded region is the confidence interval for the field. The value of α was defined to be -1, just as in Franck & McGaugh (2017). The values of m^* and ϕ^* for each epoch can be found in Table 5.2. The last redshift range displayed ($5.0 \leq z < 7.0$) has too few galaxies to be adequately fit. By eye, only the redshift ranges of $2.0 \leq z < 2.25$ and $2.75 \leq z < 3.0$ have confidence intervals that can be reliably distinguished between overdense and field populations. A few other epochs ($2.5 \leq z < 2.75$, $4 \leq z < 5$) have a slight excess population of overdense galaxies in the brightest bin, which offer a hint of discrepancy with respect to the field galaxies.

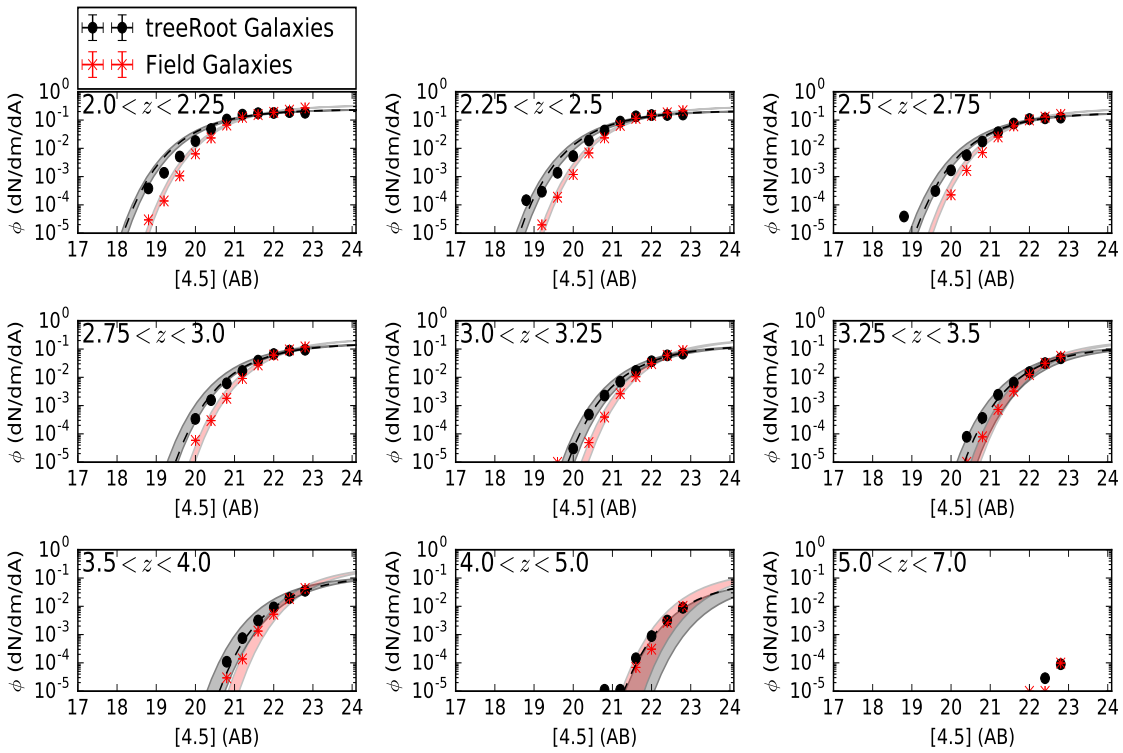

Figure 5.3: Sibling plot to Figure 5.2. In this instance, the black points are protocluster galaxies as identified within the Millennium simulation via their *treeRootIds* and not from the CCPC overdensity finder. Table 5.3 contains the Schechter fits to these simulated data sets. The separation between field and protocluster galaxies is apparent at nearly every temporal slice, in stark contrast to Figure 5.2. The only two redshift ranges where the 2σ confidence intervals meaningfully overlap in the scaled distributions are at $3.25 \leq z < 3.5$ and $4 \leq z < 5$. We will briefly note here that the difference in m^* is relatively small, and would be difficult to measure in the CCPC data.

method has been used extensively for clusters and protoclusters of galaxies at high redshift (Andreon, 2006; Muzzin et al., 2008; Mancone et al., 2010; Andreon, 2013; Wylezalek et al., 2014; Franck & McGaugh, 2017).

To provide as comparable of a result as possible to the real data, the Schechter fitting routine is identical to that of Franck & McGaugh (2017): the magnitude bin sizes (0.4 mag), minimum magnitude ([4.5] $\leq$ 23.2 AB mags), faint-end slope defined as $\alpha = -1$, and the use of the Levenberg-Marquardt algorithm in SciPy's *curvefit* package to optimize the values of $m^*(z), \phi^*(z)$. The simulated galaxies from all 9 mock lightcones are included in each fit, although the number of galaxies in a single cone ($N \sim 10^5$) is orders of magnitude larger than the number of CCPC galaxies ($N \sim 2000$). Individual mock data sets show little variation from the ensemble. To generate the 2σ uncertainty values in the parameters, we bootstrap each fit 10^4 times.

In this work, we have identified three sub-samples of the mock data, and fit the Schechter function to each. The first group is composed of galaxies identified within the overdensities found by the CCPC candidate finder algorithm (Franck & McGaugh, 2016b,a). Hereafter we will refer to these as the 'Overdense Galaxies', or variations thereof. Using the Millennium Simulations *treeRootIds* in order to correctly identify galaxies that *will* become part of a cluster system at $z = 0$, we compile another sample. These will typically be referred to as *treeRootId* galaxies. The 'Field Galaxies' will be all galaxies in the lightcones not associated with an overdense sample. For example, the 'Field Galaxies' in the mock data sets will be all galaxies not identified by a cluster *treeRootId* when the two are compared. A summary table of the samples can be found in Table 5.1.

The results of the fits are included in Table 5.2 ('Overdense Galaxies') and Table 5.3 ('*treeRootId* Galaxies'). The tables show the m^* and ϕ^* evolution of the sub-samples for redshifts $2 < z < 7$. Included are the m^* values calculated in Franck & McGaugh (2017)

for the CCPC. The magnitude shift between the candidate protocluster galaxies and the field samples are shown under Δm^*. 'Overdense Galaxies' (Table 5.2), when compared to the field, show little variation within the confines of the m^* uncertainties. In other words, galaxies found via the overdensity algorithm are indistinguishable from field galaxies, just as Franck & McGaugh (2017) found. The *treeRootId* galaxies have characteristic magnitudes at many redshifts that *are* statistically distinguishable from their field counterparts in the simulation. However, the relative differences between these simulated protocluster and field galaxies is ~ -0.4 mags, which is smaller than the uncertainties in the m^* values determined observationally in Franck & McGaugh (2017). Even if one were able to know, *a priori*, which galaxies were in protoclusters and which were field interlopers, the resulting luminosity functions would likely remain indistinguishable with existing data.

The LFs and their respective fits are plotted in Figure 5.2 and Figure 5.3 for 'Overdense' and *treeRootId* galaxies, respectively. The divisions between the field and (candidate) protocluster galaxies appear more significant than in the Tables. Without photometric measurement errors and with a sample size 10^4 times that of the CCPC, the magnitude distributions are understandably neater than in Franck & McGaugh (2017). The galaxies selected by the CCPC algorithm only rarely ($2.0 \leq z < 2.25$, $2.75 \leq z < 3.0$) show a discernible difference with respect to the field galaxies. In a few instances, the brightest magnitude bin can show a marginally higher surface density for the 'Overdense' objects. Notably, this is not true for the *treeRootId* comparison. Only two epochs ($3.25 \leq z < 3.5$, $4 \leq z < 5$) show values of m^* between the field and protocluster galaxies to be statistically *similar*.

Perhaps the most striking result from these sub-samples is the level of discrepancy between the lightcones and the observations in the CCPC. The last column in each table is the difference from the Overdense/*treeRootId* characteristic magnitude and that of the CCPC at the same redshifts. The mock data are uniformly fainter than the observations.

At $2.5 \leq z < 2.75$ and $2.75 \leq z < 3.0$, the change is ≈ 0.5 magnitudes, similar to the 2σ uncertainty in the CCPC values. At all other redshift ranges, the discrepancy is clear. The differences become as large as 3 magnitudes. The implications of this result will be examined in Section 5.4, and may be related to the "impossibly early" galaxy problem (Steinhardt et al., 2016).

Simulation $m^*(z)$ values get fainter as the redshift increases. This evolution is present in the Field, 'Overdense', and *treeRootId* sub-samples, with m^* fainter by $2.6 < \Delta m^* < 3.4$ mag between $2 < z < 7$. The CCPC m^* value is consistent, within its uncertainty, with no evolution (Franck & McGaugh, 2017).

5.4 Discussion

Here we discuss possible reasons for the differences between the observed data and the expectations of the Millennium Run. In many cases, no satisfactory explanation can be found. The basic discrepancies include: (1) the simulations lack bright galaxies in nearly all redshift bins when compared to observations, (2) the Schechter functions of simulated data have a fainter characteristic magnitude (m^*) the observations can account for at each epoch, (3) the MR m^* expectation at 4.5μm dims by several magnitudes from $z = 2 \rightarrow 6$, while the real data remain roughly constant.

5.4.1 Luminosity Function Comparisons

It was apparent in Section 5.3 that the characteristic magnitudes m^* of 'Overdense' Galaxies (in Table 5.2 and Figure 5.2), which are simulated galaxies identified as part of overdense systems by the CCPC algorithm, were rarely statistically different than the 'Field' galaxies. This is notably without any photometric or shot noise that would be present in

real survey data. *treeRootId* protocluster galaxies within the simulation had a m^* separation between the field that was more apparent than the Overdense systems (Table 5.3 and Figure 5.3). However, the difference was of the same order as the uncertainty in the measured value of the CCPC observations (Franck & McGaugh, 2017). Even if a galaxy were flagged as a protocluster member in a real data set, the simulations suggest that it would be difficult to measure a difference in m^*. As more complete spectroscopic surveys grow in number, and the precision of the photometry becomes better, the lightcone data hints that it is *possible* to measure such a discrepancy. However, separating real protocluster galaxies from field sources remains a difficult problem, as they have few distinguishing properties (Contini et al., 2016).

If the simulation is a reasonable match to our observable universe, galaxies identified by the CCPC algorithm in the simulations would be expected to show no difference in magnitude with respect to the field galaxies. Observationally, this is precisely what was found in Franck & McGaugh (2017). There is additional evidence, at both low (Ellison et al., 2009) and at high redshift (Hatch et al., 2017), that group scale interactions are more important in determining the environmental evolution of galaxies than the cluster scale. This is particularly true at high redshift, where the protocluster's dynamical time is greater than the age of the universe.

The previous paragraph rests on an assumption of equivalence between the simulation and what is observed. There appears to be an irreconcilable difference between the $m^*(z)$ in the CCPC and the mock data sets. Table 5.2 and Table 5.3 both show a multi-magnitude difference in the individual measurements of m^* at various epochs. The characteristic magnitude of the CCPC remains relatively flat at $z > 2$, while the MR lightcones show a steady decrease in brightness as the redshift increases. In effect, bright galaxies are in place earlier than anticipated by the MR. We will examine a variety of possible solutions to this model

discrepancy in the following subsections.

It may be possible that the m^* values estimated in Franck & McGaugh (2017) are a product of spectroscopic and photometric incompleteness, or careful curation of the data set. As the depth of observations also vary between fields, perhaps adjusting the limiting [4.5] magnitudes might produce different results for the characteristic magnitude. To test this possibility, we fit the Schechter function to data of varying limiting magnitude limits ($22 < [4.5] < 24$ with bin size 0.4 mags). We compared the mean of these values with the nominal CCPC value (faintest bin of $[4.5] < 23.2$, located in Table 5.2). On average, the Δm^* value changes by 0.2 magnitudes in this experiment, where the limiting magnitude was altered by a full 2 magnitudes. At most, the value of the observed m^* changes by 0.7 magnitudes at the highest redshift bin ($5 < z < 7$), where the Millennium is not able to even make a prediction. This does not alter our conclusions.

The depth of the CCPC observations are not responsible for the discrepancy between observations and simulations. The discrepancy is primarily due to the absence of galaxies at high redshift in Millennium that are as bright as those observed. The existence of bright galaxies in the observed universe is not contingent on the well-founded concerns about sample depth, nor does the fitted value of m^* depend sensitively on it.

5.4.2 Stellar Models

The magnitude of simulated galaxies were built on the SAM of Henriques et al. (2015). Could the mystery between the brighter observed m^* values, which are nearly constant at redshifts $z \geq 2$, and simulations be rectified when a different model is adopted? We tested two additional models, Somerville et al. (2008) and Croton et al. (2016). In short, although the details of the SAMs vary, and some redshift bins are slightly closer to the CCPC m^* values than others, neither of the SAMs is a clear match.

For instance, the Somerville et al. (2008) SAM $m^*_{[4.5]}$ differs from the CCPC by a mere 0.2 mags at $2.5 \leq z < 2.75$, but by 3.6 magnitudes at redshifts greater than 4. Croton et al. (2016) provide a slightly better fit at high redshifts ($\Delta m^* < 3$ mag), but is worse at earlier epochs. The average m^* difference between the SAM and the CCPC is $\Delta m^* = 1.61, 1.97, 1.31$ for Henriques et al. (2015), Croton et al. (2016), and Somerville et al. (2008), respectively. The choice of SAM appears to have little systematic difference when compared to the observed values in the CCPC.

Henriques et al. (2012) examine the predictions of the Guo et al. (2011) SAM with respect to observed data sets. They found that the lightcones are not well matched to the high redshift observations when using the stellar population models of Bruzual & Charlot (2003). The predictions were more consistent with observations when using the models from Maraston (2005). We use the lightcones in Henriques et al. (2015) imprinted with this latter stellar model, which addresses thermally-pulsating asymptotic giant branch stars (TP-AGBs) in a way that Bruzual & Charlot (2003) does not. The net effect is an increased emission of rest-frame NIR light in stellar populations of order a Gyr in age. The Universe is only a few Gyrs old at $z \approx 2$. Locally, the Maraston (2005) models over-predict NIR light by upwards of 50% (Fig 3; Kriek et al., 2010), potentially exacerbating the problem presented in this work. This NIR light is redshifted into the *Spitzer* bandpasses at the redshifts examined in Franck & McGaugh (2017) and in this work. Presumably the models should be capturing this enhanced component, yet we still see a discrepancy in the sense that the models are still too faint. *The discrepancy between the model and observation remains, despite the inclusion of these updated stellar population models.*

Table 5.4. 4.5μm Bright Galaxy Number Density

Redshift Range	CCPC m^* (AB)	$n_{GOODS-S}$ $(deg)^{-2}$	n_{AGES} $(deg)^{-2}$	n_{MR} $(deg)^{-2}$
$2 \leq z < 2.25$	19.73	84	11	3
$3 \leq z < 3.25$	21.08	89	5	2
$4 \leq z < 5$	20.71	119	1	0

Note. — The number density of galaxies brighter than the $m^*(z)$ in [4.5] AB magnitudes (2^{nd} column) from Franck & McGaugh (2017) in three redshift bins. The number density in the 3^{rd} column (n_{CCPC}) is data from the CANDELS GOODS-s field, while the 4^{th} column is an AGN sample (Kochanek et al., 2012). The final column is the number density of galaxies in the mock lightcones from Henriques et al. (2015). The number density of bright objects in the CCPC data is nearly always an order of magnitude larger than the AGN sample. The AGN sample is a factor of 3 larger than the simulated data set. This simple comparison illustrates that the lower number density of very bright galaxies in the Millennium Run versus the CCPC cannot wholly be attributed to a plethora of AGN contributing non-stellar emission. The R_c selection effect is unimportant. The lightcones in their entirety do not significantly increase the number density.

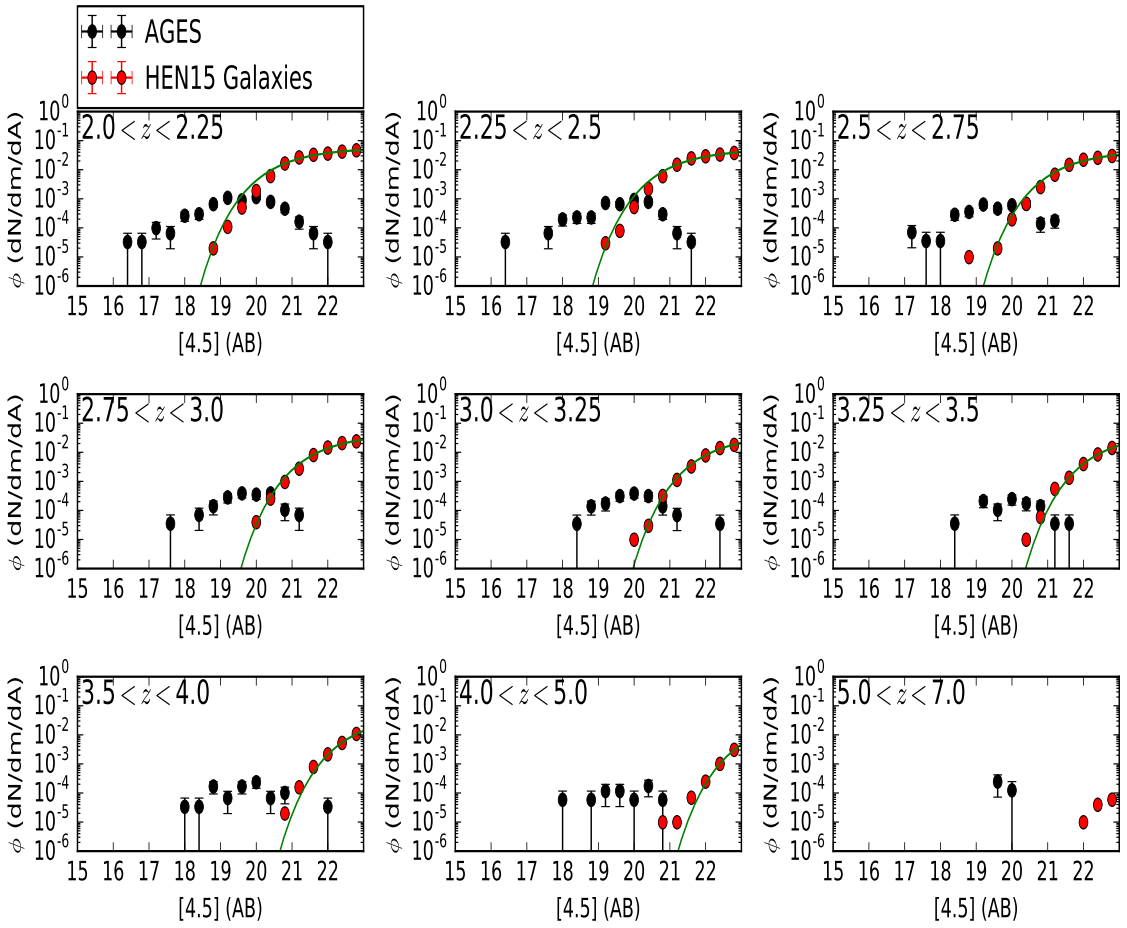

Figure 5.4: Similar to Figure 5.2, with the overdense galaxies, as identified by the CCPC algorithm on the simulated lightcones (Henriques et al., 2015), shown as red points. The black points are from the AGES survey of AGN with $z > 2$ (Kochanek et al., 2012). The vertical axes are the number of galaxies per magnitude bin per square arcminute ($dN/dm/dA$). The black points look fairly similar to the hyperluminous sources discussed in Franck & McGaugh (2017) and in Section 5.4. This suggests that a number of the very bright CCPC galaxies are AGN with significant amounts of non-stellar emission. To examine the possibility that the AGN contamination of the observed data significantly influenced the Schechter fit value of m^*, we combined the lightcone data with the AGES survey and solved for the characteristic luminosity. The fit is shown as a green line. The difference between the m^* values in Table 5.2 and this composite AGN+Simulation population is, on average, a mere 0.25 magnitudes brighter. Compared to the discrepancy between the CCPC and lightcone m^* parameters, which are $0.5 < \Delta m^* < 3$, the AGN are not likely to be the primary contributing variable to the difference.

5.4.3 AGN Emission

It is conceivable that non-stellar emission from Active Galactic Nuclei (AGN) might have some influence on the characteristic magnitude $m^*(z)$. The SAM of Henriques et al. (2015) include prescriptions for AGN feedback, but these objects do not contribute to the NIR/optical emission of the simulated host galaxy. There is some evidence that AGN are enhanced in protoclusters (Casey et al., 2015), but this is not a universal finding (Mazzucchelli et al., 2017; Lemaux et al., 2017; Uchiyama et al., 2017). To test this possibility, we used all spectroscopically confirmed AGN from the The AGN and Galaxy Evolution Survey (hereafter AGES; Kochanek et al., 2012) in the same redshift ranges as the CCPC, over 7.7 square degrees. As an initial test, we computed the number density of galaxies per square degree brighter than the $m^*(z)$ value found in Franck & McGaugh (2017). We used only galaxies found within the CANDELS GOODS-S field ($\sim$[0.45 deg]^2) for the CCPC comparison, as it represents the deepest and most complete field in our sample.

We are focusing on galaxies brighter than $m^*(z)$ in the infrared, which should mitigate completeness issues. In Table 5.4, we tabulate the number density of bright galaxies in three different redshift bins, comparing AGN, CCPC galaxies, and the predictions of the Henriques et al. (2015) model galaxies. At a minimum, the number density of bright CCPC galaxies is $\sim 8\times$ that of AGN. In turn, there are nearly 30 times the number density of bright CCPC galaxies compared to mock galaxies in the lightcones. While AGN may account for a handful of CCPC sources (factor of four), the discrepancy between the observations and the simulated galaxies cannot be fully accounted for in this manner.

As a second test of possible AGN influence, we measured the characteristic $m^*(z)$ of a population of AGN imprinted upon a smooth stellar mass function of mock data. The Luminosity Functions (dN/dm/dA) of the lightcone galaxies (Henriques et al., 2015) and

the AGES data (Kochanek et al., 2012) were combined into a composite sample. For each redshift bin, we fit the Schechter function to these data in the same manner as before. On average, the AGN-infused data set have brighter $m^*(z)$ of a mere 0.25 magnitudes. The AGN do not significantly shift the fit to account for discrepancy between the CCPC data and the Millennium Run expectations. This is shown in Figure 5.4. The Schechter fitting function mostly ignores these bright AGN in determining m^*, which was also the case in Franck & McGaugh (2017).

As a final attempt to identify the effect of AGN contamination, we subtracted the AGES LFs from the CANDELS GOODS-S LFs used in the CCPC. However, within the sub-sample's large 2σ uncertainty, the two values are in agreement (Franck & McGaugh, 2017). After subtracting the AGN distribution ($dN/dm/dA$) from the GOODS-S data, we fit the Schechter function to the resulting LF. On average, the difference between the CANDELS GOODS-S m^* data and the AGN-free data was $\Delta 0.01$ mags. The difference between the GOODS-S (sans-AGN) characteristic magnitudes and the mock data are approximately 1.5 magnitudes at redshifts greater than $z = 3.25$. *AGN are not meaningfully widening the gap of the characteristic magnitude between the simulations and observations.*

Removing all selection effects from the lightcones ($R_c < 25.5$ was our original magnitude limit) does nothing to bridge the considerable divide between the number density of observed and simulated galaxies. The simulation is simply not producing bright objects. The galaxies in the observed sample that would suffer the least from incompleteness or sample selection are not replicated by the simulation.

The [3.6] and [4.5] LFs presented in Franck & McGaugh (2017) contained a few objects (∼1-3 per bin) at every redshift range that were much brighter than the Schechter function fits predict. These were referred to as 'hyper-luminous' sources, as they could be 4 magnitudes brighter than $m^*(z)$. Some of these were found to be AGN, some were attributed to be

possible low-redshift interloping galaxies, while others could be unidentified AGN in the redshift catalogs we drew our sample from. It is also possible that these systems are simply very massive galaxies that are not well fit by the empirical Schechter relation. Although rare, galaxies of this stellar age and mass are currently not predicted by ΛCDM simulations. Steinhardt et al. (2016) point out that some higher-redshift galaxies are too massive to exist in a ΛCDM universe.

The AGES Survey of AGN (Kochanek et al., 2012) shows a similar *Spitzer* magnitudes and number densities to the hyperluminous sources in the CCPC (Figure 5.4). This suggests that many of those sources are likely AGN, flagged or not, in the spectroscopic catalogs that were used to identify high redshift structures in Franck & McGaugh (2016b) and Franck & McGaugh (2016a). Given the complexity of AGN distribution and non-stellar emission, it is not surprising to find a lack of these sources in the Henriques et al. (2015) lightcones. However, it bears repeating that the number density of CCPC galaxies brighter than m^* in the CANDELS GOODS-S field is significantly larger than the AGN sample, which are, in turn, much larger than the simulated galaxy number density from the lightcones (Table 5.4). *There exist more bright galaxies than expected in large scale simulations, and they are not all AGN.*

This work is not isolated in finding discrepancies between the LFs of observations and simulations. Stellar masses from Pérez-González et al. (2008) and Marchesini et al. (2009) were placed on the Millennium SAM stellar mass functions in Guo et al. (2011) as a function of redshift. These showed a similar trend, in that the observed stellar masses of galaxies at high redshifts were systematically larger and not well fit by a Schechter function. The UV luminosity functions of Bowler et al. (2014); Bouwens et al. (2015) also show similar behavior at high redshift, and can be at least partially attributed to gravitational lensing (Ono et al., 2017). Recently, Glazebrook et al. (2017) spectroscopically confirmed an old,

quiescent galaxy at $z = 3.7$ with a stellar mass of $M > 10^{11}\ M_{\odot}$. Such an object does not exist in the SAM from Henriques et al. (2015), (cf. Rong et al., 2017). There are indications that some massive $z > 5$ passive galaxies exist (Mawatari et al., 2016), but these lack spectroscopic confirmation at this time. The star-forming galaxies in the CCPC may potentially be the progenitors of these high stellar mass, passive galaxies. It appears that the excess of bright galaxies we find in the CCPC are consistent with these other studies.

5.4.4 Redshift Evolution of m^*

The characteristic magnitude m^* at 3.6 and 4.5 μm between $2 < z < 6.6$ was found to be approximately constant in Franck & McGaugh (2017). At these redshifts, the rest-frame NIR m^* is consistent with an old ($z_f \geq 7$), simple stellar population (SSP) evolving passively to $z = 0$ (Percival et al., 2009). Lower redshift samples of protoclusters and clusters selected from the literature (Mancone et al., 2010; Wylezalek et al., 2014) were added to the distribution. The trend remained the same: the characteristic magnitude of galaxies appears to be consistent with passive evolution from $0.3 < z < 6.6$. Andreon (2006) and Andreon (2013) measured the [3.6] $m^*(z)$ of clusters from $0.2 \leq z \leq 1.75$, which also transition smoothly with the results presented in Wylezalek et al. (2014) and Franck & McGaugh (2017). High redshift cluster galaxies appear to be consistent with an old stellar population that varies little, or an old population ($z_f = 10$) that has a quick ($\tau = 1$ Gyr) exponential decay. *Most importantly, while CCPC galaxies are generally expected to be star forming galaxies, they do not appear to be rapidly assembling stellar mass at these epochs.*

We stress that this is not a longitudinal study, matching progenitor clusters at high redshift with their lower-redshift counterparts, but it is intriguing that $m^*(z)$ is consistent with passive evolution. Chiang et al. (2017), using the Henriques et al. (2015) SAM, estimate

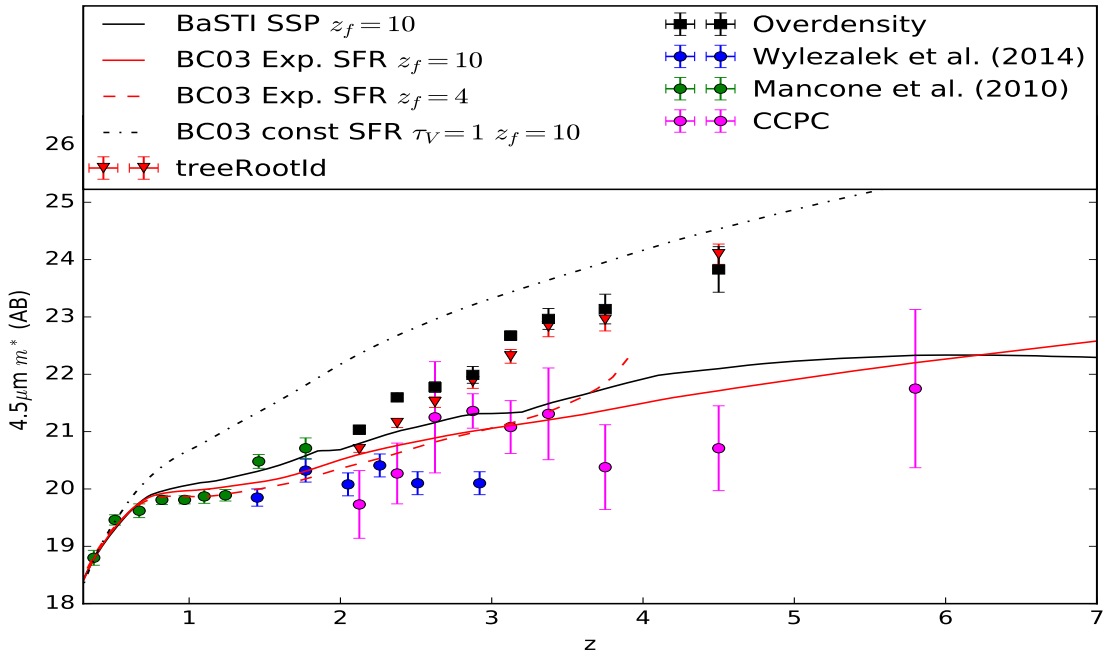

Figure 5.5: The evolution of the characteristic [4.5] magnitude of galaxies and their simulated counterparts as a function of redshift. Red triangles represent simulated protocluster galaxies as identified within the Millennium simulation by their *treeRootId*. Overdense galaxies (black squares) are identified in the Henriques et al. (2015) lightcones using the CCPC algorithm, and then fitted to a Schechter function to estimate $m^*(z)$ values at 4.5μm, as seen in Table 5.2. As a comparison, $m^*(z)$ for the CCPC are plotted as magenta points, along with galaxies in clusters and protoclusters at lower redshifts from the literature (Mancone et al., 2010; Wylezalek et al., 2014). Galaxy stellar population models constructed using EzGal are plotted as comparisons, but we are not assuming that these protoclusters and clusters are a progenitor-descendant matched sample. The mock data has a $m^*(z)$ trend that looks more similar to a constant star formation model (dust extinction of $\tau_V = 1$) shown by a dash-dot red line (Bruzual & Charlot, 2003) up to $z \sim 2$, where it might merge into a passive evolution model born at high redshift ($z_f > 7$). The simulations predict a large stellar mass assembly between $2 < z < 6.6$, while these data seem best fit by a massive, old population of galaxies.

that by $z = 2$, protocluster galaxies have grown half of their final stellar mass during a period of robust star formation prior to being quenched. Yet we see no distinct trend of this in the data, as Fig 5.5 shows. The results of Chiang et al. (2017) also suggest that in the simulations, the field mass assembly occurs Gyrs after the protocluster galaxies, which is also not seen in the field samples of Franck & McGaugh (2017).

The evolution of $m^*(z)$ can be compared to the results in the MR simulation for galaxies found in overdensities via the CCPC algorithm (Table 5.2) or in protocluster galaxies identified via their *treeRootIds* (Table 5.3). Fig 5.5 plots the [4.5] m^* values from the literature and the CCPC. Included are various galaxy population models constructed with EzGal (Mancone & Gonzalez, 2012), including SSPs and constant star formation models (BC03 hereafter; Bruzual & Charlot, 2003). For both samples, the mock data predicts that a rapid evolution of galaxies should exist at redshifts greater than 2, where the $m^*(z)$ value grows brighter by roughly 3 magnitudes in a few Gyrs. This evolution looks similar in shape to the BC03 constant star formation model. The observed galaxies show no such behavior, and are well fit by old galaxy populations with either a single burst of star formation (SSP), or a BC03 exponentially decaying star formation rate (Tinsley, 1972) with $\tau = 1$ Gyr. Table 5.2 and Table 5.3 both show a $\Delta m^*(z) > 3$ between those implied by the simulations and the observed values at the highest redshift bins, which are much larger than their respective errors. The m^* values of Wylezalek et al. (2014) are in greater contention with the Millennium data than the CCPC data at $2 < z \leq 3$.

This result is not easy to explain away. Earlier in this manuscript it was shown that AGN cannot influence the characteristic magnitude enough to account for the large discrepancy in brightness. There is a large, genuine deficit of bright galaxies in the lightcones.

5.4.5 Overdensity Comparisons to the CCPC

In Franck & McGaugh (2016b,a), each candidate protocluster was assigned a probability of becoming a cluster by $z = 0$ based on its galaxy overdensity (δ_{gal}) and redshift. We took these estimates from analysis of the MR by Chiang et al. (2013), in which they plotted the fraction of galaxy overdensities that will collapse into clusters by $z = 0$ for a variety of redshifts and tracer populations (see their Fig 8). Analysis of our candidate protoclusters in the CANDELS GOODS-S suggested that their assigned probabilities were much too conservative, based on the number density of clusters in a similar volume (Franck & McGaugh, 2016a). There should be nearly 50 protoclusters within the volume probed, but the sum of our collapse probabilities suggest that only 3 CCPC systems would become clusters.

As reported in Section 5.3, the median overdensity is $\delta_{gal} = 6.43$ for systems found in the Henriques et al. (2015) lightcone overdensities. The values in Franck & McGaugh (2016b) and Franck & McGaugh (2016a) are $\delta_{gal} \sim 2$ and 3, respectively. In CANDELS GOODS-S field, with the deepest and most complete spectroscopic data set in the CCPC, the median overdensity for CCPC systems is even lower ($\delta_{gal} \sim 1.59$). This difference may be a clue as to why the probabilities of cluster collapse reported for the CCPC candidates was so low when compared to the expected number of true-positive overdensities that we should have identified. Larger overdensities in the Chiang et al. (2013) analysis of the MR are much more likely to form clusters at $z = 0$, and more massive ones at that. It therefore seems plausible that a large component of this discrepancy can be directly related to the difference in the spectroscopic catalogs versus the mock galaxy catalogs. For the entire CCPC catalog, more than 18% of candidate systems have $\delta_{gal} < 1$, while only 0.3% of simulated overdensities can make the same claim. The overdensities in the MR and in the CCPC do not share the same distribution function or mean, with the simulated overdensities

typically exhibiting δ_{gal} of a factor of $\geq 2\times$ larger.

There are two probable reasons for this discrepancy, both of which can play a role. The first is that the methods used by Chiang et al. (2013) for determining δ_{gal} and finding structure are not the same as those in the CCPC, and the volumes/geometries probed are different. Secondly, it is difficult to match the galaxy populations spectroscopically identified with a simulated set of observations in a SAM. The first of these issues can be amended, as we have run the Henriques et al. (2015) lightcones through our own overdensity finder algorithm to mimic the observed CCPC. As mentioned previously, the galaxies themselves appear to be in less dense environments in the real universe compared to the expected overdensities in the MR. *Considering the orders of magnitude difference in bright galaxy number densities found in this work, a factor of 2-3 in overdensity variation is not surprising.*

It appears that the low collapse probabilities assigned to CCPC galaxies may be more of a product of the mapping between the universe and the simulations, rather than a total failure of the algorithm. When used on the simulated data, our methods find much larger overdensities and fewer marginal detections than were found in the actual data. The models are over-predicting the contrast.

Within the simulations, we find little magnitude dependence on δ_{gal}. To test the possibility that the discrepancy in overdensity was related to the source stellar mass, we recomputed the δ_{gal} of each candidate simulated overdensity over a range of minimum [4.5] magnitudes. Recall that the median overdensity of candidate systems in the simulation with no *Spitzer* limiting magnitude was $\delta_{gal} \sim 6$. Limiting the overdensity calculation to only simulated objects of [4.5] < 23.8, 22.8, 21.8 resulted in median δ_{gal}'s of 5.2, 5.9, 7.0, respectively. In other words, calculating the overdensity of a more biased, more massive object resulted in slightly larger δ_{gal}. This is precisely what is expected in simulations (Chiang et al., 2013). *The key takeaway is that the discrepancy between the CCPC δ_{gal} values and the simulations*

δ_{gal} cannot be bridged by altering the limiting magnitude within the simulation.

This leads to an interesting conundrum. Galaxies are biased tracers of the matter density (δ_m) of the universe, and are typically parameterized as $\delta_{gal} = b\delta_m$, with b as the bias parameter. The value of b has been estimated for high redshift sources observationally (Gawiser et al., 2007) and within simulations (Chiang et al., 2013). Greater baryonic mass galaxies (e.g., brighter galaxies) are *more biased* tracers of the underlying dark matter density, so their bias parameter is larger. The CCPC galaxies have been shown to have brighter magnitudes at all redshifts than their simulated counterparts, suggesting they are potentially more biased. In addition, the median galaxy overdensity of the CCPC is a factor of 2 smaller than in the lightcones. Both of these factors suggest that the matter overdensities δ_m found in the Millennium Run are at least a factor of 2 larger, with the difference in the bias parameter playing a small role in widening the gap.

5.4.6 Protocluster Detection Efficiency of the CCPC Algorithm in the Millennium Run

A key question is what percentage of galaxy overdensities identified by the CCPC algorithm contain galaxies that will become part of a cluster at $z = 0$? Additionally, knowing what protoclusters are identified/missed in the volume probed in deep spectroscopic surveys would aid in refining the algorithm. Are the most massive protocluster systems found by their comparatively larger δ_{gal} values, leaving the weaker overdensities buried in the noise? It is not possible to continuously observe a single system evolve over 10 Gyrs, but the simulations might provide a clue as to the nature of the CCPC overdensities.

The methodology is relatively simple: compare the *treeRootIds* of galaxies identified within the overdense regions the CCPC algorithm found, to the *treeRootIds* of galaxies associated with clusters at $z = 0$. We can also establish the number of clusters that have

at least one galaxy which exists within the lightcone we are probing, as well as the mass distribution of protoclusters identified.

The CCPC methodology on the MR data is able to correctly identify $\sim 1/3$ of all protocluster galaxies. The percentage of simulated galaxies in overdensities that were correctly identified as $z = 0$ cluster progenitors is on average $36\% \pm 4\%$ across the 9 lightcones analyzed. The fraction of unique clusters that have one or more galaxies within a lightcone, which are flagged as part of a CCPC overdensity, is 0.33 ± 0.03 of the total number of unique clusters. Of the simulated galaxies that reside in clusters by $z = 0$, the CCPC algorithm identifies 33% as protoclusters at high redshift. Diener et al. (2013) found that a similar fraction (~ 0.35) of the overdensities they identified became clusters in the mock data they tested their algorithm against. While this may initially appear to be a quite low fraction of recovered protoclusters, we should note that the relative overdensities of these systems are only a factor of a few larger than the mean density of the universe at these epochs and are quite spatially diffuse (Chiang et al., 2013). This hints at a selection bias towards the most overdense, massive systems being detected.

There is a only small difference in the median $z = 0$ cluster mass (descendant mass) of the protoclusters we identified as overdensities versus those that were missed. The median mass of successfully identified overdensities was $M_{z=0} \sim 1.7 \times 10^{14}\ h^{-1}\ M_{\odot}$, while the missed protoclusters' belonged to systems of descendant mass $M_{z=0} \sim 1.4 \times 10^{14}\ h^{-1}\ M_{\odot}$. This discrepancy was identified in all lightcones, with little inherent scatter. A two-sided Kolmogorov-Smirnov (KS) test showed no distinguishable difference in the respective mass distributions. In general, the algorithm identified the most massive protoclusters in the lightcones ($M_{z=0} > 10^{15}\ h^{-1}\ M_{\odot}$), of which there are only 8 in the entire simulation. This skews all of the median values slightly to the more massive end, but with no statistical significance. This finding remains true with the stellar mass of protocluster galaxies identified with our

algorithm. They have a median value of $M_\star \sim 3.4 \times 10^9\ h^{-1}\ M_\odot$, while their un-identified counterparts (that will become cluster galaxies) have a median of $M_\star \sim 3.3 \times 10^9\ h^{-1}\ M_\odot$. Their respective distributions are also equivalent from the perspective of a KS test.

There are some systematic and unavoidable issues present in this analysis. In both the real survey data and the simulated lightcones, edge effects can be a challenge. For instance, if a protocluster were to be grazed by a simulation, in that only a single progenitor galaxy happened to be caught in the lightcone, then it would not be identified as a galaxy overdensity. It would then be classified as a missed protocluster, lowering the fraction of recovered systems. The stellar mass limit of galaxies at $z = 0$ from which we identify *treeRootIds* is $M_{\star,z=0} \geq 10^{10}\ h^{-1}\ M_\odot$, with the expectation that many protocluster galaxies will grow to this mass over time. However, some spectroscopically detectable high redshift protocluster galaxies may have masses $M_\star \geq 10^9\ h^{-1}\ M_\odot$ but do not grow above the minimum threshold by $z = 0$. These would also be flagged as galaxies and/or protoclusters missed by our algorithm. A final consideration is that at redshifts $z > 4$, the simulation predicts very few galaxies to exist. The CCPC algorithm requires at least 4 galaxies to exist within the search volume to be considered a candidate. In order to calculate a galaxy overdensity δ_{gal}, nearby field sources have to be present for a comparison to be made. If these requirements are not met, then the algorithm cannot label the volume as a protocluster candidate, even if the Millennium Run has flagged these as protocluster galaxies. This also can lower the efficiency metric of the CCPC search algorithm.

In some cases, the calculated recovery rate can be artificially boosted. More than one cluster can be contained within the search volume in some cases, which would boost the recovery rate for the CCPC algorithm. In the case of the most massive cluster in the Millennium Simulation ($M_{z=0} \sim 2.3 \times 10^{15}\ h^{-1}\ M_\odot$), there are 2 additional clusters within the CCPC search volume of smaller mass ($M_{z=0} \sim 1.5 \times 10^{14}$ and $2.8 \times 10^{14}\ h^{-1}\ M_\odot$) all within

$\sim$ 10 h^{-1} cMpc of each other. This is obviously an anomalous case in which the cluster density (and also mass density) is much greater than the average within the MR. It also suggests that it can occur in the universe, theoretically.

If we restrict the overdensity search radius to 10 cMpc (down from 20 in the original algorithm) and ±10 cMpc along the line of sight, one might assume that the completeness would increase substantially, as fewer field galaxies might be included in the overdense regions. In fact, the percentage of successfully identified galaxies decreases to $27\% \pm 2\%$. We recover a similar percentage of all protocluster galaxies within the lightcones, on average. This slight decrease in performance was balanced by the increase in the number of clusters identified ($45\% \pm 1\%$). As before, the caveats mentioned prior still apply. However, if the goal is to get the largest number of clusters identified, it might be reasonable to decrease the search volume. If a protocluster sample wanted the greatest fraction of galaxies within the sample to end up as cluster galaxies, then perhaps the larger search volume should be employed.

5.4.7 Overlooked Protoclusters

What are the defining characteristics of the missed protocluster galaxies (as identified by their *treeRootId*) when compared to the galaxies identified by the CCPC algorithm? Why were these simulated galaxies passed over as potential centers of protoclusters? To provide some context, we examined each individual mock galaxy that was missed in the lightcones as if it were the center of a candidate protocluster with the CCPC algorithm. There are approximately 12,000 missed protocluster galaxies per lightcone. We then computed the volume's richness, overdensity δ_{gal}, and velocity dispersion.

This approach requires some caveats. Some of these missed galaxies may belong to the same cluster as galaxies that *were* identified, but were outside of the search volume and

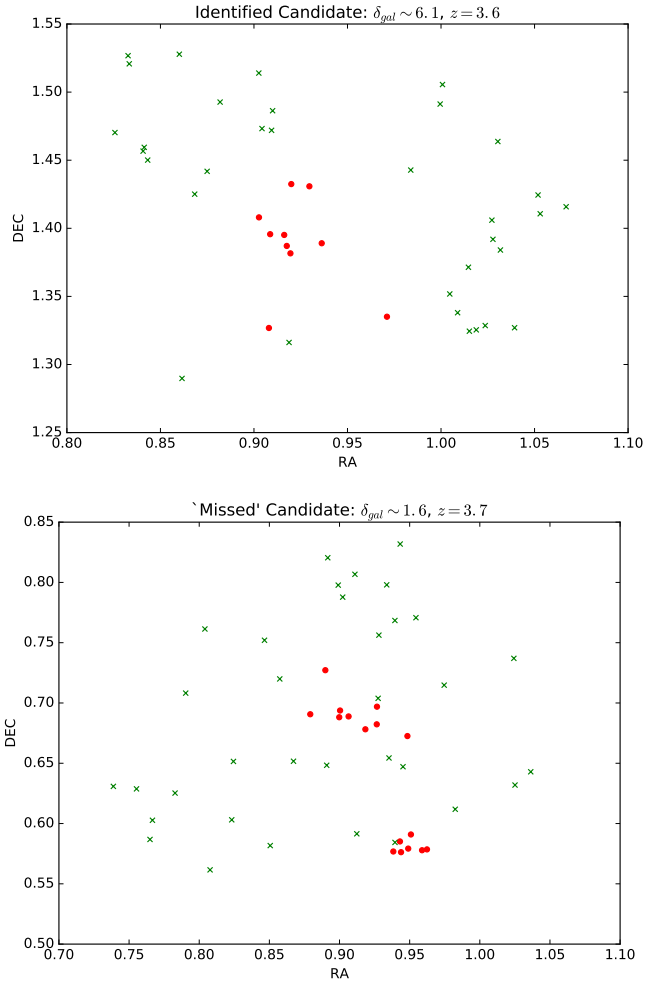

Figure 5.6: *Left:* A simulated protocluster correctly identified by the CCPC algorithm at $z = 3.6$. It has an overdensity typical of a candidate within the lightcones ($\delta_{gal} \sim 6.4$). The volume shown is that of the CCPC search radius ($R = 20$ cMpc) and along the line of sight ($dz = 40$ cMpc). Red points are simulated protocluster galaxies as identified by their *treeRootId*. The green $\times$ symbols are field galaxies not bound to a cluster by $z = 0$. *Right:* Simulated galaxies surrounding a protocluster mock galaxy that was missed by the CCPC algorithm at $z \sim 3.7$. It is found in a much lower overdensity, typical of these missed systems ($\delta_{gal} \sim 1.6$). The only identifiable trait that distinguishes these two systems is their overdensity, which manifests itself along the line of sight.

were therefore 'missed'. Many of these missed galaxies exist in the same volume as one another, and therefore may be associated. This can affect the statistics, as regions with many missed galaxies will lead to N measurements of similar δ_{gal}. For example, a volume with 50 galaxies inside it may have a roughly similar overdensity measurement surrounding it with our method, and will have approximately the same richness, for each iteration of the analysis. Thus the median or mean could be shifted to these high richness regions.

The median overdensity value for these 'missed' protocluster galaxies is $\delta_{gal} \sim 1.6$. Only 17% of these galaxies have overdensities greater than the median CCPC value ($\delta_{gal} \sim 3$), while nearly 4% would fail the minimum overdensity requirement of $\delta_{gal} > 0.25$ necessary for CCPC consideration. Only 2.4% of these missed galaxies have overdensity values greater than the median value of the overdensities found in the lightcones with the CCPC algorithm ($\delta_{gal} \sim 6.4$). The velocity dispersion of the mock galaxies surrounding the missed protocluster systems is $\sigma \sim 719 \pm 119$ km s^{-1}. This is larger than the value around the overdensities found by our algorithm ($\sigma \sim 550$ km s^{-1}), but within the uncertainty.

It appears that the reason the vast majority of these missed, mock protocluster galaxies are overlooked is that they are not located in the peak overdensity regions. As the algorithm explores the volume space, it is looking for the most overdense centers of galaxy systems, and not simply all overdensities. In every other respect, these systems are unremarkable. At least in the lightcones, protocluster galaxies (as identified by their *treeRootId*) have similar stellar mass, magnitudes, and velocity dispersions, whether or not they were identified by our algorithm as potential protocluster galaxies. Figure 5.6 shows two examples of protocluster galaxies at $z \sim 3.6$ as 'viewed' on the lightcone. One shows an example of a positive detection by our algorithm with an overdensity value similar to the median value of all overdensities found in the simulated lightcones ($\delta_{gal} \sim 6.4$). The other example is a 'missed' detection of simulated protocluster galaxies. It has a lower overdensity of $\delta_{gal} \sim 1.6$, the

same as the median of all missed systems. In each plot, the simulated galaxies that will be bound to a cluster by $z = 0$ are shown as red dots, while interlopers are green $\times$s. The two overdensities look quite similar as seen within the lightcones, with the major distinguishing feature being the δ_{gal} value.

This is a double-edged sword. On the positive side, this suggests that the CCPC algorithm is effective in picking out regions of the highest density in a survey, and can legitimately find structure in the lightcones. However, it also means that our methodology is missing nearly 2/3 of all protocluster galaxies, and that these systems do not reside in the densest environments. As a silver lining to this latter statement, the model galaxies not selected do not appear to be that different from the galaxies selected as part of an overdensity.

5.4.8 Spectroscopic Completeness Comparisons

It is not expected that any spectroscopic survey is complete. In our analysis of the Millennium Simulation, we have not made any attempt to match the complex and heterogeneous completeness function of the CCPC surveys to the simulated data sets. What would be the results of the CCPC algorithm on a Henriques et al. (2015) lightcone if its spectroscopic completeness were lowered? Could the large discrepancy in the characteristic magnitude between the CCPC and the simulation be rectified? To test this, we assumed a flat completeness rate of 50% and randomly drew 1/2 of all mock galaxies from the 9 lightcones ($R_c < 25.5$). We ran the entire CCPC pipeline on the incomplete mock data sets, as was done with the full lightcones. Little variation was found between the two results.

The largest difference between the incomplete simulation sample and the full Millennium lightcone is the median overdensity. The full sample has a typical overdensity of $\delta_{gal} \sim 6.4$, while the median overdensity of the incomplete sample is $\delta_{gal} \sim 8.4$. As field galaxies become more rarefied, the relative overdensity must increase. As noted earlier,

candidates at higher redshift can have few field galaxies along the line of sight, thus inflating the overdensities. This is especially true with the incomplete sample.

The Schechter parameters $m^*(z)$ do not significantly change. Reducing the number of mock galaxies through spectroscopic incompleteness cannot increase the number density of bright, simulated objects. There are simply not enough predicted bright galaxies at high redshift in the Millennium Simulation to match the results in Franck & McGaugh (2017). Indeed, the highest redshift bins had too few mock galaxies to effectively fit a Schechter function to. For mock data at $z < 3.25$, the average difference in m^* between the overdensities in the incomplete and full Millennium sample (Table 5.2) was $< \Delta m^* >= 0.0 \pm 0.2$.

5.4.9 Future Protocluster Survey Recommendations

What prescriptions could be given, if any, to a future protocluster survey in the coming age of 'Big Data' Astronomy, based on the findings of this work? We once again emphasize that this is wholly dependent on the models matching the actual universe. This matching deserves further scrutiny prior to any implementation, as we have shown in this work. However, we can provide some guidance based on the assumption that the model *is* representative of the observations, in general terms.

The CCPC method was tailored to identify the most massive protoclusters in the universe, which are expected to have diameters of order 40 cMpc (Chiang et al., 2013; Muldrew et al., 2015). In past protocluster surveys, sometimes by search design and many times because of an instrument's field of view, the survey volume is rather smaller (Venemans et al., 2007; Diener et al., 2013; Wylezalek et al., 2014). As most protoclusters at $z > 2$ are thought to be of lower mass, and thus would be of smaller size, it could be argued that a smaller diameter search radius should be used to capture more candidate protoclusters.

Decreasing the search radius of the algorithm to $R = 10$ cMpc results in a greater

fraction of recovered cluster progenitors (0.45) compared to the 30% recovery rate in the original version. This makes sense on an intuitive level. If the goal is to identify as many protoclusters as possible, it might be reasonable to look for overdensities on a volume scale which is common to low-mass systems. If the survey aim is to identify the largest fraction of systems, there is reason to aim for smaller radii.

However, there are two significant downsides to this approach. Primarily, the results presented here suggest that this change will lead to a lower fraction of simulated protocluster member recovery, from roughly 0.25 with $R = 10$ to 0.35 with $R = 20$ cMpc. If the observer is attempting to gather the greatest number of protocluster *galaxies* for analysis, as was the aim in Franck & McGaugh (2017), reducing the search radius might be counter-productive.

Another significant issue is detection efficiency. Previous work by Chiang et al. (2013) show that MR protoclusters have different galaxy overdensity distributions based on their $z = 0$ cluster mass, with larger mass clusters having statistically larger overdensities throughout their existence. Less massive progenitor clusters with lower values of δ_{gal} are more difficult to differentiate from the noise. Making this a more problematic scenario is that the CCPC mean overdensity value is a factor of 2 smaller than what was found by analyzing the lightcones in the same way. This suggests it might be even more difficult to disentangle structure from the noisy spectroscopic data-sets than the simulations suggest.

5.5 Summary

In this paper, we applied the algorithms used to detect, characterize, and analyze candidate protoclusters in the CCPC to mock lightcones from the Millennium Run simulation combined with the models of Henriques et al. (2015). We identified significant discrepancies between the theoretical expectations and what was observed. Our findings can be

summarized as follows:

- The number density of galaxies brighter than $m^*(z)$ is orders of magnitude larger for the CANDELS GOODS-S field than in the mock lightcones (Table 5.4), consistent with Franck & McGaugh (2017). This is not sensitive to sample selection.

- The $4.5\mu m$ characteristic magnitudes $m^*(z)$ estimated by fitting a Schechter function to observed galaxies are brighter at all epochs than the Millennium Run predictions (Fig 5.1, Tables 5.2, 5.3).

- The [4.5] m^* for the CCPC galaxies is nearly constant with redshift, while the simulations suggest the characteristic magnitude of mock galaxies should be changing rapidly at $z \geq 2$ (Fig 5.5). The observed variation of m^* with redshift is consistent with the passive evolution of galaxies formed at high redshift ($z_f > 7$) but not with the model lightcones above $z > 2$.

- Differences between the populations of Overdense (as found by the CCPC algorithm) and Field mock galaxies is negligible (Figure 5.2, Table 5.2)

- Even when the simulated galaxies are known to be part of protoclusters by their *tree-RootId*, the difference with respect to the field is rarely larger than the photometric errors found in observational studies (Figure 5.3, Table 5.3)

- The overdensities found in MR have a δ_{gal} that is a factor of ≥ 2 larger than what was found in the CCPC data, which does not depend strongly on the limiting magnitude. This may indicate that the contrast between overdense regions and the field are less pronounced in the real universe.

- The CCPC algorithm captured more than 30% of the mock protocluster galaxies, as traced by their *treeRootId*, that are produced in the simulated lightcones. The remaining galaxies belong to median overdensities nearly a factor of 4 smaller, and thus lack the contrast to stand out from the noise.

The current generation of simulations diverge from the observed properties of galaxies at $z \geq 2$. Real galaxies are brighter than anticipated at all epochs examined in this work. This is not a subtle effect, amounting to more than a 3 magnitude discrepancy at $z \geq 4$.

Chapter 6

Summary

The Candidate Cluster and Protocluster Catalog (CCPC) is a data mining project that leverages open astronomical databases to systematically identify and understand high redshift protoclusters and their galaxies. With these open data sets, we could vastly expand the scope of previous research projects. The CCPC also aimed to standardize the search process so that relevant systems could be compared on an equal basis. Only then could we quantitatively measure the properties and evolution of these high redshift overdensities.

The pilot study of the CCPC, published in (Franck & McGaugh, 2016b), started with data mining 14,000 galaxies with spectroscopic galaxies out of NASA's Extraglactic Database (NED) between $2.75 < z < 3.75$. This redshift range corresponds to the onset of typical Lyman Break Galaxy selection (Steidel et al., 1998).

We built an overdensity finding algorithm with a minimum criteria of $N \geq 4$ galaxies within a search radius of $R = 20$ cMpc and length of $\Delta z = 40$ cMpc. For each region that satisfied this richness metric, we computed its overdensity, δ_{gal}, with respect to field galaxies. Any region that had an overdensity $\delta_{gal} \geq 0.25$ was considered a candidate. Based on their overdensity and redshift, we computed the probability that each candidate was a bona fide protocluster based on results in the Millennium Simulation (Chiang et al., 2013).

In this first iteration:

- We identified 43 candidate protoclusters. More than 10 of these are also confirmed in the literature.
- The median overdensity is $\delta_{gal} \sim 2$, with a median richness of 11 galaxies.
- Twelve candidates have more than an 85% probability of being associated with a cluster at $z = 0$.
- This catalog represents the greatest number of spectroscopically confirmed protoclusters in the literature.

The second implementation of the CCPC algorithm was on a data set of expanded redshift range ($2 < z < 7$). Once again, all the data were mined through NED of spectroscopically confirmed galaxies. This work was published in Franck & McGaugh (2016a). At this point that we have a data set with enough redshift coverage to follow the evolutionary development analysis of protoclusters and their galaxies.

- The total number of CCPC candidates is increased to 216 with the addition of 173 new systems in CCPC2. This represents the largest protocluster catalog to date.
- We identify the highest redshift protocluster confirmed spectroscopically at $z = 6.56$. It has 9 galaxy members and an overdensity of $\delta_{gal} \sim 4.2$.
- In the CCPC candidate list, there are three 'super-structures' at $z \sim 2.3, 3.5$ and 6.55. These are groups of CCPC candidates that were found to nearly overlap one another in a continuous distribution of overdensity. These complexes can be nearly 120 cMpc in length if they are associated with one another. They may be primordial superclusters.

There are many unanswered questions in galaxy formation theories. Are protocluster galaxies more massive than their less dense counterparts? Do they form more stars, form earlier, evolve more quickly? Do we see an epoch at which galaxies in a given environment undergo significant growth or quenching? These are all testable hypotheses that are provided by theoretical models (Chiang et al., 2013; Muldrew et al., 2015) and can be measured with current technology.

The studies that attempt to measure the influence of dense environments at high redshift have found no clear picture. For individual protocluster systems, there may be some evidence that galaxies look slightly different when compared to the field population (Hatch et al., 2011; Casey et al., 2015; Dey et al., 2016; Zheng et al., 2016; Hatch et al., 2017). However, *what* that difference is and its significance can differ among studies.

Interestingly, bulk analysis of galaxies at high redshifts paint a different picture. In general, galaxies do not seem to vary much in different density environments (Diener et al., 2013, 2015; Zhao et al., 2017; Ownsworth et al., 2016). These bulk studies offer us a glimpse not only of evolution across densities and times, but also provide a feature check on previous hypotheses. For instance, if a single case study suggests that dense galaxies have greater stellar mass, it could be true for that system alone. A bulk study can test that feature across epochs and determine if it is a broad component of *all* protocluster galaxies. The tendency for results of an observed difference between properties of protocluster and field galaxies to fade away with improved statistics is an important test that the CCPC is ideally suited to corroborate.

The CCPC has considerable leverage in comparison to some previous, multi-epoch data sets that concluded with a null result. For instance, the work of Diener et al. (2013) is bound to a singular survey (COSMOS) and is limited to redshifts $z < 3$. The CCPC has data from many deep fields, and extends to much earlier epochs ($z \leq 6.6$). The redshift range of $2 <$

$z < 6.6$ covers more than 2 Gyrs of rapid galaxy evolution (Madau & Dickinson, 2014) and structure growth (Chiang et al., 2013). This time period forms a crucible from which galaxy model predictions must emerge unscathed when compared to the observations. Chapters 4 and 5, the first of which is published in Franck & McGaugh (2017), begin to explore the growth and evolution of these CCPC sources.

The CCPC contains more than 2000 spectroscopic galaxies within its catalog of protoclusters. In Chapter 4, we identify nearly 3/4 of these galaxies in images found within the *Spitzer Heritage Archive* at 3.6 and 4.5μm wavelengths. At these redshifts, the bandpasses are measuring the rest-frame NIR light from the galaxies, which is then used as a mass-proxy. These data were then binned according to their redshift. Their resulting luminosity functions were then fit with a Schechter function to measure their characteristic magnitude $m^*(z)$. We simultaneously measured all galaxies with spectroscopic redshifts that were not part of a protocluster. These we considered our 'field' sample. We compared the evolution of $m^*(z)$ with a variety of stellar population models to estimate the history of these galaxies.

- The Luminosity Functions of CCPC and field galaxies were statistically similar to one another at all redshifts. This suggests that there is little difference in the stellar mass of galaxies at high redshift based on the local galaxy density.

- The characteristic magnitude m^* evolves little from $z = 2$ to $z = 7$ for both the field and protocluster samples.

- A simple stellar population born at high redshift $z_f > 7$ is consistent with the CCPC. Results from past studies (Mancone et al., 2010; Wylezalek et al., 2014) were also consistent with our results. Constant star formation evolutionary models were disfavored.

The last chapter of the thesis (Chapter 5) used the algorithms described in the first few chapters to identify and characterize the simulated protoclusters of the Millennium Run (Springel et al., 2005) and the lightcones developed for it (Henriques et al., 2015). The lightcones were the main basis of comparison, as they most closely resemble the data as was observed. If the models are fair representations of reality, they may provide clues as to the nature and evolution of what is observed. If they do not, the discrepancies can provide impetus for the modeling community to better match the Universe.

We first identified structure within 9 lightcones of $\sim$ 1 degree radius between redshifts $2 < z < 7$, using the same overdensity finding algorithm as in Franck & McGaugh (2016a). We compared the galaxy overdensities δ_{gal} with those found in Franck & McGaugh (2016b,a). We 'observed' the lightcones at *Spitzer* wavelengths and characterized the luminosity functions. We then compared these results using the identical procedure in Franck & McGaugh (2017). By tracking the growth of structure in the Millennium Simulation, we evaluated the precision with which the CCPC algorithm was able to identify simulated galaxies that will become cluster galaxies by $z = 0$. Our results can be summarized as follows:

- The median overdensities found within the simulation were more than a factor of 2 larger than those found in Franck & McGaugh (2016b). This may imply that the simulation has greater contrast between field and overdense galaxies than the observations show.

- The observed data show orders of magnitudes greater number density of bright galaxies at 4.5μm wavelengths when compared to the expectations of the simulation. This holds true at all epochs, but particularly at the highest redshifts.

- The characteristic magnitude at 4.5μm (m^*) in the real data can be as much as 3 mag-

nitudes brighter than the simulations predict. This trend holds true across multiple semi-analytic models (Somerville et al., 2008; Croton et al., 2016; Henriques et al., 2015) and at all redshifts.

- The evolution of m^* in the CCPC is nearly constant with redshift from $2 < z < 6.6$. This trend is consistent with a simple stellar population model that formed at $z_f > 7$. In contrast, the simulation m^* grows by magnitudes over the same time period, and can be best fit with a constant star formation model.

- Finally, we evaluated the ability of the CCPC algorithm to detect simulated protocluster galaxies. The algorithm correctly identified roughly 1/3 of all protocluster galaxies within the lightcones.

The CCPC expanded the number of protoclusters from 10s to 100s of systems. Most fundamentally, by estimating their stellar mass with *Spitzer* observations, we have shown that these galaxies appear to evolve consistent with stellar models of passive evolution and formed very early in the universe ($z_f > 7$). This is in stark contrast to the expectations from theory, which suggest that galaxies should be growing in mass rapidly between $2 < z < 7$. The simulations also expect many fewer bright galaxies than are actually observed. This thesis work strongly suggests that galaxies are more massive and form earlier than current galaxy formation models predict. ΛCDM predicts that galaxies grow hierarchically, but observations have shown there exists an already massive, quiescent population of galaxies is already in existence at high redshift.

The ramifications from this result are potentially serious. At a minimum, the semi-analytic models associated with the Millennium simulation need revision. Perhaps hydrodynamical models are better suited to matching the high redshift galaxies, but that remains to

be seen. Whether ΛCDM can provide a satisfactory explanation for the observed, massive protocluster galaxies poses a fundamental test of the current structure formation paradigm.

Bibliography

Abell, G. O. 1958, Astrophysical Journals, 3, 211

Abraham, R. G., et al. 2004, *Astronomical Journal*, 127, 2455

Abramson, L. E., Gladders, M. D., Dressler, A., Oemler, Jr., A., Poggianti, B., & Vulcani, B. 2016, *Astrophysical Journal*, 832, 7

Adams, J. J., et al. 2011, *Astrophysical Journal, Supplement*, 192, 5

Adams, S. M., Martini, P., Croxall, K. V., Overzier, R. A., & Silverman, J. D. 2015, *Monthly Notices of the Royal Astronomical Society*, 448, 1335

Akiyama, M. 2005, *Astrophysical Journal*, 629, 72

Anderson, S. F., & Margon, B. 1987, *Astrophysical Journal*, 314, 111

Anderson, S. F., et al. 2001, *Astronomical Journal*, 122, 503

Ando, M., Ohta, K., Iwata, I., Watanabe, C., Tamura, N., Akiyama, M., & Aoki, K. 2004, *Astrophysical Journal*, 610, 635

Andreon, S. 2006, *Astronomy and Astrophysics*, 448, 447

—. 2013, *Astronomy and Astrophysics*, 554, A79

Astropy Collaboration et al. 2013, *Astronomy and Astrophysics*, 558, A33

Bahcall, N. A., & Fan, X. 1998, *Astrophysical Journal*, 504, 1

Baldry, I. K., Glazebrook, K., Brinkmann, J., Ivezić, Ž., Lupton, R. H., Nichol, R. C., & Szalay, A. S. 2004, *Astrophysical Journal*, 600, 681

Balestra, I., et al. 2010, Astronomy and Astrophysics, 512, A12

Barger, A. J., Cowie, L. L., Bautz, M. W., Brandt, W. N., Garmire, G. P., Hornschemeier, A. E., Ivison, R. J., & Owen, F. N. 2001a, *Astronomical Journal*, 122, 2177

Barger, A. J., Cowie, L. L., Brandt, W. N., Capak, P., Garmire, G. P., Hornschemeier, A. E., Steffen, A. T., & Wehner, E. H. 2002, *Astronomical Journal*, 124, 1839

Barger, A. J., Cowie, L. L., Mushotzky, R. F., & Richards, E. A. 2001b, *Astronomical Journal*, 121, 662

Barger, A. J., Cowie, L. L., & Richards, E. A. 2000, *Astronomical Journal*, 119, 2092

Barger, A. J., Cowie, L. L., & Wang, W.-H. 2008, *Astrophysical Journal*, 689, 687

Barger, A. J., et al. 2003, *Astronomical Journal*, 126, 632

Bauer, F. E., et al. 2002, *Astronomical Journal*, 123, 1163

Baum, W. A. 1959, *Publications of the ASP*, 71, 106

Beckwith, S. V. W., et al. 2006, *Astronomical Journal*, 132, 1729

Beers, T. C., Flynn, K., & Gebhardt, K. 1990, *Astronomical Journal*, 100, 32

Bertincourt, B., et al. 2009, *Astrophysical Journal*, 705, 68

Bielby, R., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 430, 425

Bleem, L. E., et al. 2015, Astrophysical Journals, 216, 27

Bolzonella, M., Miralles, J.-M., & Pelló, R. 2000, *Astronomy and Astrophysics*, 363, 476

Bond, N. A., Feldmeier, J. J., Matković, A., Gronwall, C., Ciardullo, R., & Gawiser, E. 2010, *Astrophysical Journal, Letters*, 716, L200

Bond, N. A., Gawiser, E., Guaita, L., Padilla, N., Gronwall, C., Ciardullo, R., & Lai, K. 2012, *Astrophysical Journal*, 753, 95

Bond, N. A., Gawiser, E., & Koekemoer, A. M. 2011, *Astrophysical Journal*, 729, 48

Bond, N. A., et al. 2014, *Astrophysical Journal*, 791, 18

Bonzini, M., et al. 2012, *Astrophysical Journal, Supplement*, 203, 15

Bothwell, M. S., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 429, 3047

Boutsia, K., et al. 2011, *Astrophysical Journal*, 736, 41

Bouwens, R. J., et al. 2015, *Astrophysical Journal*, 803, 34

Bower, R. G., Lucey, J. R., & Ellis, R. S. 1992, Monthly Notices, 254, 601

Bowler, R. A. A., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 2810

Brammer, G. B., van Dokkum, P. G., Illingworth, G. D., Bouwens, R. J., Labbé, I., Franx, M., Momcheva, I., & Oesch, P. A. 2013, *Astrophysical Journal, Letters*, 765, L2

Brodwin, M., et al. 2013, *Astrophysical Journal*, 779, 138

Brusa, M., et al. 2009a, *Astronomy and Astrophysics*, 507, 1277

—. 2009b, *Astrophysical Journal*, 693, 8

—. 2010, *Astrophysical Journal*, 716, 348

Bruzual, G., & Charlot, S. 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 1000

Buchner, J., et al. 2014, *Astronomy and Astrophysics*, 564, A125

Bunker, A. J., Stanway, E. R., Ellis, R. S., & McMahon, R. G. 2004, *Monthly Notices of the Royal Astronomical Society*, 355, 374

Burbidge, E. M., Junkkarinen, V. T., Koski, A. T., Smith, H. E., & Hoag, A. A. 1980, Astrophysical Journall, 242, L55

Butcher, H., & Oemler, Jr., A. 1978, Astrophysical Journal, 219, 18

—. 1984, Astrophysical Journal, 285, 426

Cai, Z., et al. 2016, *Astrophysical Journal*, 833, 135

Campos, A., Yahil, A., Windhorst, R. A., Richards, E. A., Pascarelle, S., Impey, C., & Petry, C. 1999, *Astrophysical Journal, Letters*, 511, L1

Cantalupo, S., Lilly, S. J., & Porciani, C. 2007, *Astrophysical Journal*, 657, 135

Carilli, C. L., Riechers, D., Walter, F., Maiolino, R., Wagg, J., Lentati, L., McMahon, R., & Wolfe, A. 2013, *Astrophysical Journal*, 763, 120

Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abraham, R., Gravel, P., Morris, S., & Pritchet, C. J. 1996, *Astrophysical Journal*, 462, 32

Casey, C. M., Chapman, S. C., Smail, I., Alaghband-Zadeh, S., Bothwell, M. S., & Swinbank, A. M. 2011, *Monthly Notices of the Royal Astronomical Society*, 411, 2739

Casey, C. M., et al. 2012, *Astrophysical Journal*, 761, 139

—. 2015, *Astrophysical Journal, Letters*, 808, L33

Castro-Rodríguez, N., & López-Corredoira, M. 2012, *Astronomy and Astrophysics*, 537, A31

Cen, R., Bahcall, N. A., & Gramann, M. 1994, Astrophysical Journall, 437, L51

Cen, R., & Ostriker, J. P. 1994, *Astrophysical Journal*, 429, 4

Chabrier, G. 2003, *Publications of the ASP*, 115, 763

Chapman, S. C., Scott, D., Windhorst, R. A., Frayer, D. T., Borys, C., Lewis, G. F., & Ivison, R. J. 2004a, *Astrophysical Journal*, 606, 85

Chapman, S. C., Smail, I., Blain, A. W., & Ivison, R. J. 2004b, *Astrophysical Journal*, 614, 671

Chapman, S. C., Smail, I., Windhorst, R., Muxlow, T., & Ivison, R. J. 2004c, *Astrophysical Journal*, 611, 732

Charlot, S., & Fall, S. M. 2000, *Astrophysical Journal*, 539, 718

Chiang, Y.-K., Overzier, R., & Gebhardt, K. 2013, *Astrophysical Journal*, 779, 127

—. 2014, *Astrophysical Journal, Letters*, 782, L3

Chiang, Y.-K., Overzier, R. A., Gebhardt, K., & Henriques, B. 2017, ArXiv e-prints, 1705.01634

Chiang, Y.-K., et al. 2015, *Astrophysical Journal*, 808, 37

Ciardullo, R., et al. 2012, *Astrophysical Journal*, 744, 110

Civano, F., et al. 2011, *Astrophysical Journal*, 741, 91

Clements, D. L., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 1193

Colbert, J. W., Malkan, M. A., & Rich, R. M. 2006, *Astrophysical Journal*, 648, 250

Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, *Astronomical Journal*, 126, 1183

Conselice, C. J., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 80

Contini, E., De Lucia, G., Hatch, N., Borgani, S., & Kang, X. 2016, *Monthly Notices of the Royal Astronomical Society*, 456, 1924

Cooke, E. A., Hatch, N. A., Muldrew, S. I., Rigby, E. E., & Kurk, J. D. 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 3262

Cooke, E. A., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 2318

Cooper, M. C., et al. 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 1058

Coppin, K. E. K., et al. 2008, *Monthly Notices of the Royal Astronomical Society*, 389, 45

—. 2010, *Monthly Notices of the Royal Astronomical Society*, 407, L103

Cowie, L. L., Barger, A. J., Hu, E. M., Capak, P., & Songaila, A. 2004, *Astronomical Journal*, 127, 3137

Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, *Astronomical Journal*, 112, 839

Cowie, L. L., Songaila, A., Kim, T.-S., & Hu, E. M. 1995, *Astronomical Journal*, 109, 1522

Cristiani, S., & D'Odorico, V. 2000, *Astronomical Journal*, 120, 1648

Cristiani, S., et al. 2000, *Astronomy and Astrophysics*, 359, 489

Croft, S., Kurk, J., van Breugel, W., Stanford, S. A., de Vries, W., Pentericci, L., & Röttgering, H. 2005, *Astronomical Journal*, 130, 867

Croton, D. J., et al. 2016, *Astrophysical Journal, Supplement*, 222, 22

Cucciati, O., et al. 2014, *Astronomy and Astrophysics*, 570, A16

Curtis-Lake, E., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 422, 1425

Daddi, E., Dannerbauer, H., Krips, M., Walter, F., Dickinson, M., Elbaz, D., & Morrison, G. E. 2009, *Astrophysical Journal, Letters*, 695, L176

Daddi, E., et al. 2005, *Astrophysical Journal*, 626, 680

Dannerbauer, H., Lehnert, M. D., Lutz, D., Tacconi, L., Bertoldi, F., Carilli, C., Genzel, R., & Menten, K. M. 2004, *Astrophysical Journal*, 606, 664

Dannerbauer, H., et al. 2006, *Astrophysical Journal, Letters*, 637, L5

Davies, L. J. M., Bremer, M. N., Stanway, E. R., Birkinshaw, M., & Lehnert, M. D. 2010, *Monthly Notices of the Royal Astronomical Society*, 408, L31

Davis, M., & Peebles, P. J. E. 1983, Astrophysical Journal, 267, 465

Dawson, S., Rhoads, J. E., Malhotra, S., Stern, D., Wang, J., Dey, A., Spinrad, H., & Jannuzi, B. T. 2007, *Astrophysical Journal*, 671, 1227

Dawson, S., Spinrad, H., Stern, D., Dey, A., van Breugel, W., de Vries, W., & Reuland, M. 2002, *Astrophysical Journal*, 570, 92

Dawson, S., Stern, D., Bunker, A. J., Spinrad, H., & Dey, A. 2001, *Astronomical Journal*, 122, 598

Dawson, S., et al. 2004, *Astrophysical Journal*, 617, 707

De Breuck, C., et al. 2004, *Astronomy and Astrophysics*, 424, 1

de Bruyn, A. G., O'Dea, C. P., & Baum, S. A. 1996, *Astronomy and Astrophysics*, 305, 450

De Lucia, G. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 379, Cosmic Frontiers, ed. N. Metcalfe & T. Shanks, 257

De Lucia, G., & Blaizot, J. 2007, Monthly Notices, 375, 2

Devriendt, J. E. G., Guiderdoni, B., & Sadat, R. 1999, *Astronomy and Astrophysics*, 350, 381

Dey, A., Lee, K.-S., Reddy, N., Cooper, M., Inami, H., Hong, S., Gonzalez, A. H., & Jannuzi, B. T. 2016, *Astrophysical Journal*, 823, 11

Diener, C., et al. 2013, *Astrophysical Journal*, 765, 109

—. 2015, *Astrophysical Journal*, 802, 31

Digby-North, J. A., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 407, 846

Dijkstra, M., & Westra, E. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 2343

Djorgovski, S. G., Stern, D., Mahabal, A. A., & Brunner, R. 2003, *Astrophysical Journal*, 596, 67

Dobrzycki, A., & Bechtold, J. 1996, *Astrophysical Journal*, 457, 102

Doherty, M., et al. 2010, *Astronomy and Astrophysics*, 509, A83

Donley, J. L., Rieke, G. H., Alexander, D. M., Egami, E., & Pérez-González, P. G. 2010, *Astrophysical Journal*, 719, 1393

Donley, J. L., Rieke, G. H., Pérez-González, P. G., Rigby, J. R., & Alonso-Herrero, A. 2007, *Astrophysical Journal*, 660, 167

Dow-Hygelund, C. C., et al. 2007, *Astrophysical Journal*, 660, 47

Dressler, A. 1980, Astrophysical Journal, 236, 351

Eales, S., Bertoldi, F., Ivison, R., Carilli, C., Dunne, L., & Owen, F. 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 169

Eisenhardt, P. R. M., et al. 2008, *Astrophysical Journal*, 684, 905

Eke, V. R., Cole, S., & Frenk, C. S. 1996a, Monthly Notices, 282

Eke, V. R., Cole, S., Frenk, C. S., & Navarro, J. F. 1996b, Monthly Notices, 281, 703

Ellison, S. L., Pettini, M., Steidel, C. C., & Shapley, A. E. 2001, *Astrophysical Journal*, 549, 770

Ellison, S. L., Simard, L., Cowan, N. B., Baldry, I. K., Patton, D. R., & McConnachie, A. W. 2009, *Monthly Notices of the Royal Astronomical Society*, 396, 1257

Elmegreen, B. G., & Elmegreen, D. M. 2010, *Astrophysical Journal*, 722, 1895

Elston, R., Bechtold, J., Hill, G. J., & Ge, J. 1996, *Astrophysical Journal, Letters*, 456, L13

Erb, D. K., Bogosavljević, M., & Steidel, C. C. 2011, *Astrophysical Journal, Letters*, 740, L31

Erb, D. K., Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., Hunt, M. P., Moorwood, A. F. M., & Cuby, J.-G. 2003, *Astrophysical Journal*, 591, 101

Erb, D. K., Steidel, C. C., Shapley, A. E., Pettini, M., & Adelberger, K. L. 2004, *Astrophysical Journal*, 612, 122

Erb, D. K., Steidel, C. C., Shapley, A. E., Pettini, M., Reddy, N. A., & Adelberger, K. L. 2006, *Astrophysical Journal*, 646, 107

Evrard, A. E. 1989, Astrophysical Journal, 341, L71

Faber, S. M. 1973, Astrophysical Journal, 179, 731

Faber, S. M., et al. 2007, *Astrophysical Journal*, 665, 265

Fasano, G., Cristiani, S., Arnouts, S., & Filippi, M. 1998, *Astronomical Journal*, 115, 1400

Fassbender, R., et al. 2014, Astronomy and Astrophysics, 568, A5

Fernández-Soto, A., Lanzetta, K. M., Chen, H.-W., Pascarelle, S. M., & Yahata, N. 2001, *Astrophysical Journal, Supplement*, 135, 41

Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, *Astrophysical Journal*, 513, 34

Ferreras, I., et al. 2012, Astronomical Journal, 144, 47

Feruglio, C., Daddi, E., Fiore, F., Alexander, D. M., Piconcelli, E., & Malacaria, C. 2011, *Astrophysical Journal, Letters*, 729, L4

Finkelstein, S. L., Malhotra, S., Rhoads, J. E., Hathi, N. P., & Pirzkal, N. 2009a, *Monthly Notices of the Royal Astronomical Society*, 393, 1174

Finkelstein, S. L., Rhoads, J. E., Malhotra, S., & Grogin, N. 2009b, *Astrophysical Journal*, 691, 465

Finkelstein, S. L., et al. 2011, *Astrophysical Journal*, 735, 5

Finlator, K., Davé, R., & Oppenheimer, B. D. 2007, *Monthly Notices of the Royal Astronomical Society*, 376, 1861

Fiore, F., et al. 2012, *Astronomy and Astrophysics*, 537, A16

Fontanot, F., Cristiani, S., Monaco, P., Nonino, M., Vanzella, E., Brandt, W. N., Grazian, A., & Mao, J. 2007, *Astronomy and Astrophysics*, 461, 39

Francis, P. J., et al. 1996, Astrophysical Journal, 457, 490

Franck, J. R., & McGaugh, S. S. 2016a, *Astrophysical Journal*, 833, 15

—. 2016b, *Astrophysical Journal*, 817, 158

—. 2017, *Astrophysical Journal*, 836, 136

Franck, J. R., McGaugh, S. S., & Schombert, J. M. 2015, *Astronomical Journal*, 150, 46

Frank, S., Appenzeller, I., Noll, S., & Stahl, O. 2003, *Astronomy and Astrophysics*, 407, 473

Fritz, A., et al. 2014, Astronomy and Astrophysics, 563, A92

Fry, J. N. 1996, *Astrophysical Journal, Letters*, 461, L65

Fynbo, J. P. U., Ledoux, C., Möller, P., Thomsen, B., & Burud, I. 2003, *Astronomy and Astrophysics*, 407, 147

Fynbo, J. U., Möller, P., & Thomsen, B. 2001, *Astronomy and Astrophysics*, 374, 443

Galametz, A., Stern, D., Stanford, S. A., De Breuck, C., Vernet, J., Griffith, R. L., & Harrison, F. A. 2010, Astronomy and Astrophysics, 516, A101

Galametz, A., et al. 2012, *Astrophysical Journal*, 749, 169

—. 2013, *Astronomy and Astrophysics*, 559, A2

Gavignaud, I., et al. 2006, *Astronomy and Astrophysics*, 457, 79

—. 2008, *Astronomy and Astrophysics*, 492, 637

Gawiser, E., et al. 2007, *Astrophysical Journal*, 671, 278

Georgakakis, A., et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 371, 221

Georgantopoulos, I., Rovilos, E., Xilouris, E. M., Comastri, A., & Akylas, A. 2011, *Astronomy and Astrophysics*, 526, A86

Gladders, M. D., Oemler, A., Dressler, A., Poggianti, B., Vulcani, B., & Abramson, L. 2013, *Astrophysical Journal*, 770, 64

Glazebrook, K., et al. 2017, *Nature*, 544, 71

Gnerucci, A., et al. 2011, *Astronomy and Astrophysics*, 528, A88

Gobat, R., et al. 2011, Astronomy and Astrophysics, 526, A133

—. 2012, *Astrophysical Journal, Letters*, 759, L44

—. 2013, *Astrophysical Journal*, 776, 9

Gonzalez, A. H., Tran, K.-V. H., Conbere, M. N., & Zaritsky, D. 2005, *Astrophysical Journal, Letters*, 624, L73

Granato, G. L., Ragone-Figueroa, C., Domínguez-Tenreiro, R., Obreja, A., Borgani, S., De Lucia, G., & Murante, G. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1320

Grazian, A., et al. 2006, *Astronomy and Astrophysics*, 449, 951

—. 2012, *Astronomy and Astrophysics*, 547, A51

Grogin, N. A., et al. 2011, *Astrophysical Journal, Supplement*, 197, 35

Gronwall, C., et al. 2007, *Astrophysical Journal*, 667, 79

Grove, L. F., Fynbo, J. P. U., Ledoux, C., Limousin, M., Møller, P., Nilsson, K. K., & Thomsen, B. 2009, *Astronomy and Astrophysics*, 497, 689

Gunn, J. E., & Gott, III, J. R. 1972, *Astrophysical Journal*, 176, 1

Guo, Q., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 101

Hainline, K. N., Shapley, A. E., Greene, J. E., & Steidel, C. C. 2011, *Astrophysical Journal*, 733, 31

Hainline, L. J., Blain, A. W., Greve, T. R., Chapman, S. C., Smail, I., & Ivison, R. J. 2006, *Astrophysical Journal*, 650, 614

Hall, P. B., & Green, R. F. 1998, *Astrophysical Journal*, 507, 558

Hall, P. B., et al. 2001, Astronomical Journal, 121, 1840

Hamann, F., Barlow, T. A., & Junkkarinen, V. 1997, *Astrophysical Journal*, 478, 87

Harrison, C. M., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 1073

Hatch, N. A., Cooke, E. A., Muldrew, S. I., Hartley, W. G., Almaini, O., Conselice, C. J., & Simpson, C. J. 2016, *Monthly Notices of the Royal Astronomical Society*

—. 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 876

Hatch, N. A., Kurk, J. D., Pentericci, L., Venemans, B. P., Kuiper, E., Miley, G. K., & Röttgering, H. J. A. 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 2993

Hathi, N. P., Malhotra, S., & Rhoads, J. E. 2008, *Astrophysical Journal*, 673, 686

Hawkins, M. R. S., & Veron, P. 1993, *Monthly Notices of the Royal Astronomical Society*, 260, 202

Hayashi, M., Kodama, T., Koyama, Y., Tadaki, K.-I., & Tanaka, I. 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 2670

Hayashi, M., Kodama, T., Tadaki, K.-i., Koyama, Y., & Tanaka, I. 2012, *Astrophysical Journal*, 757, 15

Hayashi, M., et al. 2009, *Astrophysical Journal*, 691, 140

Hayashino, T., et al. 2004, *Astronomical Journal*, 128, 2073

Henriques, B. M. B., White, S. D. M., Lemson, G., Thomas, P. A., Guo, Q., Marleau, G.-D., & Overzier, R. A. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 2904

Henriques, B. M. B., White, S. D. M., Thomas, P. A., Angulo, R., Guo, Q., Lemson, G., Springel, V., & Overzier, R. 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 2663

Henry, A. L., Martin, C. L., Dressler, A., Sawicki, M., & McCarthy, P. 2012, *Astrophysical Journal*, 744, 149

Hewett, P. C., & Wild, V. 2010, *Monthly Notices of the Royal Astronomical Society*, 405, 2302

Hill, G. J., et al. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki, 115

Hodge, J. A., Carilli, C. L., Walter, F., Daddi, E., & Riechers, D. 2013, *Astrophysical Journal*, 776, 22

Hornschemeier, A. E., et al. 2001, *Astrophysical Journal*, 554, 742

Hu, E. M., Cowie, L. L., Barger, A. J., Capak, P., Kakazu, Y., & Trouille, L. 2010, *Astrophysical Journal*, 725, 394

Hu, E. M., Cowie, L. L., Capak, P., McMahon, R. G., Hayashino, T., & Komiyama, Y. 2004, *Astronomical Journal*, 127, 563

Hu, E. M., & McMahon, R. G. 1996, *Nature*, 382, 231

Hu, E. M., McMahon, R. G., & Cowie, L. L. 1999, *Astrophysical Journal, Letters*, 522, L9

Hu, E. M., McMahon, R. G., & Egami, E. 1996, *Astrophysical Journal, Letters*, 459, L53

Huchra, J. P., & Geller, M. J. 1982, *Astrophysical Journal*, 257, 423

Husband, K., Bremer, M. N., Stanway, E. R., Davies, L. J. M., Lehnert, M. D., & Douglas, L. S. 2013, *Monthly Notices of the Royal Astronomical Society*, 432, 2869

Husband, K., Bremer, M. N., Stott, J. P., & Murphy, D. N. A. 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 421

Inoue, A. K., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 411, 2336

Ishigaki, M., Ouchi, M., & Harikane, Y. 2016, *Astrophysical Journal*, 822, 5

Ivison, R. J., Papadopoulos, P. P., Smail, I., Greve, T. R., Thomson, A. P., Xilouris, E. M., & Chapman, S. C. 2011, *Monthly Notices of the Royal Astronomical Society*, 412, 1913

Iwata, I., Inoue, A. K., & Burgarella, D. 2005, *Astronomy and Astrophysics*, 440, 881

Iwata, I., Ohta, K., Tamura, N., Akiyama, M., Aoki, K., Ando, M., Kiuchi, G., & Sawicki, M. 2007, *Monthly Notices of the Royal Astronomical Society*, 376, 1557

Iye, M., et al. 2006, *Nature*, 443, 186

Jangren, A., Wegner, G., Salzer, J. J., Werk, J. K., & Gronwall, C. 2005, *Astronomical Journal*, 130, 496

Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jiang, L., et al. 2011, *Astrophysical Journal*, 743, 65

—. 2013, *Astrophysical Journal*, 772, 99

Kajino, H., et al. 2009, *Astrophysical Journal*, 704, 117

Kakazu, Y., Cowie, L. L., & Hu, E. M. 2007, *Astrophysical Journal*, 668, 853

Kang, E., & Im, M. 2009, *Astrophysical Journal, Letters*, 691, L33

—. 2015, Journal of Korean Astronomical Society, 48, 21

Kashikawa, N., et al. 2006, *Astrophysical Journal*, 648, 7

—. 2011, *Astrophysical Journal*, 734, 119

Katz, H., McGaugh, S., Teuben, P., & Angus, G. W. 2013, *Astrophysical Journal*, 772, 10

Keel, W. C., Cohen, S. H., Windhorst, R. A., & Waddington, I. 1999, Astronomical Journal, 118, 2547

Keel, W. C., Wu, W., Waddington, I., Windhorst, R. A., & Pascarelle, S. M. 2002, *Astronomical Journal*, 123, 3041

Keisler, R., et al. 2011, Astrophysical Journal, 743, 28

Kilic, M., Mendez, R. A., von Hippel, T., & Winget, D. E. 2005, *Astrophysical Journal*, 633, 1126

Kirkpatrick, A., et al. 2012, *Astrophysical Journal*, 759, 139

Kitzbichler, M. G., & White, S. D. M. 2007, *Monthly Notices of the Royal Astronomical Society*, 376, 2

Knudsen, K. K., van der Werf, P. P., & Kneib, J.-P. 2008, *Monthly Notices of the Royal Astronomical Society*, 384, 1611

Kocevski, D. D., Lubin, L. M., Gal, R., Lemaux, B. C., Fassnacht, C. D., & Squires, G. K. 2009, *Astrophysical Journal*, 690, 295

Kochanek, C. S., et al. 2012, *Astrophysical Journal, Supplement*, 200, 8

Koyama, Y., Kodama, T., Tadaki, K.-i., Hayashi, M., Tanaka, M., Smail, I., Tanaka, I., & Kurk, J. 2013, *Monthly Notices of the Royal Astronomical Society*, 428, 1551

Kriek, M., et al. 2008, *Astrophysical Journal*, 677, 219

—. 2010, *Astrophysical Journal, Letters*, 722, L64

Kroupa, P. 2002, Science, 295, 82

Krumpe, M., et al. 2008, *Astronomy and Astrophysics*, 483, 415

Kuiper, E., Venemans, B. P., Hatch, N. A., Miley, G. K., & Röttgering, H. J. A. 2012, *Monthly Notices of the Royal Astronomical Society*, 425, 801

Kuiper, E., et al. 2011, *Monthly Notices of the Royal Astronomical Society*, 417, 1088

Kulas, K. R., Shapley, A. E., Kollmeier, J. A., Zheng, Z., Steidel, C. C., & Hainline, K. N. 2012, *Astrophysical Journal*, 745, 33

Kurk, J. D., Pentericci, L., Overzier, R. A., Röttgering, H. J. A., & Miley, G. K. 2004a, *Astronomy and Astrophysics*, 428, 817

Kurk, J. D., Pentericci, L., Röttgering, H. J. A., & Miley, G. K. 2004b, *Astronomy and Astrophysics*, 428, 793

Lanzetta, K. M., Yahil, A., & Fernández-Soto, A. 1996, *Nature*, 381, 759

Larson, R. B., Tinsley, B. M., & Caldwell, C. N. 1980, *Astrophysical Journal*, 237, 692

Law, D. I. 2008, NED: Public Communication, 1

Law, D. R., Steidel, C. C., Shapley, A. E., Nagy, S. R., Reddy, N. A., & Erb, D. K. 2012, *Astrophysical Journal*, 745, 85

Le Fevre, O., Deltorn, J. M., Crampton, D., & Dickinson, M. 1996, *Astrophysical Journal, Letters*, 471, L11

Le Fèvre, O., et al. 2004, *Astronomy and Astrophysics*, 428, 1043

—. 2005, *Astronomy and Astrophysics*, 439, 845

Lee, K.-S., Dey, A., Hong, S., Reddy, N., Wilson, C., Jannuzi, B. T., Inami, H., & Gonzalez, A. H. 2014, *Astrophysical Journal*, 796, 126

Lehmer, B. D., et al. 2009a, *Astrophysical Journal*, 691, 687

—. 2009b, *Monthly Notices of the Royal Astronomical Society*, 400, 299

Lemaux, B. C., et al. 2009, *Astrophysical Journal*, 700, 20

—. 2017, ArXiv e-prints, 1703.10170

Lemoine-Busserolle, M., Bunker, A., Lamareille, F., & Kissler-Patig, M. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 1657

Lilly, S. J., et al. 2007, *Astrophysical Journal, Supplement*, 172, 70

Lowenthal, J. D., Koo, D. C., Simard, L., & van Kampen, E. 2009, *Astrophysical Journal*, 703, 198

Lowenthal, J. D., et al. 1997, *Astrophysical Journal*, 481, 673

Lu, L., Sargent, W. L. W., & Barlow, T. A. 1998, *Astronomical Journal*, 115, 55

Luo, B., et al. 2008, *Astrophysical Journal, Supplement*, 179, 19

Ly, C., et al. 2009, *Astrophysical Journal*, 697, 1410

Madau, P., & Dickinson, M. 2014, *Annural Reviews Astronomy*, 52, 415

Magdis, G. E., Rigopoulou, D., Huang, J.-S., & Fazio, G. G. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 1521

Mainieri, V., et al. 2008, *Astrophysical Journal, Supplement*, 179, 95

Maiolino, R., Caselli, P., Nagao, T., Walmsley, M., De Breuck, C., & Meneghetti, M. 2009, *Astronomy and Astrophysics*, 500, L1

Malhotra, S., et al. 2005, *Astrophysical Journal*, 626, 666

Malkan, M. A., Teplitz, H., & McLean, I. S. 1996, Astrophysical Journall, 468, L9

Mancone, C. L., & Gonzalez, A. H. 2012, *Publications of the ASP*, 124, 606

Mancone, C. L., Gonzalez, A. H., Brodwin, M., Stanford, S. A., Eisenhardt, P. R. M., Stern, D., & Jones, C. 2010, *Astrophysical Journal*, 720, 284

Mantz, A. B., et al. 2014, *Astrophysical Journal*, 794, 157

Maraston, C. 2005, *Monthly Notices of the Royal Astronomical Society*, 362, 799

Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., Franx, M., Labbé, I., & Wuyts, S. 2009, *Astrophysical Journal*, 701, 1765

Martin, C. L., Sawicki, M., Dressler, A., & McCarthy, P. 2008, *Astrophysical Journal*, 679, 942

Maschietto, F., et al. 2008, *Monthly Notices of the Royal Astronomical Society*, 389, 1223

Matsuda, Y., et al. 2005, *Astrophysical Journal, Letters*, 634, L125

—. 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 2041

Mawatari, K., Yamada, T., Fazio, G. G., Huang, J.-S., & Ashby, M. L. N. 2016, *Publications of the ASJ*, 68, 46

Mazzucchelli, C., Bañados, E., Decarli, R., Farina, E. P., Venemans, B. P., Walter, F., & Overzier, R. 2017, *Astrophysical Journal*, 834, 83

McCarthy, P. J., Kapahi, V. K., van Breugel, W., Persson, S. E., Athreya, R., & Subrahmanya, C. R. 1996, *Astrophysical Journal, Supplement*, 107, 19

McGaugh, S. S. 1996, *Monthly Notices of the Royal Astronomical Society*, 280, 337

—. 2004, *Astrophysical Journal*, 611, 26

—. 2015, Canadian Journal of Physics, 93, 250

McIntosh, D. H., Impey, C. D., & Petry, C. E. 2004, *Astronomical Journal*, 128, 544

McLure, R. J., et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 372, 357

Mehlert, D., et al. 2002, *Astronomy and Astrophysics*, 393, 809

Mei, S., et al. 2009, *Astrophysical Journal*, 690, 42

Melnyk, O., et al. 2013, *Astronomy and Astrophysics*, 557, A81

Menéndez-Delmestre, K., et al. 2009, *Astrophysical Journal*, 699, 667

Mignoli, M., et al. 2005, *Astronomy and Astrophysics*, 437, 883

Mihos, J. C., & Hernquist, L. 1996, *Astrophysical Journal*, 464, 641

Möller, P., & Fynbo, J. U. 2001, *Astronomy and Astrophysics*, 372, L57

Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, *Nature*, 379, 613

Moorwood, A. F. M., van der Werf, P. P., Cuby, J. G., & Oliva, E. 2000, *Astronomy and Astrophysics*, 362, 9

Mortonson, M. J., Hu, W., & Huterer, D. 2011, Physical Review Letters, 83, 023015

Moth, P., & Elston, R. J. 2002, *Astronomical Journal*, 124, 1886

Muchovej, S., et al. 2007, Astrophysical Journal, 663, 708

Muldrew, S. I., Hatch, N. A., & Cooke, E. A. 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 2528

Murayama, T., et al. 2007, *Astrophysical Journal, Supplement*, 172, 523

Muzzin, A., Wilson, G., Lacy, M., Yee, H. K. C., & Stanford, S. A. 2008, *Astrophysical Journal*, 686, 966

Nagao, T., et al. 2004, *Astrophysical Journal, Letters*, 613, L9

—. 2005, *Astrophysical Journal*, 634, 142

—. 2007, *Astronomy and Astrophysics*, 468, 877

Neri, R., Downes, D., Cox, P., & Walter, F. 2014, *Astronomy and Astrophysics*, 562, A35

Nestor, D. B., Shapley, A. E., Steidel, C. C., & Siana, B. 2011, *Astrophysical Journal*, 736, 18

Noirot, G., et al. 2016, *Astrophysical Journal*, 830, 90

Noll, S., et al. 2004, *Astronomy and Astrophysics*, 418, 885

Nusser, A. 2002, *Monthly Notices of the Royal Astronomical Society*, 331, 909

Oemler, Jr., A., Dressler, A., Gladders, M. G., Fritz, J., Poggianti, B. M., Vulcani, B., & Abramson, L. 2013, *Astrophysical Journal*, 770, 63

Ogura, K., et al. 2017, *Publications of the ASJ*, 69, 51

Ono, Y., et al. 2012, *Astrophysical Journal*, 744, 83

—. 2017, ArXiv e-prints, 1704.06004

Onodera, M., Arimoto, N., Daddi, E., Renzini, A., Kong, X., Cimatti, A., Broadhurst, T., & Alexander, D. M. 2010, *Astrophysical Journal*, 715, 385

Osmer, P. S., Porter, A. C., & Green, R. F. 1994, *Astrophysical Journal*, 436, 678

Ouchi, M., et al. 2003, *Astrophysical Journal*, 582, 60

—. 2004a, *Astrophysical Journal*, 611, 660

—. 2004b, *Astrophysical Journal*, 611, 685

—. 2005, *Astrophysical Journal, Letters*, 620, L1

—. 2008, *Astrophysical Journal, Supplement*, 176, 301

—. 2009, *Astrophysical Journal*, 696, 1164

—. 2010, *Astrophysical Journal*, 723, 869

Overzier, R. A. 2016, Astronomy and Astrophysics Reviews, 24, 14

Ownsworth, J. R., Conselice, C. J., Mundy, C. J., Mortlock, A., Hartley, W. G., Duncan, K., & Almaini, O. 2016, *Monthly Notices of the Royal Astronomical Society*, 461, 1112

Papovich, C. 2008, *Astrophysical Journal*, 676, 206

Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, *Astrophysical Journal*, 559, 620

Papovich, C., Dickinson, M., Giavalisco, M., Conselice, C. J., & Ferguson, H. C. 2005, *Astrophysical Journal*, 631, 101

Papovich, C., et al. 2010, *Astrophysical Journal*, 716, 1503

—. 2012, Astrophysical Journal, 750, 93

Park, S. Q., et al. 2010, *Astrophysical Journal*, 717, 1181

Paterno-Mahler, R., Blanton, E. L., Ashby, M. L. N., Brodwin, M., Wing, J. D., Anand, G., Decker, B., & Golden-Marx, E. 2016, ArXiv e-prints, 1611.00746

Patnaik, A. R., Browne, I. W. A., Walsh, D., Chaffee, F. H., & Foltz, C. B. 1992, *Monthly Notices of the Royal Astronomical Society*, 259, 1P

Peebles, P. J. E. 1970, *Astronomical Journal*, 75, 13

Peng, Y.-j., et al. 2010, *Astrophysical Journal*, 721, 193

Pentericci, L., Kurk, J. D., Carilli, C. L., Harris, D. E., Miley, G. K., & Röttgering, H. J. A. 2002, *Astronomy and Astrophysics*, 396, 109

Pentericci, L., Roettgering, H. J. A., Miley, G. K., Carilli, C. L., & McCarthy, P. 1997, Astronomy and Astrophysics, 326, 580

Pentericci, L., et al. 2000, *Astronomy and Astrophysics*, 361, L25

—. 2011, *Astrophysical Journal*, 743, 132

Percival, S. M., Salaris, M., Cassisi, S., & Pietrinferni, A. 2009, *Astrophysical Journal*, 690, 427

Pérez-González, P. G., et al. 2008, *Astrophysical Journal*, 675, 234

Péroux, C., Storrie-Lombardi, L. J., McMahon, R. G., Irwin, M., & Hook, I. M. 2001, *Astronomical Journal*, 121, 1799

Peter, A. H. G., Shapley, A. E., Law, D. R., Steidel, C. C., Erb, D. K., Reddy, N. A., & Pettini, M. 2007, *Astrophysical Journal*, 668, 23

Petitjean, P., & Srianand, R. 1999, *Astronomy and Astrophysics*, 345, 73

Petry, C. E., Impey, C. D., & Foltz, C. B. 1998, *Astrophysical Journal*, 494, 60

Phillips, A. C., Guzmán, R., Gallego, J., Koo, D. C., Lowenthal, J. D., Vogt, N. P., Faber, S. M., & Illingworth, G. D. 1997, *Astrophysical Journal*, 489, 543

Pirzkal, N., et al. 2013, *Astrophysical Journal*, 772, 48

Planck Collaboration et al. 2013, ArXiv e-prints, 1303.5080

—. 2014, *Astronomy and Astrophysics*, 1303.5076, A16

Polletta, M., et al. 2008, *Astronomy and Astrophysics*, 492, 81

Pope, A., et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 370, 1185

Postman, M., Lubin, L. M., & Oke, J. B. 2001, *Astronomical Journal*, 122, 1125

Prescott, M. K. M., Kashikawa, N., Dey, A., & Matsuda, Y. 2008, *Astrophysical Journal, Letters*, 678, L77

Press, W. H., & Davis, M. 1982, *Astrophysical Journal*, 259, 449

Press, W. H., & Schechter, P. 1974, *Astrophysical Journal*, 187, 425

Rafelski, M., Wolfe, A. M., & Chen, H.-W. 2011, *Astrophysical Journal*, 736, 48

Rafelski, M., Wolfe, A. M., Cooke, J., Chen, H.-W., Armandroff, T. E., & Wirth, G. D. 2009, *Astrophysical Journal*, 703, 2033

Raiter, A., Fosbury, R. A. E., & Teimoorinia, H. 2010, *Astronomy and Astrophysics*, 510, A109

Rakos, K. D., & Schombert, J. M. 1995, *Astrophysical Journal*, 439, 47

Ramos Almeida, C., Rodríguez Espinosa, J. M., Barro, G., Gallego, J., & Pérez-González, P. G. 2009, *Astronomical Journal*, 137, 179

Ranalli, P., et al. 2013, *Astronomy and Astrophysics*, 555, A42

Rangel, C., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 3630

Reddy, N. A., Steidel, C. C., Erb, D. K., Shapley, A. E., & Pettini, M. 2006, *Astrophysical Journal*, 653, 1004

Reimers, D., Clavel, J., Groote, D., Engels, D., Hagen, H. J., Naylor, T., Wamsteker, W., & Hopp, U. 1989, *Astronomy and Astrophysics*, 218, 71

Reshetnikov, V. P., & Vasil'Ev, A. A. 2002, Astronomy Letters, 28, 1

Rettura, A., et al. 2014, *Astrophysical Journal*, 797, 109

Rhoads, J. E., et al. 2004, *Astrophysical Journal*, 611, 59

—. 2005, *Astrophysical Journal*, 621, 582

—. 2009, *Astrophysical Journal*, 697, 942

Riechers, D. A., et al. 2006, *Astrophysical Journal*, 650, 604

—. 2011, *Astrophysical Journal, Letters*, 733, L11

Rigby, E. E., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 437, 1882

Rigby, J. R., et al. 2008, *Astrophysical Journal*, 675, 262

Rodney, S. A., et al. 2014, *Astronomical Journal*, 148, 13

Rodriguez-Pascual, P. M., de La Fuente, A., Sanz, J. L., Recondo, M. C., Clavel, J., Santos-Lleo, M., & Wamsteker, W. 1995, *Astrophysical Journal*, 448, 575

Roettgering, H. J. A., van Ojik, R., Miley, G. K., Chambers, K. C., van Breugel, W. J. M., & de Koff, S. 1997, *Astronomy and Astrophysics*, 326, 505

Rong, Y., Jing, Y., Gao, L., Guo, Q., Wang, J., Sun, S., Wang, L., & Pan, J. 2017, ArXiv e-prints, 1704.00012

Ross, N. P., et al. 2012, *Astrophysical Journal, Supplement*, 199, 3

Rottgering, H. J. A., Hunstead, R. W., Miley, G. K., van Ojik, R., & Wieringa, M. H. 1995, *Monthly Notices of the Royal Astronomical Society*, 277, 389

Saito, T., Shimasaku, K., Okamura, S., Ouchi, M., Akiyama, M., & Yoshida, M. 2006, *Astrophysical Journal*, 648, 54

Saito, T., Shimasaku, K., Okamura, S., Ouchi, M., Akiyama, M., Yoshida, M., & Ueda, Y. 2008, *Astrophysical Journal*, 675, 1076

Salimbeni, S., et al. 2009, *Astronomy and Astrophysics*, 501, 865

Saliwanchik, B. R., et al. 2015, Astrophysical Journal, 799, 137

Sand, D. J., Treu, T., Ellis, R. S., & Smith, G. P. 2005, *Astrophysical Journal*, 627, 32

Sanders, R. H. 1998, *Monthly Notices of the Royal Astronomical Society*, 296, 1009

Santini, P., et al. 2009, *Astronomy and Astrophysics*, 504, 751

Santos, M. R., Ellis, R. S., Kneib, J.-P., Richard, J., & Kuijken, K. 2004, *Astrophysical Journal*, 606, 683

Sarajedini, V. L., Koo, D. C., Klesman, A. J., Laird, E. S., Perez Gonzalez, P. G., & Mozena, M. 2011, *Astrophysical Journal*, 731, 97

Savaglio, S., Cristiani, S., D'Odorico, S., Fontana, A., Giallongo, E., & Molaro, P. 1997, *Astronomy and Astrophysics*, 318, 347

Savaglio, S., D'Odorico, S., & Moller, P. 1994, *Astronomy and Astrophysics*, 281, 331

Schaye, J., Theuns, T., Rauch, M., Efstathiou, G., & Sargent, W. L. W. 2000, *Monthly Notices of the Royal Astronomical Society*, 318, 817

Schechter, P. 1976, *Astrophysical Journal*, 203, 297

Schenker, M. A., Stark, D. P., Ellis, R. S., Robertson, B. E., Dunlop, J. S., McLure, R. J., Kneib, J.-P., & Richard, J. 2012, *Astrophysical Journal*, 744, 179

Schneider, D. P., Schmidt, M., & Gunn, J. E. 1991, *Astronomical Journal*, 101, 2004

Schombert, J., & Rakos, K. 2009, Astronomical Journal, 137, 528

Schreier, E. J., et al. 2001, *Astrophysical Journal*, 560, 127

Scott, J., Bechtold, J., & Dobrzycki, A. 2000, *Astrophysical Journal, Supplement*, 130, 37

Sealey, K. M., Drinkwater, M. J., & Webb, J. K. 1998, *Astrophysical Journal, Letters*, 499, L135

Shapley, A. E., Erb, D. K., Pettini, M., Steidel, C. C., & Adelberger, K. L. 2004, *Astrophysical Journal*, 612, 108

Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, *Astrophysical Journal*, 562, 95

Shapley, A. E., Steidel, C. C., Erb, D. K., Reddy, N. A., Adelberger, K. L., Pettini, M., Barmby, P., & Huang, J. 2005, *Astrophysical Journal*, 626, 698

Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., & Erb, D. K. 2006, *Astrophysical Journal*, 651, 688

Sharp, R. G., Sabbey, C. N., Vivas, A. K., Oemler, A., McMahon, R. G., Hodgkin, S. T., & Coppi, P. S. 2002, *Monthly Notices of the Royal Astronomical Society*, 337, 1153

Sheth, R. K. 2001, in Annals of the New York Academy of Sciences, Vol. 927, The Onset of Nonlinearity in Cosmology, ed. J. N. Fry, J. R. Buchler, & H. Kandrup, 1–12

Shim, H., Chary, R.-R., Dickinson, M., Lin, L., Spinrad, H., Stern, D., & Yan, C.-H. 2011, *Astrophysical Journal*, 738, 69

Shimakawa, R., Kodama, T., Tadaki, K.-i., Tanaka, I., Hayashi, M., & Koyama, Y. 2014, *Monthly Notices of the Royal Astronomical Society*, 441, L1

Shimasaku, K., et al. 2003, *Astrophysical Journal, Letters*, 586, L111

—. 2006, *Publications of the ASJ*, 58, 313

Silverman, J. D., et al. 2005, *Astrophysical Journal*, 618, 123

—. 2010, *Astrophysical Journal, Supplement*, 191, 124

Simcoe, R. A., Sargent, W. L. W., Rauch, M., & Becker, G. 2006, *Astrophysical Journal*, 637, 648

Simpson, J. M., et al. 2014, *Astrophysical Journal*, 788, 125

Skelton, R. E., et al. 2014, *Astrophysical Journal, Supplement*, 214, 24

Smail, I., Chapman, S. C., Blain, A. W., & Ivison, R. J. 2004, *Astrophysical Journal*, 616, 71

Söchting, I. K., Coldwell, G. V., Clowes, R. G., Campusano, L. E., & Graham, M. J. 2012, *Monthly Notices of the Royal Astronomical Society*, 423, 2436

Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, *Monthly Notices of the Royal Astronomical Society*, 391, 481

Song, M., et al. 2014, *Astrophysical Journal*, 791, 3

Songaila, A. 1998, *Astronomical Journal*, 115, 2184

Spinrad, H., Stern, D., Bunker, A., Dey, A., Lanzetta, K., Yahil, A., Pascarelle, S., & Fernández-Soto, A. 1998, *Astronomical Journal*, 116, 2617

Spitler, L. R., et al. 2014, *Astrophysical Journal, Letters*, 787, L36

Springel, V., et al. 2005, *Nature*, 435, 629

Srianand, R. 1996, *Astrophysical Journal*, 462, 68

Stalin, C. S., Petitjean, P., Srianand, R., Fox, A. J., Coppolani, F., & Schwope, A. 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 294

Stanford, S. A., Dickinson, M., Postman, M., Ferguson, H. C., Lucas, R. A., Conselice, C. J., Budavári, T., & Somerville, R. 2004, *Astronomical Journal*, 127, 131

Stanford, S. A., Eisenhardt, P. R., & Dickinson, M. 1998, *Astrophysical Journal*, 492, 461

Stanford, S. A., Eisenhardt, P. R. M., & Dickinson, M. 1995, *Astrophysical Journal*, 450, 512

Stanway, E. R., Bunker, A. J., McMahon, R. G., Ellis, R. S., Treu, T., & McCarthy, P. J. 2004a, *Astrophysical Journal*, 607, 704

Stanway, E. R., et al. 2004b, *Astrophysical Journal, Letters*, 604, L13

—. 2007, *Monthly Notices of the Royal Astronomical Society*, 376, 727

Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, *Astrophysical Journal, Letters*, 728, L2

Stark, D. P., Schenker, M. A., Ellis, R., Robertson, B., McLure, R., & Dunlop, J. 2013, *Astrophysical Journal*, 763, 129

Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., Pettini, M., & Kellogg, M. 1998, *Astrophysical Journal*, 492, 428

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, *Astrophysical Journal*, 519, 1

Steidel, C. C., Adelberger, K. L., Shapley, A. E., Erb, D. K., Reddy, N. A., & Pettini, M. 2005, *Astrophysical Journal*, 626, 44

Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2000, *Astrophysical Journal*, 532, 170

—. 2003, *Astrophysical Journal*, 592, 728

Steidel, C. C., Shapley, A. E., Pettini, M., Adelberger, K. L., Erb, D. K., Reddy, N. A., & Hunt, M. P. 2004, *Astrophysical Journal*, 604, 534

Steinhardt, C. L., Capak, P., Masters, D., & Speagle, J. S. 2016, *Astrophysical Journal*, 824, 21

Stern, D., Holden, B., Stanford, S. A., & Spinrad, H. 2003, Astronomical Journal, 125, 2759

Stern, D., & Spinrad, H. 1999, *Publications of the ASP*, 111, 1475

Storrie-Lombardi, L. J., McMahon, R. G., Irwin, M. J., & Hazard, C. 1996, *Astrophysical Journal*, 468, 121

Storrie-Lombardi, L. J., & Wolfe, A. M. 2000, *Astrophysical Journal*, 543, 552

Straughn, A. N., et al. 2008, *Astronomical Journal*, 135, 1624

—. 2011, *Astronomical Journal*, 141, 14

Strazzullo, V., et al. 2016, *Astrophysical Journal, Letters*, 833, L20

Suwa, T., Habe, A., & Yoshikawa, K. 2006, *Astrophysical Journal, Letters*, 646, L5

Swinbank, A. M., Smail, I., Chapman, S. C., Blain, A. W., Ivison, R. J., & Keel, W. C. 2004, *Astrophysical Journal*, 617, 64

Szokoly, G. P., et al. 2004, *Astrophysical Journal, Supplement*, 155, 271

Tadaki, K.-I., Kodama, T., Koyama, Y., Hayashi, M., Tanaka, I., & Tokoku, C. 2011, *Publications of the ASJ*, 63, 437

Tan, Q., et al. 2014, *Astronomy and Astrophysics*, 569, A98

Tanaka, M., et al. 2013, *Astrophysical Journal*, 772, 113

Tang, Y., Giavalisco, M., Guo, Y., & Kurk, J. 2014, *Astrophysical Journal*, 793, 92

Taniguchi, Y., et al. 2005, *Publications of the ASJ*, 57, 165

—. 2009, *Astrophysical Journal*, 701, 915

Tanner, A. M., Bechtold, J., Walker, C. E., Black, J. H., & Cutri, R. M. 1996, *Astronomical Journal*, 112, 62

Tapken, C., Appenzeller, I., Noll, S., Richling, S., Heidt, J., Meinköhn, E., & Mehlert, D. 2007, *Astronomy and Astrophysics*, 467, 63

Tapken, C., et al. 2006, *Astronomy and Astrophysics*, 455, 145

Teplitz, H. I., et al. 2007, *Astrophysical Journal*, 659, 941

Tinsley, B. M. 1972, *Astronomy and Astrophysics*, 20, 383

Tody, D. 1986, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 627, Instrumentation in astronomy VI, ed. D. L. Crawford, 733

Tody, D. 1993, in Astronomical Society of the Pacific Conference Series, Vol. 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 173

Topping, M. W., Shapley, A. E., & Steidel, C. C. 2016, *Astrophysical Journal, Letters*, 824, L11

Toshikawa, J., et al. 2014, *Astrophysical Journal*, 792, 15

—. 2016, *Astrophysical Journal*, 826, 114

Tozzi, P., et al. 2009, *Astrophysical Journal*, 698, 740

Tran, K.-V. H., van Dokkum, P., Illingworth, G. D., Kelson, D., Gonzalez, A., & Franx, M. 2005, *Astrophysical Journal*, 619, 134

Treister, E., et al. 2009, *Astrophysical Journal*, 693, 1713

Trenti, M., & Stiavelli, M. 2008, *Astrophysical Journal*, 676, 767

Trenti, M., et al. 2012, *Astrophysical Journal*, 746, 55

Tripp, T. M., Lu, L., & Savage, B. D. 1997, *Astrophysical Journal, Supplement*, 112, 1

Trouille, L., Barger, A. J., Cowie, L. L., Yang, Y., & Mushotzky, R. F. 2008, *Astrophysical Journal, Supplement*, 179, 1

Trump, J. R., et al. 2009, *Astrophysical Journal*, 696, 1195

—. 2011, *Astrophysical Journal*, 743, 144

—. 2014, *Astrophysical Journal*, 793, 101

Uchimoto, Y. K., et al. 2012, *Astrophysical Journal*, 750, 116

Uchiyama, H., et al. 2017, ArXiv e-prints, 1704.06050

Utsumi, Y., Goto, T., Kashikawa, N., Miyazaki, S., Komiyama, Y., Furusawa, H., & Overzier, R. 2010, *Astrophysical Journal*, 721, 1680

van Breukelen, C., Jarvis, M. J., & Venemans, B. P. 2005, *Monthly Notices of the Royal Astronomical Society*, 359, 895

van de Sande, J., Kriek, M., Franx, M., Bezanson, R., & van Dokkum, P. G. 2014, *Astrophysical Journal, Letters*, 793, L31

van Dokkum, P. G. 2005, *Astronomical Journal*, 130, 2647

van Dokkum, P. G., & Franx, M. 2001, *Astrophysical Journal*, 553, 90

van Dokkum, P. G., et al. 2006, *Astrophysical Journal, Letters*, 638, L59

Vanzella, E., et al. 2002, *Astronomy and Astrophysics*, 396, 847

—. 2004, *Astronomy and Astrophysics*, 423, 761

—. 2005, *Astronomy and Astrophysics*, 434, 53

—. 2006, *Astronomy and Astrophysics*, 454, 423

—. 2008, *Astronomy and Astrophysics*, 478, 83

—. 2010, *Astrophysical Journal*, 725, 1011

Venemans, B. P. 2005, PhD thesis, Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

Venemans, B. P., et al. 2002, *Astrophysical Journal, Letters*, 569, L11

—. 2004, *Astronomy and Astrophysics*, 424, L17

—. 2005, *Astronomy and Astrophysics*, 431, 793

—. 2007, *Astronomy and Astrophysics*, 461, 823

Vernet, J., & Cimatti, A. 2001, *Astronomy and Astrophysics*, 380, 409

Veron, P., & Hawkins, M. R. S. 1995, *Astronomy and Astrophysics*, 296, 665

Vikhlinin, A., Kravtsov, A., Forman, W., Jones, C., Markevitch, M., Murray, S. S., & Van Speybroeck, L. 2006, Astrophysical Journal, 640, 691

Vikhlinin, A., et al. 2009, Astrophysical Journal, 692, 1060

Visvanathan, N., & Sandage, A. 1977, Astrophysical Journal, 216, 214

Vogel, S., & Reimers, D. 1995, *Astronomy and Astrophysics*, 294, 377

Wampler, E. J., Williger, G. M., Baldwin, J. A., Carswell, R. F., Hazard, C., & McMahon, R. G. 1996, *Astronomy and Astrophysics*, 316, 33

Wang, J.-X., Malhotra, S., Rhoads, J. E., Zhang, H.-T., & Finkelstein, S. L. 2009, *Astrophysical Journal*, 706, 762

Wang, T., et al. 2016, *Astrophysical Journal*, 828, 56

Wang, W.-H., Cowie, L. L., & Barger, A. J. 2004, *Astrophysical Journal*, 613, 655

—. 2006, *Astrophysical Journal*, 647, 74

Watabe, Y., Risaliti, G., Salvati, M., Nardini, E., Sani, E., & Marconi, A. 2009, *Monthly Notices of the Royal Astronomical Society*, 396, L1

Webb, T. M. A., Lilly, S. J., Clements, D. L., Eales, S., Yun, M., Brodwin, M., Dunne, L., & Gear, W. K. 2003, *Astrophysical Journal*, 597, 680

Webb, T. M. A., et al. 2006, *Astrophysical Journal, Letters*, 636, L17

Weiner, B. J., et al. 2005, *Astrophysical Journal*, 620, 595

Weymann, R. J., Stern, D., Bunker, A., Spinrad, H., Chaffee, F. H., Thompson, R. I., & Storrie-Lombardi, L. J. 1998, *Astrophysical Journal, Letters*, 505, L95

Whitaker, K. E., et al. 2011, *Astrophysical Journal*, 735, 86

Willis, J. P., et al. 2013, Monthly Notices, 430, 134

Wirth, G. D., et al. 2004, *Astronomical Journal*, 127, 3121

Wolf, C., et al. 1999, *Astronomy and Astrophysics*, 343, 399

Wolfe, A. M., & Prochaska, J. X. 2000, *Astrophysical Journal*, 545, 591

Worseck, G., & Wisotzki, L. 2006, *Astronomy and Astrophysics*, 450, 495

Worseck, G., Wisotzki, L., & Selman, F. 2008, *Astronomy and Astrophysics*, 487, 539

Wright, E. L. 2006, *Publications of the ASP*, 118, 1711

Wuyts, S., Labbé, I., Schreiber, N. M. F., Franx, M., Rudnick, G., Brammer, G. B., & van Dokkum, P. G. 2008, *Astrophysical Journal*, 682, 985

Wuyts, S., van Dokkum, P. G., Franx, M., Förster Schreiber, N. M., Illingworth, G. D., Labbé, I., & Rudnick, G. 2009, *Astrophysical Journal*, 706, 885

Wuyts, S., et al. 2007, *Astrophysical Journal*, 655, 51

Wylezalek, D., et al. 2013, *Astrophysical Journal*, 769, 79

—. 2014, *Astrophysical Journal*, 786, 17

Xu, C., et al. 2007, *Astronomical Journal*, 134, 169

Xue, Y. Q., et al. 2010, *Astrophysical Journal*, 720, 368

Yamada, T., Matsuda, Y., Kousai, K., Hayashino, T., Morimoto, N., & Umemura, M. 2012, *Astrophysical Journal*, 751, 29

Yang, Y., Zabludoff, A., Jahnke, K., Eisenstein, D., Davé, R., Shectman, S. A., & Kelson, D. D. 2011, *Astrophysical Journal*, 735, 87

York, D. G., Yanny, B., Crotts, A., Carilli, C., Garrison, E., & Matheson, L. 1991, *Monthly Notices of the Royal Astronomical Society*, 250, 24

Yoshida, M., et al. 2006, *Astrophysical Journal*, 653, 988

Yoshikawa, T., et al. 2010, *Astrophysical Journal*, 718, 112

Yuan, T., et al. 2014, *Astrophysical Journal, Letters*, 795, L20

Zel'dovich, Y. B. 1970, *Astronomy and Astrophysics*, 5, 84

Zeldovich, Y. B., & Sunyaev, R. A. 1969, *Astrophysics and Space Science*, 4, 301

Zhao, D., Conselice, C. J., Aragón-Salamanca, A., Almaini, O., Hartley, W. G., Lani, C., Mortlock, A., & Old, L. 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 1393

Zheng, W., et al. 2004, *Astrophysical Journal, Supplement*, 155, 73

Zheng, Z.-Y., Malhotra, S., Rhoads, J. E., Finkelstein, S. L., Wang, J.-X., Jiang, C.-Y., & Cai, Z. 2016, *Astrophysical Journal, Supplement*, 226, 23

Zheng, Z.-Y., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 431, 3589

Zwicky, F. 1937, Astrophysical Journal, 86, 217