Doctoral Dissertation
博士論文

Early Stages of Galaxy Formation Probed by Optical Observations and the Machine Learning Technique (可視光観測と機械学習で探る初期の銀河形成)

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2019
令和元年 12 月博士 (理学) 申請

Department of Physics, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
物理学専攻

Takashi Kojima
小島 崇史

Copyright © 2020 by
All rights reserved

Ph.D. Thesis Committee:

Dr. Satoshi Yamamoto (University of Tokyo); Chair

—alphabetical order—

Dr. Masashi Chiba (Tohoku University)

Dr. Tadayuki Takahashi (Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo)

Dr. Hidenobu Yajima (University of Tsukuba)

Dr. Naoki Yoshida (Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo)

Thesis Advisor:

Dr. Masami Ouchi (Institute for Cosmic Ray Research, University of Tokyo; National Astronomical Observatory of Japan)

This thesis is based on the following papers:

PAPER-I: Takashi Kojima, Masami Ouchi, Michael Rauch, Yoshiaki Ono, Yuki Isobe, Seiji Fujimoto, Yuichi Harikane, Takuya Hashimoto, Masao Hayashi, Yutaka Komiyama, Haruka Kusakabe, Ji Hoon Kim, Chien-Hsiu Lee, Shiro Mukae, Tohru Nagao, Masato Onodera, Takatoshi Shibuya, Yuma Sugahara, Masayuki Umemura, Kiyoto Yabe (Submitted to Astrophysical Journal; under review.)
"Subaru/HSC Machine Learning Survey for Extremely Metal-Poor Galaxies. I. A New Successful Approach for Identifying a Galaxy with $M_* < 10^6 M_\odot$ and $0.021 Z_\odot$"

PAPER-II: Takashi Kojima, Masami Ouchi, Michael Rauch, Yoshiaki Ono, Yuki Isobe, Seiji Fujimoto, Yuichi Harikane, Takuya Hashimoto, Masao Hayashi, Yutaka Komiyama, Haruka Kusakabe, Ji Hoon Kim, Chien-Hsiu Lee, Shiro Mukae, Tohru Nagao, Masato Onodera, Takatoshi Shibuya, Yuma Sugahara, Masayuki Umemura, Kiyoto Yabe (To be Submitted to Astrophysical Journal.)
"Subaru/HSC Machine Learning Survey for Extremely Metal-Poor Galaxies. II."

ACKNOWLEDGEMENTS

First of all, I would like to express my best gratitude to my supervisor, Masami Ouchi, for his strong supports for my research activities. He has spent great amount of time for me even when he is extremely busy. His every advice makes me stronger step by step, as a person who struggles to discover something new. I have learnt from Masami that my research activities are supported by many people including persons who I have never met, and begun to consider how people think about every science research, what people really want to know, how I convey my ideas to others, and what I can do for communities as a member. I find that these ways of thinking are quite essential to create values in human communities. I would like to try to contribute to human communities by making most of what I have learnt from Masami, even after I move to another field from next April.

I would like to express my special gratitude to Yoshiaki Ono, Ken Mawatari, Yuichi Harikane, and Seiji Fujimoto. They have been always my supporters in my research activities. Every one of them spent me a lot of time even if they were very busy. I really appreciate their care and thoughtfulness when I had difficulties on my research. I discuss varieties of science topics with Yoshiaki, where he does not give me an answer, but give me questions that always lead to the correct answer. I experience how a good researcher thinks and reaches an answer through the discussion with him. Yuichi's and Seiji's research styles have also affected my research style. Yuichi and Seiji always make the best effort to maximize their research opportunity, which leads to very good papers attracting attention from researchers all over the world. Their active styles stimulate my desire to do research more and more. Ken is kind enough to get me out of my room to play tennis together, which let me switch my mind especially when I have a difficulty in my research activities.

I have also been supported by other collaborators, Kimihiko Nakajima, Michael Rauch, Takuya Hashimoto, Masao Hayashi, Yutaka Komiyama, Haruka Kusakabe, Ji Hoon Kim, Chien-Hsih Lee, Tohru Nagao, Masato Onodera, Masayuki Umemura, and Kiyoto Yabe. I really appriciate their contribution, which makes our papers much better in terms of English writing and scientific discussion. Especially, Michael has

kindly started collaboration with us and provided us with opportunity to conduct MagE and LDSS-3 spectroscopy for our EMPG candidates. Michael also invited me to Carnegie Observatories on December, 2018, and helped me with solving issues of our MagE data.

I would like to express my gratitude to former and present post-docs and alumni in our group: Tomoki Saito, Suraphong Yuma, Takatoshi Shibuya, Momose Rieko, Mariko Kubo, Florent Duval, Yi-Kuan Chang, Ken Mawatari, Jun Toshikawa, Akira Konno, Yoshiaki Naito, Masafumi Ishigaki, Hiroko Tamazawa, and Haibin Zhang. They always helped me when I was in trouble, and give me critical advices leading to correct answers.

I would like to thank all the students in our group, Ryo Higuchi, Hilmi Miftahul, Shotaro Kikuchihara, Ryo Itoh, Ryota Kakuma, Karin Shimodate, Yechi Zhang, Yuki Isobe, Nao Sakai, and Dongsheng Sun for everyday life, in which we chatted something funny, discussed science issues seriously, and hold parties in our institution.

I also thank all of my friends who I met in my Master and Ph.D. courses. Special thanks go to my colleagues, Shiro Mukae, Yuma Sugahara, Satoshi Kikuta, Yasunori Terao, Tadafumi Matsuno, Ryo Ando, and Tomohiro Inada. I would not forget every moment we spent together in the last five years. We visited many places in Japan or in the world together to attend a science conference or as a private trip. We not only chatted funny things, but also talked on science matters seriously, sometimes criticizing one another. They have been always my good rivals, who stimulate my scientific motivation every time.

I would like to express my special thanks to Daichi Kashino. I really enjoyed discussing every science topic with him. He also looks enjoyable every time we discuss, and such atmosphere reminds me of my scientific interests.

I really appreciate every kindness of Yabe Kiyoto, Rieko Momose, and Ryusei Yabe, always welcoming me to their home when I was facing difficulties or experienced something tough. I was very often heeled by their welcoming atmosphere and kindness, which kept my motivation to conduct tough research activities.

I thank many researches I have met in my career, Kentaro Aoki, Lennox Cowie, Mark Dijikstra, Richard Ellis, Dawn Erb, Andrea Ferrara, Tomotsugu Goto, Akio

Inoue, Koki Kakiichi, Daichi Kashino, Lisa Kewley, Taddy Kodama, Adam Leroy, Matthew Malkan, Crystal Martin, Daniel Masters, Takahiro Morishita, Kentaro Motohara, Tohru Nagao, Yuri Nishimura, Masato Onodera, Ryan Sanders, Alice Shapley, Rhythm Shimakawa, Kazuhiro Shimasaku, Daniel Stark, Charles Steidel, Koki Terao, and Claudia Lagos Urbina for their critical comments and powerful supports in my research activities. I would like to express my special thanks to Daniel Kelson and Kentaro Aoki. Daniel Kelson made great efforts in helping me reduce and calibrate our MagE data when I stayed Carnegie Observatories in 2018. The MagE data have become critical part of my thesis. Kentaro Aoki supported our FOCAS observations, in which one of the most important discoveries has been established by his great contribution.

I am grateful to the members of my thesis committee, Satoshi Yamamoto, Masashi Chiba, Tadayuki Takahashi, Hidenobu Yajima, and Naoki Yoshida for their critical, constructive comments to improve my thesis. I have been able to improve my research largely thanks to their professional comments.

Finally, I thank my family for their supports. I would like to thank my elder sisters, Mio Kojima and Aki Kojima, who usually make me feel relaxed and friendly. They provide me the place where I can always go back. I am grateful to my mother, Masako Kojima for her every support throughout my life. I have done everything I wanted to do until my Ph.D. course thanks to her supports. If it were no for her great supports, I would not be what I am today. I would like to dedicate this doctoral dissertation to my late father, Yukio Kojima, who has led me to the Ph.D. course and the new field from the next April.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. The

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

We also thank staffs of the Las Campanas observatories, the Subaru telescope, and the Keck observatories for helping us with our observations. The observations were carried out within the framework of Subaru-Keck time exchange program, where the travel expense was supported by the Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. I wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. I am most fortunate to have the opportunity to conduct observations from the great mountain, Maunakea.

This work is supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, as well as KAKENHI Grant-in-Aid for Scientific Research (A) (15H02064, 17H01110, and 17H01114) through Japan Society for the Promotion of Science (JSPS), as well as JSPS KAKENHI Grant Numbers, 18K13578, and 17K14257. This thesis is especially supported by JSPS KAKENHI Grant Numbers, 18J12840.

ABSTRACT

Hydrodynamical simulation predicts that cluster-like galaxies with low stellar masses ($\sim 10^4$–$10^6 M_\odot$), low metallicities ($\sim$0.1–1% $Z_\odot$), and high specific star-formation rates (sSFR $\sim$ 100 Gyr^{-1}) are formed at $z \gtrsim$ 10–20, and experiencing an early stage of the galaxy formation. Although cluster-like galaxies are important to understand the stellar population and star formation history in the early-stage galaxy formation, data used in previous observational studies are not deep enough to discover cluster-like galaxies.

We have initiated a new survey for local extremely metal-poor galaxies (EMPGs) with Subaru/Hyper Suprime-Cam (HSC) large-area ($\sim$ 500 deg^2) optical images reaching a 5σ limit of $\sim$ 26 magnitude, about 100 times deeper than the one of Sloan Digital Sky Survey (SDSS). To select $Z/Z_\odot < 0.1$ EMPGs from $\sim$ 40 million sources detected in the Subaru images, we first develop a machine-learning (ML) classifier based on a deep neural network algorithm with a training data set consisting of optical photometry of galaxy, star, and QSO models. We test our ML classifier with SDSS objects having spectroscopic metallicity measurements, and confirm that our ML classifier accomplishes 86%-completeness and 46%-purity EMPG classifications with photometric data. Applying our ML classifier to the photometric data of the Subaru sources as well as faint SDSS objects with no spectroscopic data, we obtain 27 and 86 EMPG candidates from the Subaru and SDSS photometric data, respectively.

We conduct optical follow-up spectroscopy for 10 out of our EMPG candidates with Magellan/LDSS-3+MagE, Keck/DEIMOS, and Subaru/FOCAS, and find that the 10 EMPG candidates are star-forming galaxies at $z = 0.007 - 0.03$ with large Hβ equivalent widths of 104–265 Å, stellar masses of $\log(M_\star/M_\odot)$=5.0–7.1, and high specific star-formation rates of $\sim$300 Gyr^{-1}, which are similar to those of early galaxies at $z \gtrsim 6$ reported recently. We spectroscopically confirm that 3 out of 10 candidates are truly EMPGs with $Z/Z_\odot < 0.1$, one of which is HSC J1631+4426, the most metal-poor galaxy with $Z/Z_\odot = 0.019$ so far identified ever.

The number density of our metal-poor galaxies is $\sim 10^{-4}$–10^{-5} Mpc^{-3}, suggesting that such metal-poor galaxies are very rare in the local universe. Our metal-poor

galaxies reside relatively isolated environment (i.e., nearest neighborhood distances $\sim$ 3.83 Mpc in median), which is confirmed by a Kolmogorov-Smirnov test ($p = 1.9 \times 10^{-3}$).

We find that α-element ratios of Ne/O and Ar/O show almost constant values of log(Ne/O)~-0.8 and log(Ne/O)~-2.4, respectively, while low N/O ratios, log(N/O)$\lesssim-1.4$, suggest the undergoing primary nucleosynthesis due to their low metallicity and young stellar population. Two EMPGs show Fe/O ratios 0.5–1.0 dex higher than Galactic stars and the Fe/O evolution model calculation at fixed metallicity, including the EMPG with the solar Fe/O ratio and 0.019 $(O/H)_\odot$. We conclude that the high Fe/O ratios are attributed to SNe of very massive stars above 300 $M_\odot$.

To probe ionizing radiation in our metal-poor galaxies, we inspect emission lines of Hβ, [O II]3727, [Ar III]4740, [O III]5007, [Ar IV]7136, and He II4686 lines, which are sensitive to ionizing photon in the range of 13.6–54.4 eV. We find that interstellar medium of our metal-poor galaxies is highly ionized characterized by strong [O III]5007, [Ar IV]7136, and He II4686 emission lines. High mass X-ray binary (HMXB) models explain He II4686 intensities of half of our metal-poor galaxies, including ones in very early phases of HMXB evolution ($\lesssim$5 Myr) before or in the middle of the first compact object formation. The other half of our metal-poor galaxies, including two EMPGs of $Z/Z_\odot < 0.1$, are not explained by either the HMXB models or the latest binary stellar synthesis models, where their high He II4686/Hβ ratios are not reproduced at their very young ages, $\lesssim$5 Myr suggested by $EW_0(H\beta)$>100 Å. To explain strong He II4686 emission from very young, metal-poor galaxies, we newly suggest possibilities of metal-poor AGNs and very massive stars beyond 300 $M_\odot$. Especially, EMPGs might be able to form very massive star beyond 300 $M_\odot$ due to their extremely metal-poor gas, which eventually form intermediate-mass black holes and thus HMXBs as early as $\sim$2 Myr. To verify our scenarios, self-consistent HMXB modeling is necessary with a higher maximum stellar mass cut above $\sim$300 $M_\odot$.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Theoretical Prediction

The early universe is dominated by a large number of young, low-mass, metal-poor galaxies. Theoretically, first galaxies are formed at $z \sim 10$–20 from gas already metal-enriched by Pop-III (i.e., metal free) stars. According to hydrodynamical simulation (e.g., Wise et al., 2012), first galaxies are created in dark matter (DM) mini halos with $\sim 10^8 M_\odot$ and have low stellar masses ($\log(M_\star/M_\odot)\sim$4–6), low metallicities ($Z \sim 0.1$–$1\% Z_\odot$), and high specific star-formation rates (sSFR ~ 100 Gyr^{-1}) at $z \sim 10$. The typical stellar mass is remarkably small, comparable to those of star clusters. Such cluster-like galaxies are undergoing an early stage of the galaxy formation. One of ultimate goals is to understand the early-stage galaxy formation by probing the cluster-like galaxies. The cluster-like galaxies are important in that they are building blocks of the galaxy formation hierarchy.

1.2 Observational Background

Even the current best telescopes do not reach sensitivities high enough to detect the very low-mass galaxies ($\log(M_\star/M_\odot)\sim$4–6) at $z \gtrsim 10$ described in Section 1.1. In the last decade, observational efforts have been made in three ways to reach as low-mass galaxies as possible. First, the gravitational lensing has been used to detect relatively low-mass galaxies at $z \sim 2$–7. Lensed, low-mass galaxies with $\log(M_\star/M_\odot)\sim$6–9 are spectroscopically confirmed at $z \sim 2$–3 (Christensen et al., 2012a,b; Stark et al., 2014; Vanzella et al., 2017) showing very high equivalent widths of [O III]+Hβ lines ($\gtrsim$1000 Å) and low metallicity ($\sim 0.1 Z_\odot$). At $z \sim 6$–7, Stark et al. (2015) and Mainali et al. (2017) have spectroscopically confirmed rest-frame ultra-violet (UV) faint galaxies with $M_{\rm UV}\sim -20$, which corresponds to $\log(M_\star/M_\odot)\sim$9 (Kikuchihara et al., 2019). The galaxies of Stark et al. (2015) and Mainali et al. (2017) also show strong emission lines, such as C III]λ1909, in the rest-frame UV spectra. Stellar

synthesis and photoionization models (Inoue, 2011) demonstrate that the rest-frame equivalent width of Hα line, EW$_0$(Hα), can reach $\sim$1,000–3,000 Å for stellar ages of $\lesssim$100 Myr, for example. Thus, these observations suggest that low-mass galaxies at $z \sim 2$–7 are characterized by strong emission lines with young stellar ages. However, the stellar mass ranges of the previous studies above ($\log(M_\star/M_\odot)\sim$6–9) are not as low as cluster-like galaxies ($\log(M_\star/M_\odot)\sim$4–6) undergoing an early-stage galaxy formation.

Second, the stacking analysis is another powerful tool to probe low-mass, faint galaxies. Harikane et al. (2018) have created stack images of a large number of Lyα emitters (LAEs) at $z \sim 5$–7 discovered in a deep, narrow-band survey of Hyper Suprime-Cam (HSC; Miyazaki et al., 2012, 2018; Komiyama et al., 2018; Kawanomoto et al., 2018; Furusawa et al., 2018) installed on the Subaru telescope. Harikane et al. (2018) have also stacked images of Spitzer/IRAC (3.5 and 4.5 μm), which show board band excesses caused by very large equivalent widths of [O III]λ5007, Hβ, and Hα emission lines. Harikane et al. (2018) have also estimated average metallicities of their faint LAEs from the [O III]λ5007, Hβ, and Hα flux ratios obtained from IRAC stack images, finding that an average metallicity of their faint LAEs is $\sim$0.03 $Z_\odot$. The sample of Harikane et al. (2018) reaches as faint as $M_{\rm UV}\sim -20$ corresponding to $\log(M_\star/M_\odot)\sim$9 (Kikuchihara et al., 2019). Even the combination of the deep imaging and stacking analysis cannot reach very low-mass, cluster-like galaxies ($\log(M_\star/M_\odot)\sim$4–6) described in Section 1.1.

Third, low-mass, metal-poor galaxies with strong emission lines have been discovered in the local universe (e.g., Cardamone et al., 2009). Such local galaxies are regarded as local analogs of high-z galaxies because they have low stellar masses, low metallicities, and strong emission line equivalent widths similar to the high-z galaxies. To understand the nature of low mass galaxies at high redshifts, local analogs have been investigated in the last decade. In this thesis, we also aim to understand the early-stage galaxy formation of high-z, cluster-like galaxies by probing local analogs. Below, we describe previous local-analog studies to highlight the background of the local analogs.

Metal-poor galaxies with a large EW$_0$([O III]λ5007) have been found by the broad-

band excess technique in the data of Sloan Digital Sky Survey (SDSS; York et al., 2000). For example, Cardamone et al. (2009) have reported metal-poor, actively star-forming galaxies at $z \sim 0.3$ selected from the SDSS data, which have been named "green pea galaxies" (GPs) after their compact size and intrinsically green color caused by the large EW$_0$([O III]λ5007) up to $\sim$1500 Å. Yang et al. (2017b) have also discovered metal-poor, highly star-forming galaxies at $z \sim 0.04$ in the SDSS data selected with the g-band excess with the very large EW$_0$([O III]λ5007)$\sim$500–2500 Å. The galaxies found by Yang et al. (2017b) have been nicknamed "blueberry galaxies" (BBs). Typical stellar mass ranges of GPs and BBs are $\log(M_\star/M_\odot)$=8–9 and 7–8, respectively.

Typical metallicities of these GPs/BBs are 12+log(O/H)=8.0$\pm$0.3, which fall into a *moderate* metallicity range compared to extremely metal-poor galaxies (EMPGs) such as J0811+4730 (Izotov et al., 2018b), SBS0335$-$052 (e.g., Izotov et al., 2009), AGC198691 (Hirschauer et al., 2016), J1234+3901 (Izotov et al., 2019b), LittleCub (Hsyu et al., 2017), DDO68 (Pustilnik et al., 2005; Annibali et al., 2019), IZw18 (e.g., Izotov & Thuan, 1998; Thuan & Izotov, 2005), and LeoP (Skillman et al., 2013) in the range of 12+log(O/H)$\sim$7.0–7.2. Stellar synthesis and photoionization models (Inoue, 2011) suggest that the EW$_0$([O III]λ5007) and an [O III]λ5007/Hα flux ratio take maximum values around 12+log(O/H) $\sim$ 8.0 as shown in Figure 1.1. Thus, galaxies selected with a single broadband excess such as GPs/BBs may be somewhat biased towards a large EW$_0$([O III]λ5007), i.e., a *moderate* metallicity of 12+log(O/H) $\sim$ 8.0. The models of Inoue (2011) demonstrate that the [O III]λ5007/Hα ratio monotonically decreases with decreasing metallicity in the range of 12+log(O/H) $<$ 8.0 (Figure 1.1) simply because the oxygen element becomes deficient. On the other hand, in the range of 12+log(O/H)$>$8.0, the [O III]λ5007/Hα ratio also decreases with increasing metallicity due to a low electron temperature caused by the efficient metal cooling and a low ionization state of inter-stellar medium (ISM), where O^{2+} ions become O^+ ions or neutral oxygen atoms. Indeed, as shown in Figure 1.1, representative metal-poor galaxies (e.g., Thuan & Izotov, 2005; Izotov et al., 2009; Skillman et al., 2013; Hirschauer et al., 2016; Izotov et al., 2018b, 12+log(O/H)$\sim$7.0–7.2) have a ratio of [O III]λ5007/Hα=0.4–1.0. The [O III]λ5007 line is no longer the strongest emission

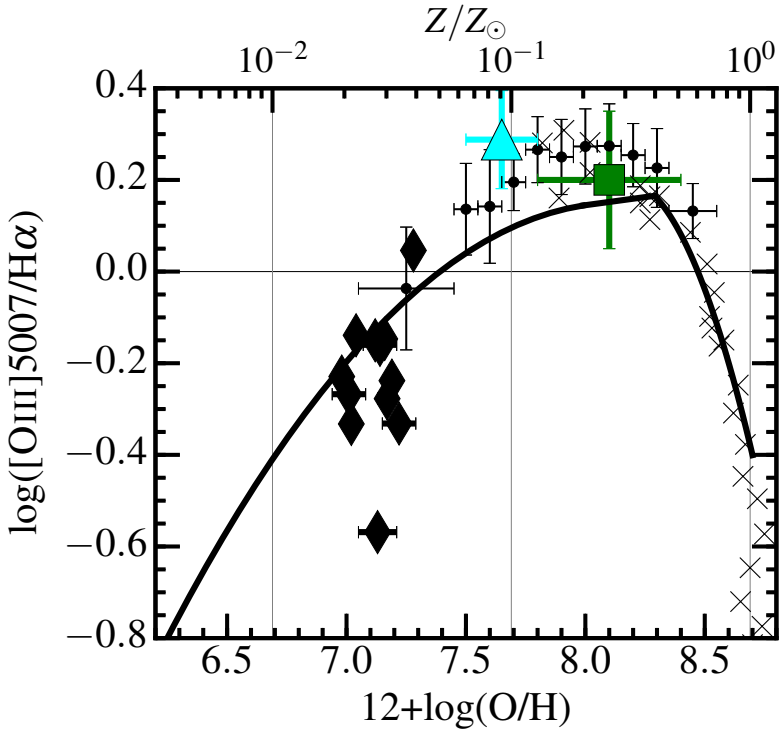

Figure 1.1 [OIII]/Hα ratio as a function of gas-phase metallicity. The cross marks (Andrews & Martini, 2013) and dots (Nagao et al., 2006) present the average of local star-forming galaxies. We also show the typical values of SDSS GPs (green square; Cardamone et al., 2009; Amorín et al., 2010) and BBs (cyan triangle; Yang et al., 2017b). The solid line is a theoretical prediction (Inoue, 2011). The diamonds are representative metal-poor galaxies, J0811+4730 (Izotov et al., 2018b), SBS0335−052 (e.g., Izotov et al., 2009), AGC198691 (Hirschauer et al., 2016), J1234+3901 (Izotov et al., 2019b), LittleCub (Hsyu et al., 2017), DDO68 (Pustilnik et al., 2005; Annibali et al., 2019), IZw18 (e.g., Izotov & Thuan, 1998; Thuan & Izotov, 2005), and LeoP (Skillman et al., 2013) in the range of 12+log(O/H)∼7.0–7.2. The GPs and BBs show *moderate* metallicities around 12+log(O/H)∼8.0, which correspond to a high [OIII]/Hα ratio of ∼2, while the representative EMPGs have relatively low [OIII]/Hα ratio of ∼0.3–1.0.

line in an optical spectrum of an EPMG as demonstrated in Figure 1.2. Thus, optical spectra of the EMPGs are characterized by multiple strong emission lines such as hydrogen Balmer lines and [O III]$\lambda\lambda$4959,5007. The strong emission lines of EMPGs cause g- and r-band excesses at $z \lesssim 0.03$.

Recently, Hsyu et al. (2018) and Senchyna & Stark (2019) have started metal-poor galaxy surveys with the SDSS data, where they have selected objects that show g- and r-band excesses. Unfortunately, the EMPGs have similar colors to those of other types of objects (i.e., blue stars, transient objects) on a classical color-color diagram of $r - i$ vs. $g - r$ (Hsyu et al., 2018; Senchyna & Stark, 2019).

1.3 Scientific Motivation

Element abundances are important to understand the stellar population and the star-formation history of galaxies in the early-phase star formation. Especially, iron and nitrogen are good tracers of the past star formation and the stellar population because these elements are produced and ejected by the different stellar population at different ages. Below we briefly describe processes of the iron and nitrogen enrichment, which are key to unveil the stellar population of cluster-like galaxies undergoing the early-stage galaxy formation.

First, iron elements are effectively produced and released into ISM by type-Ia supernovae (SNe) $\sim$1 Gyr after the star formation, which leads to an increasing iron-to-oxygen ratio (Fe/O) as a function of time. As reported in studies of Galactic bulge stars (Bensby & Feltzing, 2006) and Galactic thick disk stars (Lecureur et al., 2007; Bensby et al., 2013), the increasing Fe/O trend is only seen at a high stellar metallicity range of $Z_*/Z_\odot \gtrsim 0.2$ (corresponding to 12+log(O/H)$\gtrsim$8.0, see also Figure 10.10 for reference). Below 12+log(O/H)$\sim$8.0 (or $\lesssim$1 Gyr), the core-collapse SNe mainly contribute to the production and release of iron and oxygen. The core-collapse SNe do not release much iron elements compared to oxygen, because iron elements are produced at the core of a massive star and fall into a black hole (BH). In the iron-enrichment mechanism described above, low-mass, metal-poor, young SFGs are expected to have a low Fe/O ratio due to their young ages.

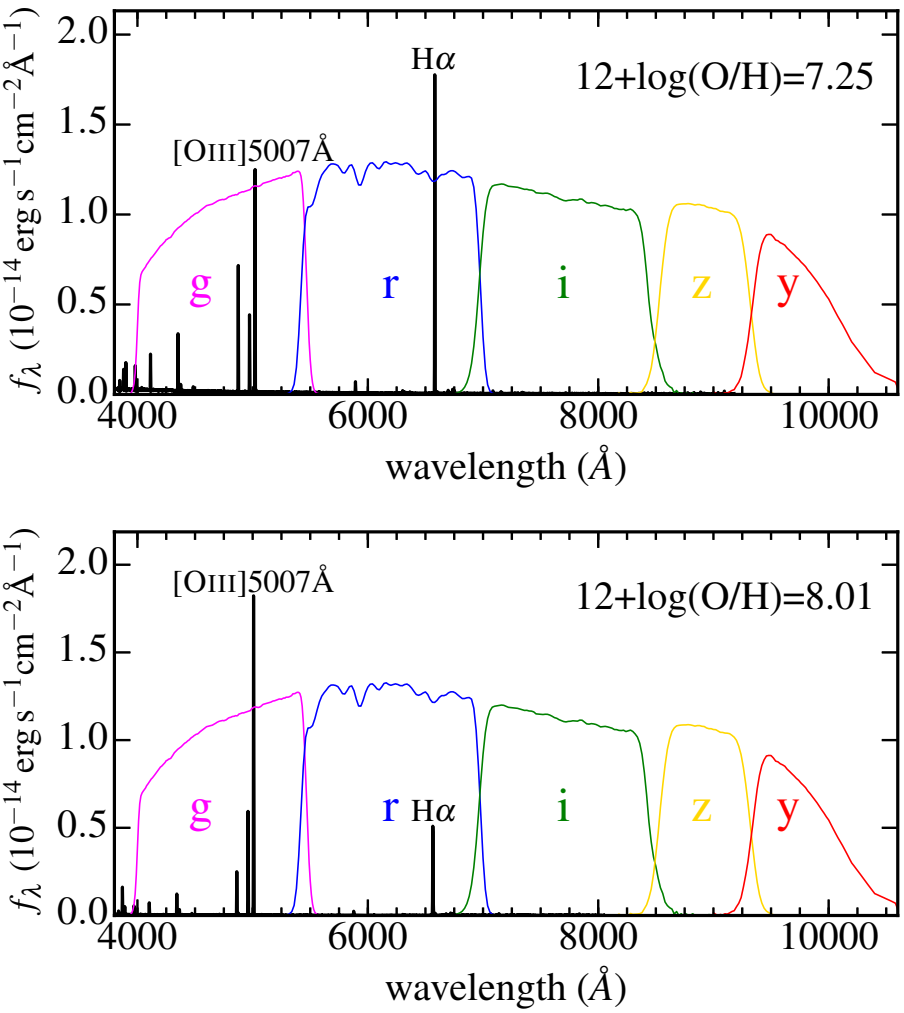

Figure 1.2 Top: Spectrum example of an EMPG with a very low metallicity of 12+log(O/H)=7.25 (Kniazev et al., 2003). Bottom: Same as top panel, but for a GP with a moderate metallicity of 12+log(O/H)=8.01 (Jaskot & Oey, 2013). We show the GP spectrum in the rest-frame for an easy comparison with the EMPG spectrum. The color curves are throughput curves of HSC *grizy*-band filters for reference. In the optical spectrum of this typical EMPG (top panel), Hα is the strongest line.

Second, nitrogen elements are tracers of activities of massive stars and low- and intermediate-mass stars. As suggested by previous studies (Pérez-Montero & Contini, 2009; Pérez-Montero et al., 2013; Andrews & Martini, 2013), nitrogen-to-oxygen ratios (N/O) of SFGs present a plateau ($\log(\mathrm{N/O})\sim-1.5$) in the range of $12+\log(\mathrm{O/H})\lesssim8.0$ and a positive slope at higher metallicities as a function of metallicity (see also the panel (c) of Figure 10.9 for reference). Model calculation of the N/O evolution (e.g., Vincenzo et al., 2016) also supports this trend, showing the plateau and positive slope of N/O. The plateau is basically resulted from the primary nucleosynthesis of massive stars, while the positive slope is mainly attributed to the secondary nucleosynthesis of low- and intermediate-mass stars (e.g., Vincenzo et al., 2016). At low metallicities, nitrogen elements are mainly produced after the formation of a heavy-element core (e.g., O and C) in a massive star. The produced nitrogen is ejected into ISM by SNe for stars more massive than $\sim8\ M_{\odot}$. At high metallicities, nitrogen is produced by the CNO-cycle in low- and intermediate-mass stars, and ejected through stellar winds during the asymptotic giant branch (AGB) phase, $\sim$1 Gyr after the star formation. In the nitrogen-enrichment mechanism, low-mass, metal-poor, young SFGs have a low N/O ratio because of their low metallicities and young ages.

Ionizing radiation is another key to understand the stellar population of galaxies in the early-phase star formation. Ionizing radiation is contributed by massive stars and/or a hot accretion disk around compact objects such as BHs. Observational studies (López-Sánchez & Esteban, 2008; Shirazi & Brinchmann, 2012; Senchyna et al., 2017) have suggested that SFGs show strong He II4686 emission lines in the range of $12+\log(\mathrm{O/H})<8.0$. The He II4686 line is a recombination line resulted from the He^{+} ionization (Equation 1.1), the He^{2+} recombination (Equation 1.2), and the subsequent cascade emission process (Equation 1.3).

$$\mathrm{He}^{+} + \gamma \longrightarrow \mathrm{He}^{2+} + \mathrm{e}^{-} \tag{1.1}$$

$$\mathrm{He}^{2+} + \mathrm{e}^{-} \longrightarrow (\mathrm{He}^{+})^{*} + \gamma \tag{1.2}$$

$$(\mathrm{He}^{+})^{*} \longrightarrow \mathrm{He}^{+} + \gamma \tag{1.3}$$

The He II4686 line is sensitive to ionizing photons above 54.4 eV because the ionization

potential of He^+ is 54.4 eV. Such high-energy ionizing photons are not abundant in radiation of O- and B-type hot stars. Thus, under the stellar radiation, the He II4686 emission line is very weak compared to an Hβ line (i.e, low He II4686/Hβ ratio). The typical He II4686/Hβ ratio is $\sim$1/100 for local SFGs above 12+log(O/H)=8.0 (see also Figure 10.13 for reference), while the He II/Hβ ratio increases as metallicity decreases at 12+log(O/H)<8.0 (López-Sánchez & Esteban, 2008; Shirazi & Brinchmann, 2012; Senchyna et al., 2017). Xiao et al. (2018) have created nebular emission models with the combination of the photoionization code CLOUDY and the BPASS (Binary Population and Spectral Synthesis) code (Stanway et al., 2016; Eldridge et al., 2017). Schaerer et al. (2019) demonstrate that the models of Xiao et al. (2018) predict He II4686/H$\beta\sim$1/1000, which is well below the observed He II4686/Hβ ratios ($\gtrsim$1/100). This means that the main contributors of He II4686 are not hot stars. Schaerer et al. (2019) have estimated He II4686/Hβ ratios with high mass X-ray binary (HMXB) models of Fragos et al. (2013a) and Fragos et al. (2013b), and suggested that high He II4686/Hβ ratios can be partly explained by the HMXB models. However, the HMXB models still do not explain the high He II4686/Hβ ratios for galaxies with a high Hβ equivalent width, $EW_0(H\beta)$>100 Å (i.e., younger than 5 Myr). Schaerer et al. (2019) suggest other possible contribution of old stellar population and/or shock-heated gas. The main contributor of the He II4686 emission is still under debate.

1.4 Scope of this Thesis

In this study, we target EMPGs with strong emission lines in the local universe ($z \lesssim 0.03$), which may be a local analog of a high-z star-forming galaxy (Section 1.2). Because such galaxies are intrinsically faint and rare in the local universe, wide-field, deep imaging data are necessary. However, the SDSS spectroscopy data used in the previous studies are not deep enough to discover EMPGs with $\log(M_\star/M_\odot)$=4–6, which may be cluster-like galaxies experiencing the early-stage galaxy formation described in Section 1.1. To discover very faint EMPGs that the previous SDSS surveys could not find out, deeper, wide-field imaging data have been expected. In March

2014, the Subaru telescope has started a large-area ($\sim$1,400 deg^2), deep survey with Subaru/HSC, called HSC Subaru Strategic Program (HSC SSP; Aihara et al., 2018). The goal of our research is to discover faint EMPGs by exploiting the Subaru HSC-SSP data, whose i-band limiting magnitude ($i_{\rm lim}\sim$26 mag) is $\sim$5 mag deeper than the one of the SDSS data ($i_{\rm lim}\sim$21 mag). We also use the SDSS data to complement brighter EMPGs in this study.

This thesis also aim to establish a selection more efficient than the classical color-color diagram explained in Section 1.2. We establish a more efficient selection with the Machine Learning (ML) technique.

Our scientific goals are basically element abundances such as iron and nitrogen and hard ionizing radiation probed by the He II4686 line described in Section 1.3. To achieve these goals, we conduct deep spectroscopy to detect (or strongly constrain) very faint emission lines of [O III]4363, [Fe III]4658, [N II]6584, and He II4686. The [O III]4363, [Fe III]4658, and [N II]6584 lines are necessary to estimate the metallicity, Fe/O and N/O, respectively. The He II4686 is needed to probe high energy ionizing photons above 54.4 eV.

The outline of this thesis is as follows. Chapter 2 describes theoretical backgrounds of this thesis, mainly focusing on nebular physics. In Chapter 3, telescopes and instruments used in this study are explained. In Chapter 4, we explain the Subaru HSC-SSP data as well as how we construct a source catalog from the HSC-SSP and SDSS data. Chapter 5 explains our new selection technique that we develop with the ML and shows results of a test of our ML selection. Chapter 6 explains the EMPG selection from the source catalogs. In Chapter 7, we describe our optical spectroscopy carried out for our EMPG candidates. Chapter 8 explains reduction and calibration processes of our spectroscopy data. In Chapter 9, we measure emission line fluxes and estimate galaxy properties such as stellar mass, star-formation rate, and metallicity. Chapter 10 shows results of our spectroscopy and discusses our metal-poor galaxies in terms of number density, environment, galaxy-scale properties such as $M_\star$ and SFR, element abundance ratios, and ionizing radiation. We describe conclusions of this

thesis in Chapter 11. Then Chapter 12 summarizes this thesis.

Throughout this thesis, magnitudes are on the AB system (Oke & Gunn, 1983). We adopt the following cosmological parameters, $(h, \Omega_m, \Omega_\Lambda) = (0.7, 0.3, 0.7)$. The definition of the solar metallicity is given by 12+log(O/H)=8.69 (Asplund et al., 2009). We also define an EMPG as a galaxy with 12+log(O/H)<7.69 (i.e., $Z/Z_\odot$<0.1) in this thesis, which is almost the same as in previous metal-poor galaxy studies (e.g., Kunth & Östlin, 2000; Izotov et al., 2012; Guseva et al., 2017).

CHAPTER 2
THEORETICAL BACKGROUNDS

2.1 Nebular Physics in Star-forming Galaxies

In star-forming regions of a galaxy, the diffuse ISM gas is ionized by the UV radiation from young hot stars. The region of the ionized ISM gas is called a nebula or an H II region. In theoretical models, the nebula is described by two kinds of equilibrium states.

One is the photoionization equilibrium, which is fixed by the balance between the photoionization and recombination of electrons with ions. In an H II region, hydrogen recombination lines (e.g., Lyman, Balmer, and Paschen series emission lines) are emitted during the recombination and the following cascading process, where electrons transit from high- to low-energy levels by the downward radiation. Among these hydrogen recombination lines, Balmer lines such as Hα, Hβ, Hγ, and Hδ are prominent in the rest-frame optical spectrum.

The other is the thermal equilibrium. In the nebula, ionized ISM gas is heated by the photoionization of the UV radiation from stars. The heated gas is cooled by the hydrogen recombination lines and metal lines from the H II region. The cooling rate of the emission is balanced with the heating rate of the photoionization. The most dominant cooling lines are forbidden and semi-forbidden lines from heavy element ions, such as [O II]λ3727, O III]$\lambda\lambda$1661,1666, [O III]λ4363, [O III]$\lambda\lambda$4959,5007, [N II]$\lambda\lambda$6548,6584, and [S II]$\lambda\lambda$6719,6731, which are prominent in the rest-frame optical spectrum. The forbidden and semi-forbidden lines cool the nebular efficiently because photons of these lines escape from the nebula easily due to the small cross section of these photons with ions. The collisions of free electrons are needed to produce excited ions for the emission of forbidden and semi-forbidden lines. However, the process of the collisional de-excitation also works, and reduces the number of ions at the excitation states. Thus, a low electron density environment is necessary to emit the forbidden and semi-forbidden lines. The forbidden and semi-forbidden lines are thus called collisional excitation lines. A key quantity here is a critical electron

density, which is defined by equal rates of the spontaneous emission and collisional de-excitation.

It should be noted that UV and optical emission lines are affected by the dust absorption in ISM.

2.2 Nebular Model

In this section, we describe nebular modeling.

2.2.1 Strömgren sphere

We introduce a simple model where an analytical solution can be obtained. In this model, we place a point radiation source at the center and uniform hydrogen gas around the radiation source. The hydrogen gas is ionized by ionizing photons from the radiation source, and forms a spherical H II region called Strömgren sphere (Strömgren, 1939). Figure 2.1 presents a schematic illustration of this model.

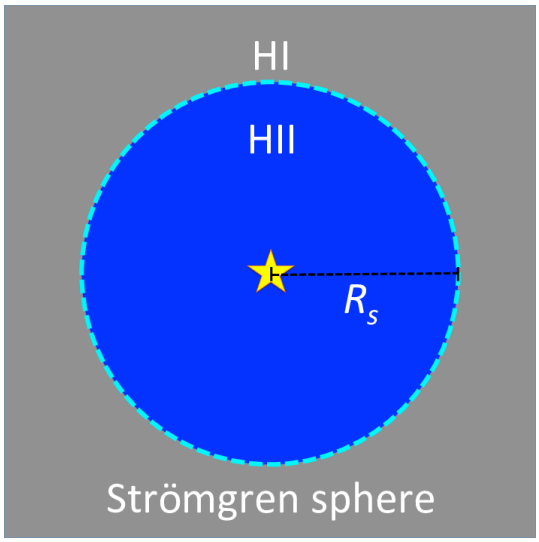

Figure 2.1 Schematic illustration of Strömgren sphere. The star mark indicates a point radiation source, while the blue and gray regions present the H II and H I regions, respectively. We mark the surface of the Strömgren sphere with a cyan dashed line, and the R_s stands for the Strömgren radius.

We assume moderately thick hydrogen gas where photons above 13.6 eV emitted during recombination are immediately reabsorbed. This assumption is called the case B recombination, which is introduced by Baker et al. (1938). A nebula usually has neutral hydrogen densities large enough so that recombination lines to level $n = 1$ have a very small probability of escaping from the nebula (Draine, 2010).

When this model reaches the photoionization equilibrium, the ionizing-photon production rate balances the recombination rate in the Strömgren sphere,

$$Q_0 = \frac{4}{3}\pi R_s^3 n_{\rm H}^2 \alpha_B \epsilon, \tag{2.1}$$

where we define the Strömgren sphere radius R_s, the hydrogen atom number density $n_{\rm H}$, the hydrogen ionizing photon production rate Q_0, the case B recombination rate α_B, and the volume filling factor ϵ. Here we introduce the ionization parameter, $q_{\rm ion}$,

$$q_{\rm ion} = \frac{Q_0}{4\pi R_s^2 n_{\rm H}}, \tag{2.2}$$

which is a ratio of ionizing photon flux at the Strömgren radius to hydrogen atom number density. Canceling out R_s in Equations (2.1) and (2.2), we obtain

$$q_{\rm ion} \propto Q_0^{1/3} n_{\rm H}^{1/3} \epsilon^{2/3}. \tag{2.3}$$

Although the sphere describes the analytical solution of homogeneous H II regions solely made of hydrogen under the photoionization equilibrium, real H II regions include various atoms of metal. We thus need more realistic models to describe a nebula.

2.2.2 Photoionization model

Another approach of nebular modeling is a numerical calculation. Using the numerical calculation package, Cloudy (Ferland et al., 1998, 2013), we construct photoionization models more realistic than the Strömgren sphere model.

In the Cloudy models, one needs to give initial conditions such as the geometrical

structure of gas and ionizing-photon source, radiation intensity and hardness, gas density, and element abundance. Based on these initial conditions, Cloudy iteratively performs a calculation until the system reaches the photoionization and thermal equilibriums. Emission line fluxes and physical properties (e.g., electron temperature and ion abundance) are derived from these calculations. Figure 2.2 presents an example of an emission line diagnostic diagram of [O III]/Hβ and [N II]/Hα calculated with Cloudy. The diagnostic diagram is called Baldwin-Phillips-Terlevich diagram (BPT diagram, Baldwin et al., 1981), which distinguishes main radiation sources of stellar or AGN radiation. The calculation above is conducted under the initial conditions of $n_e = 25\mathrm{cm}^{-3}$, 12+log(O/H)=(7.69, 7.99, 8.29, 8.59, 8.69, 8.99), and $\log q_{\mathrm{ion}}$=(7.0, 7.2, 7.4, 7.6, 7.8, 8.0, 8.2, 8.4, 8.6, 8.8, 9.0). Figure 2.2 also shows that the models (red lines) cover the observed trend of SFGs on the BPT diagram. One can also construct photoionization models under the assumption of power-low radiation from an AGN (e.g., Kewley et al., 2013). As explained here, the photoionization model is a powerful tool to constrain UV radiation fields and to identify ionization photon sources such as massive stars and AGNs.

2.3 Physical Properties of Inter-stellar Medium

Emission line fluxes depend on ISM physical properties such as gas density, electron temperature, metallicity, element abundance ratios, radiation intensity, and radiation hardness. We can estimate these ISM physical properties by utilizing these dependences. In this section, we explain how the electron temperature, metallicity, element abundance ratios, and ionization parameter are estimated from emission line fluxes with modeling. Note again that a metallicity is defined by a gas-phase oxygen-to-hydrogen ratio in the form of 12+log(O/H), as we define in Chapter 1. When we call an element abundance ratio, we indicate an element abundance ratio except for oxygen-to-hydrogen ratio in this thesis.

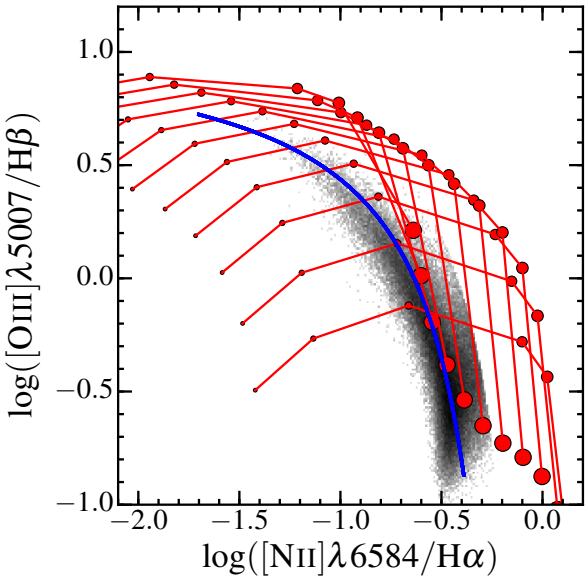

Figure 2.2 The grids of the [O III]/Hβ vs. [N II]/Hα line ratios calculated by our Cloudy models. The SDSS star-forming galaxies are shown as the gray histogram, and the blue solid line is the best-fit sequence of the SDSS galaxies (Kewley & Ellison, 2008). Colored circles present the model grid points, where the size of the circle indicates the gas-phase metallicity, from 12+log(O/H)=7.69 (smallest circles) to 12+log(O/H)=8.99 (largest circles). Grid points of the constant ionization parameter are connected with red solid lines.

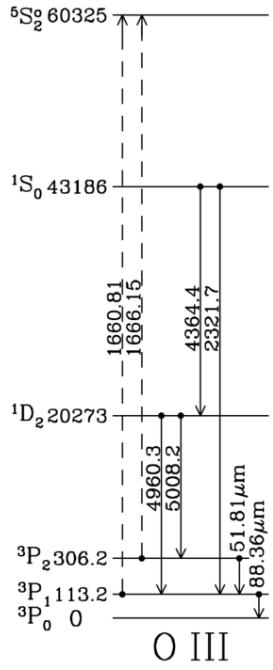

Figure 2.3 Energy levels of O^{2+} (OIII) ion and radiative transitions between energy levels (Draine, 2010). Note that emission line wavelengths are given in vacuum here although we use wavelengths in air in the text.

2.3.1 Electron temperature

The electron temperature is estimated with two collisional excitation lines of the same ion (e.g., O^{2+}), because the collisional excitation rate is determined by the electron temperature. As an example, we explain the common method of the electron temperature estimation with an [O III]λ4363/[O III]λ5007 ratio. Figure 2.3 shows energy levels of an O^{2+} ion. The [O III]λ4363 and [O III]λ5007 lines are emitted by the downward transition of $^1S_0 \rightarrow ^1D_2$ and $^1D_2 \rightarrow ^3P_2$, respectively. The critical density is 6.4×10^5 cm^{-3} for the $^1S_0$ level, and 2.8×10^7 cm^{-3} for the $^1D_2$ level. When an electron density is much lower than the critical density, the [O III]λ4363/[O III]λ5007 ratio only depends on electron temperature. Thus, an electron temperature is estimated with this [O III]λ4363/[O III]λ5007 ratio. In Figure 2.4, we present the relation between [O III]λ4363/[O III]λ5007 ratio and electron temperature of O^{2+} ions, T_e(OIII), which is calculated by a five-level model (Shaw & Dufour, 1995).

Similarly, we are able to estimate an electron temperature, using an emission line ratio of O III]$\lambda\lambda$(1661+1666)/[O III]λ5007 (Figure 2.4). The emission lines of O III]$\lambda\lambda$1661 and 1666 correspond to transitions of $^5S_2^o \rightarrow ^3P_1$ and $^5S_2^o \rightarrow ^3P_2$, respectively. The critical density of $^5S_2^o$ level is 3.4×10^{10} cm^{-3}, which is significantly higher than that of typical H II regions.

We are also able to estimate electron temperature of O^+ and N^+ ions with line ratios of [O II](7320+7330)/(3727+3729) and [N II]5755/6584, respectively. Hereafter, the electron temperatures derived from O^{2+}, O^+, and N^+ emission lines are denoted as T_e(OIII), T_e(OII), and T_e(NII), respectively.

2.3.2 Metallicity and Element abundance ratios

We explain how metallicities and element abundance ratios are estimated. Under the thermal equilibrium, an ion abundance ratio is described as a function of line flux ratio, electron temperature, and electron density (Aller, 1984). Because a typical nebula has an electron density lower than $\lesssim 10^3$ cm^{-3}, we can ignore the electron-density dependence of ion abundances. Izotov et al. (2006) has improved the relations

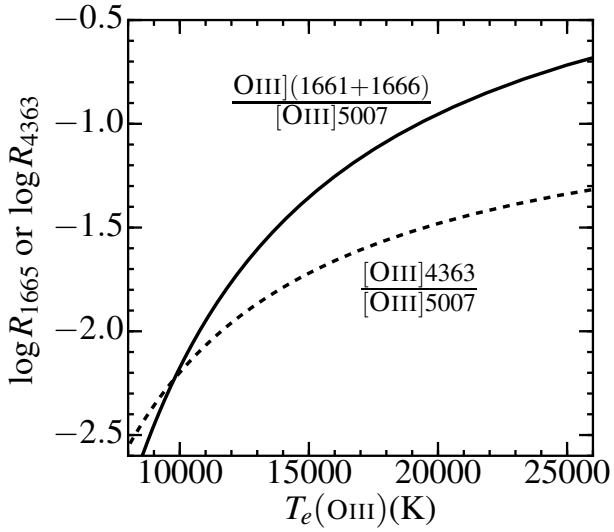

Figure 2.4 Flux ratios of $R_{1665} \equiv$ O III] $\lambda\lambda$1661,1666/[O III] λ5007 (solid line) and $R_{4363} \equiv$ [O III] λ4363/[O III] λ5007 (dashed line) as a function of electron temperature T_e(OIII). These R_{1665} and R_{4363} values are obtained by the *nebular.temden* routine of IRAF/STSDAS under the assumption of the low electron density limit.

of Aller (1984) with the latest atomic data, and derived relations[1],

$$12 + \log\left(\frac{\mathrm{O}^+}{\mathrm{H}^+}\right) = \log\left(\frac{[\mathrm{O\,II}]\,\lambda3727}{\mathrm{H}\beta}\right) + 5.961 + \frac{1.676}{t_2} - 0.40\log t_2 - 0.034 t_2, \quad (2.4)$$

$$12 + \log\left(\frac{\mathrm{O}^{2+}}{\mathrm{H}^+}\right) = \log\left(\frac{[\mathrm{O\,III}]\,\lambda\lambda4959+5007}{\mathrm{H}\beta}\right) + 6.200 + \frac{1.251}{t_3} - 0.55\log t_3 - 0.014 t_3, \quad (2.5)$$

$$12 + \log\left(\frac{\mathrm{N}^+}{\mathrm{H}^+}\right) = \log\left(\frac{[\mathrm{N\,II}]\,\lambda\lambda6548+6584}{\mathrm{H}\beta}\right) + 6.234 + \frac{0.950}{t_2'} - 0.42\log t_2' - 0.027 t_2', \quad (2.6)$$

where $t_2 = 10^{-4} T_e$(OII), $t_3 = 10^{-4} T_e$(OIII), and $t_2' = 10^{-4} T_e$(NII).

Equations (2.4) and (2.5) give metallicity defined by an oxygen-to-hydrogen abundance ratio (O/H) under the assumptions of

$$\frac{\mathrm{O}}{\mathrm{H}} = \frac{\mathrm{O}^+ + \mathrm{O}^{2+}}{\mathrm{H}^+} \quad (2.7)$$

[1]We omit negligibly small terms of n_e.

and

$$t_2 = 0.7t_3 + 0.3 \tag{2.8}$$

(Campbell et al., 1986; Garnett, 1992). Here we ignore contributions of triply-ionized and more highly ionized oxygen because these highly ionized oxygen with ionization potential of $\gtrsim$ 55eV cannot be generated by stellar UV radiation. Moreover, the assumption of Equation (2.8) is supported by photoionization models obtained by Stasińska (1982)

A nitrogen-to-oxygen abundance ratio (N/O) is obtained from Equations (2.4) and (2.6) under the assumptions of

$$\frac{\mathrm{N}}{\mathrm{O}} = \frac{\mathrm{N}^+}{\mathrm{O}^+} \tag{2.9}$$

and

$$t_2' = t_2 = 0.7t_3 + 0.3 \tag{2.10}$$

(Campbell et al., 1986; Garnett, 1992). We can assume the Equation (2.9) because nitrogen and oxygen ions have similar ionization potentials. The assumption of Equation (2.10) is also supported by the photoionization models obtained by Stasińska (1982). Other element abundance ratios are estimated similarly to the N/O ratio.

In this section, we have shown some examples of how to estimate metallicities and abundance ratios. Recent studies (i.g., Pérez-Montero, 2014; Pérez-Montero & Amorín, 2017; Esteban et al., 2014; Onodera et al., 2016; Micheva et al., 2019; Annibali et al., 2019) obtain electron temperatures and element abundances with nebular physics calculation codes of PyNeb (Luridiana et al., 2015). The PyNeb codes calculate emission line intensities by solving thermal and ionization equilibrium equations for an n-level atom. The latest atomic data are used in the PyNeb calculation. We use the PyNeb codes to estimate electron temperatures and element abundances, as we explain in Section 9.2.

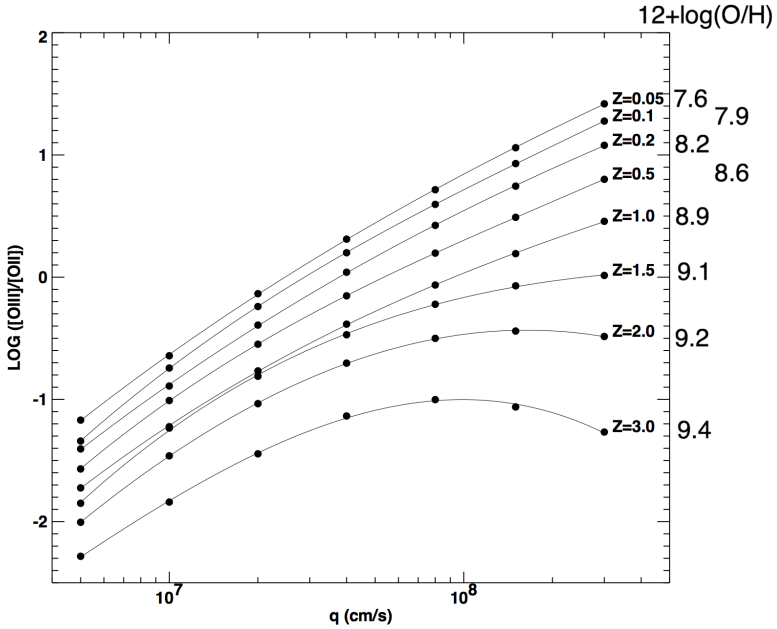

Figure 2.5 Ionization parameter vs. [O III]λ5007/[O II]λ3727 relation given by Kewley & Dopita (2002). Curves present relations obtained for 12+log(O/H)=(7.6, 7.9, 8.2, 8.9, 9.1, 9.2, 9.4). This figure is cited from Kojima et al. (2017).

2.3.3 Ionization parameter

The ionization parameter (Equation (2.2)) is estimated from an observed emission line ratio of [O III]λ5007/[O II]λ3727, which is sensitive to the ionization state of nebular gas. However, the [O III]/[O II] ratio depends not only on ionization parameter, but also on metallicity, electron density, and spectral hardness. Thus, the ionization parameter needs to be estimated by the numerical calculation that includes all of these physical parameters. For example, Kojima et al. (2017) use calculation results obtained by Kewley & Dopita (2002) to estimate ionization parameters. Figure 2.5 presents a relation between the [O III]/[O II] ratio, ionization parameter, and metallicity given in Kewley & Dopita (2002).

2.4 Dust Extinction

All of the observed emission lines are affected by the dust extinction as mentioned in Section 2.1. To estimate an intrinsic line ratio, we need to correct an observed line ratio with the effect of dust extinction. The dust extinction correction is performed with the Balmer line ratios under the assumptions of dust extinction curves of Calzetti et al. (2000), Cardelli et al. (1989), and Gordon et al. (2003, Small Magellanic Cloud). Below we give an example of the dust correction with the Calzetti et al. (2000) extinction curve.

Let $F_{\rm obs}(\lambda)$ and $F_{\rm int}(\lambda)$ be observed and intrinsic fluxes at the wavelength λ, respectively. The extinction $A(\lambda)$ is defined by

$$A(\lambda) = 2.5 \log_{10} \left(\frac{F_{\rm obs}(\lambda)}{F_{\rm int}(\lambda)} \right). \tag{2.11}$$

The normalization of $A(\lambda)$ with a color excess $E(B-V) \equiv A(B) - A(V)$ defines

$$k(\lambda) = A(\lambda)/E(B-V). \tag{2.12}$$

From Equations (2.11) and (2.12), we derive the form,

$$F_{\rm obs}(\lambda) = F_{\rm int}(\lambda) 10^{-0.4E(B-V)k(\lambda)}. \tag{2.13}$$

Calzetti et al. (1994) estimate the dust extinction curve, using the stellar continuum in starburst galaxies. This dust extinction curve is improved by Calzetti et al. (2000). Calzetti et al. (2000) obtain the wavelength dependence of dust extinction in the form of

$$\begin{aligned} k(\lambda) &= 2.659(-2.156 + 1.509/\lambda - 0.198/\lambda^2 + 0.011/\lambda^3) + 4.05 \\ &\qquad\qquad (0.12\mu{\rm m} \leq \lambda \leq 0.63\mu{\rm m}) \\ &= 2.659(-1.857 + 1.040/\lambda) + 4.05. \\ &\qquad\qquad (0.63\mu{\rm m} \leq \lambda \leq 2.20\mu{\rm m}) \end{aligned} \tag{2.14}$$

Figure 2.6 exhibits the dust extinction curves of Calzetti et al. (2000), Cardelli et al. (1989), and Gordon et al. (2003) in the cases of $E(B-V) = 0.0, 0.05, 0.1, 0.2$, and 0.4.

Under the assumption of the case B recombination (Section 2.2.1) and 10,000 K electron temperature, the intrinsic Balmer line ratios are

$$\left(\frac{{\rm H}\alpha}{{\rm H}\beta}\right)_{\rm int} = 2.86, \tag{2.15}$$

$$\left(\frac{{\rm H}\gamma}{{\rm H}\beta}\right)_{\rm int} = 0.468, \tag{2.16}$$

$$\left(\frac{{\rm H}\delta}{{\rm H}\beta}\right)_{\rm int} = 0.259, \tag{2.17}$$

for example (Osterbrock & Michael Shull, 1989). The 10,000 K is a typical value of electron temperature because the electron temperature of nebular gas is $5,000$–$25,000$ K. These Balmer line ratios of Equations (2.15)–(2.17) do not change by more than 10% in the range of $5,000$–$25,000$ K.

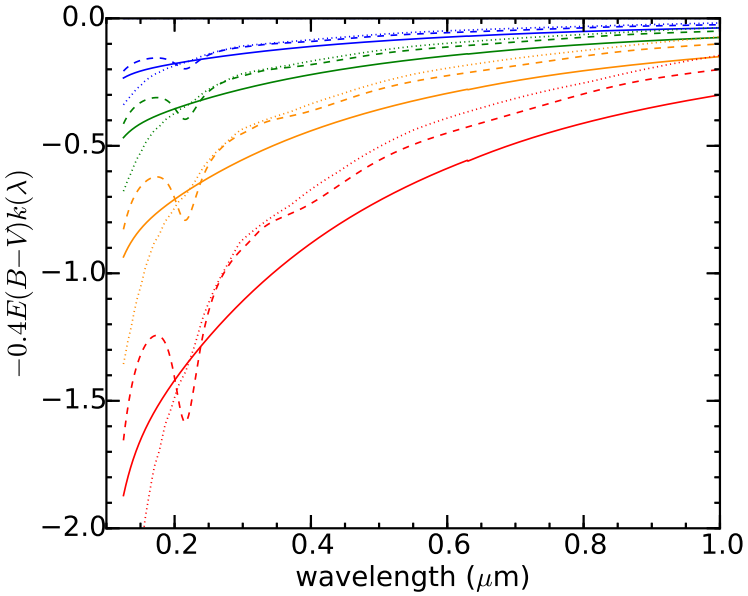

Figure 2.6 Dust extinction curves with $E(B-V)$=0.0 (dark blue), 0.05 (blue), 0.1 (green), 0.2 (orange), and 0.4 (red). Solid, dashed, and dotted lines are dust extinction curves obtained by Calzetti et al. (2000), Cardelli et al. (1989), and Gordon et al. (2003), respectively.

We obtain observed Balmer line ratios with

$$\left(\frac{\mathrm{H}\alpha}{\mathrm{H}\beta}\right)_{\mathrm{obs}} = \left(\frac{\mathrm{H}\alpha}{\mathrm{H}\beta}\right)_{\mathrm{int}} \times 10^{-0.4E(B-V)[k(\mathrm{H}\alpha)-k(\mathrm{H}\beta)]}, \tag{2.18}$$

$$\left(\frac{\mathrm{H}\gamma}{\mathrm{H}\beta}\right)_{\mathrm{obs}} = \left(\frac{\mathrm{H}\gamma}{\mathrm{H}\beta}\right)_{\mathrm{int}} \times 10^{-0.4E(B-V)[k(\mathrm{H}\gamma)-k(\mathrm{H}\beta)]}, \tag{2.19}$$

$$\left(\frac{\mathrm{H}\delta}{\mathrm{H}\beta}\right)_{\mathrm{obs}} = \left(\frac{\mathrm{H}\delta}{\mathrm{H}\beta}\right)_{\mathrm{int}} \times 10^{-0.4E(B-V)[k(\mathrm{H}\delta)-k(\mathrm{H}\beta)]}. \tag{2.20}$$

Each of Equations (2.18)–(2.20) is obtained from Equation (2.13) for two hydrogen Balmer lines. A color excess $E(B-V)$ is estimated with Equations (2.18)–(2.20), where $k(\mathrm{H}\alpha) = 3.33$, $k(\mathrm{H}\beta) = 4.60$, $k(\mathrm{H}\gamma) = 5.12$, and $k(\mathrm{H}\delta) = 5.39$ are calculated with Equation (2.14). The other Balmer lines of higher orders are also applicable in the dust extinction estimation. Once a color excess $E(B-V)$ is estimated, we can perform the dust correction with Equations (2.13) and (2.14) and obtain intrinsic emission line fluxes.

We have shown an example of the Calzetti et al. (2000) dust extinction curve, which is chosen from the dust extinction curves of Calzetti et al. (2000), Cardelli et al. (1989), and Gordon et al. (2003, Small Magellanic Cloud) described at the beginning of this section. A choice of dust extinction curve is of importance especially for dusty SFGs. However, metal-poor galaxies are usually less dusty due to insufficient dust production from metal-deficient gas. The choice of dust extinction curve does not change results significantly in the most metal-poor galaxy studies.

CHAPTER 3
TELESCOPES AND INSTRUMENTS

3.1 Subaru Telescope

Subaru Telescope is a ground-based telescope located on the summit of Mauna Kea, Hawaii. Subaru Telescope has a 8.2-meter primary mirror, which is the largest monolithic primary mirror in the world. Observations of Subaru Telescope cover optical and NIR wavelength ranges. Eight main instruments of cameras and spectrographs are installed in the locations of the prime focus, and Nasmyth and Cassegrain focal points of Subaru Telescope.

In our research, we select our science targets from the deep, wide-field survey data taken by HSC, and carry out spectroscopy with the Faint Object Camera and Spectrograph (FOCAS) on Subaru Telescope. We summarize the instrumentation of HSC and FOCAS below.

3.1.1 HSC

HSC is an optical wide-field camera installed at the prime focus of Subaru Telescope. HSC has a wide FoV of 1.5-degree diameter. The HSC detector is composed of 104 main science 2k×4k CCDs as well as 4 and 8 additional 2k×4k CCDs for the auto guide and auto focus, respectively. These CCDs were manufactured by Hamamatsu Photonics K.K. The HSC imaging covers the optical wavelength range of 4,000–11,000 Å with 5 main broad-band filters and 14 narrow-band filters as of December, 2019. A one-hour integration with a broad-band filter typically reaches a depth of ∼26–28 magnitude[1] in a good observational condition.

In our study, we use the deep, wide-field survey data of HSC SSP to select EMPG candidates as described in Chapter 1. We present the HSC configuration in Figure 3.1.

[1] 2-arcsec diameter magnitude

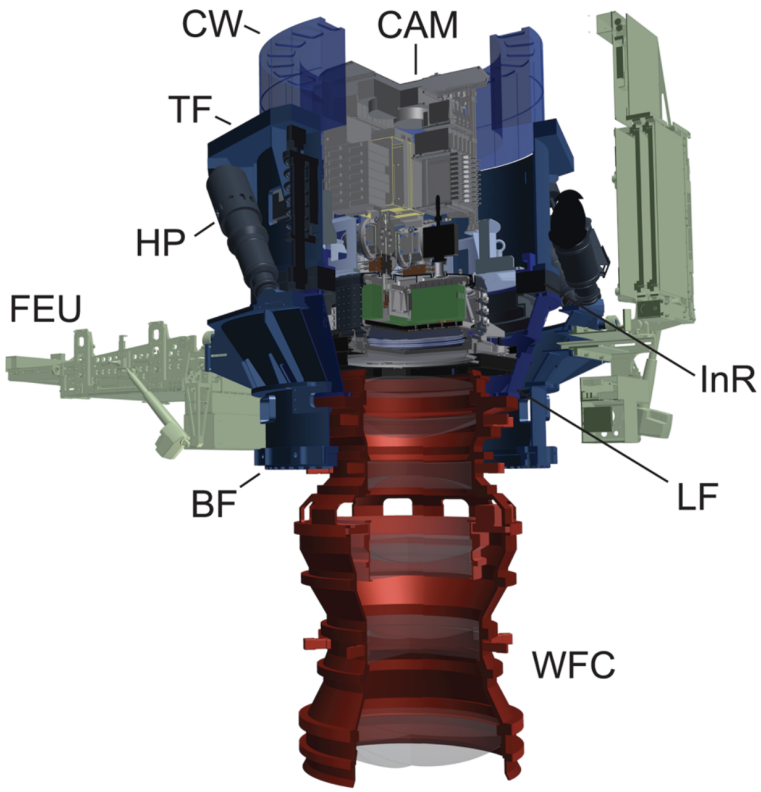

Figure 3.1 Configuration of the HSC mechanical components (Miyazaki et al., 2018). HSC is composed of three major system units of a camera (CAM), a wide-field corrector (WFC), and a main chassis body called POPT2, which holds the CAM and WFC. POPT2 consists of the Base Frame (BF), the Hexa-Pod (HP), the Top Frame (TF), the Instrument Rotator (InR), the Dewar Frame (DF), the Lens Frame (LF), and the Cable Wrapper (CW), which are detailed in Miyazaki et al. (2018).

3.1.2 FOCAS

FOCAS is an optical imaging and spectroscopy, which is installed at the Cassegrain focus of Subaru Telescope. FOCAS has a circular 6-arcmin field of view (FoV) and a spectral coverage from 3,700 to 10,000 Å. FOCAS is equipped with 5 gratings and 8 volume-phased holographic (VPH) grisms as of December, 2019, whose spectral resolutions range $R = 250$ to 7,500. FOCAS is able to conduct spectroscopy with the atmospheric dispersion corrector (ADC) placed in the Cassegrain unit. FOCAS uses two fully depleted CCDs with pixel numbers of 2k×4k, which were produced by Hamamatsu Photonics K.K. A peak total efficiency of FOCAS reaches around 40% at 6000Å.

We present the FOCAS configuration in Figure 3.2, and summarize the FOCAS characteristics in Table 3.1.

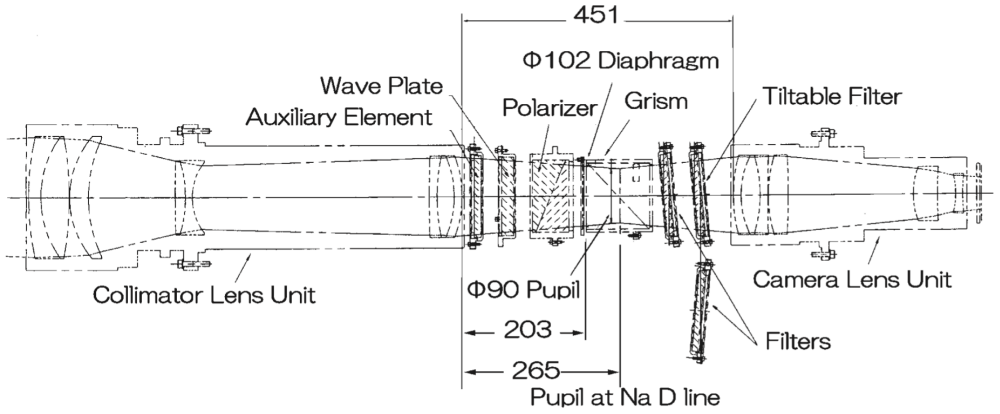

Figure 3.2 Configuration of the FOCAS mechanical components with the optical paths. This picture is cited from Kashikawa et al. (2002).

3.2 Keck Telescopes

Keck-I and Keck-II Telescopes are ground-based telescopes placed on the summit of Mauna Kea, Hawaii. Each Keck Telescope is mounted on an alt-azimuth mount, having a 10-meter primary mirror composed of 36 hexagonal segment mirrors. Observations with Keck-I and Keck-II Telescopes cover the range of wavelength from optical

to NIR. Currently, a total of eight observational instruments (optical/NIR cameras and spectrographs) are installed in the locations of the Nasmyth and Cassegrain focal points of Keck-I and Keck-II Telescopes. The combination of Keck-I/Keck-II Telescopes and these instruments provides an angular resolution of 0.04–0.4 arcsec in optical and NIR ranges.

In our research, we conducted spectroscopic observations with the Deep Imaging Multi-Object Spectrograph (DEIMOS) on Keck-II Telescope. We summarize the instrumentation of DEIMOS below.

3.2.1 DEIMOS

DEIMOS is an optical imaging and spectroscopy instrument installed at Nasmyth focus of Keck-II Telescope. The field of view (FoV) of DEIMOS is 16×4 arcmin, and DEIMOS can perform the long-slit spectroscopy or the multi-object spectroscopy (MOS). The DEIMOS gratings have spectral resolution of R =2,000–7,000, and cover a 4,000–10,000 Å wavelength range, with peak system efficiencies of $\sim$ 30%. The detector is composed of 2×4 mosaics of 2k×4k CCDs, whose pixel scale is 0.1185 ″/pixel. We present the DEIMOS configuration in Figure 3.3, and summarize the DEIMOS characteristics in Table 3.1.

3.3 Magellan Telescope

Magellan-I (Baade) and Magellan-II (Clay) Telescopes are two ground-based telescopes located in Las Campanas Observatory in Chile. Both Magellan-I and -II Telescopes have a 6.5-meter primary mirror. Seven main instruments of cameras and spectrographs are installed in Magellan-I and -II Telescopes.

In our research, we conducted spectroscopic observations with the Magellan Echellette (MagE) Spectrograph and the Low Dispersion Survey Spectrograph (LDSS-3) on Magellan-I and Magellan-II Telescopes, respectively. We summarize the instrumentation of MagE and LDSS3 below.

[2]https://www2.keck.hawaii.edu/inst/deimos/

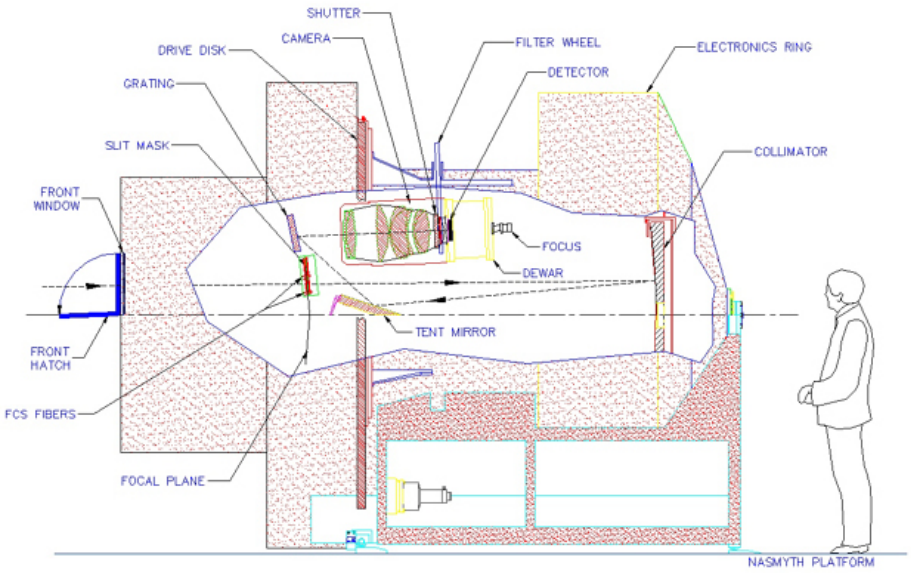

Figure 3.3 Configuration of the DEIMOS mechanical components with the optical path.[2]

3.3.1 MagE

MagE is an optical echellette spectrograph installed on the folded port 2 of Magellan-I Telescope. MagE has been designed to target a single object with a full optical wavelength coverage and a moderately high spectral resolution power. MagE provides the wavelength range of 3,200–10,000 Å with spectral resolutions of R ∼1,000 to 8,000. A single-slit length of MagE is 10-arcsec, while the slit width is selected from 0.5 to 5 arcsec. MagE has a peak efficiency of ∼20% at around 6,000 Å. Figure 3.4 demonstrates an example of a MagE science frame, where echelle spectral orders and corresponding typical wavelengths are shown with numbers. We present the MagE configuration in Figure 3.5, and summarize the MagE characteristics in Table 3.1.

3.3.2 LDSS-3

LDSS-3 is an optical imaging and spectroscopy instrument installed at Nasmyth focus of Magellan-II Telescope. LDSS-3 can perform the long-slit spectroscopy or the multi-object spectroscopy in a 8.3-arcmin diameter FoV. The LDSS-3 is equipped with 3

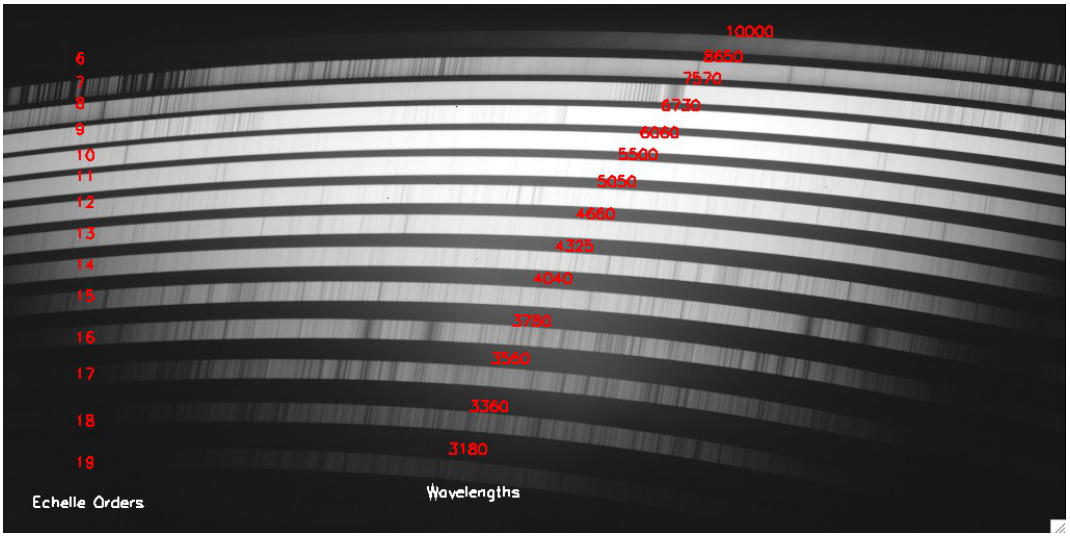

Figure 3.4 An example of a MagE science frame. Echelle spectral orders and corresponding typical wavelengths are shown with numbers. This picture is cited from Marshall et al. (2008).

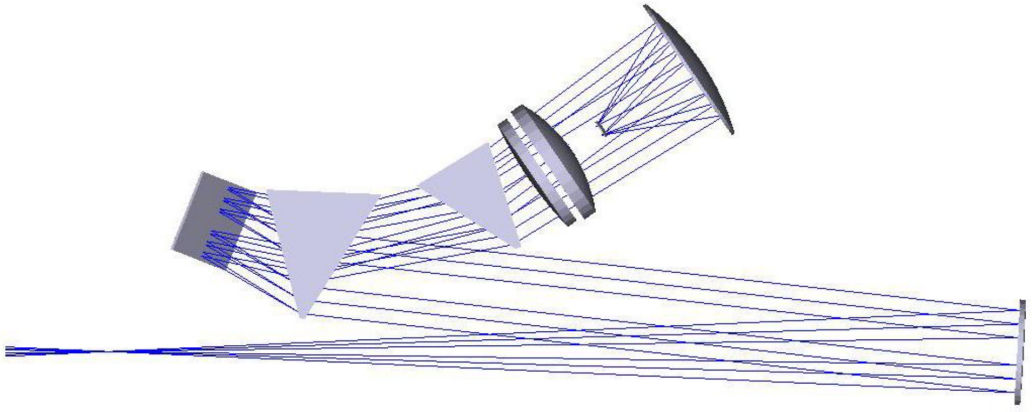

Figure 3.5 Configuration of the MagE mechanical components with the optical path. This picture is cited from Marshall et al. (2008).

VPH grisms, whose spectral resolutions are R =900–1,800 in a wavelength coverage of 3,600–10,000 Å with a peak total system efficiency around 35%. A detector of LDSS-3 consists of a 2k × 4k CCD, whose pixel scale is 0.189 ″/pixel. We summarize the LDSS-3 characteristics in Table 3.1.

Table 3.1 Spectrographs of Subaru/FOCAS, Keck/DEIMOS, Magellan/MagE, and Magellan/LDSS-3

Instrument	Wavelength coverage (μm)	Spectral resolution ($R \equiv \lambda/\Delta\lambda$)	FoV (arcmin)
Subaru/FOCAS	0.37–1.0	250–7,500	ϕ=6
Keck/DEIMOS	0.40–1.0	2,000–7,000	16×4
Magellan/MagE	0.32–1.0	1,000–8,000	*single slit*
Magellan/LDSS-3	0.36–1.0	900–1,800	ϕ=8.3

CHAPTER 4
DATA

We explain HSC-SSP imaging data used in this study in Section 4.1. We construct source catalogs from the HSC-SSP and SDSS data in Sections 4.2 and 4.3, respectively. These source catalogs and following selection processes are summarized in Figure 4.1.

4.1 HSC-SSP Imaging Data

We use the HSC-SSP internal data of S17A and S18A data releases, which are taken from 2014 March to 2017 May and from 2014 March to 2018 Jan, respectively. The internal S17A+S18A data are explained in the second data release (DR2) paper of HSC SSP (Aihara et al., 2019). The HSC-SSP survey data are taken in three layers of Wide, Deep, and UltraDeep. In the HSC-SSP S17A and S18A data releases, images were reduced with the HSC pipeline, `hscPipe` v5.4 and v6.7 (Bosch et al., 2018), respectively, with codes of the Large Synoptic Survey Telescope (LSST) software pipeline (Ivezić et al., 2019; Axelrod et al., 2010; Jurić et al., 2015). The pipeline conducts the bias subtraction, flat fielding, image stacking, astrometry and zero-point magnitude calibration, source detection, and magnitude measurement. The `hscPipe` v6.7 (S18A) uses the global background subtraction, a lower detection threshold, a new artifact rejection algorithm, the different co-add weighting between old and new i/r-band filters, and the updated way of the point-spread-function (PSF) estimation (detailed in Aihara et al., 2019). These pipeline differences slightly change the detection and magnitude measurements, which may affect our classification results. Indeed, as we will explain later in Section 6.1, we find that part of EMPG candidates are selected only in either of S17A or S18A data, which is caused by the different `hscPipe` versions between S17A and S18A. To maximize the EMPG sample size, we use both S17A and S18A data in this study. The details of the observations, data reduction, and detection and photometric catalog are described in Aihara et al. (2019) and Bosch et al. (2018). We use `cmodel` magnitudes corrected for Milky-Way dust extinction to estimate the total magnitudes of a source. See the detailed algorithm

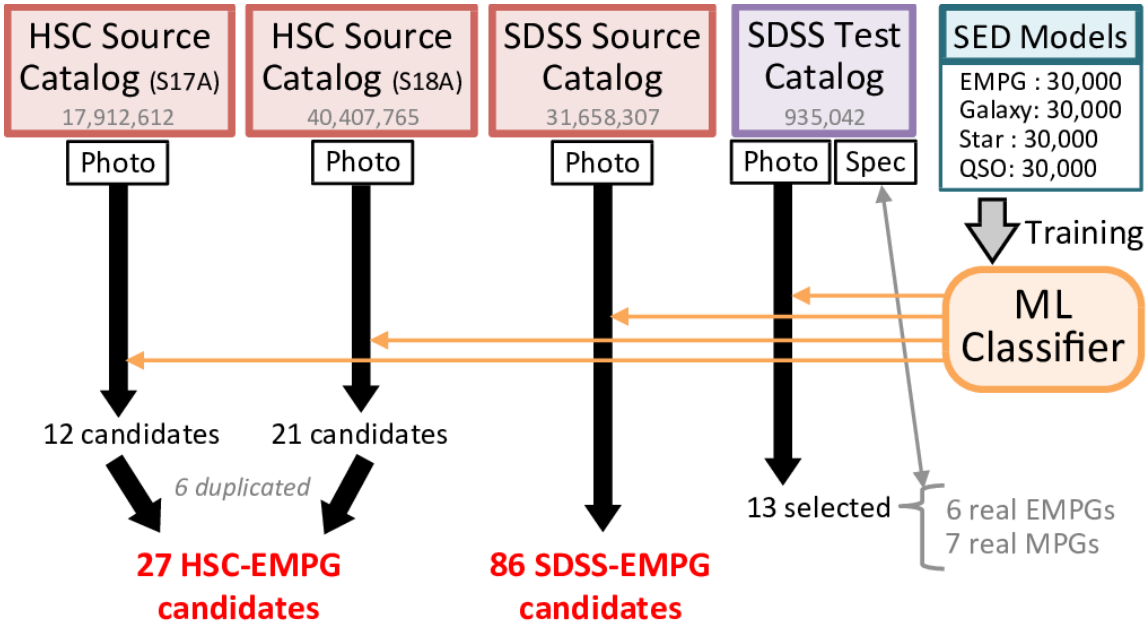

Figure 4.1 Picture of our selection flow. We select our EMPG candidates from the HSC source catalogs (Section 4.2) and SDSS source catalog (Section 4.3), which consist of photometric data. To test our ML classifier, we use the SDSS test catalog (Section 5.3.1), which is composed of photometry+spectroscopy data. Our ML classifier (Section 5.2) is trained by SED models of galaxies, stars, and QSOs (Section 5.2.3). We do not use existing observational data in the training because we target very faint EMPGs that no previous survey has been able to discover. Part of details are omitted in this flow for simplicity. See the details in each section.

of the `cmodel` photometry in Bosch et al. (2018).

4.2 HSC Source Catalog

We explain how we construct an HSC source catalog, from which we select EMPG candidates. We derive sources from Wide field of HSC-SSP data. We use isolated or cleanly deblended sources that fall within *griz*-band images. We also require that none of the pixels in their footprints are interpolated, none of the central 3×3 pixels are saturated, none of the central 3×3 pixels suffer from cosmic rays, and there are no bad pixels in their footprints. Then we exclude sources whose `cmodel` magnitude or centroid position measurements have a problem. We require detection in the *griz*-band images. We mask sources close to a bright star (Coupon et al., 2018; Aihara et al., 2019). Here we select objects whose photometric measurements are brighter than 5σ limiting magnitudes, $g < 26.5$, $r < 26.0$, $i < 25.8$, and $z < 25.2$ mag, which are estimated by Ono et al. (2018) with 1.5-arcsec diameter circular apertures. This study does not use y-band photometry because the y-band limiting magnitude is shallower ($y = 24.5$ mag; Ono et al., 2018) than the other 4 bands and the y-band imaging has not yet been completed in part of the survey area that we use this time. We also require that the photometric measurement errors are less than 0.1 mag in *griz* bands. Here the photometric measurement errors are given by `hscPipe`. Finally, we obtain 17,912,612 and 40,407,765 sources in total from the HSC-SSP S17A and S18A data, respectively. The effective area is 205.82 and 508.84 deg^2 in the HSC-SSP S17A and S18A data, respectively. Note again that there is overlap between S17A and S18A data (see also Sections 4.1 and 6.1). Table 4.1 summarizes the selection criteria that we apply to make the HSC source catalog.

Table 4.1 Selection criteria in our source catalog construction.

Parameter	Value	Band	Comment
`isprimary`	True	—	Object is a primary one with no deblended children.
`detect_ispatchinner`	True	—	Object falls on the inner region of a coadd patch.
`detect_istractinner`	True	—	Object falls on the inner region of a coadd tract.
`pixelflags_edge`	False	*griz*	Object locates within images.
`pixelflags_interpolatedcenter`	False	*griz*	None of the central 3×3 pixels of an object is interpolated.
`pixelflags_saturatedcenter`	False	*griz*	None of the central 3×3 pixels of an object is saturated.
`pixelflags_crcenter`	False	*griz*	None of the central 3×3 pixels of an object is masked as cosmic ray.
`pixelflags_bad`	False	*griz*	None of the pixels in the footprint of an object is labeled as bad.
`cmodel_flag`	False	*griz*	Cmodel flux measurement has no problem.
`merge_peak`	True	*griz*	Detected in *griz* bands.
`mask_bright_objectcenter`[a]	False	*griz*	No bright stars near an object.
`cmodel_mag`	< 26.5	g	g-band cmodel magnitudes are smaller than the 5σ limiting magnitude.
	< 26.0	r	r-band cmodel magnitudes are smaller than the 5σ limiting magnitude.
	< 25.8	i	i-band cmodel magnitudes are smaller than the 5σ limiting magnitude.
	< 25.2	z	z-band cmodel magnitudes are smaller than the 5σ limiting magnitude.
`cmodel_magsigma`	< 0.1	*griz*	Errors of *griz*-band cmodel magnitudes are less than 0.1 mag.

[a] Only used in the S18A catalog because we find many contaminants in S18A data caused by nearby bright stars, while we do not in S17A.

4.3 SDSS Source Catalog

We construct a SDSS source catalog from the 13th release (DR13; Albareti et al., 2017) of the SDSS photometry data. Although the SDSS data are ∼5 mag shallower ($i_{\rm lim}$∼21 mag) than HSC-SSP data ($i_{\rm lim}$∼26 mag), we also select EMPG candidates from the SDSS data to complement brighter EMPGs. Here we select objects whose photometric measurements are brighter than SDSS limiting magnitudes, $u < 22.0$, $g < 22.2$, $r < 22.2$, $i < 21.3$, and $z < 21.3$ mag[1]. We only obtain objects whose magnitude measurement errors are <0.1 mag in *ugriz* bands. We use `Modelmag` for the SDSS data. Among flags in the `PhotoObjALL` catalog, we require that a `clean` flag is "1" (i.e., *True*) to remove objects with photometry measurement issues. The `clean` flag[2] eliminates the duplication, deblending/interpolation problems, suspicious detections, and detections at the edge of an image. We also remove objects with a *True* comic-ray flag and/or a *True* blended flag, which often mimics a broad-band excess in photometry. We reject relatively large objects with a ninety-percent petrosian radius greater than 10 arcsec to eliminate contamination of HII regions in nearby spiral galaxies. Finally, we derive 31,658,307 sources in total from the SDSS DR13 photometry data. The total unique area of SDSS DR13 data is 14,555 deg^2.

[1]Magnitudes reaching 95% completeness, which are listed in `https://www.sdss.org/dr13/scope/`

[2]Details are described in `http://www.sdss.org/dr13/algorithms/photo_flags_recommend/`

CHAPTER 5
CONSTRUCTION OF CLASSIFIER

In this section, we construct a classifier based on the ML, which will be applied to the HSC-SSP and SDSS source catalogs to select EMPGs. We target galaxies that have a metallicity of 12+log(O/H)=6.69–7.69 (i.e., 1–10% solar metallicity) with a rest-frame Hα equivalent width of $EW_0(H\alpha)$>1000 Å. The basic idea of our selection technique is that we build an object classifier that separates EMPG candidates from other types of objects, such as non-EMPG galaxies[1], stars, and QSOs. We construct the object classifier with a deep neural network (DNN; Lecun et al., 2015). In Section 5.1, we discuss typical colors of EMPGs to show how we determine the ranges of metallicity, equivalent width, and redshift of EMPGs that we target in this study. Section 5.2 explains how we construct our ML classifier that distinguishes EMPGs from non-EMPG galaxies, stars, and QSOs. In Section 5.3, we test our ML classifier with the SDSS photometry+spectroscopy data (i.e., data of SDSS objects that are detected in photometry and observed in spectroscopy) to check whether our ML classifier successfully selects EMPGs.

5.1 EMPG Colors

We examine typical colors of EMPGs in the literature. This study only focuses on EMPGs at $z \lesssim 0.03$, where the [O III]+Hβ and Hα lines fall on the g-band and r-band, respectively.

We have compiled SDSS metal-poor galaxies at $z < 0.03$ with 12+log(O/H)<7.69 from the literature (Kunth & Östlin, 2000; Kniazev et al., 2003; Guseva et al., 2007; Izotov & Thuan, 2007; Izotov et al., 2009; Pustilnik et al., 2010; Izotov et al., 2012; Pilyugin et al., 2012; Sánchez Almeida et al., 2016; Guseva et al., 2017). Figure 5.1 shows these SDSS metal-poor galaxies on the $r - i$ vs. $g - r$ diagram, whose $EW_0(H\alpha)$ values are in the ranges of 0—300, 300–800, 800–1,200, and >1,200 Å. In

[1]We define a non-EMPG galaxy as a galaxy that does not satisfy the EMPG condition, 12+log(O/H)<7.69.

Figure 5.1, metal-poor galaxies with a higher $EW_0(H\alpha)$ have a smaller $r-i$ value with $g-r \sim 0$ due to the g- and r-band excesses caused by strong nebular emission lines (top panel of Figure 1.2). This trend is also supported by the stellar synthesis and photoionization models as shown with solid lines in Figure 5.1. These g- and r-band excesses are typical for EMPGs with strong emission lines, which basically enables us to separate EMPGs from other types of objects (e.g., galaxies, stars, and QSOs) only with photometric data. In addition, as described in Chapter 1, EMPGs with strong emission lines are expected to be local analogs of high-z SFGs because high-z SFGs have a high sSFR, which corresponds to high emission-line equivalent widths, and a low metallicity.

As described in Chapter 1, there are many contaminants in EMPG candidates selected with the classical color-color selection. Figure 5.2 shows the SDSS EMPGs with $EW_0(H\alpha)$>800 Å on the $r-i$ vs. $g-r$ diagram as well as the SDSS source catalog created in Section 4.3. Figure 5.2 demonstrates that the positions of the EMPGs are overlapped by many sources on the $r-i$ vs. $g-r$ diagram. With the visual inspection, we find that most of the overlapping sources are contaminants such as stars and artifacts. Thus, we suggest that the classical color-color diagram is not effective for selecting EMPGs.

We also compare EMPGs with GPs and BBs on the $r-i$ vs. $g-r$ diagram. Figure 5.3 suggests that the EMPGs, GPs and BBs fall on the different spaces on the $r-i$ vs. $g-r$ diagram. The solid and dotted lines are the selection criteria of the GPs (Cardamone et al., 2009) and BBs (Yang et al., 2017b), respectively. We also show the model calculations of galaxies with 12+log(O/H)=8.00 (GP/BB) and 6.69 (EMPG) with gray and blue solid lines, respectively (see model details in Section 5.2.3). The model calculations of galaxies with 12+log(O/H)=6.69 (EMPG) do not fall on neither the GP nor BB selection criterion, which basically means that the GP/BB selection criteria cannot select EMPGs with 12+log(O/H)=6.69. In other words, the selection with g- and r-band excesses is one of the limited windows to search for EMPGs in the local to low-z universe.

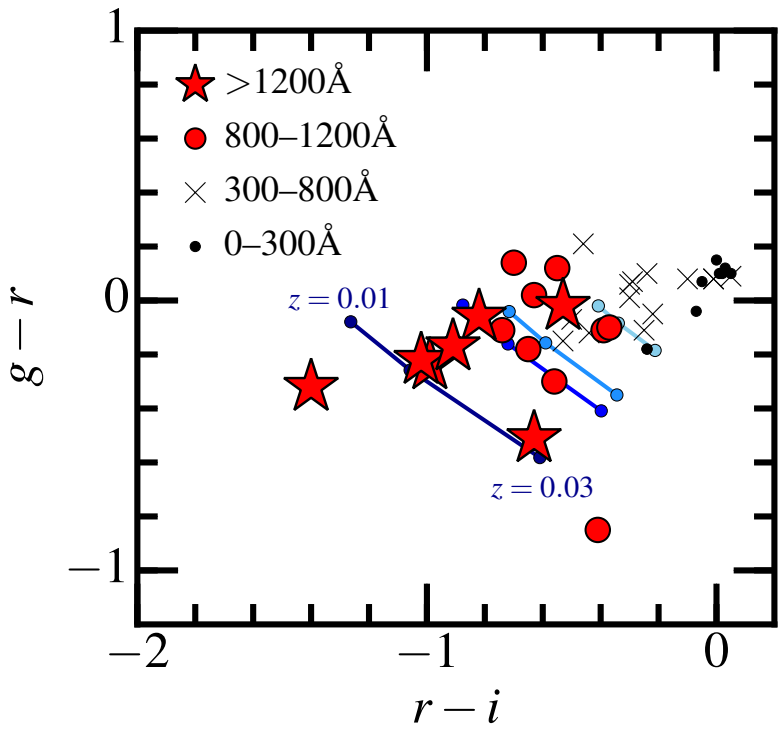

Figure 5.1 Color-color diagram of $g-r$ vs. $r-i$ for previously reported metal-poor galaxies with 12+log(O/H)<7.69 at $z < 0.03$. The red stars, red circles, black crosses, and black dots present SDSS metal-poor galaxies with $EW_0(H\alpha)$>1,200 Å, $EW_0(H\alpha)$=800–1,200 Å, $EW_0(H\alpha)$=300–800 Å, and $EW_0(H\alpha)$=0–300 Å, respectively. The four blue sold lines present our BEAGLE model calculations (Chevallard & Charlot, 2016) with $EW_0(H\alpha)$∼2,500, 1,500, 1,000, and 500 Å (from dark blue to light blue) under the assumption of 12+log(O/H)=7.50. Our BEAGLE models here are calculated in the same manner as the EMPG models in Section 5.2.3. On the blue solid lines, redshifts are indicated with dots ($z = 0.01$, 0.02, and 0.03 from upper left to lower right). The SDSS metal-poor galaxies with a larger $EW_0(H\alpha)$ show smaller $r-i$ values due to the strong Hα-line contribution in an r-band magnitude, which are consistent with the BEAGLE model calculations.

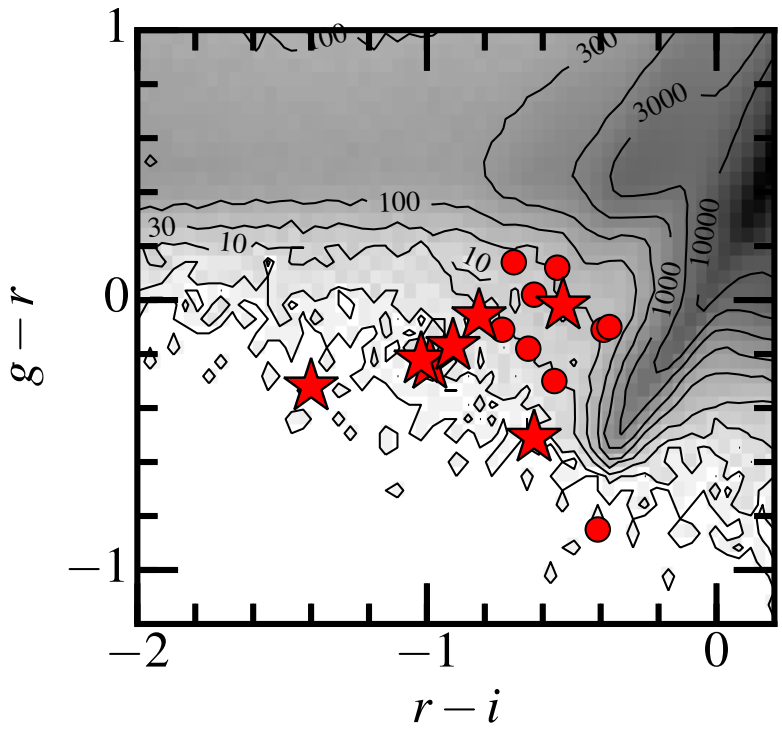

Figure 5.2 The same as Figure 5.1, but only with the SDSS EMPGs with $EW_0(H\alpha)$>800 Å. The black mesh and contours present a two-dimensional histogram of the SDSS source catalog (discussed later in Section6.2) . The contours indicate the number of sources (N=1, 3, 10, 30,..., 10000) in each bin with a size of Δm=0.04 mag. On this color-color diagram, the EMPGs overlap largely with many SDSS sources, most of which are contaminants such as stars.

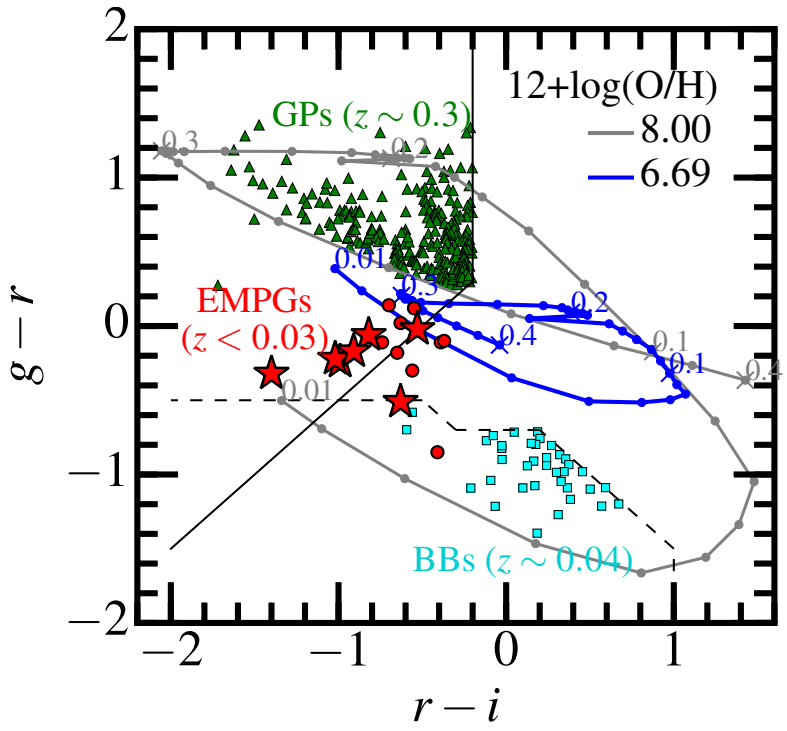

Figure 5.3 Model calculation of $g-r$ and $r-i$ colors for galaxies with $EW_0(H\alpha)\sim$2,000 Å that have 12+log(O/H)=6.69 (blue solid line; EMPGs) and 12+log(O/H)=8.00 (gray solid line; GPs). Dots and crosses are placed in every step of Δz=0.01 and 0.1, respectively. The green triangles and cyan squares are GPs at $z \sim 0.3$ (Cardamone et al., 2009) and BBs at $z \sim 0.04$ (Yang et al., 2017b), respectively. The black solid and dotted lines are boundaries used to select GPs at $z \sim 0.3$ (Cardamone et al., 2009) and BBs at $z \sim 0.04$ (Yang et al., 2017b), respectively. EMPGs derived from the literature are shown with the red stars and circles (see Figure 5.1). The black solid and dotted lines indicate selection criteria for GPs (Cardamone et al., 2009) and BBs (Yang et al., 2017b), respectively.

5.2 ML Classifier

We construct an object classifier with the DNN that separates EMPGs from other types of objects. In the following sections, we explain the merits of DNN (Section 5.2.1), how the ML classifier is constructed (Section 5.2.2) and how the training sample is generated from models (Section 5.2.3). We train the ML classifier with spectral energy distribution (SED) models because very faint EMPGs ($\gtrsim$21 mag) have not been discovered yet, and it is impossible to make a training sample from existing data.

5.2.1 Merits

There are four merits of the use of the DNN as shown below.

i) We can select objects in the multi-dimensional photometry space (e.g., in a *grizy* 5-dimensional space), while classical selections use a projection onto a 2-dimensional color-color diagram. As discussed in Section 5.1, EMPGs are overlapped by many sources on a color-color diagram of $g-r$ vs. $r-i$ (Figure 5.2) for instance. If we use criteria in a multi-dimensional space, we can eliminate such overlapping sources more efficiently in principle.

ii) We can use a non-linear boundary that separates object types. The DNN can determine a non-linear boundary thanks to a non-linear function, called an activation function, which is used in the DNN structure (see Section 5.2.2). Although classical selections try to separate object types with a straight line on a color-color diagram, different types objects are not always separated by such a simple, straight line. The use of a non-linear boundary usually reduces the contamination and increases the completeness.

iii) A boundary is optimized by the DNN algorithm, albeit the classical boundaries are determined by eyes. The DNN enables the objective determination of the boundaries. Figure 5.4 is a schematic illustration of merits i) to iii).

iv) The DNN selection is very fast. Indeed, in principle, we are able to select EMPG candidates by fitting with SED models of galaxies, stars, and QSOs in a wide range of parameters. However, such SED fitting takes much longer time than the

DNN. Our DNN classifier requires only less than several minutes to train itself and classify sources once we produce SED models of galaxies, stars, and QSOs.

Generally speaking, these merits are also obtained with another ML algorithm such as support vector machine (SVM). However, we only focus on the use of the DNN in this thesis because our purpose is to construct an EMPG sample, not to find the most efficient selection methods. In Section 10.1, we spectroscopically confirm that the success rate of our ML classifier is over $\gtrsim 50\%$, which is high enough to construct an EMPG sample. Thus, the comparison between the DNN and other ML techniques is out of the scope of this thesis.

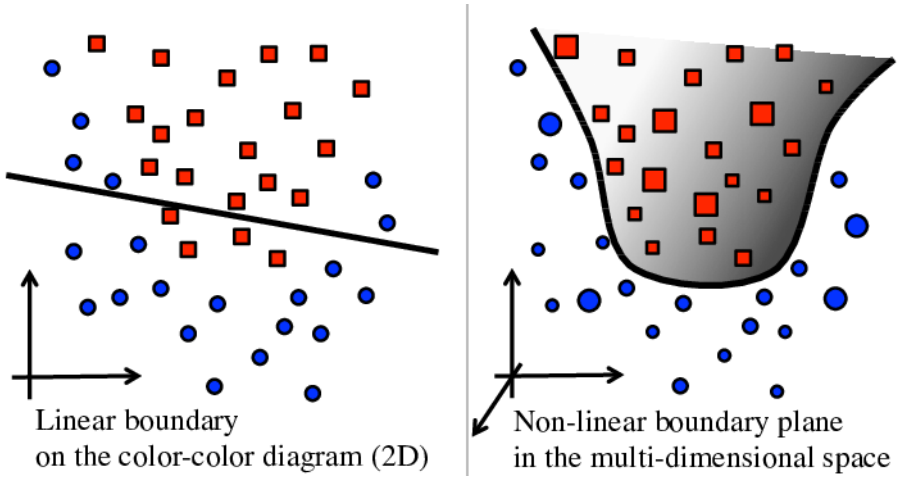

Figure 5.4 Schematic illustrations of two methods of the object classification. Left—The object classification on the color-color diagram with a linear boundary. Two different types of objects are presented with the squares and circles. Right—The object classification in the multi-dimensional space (e.g., 5 dimensional space of *grizy*-band magnitudes) with a non-linear boundary. The DNN classification corresponds to the right illustration, while the left panel demonstrates a classical color-color selection. The size of circles and squares presents a distance on the line of sight.

5.2.2 Structure

We construct an object classifier that distinguishes four object types of EMPGs, non-EMPG galaxies, stars, and QSOs. For every source input, the classifier calculates probabilities of the four types and chooses only one type whose probability is the

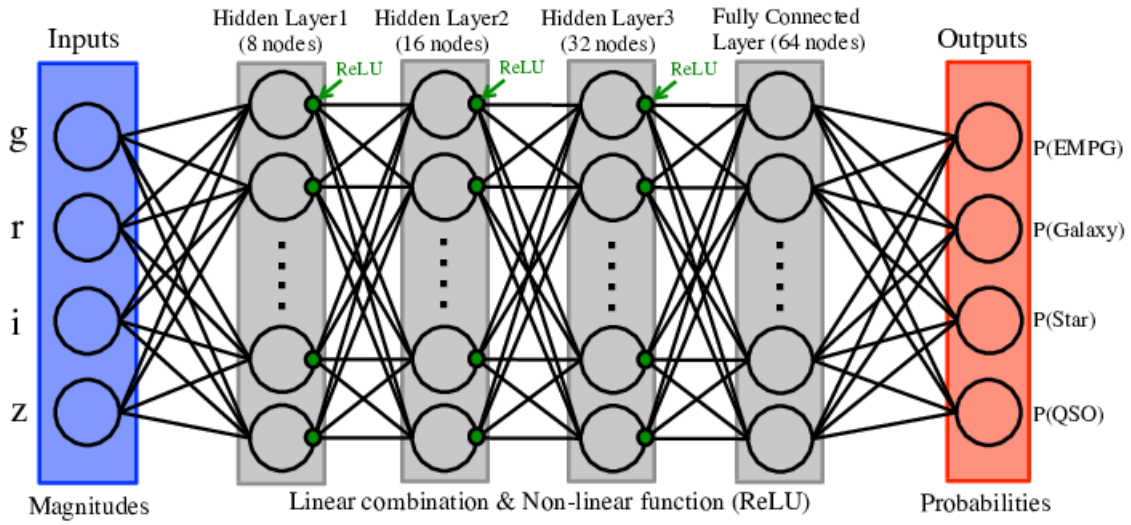

Figure 5.5 Schematic illustration of the structure of our ML classifier based on the DNN. The nodes (blank circles) and branches (solid lines) represent a linear combination. The green circles are ReLU activation functions.

highest in the four. We conduct calculation with an open-source software library for ML, `Tensorflow`[2]. Its detailed structure and training process are explained below.

The object classifier is constructed with the DNN that consists of three hidden layers and one fully connected layer. Figure 5.5 is an schematic illustration of the structure of our classifier. The three hidden layers and one fully connected layer have 8/16/32 and 64 nodes, respectively. As Figure 5.5 shows, these nodes are connected with branches, which represent a linear-combination calculation. Each node in the hidden layers is followed by an activation function called rectified linear unit (ReLU; Morandi et al., 2012). The activation function, ReLU, is a non-linear function, which is essential to construct a deep-layer structure. In the fully connected layer, 10% of the nodes are dropped at random to avoid over-fitting.

Inputs of our classifier are four (five) photometric magnitudes of HSC *griz* bands (SDSS *ugriz* bands). We do not use HSC *y*-band photometry, which is shallower than the other bands, to reach as faint magnitudes as possible. After calculations, the classifier outputs four probabilities of the EMPG, non-EMPG galaxy, star, and

[2]https://www.tensorflow.org

QSO and chooses only one type whose probability is the highest among the four. Here we obtain probabilities with the `softmax` function.

The structure of the neural network is optimized so that sums of output errors are minimized. We optimize our classifier with a training sample, in which object types are already known beforehand. The optimization process is usually called "training". The cross entropy and Adam optimizer (Kingma & Ba, 2015) are used to evaluate and minimize the errors in the training, respectively. To train our classifier, we prepare a training sample with the SED models, which will be detailed in Section 5.2.3. Then the training sample is divided into two independent data sets. Here, 80% of the training sample is used as training data and the other 20% as check data. We use the training data to train the neural network, while the check data are prepared to check whether the classifier successfully separates EMPGs. In every step, 100 models are randomly derived from the training sample and used to train the neural network. This training step is repeated 10,000 times, until which the errors converges at minimum values.

Success cases are defined by the true-positive EMPGs (i.e., a real EMPG classified as an EMPG) and its true-negative (i.e., a real galaxy/star/QSO classified as a galaxy/star/QSO) as summarized in Figure 5.6. This is because we only focus on whether an object is an EMPG or not. In other words, we ignore mistakes in the classification among galaxies, stars, and QSOs. The ignorance little affects our results of the classification between EMPGs and objects not EMPGs. We define a success rate by the number of success cases over the number of total classification. After the 10,000 steps in the training, the success rate converges at $\gtrsim$99.5%.

5.2.3 Training Sample

We prepare the training sample that is used to train ML classifier explained in Section 5.2.2. The training sample consists of photometric magnitudes calculated from models of EMPGs, non-EMPG galaxies, stars, and QSOs. The photometric magnitudes are calculated from the SED models by convoluting the SEDs with the throughput curves of the HSC broadband filters (Kawanomoto et al., 2018) or the SDSS broadband filters

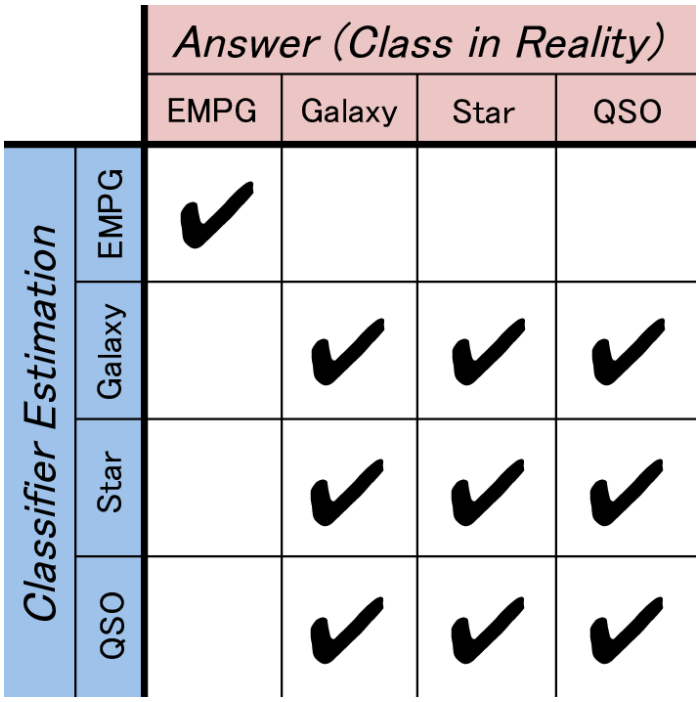

Figure 5.6 Matrix explains a definition of success cases. The columns and rows correspond to answers in reality and estimations made by the classifier, respectively. Check marks are placed in boxes of success cases. We ignore mistakes in the classification between galaxies, stars, and QSOs because we only aim to select EMPGs.

(Fukugita et al., 1996). Below, we detail the models of the EMPG, non-EMPG galaxy, star, and QSO.

1) EMPG model: We explain how we obtain the EMPG models and calculate their photometric magnitudes. We generate EMPG SEDs with the SED interpretation code, BEAGLE (Chevallard & Charlot, 2016). The BEAGLE code calculates both the stellar continuum and the nebular emission (line + continuum) in a self-consistent manner, using the stellar population synthesis code (Bruzual & Charlot, 2003) and the photoionization code, CLOUDY (Ferland et al., 2013). The BEAGLE codes use the CLOUDY photoionization models produced by Gutkin et al. (2016), where the photoionization calculations are stopped when the electron density falls below 1% of the hydrogen density or if the temperature falls below 100 K. In the CLOUDY photoionization models, we assume the solar carbon-to-oxygen abundance ratio (C/O) and the metallicity-dependent nitrogen-to-oxygen abundance ratio (N/O) given in Gutkin et al. (2016). In the BEAGLE calculation, we change five parameters of stellar mass, maximum stellar age (called just "age" hereafter), gas-phase metallicity, ionization parameter, log U, and redshift as shown below.

- $\log(M_\star/M_\odot)$=(4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0)
- log(age/yr)=(6.00, 6.25, 6.50, 6.60, 6.70, 6.80, 6.90, 7.00, 7.25, 7.50, 7.75)
- 12+log(O/H)=(6.69, 7.19, 7.69)
- log U=(-2.7, -2.5, -2.3)
- redshift=(0.01, 0.02)

These stellar mass, age, and gas-phase metallicity cover typical values of EMPGs. A stellar metallicity is matched to a gas-phase metallicity here. The ionization parameter is defined by a ratio of hydrogen ionizing-photon flux, S_{H^0} and hydrogen gas density, n_H, normalized by speed of light, c,

$$\log U \equiv \frac{S_{H^0}}{c\, n_H}. \tag{5.1}$$

We choose ionization parameters of log U=(-2.7, -2.5, -2.3), which are typical values for metal-poor galaxies as demonstrated in Figure 5.7. The constant star-formation history is assumed in the model. Here, we also assume no dust attenuation because we target very metal-poor galaxies, where the dust production is insufficient. Indeed, representative metal-poor galaxies (e.g., Thuan & Izotov, 2005; Izotov et al., 2009; Skillman et al., 2013; Hirschauer et al., 2016; Izotov et al., 2018b, 12+log(O/H)$\sim$7.0–7.2) show a negligibly small dust attenuation with a color excess of $E(B-V) \sim 0$. The Chabrier (2003) stellar initial mass function (IMF) is applied in the BEAGLE code (Chevallard & Charlot, 2016). In total, we generate 2,178 (=11×11×3×3×2) SEDs with the parameters described above. For each SED, the BEAGLE code also calculates the photometric magnitudes with response curves of the HSC and SDSS filters, as well as emission line equivalent widths. From the 2,178 model SEDs, we only select 1,397 models that satisfy $i < 26$ mag and EW$_0$(Hα)>1,000 Å. The 26 mag corresponds to about an i-band limiting magnitude of the HSC imaging data. The EW$_0$(Hα)>1,000 Å corresponds to an age $\lesssim$ 10–100 Myr in this metallicity range, 12+log(O/H)<7.69 as shown in Figure 5.8. Then, to take the magnitude errors of 0.1 mag (see Sections 4.2 and 4.3) into consideration in models, we generate random numbers under the assumption of the normal distribution with $\sigma = 0.1$ and add them to the photometric magnitudes. Here we generate 30 sets of random numbers for each model[3]. Thus, we obtain a total of 41,910 (=1,397×30) models including magnitude errors. We do not use models that satisfy z =0.02–0.03 because we find that a contamination rate increases in that case.

2) Galaxy model (non-EMPG): We introduce two types of non-EMPG galaxies: normal SFGs and GPs. First, we generate SEDs of normal SFGs with the BEAGLE code similarly to the EMPG models. In the calculation, we change five parameters of stellar mass, age, metallicity, redshift, and V-band dust-attenuation optical depth ($\tau_{\rm V}$) as shown below, assuming a bursty star-formation history.

- log($M_\star/M_\odot$)=8.0, 9.0, 10.0, 11.0

[3]We remove random numbers beyond $\sigma = 0.1$ from models as we eliminate sources with $\sigma > 0.1$ in the source catalogs (Sections 4.2 and 4.3). We continue to generate random numbers until the total number becomes 250.

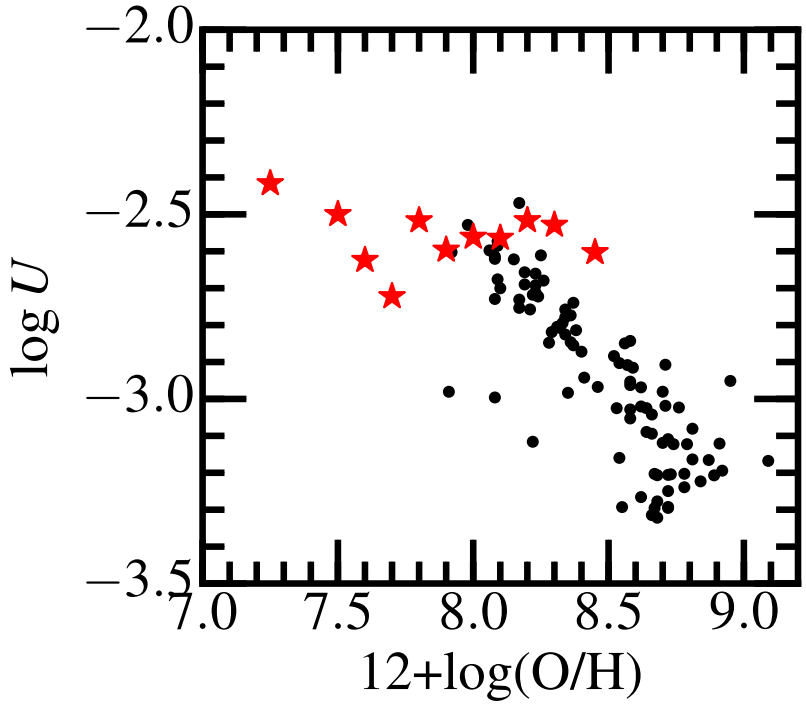

Figure 5.7 Ionization parameters of typical local galaxies as a function of metallicity. Red stars represent averages of the local metal-poor galaxies Nagao et al. (2006, sample A+B). Black dots are obtained from the SDSS composite spectra of Andrews & Martini (2013). Metallicities are based on the electron temperature measurements. Ionization parameters are calculated assuming the photoionization model of Kewley & Dopita (2002). This figure suggests that galaxies with 12+log(O/H)=7.0–7.5 have $\log U \sim -2.5$.

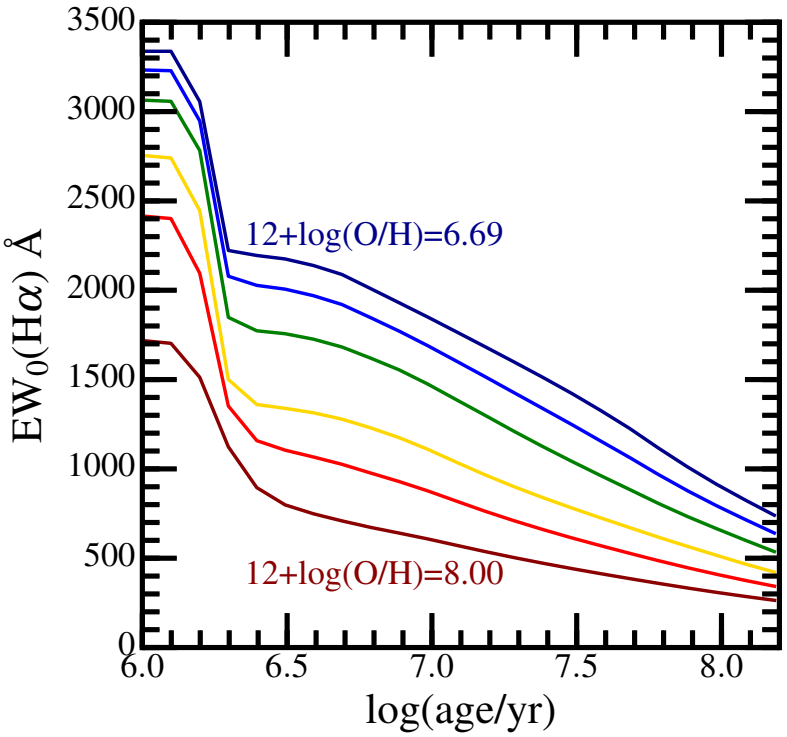

Figure 5.8 $EW_0(H\alpha)$ values as a function of age. The colors of the solid lines correspond to metallicities, 12+log(O/H)=6.69, 6.94, 7.19, 7.50, 7.69, and 8.00 from dark blue to dark red. These relations are provided by the BEAGLE models under the assumption of the constant star-formation.

- log(age/yr)=8.5, 9.0, 9.5, 10.0
- 12+log(O/H)=8.19, 8.69, 8.89
- redshift=0.02, 0.04, 0.06, 0.08, 0.10
- log($\tau_{\rm V}$)=0.0, 2.0

These stellar mass, age, metallicity, and V-band dust-attenuation optical depth are selected from typical values of local SFGs. We fix an ionization parameter to log U=−3.0, which is a value representative of local galaxies as demonstrated in Figure 5.7. In total, we generate 480 SEDs with the parameters described above. The photometric magnitudes are calculated in the same manner as the EMPG models. From the 480 models, we only select models that satisfy $i < 26$ mag. After the i-band magnitude selection, 471 models remain. We introduce magnitude errors of 0.1 mag similarly to the EMPG models, generating 10 sets of random numbers for 471 models. Then we have 47,100 normal-SFG models in total including magnitude errors.

Second, we also create GP SEDs with the BEAGLE code with following 4 parameters.

- log($M_\star/M_\odot$)=7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0
- log(age/yr)=6.0, 6.1, 6.2,..., 8.0
- log U=(−3.0, −2.5, −2.0)
- redshift=0.08, 0.09, 0.10,..., 0.40

Metallicity is fixed at 12+log(O/H)=8.00, which is typical value of GPs (see Figure 1.1). We also assume dust free in the GP models. We obtain 14,553(=7×21×3×21) models. From the 14,553 models, we use 3,234 models that satisfy $i < 26$ mag. We also introduce 0.1-mag errors in magnitude as described above, generating 10 sets of random numbers for 3,234 models. Then we have 32,340 GP models in total.

We combine the 47,100 normal-SFG models and 32,340 GP models into 79,440 models of non-EMPG galaxies. We do not include galaxies over $z = 0.1$ in the non-EMPG galaxy models because galaxies at $z > 0.1$ have quite different colors from

those of low-z EMPGs. The inclusion of high-z galaxy models does not change our result in principle.

3) Star model: We use stellar SED models of Castelli & Kurucz (2004), where 53 types of stars are modeled from O-type to M-type. SEDs are calculated in a metallicity range of $\log(Z/Z_{\odot}) = (-2.5, -2.0, -1.5, -1.0, -0.5, \pm 0.0, +0.2, +0.5)$ for each stellar type. Thus, we obtain 424 (=53×8) model SEDs in total. These star model SEDs are derived from the STScI web site[4]. Assuming the HSC and SDSS filters, we calculate $u-i$, $g-i$, $r-i$, and $z-i$ colors from the 424 model SEDs. Then, we determine i-band magnitudes, selecting 10 values in the range of i = 15–26 mag at regular intervals. Multiplying the 424 sets of $u-i$, $g-i$, $r-i$, and $z-i$ colors and the 10 i-band magnitudes, we generate 4,240 (=424×10) sets of star models with photometric magnitudes. In addition, we also introduce magnitude errors (0.1 mag) similarly to the EMPG models, obtaining 42,400 (=4,240×10) star models in total.

4) QSO model: We use a composite spectrum of QSOs at $1 < z < 2.1$ that are observed by the X-SHOOTER spectrograph installed on Very Large Telescope (VLT) (Selsing et al., 2016). This composite spectrum covers the wide wavelength range of 1,000–11,000 Å in the rest-frame. From this composite spectrum, we generate mock spectra by varying three parameters as follows: the power law index (α) of an intrinsic near UV (NUV) slope $f_{\lambda} \propto \lambda^{\alpha}$, the V-band dust-attenuation optical depth, and the redshift.

- α=−2.0, −1.5, −1.0, −0.5
- $\tau_{\rm V}$=0.0, 0.5, 1.0
- redshift=0.1, 0.2, 0.3, ..., 3.0.

The intrinsic NUV slope and V-band dust-attenuation optical depth of typical QSOs are well covered by the parameters above (e.g., Telfer et al., 2002; Selsing et al., 2016). Then we get 360(=4×3×30) QSO model SEDs in total. Similarly to the star models, we calculate $u-i$,$g-i$, $r-i$, and $z-i$ colors from the 360 model SEDs. Here, we take 10 values of i-band magnitude in the range of i = 15–26 at

[4] `ftp://ftp.stsci.edu/cdbs/grid/ck04models`

regular intervals. From both the 360 sets of $g-i$, $r-i$, and $z-i$ colors and the 10 i-band magnitudes, we generate 3,600 sets of $ugriz$-band model magnitudes. We also introduce magnitude errors (0.1 mag) similarly to the EMPG models, obtaining 36,000 (=3,600×10) QSO models in total.

In this section, we have generated 41,910, 79,440, 42,400, and 36,000 models for the EMPG, non-EMPG galaxy, star, and QSO, respectively. Selecting 30,000 models from each of EMPG, non-EMPG galaxy, star, and QSO, we obtain a training sample composed of 120,000 (=30,000×4) models in total. Figure 5.9 shows models of EMPGs, non-EMPG galaxies, stars, and QSOs on the projected color-color diagrams of $g-r$ vs. $r-i$ and $r-i$ vs. $i-z$. Here, EMPGs are overlapped with non-EMPG galaxies and stars on these projected color-color diagrams, which potentially causes contamination in the EMPG selection.

5.3 Test with SDSS Data

Before we apply our ML classifier to the HSC-SSP and SDSS source catalogs, we test if our classifier successfully distinguishes EMPGs from other types of objects (non-EMPG galaxy, star, or QSO). We carry out the test with SDSS data whose sources are detected in photometry and observed in spectroscopy. Such SDSS data are hereafter referred to as SDSS photometry+spectroscopy data in this thesis. The SDSS photometry+spectroscopy data set is a good test sample because we can easily check object types (galaxy, star, or QSO) and metallicities in their spectra. We can also see if a source satisfies the EMPG condition of 12+log(O/H)<7.69. We do not expect to discover unconfirmed EMPGs from the SDSS photometry+spectroscopy data, because the SDSS photometry+spectroscopy data have been intensively investigated by many authors (e.g., Sánchez Almeida et al., 2016; Guseva et al., 2017). Keep in mind that our final goal is to discover unconfirmed EMPGs in the HSC data, whose limiting magnitude is $\gtrsim$5 mag deeper than the SDSS data. Here we explain how we create a SDSS test catalog in Section 5.3.1 and the test results are described in Section 5.3.2.

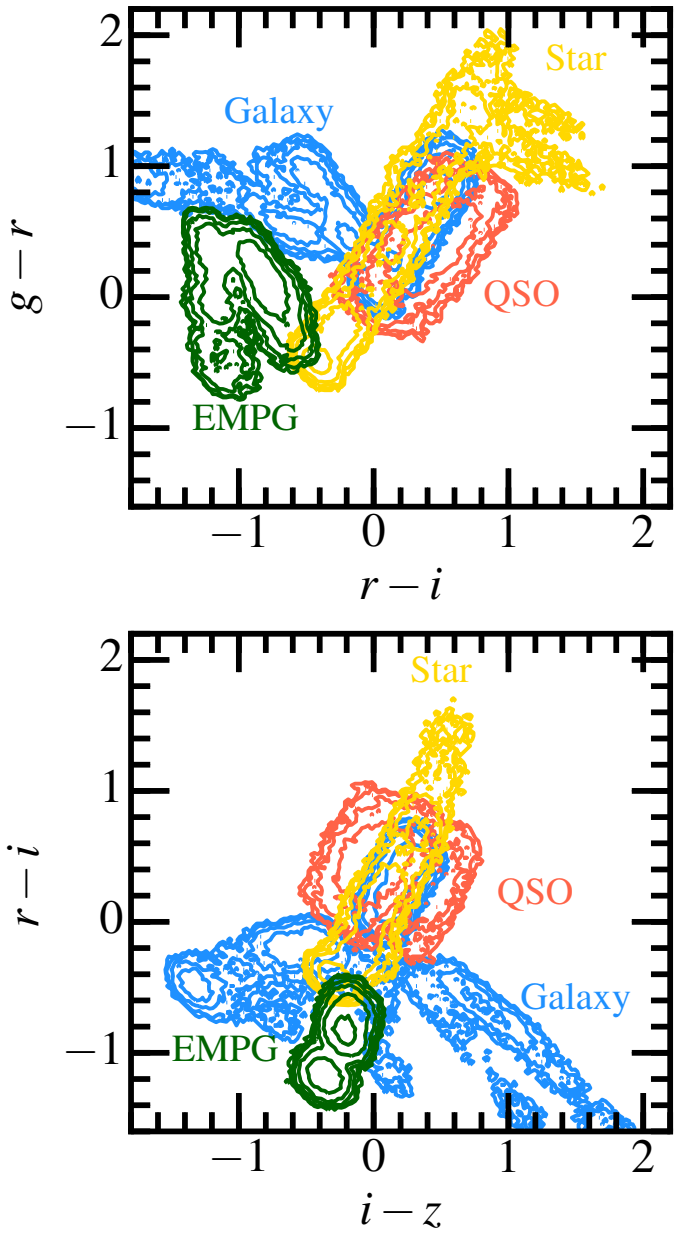

Figure 5.9 Top: Model templates of EMPGs (green), non-EMPG galaxies (blue), stars (yellow), and QSOs (red) on the color-color diagrams of $g-r$ vs. $r-i$. The contours show the number of these models (N=1, 3, 10 ,30, 100, 300) in each bin with a size of Δm=0.025 mag. Bottom: Same as the top panel, but for $r-i$ vs. $i-z$.

5.3.1 SDSS Test Catalog

We construct a *SDSS test catalog* from the SDSS DR13 data. The SDSS DR13 data is based on the SDSS-I through SDSS-IV data, which contain the extragalactic spectroscopy data from the Sloan Digital Sky Survey (SDSS-I, York et al., 2000), the Sloan Digital Sky Survey II (SDSS-II, Abazajian et al., 2009), the Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al., 2013), and extended Baryon Oscillation Spectroscopic Survey (eBOSS, Dawson et al., 2016). In the SDSS data, typical wavelength ranges are 3,800–9,200 Å or 3,650-10,400 Å and a typical spectral resolution is R=1,500–2,500. We only select objects crossmatched with the photometry catalog of `PhotoObjAll` and the spectroscopy catalogs of `SpecObjAll`, `galSpecExtra`, and `galSpecLine`. Then we construct the SDSS *test* catalog in the same way as the SDSS *source* catalog described in Section 4.3. The SDSS test catalog is composed of 935,042 sources (579,961 galaxies, 327421 stars, and 27,660 QSOs) in total. The spectroscopic effective area of the SDSS data is 9,376 deg^2.

5.3.2 Tests

We apply our ML classifier (Section 5.2) to the SDSS test catalog (Section 5.3.1). Then we have obtained thirteen EMPG candidates from the SDSS test catalog. We have checked their object classes (galaxy, star, and QSO) that are given in the `SpecObjAll` catalog based on spectroscopy. Based on the images, spectra and the object classes, we identify all of the thirteen candidates as galaxies.

For the thirteen galaxies, we obtain redshift, EW$_0$(Hα), and EW$_0$(Hβ) values from `SpecObjAll` and `galSpecLine` catalogs. Metallicities of the thirteen galaxies are derived from the literature (Kunth & Östlin, 2000; Kniazev et al., 2003; Guseva et al., 2007; Izotov & Thuan, 2007; Izotov et al., 2009; Pustilnik et al., 2010; Izotov et al., 2012; Pilyugin et al., 2012; Sánchez Almeida et al., 2016; Guseva et al., 2017). Note that these metallicities are calculated based on electron temperature measurements. We find that six out of the thirteen galaxies are EMPGs with a high Hα equivalent width, EW$_0$(Hα)>800Å. Although the other seven galaxies do not fulfill the EMPG definition, they still have low metallicities, 12+log(O/H)=7.8–8.3. As we have ex-

pected, all of the thirteen galaxies show large $EW_0(H\alpha)$ values (750–1,700 Å). We summarize their object classes, redshifts, $EW_0(H\alpha)$, $EW_0(H\beta)$, and metallicities in Table 5.1.

The success rate, or a purity, of our EMPG selection is 46% (6/13) for the SDSS test catalog. It is worth noting that the other 54% galaxies (7/13) also show a low metallicity as described above. In the local metal-poor galaxy sample obtained from the literature in Section 5.1, we find 7 EMPGs, which are also included in the SDSS test catalog and have a high equivalent width, $EW_0(H\alpha)$>800Å, necessary to be selected by the *gr*-band excess technique. In other words, we have successfully selected the 6 EMPGs as above out of the 7 known high-EW EMPGs in the SDSS test catalog, which suggests that our selection reaches 86% (6/7) completeness. Thus, we conclude that our selection method has successfully selected EMPGs and EMPG-like galaxies from the SDSS test catalog.

Table 5.2 shows the probabilities of EMPG, galaxy, star, and QSO (P(EMPG), P(Galaxy), P(Star), and P(QSO)) estimated by the ML classifier for the 7 known EMPGs in the SDSS test catalog. Here, for easy comparison, we show probabilities obtained before we apply the `softmax` function. The 6 known EMPGs that are successfully selected from the SDSS test catalog show that P(Star) is the second highest among the four probabilities. The other known EMPG not selected by the ML classifier has been classified as a star with P(Star)=0.583. The relatively high P(Star) probabilities mean that stars have similar colors to EMPGs.

Because we have confirmed that our ML classifier successfully selects EMPGs from the SDSS test catalog as described above, we do not improve it further in this thesis. However, the ML classifier still has potential for the further improvement by introducing additional information such as size and morphology in the future step. We may be able to apply the ML technique to select other types of galaxies, stars, AGNs, and so on.

Table 5.1 Parameters of EMPG candidates selected in the SDSS test

#	ID	class	redshift	EW$_0$(Hα) Å	EW$_0$(Hβ) Å	12+log(O/H)	citation
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	J012534.2+075924.5	EMPG	0.010	1351	242	**7.58**	P12
2	J080758.0+341439.3	EMPG	0.022	1277	252	**7.69**	SA16
3	J082555.5+353231.9	EMPG	0.003	1441	238	**7.45**	K03
4	J104457.8+035313.1	EMPG	0.013	1462	276	**7.44**	K03
5	J141851.1+210239.7	EMPG	0.009	1153	215	**7.50**	SA16
6	J223831.1+140029.8	EMPG	0.021	953	183	**7.43**	P12
7	J001428.8−004443.9	Galaxy	0.014	995	184	8.05	P12
8	J025346.7−072344.0	Galaxy	0.005	787	138	7.97	K04
9	J115804.9+275227.2	Galaxy	0.011	750	119	8.34	B08
10	J125306.0−031258.8	Galaxy	0.023	1291[1]	236	8.08	K04
11	J131447.4+345259.7	Galaxy	0.003	1700	293	8.14	B08
12	J132347.5−013252.0	Galaxy	0.022	1458	248	7.77	K04
13	J143905.5+364821.9	Galaxy	0.002	766[1]	140	7.94	P12

(1): Number. (2): ID. (3): Object class. Galaxies with 12+log(O/H)<7.69 are classified as an EMPG. (4): Redshift. (5), (6): Rest-frame equivalent widths of Hα and Hβ emission lines. These values are obtained from the SDSS DR13 catalog. (7): Gas-phase metallicity obtained with the electron temperature measurement. (8): Citation from which the metallicity values are derived—P12: Pilyugin et al. (2012), K03: Kniazev et al. (2003), K04: Kniazev et al. (2004), B04: Brinchmann et al. (2008), SA16: Sánchez Almeida et al. (2016).

[a] No reliable EW_0(Hα) measurements are given due to pixel issues on the spectrum. Instead, we estimate EW_0(Hα) values from EW_0(Hβ) measurements and the empirical relation of EW_0(Hα)=5.47×EW_0(Hβ), which is obtained from metal-poor galaxies in the literature (Figure 10.2).

Note that we highlight values of equivalent widths and metallicities that satisfy the EMPG conditions in bold font.

Table 5.2 Estimated Probabilities for known EMPGs in the SDSS test catalog

#	ID	Answer	Estimate	P(EMPG)	P(Galaxy)	P(Star)	P(QSO)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	J012534.2+075924.5	EMPG	EMPG	**0.995**	0.000	0.005	0.000
2	J080758.0+341439.3	EMPG	EMPG	**0.584**	0.002	0.414	0.000
3	J082555.5+353231.9	EMPG	EMPG	**0.932**	0.005	0.063	0.000
4	J104457.8+035313.1	EMPG	EMPG	**0.995**	0.000	0.005	0.000
5	J141851.1+210239.7	EMPG	EMPG	**0.733**	0.025	0.243	0.000
6	J223831.1+140029.8	EMPG	EMPG	**0.643**	0.005	0.352	0.000
7	J013352.6+134209.5	EMPG	Star	0.035	0.382	**0.583**	0.000

(1): Number. (2): ID. (3): Real object class. This is an answer confirmed with spectroscopy data. (4): Object class estimated by the ML classifier.. (5)–(8): Probabilities of EMPG, galaxy, star, and QSO in the estimation of the ML classifier. The highest probability is highlighted in bold font. Here, for easy comparison, we show probabilities obtained before we apply the `softmax` function. Note that the probabilities shown in this table are just the linear expression, which are mathematically equivalent to the probabilities obtained after we apply the `softmax` function.

CHAPTER 6
SELECTION

In Section 5.3, we have confirmed that our object classifier works well with the SDSS test catalog. Thus, we expect that our object classifier can also select EMPGs in the HSC-SSP data. In Sections 6.1 and 6.2, we choose EMPG candidates from the HSC-SSP and SDSS source catalogs (Sections 4.2 and 4.3) with our ML classifier (Section 5.2). Hereafter, these candidates chosen from the HSC-SSP and SDSS source catalogs are called "HSC-EMPG candidates" and "SDSS-EMPG candidates", respectively.

6.1 EMPG Candidates from the HSC Data

In Section 4.2, we have created the HSC-SSP source catalog, which consists of 17,912,612 and 40,407,765 sources in the S17A and S18A source catalogs, respectively. As noted in Section 4.2, the sources selected from S17A and S18A data at this point are partly duplicated, but the duplication will be removed in the last step of the selection. In this section, we select EMPG candidates from the HSC-SSP source catalog in four steps described below.

In the first step, we coarsely remove sources based on blending, extendedness, and color before we apply our ML classifier. We remove sources whose photometry is strongly affected by back/foreground objects as follows. Fluxes of a source and a back/foreground object are measured at the central position of the source, and when a flux of the back/foreground object exceeds 50% of the source flux, the source is removed. We only select extended sources whose `extendedness_value` flags are 1 in all of the *griz* bands. The `hscPipe` labels a point source and an extended source as `extendedness_value`=0 and 1, respectively. The `hscPipe` defines a point source when a PSF magnitude and a `cmodel` magnitude matches ($m_{\rm psf}-m_{\rm cmodel}<0.0164$, Bosch et al., 2018). To save calculation time, we remove sources whose colors are apparently different from EMPGs. In other words, we choose sources that satisfy all of Equations (6.1)–(6.3) below.

$$r - i \ < \ -0.3 \tag{6.1}$$

$$i - z < 0.2 \tag{6.2}$$

$$g - r < -0.3125(r - i) + 0.1375 \tag{6.3}$$

After these selection criteria, 680 and 2,494 sources remain from the S17A and S18A data, respectively. This first-step source removal effectively reduces the calculation time in the ML classifier in the second step below.

In the second step, we apply the ML classifier constructed in Section 5.2 to the HSC-SSP source sample. The ML classifier selects 32 and 57 sources out of the 680 (S17A) and 2,494 (S18A), respectively.

In the third step, we remove transient objects by checking the g, r-band multi-epoch images. We measure fluxes in each epoch and calculate an average and a standard deviation of these flux values. If the standard deviation becomes larger than 25% of the average value, we regard the source as a transient object and eliminate it from the sample.

In the last step, we inspect a *gri*-composite image. Here we remove apparent H II regions inside a large star-forming galaxy, sources affected by a surrounding bright star, and apparently red sources. The apparently red sources are mistakenly selected due to an issue in the `cmodel` photometry. Indeed, they show red colors ($r - i > 0.0$) in the 1.0-arcsec aperture photometry, while they become blue in the `cmodel` photometry. In the inspection of multi-epoch images and *gri*-composite images, we have removed 20 and 36 sources from the S17A and S18A data, respectively.

Eventually, we thus obtain 12 and 21 HSC-EMPG candidates from the S17A and S18A catalogs, respectively. We find that 6 out of the HSC-EMPG candidates are duplicated between the S17A and S18A catalogs. Thus, the number of our independent HSC-EMPG candidates is 27 (=12+21−6). A magnitude range of the 27 HSC-EMPG candidates is $i = 19.3$–24.3 mag.

Out of the 27 candidates, we find 6 candidates that are selected in S17A but not selected again in S18A. Four out of the 6 candidates are slightly redder in S18A than in S17A and thus not selected in S18A. The other two are removed in S18A due to flags related to a cosmic ray or a nearby bright star. Such differences arise probably due to the different pipeline versions between S17A and S18A as described in Section

4.1. We check images and photometry of these 6 candidates individually. Then we confirm that these 6 candidates seem to have no problem as an EMPG candidate.

6.2 EMPG Candidates from the SDSS Data

In Section 4.3, we have constructed the SDSS source catalog consisting of 31,658,307 sources. In this section, we try to select EMPG candidates from the SDSS source catalog similarly to the HSC source catalog in Section 6.1.

First, we remove sources that have colors apparently different from EMPGs with equations (6.1)–(6.3). Then we apply our ML classifier to the SDSS source catalog and our classifier has selected 107 sources. Checking *gri*-composite images, we eliminate apparent H II regions in a spiral galaxy, sources affected by a surrounding bright star, and apparently red sources. We also remove sources if the corresponding composite image shows an apparent problem, which may be caused by an incorrect zero-point magnitude. In the visual inspection above, 21 sources have been removed. Finally, we derive 86 SDSS-EMPG candidates from the SDSS source catalog, whose i-band magnitudes range $i = 14.8$–20.9 mag.

Crossmatching with the SDSS spectra data, we find that 17 out of the 86 candidates already have an SDSS spectrum. These 17 spectra show strong nebular emission lines from galaxies at $z = 0.002$–0.026, 15 out of which have been already reported with a metallicity measurement in the range of 12+log(O/H)=7.44–8.22 (Kniazev et al., 2003, 2004; Izotov et al., 2007; Engelbracht et al., 2008; Izotov et al., 2012; Shirazi & Brinchmann, 2012; Sánchez Almeida et al., 2016; Izotov & Thuan, 2016). Seven out of the 15 galaxies satisfy the EMPG condition, 12+log(O/H)<7.69. All of the 6 EMPGs chosen in our classifier test (Section 5.3) are selected again here. Another object out of the 86 candidates is HSC J1429−0110, which is also selected as an HSC-EMPG candidate in Section 6.1.

CHAPTER 7
SPECTROSCOPY

In this section, we explain our spectroscopy for the 10 EMPG candidates carried out with 4 spectrographs of the Low Dispersion Survey Spectrograph 3 (LDSS-3) and the Magellan Echellette Spectrograph (MagE, Marshall et al., 2008) on Magellan telescope, the Deep Imaging Multi-Object Spectrograph (DEIMOS, Faber et al., 2003) on Keck-II telescope, and the Faint Object Camera And Spectrograph (FOCAS, Kashikawa et al., 2002) on Subaru telescope.

7.1 Magellan/LDSS-3 Spectroscopy

We conducted spectroscopy for the one HSC-EMPG candidate (HSC J1429−0110) on 2018 June 12 with LDSS-3 at Magellan telescope (PI: M. Rauch). We used the VPH-ALL grism with the $0\farcs75\times4'$ long-slit, which was placed at the offset position two-arcmin away from the center of the long-slit mask so that the spectroscopy could cover the bluer side. The exposure time was 3,600 seconds. The spectroscopy covered λ ∼3,700–9,500 Å with the spectral resolution of $R \equiv \lambda/\Delta\lambda \sim 860$. The A0-type standard star CD-32 9972 (RA=14:11:46.37, Dec.=−33:03:14.3 in J2000) was also observed. The sky was clear during the observation with seeing sizes of $0\farcs6$–$0\farcs9$.

7.2 Magellan/MagE Spectroscopy

We carried out spectroscopy for the two HSC-EMPG candidates (HSC J2314+0154 and HSC J1142−0038) and the six SDSS-EMPG candidates (SDSS J0002+1715, SDSS J1642+2233, SDSS J2115−1734, SDSS J2253+1116, SDSS J2310−0211, and SDSS J2327−0200) on 2018 June 13 with MagE of Magellan telescope (PI: M. Rauch). We used the echellette grating with the $0\farcs85\times10''$ or $1\farcs2\times10''$ longslits. The exposure time was 1,800 or 3,600 seconds, depending on luminosities of the candidates. The MagE spectroscopy covered λ ∼3,100–10,000 Å with the spectral resolution of $R \equiv \lambda/\Delta\lambda \sim 4{,}000$. The A0-type standard star CD-32 9972 (RA=14:11:46.37, Dec.=−33:03:14.3 in J2000) and the DOp-type standard star Feige 110 (RA=23:19:58.39,

Dec.=−05:09:55.8 in J2000) were also observed. The sky was clear during the observation with seeing sizes of 0.″8–1.″5.

7.3 Keck/DEIMOS Spectroscopy

We conducted spectroscopy for the one HSC-EMPG candidate (HSC J1631+4426) as a filler target on 2018 August 10 with DEIMOS of the Keck-II telescope (PI: Y. Ono). We used the multi-object mode with the 0.″8 slit width. The exposure time was 2,400 seconds. We used the 600ZD grating and the BAL12 filter with a blaze wavelength at 5,500 Å. The DEIMOS spectroscopy covered λ ∼3,800–8,000 Å with the spectral resolution of ∼4 Å in FWHM. The A0-type standard star G191B2B (RA=05:05:30.6, Dec.=+52:49:54 in J2000) was also observed. The sky was clear during the observation with seeing sizes of 0.″5.

7.4 Subaru/FOCAS Spectroscopy

We carried out deep spectroscopy for the one HSC-EMPG candidate (HSC J1631+4426) on 2019 May 13 with FOCAS installed on the Subaru telescope (PI: T. Kojima). HSC J1631+4426 was observed again with FOCAS in a longer integration time of 10,800 sec. We used the long slit mode with the 2.″0 slit width. The exposure time was 10,800 seconds (=3 hours). We used the 300R grism and the L550 filter with a blaze wavelength at 7,500 Å in a 2nd order. The FOCAS spectroscopy covered λ ∼3,400–5,250 Å with the spectral resolution of $R \equiv \lambda/\Delta\lambda$=400 with the 2.″0 slit width. The O-type subdwarf BD+28 4211 (RA=21:51:11.07, Dec.=+28:51:51.8 in J2000) was also observed as a standard star. The sky condition was clear during the observation with a seeing size of 0.″6.

The LDSS-3, MagE, DEIMOS, and FOCAS observations are summarized in Table 7.1.

Table 7.1 Summary of LDSS3, MagE, DEIMOS, and FOCAS observations

ID	R.A.	Dec.	slit width (arcsec)	exposure (sec)	seeing (arcsec)
LDSS3 observation					
HSC J1429−0110	14:29:48.61	−01:10:09.67	0.75	3,600	0.8
MagE observation					
HSC J2314+0154	23:14:37.55	+01:54:14.27	0.85	3,600	0.9
HSC J1142−0038	11:42:25.19	−00:38:55.64	0.85	3,600	0.8
SDSS J0002+1715	00:02:09.94	+17:15:58.65	1.2	1,800	1.5
SDSS J1642+2233	16:42:38.45	+22:33:09.09	0.85	1,800	1.0
SDSS J2115−1734	21:15:58.33	−17:34:45.09	0.85	1,800	1.1
SDSS J2253+1116	22:53:42.41	+11:16:30.62	1.2	1,800	1.2
SDSS J2310−0211	23:10:48.84	−02:11:05.74	1.2	1,800	1.0
SDSS J2327−0200	23:27:43.69	−02:00:55.89	1.2	1,800	1.0
DEIMOS observation					
HSC J1631+4426	16:31:14.24	+44:26:04.43	0.80	2,400	0.5
FOCAS observation					
HSC J1631+4426	16:31:14.24	+44:26:04.43	2.0	10,800	0.6

CHAPTER 8
REDUCTION AND CALIBRATION OF SPECTROSCOPIC DATA

We explain how we reduced and calibrated the spectroscopic data of Magellan/LDSS-3, Magellan/MagE, Keck/DEIMOS, Subaru/FOCAS in Sections 8.1–8.4, respectively.

8.1 LDSS-3 Data

We used the IRAF package to reduce and calibrate the data taken with LDSS-3 (Section 3.3.2). The reduction and calibration processes include the bias subtraction, flat fielding, one-dimensional (1D) spectrum subtraction, sky subtraction, wavelength calibration, flux calibration, and atmospheric-absorption correction. A one-dimensional spectrum was derived from an aperture centered on the blue compact component of our EMPG candidates. A standard star, CD-32 9972 was used in the flux calibration. The wavelengths were calibrated with the HeNeAr lamp. Atmospheric absorption was corrected with the extinction curve at Cerro Tololo Inter-American Observatory (CTIO). We used the CTIO extinction curve because Magellan Telescopes were located at Las Campanas Observatory, which neighbored the site of CTIO in Chili at a similar altitude.

In our spectroscopy, a slit was not necessarily placed perpendicular to the horizon (i.e., at a parallactic angle), but instead chosen to include extended substructure in our EMPG candidates. Thus, part of our spectra may have been affected by atmospheric refraction. Because targets are acquired with an R-band camera in the LDSS-3 observation, red light falls on the center of the slit while blue light might drop out of the slit. Thus, the atmospheric refraction can cause a wavelength-dependent slit loss. To estimate the wavelength-dependent slit loss $SL(\lambda)$ carefully, we made a model of the atmospheric refraction. We assumed the atmospheric refraction measured at La Silla in Chile (Filippenko, 1982), where the atmospheric condition was similar to Las Campanas in terms of the altitude and humidity. The model took into consideration a parallactic angle, a slit position angle, an air mass, and a seeing size at the time of

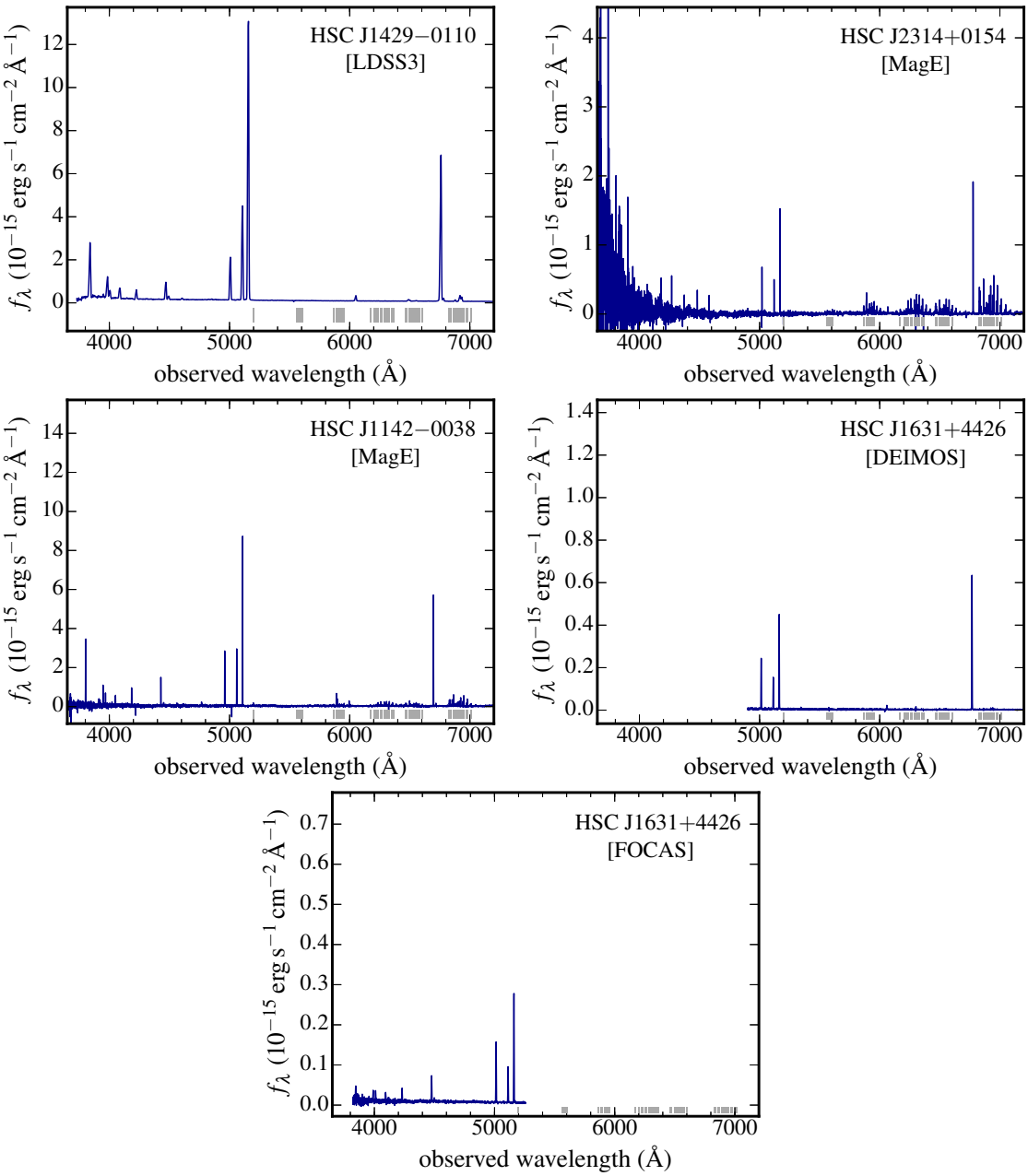

Figure 8.1 Spectra of our 4 HSC EMPGs. Positions of sky emission lines are indicated by gray vertical lines at the bottom of each panel. We mask part of strong sky lines that are not subtracted very well. Here we show two spectra of the same target, HSC J1631+4426, for which we conduct spectroscopy both with Keck/DEIMOS (red side, $\lambda \gtrsim 5{,}000$ Å) and Subaru/FOCAS (blue side, $\lambda \lesssim 5{,}000$ Å).

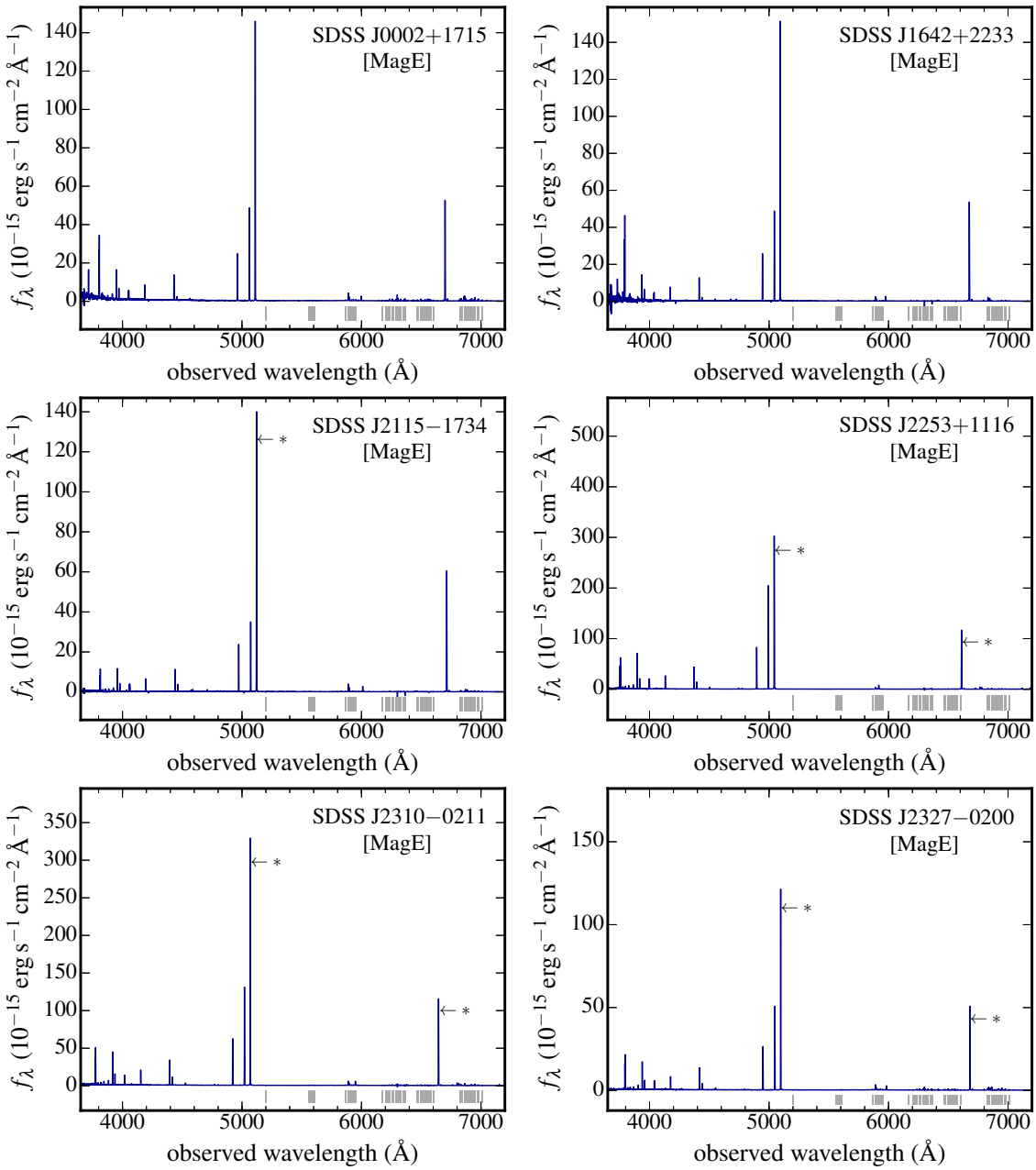

Figure 8.2 Same as Figure 8.1, but for our 6 SDSS EMPGs. We indicate part of emission lines with asterisks that may be underestimated because of the saturation. The saturation depends on the strength of an emission line and its position in each spectral order of the echellette spectroscopy because an edge (a center) of each order has a low (high) sensitivity.

exposures. An object size was broadened with a Gaussian convolution. We assumed the wavelength dependence for the seeing size $\propto \lambda^{-0.2}$, where the seeing size was measured in R-band. We integrated a model surface brightness $B(\lambda)$ on the slit to estimate an observed flux density F_λ^{obs} as a function of wavelength. Then we estimated the $SL(\lambda)$ by comparing the observed flux density F_λ^{obs} and total flux density F_λ^{tot} predicted in the model.

$$SL(\lambda) = 1 - F_\lambda^{obs}/F_\lambda^{tot} \tag{8.1}$$

The obtained SL values for HSC J1429−0110 were SL(4,000Å)=1.74 and SL(7,000Å) =1.61, for example. The SL ratio was SL(4,000Å)/SL(7,000Å)=1.08, correcting emission line ratios up to ∼10% between 4,000 and 7,000Å. Then we corrected the spectrum with SL(λ) and obtained the slit-loss corrected spectrum. We estimated multiple color excesses $E(B-V)$ from multiple pairs of Balmer lines and confirmed that these $E(B-V)$ values were consistent between them within error bars.

8.2 MagE Data

To reduce the raw data taken with MagE, we used the MagE pipeline from Carnegie Observatories Software Repository[1]. The MagE pipeline has been developed on the basis of the `Carpy` package (Kelson et al., 2000; Kelson, 2003). The bias subtraction, flat fielding, scattered light subtraction, two-dimensional (2D) spectrum subtraction, sky subtraction, wavelength calibration, cosmic-ray removal, 1D-spectrum subtraction were conducted with the MagE pipeline. Details of these pipeline processes are described on the web site of Carnegie Observatories Software Repository mentioned above. In the sky subtraction, we used a sky-line reference mask (i.e., a mask targeting a blank sky region with no object). One-dimensional spectra were subtracted by summing pixels along the slit-length direction on a 2D spectrum.

We conducted the flux calibration with the standard star, Feige 110, using IRAF routines. Wavelengths were calibrated with emission lines of the ThAr lamp. Spectra of each order were calibrated separately and combined with the weight of electron

[1] `https://code.obs.carnegiescience.edu`

counts to generate a single 1D spectrum. Atmospheric absorption was corrected in the same way as in Section 8.1.

In the MagE spectroscopy, we also placed a slit along a sub-structure of our EMPGs regardless of a parallactic angle. We also corrected the wavelength-dependent slit loss carefully in the same manner as the LDSS-3 spectroscopy described in Section 8.1.

8.3 DEIMOS Data

We used the IRAF package to reduce and calibrate the data taken with DEIMOS (Section 3.2.1). The reduction and calibration processes were the same as the LDSS-3 data explained in Section 8.1. A standard star, G191B2B was used in the flux calibration. Wavelengths were calibrated with the NeArKrXe lamp. Atmospheric absorption was corrected under the assumption of the extinction curve at Mauna Kea Observatories. It should be noted that part of flat and arc frame have been affected by stray light[2]. In our observation, a spectrum was largely affected in the wavelength range of λ =4,400–4,900 Å. Thus, we only used a spectrum within the wavelength range of λ >4,900 Å, which was free from the stray light. We ignore the effect of the atmospheric refraction here because we only use a red side (λ >4,900 Å) of DEIMOS data, which is insensitive to the atmospheric refraction. We also confirm that the effect of the atmospheric refraction is negligible with the models described in Section 8.1. In the DEIMOS data, we only used line flux ratios normalized to an Hβ flux. Emission line fluxes were scaled with an Hβ flux by matching an Hβ flux obtained with DEIMOS to one obtained with FOCAS (see Section 8.4). Note again that we have conducted spectroscopy for HSC J1631+4426 both with DEIMOS and FOCAS.

[2]As of September 2018, a cause of the stray light has not yet been identified according to a support astronomer at W. M. Keck Observatory (private communication). It is reported that the stray light pattern appears on a blue-side of CCD chips when flat and arc frames are taken with a grating tilted towards blue central wavelengths.

8.4 FOCAS Data

We used the IRAF package to reduce and calibrate the data taken with FOCAS (Section 7.4). The reduction and calibration processes were the same as the LDSS-3 data explained in Section 8.1. A standard star, BD+28 4211 was used in the flux calibration. Wavelengths were calibrated with the ThAr lamp. Atmospheric absorption was corrected in the same way as in Section 8.3. Our FOCAS spectroscopy covered λ ∼3,800–5,250 Å, which was complementary to the DEIMOS spectroscopy described in Section 8.3, whose spectrum was reliable only in the range of λ >4900 Å. We ignore the atmospheric refraction here because FOCAS is equipped with the atmospheric dispersion corrector. Because an Hβ line was observed twice in FOCAS and DEIMOS spectroscopy, we used an Hβ line flux to scale the emission line fluxes obtained in the DEIMOS observation (see Section 8.3).

We show spectra of the 4 HSC-EMPG candidates and 6 SDSS-EMPG candidates obtained with LDSS-3, MagE, DEIMOS, and FOCAS spectrographs in Figures 8.1 and 8.2.

CHAPTER 9
ANALYSIS

In this section, we explain the emission line measurement (Section 9.1) and the estimation of galaxy properties (Section 9.2). Here we estimate stellar masses, star-formation rates, emission-line equivalent widths, electron temperatures, gas-phase metallicities, and gas-phase ion abundance ratios of our 10 EMPG candidates confirmed in our spectroscopy.

9.1 Emission Line Measurements

We measure central wavelengths and emission-line fluxes with a best-fit Gaussian profile using the IRAF routine, `splot`. We also estimate flux errors, which originate from read-out noise and photon noise of sky+object emission. We measure observed equivalent widths (EWs) of emission lines with the same IRAF routine, `splot` and convert them into the rest-frame equivalent widths (EW_0). Redshifts are estimated by comparing the observed central wavelengths and the rest-frame wavelengths in the air of strong emission lines. Generally speaking, when the slit or fiber spectroscopy is conducted for a spatially resolved object, one obtains a spectrum only inside a slit or fiber, which may not represents an average spectrum of its whole system. However, we find that our emission-line estimation represents a whole galaxy that we are focusing on, because our metal-poor galaxies have a size comparable to or slightly larger than the seeing size of the observations. The sizes of our metal-poor galaxies will be discussed in Isobe et al. in prep.

Color excesses, E(B−V) are estimated with the Balmer decrement under the assumptions of the dust extinction curve given by Cardelli et al. (1989) and the case B recombination (T_e=20,000 K and electron density of n_e=100 cm^{-3}). We use several Balmer lines, except for lines with any flux measurement issues and/or possible contamination from another emission line. We find that most of our 10 EMPG candidates show E(B−V)$\sim$0 due to the low metallicities. We correct the observed fluxes against the dust extinction and estimate the intrinsic, dust-corrected fluxes. In the

flux estimation, we ignore the contribution of stellar atmospheric absorption around Balmer lines because our galaxies have very large equivalent widths compared to the expected absorption equivalent width. We have confirmed consistency between observed emission-line fluxes and those estimated in the process of photometric SED fitting. The photometric SED fitting is detailed in Section 9.2. We summarize redshifts and corrected fluxes in Tables 9.1 and 9.2, respectively.

Table 9.1 R.A., Dec., redshifts, and photometric magnitudes of our targets

#	ID	R.A.	Dec.	redshift	u	g	r	i	z	y
					mag	mag	mag	mag	mag	mag
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	HSC J1429−0110	14:29:48.61	−01:10:09.67	0.02980	—	18.14	18.65	19.38	19.47	18.92
2	HSC J2314+0154	23:14:37.55	+01:54:14.27	0.03265	—	21.94	21.95	22.76	22.57	22.38
3	HSC J1142−0038	11:42:25.19	−00:38:55.64	0.02035	—	21.39	21.62	22.42	22.28	22.01
4	HSC J1631+4426	16:31:14.24	+44:26:04.43	0.03125	—	21.84	21.88	22.52	22.75	22.39
5	SDSS J0002+1715	00:02:09.94	+17:15:58.65	0.02083	18.48	17.61	18.05	18.61	18.57	—
6	SDSS J1642+2233	16:42:38.45	+22:33:09.09	0.01725	18.50	17.99	18.38	19.01	19.14	—
7	SDSS J2115−1734	21:15:58.33	−17:34:45.09	0.02296	19.59	18.49	19.00	19.67	19.57	—
8	SDSS J2253+1116	22:53:42.41	+11:16:30.62	0.00730	17.91	16.62	17.07	18.08	18.12	—
9	SDSS J2310−0211	23:10:48.84	−02:11:05.74	0.01245	18.12	17.19	17.46	17.97	18.02	—
10	SDSS J2327−0200	23:27:43.69	−02:00:55.89	0.01812	19.02	18.16	18.47	19.26	19.25	—

(1): Number. (2): ID. (3): RA. (4): Dec. (5): Redshift. Typical uncertainties are $\Delta z \sim 10^{-6}$. (6)–(11): Magnitudes of *ugrizy* broad-band photometry. Photometry of our HSC-EMPGs is given with HSC `cmodel` magnitudes, while we use SDSS `model` magnitudes in the photometry of our SDSS-EMPG.

Table 9.2 Flux measurements

#	ID	[O II]3727	[O II]3729	$[O\,II]_{tot}$	H13	H12	H11	H10	H9
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	HSC J1429−0110	—	—	1.6033±0.0191	<0.0127	<0.0093	<0.0089	<0.0079	0.0526±0.0066
2	HSC J2314+0154	<0.1409	<0.1338	<0.1943	<0.1193	<0.1042	<0.1148	<0.0806	<0.0677
3	HSC J1142−0038	0.6806±0.0081	0.9182±0.0084	1.5988±0.0116	0.0326±0.0068	0.0363±0.0067	0.0665±0.0079	0.0439±0.0066	0.0613±0.0063
4	HSC J1631+4426	—	—	0.4191±0.0222	<0.0149	<0.0134	<0.0118	<0.0114	0.0396±0.0099
5	SDSS J0002+1715	0.7428±0.0050	1.0308±0.0051	1.7736±0.0071	<0.0033	<0.0049	<0.0036	0.0599±0.0030	0.0842±0.0025
6	SDSS J1642+2233	1.0137±0.0071	1.4656±0.0074	2.4793±0.0103	<0.0048	<0.0046	<0.0074	<0.0041	0.0662±0.0033
7	SDSS J2115−1734	0.3136±0.0025	0.4315±0.0028	0.7451±0.0038	0.0258±0.0019	0.0316±0.0030	0.0391±0.0019	0.0407±0.0018	0.0544±0.0017
8	SDSS J2253+1116	0.3944±0.0011	0.5506±0.0012	0.9450±0.0016	0.0236±0.0006	0.0340±0.0005	0.0408±0.0005	0.0559±0.0005	0.0759±0.0005
9	SDSS J2310−0211	0.4337±0.0010	0.5868±0.0012	1.0205±0.0016	0.0236±0.0006	0.0273±0.0006	0.0375±0.0006	0.0568±0.0006	0.0744±0.0006
10	SDSS J2327−0200	0.4507±0.0011	0.5996±0.0012	1.0503±0.0016	0.0248±0.0006	0.0336±0.0006	0.0399±0.0008	0.0529±0.0006	0.0719±0.0006

#	[Ne III]3869	[Ne III]3967	H7	Hδ	Hγ	[O III]4363	[Fe III]4658	HeII4686	[Ar IV]4711
(1)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
1	0.6178±0.0068	—	0.3146±0.0046 [a]	0.2105±0.0032	0.4471±0.0027	0.0890±0.0022	0.0093±0.0017	0.0136±0.0017	0.0200±0.0016
2	<0.0583	<0.0387	<0.0383	0.2599±0.0309	0.4636±0.0166	<0.0160	<0.0114	<0.0104	<0.0094
3	0.2933±0.0063	0.0774±0.0054	0.1716±0.0054	0.2736±0.0051	0.4820±0.0043	0.0596±0.0039	<0.0029	<0.0030	<0.0041
4	0.1848±0.0095	—	0.1728±0.0075 [a]	0.2424±0.0057	0.4298±0.0045	0.0753±0.0044	0.0089±0.0037	0.0226±0.0037	<0.0034
5	0.4664±0.0030	0.1500±0.0021	0.1627±0.0019	0.2604±0.0017	0.4692±0.0014	0.0638±0.0009	0.0066±0.0005	0.0086±0.0005	0.0079±0.0006
6	0.4387±0.0036	0.1200±0.0025	0.1503±0.0024	0.2608±0.0019	0.4448±0.0016	0.0654±0.0010	0.0064±0.0006	0.0120±0.0006	0.0044±0.0011
7	0.4046±0.0022	0.1305±0.0016	0.1452±0.0015	0.2463±0.0015	0.4515±0.0015	0.1342±0.0011	0.0086±0.0006	0.0263±0.0011	0.0187±0.0008
8	0.6412±0.0011	0.1882±0.0006	0.1612±0.0006	0.2567±0.0005	0.4652±0.0006	0.1413±0.0004	0.0036±0.0001	0.0030±0.0002	0.0212±0.0002
9	0.5318±0.0010	0.1576±0.0006	0.1628±0.0006	0.2652±0.0006	0.4839±0.0007	0.1480±0.0004	0.0045±0.0002	0.0048±0.0002	0.0170±0.0002
10	0.5072±0.0011	0.1514±0.0007	0.1745±0.0007	0.2593±0.0007	0.4739±0.0008	0.1281±0.0005	0.0063±0.0003	0.0078±0.0003	0.0151±0.0003

(1): Number. (2): ID. (3)–(36): Dust-corrected emission-line fluxes normalized to an Hβ line flux in the unit of $\mathrm{erg\,s^{-1}\,cm^{-2}\,\AA^{-1}}$. Upper limits are given with a 1σ level. Lines suffering from saturation or affected by sky emission lines are shown as no data here. $[O\,II]_{tot}$ represents a sum of [O II]3727 and [O II]3729 fluxes. If the spectral resolution is not high enough to resolve [O II]3727 and [O II]3729 lines, we only show $[O\,II]_{tot}$ fluxes. The notations of H13, H12,... H8, and H7 are higher orders of hydrogen Balmer lines, which correspond to transitions from hydrogen energy levels of n = 13,12,..., 8 and 7 to n = 2, respectively.

[a] A sum of [Ne III]3867 and H7 fluxes because they are blended due to the low spectral resolution.

Table 9.2 Flux measurements

#	ID	[Ar IV]4740	Hβ	[O III]4959	[O III]5007	HeI5876	[O I]6300	[S III]6312	[N II]6548
(1)	(2)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)
1	HSC J1429−0110	0.0095±0.0016	1.0000±0.0024	2.1126±0.0029	6.2972±0.0046	0.0951±0.0009	0.0304±0.0008	0.0110±0.0008	<0.0020
2	HSC J2314+0154	<0.0101	1.0000±0.0100	0.6962±0.0072	2.0769±0.0088	0.1292±0.0041	0.1196±0.0049	<0.0044	<0.0032
3	HSC J1142−0038	<0.0036	1.0000±0.0052	1.0276±0.0047	3.0814±0.0065	0.1111±0.0031	0.0529±0.0031	<0.0030	0.0199±0.0024
4	HSC J1631+4426	<0.0035	1.0000±0.0037	0.5656±0.0034	1.7453±0.0039	0.1021±0.0067	—	<0.0063	<0.0059
5	SDSS J0002+1715	0.0088±0.0006	1.0000±0.0016	1.9665±0.0019	5.9318±0.0032	0.1109±0.0005	0.0234±0.0004	0.0163±0.0003	0.0192±0.0003
6	SDSS J1642+2233	0.0078±0.0007	1.0000±0.0017	1.8353±0.0021	5.7288±0.0034	0.1087±0.0006	0.0263±0.0004	0.0191±0.0004	0.0171±0.0004
7	SDSS J2115−1734	0.0173±0.0007	1.0000±0.0019	1.6656±0.0022	—	0.1138±0.0007	0.0182±0.0005	0.0183±0.0005	0.0128±0.0004
8	SDSS J2253+1116	0.0169±0.0002	1.0000±0.0007	2.5060±0.0011	—	0.1117±0.0002	—	—	0.0116±0.0001
9	SDSS J2310−0211	0.0139±0.0002	1.0000±0.0009	2.1445±0.0012	—	0.1046±0.0003	0.0241±0.0002	0.0124±0.0002	0.0107±0.0002
10	SDSS J2327−0200	0.0112±0.0003	1.0000±0.0011	2.0095±0.0014	—	0.1128±0.0004	0.0279±0.0003	0.0152±0.0002	0.0131±0.0002

#	Hα	[N II]6584	HeI6678	[S II]6716	[S II]6731	HeI7065	[Ar III]7136	[O II]7320	[O II]7330
(1)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)
1	2.5560±0.0020	0.0527±0.0018	0.0307±0.0007	0.0817±0.0008	0.0631±0.0009	0.0285±0.0008	0.0588±0.0008	0.0153±0.0007	0.0097±0.0007
2	2.8066±0.0067	0.0212±0.0030	—	0.0522±0.0033	<0.0055	—	<0.0086	<0.0047	<0.0062
3	2.7209±0.0057	0.0864±0.0026	0.0275±0.0021	0.1624±0.0024	0.0852±0.0027	<0.0030	0.0552±0.0043	<0.0036	<0.0037
4	2.7067±0.0117	<0.0057	0.0247±0.0070	—	—	—	—	<0.0060	<0.0068
5	2.8048±0.0017	0.0568±0.0004	0.0299±0.0003	0.1119±0.0004	0.0597±0.0003	0.0258±0.0003	—	—	—
6	2.8131±0.0021	0.0538±0.0004	0.0269±0.0004	0.1059±0.0006	0.0830±0.0004	0.0145±0.0004	0.0644±0.0005	0.0189±0.0004	0.0150±0.0004
7	3.0097±0.0024	0.0336±0.0004	—	0.0567±0.0004	0.0506±0.0005	0.0397±0.0005	0.0508±0.0007	0.0120±0.0005	0.0092±0.0004
8	—	0.0329±0.0001	0.0297±0.0001	0.0647±0.0001	0.0507±0.0001	0.0356±0.0001	0.0508±0.0002	0.0155±0.0001	0.0103±0.0001
9	—	0.0278±0.0002	0.0267±0.0002	0.0770±0.0002	0.0560±0.0002	0.0321±0.0002	0.0474±0.0002	0.0128±0.0002	0.0099±0.0002
10	—	0.0321±0.0002	0.0293±0.0002	0.0878±0.0003	0.0649±0.0003	0.0383±0.0003	0.0536±0.0003	0.0169±0.0002	—

Continued.

Table 9.3 Parameters of our metal-poor galaxies

#	ID	EMPG?	$EW_0(H\beta)$	12+log(O/H)		$\log(M_\star)$	log(SFR)	$E(B-V)$	σ	D_{near}
			Å	direct	empirical	$M_\odot$	$M_\odot\,yr^{-1}$	mag	$km\,s^{-1}$	Mpc
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	HSC J1429−0110	*no*	$172.6^{+0.7}_{-0.6}$	8.30 ± 0.02	—	$6.55^{+0.13}_{-0.09}$	0.390 ± 0.001	0.31 ± 0.02	—	1.56
2	HSC J2314+0154	*yes*	$213.3^{+23.4}_{-17.6}$	—	7.224 ± 0.008	5.17 ± 0.01	-0.859 ± 0.001	0.27 ± 0.02	25.2±0.8	2.25
3	HSC J1142−0038	*no*	$111.9^{+1.4}_{-1.3}$	7.72 ± 0.03	—	$4.95^{+0.04}_{-0.01}$	-1.066 ± 0.002	$0.00^{0.01}_{0.00}$	25.5±0.8	2.28
4	HSC J1631+4426	*yes*	$123.5^{+3.5}_{-2.8}$	6.97 ± 0.03	7.197 ± 0.009	$5.89^{+0.10}_{-0.09}$	-1.442 ± 0.006	0.02 ± 0.02	—	1.54
5	SDSS J0002+1715	*no*	103.9 ± 0.2	8.22 ± 0.01	—	7.06 ± 0.03	-0.002 ± 0.001	$0.000^{0.004}_{0.000}$	30.7±0.9	3.05
6	SDSS J1642+2233	*no*	$153.7^{+0.5}_{-0.4}$	8.50 ± 0.01	—	$6.06^{+0.03}_{-0.13}$	-0.187 ± 0.001	$0.00^{0.01}_{0.00}$	32.5±1.0	0.49
7	SDSS J2115−1734	*yes*	$214.0^{+0.9}_{-0.8}$	7.69 ± 0.01	—	6.56 ± 0.02	0.188 ± 0.001	0.09 ± 0.03	31.4±0.9	17.69
8	SDSS J2253+1116	*no*	264.7 ± 0.3	8.005 ± 0.002	—	5.78 ± 0.01	-0.541 ± 0.001	$0.000^{0.002}_{0.000}$	23.0±0.7	14.33
9	SDSS J2310−0211	*no*	127.6 ± 0.2	$7.933^{+0.003}_{-0.004}$	—	6.99 ± 0.03	-0.162 ± 0.001	$0.000^{0.003}_{0.000}$	18.1±0.5	2.15
10	SDSS J2327−0200	*no*	111.0 ± 0.2	$7.908^{+0.004}_{-0.005}$	—	$6.51^{+0.02}_{-0.03}$	-0.180 ± 0.001	$0.000^{0.003}_{0.000}$	17.8±0.5	4.00

(1): Number. (2): ID. (3): Whether or not an object satisfies the EMPG definition, 12+log(O/H)<7.69. If yes (no), we write *yes* (*no*) in the column. (4): Rest-frame equivalent width of an Hβ emission line. (5)–(6): Gas-phase metallicity obtained with the direct method and the empirical relation of Izotov et al. (2019a). (7): Stellar mass. (8): Star-formation rate. (9): Color excess. (10): Velocity dispersion obtained from an Hβ emission line. An instrumental velocity dispersion is already removed. Note that the emission line broadening from a galaxy rotation is not eliminated. (11): Distance to the nearest neighborhood selected in the SDSS DR13 spectroscopic catalog.

9.2 Galaxy Properties

In this section, we estimate stellar masses, star-formation rates (SFRs), electron temperatures ($T_{\rm e}$), gas-phase metallicities (O/H), and gas-phase element abundance ratios of our EMPG candidates.

We estimate stellar masses of our EMPG candidates with the SED interpretation code BEAGLE (Chevallard & Charlot, 2016). We use *grizy* (*ugriz*) band photometry to estimate stellar masses of EMPG candidates selected from the HSC (SDSS) source sample. The constant star-formation history is assumed in the model. We run the BEAGLE code with 4 free parameters of stellar mass, maximum stellar age, ionization parameter, and metallicity, while we fix a redshift determined in our spectroscopy. We assume dust free to reduce the calculation time. When we add a dust extinction to free parameters in a stage of the rough parameter estimation, the dust extinction estimates actually become zero. Finally, we obtain estimates of stellar mass and maximum stellar age in the range of $\log(M_\star/M_\odot)$=4.95–7.06 and $t_{\rm age,max}$=3.4–51 Myr.

SFRs are estimated with the dust-corrected Hα fluxes under the assumption of the star-formation history of Kennicutt (1998). Here we assume that the Hα emission line is dominantly contributed by the photoionization caused by ionizing photon radiation from massive stars. If the Hα line is saturated, we use an Hβ line instead. The estimated SFRs of our EMPGs range $\log({\rm SFR}/M_\odot\,{\rm yr}^{-1})$=($-$0.86)–0.39.

We estimate electron temperatures of O^{2+} ($T_{\rm e}$(OIII)) and O^+, ($T_{\rm e}$(OII)), using line ratios of [O III]4363/5007 and [O II](3727+3729)/(7320+7330), respectively. A physical idea of the electron temperature measurement is described in Section 2.3.1. We use nebular physics calculation codes of PyNeb (Luridiana et al., 2015) to estimate electron temperatures. If an [O II]5007 line is saturated, we estimate an [O II]5007 flux with

$$[\mathrm{O\,III}]5007 = 2.98 \times [\mathrm{O\,III}]4959, \tag{9.1}$$

which is strictly determined by the Einstein A coefficient. If either of [O II]7320 or [O II]7330 line is detected, we estimate a total flux of [O II](7320+7330) with a relation of

$$[\mathrm{O\,II}]7330 = 0.56 \times [\mathrm{O\,II}]7320. \tag{9.2}$$

We have confirmed that Equation (9.2) holds with very little dependence on $T_{\rm e}$ and $n_{\rm e}$, using PyNeb (Luridiana et al., 2015). If none of [O II]7320,7330 line is detected, we estimate $T_{\rm e}$(OII) from an empirical relation of

$$T_{\rm e}({\rm O\,II}) = 0.7 \times T_{\rm e}({\rm O\,III}) + 3000, \tag{9.3}$$

which has been confirmed by Campbell et al. (1986) and Garnett (1992). We also assume

$$T_{\rm e}({\rm S\,III}) = 0.83 \times T_{\rm e}({\rm O\,III}) + 1700, \tag{9.4}$$

to estimate electron temperatures associated with S^{2+} ions (Garnett, 1992). We regard $T_{\rm e}$(OIII), $T_{\rm e}$(OII), and $T_{\rm e}$(SIII) as representative electron temperature associated with ions in high, intermediate, and low ionization states, respectively. The electron temperature measurements are summarized in Table 9.4.

We estimate metallicities based on electron temperature measurements, which are so-called $T_{\rm e}$-metallicities. The $T_{\rm e}$-metallicity is one of the most robust metallicity measurements based on the nebular physics, as explained in Section 2.3.2. Hereafter, we call the $T_{\rm e}$-metallicity just "metallicity" unless we describe explicitly. We also use PyNeb to estimate metallicities instead of the formulae of Section 2.3.2 because atomic data are updated in the PyNeb codes. We do not estimate a $T_{\rm e}$-based metallicity of HSC J2314+0154 because none of the $T_{\rm e}$(OIII), $T_{\rm e}$(OII), and $T_{\rm e}$(SIII) is estimated due to no detection of [O III]4363.

For comparison, we also estimate metallicities of HSC J2314+0154 and HSC J1631+4426 with a calibrator obtained by Skillman (1989). The Skillman (1989) calibrator is calibrated between an emission line index of

$$R_{23} \equiv \frac{[{\rm O\,II}]3727,3729 + [{\rm O\,III}]4959,5007}{{\rm H}\beta} \tag{9.5}$$

and $T_{\rm e}$-based metallicities of metal-poor galaxies. Recently, Izotov et al. (2019a) recalibrate a relation between R_{23} and $T_{\rm e}$-based metallicities with latest data of EMPGs. Figure 1-*a* of Izotov et al. (2019a) demonstrates that the Skillman (1989) calibrator well reproduces metallicities of EMPGs. The Skillman (1989) calibrator is applicable

in the low metallicity range of 12+log(O/H)<7.4, which corresponds to $\log(R_{23}) \lesssim 0.7$. Because our galaxies except for HSC J2314+0154 and HSC J1631+4426 do not satisfy $\log(R_{23}) \lesssim 0.7$, we do not estimate metallicities of the other 8 galaxies with the Skillman (1989) calibrator. The estimates of stellar masses, star-formation rates, and gas-phase metallicities are summarized in Table 9.3.

We estimate element abundance ratios of neon-to-oxygen (Ne/O), argon-to-oxygen (Ar/O), nitrogen-to-oxygen (N/O), and iron-to-oxygen (Fe/O) in a similar way to Izotov et al. (2006). First, we estimate ion abundance ratios of Ne^{2+}/H^+, Ar^{3+}/H^+, Ar^{2+}/H^+, N^+/H^+, and, Fe^{2+}/H^+ with the PyNeb codes. Because different ions reside in different parts of an H II region, we choose one of the T_e(OIII), T_e(SIII), and T_e(OII) to estimate ion abundances of each ion according to their ionization potential. We use T_e(OIII) to estimate abundances of O^{2+}, Ne^{2+}, and Ar^{3+}. We adopt T_e(SIII) in the estimation of Ar^{2+} abundances. We apply T_e(OII) to estimate abundances of the other lowly ionized ions, O^+, N^+, and Fe^{2+}. Table 9.6 summarizes types of electron temperature, T_e(OIII), T_e(SIII), and T_e(OII), adopted to estimate abundances of each ion. Second, we convert the ion abundances into element abundances with ionization correction factors (ICFs) of Izotov et al. (2006) shown below.

$$\frac{\mathrm{Ne}}{\mathrm{H}} = \frac{\mathrm{Ne}^+}{\mathrm{H}^+} + ICF(\mathrm{Ne}^+) \tag{9.6}$$

$$\frac{\mathrm{Ar}}{\mathrm{H}} = \frac{\mathrm{Ar}^{3+} + \mathrm{Ar}^{2+}}{\mathrm{H}^+} + ICF(\mathrm{Ar}^{3+} + \mathrm{Ar}^{2+}) \tag{9.7}$$

$$\frac{\mathrm{N}}{\mathrm{H}} = \frac{\mathrm{N}^+}{\mathrm{H}^+} + ICF(\mathrm{N}^+) \tag{9.8}$$

$$\frac{\mathrm{Fe}}{\mathrm{H}} = \frac{\mathrm{Fe}^{2+}}{\mathrm{H}^+} + ICF(\mathrm{Fe}^{2+}) \tag{9.9}$$

The ICFs are based on H II region models of Stasińska & Izotov (2003) and are given as a function of $v = O^+/(O^{2+}+O^+)$ or $w = O^{2+}/(O^{2+}+O^+)$. Finally, we obtain Ne/O, Ar/O, N/O, and Fe/O ratios by dividing Ne/H, Ar/H, N/H, and Fe/H by O/H (i.e., metallicity). We summarize the element abundance ratios in Table 9.5.

Table 9.4 Electron temperature measurements

#	ID	T_e(O III) (10^4 K)	T_e(O II) (10^4 K)
(1)	(2)	(3)	(4)
1	HSC J1429−0110	$1.301^{+0.014}_{-0.012}$	$0.883^{+0.015}_{-0.012}$
2	HSC J2314+0154	$1.484^{+0.043}_{-0.037}$	—
3	HSC J1142−0038	—	—
4	HSC J1631+4426	$2.331^{+0.095}_{-0.091}$	—
5	SDSS J0002+1715	$1.174^{+0.007}_{-0.005}$	—
6	SDSS J1642+2233	$1.200^{+0.006}_{-0.007}$	0.837 ± 0.007
7	SDSS J2115−1734	1.751 ± 0.007	$1.242^{+0.024}_{-0.016}$
8	SDSS J2253+1116	$1.472^{+0.002}_{-0.001}$	1.208 ± 0.006
9	SDSS J2310−0211	1.619 ± 0.002	$1.066^{+0.007}_{-0.006}$
10	SDSS J2327−0200	$1.557^{+0.004}_{-0.002}$	$1.147^{+0.009}_{-0.010}$

(1): Number. (2): ID. (3), (4): Electron temperatures of O^{2+} and O^{+}.

Table 9.5 Element abundance ratios

#	ID	log(Ne/O)	log(Ar/O)	log(N/O)	log(Fe/O)
(1)	(2)	(3)	(4)	(5)	(6)
1	HSC J1429−0110	$-0.659^{+0.006}_{-0.007}$	-2.605 ± 0.026	$-1.720^{+0.020}_{-0.023}$	$-1.968^{+0.091}_{-0.075}$
2	HSC J2314+0154	$-0.717^{+0.008}_{-0.010}$	< -2.217	$-1.277^{+0.019}_{-0.014}$	< -2.102
3	HSC J1142−0038	—	—	—	—
4	HSC J1631+4426	$-0.703^{+0.020}_{-0.019}$	—	< -1.572	$-1.177^{+0.173}_{-0.313}$
5	SDSS J0002+1715	$-0.711^{+0.003}_{-0.002}$	—	$-1.628^{+0.003}_{-0.004}$	$-2.106^{+0.035}_{-0.028}$
6	SDSS J1642+2233	$-0.778^{+0.004}_{-0.005}$	-2.707 ± 0.015	$-1.943^{+0.006}_{-0.009}$	$-2.324^{+0.027}_{-0.038}$
7	SDSS J2115−1734	$-0.784^{+0.004}_{-0.003}$	$-2.211^{+0.006}_{-0.011}$	$-1.427^{+0.009}_{-0.010}$	-1.584 ± 0.026
8	SDSS J2253+1116	-0.722 ± 0.001	-2.394 ± 0.002	-1.582 ± 0.003	$-2.081^{+0.020}_{-0.021}$
9	SDSS J2310−0211	$-0.783^{+0.001}_{-0.002}$	-2.445 ± 0.004	$-1.720^{+0.005}_{-0.004}$	$-2.041^{+0.019}_{-0.023}$
10	SDSS J2327−0200	$-0.756^{+0.001}_{-0.002}$	$-2.392^{+0.007}_{-0.006}$	$-1.634^{+0.006}_{-0.004}$	$-1.891^{+0.022}_{-0.017}$

(1): Number. (2): ID. (3)–(6): Element abundance ratios of Ne/O, Ar/O, N/O, and Fe/O. Upper limits are given with a 2σ confidence level.

Table 9.6 Electron Temperature Adopted in the Ionic Abundance Estimation

Ion	Ionization Potential (eV)	Temperature
(1)	(2)	(3)
O^{2+}	35.1	T_e(OIII)
Ne^{2+}	41.0	T_e(OIII)
Ar^{3+}	40.7	T_e(OIII)
Ar^{2+}	27.6	T_e(SIII)
O^{+}	13.6	T_e(OII)
N^{+}	13.6	T_e(OII)
Fe^{2+}	16.2	T_e(OII)

(1) Ions. (2) Ionization potentials needed to produce the corresponding ions. (3) Type of electron temperature adopted in the ionic abundance estimation.

CHAPTER 10
RESULTS AND DISCUSSIONS

In Section 10.1, we describe the results of the object class identification for our HSC-EMPG and SDSS-EMPG candidates, and show the distribution of $EW_0(H\beta)$ and metallicity to characterize our sample. We also investigate the cosmic number density of our metal poor galaxies (Section 10.2) and their environment (Section 10.3). We show the stellar mass and SFR ($M_\star$-SFR) and the stellar-mass and metallicity ($M_\star$-Z) relations of our EMPG candidates in Sections 10.4 and 10.5, respectively. The velocity dispersions of our sample is presented and discussed in Section 10.6. In Section 10.7, we discuss the possibility of the AGN/shock contribution on the diagram of [N II]/Hα and [O III]/Hβ emission line ratios, so-called the Baldwin-Phillips-Terlevich diagram (BPT diagram, Baldwin et al., 1981). Section 10.8 shows element abundance ratios of Ne/O, Ar/O, N/O, and Fe/O and discusses physical mechanisms that may be able to explain the obtained abundance ratios. In Section 10.9, we investigate the ISM ionization state of our metal-poor galaxies through the emission lines of [O II]3727, [Ar III]4740, [O III]5007, [Ar IV]7136, He II4686, and Hβ.

10.1 Object Class Identification

As described in Chapter 7, we conducted spectroscopy for 4 out of 27 HSC-EMPG candidates and 6 out of 86 SDSS-EMPG candidates. We find that all of the observed 10 EMPG candidates are identified as real galaxies with strong emission lines. We show spectra of the 4 HSC-EMPG candidates and 6 SDSS-EMPG candidates that exhibit strong emission lines in Figures 8.1 and 8.2, respectively. Two spectra are shown for HSC J1631+4426 because we have conducted spectroscopy both with Keck/DEIMOS and Subaru/FOCAS for this object.

Figure 10.1 shows the distribution of metallicity and $EW_0(H\beta)$ of our EMPG candidates (red stars). We find that our sample covers a wide range of metallicity, 12+log(O/H)=7.0–8.3 (i.e., 0.02–0.3 $Z_\odot$) and that 3 out of our 10 candidates satisfied the EMPG criteria of 12+log(O/H)<7.69, while the other 7 candidates do not satisfy

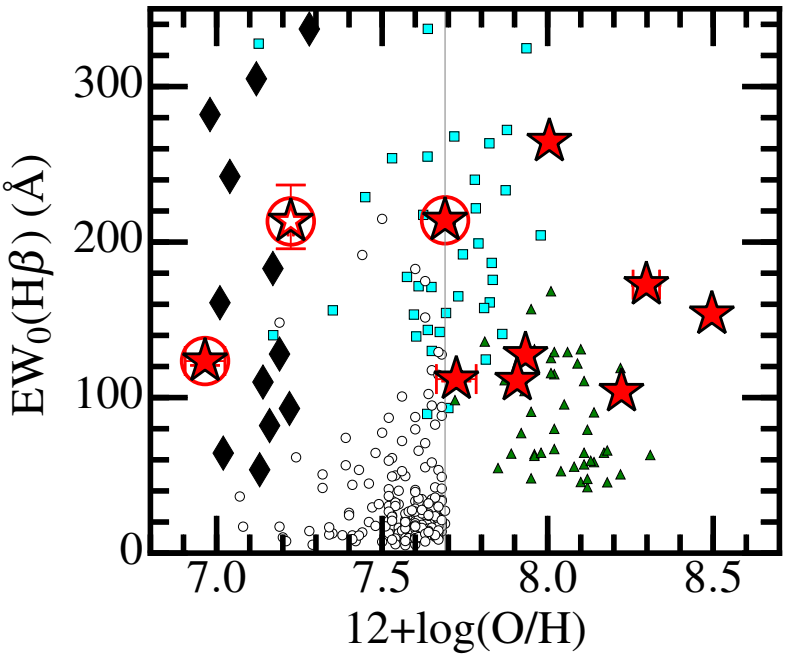

Figure 10.1 $EW_0(H\beta)$ as a function of metallicity of our metal-poor galaxies from HSC-EMPG and SDSS-EMPG source catalogs (red stars). Galaxies that satisfy the EMPG condition in our metal poor galaxy sample are marked with a large circle. The solid lines indicate criterion of EMPGs, 12+log(O/H)<7.69. Open star indicates a galaxy whose metallicity is obtained with the empirical relation of Skillman (1989), not with the T_e method due to no detection of an [O III]4363 line (Section 9.2). We also present GPs (Yang et al., 2017a, green triangle), BBs (Yang et al., 2017b, cyan square), and metal-poor galaxies (Sánchez Almeida et al., 2016, open circle) from the literature for comparison. With diamonds, we show representative metal-poor galaxies, J0811+4730 (Izotov et al., 2018b), SBS0335−052 (e.g., Izotov et al., 2009), AGC198691 (Hirschauer et al., 2016), J1234+3901 (Izotov et al., 2019b), LittleCub (Hsyu et al., 2017), DDO68 (Pustilnik et al., 2005; Annibali et al., 2019), IZw18 (e.g., Izotov & Thuan, 1998; Thuan & Izotov, 2005), and LeoP (Skillman et al., 2013) of 12+log(O/H)∼7.0–7.2.

the criteria. Remarkably, one of the 3 EMPGs, HSC J1631+4426 has a metallicity of 12+log(O/H)=6.97 (i.e., 0.019 $Z_\odot$), which is one of the lowest metallicities reported ever (e.g., J0811+4730, 12+log(O/H)=6.98, Izotov et al., 2018b). We also find that 2 out of the 3 EMPGs, HSC J2314+0154 and HSC J1631+4426, are selected from the HSC data and have i-band magnitudes of 22.8 and 22.5 mag. We argue that these 2 faint EMPGs are selected thanks to the deep HSC data, which suggests that the deep HSC data are advantageous to select very faint EMPGs. It should be also noted that the other 7 galaxies out of the EMPG definition still show a low metallicity ($Z/Z_\odot \sim 0.1$–0.3).

In Figure 10.1, we also show GPs (Yang et al., 2017a, green triangle), BBs (Yang et al., 2017b, cyan square), and local metal-poor galaxies (Sánchez Almeida et al., 2016, SA16 hereafter, open circle) for comparison. We also compare them with the representative metal-poor galaxies in the range of 12+log(O/H)$\sim$7.0–7.2, J0811+4730 (Izotov et al., 2018b), SBS0335−052 (e.g., Izotov et al., 2009), AGC198691 (Hirschauer et al., 2016), J1234+3901 (Izotov et al., 2019b), LittleCub (Hsyu et al., 2017), DDO68 (Pustilnik et al., 2005; Annibali et al., 2019), IZw18 (e.g., Izotov & Thuan, 1998; Thuan & Izotov, 2005), and LeoP (Skillman et al., 2013) with diamonds in Figure 10.1. We use $EW_0(H\beta)$ instead of $EW_0(H\alpha)$, which is used to select high-EW EMPGs in the models (Section 5.2.3), because some of Hα emission lines are saturated in our observation. Note that the EW condition used in the model, $EW_0(H\alpha)$>1,000 Å, corresponds to $EW_0(H\beta)$>200 Å under the assumption of the tight correlation between $EW_0(H\alpha)$ and $EW_0(H\beta)$ as demonstrated in Figure 10.2. We find that our metal-poor galaxy sample covers a high $EW_0(H\beta)$ range of $\sim$100–300 Å. Most of BBs and the representative metal-poor galaxies also show high equivalent widths of $\sim$100–300 Å. These high $EW_0(H\beta)$ values ($\sim$100–300 Å) are in contrast to the metal-poor galaxy sample of SA16, in which most galaxies show $EW_0(H\beta) \lesssim 100$ Å. As suggested in Figure 5.8, galaxies that consist of younger stellar population have higher equivalent widths of Balmer emission lines. Thus, the high $EW_0(H\beta)$ values may suggest that our metal-poor galaxies, BBs, and the representative metal-poor galaxies possess younger stellar population than the metal-poor galaxies of SA16.

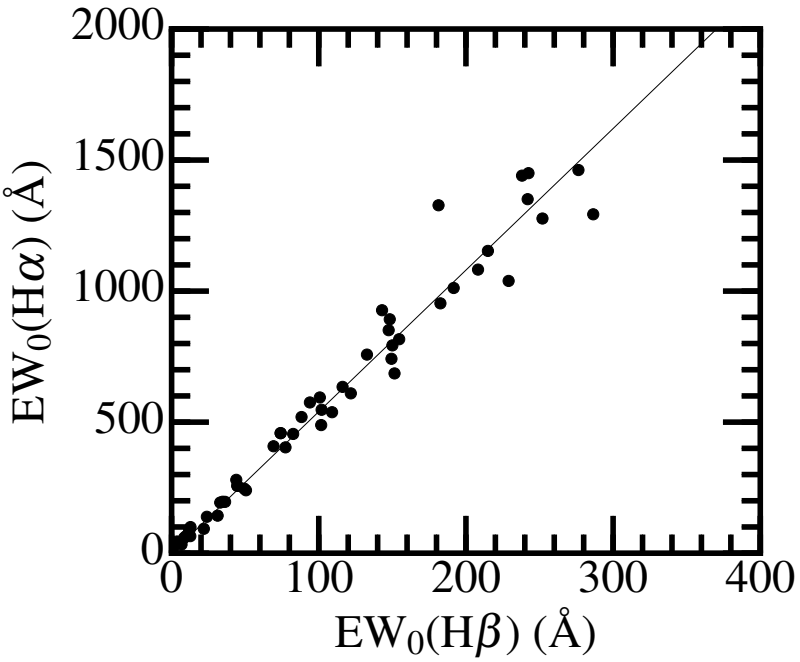

Figure 10.2 Relation between rest-frame equivalent widths of Hα and Hβ for metal-poor galaxies in the literature (Kunth & Östlin, 2000; Kniazev et al., 2003; Guseva et al., 2007; Izotov & Thuan, 2007; Izotov et al., 2009; Pustilnik et al., 2010; Izotov et al., 2012; Pilyugin et al., 2012; Sánchez Almeida et al., 2016; Guseva et al., 2017). The best least-square fit is shown with a solid line, which is $\mathrm{EW}_0(\mathrm{H}\alpha)$=5.47×$\mathrm{EW}_0(\mathrm{H}\beta)$. Because metal-poor galaxies are less dusty, a flux ratio of $F(\mathrm{H}\alpha)/F(\mathrm{H}\beta)$ becomes almost constant ($\sim$2.7–3.0, determined by the case B recombination) in the most case. In addition, a ratio of continuum level at Hα and Hβ, $f_{\lambda,0}$(6563 Å)/$f_{\lambda,0}$(4861 Å) always becomes $\sim$0.5 because the continuum slope differs little among metal-poor galaxies at wavelength of λ>4000 Å. Thus, the tight relation between $\mathrm{EW}_0(\mathrm{H}\alpha)$ and $\mathrm{EW}_0(\mathrm{H}\beta)$ is only applicable to metal-poor galaxies.

10.2 Number Density

We roughly estimate the number density of our metal-poor galaxies selected from the HSC and SDSS data. The HSC and SDSS broad-band filters can select EMPGs at $z < 0.035$ and $z < 0.030$, respectively, which correspond to 149 and 128 Mpc in cosmological physical distance. The redshift range difference ($z < 0.035$ and $z < 0.030$) is caused by the different response curves of the HSC and SDSS broad-band filters. Because we have selected 27 (86) EMPG candidates from the HSC (SDSS) data, whose effective observation area is 509 (14,355) deg^2, within $z < 0.035$ ($z < 0.030$), we obtain the number density, 1.5×10^{-4} (2.8×10^{-5}) Mpc^{-3}, from the HSC (SDSS) data. As suggested by previous surveys (Cardamone et al., 2009; Yang et al., 2017b), we confirm again that the metal-poor galaxies with strong emission lines are rare in the local universe. We also find that the number density of metal-poor galaxies is $\times10$ times higher in the HSC data than in the SDSS data. This difference is explained by the facts that fainter galaxies are more abundant and that our HSC metal-poor galaxies (median: $i \sim 22.5$ mag) are ~ 30 times fainter than our SDSS metal-poor galaxies (median: $i \sim 18.8$ mag).

10.3 Environment

To characterize the environment of our metal-poor galaxies, we compare distances to the nearest neighborhood galaxy (D_{near}) of our metal-poor galaxies and local, typical SFGs. The local, typical SFGs are randomly chosen from galaxies at $z = 0.03$–0.05 in the SDSS DR13 spectroscopic catalog. We sample 1,000 galaxies as local, typical SFGs here. We calculate distances from an object to surrounding galaxies on the basis of the SDSS DR13 spectroscopic catalog and identify the nearest neighbor. The D_{near} values of our metal-poor galaxies range from 0.49 to 17.69 Mpc, which are summarized in Table 9.3. Figure 10.3 compares the D_{near} distributions of our metal-poor galaxies and local, typical SFGs. An average D_{near} value of our metal-poor galaxies is 3.83 Mpc, which is about 2.5 times larger than that of local, typical SFGs (1.52 Mpc). We also find that 9 out of our 10 metal-poor galaxies have D_{near} values larger than the

average of local, typical SFGs (i.e., D_{near}>1.52 Mpc). Statistically, a Kolmogorov-Smirnov test rejects the null hypothesis (i.e., the distributions of the two samples are the same) with a p-value of 1.9×10^{-3}, suggesting that these distributions are significantly different. Thus, we conclude that our metal-poor galaxies exist in the relatively isolated environment compared to the local, typical SFGs. According to Yang et al. (2017b), their BB galaxy sample also shows significantly larger distances to their nearest neighborhood. Filho et al. (2015) also report that most of metal poor galaxies are found in low-density environments. In addition, as describe above, 9 out of our 10 metal-poor galaxies show D_{near} values larger than 1.5 Mpc, which the size of Local Group. Thus, our result also suggests that most of our metal-poor galaxies do not belong to galaxy groups as large as Local Group. These observational results suggest that the metal-poor galaxies have started an intensive star-formation in an isolated environment.

The formation mechanism of metal-poor galaxies in the local universe is an open question. One possible explanation is that pristine (or extremely metal poor) gas had been prevented from forming stars until recently in low-density regions by cosmic UV background, and the star formation was suddenly triggered by the collapse or inflow of metal-poor gas. Sánchez Almeida et al. (2013, 2015) has investigated tadpole galaxies, which is one of the typical metal-poor galaxy populations, and found that a blue head of tadpole galaxies has significantly lower metallicity than the rest of the galaxy body by factors of 3–10. The Northern Extended Millimeter Array (NOEMA) mm-wave interferometer has revealed that a tadpole galaxy possesses molecular gas at its head (Elmegreen et al., 2018). Filho et al. (2013) demonstrate that metal poor galaxies are surrounded by asymmetric H I gas, which can be shaped by the accretion of metal-poor gas. However, further investigations are necessary to unveil the mechanisms that prevent star formation for a long time and suddenly trigger the intensive star formation from observational and theoretical points of view.

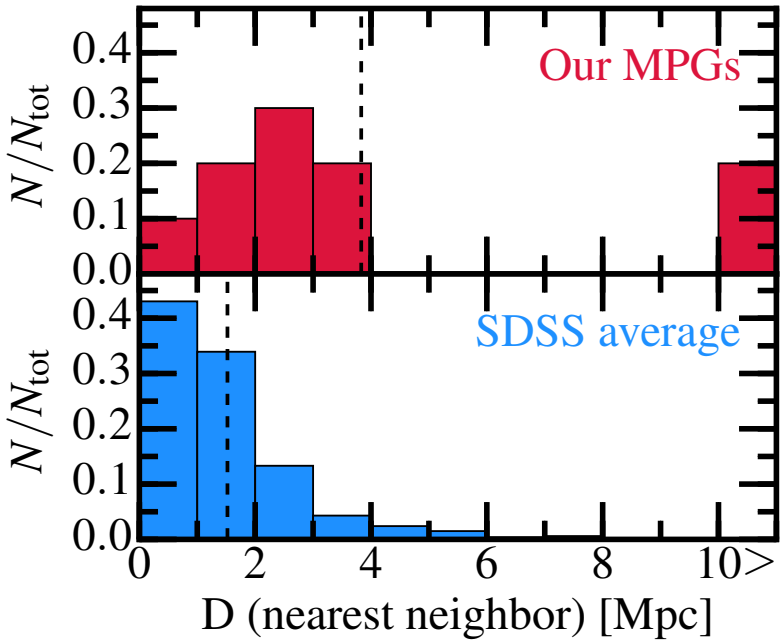

Figure 10.3 Normalized histogram of the nearest neighborhood distances of our ten metal-poor galaxies (top panel) and local, typical SFGs obtained from SDSS (bottom panel). The number of galaxies in each bin (N) is normalized by the total number of galaxies ($N_{\rm tot}$). The dashed lines indicate average values of the nearest neighborhood distances of our metal-poor galaxies (3.83 Mpc) and typical SFGs (1.52 Mpc). The bin between 10 and 11 represents the number of galaxies whose nearest neighborhood distance is beyond 10 Mpc.

10.4 $M_\star$-SFR Relation

Figure 10.4 shows SFRs and stellar masses of our metal-poor galaxies, BBs, GPs, metal-poor galaxies of SA16, and the representative metal-poor galaxies from the literature. Our metal-poor galaxies, BBs, GPs, and the most of representative metal-poor galaxies have higher SFRs than typical $z \sim 0$ galaxies (i.e., $z \sim 0$ star-formation main sequence) for a given stellar mass. In other words, they have a higher specific SFR (sSFR) than those given by $z \sim 0$ main sequence. Especially, our metal-poor galaxies have low stellar mass values in the range of $\log(M_\star/M_\odot)$<6.0, which are lower than BBs, GPs, and metal-poor galaxies of SA16.

The stellar masses of our metal-poor galaxies fall on the typical stellar-mass range of globular clusters, i.e., $\log(M_\star/M_\odot)$∼4–6. Thus, one may guess that these metal-poor galaxies might be globular clusters that have been formed very recently. However, further investigation is necessary to understand the association between metal-poor galaxies and globular clusters, which will be discussed in Isobe et al. in prep.

The solid lines in Figure 10.4 show the main sequences of typical galaxies at $z \sim 2$ (Shivaei et al., 2016) and $z \sim 4$–5 (Shim et al., 2011). As suggested by solid lines, the main sequence evolves towards high SFR for a given stellar mass with increasing redshift. Our metal-poor galaxies have higher SFRs for a given $M_\star$ than the $z \sim 0$ main sequence, falling onto the extrapolation of the $z \sim 4$–5 main sequence. Our metal-poor galaxies have as high sSFR values as low-$M_\star$ galaxies at $z \gtrsim 3$ and local Lyman continuum (LyC) leakers (e.g., $\log(\mathrm{sSFR/Gyr^{-1}})$∼1–3, Ono et al., 2010; Vanzella et al., 2017; Izotov et al., 2018b; Shim et al., 2011). Table 10.1 summarizes sSFR values of our metal-poor galaxies and other galaxies populations from the literature for reference. Based on the high sSFRs, we suggest that our metal-poor galaxies are undergoing intensive star formation comparable to the low-$M_\star$ SFGs at $z \gtrsim 3$.

Note that our SFR estimates are obtained under the simple assumption of Kennicutt (1998) because we only have optical observational results for now. The simple assumption can be broken in the very young ($\lesssim$10 Myr), metal-poor, low-$M_\star$ galaxies because the conversion factor is sensitive to the IMF, the star-formation history, the

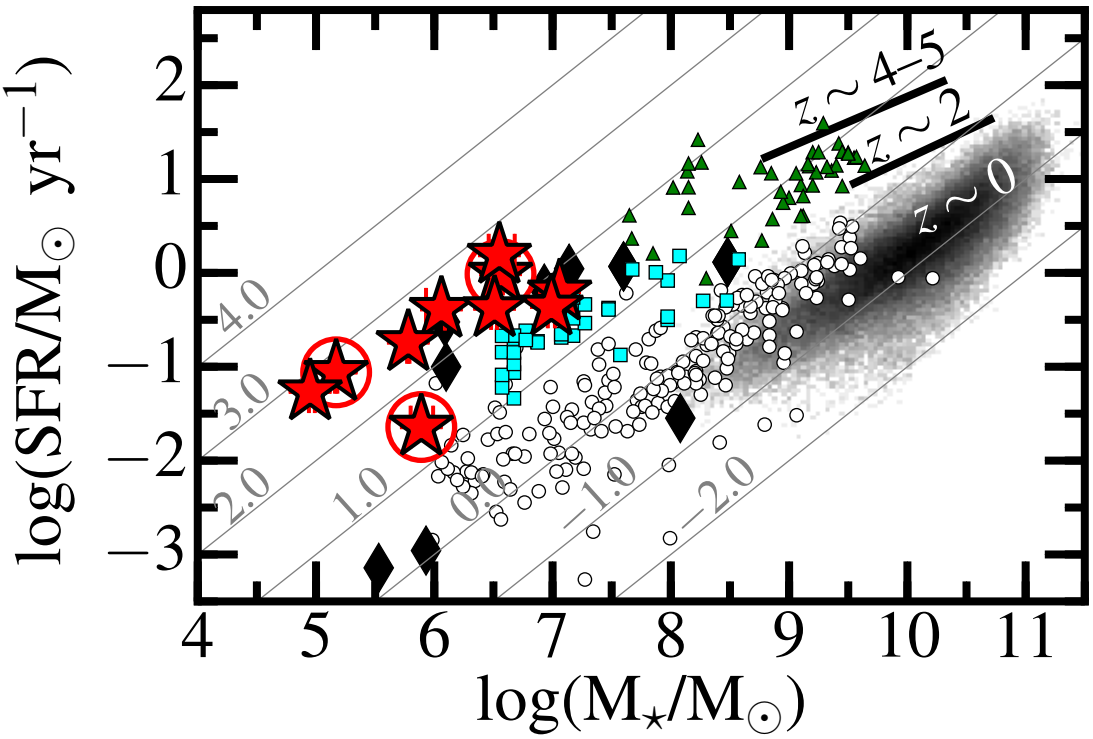

Figure 10.4 Stellar mass and SFR of our metal-poor galaxies with GPs, BBs, and local metal-poor galaxies. Symbols are the same as in Figure 10.1. We also show the stellar-mass and SFR distribution of typical $z \sim 0$ SFGs (i.e., $z \sim 0$ main sequence; black mesh), which we derive from the value-added catalog of SDSS DR7 (Kauffmann et al., 2003; Brinchmann et al., 2004; Salim et al., 2007). The solid lines represent the main sequences at $z \sim 2$ and $z \sim 4$–5 (Shivaei et al., 2016; Shim et al., 2011). The SFRs of Shivaei et al. (2016) and Shim et al. (2011) are estimated based on the Hα flux. We convert stellar masses and SFRs derived from the literature into those of the Chabrier (2003) IMF, applying conversion factors obtained by Madau & Dickinson (2014). Gray solid lines and accompanying numbers indicate $\log(\mathrm{sSFR/Gyr^{-1}})$=($-2.0$, -1.0,..., 4.0). The stellar masses and SFRs of the representative metal-poor galaxies are derived from the literature (Izotov et al., 2018b; Hirschauer et al., 2016; Izotov et al., 2019b; Hsyu et al., 2017; Hunt et al., 2015; Sacchi et al., 2016; Rhode et al., 2013; Annibali et al., 2013)

Table 10.1 Typical Values of Specific SFR

Population	log(sSFR) (Gyr^{-1})	redshift	Ref.
(1)	(2)	(3)	(4)
Our EMPGs	2.45	0.007–0.03	**This work**
EMPGs (SDSS DR7)	0.34	$\lesssim 0.1$	SA16
BBs	1.39	~ 0.05	Y17b
GPs	1.38	~ 0.3	Y17a
LyC leaker ($f_{\mathrm{esc}}^{\mathrm{LyC}}$=0.46)	1.29	0.37	I18
Main Sequence ($z \sim 0$)	-0.20	~ 0	SDSS DR7
Main Sequence ($z \sim 2$)	~ 0.0–0.5	~ 2	S16
Main Sequence ($z \sim 4$–5)	~ 1.0–1.5	~ 4–5	S11
low-$M_\star$ SFG ($z \sim 3$)	1.10/1.80	3.12	V16
Little Blue Dots	$\gtrsim$2.0	2–5	E17
LAEs ($z = 5.7$)	3.05	5.7	O10
LAEs ($z = 6.6$)	3.05	6.6	O10

(1) Galaxy Population. (2) Average of sSFR in the unit of $\log(\mathrm{Gyr}^{-1})$. We calculate a linear average of each sample here. (3) Typical redshift. (4) References of sSFR—SDSS DR7: (Kauffmann et al., 2003; Brinchmann et al., 2004; Salim et al., 2007), SA16: Sánchez Almeida et al. (2016), Y17b: Yang et al. (2017b), Y17a: Yang et al. (2017a), I18: Izotov et al. (2018a), S16: Shivaei et al. (2016), S11: Shim et al. (2011), V16: Vanzella et al. (2017), E17: Elmegreen & Elmegreen (2017), O10: Ono et al. (2010).

metallicity, and the escape fraction and dust absorption of ionizing photons (Kennicutt, 1998). Other SFR uncertainties may arise from additional ionizing photon sources, such as a low-luminosity AGN, shock-heated gas, galactic outflows, and X-ray binaries, which are not included in the stellar synthesis models used in the calibration between the SFR and luminosity of hydrogen recombination lines (Kennicutt, 1998). Further multi-wavelength observations are required to understand the star-formation rate, history, and mechanism of very young, metal-poor, low-$\mathrm{M}_\star$ galaxies.

10.5 $M_\star$-Z Relation

Figure 10.5 exhibits a mass-metallicity ($M_\star$-Z) relation of our metal-poor galaxies. Our metal-poor galaxies are located around the low-mass end of $\log(M_\star/M_\odot)$=5–7 among metal-poor galaxy samples of BBs, GPs, the S16 metal-poor galaxies, and the representative metal-poor galaxies in Figure 10.5. Metallicities of our metal-poor galaxies extend in a relatively wide range, 12+log(O/H)$\sim$7.0–8.3. The gray shaded regions in Figure 10.5 represent the 68 and 95-percentile distributions of local SFGs of

Zahid et al. (2012), who have reported that the metallicity scatter of galaxies becomes larger with decreasing metallicity for a given mass. Although the extrapolation is applied below $\log(M_\star/M_\odot)$=8.4 here, 5 out of our metal-poor galaxies fall in the 68-percentile distribution of the local $M_\star$-Z relation.

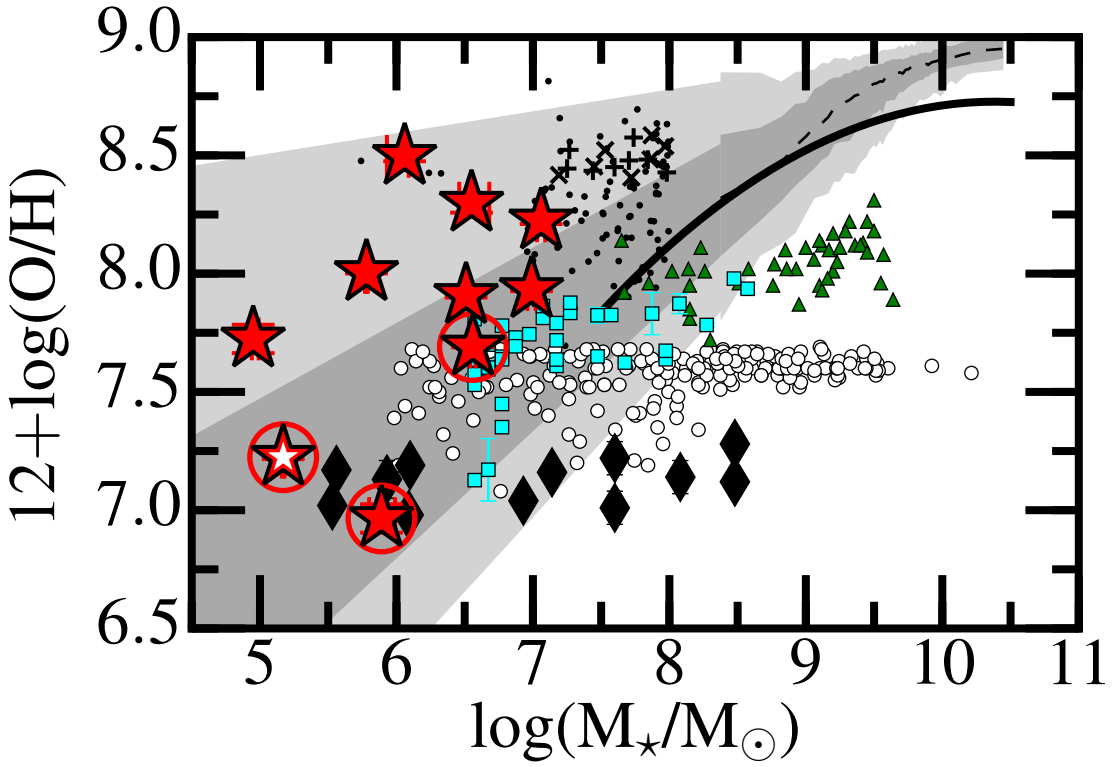

Figure 10.5 Mass-metallicity relation of our metal-poor galaxies. Symbols are the same as in Figure 10.1. The solid and dashed lines indicate averaged local SFGs given by Andrews & Martini (2013) and Zahid et al. (2012) from SDSS data, respectively. The dark-gray and light-gray shaded regions represent the 68 and 95-percentile distributions of SFGs of Zahid et al. (2012), although the extrapolation is applied below $\log(M_\star/M_\odot)$=8.4. We also show relatively metal-enriched dwarfs of Peeples et al. (2008, cross) and Zahid et al. (2012, plus) from SDSS, as well as DEEP2 galaxies of Zahid et al. (2012, dot) in the stellar mass range of $\log(M_\star/M_\odot)$<8.0. Typical metallicity error of our metal-poor galaxies is Δ(O/H)$\sim$0.01 dex.

Interestingly, we find that the other 5 metal-poor galaxies of ours are located above the 68-percentile distribution given by Zahid et al. (2012), i.e., higher metallicities for a given stellar mass. We refer the 5 metal-poor galaxies located above the 68-percentile distribution as "above-MZ galaxies" hereafter. Our above-MZ galaxies

have moderate metallicities of 12+log(O/H)$\sim$8.0 in spite of their very low-M$_\star$ (i.e., log($M_\star/M_\odot$)=5–7). A possible explanation of these above-MZ galaxies has been given by Zahid et al. (2012) and Peeples et al. (2008). In Figure 10.5, we also show the low-z galaxy samples of Zahid et al. (2012) and Peeples et al. (2008) in the stellar mass range of log($M_\star/M_\odot$)<8.0. In a sample from the DEEP2 survey, Zahid et al. (2012) have found galaxies with metallicity higher than the local $M_\star$-Z relation (Figure 10.5) and a higher SFR for a given stellar mass, which is similar to our above-MZ galaxies. Zahid et al. (2012) have also found counterpart galaxies of their DEEP2 galaxies (i.e., above both the $M_\star$–Z and $M_\star$–SFR relations) in the SDSS data (Figure 10.5). Zahid et al. (2012) have argued that their DEEP2 and SDSS galaxies may be transitional objects from gas-rich dwarf irregulars to gas-poor dwarf spheroidals and ellipticals suggested by Peeples et al. (2008). With SDSS data, Peeples et al. (2008) has also investigated local galaxies whose metallicities are higher than the local $M_\star$-Z relation. Unlike our above-MZ galaxies and Zahid et al. (2012) galaxies, the Peeples et al. (2008) sample show redder colors and lower SFRs consistent with the local $M_\star$–SFR relation. Peeples et al. (2008) have claimed that the Peeples et al. (2008) galaxies may be in a later stage of the transition from gas-rich dwarf irregulars to gas-poor dwarf spheroidals and ellipticals, and that the gas deficit leads to the low SFRs and high metallicity. It should be noted that Zahid et al. (2012) and Peeples et al. (2008) galaxies are located in relatively isolated environment, similarly to our above-MZ galaxies. If our above-MZ galaxies are explained by an early stage of the transition, our above-MZ galaxies may be loosing (or have lost) gas despite of their very recent star formation suggested by the high EW$_0$(Hβ) (Section 10.1). The gas loss can be caused by the galactic outflow triggered by supernovae (SNe) in young, starburst galaxies like our above-MZ galaxies. However, to characterize these above-MZ galaxies, more observations are necessary such as far-infrared and radio observations which trace emission from molecular gas, H I gas, and the SNe.

Another possible explanation of the moderate metallicities is that SNe contribute to the metal enrichment by ejecting metal-rich gas into ISM. Because our metal-poor galaxies have very low stellar masses, local events such as SNe may have a global impact over a whole galaxy within a very short time. The time scale of the SN metal

enrichment should be shorter than the time scale of the primordial gas inflow, which dilutes ISM and lowers a metallicity. We need to verify with simulation that SNe actually increase a metallicity by ∼0.5–1.0 dex in a galaxy with $\log(M_\star/M_\odot)$∼5–7.

Figure 10.6 demonstrates the low-$M_\star$, low-metallicity ends of the $M_\star$-Z relation. Here we compare our metal-poor galaxies with the representative metal-poor galaxies. Among the representative metal-poor galaxies, we find that our HSC J1631+4426 (12+log(O/H)=6.97, i.e., $Z/Z_\odot$=0.019) has the lowest metallicity reported ever. The metallicity of HSC J1631+4426 is comparable with that of J0811+4730 (Izotov et al., 2018b) within a error bar. The metallicity of our HSC J1631+4426 is as low as those of J0811+4730 (Izotov et al., 2018b), AGC198691 (Hirschauer et al., 2016), SBS0335−052 (e.g., Izotov et al., 2009), and J1234+3901 (Izotov et al., 2019b). We emphasis that the discovery of the very faint EMPG, HSC J1631+4426 (i=22.5 mag) has been enabled by the deep, wide-field HSC-SSP data, which the previous SDSS surveys could not achieve. Note that this thesis presents just the first spectroscopic result of 4 out of the 27 HSC-EMPG candidates. We expect to discover more EMPGs from our HSC-EMPG candidates in the undergoing spectroscopy.

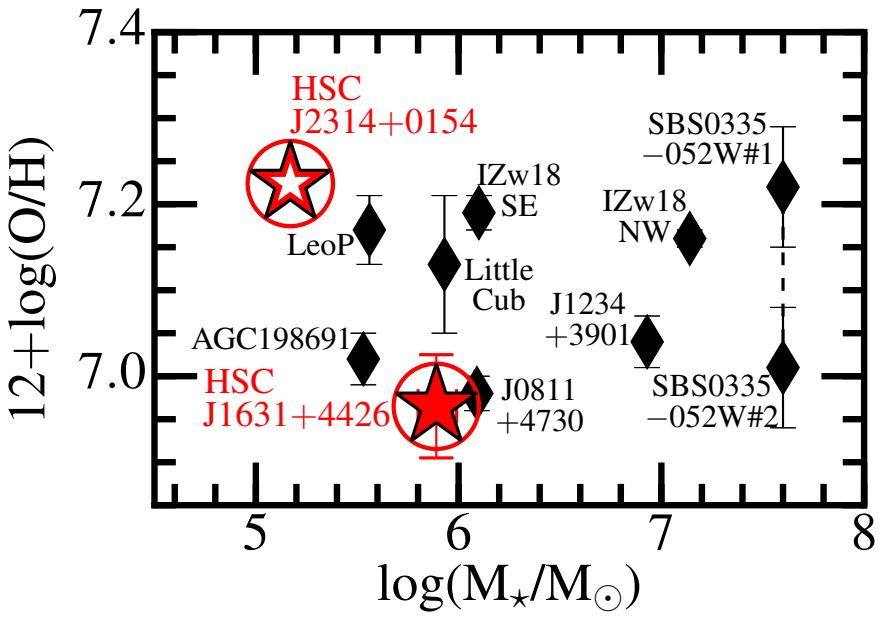

Figure 10.6 Same as Figure 10.5, but zoom-in around the low-$M_\star$, low-metallicity ends.

10.6 Velocity Dispersion

We estimate velocity dispersions of our 8 metal-poor galaxies observed with MagE. Note that LDSS-3, DEIMOS, and FOCAS do not have a spectral resolution high enough to resolve emission lines of our very low-mass sample. We measure emission-line widths of our metal-poor galaxies by the Gaussian fit to an Hβ emission line, obtaining $\sigma_{\rm obs}$=36.5–45.3 km s^{-1}. We also measure the instrumental line broadening with arc-lamp frames and find $\sigma_{\rm inst}$=26.4 and 33.3 km s^{-1} with the slit widths of 0.85 and 1.20 arcsec, respectively. Assuming that the instrumental line broadening is given in the form of a Gaussian function, we obtain intrinsic velocity dispersions, σ of our metal-poor galaxies by calculating

$$\sigma = \sqrt{\sigma_{\rm obs}^2 - \sigma_{\rm inst}^2}. \tag{10.1}$$

The obtained values are in the range of σ=27.8–32.5 km s^{-1} (Table 9.3). We do not remove an effect of the emission line broadening caused by the dynamical galaxy rotation because the spectral resolution of MagE is still not enough to separate the rotation and the dispersion. Thus, our estimates may provide upper limits on the velocity dispersions.

Top panel of Figure 10.7 demonstrates velocity dispersions of our metal poor galaxies as a function of V-band absolute magnitude in comparison with stellar velocity dispersions of massive galaxies (Prugniel & Simien, 1996), dwarf galaxies (a compiled catalog of Lin & Ishak, 2016), and globular clusters (Harris, 1996). We find that our metal-poor galaxies fall on a velocity-dispersion sequence made of massive galaxies, dwarf galaxies, and globular clusters in the top panel of Figure 10.7. The compiled dwarf galaxy catalog of Lin & Ishak (2016) are derived from the literatures on dwarf galaxies in Local Group ($\lesssim$3 Mpc) reported by McConnachie (2012), Kirby et al. (2015b,a), Simon et al. (2015), and Martin et al. (2016). Our metal-poor galaxies may be the first example of the lowest-$M_\star$ galaxies outside Local Group whose velocity dispersions are strongly constrained down to $\sigma\sim$30.0 km s^{-1}. It should be noted that the velocity dispersions of our metal-poor galaxies trace the gas kinemat-

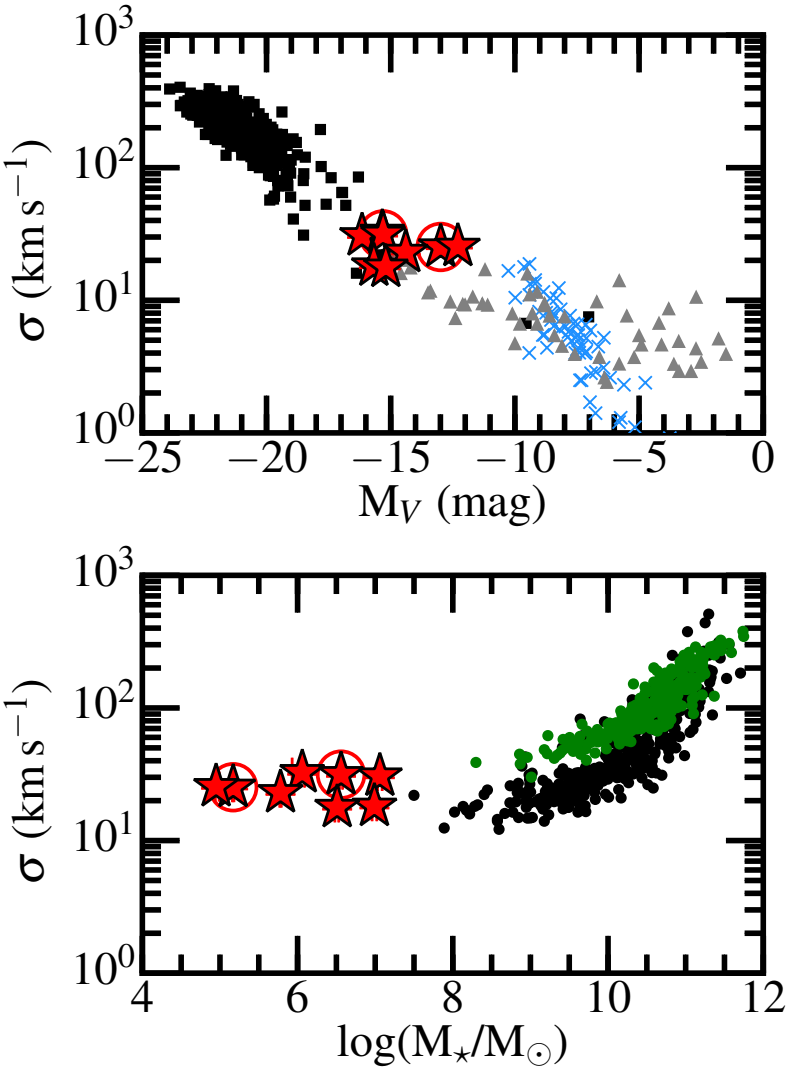

Figure 10.7 Top: Velocity dispersion as a function of optical magnitudes. Stars are the same as in Figure 10.1. Squares and triangles represent stellar velocity dispersions of bright galaxies and local faint galaxies from the literature, which are compiled by Prugniel & Simien (1996) and Lin & Ishak (2016), respectively. We also show stellar velocity dispersions of globular clusters (Harris, 1996) with blue crosses. To estimate the continuum level in V-band of our metal poor galaxies, we use an i-band magnitude instead of g- or r-band magnitudes because the g and r band are strongly affected by strong emission lines. Here we assume a flat continuum from V to i bands in the unit of erg s^{-1} cm^{-2} Hz^{-1}. We confirm that this assumption is correct within $\sim$0.2 mag by looking at a continuum in MagE spectra of our metal-poor galaxies. Bottom: Same as top panel, but as a function of stellar mass. Black and green circles represent velocity dispersions obtained with stellar and nebular lines, respectively, from the Sydney-AAO Multi-object Integral field spectrograph (SAMI) galaxy survey (Barat et al., 2019). All the velocity dispersion measurements are corrected for an instrumental broadening.

ics while those of the massive galaxies, dwarf galaxies, and globular clusters shown here are estimated simply from the stellar kinematics. Indeed, as shown in bottom panel of Figure 10.7, the velocity dispersions of gas (green circles) and stars (black circles) are different in $\log(M_\star/M_\odot)\sim$8–11 by a factor of 1.0–1.3 (Barat et al., 2019), although the low-$M_\star$ range of our metal-poor galaxies is not investigated for now. The combination of high sensitivity integral-field spectroscopy and high spectral resolution ($R\gtrsim$10,000) spectroscopy may be necessary to unveil the kinematics of the very low-$M_\star$ galaxies by resolving the gas and stellar components and the rotation and dispersion.

10.7 BPT Diagram

Figure 10.8 is an emission line diagnostic diagram of [N II]/Hα and [O III]/Hβ (i.e., BPT diagram, see Section 2.2.2) for our metal-poor galaxies. Our metal-poor galaxies fall on the SFG region defined by the maximum photoionization models under the stellar radiation (Kewley et al., 2001). We do not find any evidence that our metal-poor galaxies are affected by an AGN or shock heating from the optical emission line ratios. However, Kewley et al. (2013) suggest that metal-poor gas heated by the AGN radiation or shock also show emission-line ratios falling on the SFG region defined by Kewley et al. (2001). We thus do not exclude the possibility of the existence of a metal-poor AGN or shock heating of metal-poor gas.

10.8 Element Abundance Ratios

We show element abundance ratios of neon, argon, nitrogen, and iron to oxygen (Ne/O, Ar/O, N/O, and Fe/O) of our metal-poor galaxies and compare them with literature in this Section. Figure 10.9 shows the Ne/O, Ar/O, N/O, and Fe/O ratios as a function of metallicity, 12+log(O/H). We discuss these element abundance ratios in the following subsections.

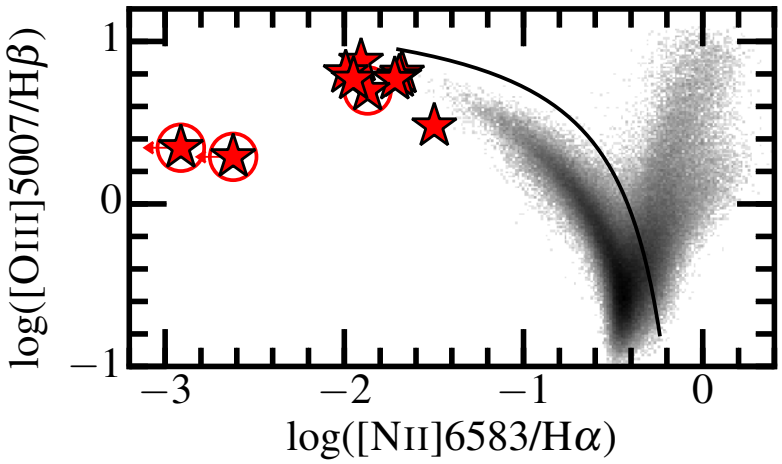

Figure 10.8 Our metal-poor galaxies on the BPT diagram (red stars). The black mesh represents $z \sim 0$ SFGs and AGNs derived from the emission-line catalog of SDSS DR7 (Tremonti et al., 2004). The solid curve indicates the maximum photoionization models that can be achieved under the assumption of stellar radiation (Kewley et al., 2001). The region below the solid curve is defined by the SFG region, while the upper right side is defined by the AGN region.

10.8.1 Ne/O and Ar/O ratios

Izotov et al. (2006) report that Ne/O, Ar/O ratios little depend on metallicity because the neon, argon, and oxygen are all α elements, which are produced by the nuclear fusion of α particles inside a star. As shown in the panels (a) and (b) of Figure 10.9, we find that the Ne/O, Ar/O ratios of our metal-poor galaxies are consistent with those of local galaxies reported by Izotov et al. (2006) within a scatter. Especially, the Ne/O ratios show consistency within the small scatter between our sample and the Izotov et al. (2006) sample as shown in the panel (a) of Figure 10.9. The consistency suggests that our metal-poor galaxy sample also shows no metallicity dependence of Ne/O and Ar/O ratios. Although the Ar/O ratio might slightly decrease in the range of 12+log(O/H)$\gtrsim$8.2 in our sample and the Izotov et al. (2006) sample, we do not discuss it further in this thesis because we cannot rule out the possibility of the underlying systematics in the Ar/O estimation at 12+log(O/H)$\gtrsim$8.2, and the current sample size is not large enough to reach a definitive conclusion.

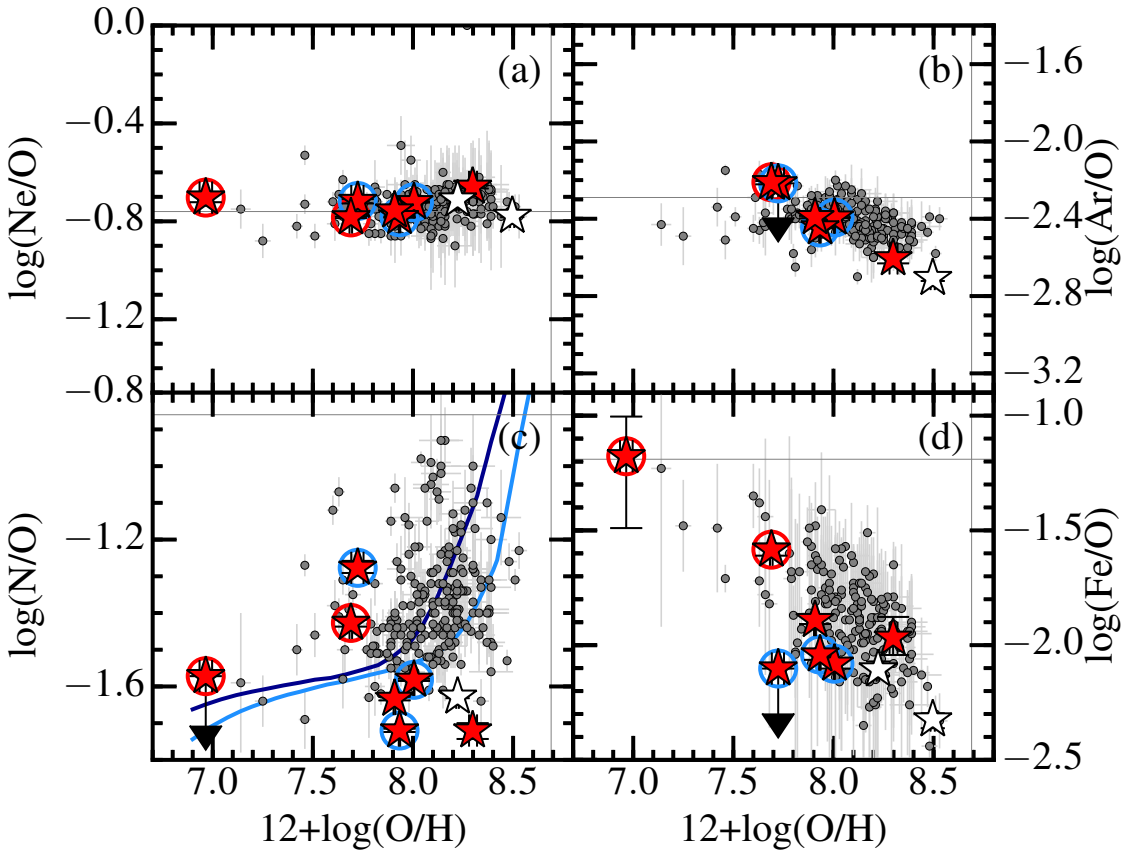

Figure 10.9 Element abundance ratios of nitrogen, neon, argon, and iron to oxygen (N/O, Ne/O, Ar/O, and Fe/O) as a function of metallicity. Symbols are the same as 10.1. White open star marks are our metal-poor galaxies that might include the possible systematics in the slit loss correction. Blue circles are galaxies showing deviations in emission line ratios, which will be described in Section 10.9. Gray dots represent local galaxies of Izotov et al. (2006). Solid lines in the panel (c) are the model calculation of the N/O evolution (Vincenzo et al., 2016) with star-formation efficiencies of 0.5 (dark blue) and 1.0 (light blue) Gyr^{-1}. Gray vertical and horizontal lines indicate solar abundance ratios and metallicity (Asplund et al., 2009).

10.8.2 N/O ratio

As suggested by previous studies (Pérez-Montero & Contini, 2009; Pérez-Montero et al., 2013; Andrews & Martini, 2013), N/O ratios of SFGs present a plateau ($\log(\mathrm{N/O})\sim-1.5$) in the range of $12+\log(\mathrm{O/H})\lesssim8.0$ and a positive slope at higher metallicities as a function of metallicity. The panel (c) of Figure 10.9 presents model calculation of the N/O evolution (Vincenzo et al., 2016), which also shows the plateau and positive slope. The plateau is basically resulted from the primary nucleosynthesis of massive stars, while the positive slope is mainly attributed to the secondary nucleosynthesis of low- and intermediate-mass stars (e.g., Vincenzo et al., 2016). We briefly describe the two nitrogen production processes below.

- Primary nucleosynthesis: Inside a metal-poor star, protons are burned through the proton-proton (*p*-*p*) chain reaction, and little nitrogen is produced at this stage. Nitrogen elements are mainly produced after the formation of a heavy-element core (e.g., O and C) and ejected into ISM by SNe, for stars more massive than $\sim8\ M_{\odot}$.

- Secondary nucleosynthesis: Metal-rich stars efficiently burn hydrogen through the carbon-nitrogen-oxygen (CNO) cycle, where nitrogen elements accumulate because ^{14}N fusion ($^{14}\mathrm{N}+p\rightarrow{}^{15}\mathrm{O}+\gamma$) is the slowest process in the CNO cycle. Then nitrogen is ejected through stellar winds during the asymptotic giant branch (AGB) phase, $\sim$1 Gyr after the birth of low- and intermediate-mass stars.

As shown in the panel (c) of Figure 10.9, most of our metal-poor galaxies have N/O ratios lower than $\log(\mathrm{N/O})=-1.4$ (i.e., less than $\sim$30 percent of the solar N/O ratio). Especially, HSC J1631+4426 has a strong, 2σ upper limit of $\log(\mathrm{N/O})<-1.57$, which will be discussed in Section 10.8.3. These low N/O ratios suggest that our metal-poor galaxies have not yet started the secondary nucleosynthesis due to their low metallicities and young stellar ages.

10.8.3 Fe/O ratio

In the panel (d) of Figure 10.9, we find that our metal-poor galaxies show a decreasing trend in Fe/O ratio as metallicity increases. The same decreasing Fe/O trend is found in a star-forming-galaxy sample of Izotov et al. (2006). Most of our metal-poor galaxies have Fe/O ratios comparable to a star-forming-galaxy sample of Izotov et al. (2006). Two EMPGs, HSC J1631+4426 and SDSS J2115−1734 (encircled by a red circle) show relatively high Fe/O ratios, $\log(\mathrm{Fe/O})>-1.6$, among our metal poor galaxies. Interestingly, we find that HSC J1631+4426 has a Fe/O ratio comparable to the solar Fe/O ratio, $\log(\mathrm{Fe/O})_\odot=-1.19$, although HSC J1631+4426 is one of the lowest metallicity galaxy with $\sim$0.019 $(\mathrm{O/H})_\odot$.

To characterize the two EMPGs with a high Fe/O ratio, we also compare our metal poor galaxies with Galactic stars (blue squares) in Figure 10.10. The solid line here represents an Fe/O evolution model under the assumption that gas is enriched by massive stars with 9–100 $M_\odot$ (Suzuki & Maeda, 2018). Figure 10.10 suggests that the two EMPGs deviate from the observational results of Galactic stars and the Fe/O evolution model. Below, we discuss three scenarios that may be able to explain the Fe/O deviation of the two EMPGs.

The first scenario is the preferential dust depletion of iron, suggested by Rodriguez & Rubin (2005) and Izotov et al. (2006). Rodriguez & Rubin (2005) and Izotov et al. (2006) discuss that gas-phase Fe/O ratios decrease as a function of metallicity in the range of $12+\log(\mathrm{O/H})\lesssim8.5$ because iron elements are depleted into dust more effectively than oxygen. The depletion may become dominant in a higher metallicity range, where the dust production becomes more efficient. For dust-free (i.e., metal-poor) galaxies, gas-phase Fe/O ratios are expected to become comparable to the observation results of Galactic stars and the Fe/O evolution model. Although the dust depletion may explain the negative Fe/O slope, it does *not* explain the fact that our EMPGs show higher Fe/O ratios than the Galactic stars and model at fixed metallicity. In addition, as we have seen in Section 9.1, most of our metal-poor galaxies show $E(B-V)\sim0$ (i.e., dust free), which means that the dust depletion is not relevant to the Fe/O ratio of our sample. Thus, we rule out the first scenario.

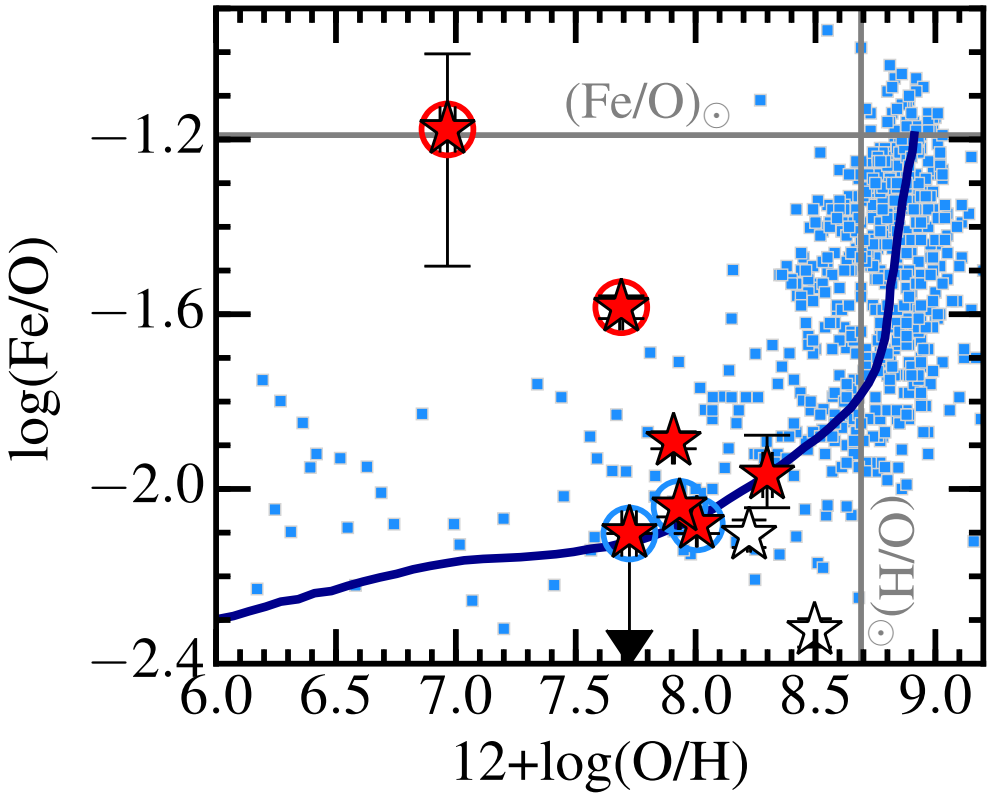

Figure 10.10 Comparison of Fe/O ratios of our metal-poor galaxies (symbols are the same as Figure 10.9) and Galactic stars (blue squares). We show observational data of Galactic stars from Cayrel et al. (2004), Gratton et al. (2003), and Bensby et al. (2013). Blue solid line represents an Fe/O evolution model under the assumption that gas is enriched by massive stars with 9–100 $M_\odot$ (Suzuki & Maeda, 2018).

The second scenario is a combination of metal enrichment and gas dilution caused by inflow. In this scenario, we assume that EMPGs are formed from metal-enriched gas with the solar metallicity and solar Fe/O ratio. If primordial gas (i.e., almost metal free) falls into the metal-enriched EMPGs, the metallicity (i.e, O/H) decreases while the Fe/O ratio does not change. It seems that this scenario may explain the Fe/O deviation of the two EMPGs. However, if the second scenario is true, both the Fe/O and N/O ratios should be solar abundances because the N/O ratio also reaches the solar N/O ratio, $\log(\mathrm{N/O})_\odot = -0.86$, at the solar metallicity. As we have seen in the panel (c) of Figure 10.9, the two deviating EMPGs, HSC J1631+4426 and SDSS J2115−1734 (encircled by a red circle) have a strong 2σ upper limit of <0.19 $(\mathrm{N/O})_\odot$ and a low value of 0.27 $(\mathrm{N/O})_\odot$, respectively. These low N/O ratios suggest that the two deviating EMPGs are experiencing the primary nucleosynthesis, not the secondary nucleosynthesis. Thus, we exclude the second scenario because the second scenario does not explain the Fe/O and N/O ratios simultaneously.

The third scenario is the contribution of very massive stars above 300 $M_\odot$. Very massive stars above 300 $M_\odot$ eject much iron at the time of core-collapse SN explosion. Ohkubo et al. (2006) have calculated yields from core-collapse SNe under the assumption of the progenitor star mass with 500–1000 $M_\odot$, obtaining $\sim$2–40 $(\mathrm{Fe/O})_\odot$. In the very massive stars above 300 $M_\odot$, an iron core grows until the iron core occupies more than 20 percent of the stellar mass. Although massive stars with 140–300 $M_\odot$ undergo thermonuclear explosions triggered by pair-creation instability (PISNe, Barkat et al., 1967), very massive stars above 300 $M_\odot$ are too massive to trigger PISNe and thus continue the iron core growth. The very massive stars above 300 $M_\odot$ eject a large amount of iron by a jet stream from the iron core during the SN explosion. On the other hand, the core-collapse SNe of typical-mass stars (10–50 $M_\odot$) eject gas with an average of $\sim$0.4 $(\mathrm{Fe/O})_\odot$ (Tominaga et al., 2007, IMF integrated in the range of 10–50 $M_\odot$), which is below the solar Fe/O ratio. Yields of type-Ia SNe calculated by Iwamoto et al. (1999) show $\sim$40 $(\mathrm{Fe/O})_\odot$. Figure 10.11 demonstrates the yields of very massive star SNe (500 $M_\odot$), typical mass SNe (10–50 $M_\odot$), and type-Ia SNe. Only the type-Ia SNe and the SNe of very massive stars ($>$300 $M_\odot$) contribute to the iron enrichment larger than the solar Fe/O ratio. As discussed in Section 1.3,

observations of Galactic stars show low Fe/O ratios in the range of 12+log(O/H)<8.0 and a rapid increase at higher metallicities. An interpretation of this trend is that the iron enrichment is dominated by core-collapse SNe at 12+log(O/H)<8.0 and becomes dominated by type-Ia SNe ∼1 Gyr after the star formation corresponding to 12+log(O/H)> 8.0 (e.g., Carretta et al., 2000; Takeda, 2003; Bai et al., 2004). As we discussed in the second scenario above, we find that the two deviating EMPGs have low N/O ratios. The low N/O ratios suggest that the two EMPGs have not yet started the secondary nucleosynthesis, which is expected to start ∼1 Gyr after the star formation (Vincenzo et al., 2016). Thus, the high Fe/O ratios of the two EMPGs are not caused by type-Ia SNe, which also occur ∼1 Gyr after the star formation. The remaining possibility is the contribution from the SNe of very massive stars above 300 $M_{\odot}$.

In summary of this subsection, we have discussed the three scenarios that may be able to explain the high Fe/O ratios of the two EMPGs. We insist that the high Fe/O ratios of the two EMPGs are attributed to the contribution from core-collapse SNe of very massive stars above 300 $M_{\odot}$. This scenario is also consistent with the fact that the two EMPGs are very young ($\lesssim$ 50 Myr).

10.9 Ionizing Radiation

10.9.1 Emission Line Ratios

We investigate ionizing radiation of our metal-poor galaxies by comparing emission line ratios of various ions. Figure 10.12 shows eight emission line ratios of [O II]3727/Hβ, [Ar III]4740/Hβ, [O III]5007/Hβ, [Ar IV]7136/Hβ, [O II]3727/He II4686, [Ar III]4740/He II4686, [O III]5007/He II4686, and [Ar IV]7136/He II4686 as a function of metallicity. We choose [O II]3727, [Ar III]4740, [O III]5007, and [Ar IV]7136 emission lines among many emission lines detected in our spectroscopy for two reasons below. The first reason is that oxygen and argon are both α elements, and thus the Ar/O abundance ratio is almost constant as we confirm in Section 10.8. Thus, emission line ratios are simply interpreted by ionizing radiation intensity and/or hardness, free

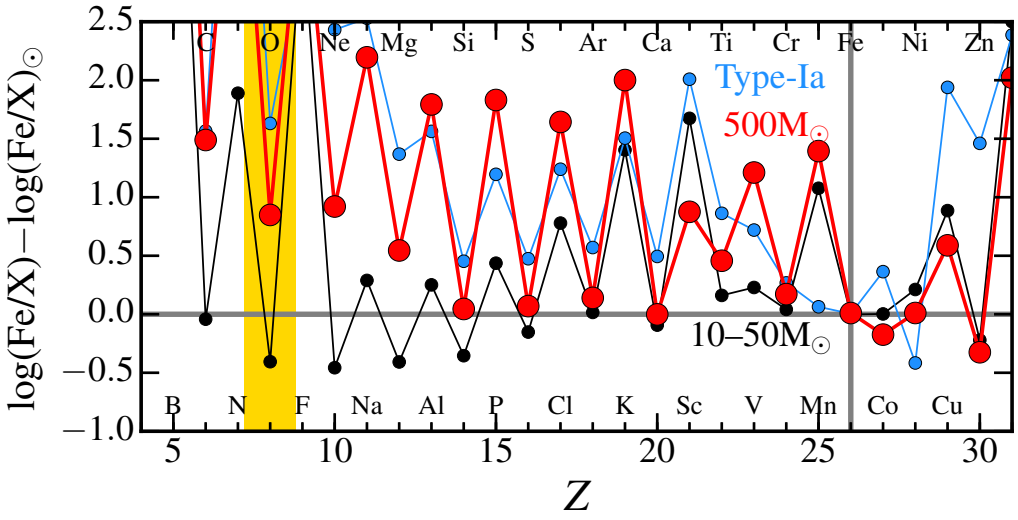

Figure 10.11 Yields of three types of SNe. For the yields, we show abundance ratios of iron to various elements from boron (atomic number $Z = 5$) to zinc ($Z = 30$), which is normalized by the solar abundance ratios (gray horizontal line). Red and black circles are yields from core-collapse SNe of 500 $M_\odot$ and 10–50 $M_\odot$ (IMF weighted) calculated by Ohkubo et al. (2006) and Tominaga et al. (2007), respectively. Blue circles show yields of type-Ia SNe estimated by Iwamoto et al. (1999). The position of Fe/O ratio is highlighted in yellow. Core-collapse SNe of 10–50 $M_\odot$ and 500 $M_\odot$ and type-Ia SNe show ~ 0.4, 7, and 40 $(\mathrm{Fe/O})_\odot$, respectively.

from variance of element abundance ratio. The second reason is that the four lines are sensitive to ionization photons in a wide energy range from 13.6 to 40.7 eV. The [O II]3727, [Ar III]4740, [O III]5007, and [Ar IV]7136 lines are emitted via spontaneous emission after collisional excitation of O^+, Ar^{2+}, O^{2+}, and Ar^{3+}, respectively. The O^+, Ar^{2+}, O^{2+}, and Ar^{3+} ions are produced by the ionization of O^0, Ar^+, O^+, and Ar^{2+}, whose ionization potentials are 13.6, 27.6, 35.1, and 40.7 eV, respectively. Table 10.2 summarizes these emission line processes and corresponding photon energy required to emit these lines.

Table 10.2 Summary of Emission Line Process, Ionization Process and Ionization Potential

Line	Emission Process	Ionization Process	Ionization Potential (eV)
[O II]3727	Collisional Excitation	$O^0+\gamma \rightarrow O^+$	13.6
[Ar III]4740	Collisional Excitation	$Ar^{2+}+\gamma \rightarrow Ar^{3+}$	27.6
[O III]5007	Collisional Excitation	$O^++\gamma \rightarrow O^{2+}$	35.1
[Ar IV]7136	Collisional Excitation	$Ar^{3+}+\gamma \rightarrow Ar^{4+}$	40.7
Hβ	Recombination	$H^0+\gamma \rightarrow H^+$	13.6
He II4686	Recombination	$He^++\gamma \rightarrow He^{2+}$	54.4

We normalize the four oxygen and argon lines by Hβ and He II4686 lines, because the hydrogen and helium are the most elementary element and abundant in the universe. The Hβ and He II4686 lines are recombination lines, which are emitted during the cascade process in the recombination of H^+ and He^{2+}. The H^+ and He^{2+} ions require ionizing photons above 13.6 eV and 54.4 eV because the ionization potentials of H^0 and He^+ are 13.6 eV and 54.4 eV, respectively. We also summarize the emission line processes and their corresponding ionization potentials in Table 10.2.

In panels of (a)–(d) of Figure 10.12, we show local, average SFGs of Andrews & Martini (2013, AM13 hereafter) with black circles. We regard the AM13 SFGs as local averages because the AM13 sample is obtained by the SDSS composite spectra in bins of wide SFR and stellar-mass ranges. In panels of (a)–(d), the AM13 SFGs form sequences as a function of metallicity. The sequences of [O II]3727/Hβ and [Ar III]4740/Hβ show peaks at around 12+log(O/H)$\sim$8.7 and 8.3, respectively. The [O III]5007/Hβ and [Ar IV]7136/Hβ ratios may also have peaks around 12+log(O/H)

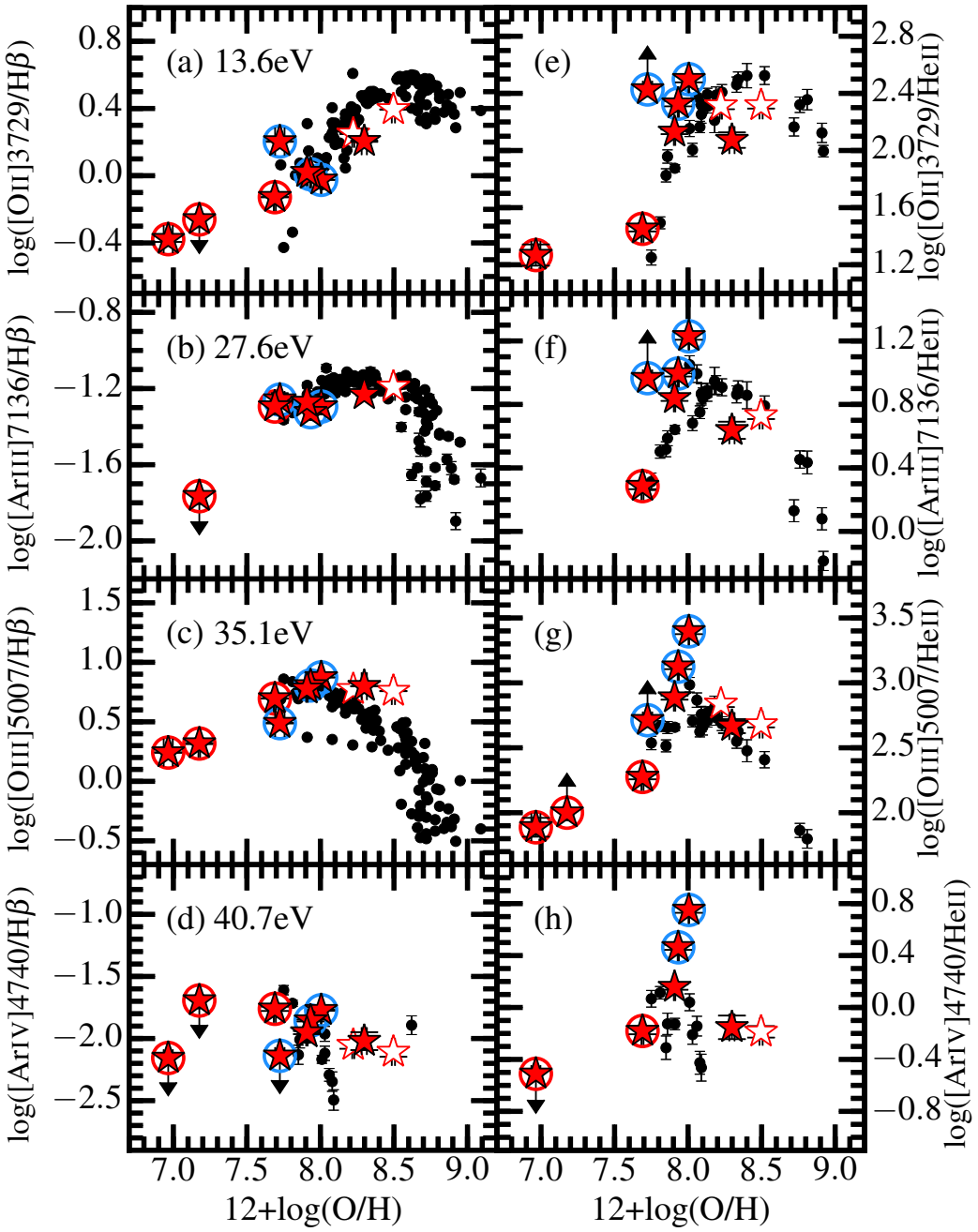

Figure 10.12 Emission line ratios of [O II]3727, [Ar III]4740, [O III]5007, and [Ar IV]7136 to Hβ in panels (a)–(d) and those to He II4686 in panels (e)–(h). Symbols are the same as in Figure 10.9. We emphasis with blue circles three metal poor galaxies which show the deviation from the local averages of Andrews & Martini (2013) by $\gtrsim$0.4 dex for a given metallicity in any of panels (e)–(h). These ratios are dust corrected. The ionization potentials of O^0, Ar^{2+}, O^+, and Ar^{3+} ions (13.6, 27.6, 35.1, and 40.7 eV, respectively) are presented in panels (a)–(d). Black circles represent averages of local SFGs obtained with SDSS composite spectra (Andrews & Martini, 2013).

$\sim$8.0 and 12+log(O/H)$\sim$7.2–7.7 by interpolating AM13 SFGs and our metal-poor galaxies. Recalling that the [O II]3727, [O III]5007, [Ar III]4740, and [Ar IV]7136 lines are sensitive to ionizing photon above 13.6, 35.1, 27.6, and 40.7 eV, respectively, we find that the peak metallicities decrease with increasing ionizing potentials of the corresponding emission lines. The peak transition demonstrates that ISM is irradiated by more intense or harder ionizing radiation in lower metallicity, as suggested by previous studies (e.g., Nakajima & Ouchi, 2014; Steidel et al., 2016; Nakajima et al., 2016; Kojima et al., 2017). We also find that our metal-poor galaxies fall on the sequences of AM13 SFGs within a scatter. Thus, we insist that our metal-poor galaxies and the AM13 SFGs have similar ionizing radiation in the range of 13.6–40.7 eV.

In panels (e)–(h) of Figure 10.12, we show [O II]3727, [O III]5007, [Ar III]4740, and [Ar IV]7136 lines normalized by the He II4686 recombination line. The He II4686 line is sensitive to ionizing photons with very high energy (>54.4 eV). Such high-energy photons are not abundant in radiation of O- and B-type hot stars. Thus, under the stellar radiation, the He II4686 emission line is very weak compared to hydrogen recombination lines such as Hβ (typically He II4686/H$\beta\sim$1/100 for local SFGs, see also Figure 10.13). In panels (e)–(h), the AM13 SFGs also show sequences similar to those in panels (a)–(d) of Figure 10.12. However, in panels (g) and (h), SDSS J2253+1116, and SDSS J2310−0211 show deviations from the sequences of [O III]5007/He II4686 and [Ar IV]7136/He II4686 by +0.2 to +0.4 dex for a fixed metallicity. HSC J1142−0038 also deviates from the sequences of [O II]3727/He II4686 and [Ar III]4740/He II4686 by more than +0.6 dex for a given metallicity. We mark these three metal-poor galaxies with blue circles in Figure 10.12. In panel (b) of Figure 10.12, we have found that SDSS J2253+1116 and SDSS J2310−0211 (two galaxies encircled by a blue circle) fall on the sequence made by local SFGs of Izotov et al. (2006) on the relation between log(Ar/O) and 12+log(O/H). Although HSC J1142−0038 only has an upper limit of log(Ar/O)<−2.22, the upper limit is still consistent with the sequence of local SFGs of Izotov et al. (2006), log(Ar/O)$\sim$−2.4. Thus, we insist that these deviations are not attributed to variation of the element abundance ratio. We infer that hard radiation above 54.4 eV is weak in the three

galaxies, compared to those of the local average galaxies. The weakness of hard radiation results in weak He II4686 emission, and thus higher ratios of [O II]3727/He II4686, [O III]5007/He II4686, [Ar III]4740/He II4686, and/or [Ar IV]7136/He II4686.

Figure 10.13 shows He II4686/Hβ ratios of our metal-poor galaxies as a function of metallicity, as well as the AM13 SFGs (black circles). The AM13 SFGs show almost constant He II4686/Hβ ratios around log(He II4686/Hβ)~-2.0 in the range of 12+log(O/H)=8.1–8.6, while the He II4686/Hβ ratios increases with decreasing metallicity in the range of 12+log(O/H)=7.7–8.1. The three deviating galaxies, SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038 (encircled by a blue circle) have He II4686/Hβ ratios significantly lower than the AM13 SFGs. Thus, we confirm that the deviations from the [O II]3727/He II4686, [O III]5007/He II4686, [Ar III]4740/He II4686, and/or [Ar IV]7136/He II4686 sequences shown in Figure 10.13 are caused by the weak radiation of hard ionizing photon above 54.4 eV.

To interpret the weak ionizing radiation above 54.4 eV, we compare our metal-poor galaxies with calculation results of Schaerer et al. (2019, S19 hereafter). In S19, He II4686 emission of local SFGs is explained with high mass X-ray binary (HMXB) models of Fragos et al. (2013a,b). HMXBs are binary systems consisting of a compact object (such as BH) and a companion star. The companion star provides gas onto the compact object, and creates a hot accretion disk around the compact object. The hot accretion disk radiates very hard, power-low radiation ranging from UV to X-ray. Fragos et al. (2013a,b) carefully calculate the HMXB evolution along the star-formation history, and predict total X-ray luminosities ($L_{\rm X}$) from a galaxy as functions of metallicity and age. S19 convert an $L_{\rm X}$/SFR ratio to the He II4686/Hβ ratio, under the simple assumptions of He II4686/Hβ=$1.74\times Q({\rm He^+})/Q({\rm H})$ (Case B recombination of 20,000K, Stasińska et al., 2015), Q(H)/SFR=9.26×10^{52} photon s^{-1}/(M$_\odot$ yr^{-1}) (Kennicutt, 1998), and hardness of $Q({\rm He^+})/L_{\rm X}=2\times10^{10}$ photon erg^{-1}. Here, the $Q({\rm He^+})$ and Q(H) are defined by ionizing photon production rates above 54.4 and 13.6 eV, respectively. S19 also use BPASS binary stellar synthesis models of Xiao et al. (2018) to associate stellar ages with EW$_0$(Hβ). Figure 10.14 compares He II4686/Hβ ratios of our metal-poor galaxies and those obtained by the S19 HMXB models (solid lines) as a function of EW$_0$(Hβ). The solid lines trace time evolution of He II4686/Hβ and

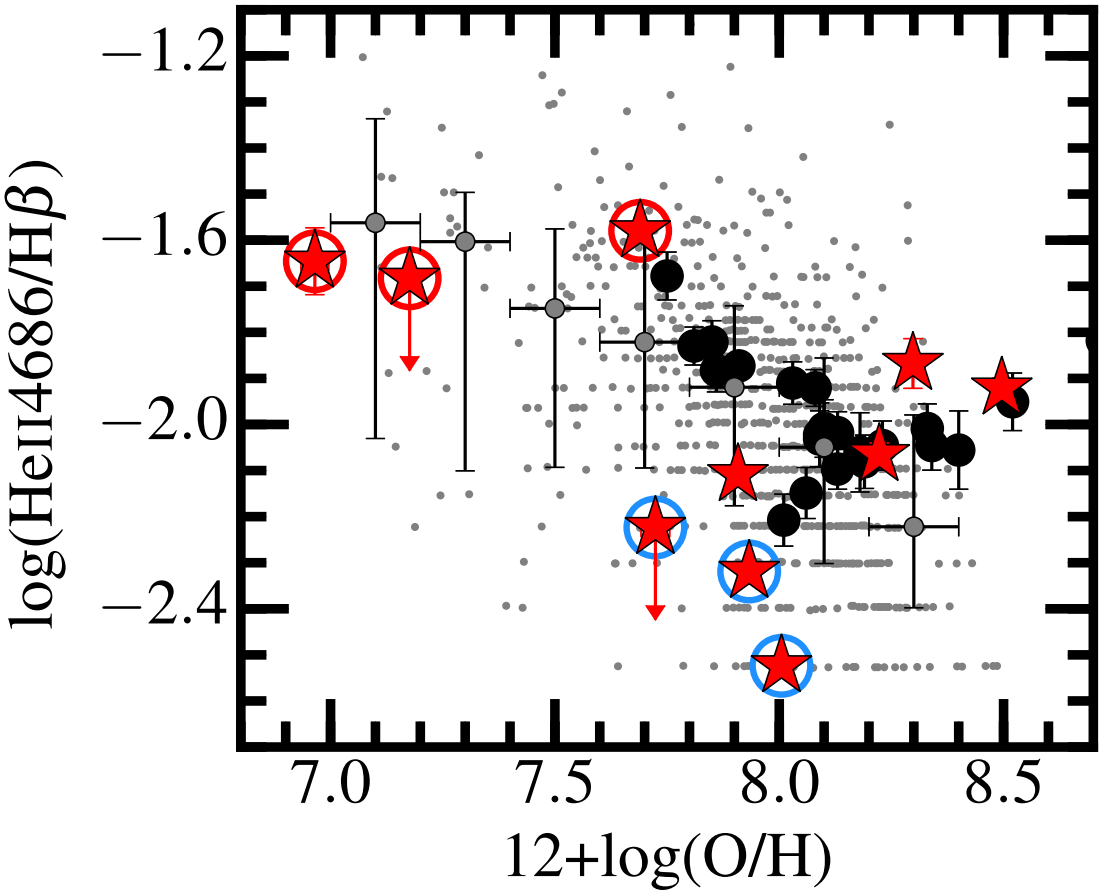

Figure 10.13 Emission line ratios of He II4686/Hβ as functions of metallicity. Symbols are the same as in Figure 10.12. Gray dots are individual local galaxies of S19 with an He II4686 detection. Gray circles and error bars show medians and 68%-percentile scatters of the S19 sample obtained in each metallicity bin, respectively.

$EW_0(H\beta)$ with different metallicities of 0.05, 0.10, 0.20, and 0.50 $Z_\odot$. An $EW_0(H\beta)$ decreases and an He II4686/Hβ ratio increases as time passes due to the stellar evolution and HMXB evolution, respectively. The HMXB models (especially 0.05 and $0.10Z_\odot$) show a rapid increase of He II4686/Hβ around $EW_0(H\beta)\sim$100–300 Å. The $EW_0(H\beta)\sim$100–300 Å corresponds to $\sim$5 Myr in the BPASS binary stellar synthesis models. The rapid increase is triggered by the first compact object formation (i.e., the first HMXB formation) after $\sim$5 Myr of the starburst. We find that the three deviating galaxies, SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038 (encircled by a blue circle), fall on or near the model lines where He II4686/Hβ ratios rapidly increase. Thus, we interpret that the low He II4686/Hβ ratios of the three deviating galaxies are explained by the very young stellar age ($\lesssim$5 Myr) before or in the middle of the HMXB formation.

It is worth noting that such low He II4686/Hβ galaxies have also been found in a local SFG sample of S19 (gray dots in Figures 10.13 and 10.14). We also exhibit their medians (gray circles) and 68%-percentile scatters (error bars) in each metallicity bin. The S19 sample includes galaxies whose He II4686/Hβ ratios are comparable to those of SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038. However, such low-He II4686/Hβ galaxies are still in a minority of the S19 sample, falling out of 68%-percentiles. The rarity of low He II4686/Hβ galaxies is explained by the very short period ($\sim$5 Myr) of the HMXB evolutionary phase. We emphases that at least 30% (3 out of 10) of our metal-poor sample are such rare, young galaxies with a low He II4686/Hβ ratio, which may be undergoing the very first phase of the HMXB evolution.

10.9.2 Strong He II4686 Line

As discussed in Section 10.9.1 (Figure 10.14), the HMXB models of S19 have quantitatively explained the He II4686/Hβ ratios of part of our metal poor galaxies including SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038. However, we find that five metal-poor galaxies are not explained by the HMXB model, which fall in the ranges of $EW_0(H\beta)>100$ Å and $\log(\text{He II}4686/H\beta)>(-2.0)$. Interestingly, three out

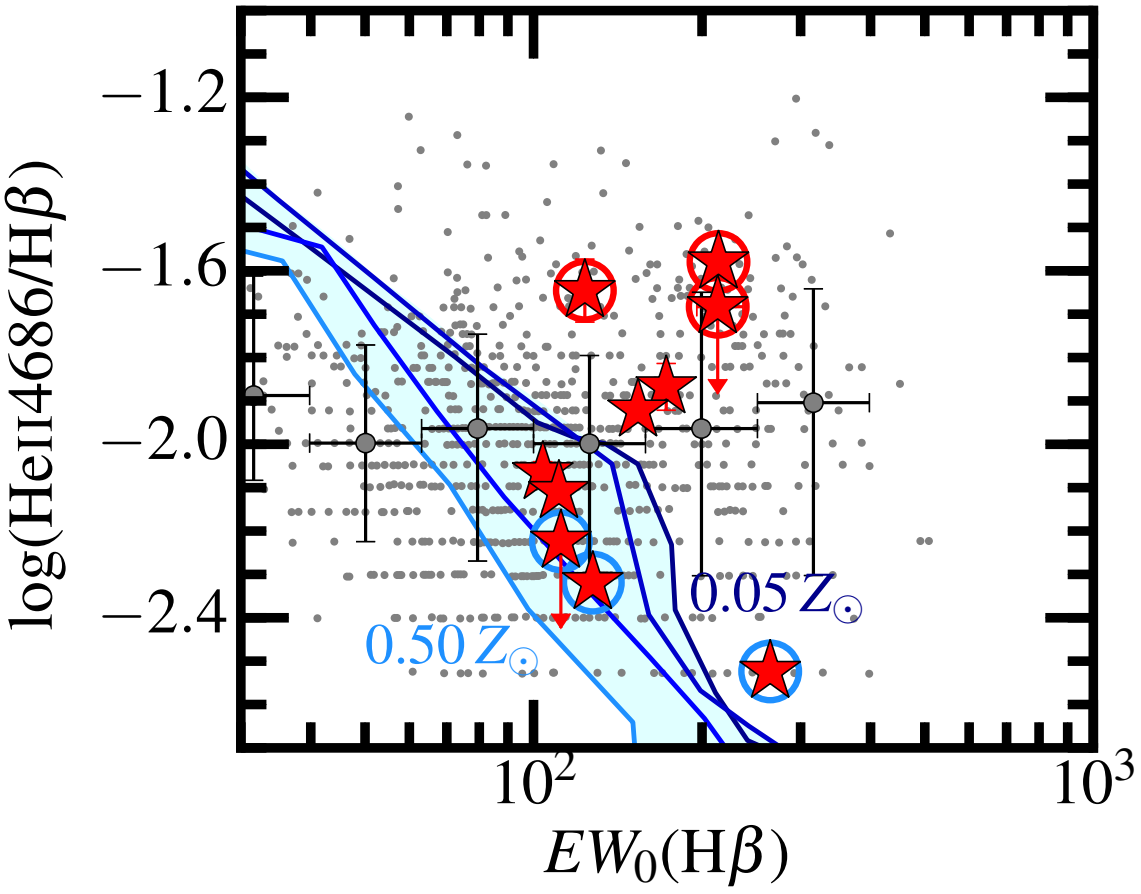

Figure 10.14 Same as Figure 10.13, but as a function of Hβ equivalent width, EW_0(Hβ). Solid lines represent the S19 HMXB models, tracing time evolution of He II4686/Hβ and EW_0(Hβ) with different metallicities of 0.05, 0.10, 0.20, and 0.50 $Z_\odot$ (from dark blue to light blue). Symbols are the same as in Figures 10.12 and 10.13.

of the five metal-poor galaxies are EMPGs, HSC J1631+4426, HSC J2314+0154, and SDSS J2115−1734 marked with a red circle. Furthermore, the S19 SFG sample also include galaxies in the ranges of $EW_0(H\beta)$>100 Å and log(He II4686/Hβ)>(−2.0). S19 have argued that other X-ray sources may appear fairly soon after the star formation ($\lesssim$5 Myr) in galaxies with high values of $EW_0(H\beta)$ and He II4686/Hβ. S19 suggest that an underlying older population or shocks could also contribute to the high He II4686/Hβ ratios.

In addition to the S19 suggestions, we propose two other possibilities here which may explain the high He II4686/Hβ ratio seen in the range of $EW_0(H\beta)$>100 Å. First, we suggest a possibility of a metal-poor AGN, which may contribute to boost the He II4686 intensity of the very young galaxies. In Section 10.7, we have confirmed that all of our metal-poor galaxies fall on the SFG region of the BPT diagram defined by the maximum photoionization models with stellar radiation (Kewley et al., 2001). However, Kewley et al. (2013) suggest that emission-line ratios calculated under the assumption of a metal-poor AGN also fall on the SFG region. Thus, we cannot exclude the possibility of a metal-poor AGN. Groves et al. (2004a,b) have constructed the photo-ionization models under the assumption of AGN-like, power-row radiation. The models of Groves et al. (2004a,b) predict very strong He II4686 emission represented by log(He II4686/Hβ)$\sim$(−1.5)–0.0. On the other hand, photo-ionization models with stellar radiation (Xiao et al., 2018) predict log(He II4686/Hβ)$\lesssim$(−2.5). To explain the observed ratios of log(He II4686/Hβ)$\sim$(−2.0), the combination of AGN and stellar radiation is required. We have checked the archival data of ROSAT and XMM, and found no detection in X-ray. This may be because the data are $\sim$2 orders of magnitudes shallower than expected X-ray luminosities ($\sim 10^{-14}$ erg s^{-1} cm^{-2}) of our metal-poor galaxy sample, which are obtained under the assumption of $L_{2\mathrm{keV}}$–M_{UV} relation of AGN (Lusso et al., 2010). Deep X-ray observations are required to constrain X-ray sources of metal-poor galaxies.

Second, we propose a possibility of very massive stars beyond 300 $M_\odot$. The HMXB models of Fragos et al. (2013a,b), discussed in Section 10.9.1, assume the Kroupa IMF (Kroupa, 2001; Kroupa & Weidner, 2003) with the maximum stellar mass (M_{max}) of 120 $M_\odot$. Thus, in the HMXB models of Fragos et al. (2013a,b),

the first HMXBs emerge $\sim$ 5 Myr after the star formation, which corresponds to a lifetime of a star with 120 $M_\odot$. On the other hand, stars more massive than 120 $M_\odot$ are expected to have a shorter life time than stars with 120 $M_\odot$. According to the theoretical study of Yungelson et al. (2008), very massive stars with 300 $M_\odot$ and 1000 $M_\odot$ die after 2.5 and 2.0 Myr after the star formation, respectively. As described in Section 10.8.3, stars between 140 and 300 $M_\odot$ undergo thermonuclear explosions triggered by PISNe (Barkat et al., 1967), and do not leave any compact object (e.g., Heger & Woosley, 2002). On the other hand, stars above 300 $M_\odot$ experience core-collapse SNe and form intermediate-mass BHs (IMBH, e.g., Ohkubo et al., 2006). Ohkubo et al. (2006) estimate BH masses of $\sim$230 and $\sim$500 $M_\odot$ for stars with initial masses of 500 and 1000 $M_\odot$, respectively. Thus, when we assume very massive stars above 300 $M_\odot$, IMBHs appear as early as $\sim$2 Myr, and part of the IMBHs may form HMXBs. Accretion disks of IMBHs emit very hard radiation including ionizing photons above 54.4 eV, which boost the He II4686 intensity. A galaxy as young as $\sim$2 Myr has $EW_0(H\beta)\sim$300–400Å according to the BPASS models. Under the assumption of very massive stars above 300 $M_\odot$, a He II4686/Hβ ratio is expected to start increasing at around $EW_0(H\beta)\sim$300–400Å. Such a model may covers the regions of $EW_0(H\beta)>100$ Å and $\log(\text{He II}4686/H\beta)>(-2.0)$ shown in Figure 10.14. Thus, we suggest that very massive stars above 300 $M_\odot$ may be able to explain the high ratios, $\log(\text{He II}4686/H\beta)>(-2.0)$ in the galaxies with $EW_0(H\beta)>100$ Å. Note again that galaxies with $\log(\text{He II}4686/H\beta)>(-2.0)$ and $EW_0(H\beta)>100$ Å includes EMPGs, HSC J1631+4426, HSC J2314+0154, and SDSS J2115−1734. These EMPGs might form very massive star beyond 300 $M_\odot$ from their extremely metal-poor gas. However, our explanation and the interpretation of S19 are based on some simple assumptions that associate the HMXB models and the BPASS stellar synthesis models. We propose to construct self-consistent SED models including the HMXB evolution with higher $M_{\rm max}$ values such as 300 $M_\odot$ than the current HMXB models.

CHAPTER 11
CONCLUSION

In this thesis, we have constructed a sample of local, metal-poor galaxies represented by low metallicities, 12+log(O/H) =6.97–8.50, low stellar masses, log($M_\star/M_\odot$)=4.95–7.06, and high specific star-formation rates (sSFR$\sim$300 Gyr^{-1}). The mass range of our metal-poor galaxies reaches down to log($M_\star/M_\odot$)$\sim$5, comparable to those of star clusters. Such cluster-like galaxies may be undergoing an early-stage galaxy formation as expected by the hydrodynamical simulation of Pop-II galaxies. In the sample, we have found 3 galaxies that satisfy the EMPG criterion, 12+log(O/H)<7.69. One of the EMPGs, HSC J1631+4426 shows one of the lowest metallicity values reported ever, 12+log(O/H)=6.97$\pm$0.08 (i.e., $Z/Z_\odot$=0.019). Our metal-poor galaxies are located in the relatively isolated environment, where the metal enrichment is inefficient. Low N/O ratios of our metal-poor galaxies suggest that our metal-poor galaxies are undergoing early phases of the nucleosynthesis (i.e., primary nucleosynthesis) due to their low metallicities and young ages.

Below, we focus only on 3 EMPGs and discuss Fe/O ratios and He II4686/Hβ ratios of the 3 EMPGs. We have found that 2 out of the 3 EMPGs, HSC J1631+4426 and SDSS J2115$-$1734[1], show Fe/O ratios 0.5–1.0 dex higher than Galactic stars and the model calculation of the Fe/O evolution at fixed metallicity. We have discussed three scenarios that may be able to explain the high Fe/O ratios, and concluded that the high Fe/O ratios are attributed to very massive stars above 300 $M_\odot$. We have also found that 3 EMPGs, HSC J2314+0154, HSC J1631+4426 and SDSS J2115$-$1734, show both high He II4686/Hβ ratios (>1/100) and high EW$_0$(Hβ) (>100 Å). These high He II4686/Hβ ratios and high EW$_0$(Hβ) are not explained by the latest binary population stellar synthesis model and the latest HMXB model, where a maximum stellar mass cut, 120 $M_\odot$ is used. On the other hand, we have suggested that very massive stars above 300 $M_\odot$ can explain the high He II4686/Hβ ratios (>1/100) and high EW$_0$(Hβ) (>100 Å). This is because stars above 300 $M_\odot$ explode as core-collapse SNe and form intermediate-mass black holes (IMBHs) $\sim$2 Myr after the star forma-

[1]HSC J2314+0154 has no element abundance measurements.

tion. Interestingly, the scenario of very massive stars above 300 $M_{\odot}$ explains both the high Fe/O ratios and the high He II4686/Hβ ratios of our EMPGs with the low metallicities ($< 0.1 Z_{\odot}$), young ages ($\lesssim$50 Myr), and very low stellar mass ($\sim 10^5$–10^6), which may be undergoing an early-stage of the galaxy formation. We conclude that very massive stars above 300 $M_{\odot}$ exist (or existed) in the cluster-like galaxies in the early-stage galaxy formation. Note that we do not exclude a possibility of metal-poor AGNs in the explanation of the high He II4686/Hβ ratios.

CHAPTER 12
SUMMARY

We search for extremely metal-poor galaxies (EMPGs) at $z \lesssim 0.03$ to construct a local sample whose galaxy properties are similar to those of high-z galaxies in the early star-formation phase (i.e., low $M_\star$, high sSFR, low metallicity, and young stellar ages). We select EMPGs from the wide-field, deep imaging data of the Subaru Strategic Program (SSP) with Hyper Sprime-Cam (HSC) in combination with the wide-field, shallow data of Sloan Digital Sky Survey (SDSS). This work is the first metal-poor galaxy survey that exploits the wide ($\sim$500 deg^2), deep ($i_{\rm lim}\sim$26 mag) imaging data of HSC SSP, with which we expect to discover rare, faint EMPGs that the previous SDSS survey could not find out. To remove contamination more efficiently than a simple color-color selection from our sample, we develop a new selection technique based on machine learning (ML). We construct a ML classifier that distinguishes EMPGs from other types of objects, which is well trained by model templates of galaxies, stars, and QSOs. By testing our ML classifier with the SDSS photometry+spectroscopy data, we confirm that our ML classifier reaches 86% completeness and 46% purity. Then our ML classifier is applied to the HSC and SDSS photometry data, obtaining 27 and 86 EMPG candidates, respectively. These EMPG candidates have a wide range of i-band magnitudes, i=14.8–24.3 mag, thanks to the combination of the SDSS and HSC data. We have conducted optical spectroscopy with Magellan/LDSS-3, Magellan/MagE, Keck/DEIMOS, and Subaru/FOCAS for 10 out of the 27+86 EMPG candidates. Our main results are summarized below.

- We find that the 10 EMPG candidates are real star-forming galaxies (SFGs) at z=0.007–0.03 with strong emission lines, whose rest-frame Hβ equivalent widths (EW$_0$) reach 104–265 Å, and a metallicity range of 12+log(O/H)=6.97–8.50. Three out of the 10 EMPG candidates satisfy the EMPG criteria of 12+log(O/H)<7.69. Note that the other 7 galaxies still show low metallicities ($\sim$0.1–0.3$Z_\odot$). We thus conclude that our new selection based on ML successfully selects real EMPGs or metal-poor, strong-line SFGs.

- The number density of our HSC metal-poor galaxies is 1.5×10^{-4} Mpc^{-3}, which is $\times10$ times higher than that of our SDSS metal-poor galaxies (2.8×10^{-5} Mpc^{-3}). This difference is explained by the fact that our HSC metal-poor galaxies (median: $i \sim 22.5$ mag) are ~ 30 times fainter than our SDSS metal-poor galaxies (median: $i \sim 18.8$ mag).

- To characterize the environment of our metal-poor galaxies, we compare nearest neighborhood distances (D_{near}) of our metal-poor galaxies with those of local, typical SFGs. The D_{near} of our metal-poor galaxies range from 0.49 to 17.69 Mpc with an average of 3.83 Mpc, which is $\sim$2.5 times larger than that of local, typical SFGs (average 1.52 Mpc). With a Kolmogorov-Smirnov test ($p = 1.9 \times 10^{-3}$), we significantly confirm that our metal-poor galaxies are located in the relatively isolated environment compared to the local, typical SFGs.

- We find that our metal-poor galaxy sample encompasses low gas-phase metallicities, 12+log(O/H)=6.97–8.50, low stellar masses, $\log(M_\star/M_\odot)$=5.0–7.1, and high specific star-formation rates (sSFR$\sim$300 Gyr^{-1}), suggesting the possibility that they are analogs of high-z, low-mass SFGs.

- We find that 5 out of our 10 metal-poor galaxies with the spectroscopically confirmation have moderate metallicities of 12+log(O/H)$\sim$8.0 in spite of their very low-$M_\star$ (i.e., $\log(M_\star/M_\odot)$=5–7), which are located above an extrapolation of the local mass-metallicity relation. One of possible explanations is that the 5 galaxies above the local mass-metallicity relation are in an early stage of the transition from gas-rich dwarf irregulars to gas-poor dwarf spheroidals or ellipticals.

- We confirm that HSC J1631+4426 shows one of the lowest metallicity values reported ever, 12+log(O/H)=6.97$\pm$0.08 (i.e., $Z/Z_\odot$=0.019).

- Our metal-poor galaxies fall on the SFG region of the BPT diagram, and we do not find any evidence that our metal-poor galaxies are affected by an AGN or shock heating from the optical emission line ratios. However, we do not exclude

the possibility of the existence of a metal-poor AGN or shock because little is known about the low-metallicity AGN or shock to date.

- We roughly measure velocity dispersions of our metal-poor galaxies with an Hβ emission line, which may trace the ionized gas kinematics. Thanks to a high spectral resolution of MagE ($R\sim$4,000), we find that our metal-poor galaxies have small velocity dispersions of σ=27.8–32.5 km s^{-1}. The velocity dispersions of our metal-poor galaxies are consistent with a relation between the velocity dispersion and V-band magnitude, which is made by a sequence of low-z bright galaxies, dwarf galaxies in Local Group, and globular clusters, although our velocity dispersion estimates can include systematic uncertainties.

- We estimate element abundance ratios of Ne/O, Ar/O, and N/O of our metal-poor galaxies, and compare them with local SFGs. We find that α-element ratios of Ne/O and Ar/O show almost constant values of log(Ne/O) ~-0.8 and log(Ne/O)~-2.4 as a function of metallicity, respectively. These constant Ne/O and Ar/O values are consistent with those of local SFGs. Most of our metal-poor galaxies have N/O ratios below log(N/O)$\lesssim-1.4$, suggesting that our metal-poor galaxies are undergoing the primary nucleosynthesis of nitrogen due to their low metallicity and young stellar population.

- We also estimate Fe/O ratios of our metal-poor galaxies, and compare them with local SFGs. Our metal-poor galaxy sample shows a decreasing Fe/O trend with increasing metallicity, consistent with results of local SFGs. We find that the two EMPGs, HSC J1631+4426 and SDSS J2115−1734, show higher Fe/O ratios than observation results of Galactic stars and model calculation of the Fe/O evolution at fixed metallicity. Especially, HSC J1631+4426 shows the solar Fe/O ratio in spite of its very low metallicity, 0.019 $(\mathrm{O/H})_\odot$. We discuss the three scenarios that may be able to explain the high Fe/O ratios with the extremely low metallicity: (1) the preferential dust depletion of iron, (2) a combination of metal enrichment and gas dilution caused by inflow, and (3) very massive stars above 300 $M_\odot$. The scenario (1) is ruled out because the

solar Fe/O ratios are not achieved by the dust depletion and we do not see any correlation between the dust extinction and the Fe/O ratios. We also exclude the scenario (2) because the observed N/O ratios are lower than the expected solar N/O ratio when the scenario (2) is true. Thus, we conclude that the high Fe/O ratios of the two EMPGs are attributed to very massive stars above 300 $M_{\odot}$, which is consistent with the young stellar ages of EMPGs ($\lesssim$ 50 Myr).

- To probe ionizing radiation in our metal-poor galaxies, we inspect emission lines from various ions covering a wide range of ionization potentials. We first choose Hβ, [O II]3727, [Ar III]4740, [O III]5007, and [Ar IV]7136 lines, which are sensitive to ionizing photon above 13.6, 13.6, 27.6, 35.1, and 40.7 eV. Our metal-poor galaxies and local, average SFGs show sequences of [O II]3727/Hβ, [Ar III]4740/Hβ, [O III]5007/Hβ, and [Ar IV]7136/Hβ as a function of metallicity, and match each other within small scatters. The match between the two samples suggests that our metal-poor galaxies and local, average SFGs have similar ionizing radiation in the range of 13.6–40.7 eV.

- We find that the sequences of [O II]3727/Hβ, [Ar III]4740/Hβ, [O III]5007/Hβ, and [Ar IV]7136/Hβ show peaks at around 12+log(O/H)$\sim$8.7, 8.3, 8.0, and 7.2–7.7, respectively. We confirm that higher ionization lines show peaks at lower metallicities. This peak transition suggests that ionizing radiation is more intense and/or harder at lower metallicities in the range of 13.6–40.7 eV.

- We also find similar sequences in [O II]3727/He II4686, [Ar III]4740/He II4686, [O III]5007/He II4686, and [Ar IV]7136/He II4686 ratios as a function of metallicity. We find that three out of our metal-poor galaxies, SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038 show deviations from the sequences (i.e., higher line ratios). We confirm that the deviations are attributed to weak He II4686 line compared to Hβ, and suggest that SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038 lacks high energy radiation sources above 54.4 eV. High mass X-ray binary (HMXB) models demonstrate that SDSS J2253+1116, SDSS J2310−0211, and HSC J1142−0038 can be in very early

phases of HMXB evolution ($\lesssim$5 Myr) before or in the middle of the first compact object formation.

- We find that five out of our metal-poor galaxies are not explained by the HMXB model of Schaerer et al. (2019), which fall in the ranges of $EW_0(H\beta)>100$ Å and $\log(He\,II4686/H\beta)>(-2.0)$. The HMXB has not yet been formed at the very young ages ($\lesssim$5 Myr) corresponding to $EW_0(H\beta)>100$ Å. Interestingly, three out of the five metal-poor galaxies are EMPGs, HSC J1631+4426, HSC J2314+0154, and SDSS J2115−1734. We suggest two new possible explanations for the five galaxies currently not explained by either the HMXB models or the latest binary stellar synthesis models, BEAGLE. One explanation is a metal-poor AGN, which may contribute to boost the He II4686 intensity of the very young galaxies even at $\lesssim$5 Myr. The other explanation is the past existence of very massive stars beyond 300 $M_\odot$. Very massive stars above 300 $M_\odot$ have very short lifetimes of $\sim$2 Myr, and form intermediate-mass black holes (IMBHs) with $\gtrsim$200 $M_\odot$ as early as $\sim$2 Myr after the star formation. Part of the IMBHs may form the HMXB system. The early HMXB onset can explain the high He II4686/Hβ ratios seen in galaxies younger than 5 Myr. EMPGs, such as HSC J1631+4426, HSC J2314+0154, and SDSS J2115−1734, might form very massive star beyond 300 $M_\odot$ from their extremely metal-poor gas. To verify our scenarios, self-consistent HMXB modeling is necessary with a higher maximum stellar mass cut above $\sim$300 $M_\odot$.

This thesis is based on the first-phase results of this EMPG study, which is expected to continue extending our understandings of high-z, low-mass galaxies. The author of this thesis has contributed to the establishment of the EMPG study by developing a new EMPG selection based on DL, proposing spectroscopic observations, reducing and analyzing the spectroscopic/photometric data, and leading discussions of HSC project 251 (PI: Takashi Kojima) consisting of more than 30 members.

REFERENCES

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, 1

Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114

Albareti, F. D., Prieto, C. A., Almeida, A., et al. 2017, ApJS, 233, 25

Aller, L. H. 1984, Astrophysics and Space Science Library, Vol. 112, Physics of Thermal Gaseous Nebulae (Dordrecht: Springer Netherlands), doi:10.1007/978-94-010-9639-3

Amorín, R. O., Pérez-Montero, E., & Vílchez, J. M. 2010, ApJ, 715, L128

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140

Annibali, F., Cignoni, M., Tosi, M., et al. 2013, AJ, 146, 144

Annibali, F., La Torre, V., Tosi, M., et al. 2019, MNRAS, 482, 3892

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, Astrophysics and Space Science, 328, 179

Axelrod, T., Kantor, J., Lupton, R. H., & Pierfederici, F. 2010, Proc. SPIE, 7740, 774015

Bai, G. S., Zhao, G., Chen, Y. Q., et al. 2004, A&A, 425, 671

Baker, J. G., Menzel, D. H., & Aller., L. H. 1938, ApJ, 88, 422

Baldwin, A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 817

Barat, D., D'Eugenio, F., Colless, M., et al. 2019, MNRAS, 487, 2924

Barkat, Z., Rakavy, G., & Sack, N. 1967, PRL, 18, 379

Bensby, T., & Feltzing, S. 2006, MNRAS, 367, 1181

Bensby, T., Yee, J. C., Feltzing, S., et al. 2013, A&A, 549, A147

Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, 1

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151

Brinchmann, J., Kunth, D., & Durret, F. 2008, A&A, 485, 657

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582

Campbell, A., Terlevich, R., & Melnick, J. 1986, MNRAS, 223, 811

Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Carretta, E., Gratton, R. G., & Sneden, C. 2000, A&A, 356, 238

Castelli, F., & Kurucz, R. L. 2004, arXiv e-prints, arXiv:0405087

Cayrel, R., Depagne, E., Spite, M., et al. 2004, A&A, 416, 1117

Chabrier, G. 2003, PASP, 115, 763

Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415

Christensen, L., Laursen, P., Richard, J., et al. 2012a, MNRAS, 427, 1973

Christensen, L., Richard, J., Hjorth, J., et al. 2012b, MNRAS, 427, 1953

Coupon, J., Czakon, N., Bosch, J., et al. 2018, PASJ, 70, 1

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44

Draine, B. T. 2010, Physics of the Interstellar and Intergalactic Medium (Princeton University Press), doi:10.2307/j.ctvcm4hzr

Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, PASA, 34, e058

Elmegreen, B. G., Herrera, C., Rubio, M., et al. 2018, ApJL, 859, L22

Elmegreen, D. M., & Elmegreen, B. G. 2017, ApJ, 851, L44

Engelbracht, C. W., Rieke, G. H., Gordon, K. D., et al. 2008, ApJ, 678, 804

Esteban, C., García-Rojas, J., Carigi, L., et al. 2014, MNRAS, 443, 624

Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, in Proc. SPIE, ed. M. Iye & A. F. M. Moorwood, Vol. 4841, 1657

Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761

Ferland, G. J., Porter, R. L., Van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137

Filho, M. E., Almeida, J. S., Muñoz-Tuñón, C., et al. 2015, ApJ, 802, 82

Filho, M. E., Winkel, B., Sánchez Almeida, J., et al. 2013, A&A, 558, A18

Filippenko, A. V. 1982, PASP, 94, 715

Fragos, T., Lehmer, B. D., Naoz, S., Zezas, A., & Basu-Zych, A. 2013a, ApJL, 776, L31

Fragos, T., Lehmer, B., Tremmel, M., et al. 2013b, ApJ, 764, 41

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Furusawa, H., Koike, M., Takata, T., et al. 2018, PASJ, 70, 1

Garnett, D. R. 1992, AJ, 103, 1330

Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279

Gratton, R. G., Carretta, E., Claudi, R., Lucatello, S., & Barbieri, M. 2003, A&A, 404, 187

Groves, B. A., Dopita, M. A., & Sutherland, R. S. 2004a, ApJS, 153, 9

—. 2004b, ApJS, 153, 75

Guseva, N. G., Izotov, Y. I., Fricke, K. J., & Henkel, C. 2017, A&A, 599, A65

Guseva, N. G., Izotov, Y. I., Papaderos, P., & Fricke, K. J. 2007, A&A, 464, 885

Gutkin, J., Charlot, S., & Bruzual, G. 2016, MNRAS, 462, 1757

Harikane, Y., Ouchi, M., Shibuya, T., et al. 2018, ApJ, 859, 84

Harris, W. E. 1996, AJ, 112, 1487

Heger, A., & Woosley, S. E. 2002, ApJ, 567, 532

Hirschauer, A. S., Salzer, J. J., Skillman, E. D., et al. 2016, ApJ, 822, 108

Hsyu, T., Cooke, R. J., Prochaska, J. X., & Bolte, M. 2017, ApJ, 845, L22

—. 2018, ApJ, 863, 134

Hunt, L. K., Garcia-Burillo, S., Casasola, V., et al. 2015, A&A, 583, A114

Inoue, A. K. 2011, MNRAS, 415, 2920

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Iwamoto, K., Brachwitz, F., Nomoto, K., et al. 1999, ApJS, 125, 439

Izotov, Y. I., Guseva, N. G., Fricke, K. J., & Henkel, C. 2019a, A&A, 623, A40

Izotov, Y. I., Guseva, N. G., Fricke, K. J., & Papaderos, P. 2009, A&A, 503, 61

Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, MNRAS, 474, 4514

Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955

Izotov, Y. I., & Thuan, T. X. 1998, ApJ, 497, 227

—. 2007, ApJ, 665, 1115

—. 2016, MNRAS, 457, 64

Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2012, A&A, 546, A122

—. 2019b, MNRAS, 483, 5491

Izotov, Y. I., Thuan, T. X., Guseva, N. G., & Liss, S. E. 2018b, MNRAS, 473, 1956

Izotov, Y. I., Thuan, T. X., & Stasińska, G. 2007, ApJ, 662, 15

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

Jurić, M., Kantor, J., Lim, K.-T., et al. 2015, arXiv e-prints, arXiv:1512.07914

Kashikawa, N., Aoki, K., Asai, R., et al. 2002, PASJ, 54, 819

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055

Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. 2018, PASJ, 70, 1

Kelson, D. 2003, PASP, 115, 688

Kelson, D. D., Illingworth, G. D., van Dokkum, P. G., & Franx, M. 2000, ApJ, 531, 159

Kennicutt, R. C. 1998, ARA&A, 36, 189

Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35

Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013, ApJ, 774, 100

Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183

Kikuchihara, S., Ouchi, M., Ono, Y., et al. 2019, arXiv e-prints, arXiv: 1905.06927

Kingma, D. P., & Ba, J. L. 2015, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, arXiv: 1412.6980

Kirby, E. N., Cohen, J. G., Simon, J. D., & Guhathakurta, P. 2015a, ApJ, 814, L7

Kirby, E. N., Simon, J. D., & Cohen, J. G. 2015b, ApJ, 810, 56

Kniazev, A. Y., Grebel, E. K., Hao, L., et al. 2003, ApJ, 593, L73

Kniazev, A. Y., Pustilnik, S. A., Grebel, E. K., Lee, H., & Pramskij, A. G. 2004, ApJS, 153, 429

Kojima, T., Ouchi, M., Nakajima, K., et al. 2017, PASJ, 69, 1

Komiyama, Y., Obuchi, Y., Nakaya, H., et al. 2018, PASJ, 70, 1

Kroupa, P. 2001, MNRAS, 322, 231

Kroupa, P., & Weidner, C. 2003, ApJ, 598, 1076

Kunth, D., & Östlin, G. 2000, A&A, 10, 1

Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436

Lecureur, A., Hill, V., Zoccali, M., et al. 2007, A&A, 465, 799

Lin, W., & Ishak, M. 2016, Journal of Cosmology and Astroparticle Physics, 2016, 025

López-Sánchez, Á. R., & Esteban, C. 2008, A&A, 491, 131

Luridiana, V., Morisset, C., & Shaw, R. A. 2015, A&A, 573, A42

Lusso, E., Comastri, A., Vignali, C., et al. 2010, A&A, 512, A34

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415

Mainali, R., Kollmeier, J. A., Stark, D. P., et al. 2017, ApJ, 836, L14

Marshall, J. L., Burles, S., Thompson, I. B., et al. 2008, in Proc. SPIE, ed. I. S. McLean & M. M. Casali, Vol. 7014, 701454

Martin, N. F., Geha, M., Ibata, R. A., et al. 2016, MNRASL, 458, L59

McConnachie, A. W. 2012, AJ, 144, 4

Micheva, G., Christian Herenz, E., Roth, M. M., Östlin, G., & Girichidis, P. 2019, A&A, 623, A145

Miyazaki, S., Komiyama, Y., Nakaya, H., et al. 2012, in Proc. SPIE, ed. I. S. McLean, S. K. Ramsay, & H. Takami, 84460Z

Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, 1

Morandi, A., Limousin, M., Sayers, J., et al. 2012, MNRAS, 425, 2069

Nagao, T., Maiolino, R., & Marconi, A. 2006, A&A, 459, 85

Nakajima, K., Ellis, R. S., Iwata, I., et al. 2016, ApJ, 831, L9

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

Ohkubo, T., Umeda, H., Maeda, K., et al. 2006, ApJ, 645, 1352

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Ono, Y., Ouchi, M., Shimasaku, K., et al. 2010, ApJ, 724, 1524

Ono, Y., Ouchi, M., Harikane, Y., et al. 2018, PASJ, 70, 2

Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42

Osterbrock, D. E., & Michael Shull, J. 1989, Physics Today, 42, 123

Peeples, M. S., Pogge, R. W., & Stanek, K. Z. 2008, ApJ, 685, 904

Pérez-Montero, E. 2014, MNRAS, 441, 2663

Pérez-Montero, E., & Amorín, R. 2017, MNRAS, 467, 1287

Pérez-Montero, E., & Contini, T. 2009, MNRAS, 398, 949

Pérez-Montero, E., Contini, T., Lamareille, F., et al. 2013, A&A, 549, A25

Pilyugin, L. S., Grebel, E. K., & Mattsson, L. 2012, MNRAS, 424, 2316

Prugniel, P., & Simien, F. 1996, A&A, 309, 749

Pustilnik, S. A., Kniazev, A. Y., & Pramskij, A. G. 2005, A&A, 443, 91

Pustilnik, S. A., Tepliakova, A. L., Kniazev, A. Y., Martin, J. M., & Burenkov, A. N. 2010, MNRAS, 401, 333

Rhode, K. L., Salzer, J. J., Haurberg, N. C., et al. 2013, AJ, 145, 149

Rodriguez, M., & Rubin, R. H. 2005, ApJ, 626, 900

Sacchi, E., Annibali, F., Cignoni, M., et al. 2016, ApJ, 830, 3

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Sánchez Almeida, J., Muñoz-Tuñón, C., Elmegreen, D. M., Elmegreen, B. G., & Méndez-Abreu, J. 2013, ApJ, 767, 74

Sánchez Almeida, J., Pérez-Montero, E., Morales-Luis, A. B., et al. 2016, ApJ, 819, 110

Sánchez Almeida, J., Elmegreen, B. G., Muñoz-Tuón, C., et al. 2015, ApJL, 810, L15

Schaerer, D., Fragos, T., & Izotov, Y. I. 2019, A&A, 622, L10

Selsing, J., Fynbo, J. P., Christensen, L., & Krogager, J. K. 2016, A&A, 585, A87

Senchyna, P., & Stark, D. P. 2019, MNRAS, 484, 1270

Senchyna, P., Stark, D. P., Vidal-García, A., et al. 2017, MNRAS, 472, 2608

Shaw, R. A., & Dufour, R. J. 1995, PASP, 107, 896

Shim, H., Chary, R. R., Dickinson, M., et al. 2011, ApJ, 738, 69

Shirazi, M., & Brinchmann, J. 2012, MNRAS, 421, 1043

Shivaei, I., Kriek, M., Reddy, N. A., et al. 2016, ApJ, 820, L23

Simon, J. D., Drlica-Wagner, A., Li, T. S., et al. 2015, ApJ, 808, 95

Skillman, E. D. 1989, ApJ, 347, 883

Skillman, E. D., Salzer, J. J., Berg, D. A., et al. 2013, AJ, 146, 3

Stanway, E. R., Eldridge, J. J., & Becker, G. D. 2016, MNRAS, 456, 485

Stark, D. P., Richard, J., Siana, B., et al. 2014, MNRAS, 445, 3200

Stark, D. P., Walth, G., Charlot, S., et al. 2015, MNRAS, 454, 1393

Stasińska, G. 1982, A&AS, 48, 299

Stasińska, G., & Izotov, Y. 2003, A&A, 397, 71

Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, A&A, 576, A83

Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159

Strömgren, B. 1939, ApJ, 89, 526

Suzuki, A., & Maeda, K. 2018, ApJ, 852, 101

Takeda, Y. 2003, A&A, 402, 343

Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, ApJ, 565, 773

Thuan, T. X., & Izotov, Y. I. 2005, ApJS, 161, 240

Tominaga, N., Umeda, H., & Nomoto, K. 2007, ApJ, 660, 516

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Vanzella, E., Balestra, I., Gronke, M., et al. 2017, MNRAS, 465, 3803

Vincenzo, F., Belfiore, F., Maiolino, R., Matteucci, F., & Ventura, P. 2016, MNRAS, 458, 3466

Wise, J. H., Turk, M. J., Norman, M. L., & Abel, T. 2012, ApJ, 745, 50

Xiao, L., Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 477, 904

Yang, H., Malhotra, S., Rhoads, J. E., et al. 2017a, ApJ, 838, 4

Yang, H., Malhotra, S., Rhoads, J. E., & Wang, J. 2017b, ApJ, 847, 38

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Yungelson, L. R., Van Den Heuvel, E. P., Vink, J. S., Zwart, S. F., & De Koter, A. 2008, A&A, 477, 223

Zahid, H. J., Bresolin, F., Kewley, L. J., Coil, A. L., & Davé, R. 2012, ApJ, 750, 120