CrossMark

Linking Extragalactic Transients and Their Host Galaxy Properties: Transient Sample, Multiwavelength Host Identification, and Database Construction

Yu-Jing Qin[1] , Ann Zabludoff[1] , Marina Kisley[2] , Yuantian Liu[3], Iair Arcavi[4,5] , Kobus Barnard[2] , Peter Behroozi[1] ,
K. Decker French[6,9] , Curtis McCully[7] , and Nirav Merchant[8]

[1] Department of Astronomy and Steward Observatory, University of Arizona, 933 N Cherry Ave., Tucson, AZ 85721, USA; qinyj@email.arizona.edu
[2] Department of Computer Science, University of Arizona, 1040 4th St., Tucson, AZ 85721, USA
[3] Department of Physics, University of Arizona, 1118 E 4th St., PO Box 210081, Tucson, AZ 85721, USA
[4] The School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel
[5] CIFAR Azrieli Global Scholars program, CIFAR, Toronto, Canada
[6] Department of Astronomy, University of Illinois, Urbana, IL 61801, USA
[7] Las Cumbres Observatory, 6740 Cortona Drive, Goleta, CA 93117, USA
[8] Data Science Institute, University of Arizona, BSRL Building (#242), 1230 N Cherry Ave., Tucson, AZ 85721, USA

Abstract

Understanding the preferences of transient types for host galaxies with certain characteristics is key to studies of transient physics and galaxy evolution, as well as to transient identification and classification in the LSST era. Here we describe a value-added database of extragalactic transients—supernovae, tidal disruption events, gamma-ray bursts, and other rare events—and their host galaxy properties. Based on reported coordinates, redshifts, and host galaxies (if known) of events, we cross-identify their host galaxies or most likely host candidates in various value-added or survey catalogs, and compile the existing photometric, spectroscopic, and derived physical properties of the host galaxies in these catalogs. This new database covers photometric measurements from the far-ultraviolet to mid-infrared. Spectroscopic measurements and derived physical properties are also available for a smaller subset of hosts. For our 36,333 unique events, we have cross-identified 13,753 host galaxies using host names, plus 4480 using host coordinates. Besides those with known hosts, there are 18,100 transients with newly identified host candidates. This large database will allow explorations of the connections of transients to their hosts, including a path toward transient alert filtering and probabilistic classification based on host properties.

Unified Astronomy Thesaurus concepts: Transient sources (1851); Gamma-ray bursts (629); Supernovae (1668); Tidal disruption (1696); Astronomy data analysis (1858); Radio transient sources (2008)

1. Introduction

Extragalactic transients cover a wide range of energetic, fast, and usually cataclysmic events observed across the electromagnetic spectrum, as well as the recently revealed gravitational wave (GW) and neutrino windows. These events include, but are not limited to, supernovae (SNe), gamma-ray bursts (GRBs), tidal distribution events (TDEs), fast radio bursts (FRBs), and GW events. With better observational facilities and computational technologies, our understanding of these transient phenomena has advanced considerably in the new century. However, their connection to the evolution history and observational properties of their host galaxies remains to be further explored.

Transients are stellar or supermassive black hole phenomena in nature. They occur on much smaller spatial scales than the typical size of galaxies. They are also ephemeral compared to timescales of galaxy evolution. Notwithstanding the dramatic contrast in spatial and temporal scales, there are examples of well-established transient-host connections:

1. Core-collapse (CC) SNe show stronger preference for late-type, star-forming galaxies than thermonuclear Type Ia supernovae (SNe Ia; e.g., van den Bergh 1959; Tammann 1978; Oemler & Tinsley 1979; Sullivan et al. 2006; Li et al. 2011; Foley & Mandel 2013; Graur et al. 2017).

2. Compared to *normal* SNe Ia, SNe Ia-91bg tend to occur in early-type hosts, while SNe Ia-91T and Ia-02cx favor late types (Taubenberger 2017, and references therein).

3. The hosts of short-duration gamma-ray bursts (SGRBs) are brighter, larger in size, more metal-rich, and, on average, more quiescent than typical hosts of long-duration gamma-ray bursts (LGRBs; Berger 2009; Fong et al. 2010; Lyman et al. 2017).

4. The rate of TDEs is dramatically boosted in post-starburst galaxies (French et al. 2016), although the underlying driving mechanism is subject to debate (Graur et al. 2018; Stone et al. 2018).

5. The recently discovered electromagnetic counterpart of GW event GW170817 (Abbott et al. 2017), also detected as an SGRB, lies in an early-type galaxy with shell-like merger features.

6. The known hosts of FRBs show rather diverse properties, with a possible excess of galaxies in the *green valley*, or galaxies with harder ionizing radiation than provided by star formation alone (e.g., Heintz et al. 2020).

This synopsis is certainly not exhaustive but nevertheless demonstrates that some transients prefer certain host environments. There may be still other connections that, while subtle, could be discovered with a large, well-curated database gleaned from archived surveys.

---

[9] Hubble Fellow.

To a certain extent, such transient-host connections can be interpreted within our existing picture of transient physics and galaxy evolution. From a statistical and empirical perspective, the event rates of optical SNe and their counterparts at other wavelengths depend on the recent star formation history and the delay time distribution of supernova explosion since star formation; theoretically, the preferred channel and mechanism of the explosion could also depend on the metallicity and dynamical environment, which directly shape the evolutionary path of progenitors and indirectly affect stellar binarity and initial mass function. For TDEs, the event rates not only are related to the integrated stellar populations of the host but also are regulated by the concentration of star formation, availability of gas near the galactic nucleus, and/or the disturbed central stellar and gas dynamics in the host. Due to the uncertainties in our theoretical models, it is unclear to what extent the rates of transients depend on the global properties of their hosts. Potential biases in transient surveys further complicate the interpretation of the observed transient-host connections. Nevertheless, the study of transient-host connections points to a better-refined view of transient phenomena and galaxy evolution, including the physics of massive stars, interacting binaries, and compact objects and the role of supermassive black holes and active nuclei in galaxy evolution.

Transient-host connections are also of practical interest. Modern transient surveys, particularly the Legacy Survey of Space and Time (LSST) at Vera C. Rubin Observatory (Ivezić et al. 2019), are expected to deliver nearly 10 million transient alerts every night in real time. For extragalactic events, over a year of operation, LSST is estimated to detect about 0.2 million SNe Ia, a smaller but comparable sample of CC SNe, and orders of magnitude fewer other uncommon transients inundated in the torrent of alerts.[10] Selecting a subset of potentially interesting events would be essential for any study that requires follow-up observations using other facilities. Often this must be done *quickly* in the early phase of transient evolution, when evanescent phenomena may give us valuable insights into their physical models. Host properties thus open up a new avenue toward automated alert filtering and classification, even *preclassification*, i.e., assigning event rates or probabilities to galaxies in the field based on their properties, before any alerts arrive. Indeed, the viability and reliability of host-based transient classification has been clearly demonstrated by the early conceptual work of Foley & Mandel (2013) and recent attempts of Gagliano et al. (2021). Even photometric data alone may effectively distinguish potential hosts of certain rare transients (e.g., French & Zabludoff 2018). Meanwhile, new-generation transient brokers for Zwicky Transient Facility (Bellm et al. 2019) and the future LSST, such as ANTARES (Narayan et al. 2018), ALeRCE (Förster et al. 2021), Lasair (Smith et al. 2019), MARS,[11] and Fink (Möller et al. 2021), are all capable of cross-matching transient alerts with external source catalogs, providing the necessary infrastructure for host-based event filtering and classification. Given this potential to provide a baseline estimate of transient probabilities (i.e., a categorical prior function of different types) or rates independent of any transient characteristics, the power of host properties must be fully utilized now for imminent transient surveys.

Exploring the connections of extragalactic transients and their host properties entails a complete census of known transient hosts, including their properties measured by various sky surveys across the electromagnetic spectrum. Existing samples and catalogs of transient host galaxies, such as the time-controlled sample of Lick Observatory Supernova Search (LOSS; Leaman et al. 2011; Li et al. 2011), also the host samples of SDSS-II Supernova Survey (Sako et al. 2018) and the Dark Energy Survey (DES) Supernova Programme (Wiseman et al. 2020), usually focus on a dedicated transient survey or a specific class of transients. The recently released GHOSTdatabase (Gagliano et al. 2021), which has a considerably larger transient sample, is mostly limited to the photometric properties of supernova hosts obtained by the Pan-STARRS (PS1) survey. There is no single catalog to date of consistently measured host properties, across a wide wavelength range, for a large and up-to-date sample of transient events. Reasons include the following:

1. Host galaxies are not always reported by transient surveys or summarized in catalogs or platforms. Consequently, a considerable fraction of transients in the literature, even with accurate sky coordinates, redshifts, and classifications, have no easily accessible host data of any kind.
2. Beyond that, new transients are continuously being discovered at ever-improving efficiencies, posing an immediate challenge to human-based host identifying and reporting. Automated host identifying algorithms, using either catalogs or images, are favored in this situation. However, only a few such algorithms have been developed up to date. (e.g., Sullivan et al. 2006; Gupta et al. 2016; Sako et al. 2018; Gagliano et al. 2021).
3. Even for transients with identifiable hosts, obtaining their properties involves substantial efforts. Some surveys conduct dedicated observations of hosts. Accessing archival data of public surveys, a more viable approach for many other cases, also requires host matching, quality control, and data compilation.

Constructing a transient-host database with better coverage of previous transient events and available host properties remains a crucial step for the in-depth exploration and analysis of transient-host connections. Here we present such a new transient-host database. We find counterparts of host galaxies in external catalogs using the reported host name designations or coordinates, as presented in the data sources. For events without reported host galaxies, we identify their best host candidates within nearby extragalactic objects, cross-matched in the various external catalogs. Our goal here is to locate the correct counterparts of transient host galaxies in various catalogs, which entails cross-matching sources in multiple catalogs, and, when necessary, ranking potential host galaxies to identify the most likely candidates. Therefore, we refer to the process as the *cross-identification* of transient host galaxies. We then collate properties of our identified host galaxies and host candidates, including photometry across UV to IR wavelengths, morphology, spectral lines and indices, and derived physical parameters such as stellar mass, metallicity, and star formation rate provided in those catalogs.

This first paper describes our database of transient host galaxies and the workflow to construct it. In Section 2, we review our data sources for transient events (published catalogs and online services) and the procedure to create a sample of

---

[10] LSST Science Book v2, Chapter 8 and Chapter 11.
[11] Mars.lco.global

unique transient events. In Section 3, we describe our method to search host properties in external catalogs and collect their information. We also discuss our strategy to identify new host galaxies for those without hosts reported in our data sources. We then describe the detailed object selection criteria and data coverage in each external catalog in Section 4, and we present the basic statistics of our transient-host pairs in Section 5. Finally, we discuss our software implementation and data format in Section 6 and summarize in Section 7.

2. The Transient Sample

This section describes our selected upstream data sources for transient events and the procedure to assemble a transient sample. We also present the basic statistics in these upstream data sources and the compiled transient sample as of 2021 June 1, i.e., the cutoff date of transient events for the initial release of our database presented in this paper.

2.1. Upstream Data Sources for Transients

Existing transient events are detected by various surveys or missions, reported at multiple circulars or platforms, and cataloged by several sites or numerous individual publications. There is no single data source that covers all previously known transient events. Therefore, we choose a few representative data sources with relatively complete records for each major transient type, organized in a machine-readable format for easier access. Below we discuss our transient data sources for each major type.

2.1.1. Supernovae

Our supernova records are imported from the Open Supernova Catalog (OSC; Guillochon et al. 2017).[12] The OSC compiles SNe in numerous published catalogs, individual works, and online services, notably the Transient Name Server (TNS),[13] Padova-Asiago Supernova Catalog (Barbon et al. 1999),[14] the Latest Supernova website[15] (Gal-Yam et al. 2013), the Weizmann Interactive Supernova data REPository (Yaron & Gal-Yam 2012),[16] and the CfA Supernova Data Archive,[17] etc. Apart from the basic information (or *metadata*) of SNe and supernova candidates, the catalog also compiles their light curves, spectra, X-ray, and radio measurements. For every single event, all related data are self-contained in a single file, where the historical versions of each file are traceable via their `git`-based version control solution.

Although great efforts are made in Guillochon et al. (2017) to collect and homogenize supernova data in various sources, for our transient sample, there are a few issues to be resolved when importing their supernova records. First, the supernova basic data (or metadata) in OSC—such as classification, sky coordinates, and redshifts—trace back to the multiple automatically updated or manually maintained reference sources. As a result, slightly different, inconsistent, or even erroneous values are often provided for the same data field (or column), requiring further selection and validation. Second, not all the

records in OSC are directly detected and reliably classified SNe. There are a large fraction of transient candidates in OSC without classification. There are also supernova remnants and historical SNe (i.e., discovered before photometric techniques became available) in the Milky Way or nearby galaxies. Long-duration GRBs in the CGRO/BATSE Catalog and grbcatalog. com, even without the associated SNe detected, are also included for the association of LGRBs with some CC SNe. Third, the type labels (including major types and detailed subtypes) in OSC can sometimes be ambiguous, incomplete, and even inconsistent. Reliable classification requires high-quality light curves and spectra covering critical phases of evolution, which are not always available. The criteria and technique for transient classification could also differ among survey programs or research groups. Even the classification scheme is evolving as our knowledge of supernova physics advances. As a result, events are not classified consistently among different sources.

For our work, we only select a subset of records in OSC, with further data curation for clarity and consistency. Records with classification, redshift *or* host information of any kind (names, or coordinates), are selected. Supernova candidates and unclassified events are omitted unless their redshifts or hosts are also reported. Those temporarily ignored candidates and unclassified events will be added back to our database during future updates once they are confirmed or classified in OSC. We also skip ancient SNe and supernova remnants, even if their types are indirectly inferred in the literature. Furthermore, we only select GRBs in OSC with the associated SNe detected (i.e., classified as SNe at the same time). Similarly, we omit records labeled as TDEs at this stage. TDEs and GRBs are added via other data sources (discussed later).

The snapshot of OSC we use here has 82,864 records, where 26,868 records are classified (i.e., with type labels assigned), and 21,592 records have host galaxy information, either name designations or coordinates. As new records and data are added into OSC by programs and users regularly, we set up an automatic script to pull the latest version of OSC hosted at github.com and update our transient records accordingly. Our database does not include detailed supplementary data of each event in OSC, including transient light curves, spectra, X-ray, and radio fluxes.

2.1.2. Tidal Disruption Events

Currently, there are only a few dozen tidal disruption events reported in the literature. For the initial release of our database, we use the summarized list in the review of French et al. (2020) as our primary data source. Additionally, we use the Open TDE Catalog,[18] a close sibling of the OSC, as our secondary data source.[19] Like OSC, the Open TDE Catalog is designed as a complete sample of known TDEs, including likely ones, in the literature or transient circulars, where basic data are compiled from various reference sources. This catalog also includes unconfirmed or possible events where alternative explanations or classifications may exist, which requires manual inspection and selection.

To incorporate these records into the database, we combined events in both data sources into a single, unified list of TDEs.

---

[12] sne.space, github.com/astrocatalogs/supernovae
[13] www.wis-tns.org
[14] graspa.oapd.inaf.it/cgi-bin/sncat.php
[15] www.rochesterastronomy.org/supernova.html
[16] wiserep.weizmann.ac.il/
[17] www.cfa.harvard.edu/supernova/SNarchive.html

---

[18] tde.space
[19] For future updates of the database, we may switch our primary data source to the Transient Name Server.

We manually match the events in these two data sources by name designations (including variants) to ensure that the combined list is free of duplicates. To verify our name-based matching and to further identify unique events in each data source, we also spatially cross-match events using an angular distance threshold of 5″. Finally, we use the metadata (name, coordinate, redshift, and host galaxy) in French et al. (2020) whenever available, but also keeping other names and aliases in the Open TDE catalog.

We also unified the type labels in our combined list of TDEs. The type labels within the Open TDE Catalog are still to be standardized. Some events are classified by their inferred progenitors or physical scenarios (e.g., "MS+SMBH"), while others are only labeled as "TDE" or "TDE?." We first assign "TDE" labels to all the events in the combined list. For events in the list of French et al. (2020), we further assign labels to indicate their detected wavelengths, like "UVOptTDE" for events detected in UV or optical bands and "XrayTDE" for X-ray detected events. Possible and likely X-ray detected events in French et al. (2020) are labeled as "PossibleX-rayTDE" and "LikelyXrayTDE," based on the samples of Auchettl et al. (2017). There are other proposed classification schemes in the literature (e.g., van Velzen et al. 2021), which could be included in the future release of our database.

There are 43 TDEs in the list of French et al. (2020) and 97 events in the Open TDE catalog (as of 2021 June 1). Excluding 41 common ones, the combined list has 98 unique events.

2.1.3. Gamma-Ray Bursts

We use the following catalogs of individual GRB missions for their relatively better localization accuracy than other catalogs.

1. Swift Catalog[20] (Lien et al. 2016) is a catalog of events detected by the Burst Alert Telescope (BAT) on Swift. BAT localizes GRBs to arcminute-level accuracy. Some events also have afterglow detection by the X-ray Telescope and UV/Optical Telescope on board, which further improves the positioning accuracy to arcsecond level.
2. Fermi/LAT Catalog[21] (Ackermann et al. 2013) includes events detected by the Large Area Telescope (LAT) on Fermi, with arcminute-level positioning accuracy. This does not include events only detected by the Gamma-Ray Burst Monitor, which has a larger error circle (degree-level in the best case).
3. INTEGRAL/IBIS Catalog (Vianello et al. 2009; Bošnjak et al. 2014) and later events localized with the INTEGRAL Burst Alert System[22] (Mereghetti et al. 2003) are events detected by the Imager on-Board the INTEGRAL Satellite (IBIS), with a typical positional error of about 2′. For high signal-to-noise events, the accuracy may reach the subarcminute level.

The catalogs listed above certainly do not cover all well-localized GRBs. Other missions, such as HETE-2, BeppoSAX, and AGILE, also report events that are well-localized via afterglows or by the associated SNe, usually with other facilities. We do not use their original catalogs due to the lack of precise coordinates there.

We also accessed several ancillary data sources for GRBs, including the catalog of the GRB Host Studies Project[23] (Savaglio 2006; Savaglio et al. 2006), the GRB list maintained by Jochen Greiner,[24] the Gamma-Ray Burst Online Index,[25] and the GRBweb catalog.[26] These ancillary data sources summarize GRB properties from multiple missions and follow-up programs, including data reported to the GCN circulars[27] that are not machine-readable at the moment. These ancillary data sources provide us with (1) potentially better coordinates than reported in the three mission catalogs above, (2) more well-localized GRBs that are not reported in those mission catalogs, and (3) most importantly, GRB redshifts measured from the afterglows, host galaxies, or associated SNe.

We use the latest online version of these mission catalogs and ancillary data sources to ensure that our GRB records are up-to-date. We combine records from these reference sources into a single list, during which we cross-match GRBs by their standardized names to eliminate duplicates. For each event, we choose the best sky coordinate (i.e., with the smallest 90% error circle) within these reference sources. Redshift and T_{90} (i.e., the time when 90% photons arrive), when available, are usually consistent across these reference sources. If not, we take the median value of all the reported measurements.

Events in our list are classified as GRB by default. According to their T_{90} values, we further classify them into long-duration GRBs (LGRBs; $T_{90} \geqslant 2$ s) and short-duration GRBs (SGRBs; $T_{90} < 2$ s). Events without T_{90} in any of our data sources are only classified as GRB. The traditional long-short dichotomy might be further refined using the spectral hardness (e.g., Zhang et al. 2012), which remains a part of our future work.

It should be emphasized that GRBs are, on average, detected at higher redshifts than other transients. The drastically decreased completeness of survey catalogs, combined with the large error in coordinates, makes the host hard or even impossible to be identified in some cases. Therefore, we use a rather conservative sample selection and catalog accessing strategy to reduce the fraction of misidentified hosts. We only include GRBs with a 90% error radius under 5″ in our sample. Meanwhile, the default search radius for GRBs without redshifts is reduced to minimize the confusion due to other nearby galaxies (Section 3). Even with more accurate coordinates and smaller search radii, the chance remains that some GRB hosts are undetected or incorrectly identified. Checking quality control metrics (Section 3.5) is always encouraged when using their compiled host properties.

Besides nine GRBs with associated optical SNe in OSC, we have 7939 unique GRBs from our selected data sources, where 1621 of them have a 90% error radius better than 5″ and are thus included in the transient sample.

2.1.4. Other Rare Transients

Our database also includes rare transients such as FRBs and GWs. These rare events are generally poorly localized, so we

---

[20] swift.gsfc.nasa.gov/archive/grb_table/fullview/
[21] fermi.gsfc.nasa.gov/ssc/observations/types/grbs/lat_grbs/table.php
[22] ibas.iasf-milano.inaf.it
[23] www.grbhosts.org
[24] https://www.mpe.mpg.de/~jcg/grbgen.html
[25] https://sites.astro.caltech.edu/grbox/grbox.php
[26] https://user-web.icecube.wisc.edu/~grbweb_public/
[27] gcn.gsfc.nasa.gov/gcn3_archive.html

only include 14 FRBs and one GW event with host galaxies reported in the literature out of over 100 FRBs and nine GWs. These events include GW170817 (Abbott et al. 2017), FRB121102 (Chatterjee et al. 2017), FRB180916 (Marcote et al. 2020), FRB180924 (Bannister et al. 2019), FRB181112 (Prochaska et al. 2019), FRB190102 (Macquart et al. 2020), FRB190523 (Ravi et al. 2019), FRB190608 (Macquart et al. 2020), and FRB190614 (Law et al. 2020).[28,29] Their metadata are referred from the Fast Radio Burst Catalogue (Petroff et al. 2016) and the list of detections made by the LIGO-Virgo-KAGRA Collaborations.[30] We also refer to the FRB Host database[31] (Heintz et al. 2020) for the host associations of known FRBs. For the rest of the events, we do not attempt to identify their host candidates.

Besides the one associated with GW170817, there are several more kilonova or *macronova* events (including candidates) reported in the literature. We label the event associated with GW170817 as "`Kilonova`." Events listed in the Open Kilonova Catalog[32] (another sibling of OSC), mostly short GRBs, are labeled as kilonova candidates ("`PossibleKilonova`"). These kilonova candidates include GRB050709 (Jin et al. 2016), GRB060614 (Yang et al. 2015), GRB080503 (Perley et al. 2009), GRB130603B (Fong et al. 2014), and GRB150101B (Troja et al. 2018).

2.1.5. Remarks on Data Source Selection

There are a few more actively maintained catalogs and online services for extragalactic transients, including the TNS,[33] which cross-identifies transient events; the Astronomer's Telegram (ATel)[34] where initial discovery, confirmation, and classification of transients are posted; and the GCN circular[35] where high energy transients are reported and archived. Many sky surveys or space missions also produce their own catalogs of detected events and candidates. There are also catalogs that are no longer being actively updated (e.g., the supernova catalog of Lennarz et al. 2012).

We use the data sources above as they are relatively complete (including events in multiple missions, surveys, and catalogs), and data are summarized and tabulated in a machine-readable form. The basic statistics of our data sources are summarized in Table 1. We may include other transient data sources for transients with records in an easily accessible format in future releases.

2.2. Construction of Transient Sample

Having transient records selected from those upstream data sources, we combine these records into a single, unified transient sample. We only keep their metadata for each event, including sky coordinates, redshift, type labels, host name, and/or coordinate (if available). We take the following

Table 1
Upstream Data Sources for Transients

Source	Number		Type
	Total	Used	
The Open Supernova Catalog	82,864	34,547	SN
French et al. (2020) Review	43	43	TDE
The Open TDE Catalog	97	96	TDE
Swift GRB Catalog	1651	1493	GRB
Fermi/LAT GRB Catalog	145	78	GRB
INTEGRAL/IBAS GRB Catalog	140	140	GRB
The Fast Radio Burst Catalogue	118	13	FRB
LIGO–Virgo–Kagra Detections	12	1	GW
The Open Kilonova Catalog	6	6	KN
GRB Host Studies (GHostS)[a]	235	⋯	GRB
Jochen Greiner's GRB List[a]	2226	⋯	GRB
Gamma-Ray Burst Online Index[a]	2082	⋯	GRB
GRBweb Database[a]	7559	⋯	GRB
FRB Host Galaxy Database[b]	19	⋯	FRB
Taubenberger (2017) Review[c]	60	⋯	SN

Notes. Number counts include all records in the original data source (*Total*) and the number of records that contributed any data in our sample (*Used*). Besides French et al. (2020), other reference sources are actively updated. Numbers are as of 2021 June 1.
[a] Ancillary data source for GRB localization, T_{90}, and redshifts.
[b] Ancillary data source for FRB host association.
[c] Ancillary data source for SNe Ia classification.

approaches to ensure that events in the combined transient sample are unique, are consistently classified, and have their best-available metadata compiled.

2.2.1. Ensuring the Uniqueness of Events

We first make our transient sample unique to each event. A transient can be independently detected, with unique identifiers assigned by different surveys, missions, or research groups, in their own naming conventions. When events are reported to circulars like TNS and ATel, there could be another assigned designation (e.g., "AT" prefix + year + alphabetic identifier) for the event. Finally, when confirmed with follow-up observations, there might be an *official* name assigned using the naming convention of the community. These aliases and designations should point to the same event in our transient sample. Meanwhile, future events could be detected by multiple messengers, and they should point to the same entry in the database, provided that clear associations are established. This is already partly done within OSC but not across all of our upstream data sources.

To make the table unique to each event, besides the preferred or most commonly used designations, we also keep a list of aliases for each event. When a new event is reported at our transient data sources, we first search the alias lists of existing events to check if this is already assigned to a record. We create a new record for this event only when all of its names are new in our transient sample. When maintaining alias lists, we also considered variants of transient designations in the literature. For example, starting on 2010 January 1, the first GRB detected on a day is always labeled as "A," even if there is no other GRB on that day (Barthelmy et al. 2009). However, such a change of a naming convention is not strictly obeyed in the literature. There are also minor variations like with or without white spaces or dashes (e.g., ASASSN versus ASAS-SN). We created possible

---

[28] Heintz et al. (2020) and Bhandari et al. (2020) have recently reported the host galaxies of FRB190611, FRB190714, FRB191001, and FRB200434, which will be included in the next release of the database.
[29] FRB150418 (Keane et al. 2016) is not included as its *afterglow*, with which the host was located, and is likely a coincident variable source (Williams & Berger 2016). Also, the case of FRB171020 (Mahony et al. 2018) is not included due to its indefinitive host association.
[30] www.ligo.org/detections.php
[31] frbhosts.org
[32] kilonova.space
[33] wis-tns.weizmann.ac.il
[34] www.astronomerstelegram.org
[35] gcn.gsfc.nasa.gov

variants of event names, even though they do not strictly follow the naming conventions of certain transient communities or survey programs. This effectively prevents duplicates when the database is updated in the future.

Finally, we create unique identifiers for each transient event in our database ("_id"); identifiers are used consistently everywhere for host cross-matching and host candidate identification.

2.2.2. Finding the Best-available Metadata

When assembling the transient sample using several different data sources, we may also have multiple coordinates or redshifts reported for the same event beside multiple type labels. Finding the best-available sky coordinates and redshifts would thus be necessary for our later host cross-matching or identification. We only use the *most referred* coordinates for SNe, i.e., the sky coordinates that are most widely used in various data sources. We do this by pairing R.A./decl. by their data sources and counting the number of primary references linked to each coordinate pair reported in OSC. When available, we always prefer the spectroscopic redshift of the host galaxy over the spectroscopic redshift of the SNe, as we are assembling a catalog of transient host galaxies. Photometric redshifts of any kind have the lowest priority. For redshift values without types indicated, we assume that they are the spectroscopic redshifts of the transients. Finally, when more than one redshift of the same priority level are reported, we take the median value.

Most GRBs have no associated SNe detected, except for a handful of cases. We always use the coordinates with the smallest 90% error circle for these events. We also include positional accuracy (90% error radius in arcseconds) for GRBs. Many GRBs have X-ray or UV/optical afterglows detected, which provides more accurate localization (arcsecond-level) than the coarse coordinates determined with wide-angle, coded mask detectors (usually arcminute level in the best cases). When available, we use the most accurate position determined using X-ray or optical afterglows. Finally, when associated optical SNe are detected, we prefer supernova coordinates and redshifts, over those reported for GRB, for the same reason.

Commonly assumed as nuclear events, the reported transient coordinates of TDEs are usually their host coordinates in optical wavelengths. Instead, a small subset of TDEs may only have X-ray determined transient coordinates reported, with relatively poor localization accuracy. We take the as-reported positions as their *nominal* host coordinates when matching objects in other catalogs (see Section 3). Using the host coordinates of their optical counterparts in our final compiled summary table is always recommended over their nominal host coordinates.

Finally, the metadata of other rare events, including FRBs and GWs, are generally unique and consistent across different reference sources, so we do not make a further selection.

2.2.3. Standardization of the Classification Scheme

To create a master sample of transients, one needs coherent and consistent classification for events observed in various wavelengths and timescales. This is a nontrivial task when working with diverse reference sources and archival records of historical events due to their minor discrepancies of classification schemes. The existing classification scheme for transients, particularly for SNe, is mostly phenomenological. Supernova types are primarily assigned based on spectral features and sometimes the shape of light curves. Meanwhile, the grouping of subtypes (taxonomy or *hierarchy* of supernova types) is usually motivated by physical models. Even most data sources follow this well-established classification scheme, the detailed type notations and *label systems* (i.e., the union of possible types labels) are not identical, especially when new phenomena are discovered or potentially new types are identified. From the viewpoint of a retrospective meta-analysis, there are a few issues to be resolved when unifying transient types in different reference sources into a single, consistent scheme.

First, the notations (or type labels) for transient classification could vary in different reference sources, especially for unusual transients that do not naturally fit the existing classification scheme or newly identified subtypes that lack a standard notation in the community. When the complexity of the supernova phenomena cannot be described in our existing scheme, the hybrid, intermediate, or transitional types are occasionally used in the literature to designate these indecisive cases. For example, SNe Ia with circumstellar interaction have been once classified as "Ia/IIn" or "Ian" by some observers for their strong narrow Balmer line emission, a definitive signature for Type IIn SNe, before they are recognized as a subgroup of SNe Ia (Ia-CSM or Ia-02ic; Silverman et al. 2013, 2016). Similarly, Type IIb SNe, whose hydrogen lines—for which Type II are assigned—fade at later stages until they eventually become SNe Ib-like. Before the popularization of "IIb" designation, these events occasionally receive "II/Ib" classification for such transitional behaviors. More commonly, several type notations are used interchangeably to refer to the same newly identified subgroup of events (e.g., Ia-06gz and Ia-09dc), which should be unified during data compilation.

Second, the label system could be different across reference sources, requiring a unified classification scheme. Some reference sources may use complex or fine-grained classification schemes, including more refined subtypes, while others may only limit to some major types. When combining type labels in multiple reference sources, as a result, there could be inconsistencies due to the classification schemes themselves. For example, the same supernova could be classified as "Ia-91bg" in some reference sources and as "Ia" in other reference sources with a more limited range of type labels. As a result, in the combined sample, some "Ia-91bg" events may have coexisting "Ia" labels assigned, but other "Ia-91bg" events may not, leading to inconsistent labeling of events. Ideally, we would use the best-refined subtype label of each event whenever possible so that transients are always uniquely and unambiguously classified to the finest details. However, in a combined sample of transients, the same event could have multiple, incompatible subtype labels, i.e., labels not along the same branch of the classification tree. This can result from disagreements among reference sources or independent detection in multiple wavelengths or by multiple messengers. Therefore, we allow the same transient event to have multiple labels. Meanwhile, we choose a hierarchical classification scheme so that variations of classification schemes can naturally fit in.

Our convectional, phenomenologically-based classification scheme for transients is not strictly connected to the driving physical mechanisms, including progenitors, exploding mechanisms, and circumstellar environments. When constructing a hierarchical classification scheme, instead of adhering to a purely

phenomenological scheme, we group transient types based on their physical origins whenever known. For example, SNe IIb are commonly considered as close siblings of hydrogen-poor SNe Ib and Ic (or *stripped-envelope SNe*), rather than a subtype of Type II supernova (e.g., Filippenko et al. 1993). Similarly, Ca-rich events, named after their strong nebular-phase calcium lines, have been previously classified as a subgroup of SNe Ib by their maximum-light spectra. Later studies revealed their thermonuclear origin—they closely resemble SNe Ia, rather than collapsing massive stars like SNe Ib (e.g., Perets et al. 2010). These subtypes are preferably grouped by their physical origins instead of spectral signatures. To attain consistency of classification when combining labels from reference sources with different classification schemes, we assign each event its best-refined subtype (there can be multiple) *and* all their *physical* parent labels so that each label in our transient sample points to a complete subset of events, including its physical subtypes.

Third, another source of inconsistency is the group of so-called peculiar events. SNe with extreme properties or unusual signatures are labeled as peculiar events in our conventional classification scheme (e.g., "Pec" in many catalogs). However, this all-inclusive label is not used consistently in different reference sources. For example, a few well-defined subtypes of SN Ia, such as Ia-91T and Ia-91bg, are sometimes (but not always) labeled as "Ia Pec." These robustly defined, relatively uncommon subtypes have characteristic observational signatures and are potentially related to the diverse channels, progenitors, or environments of SNe Ia. Many other peculiar SNe are just classified by template-matching algorithms, where the "Pec" label is inherited from the best-matching template, without other detailed and definitive type labels. Occasionally, individual papers classify events as peculiar without a new label for their uncommon properties. These unspecified peculiar SNe could just be outliers of ordinary events, but they may also contain new subtypes that are yet to be discovered. Given the context-specific meaning of "peculiar SNe," we use "Pec" as a flag for relatively uncommon subtypes *and* individual events with unusual properties instead of a subtype.

The standardization of the label system and subtype hierarchy must be done on a case-by-case basis. We discuss our detailed strategy in Appendix B. The hierarchy of transient type labels is listed in Table 2.

2.2.4. Remarks on Transient Classification

We discussed the standardization of the classification scheme in the previous subsection. Such a procedure ensures a unified label system and coherent subtype hierarchy within the database. However, transients in our combined sample are detected and classified by numerous survey programs and researchers, over decades, using a wide variety of data and techniques. Therefore, the results may suffer from possible ambiguity or uncertainty, potential methodological bias, and even human subjectivity.

First, transients are primarily classified by their photometric or spectroscopic signatures, while the quality, wavelength range, and temporal coverage of data are sometimes too insufficient to precisely and definitively subtype those events. For example, some supernova subtypes are characterized by their early-time or nebular-phase spectral features, whereas a high-quality spectrum is not always available when they are far before or after maximum. As a work-around for time-intensive

Table 2
Type Hierarchy and Number Counts of Transients

Type	Total	Host
Full Sample	36333	19268
Thermonuclear SNe (Ia)	15435	8093
Ia (unspecified)	14534	7558
Ia-91bg	135	91
Ia-91T	398	177
Ia-02cx	57	38
Ia-CSM	25	9
Ia-09dc	8	8
Ia-02es	6	5
Ia-99aa	6	3
Ia-00cx	4	3
Ca-rich	21	15
Core-collapse SNe (CC)	8437	5629
Stripped-envelope SNe (SE)	1619	1123
Type Ib-like (Ib)	757	546
Ib (unspecified)	363	270
Ibn	51	27
IIb	323	235
Type Ic (Ic)	710	481
Ic (unspecified)	672	452
Ic-BL	137	73
Ib/c	315	250
H-rich Core-collapse SNe (II)	6685	4605
II (unspecified)	5023	3514
II P	918	688
II L	55	34
IIn	681	373
Superluminous SNe (SLSN)	230	29
SLSN (unspecified)	13	1
SLSN-R	6	2
SLSN-I	154	20
SLSN-II	61	8
Tidal Disruption Events (TDE)	99	93
TDE (unspecified)	56	54
UVOptTDE	22	19
XrayTDE (+ likely events)	21	20
Gamma-ray Bursts (GRB)	1621	22
GRB (unspecified)	20	0
SGRB	115	1
LGRB	1486	21
Fast Radio Bursts (FRB)	14	14
Kilonova (optical only)	1	1
Gravitational Wave (GW)	1	1

Note. *Total* indicates the number of events with a certain type label, including events classified as subtypes. *Host* indicates the number events with known host galaxies (names or coordinates). Rows with "unspecified" in *Type* exclude events with subtype labels under the same parent type. Transient types here are not mutually exclusive. Some types may have common physical origins, and the same event may receive multiple types.

spectroscopic classification, efforts have been made to classify transients purely based on light curves of limited epochs and filters, either with or without redshifts (e.g., Frieman et al. 2008; Kessler et al. 2010; Lochner et al. 2016). The results of such classifications bear even larger uncertainties.

Second, the technique and criteria used could also bias the results. Several active transient follow-up programs rely on template-matching techniques to classify events, where a spectral library of various transient types and phases are cross-correlated with the observed spectra to determine the most-alike types (e.g., Smartt et al. 2015). Such classifications and dating of transients are implicitly affected by the coverage of types and phases in the template library and possibly even by the quality and representativeness of the individual spectra. The aforementioned photometric classification is another example of how technique and criteria may affect the results.

Third, a nonnegligible fraction of classifications are visually done by experienced observers in an empirical and perhaps subjective manner. The results may, therefore, vary from one person to another.

Indeed, the ambiguities, biases, and inconsistencies cannot be fully controlled or corrected without tracing back the original reference sources or conducting a comprehensive reanalysis of archival data. However, proper curation of existing classification does, at least, guarantee the consistency of type labels *within* our database.

2.3. Transient Sample Statistics

In this version of our database, the transient sample has 36,333 unique events, including 23,703 SNe, 1621 GRBs, 99 TDEs, 14 FRBs, and 1 GW event. The remaining ones are not explicitly classified yet. These are mostly candidate events with either a potential host or redshift reported. The numbers above do not sum up to the total number of classified events in the transient sample, as some events are detected in multiple wavelengths or messengers and thus are double-counted. Also, as we have applied a series of selection criteria, the number counts do not match their corresponding upstream data sources. Number statistics, by type labels, are summarized in Table 2 and Figure 1.

More than half of the events in our database are Type Ia SNe, where 4% of them are further classified as some subtype of SNe Ia or marked as peculiar SNe Ia. These 4% SNe Ia include well-defined subtypes, like Ia-91T, Ia-91bg, and Ia-02cx, and other less commonly detected subtypes, like Ia-09dc and peculiar events without clearly defined subtypes. CC SNe ("CC"), including Ib, Ic, II, and their subtypes, are the second most abundant major class of events in our sample, accounting for nearly one-third of our transient events. We further divide CC SNe into hydrogen-poor stripped-envelope SNe ("SE," including Ib, Ic, and IIb) and hydrogen-rich Type II SNe. SE SNe accounts for about 20% of CC SNe, which is slightly lower than the common estimate of volumetric SE-to-CC ratio (e.g., $36.5^{+5.5}_{-5.4}\%$ in Smith et al. 2011; $30.4^{+5.0}_{-4.9}\%$ in Shivvers et al. 2017) Most CC SNe are Type II, but only 24% of Type II SNe are further classified into subtypes by their light-curve shapes or spectral signatures. The remaining 5% of events in our transient sample are mostly GRBs, and the majority are LGRBs. Other rare events, such as TDEs and superluminous supernovae (SLSNe; "SLSN" under CC SNe), only account for about 1% of all events.

Within the transient sample, 28,995 events have redshifts reported in their upstream data sources. These are spectroscopic redshifts of either transient or their host galaxies. Very occasionally, host photometric redshifts are used. The redshift distributions of major transient types are quite distinct from each other (Figure 2, left). Type Ia SNe are distributed over the
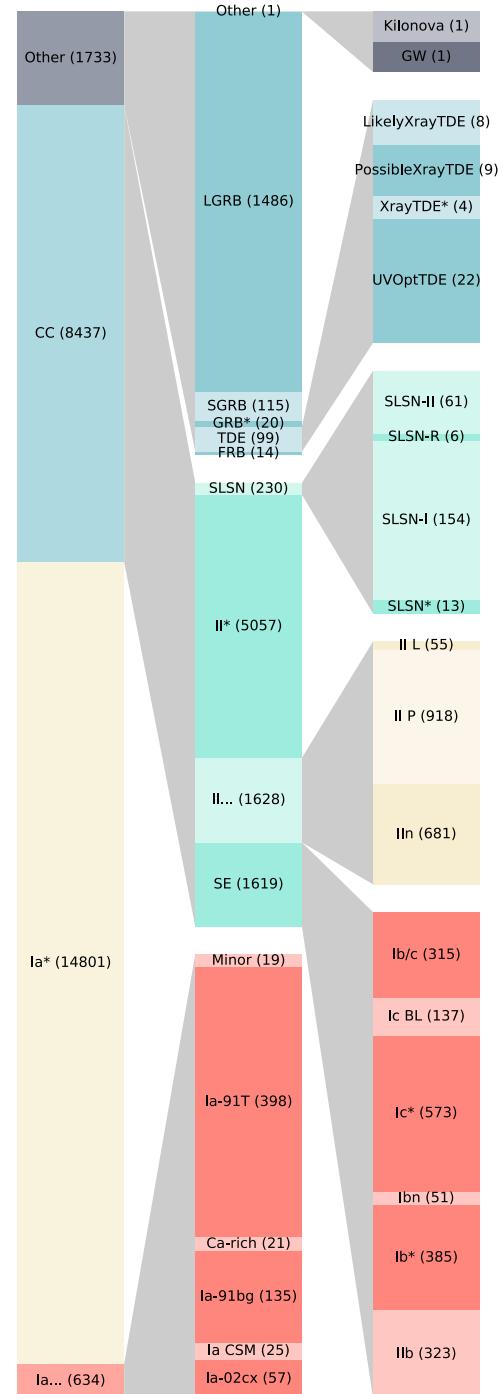

Figure 1. The number of transients under each type label, illustrated with stacked bars to show the relative sample sizes. Here, labels with ellipses (…) only count events *with* any subtype classification under this parent type, while labels with asterisks (*) count events *without* any further refined subtype label in the hierarchy (i.e., *unspecified* as defined in Table 2). Events with multiple incompatible labels (i.e., labels that are not subtypes of each other) are accounted for multiple times. Type Ia supernova is the most representative type in our sample, followed by CC SNe (Type Ib, Ic, and II). Gamma-ray bursts dominate the remaining 7% of classified events. Relatively rare events, like superluminous supernovae (SLSNe) and tidal disruption events, only account for about 1% of all records.

entire redshift range, with an extended high-redshift tail, likely due to their higher luminosities. Type II and stripped-envelope SNe are concentrated at lower redshifts, and particularly, SE SNe are at even lower redshifts than Type II SNe. Non-SN
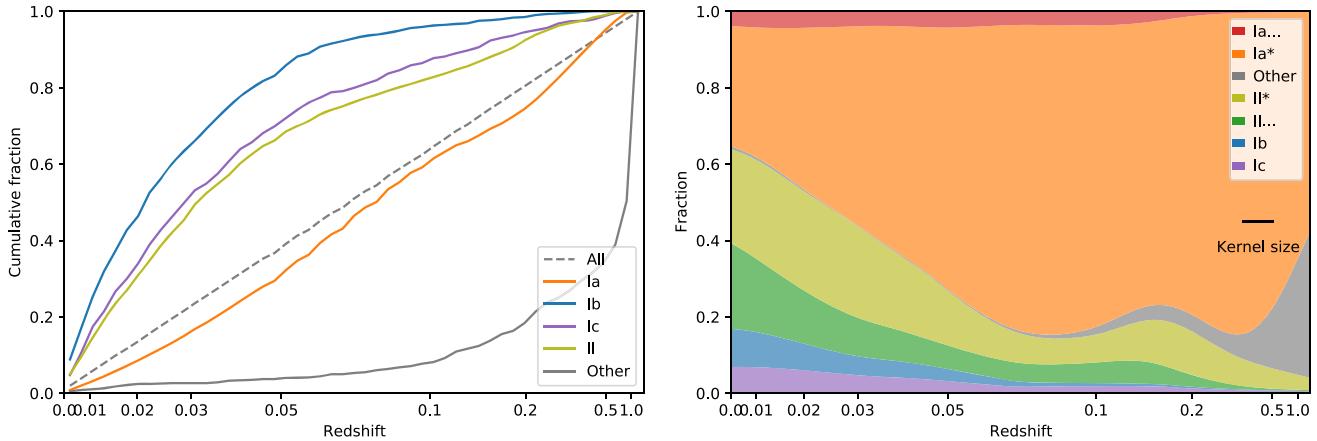
Figure 2. Left: the cumulative distribution of transient redshifts in our database. The redshift axis is stretched so that events are uniformly distributed along the axis. Major supernova types, including Type Ia, Type II, and stripped-envelope (Ib, Ic), are highlighted here. Compared to Type II and SE SNe, SNe Ia generally have higher redshifts in our sample, likely due to their higher luminosities. Meanwhile, SE SNe are distributed at even lower redshifts than Type II. Other transients, mainly GRBs with redshifts reported, are distributed at the higher redshift side of our sample. Right: the fraction of each major type across the redshift range, with similar, stretched redshift so that the shaded area scales with number counts. Labels with ellipses count events of any subtype under this parent type, while labels with asterisks only include events without any subtype labels. A large fraction of events at low redshift are CC SNe, and the relative fraction of stripped-envelope SNe (SN Ib and Ic) within CC SNe drops with increasing the redshift. The contribution of SNe Ia increases with redshift and peaks near $z \simeq 0.1$, while subtypes of SNe Ia are mostly at the lower redshifts.

events, primarily GRBs with redshifts available in our sample, also extend to very high redshifts compared to other transients. Regarding the fractions of major transient types (Figure 2, right), most events at the lower-redshift side ($z \lesssim 0.02$) are CC SNe, and the fraction of SE SNe within CC SNe drops toward higher redshifts. At medium redshifts ($z \gtrsim 0.02$), SNe Ia quickly dominate the local fraction of events, within which SNe Ia with detailed subtypes are mostly distributed at the lower-redshift side.

We further inspected the redshift and luminosity distance distributions of transients across subtypes as defined in Table 2, calculated using WMAP 9 yr cosmological parameters (Hinshaw et al. 2013). Luminosity distances are presented as distance moduli to be more closely related to the detectability of transients and can better illustrate the biases and heterogeneity in our database (Figure 3). Notably, SNe Ia in our sample have a double-peaked distance modulus distribution, where the first peak appears between $z \sim 0.05$ and 0.1, with a secondary peak between $z \sim 0.2$ and 0.5. Given the large fraction and extended redshift distribution of SNe Ia in our transient sample, the full-sample redshift distribution (in the distance modulus) is also double-peaked. Regarding subtypes of SNe Ia, we notice that subluminous Ia-91bg and Ia-02cx are more skewed toward lower redshifts than overluminous Ia-91T and Ia CSM. Ca-rich transients are distributed similarly to other low-luminosity SNe Ia. Within CC SNe, SE SNe (Ib, IIb, and Ic) have very similar, if not slightly lower, redshifts compared to Type II SNe, which dominate the total number of CC SNe. SLSNe ("SLSN") are clearly skewed toward higher redshifts, in which hydrogen-poor SLSNe (SLSN-I) are distributed farther than hydrogen-rich SLSNe (SLSN-II). Meanwhile, Type II L are at the lower-redshift end of Type II, although their peak magnitudes are not fainter than Type II P, which dominate the number count of photometrically classified Type II. For other subtypes of CC SNe, the differences are more nuanced. Type Ibn SNe are at the higher-redshift tail of SE SNe, likely due to their higher peak luminosities. Similarly, relatively brighter SNe Ic with broad emission lines (SNe Ic BL) are at slightly higher redshifts than the normal SNe Ic. Finally, rare events

like TDEs and FRBs are distributed at higher redshifts than CC SNe, and GRBs have the highest average redshift compared to other major types.

The double-peaked distance modulus distribution for SNe Ia could be a direct result of our sample heterogeneity. These transients are discovered by various surveys with significantly different sky coverage, cadence, and sensitivity. Consequently, the spatial-temporal coverage of our sample is far more complicated than a single survey. As illustrated in Figure 4, our events can be loosely grouped into low-redshift and high-redshift groups. The low-redshift group is mostly associated with wide-area, shallow surveys, with steadily improving depth over the years. Meanwhile, the high-redshift group is primarily contributed by a few deeper surveys that are more localized in their sky coverage. This is evident in their decl. angles on the sky, where low-redshift events have scattered and well-mixed decl. angles, while high-redshift events only have a few discrete ranges of values (i.e., their survey fields) that are related to their time of discovery.

Finally, the All-sky map of our transient events clearly shows the footprints of a few major deep and narrow-field supernova surveys (Figure 5), notably the SDSS Supernova Survey (Frieman et al. 2008), the Supernova Legacy Survey (Astier et al. 2006), the ESSENCE Supernova Survey (Miknaitis et al. 2007), the High-Z Supernova Search (Schmidt et al. 1998), and the Supernova Cosmology Project. These deep supernova surveys, however, only contribute a limited fraction of events in our transient sample. Wide-field transient surveys contribute the most records. Labeling transient records by individual survey programs could be interesting for certain studies, but this cannot be achieved easily due to the incompleteness of information in our upstream data sources.

The number statistics, local fraction, and redshift distribution indicate that our transient sample is highly heterogeneous. Multiple factors could contribute to such heterogeneity. First, there could be a classical Malmquist bias, where intrinsically brighter events can be observed at greater distances. A possible second-order effect is that slowly evolving events have a longer time window for detection and classification and are thus easier
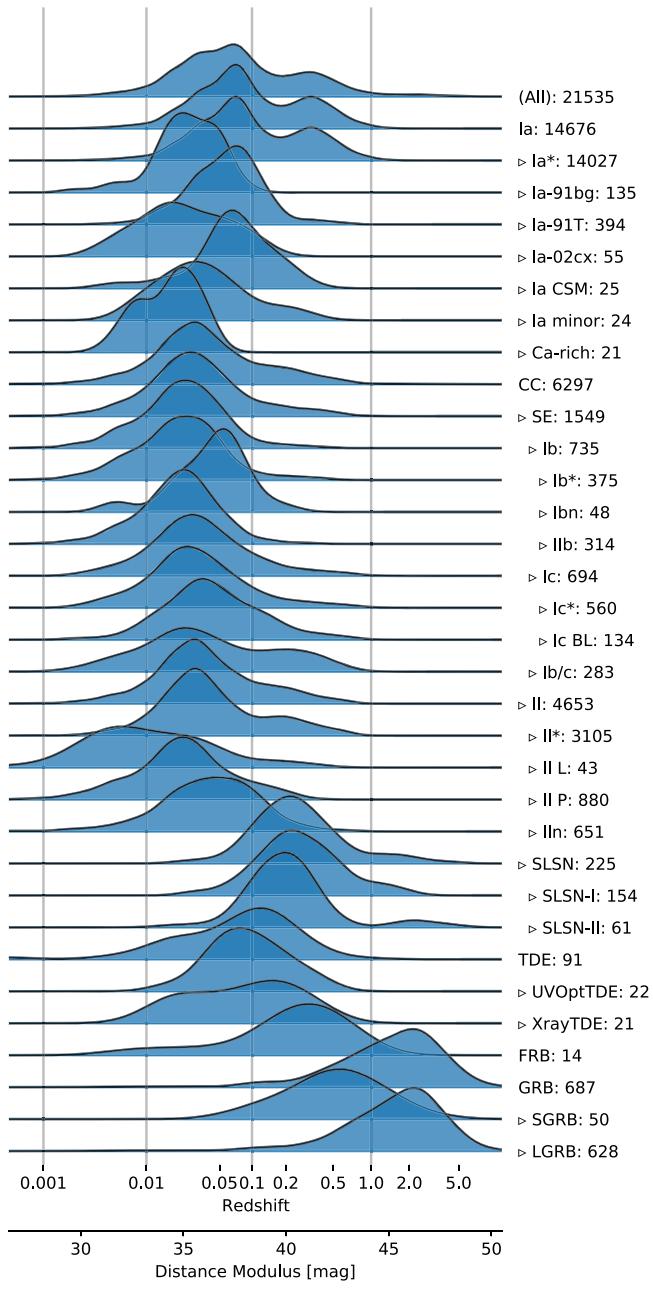
Figure 3. The distribution of redshift and luminosity distance for each type, shown as normalized density with respect to distance modulus. Labels with asterisks are events that are not further classified as any subtype of this parent type, i.e., "unspecified" events in Table 2. Luminous events, like SNe Ia and SLSNe, are generally populated at the higher-redshift side. Meanwhile, relative to their parent types, luminous subtypes (e.g., Ia-91T, Ia CSM, and Ibn) are usually at the higher-redshifts side, while faint subtypes (e.g., Ia-91bg, Ia-02cx, and Ca-rich) are usually at lower redshifts. However, the different distributions cannot be fully attributed to the classical Malmquist bias. Some types require detailed observations (e.g., densely sampled light curves, spectra) to be classified, implicitly biasing them to lower redshifts. Also, our transients are compiled from various surveys and spontaneous discoveries that do not share the same detection efficiencies.

to be identified than those that rise and fade quickly. Second, the criteria or techniques for classification may also introduce some bias, as many subtypes require high-quality spectra or even spectra at a certain phase of evolution to identify, which are often unavailable for distant, faint, or rapidly fading objects.
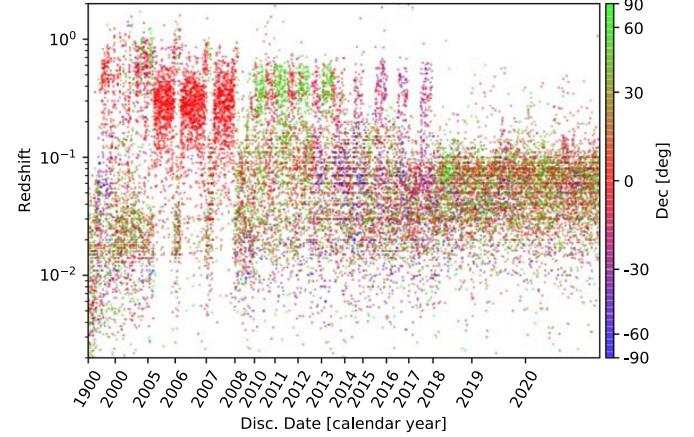
Figure 4. Year of discovery vs. redshift of our transient events, color-coded with the decl. angle. The time axis is stretched so that transients are uniformly distributed, and the vertical axis is in uniform scale of the cosmological distance modulus. The transient sample is a mixture of shallow, wide-field surveys and deep, high-redshift surveys with limited sky coverage. There exists a low-redshift, spatially scattered component peaked at $z \simeq 0.05$, with another higher redshift, spatially more localized component around $z \simeq 0.5$. Consequently, the redshift or luminosity distance distribution of our events is double-peaked, as illustrated in Figure 3.

Third, and perhaps most importantly, our master transient list is a compilation of various data sources, including spontaneous discoveries and transient surveys with vastly different detection efficiencies defined by their survey design (e.g., sky area, field-of-view, single-epoch sensitivity, and cadence). Therefore, the number count, relative fraction by redshift, and luminosity distance distribution do not represent their true cosmological event rates or that of a sensitivity-limited survey.

3. Host Cross-identification

To compile the properties of host galaxies, one needs to identify their counterparts in other data sets and then access their observed or derived properties. The procedure includes three steps: searching for potential counterparts of host galaxies in other data sets, matching potential counterparts across data sets to ensure their spatial association and correspondence relationship, and identifying the true counterparts of host galaxies.

Through this section, we use *sources* for records in other data sets, *objects* for actual galaxies and stars in the sky, *groups* for clumps of spatially associated sources in other data sets matched by our algorithm, and *fields* for patches of sky near transient coordinates containing the host and other objects.

We aim to cross-match sources in other data sets into groups correctly, then identify the counterparts of known hosts or rank the cross-matched groups to identify the best host candidates when necessary. Therefore, we refer to the entire procedure as *cross-identification*. We use *known hosts* for host galaxies cross-matched using existing host names or coordinates, and *newly identified hosts* (or simply *new hosts*) for host galaxies identified by our ranking algorithm. Many newly identified hosts are probably apparent to previous authors or observers, and some may have been identified elsewhere. However, these hosts are not reported to our upstream data sources. Our work provides independent host cross-identification for these events.

In the following subsections, we describe the detailed procedure of host cross-identification. The workflow is outlined in Figure 6.
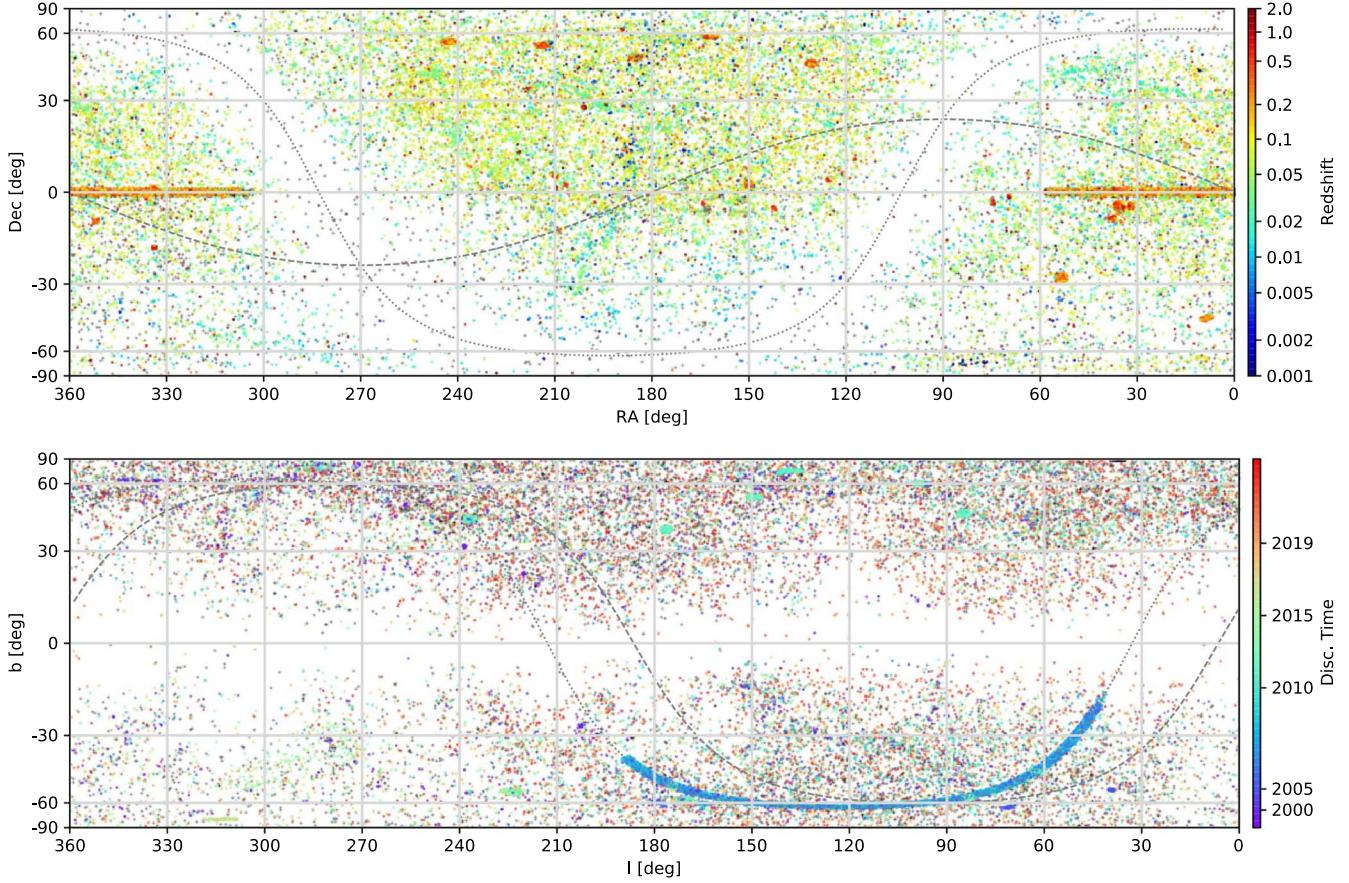
Figure 5. Top: equatorial coordinates of transients in Lambert cylindrical equal-area projection. Colors indicate their redshifts, and gray points are events without redshifts. The dashed and dotted lines trace ecliptic and Galactic planes, respectively. The fields of several deep supernova surveys are clearly visible, but they only contribute a limited fraction of events. Large-area surveys discover most events. Bottom: similar to the top panel, but showing Galactic coordinates instead. The discovery dates of events are color-coded with uniform number density over the color axis. Gray points indicate events without discovery dates. The dashed and dotted lines trace equatorial and ecliptic planes, respectively. The timespans of major deep supernova surveys, particularly the SDSS-II supernova survey near the equatorial plane, are clearly visible.
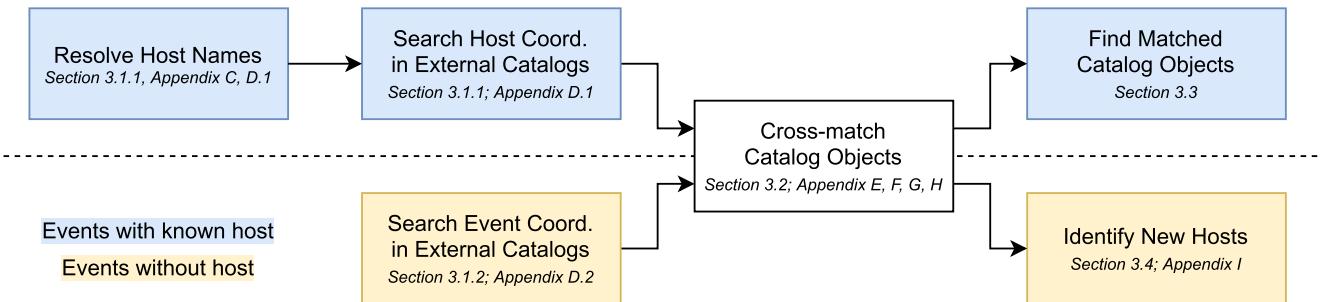
Figure 6. The general procedure to compile host properties for transient events. For events with known hosts, we resolve host names in NED and SIMBAD to obtain the best-available host coordinates and search the coordinates in external catalogs. The catalog sources that correspond to our known hosts are selected with a multi-catalog cross-matching process to avoid mismatching. For events without hosts reported, we search the transient coordinates in external catalogs with larger cone radii, cross-match all nearby catalog sources with a similar approach, and identify the most likely host candidates within cross-matched groups using a trained ranking function. The subsections are indicated on each block.

3.1. Accessing Sources in External Catalogs

We search for the counterparts of host galaxies in other data sets, including value-added catalogs (VACs) and survey catalogs, which we collectively refer to as *external catalogs*. VACs are online astronomical databases including the NASA/IPAC Extragalactic Database (NED), SIMBAD, and HyperLEDA. They collect the observed and derived properties of objects from multiple surveys and in the literature. More importantly, they are directly accessible using common name designations of astronomical objects. Survey catalogs are high-level data products of individual photometric or spectroscopic surveys, either hosted at web services or provided as static files containing directly measured properties of objects. Generally, without name-resolving services like Sesame,[36] the data of individual objects in survey catalogs are only accessible using their sky coordinates. External catalogs we use in this work are summarized in Tables 3 and 4. The number statistics and source selection criteria of these catalogs are discussed in Section 4.

---

[36] cds.u-strasbg.fr/cgi-bin/Sesame

Table 3
Photometric Catalogs

Catalog	Version	Filters	Hosted at	References
GALEX Medium-depth Imaging Survey (MIS)	GR5	FUV, NUV	Vizier	Bianchi et al. (2011)
GALEX All-sky Imaging Survey (AIS)	GR6/7	FUV, NUV	Local	Bianchi et al. (2017)
SDSS Primary Survey Objects (`photoPrimary`)	DR16	u, g, r, i, z, y	SkyServer	Ahumada et al. (2020)
Pan-STARRS 3π Survey, Stacked Object	DR2	g, r, i, z, y	MAST	Chambers et al. (2016)
DECaLS, Tractor Catalog	DR8	$g, r, z; W1–W4$	Datalab	Dey et al. (2019)
MzLS/BASS, Tractor Catalog	DR8	$g, r, z; W1–W4$	Datalab	Zou et al. (2017)
2MASS, Extended Source Catalog (XSC)	⋯	J, H, K_s	Vizier	Skrutskie et al. (2006)
2MASS, Point Source Catalog (PSC)	⋯	J, H, K_s	Vizier	Jarrett et al. (2000)
AllWISE Source Catalog	⋯	$W1–W4$	Vizier	Cutri et al. (2014)
unWISE Source Catalog	⋯	$W1, W2$	Datalab	Schlafly et al. (2019)
UKIDSS Large Area Survey (LAS)	DR9	Y, J, H, K	Vizier	Lawrence et al. (2007)
DES, Co-added Source Catalog	DR2	g, r, i, z, y	Datalab	Abbott et al. (2018)
SkyMapper, Main Table ("`master`")	DR2	u, v, g, r, i, z	Local	Wolf et al. (2018)
VHS Band-merged Multi-wave-band Catalog	DR4.1	Y, J, K_s	Vizier	McMahon et al. (2013)
SuperCOSMOS All-sky Galaxy Catalog	⋯	B, R, I	Local	Peacock et al. (2016)
NSA	1.0.1	FUV, NUV; u, g, r, i, z	Local	⋯

Table 4
Spectroscopic, Value-added, and Astrometric Catalogs

Catalog	Type	Version	Hosted At	References
SIMBAD Basic Data	VAC	⋯	simbad.u-strasbg.fr	⋯
NED Basic Data	VAC	⋯	ned.ipac.caltech.edu	⋯
HyperLEDA Astrophysical Parameters	VAC	⋯	Local	⋯
SDSS MPA-JHU Catalogs	spectroscopic	DR7	Local	Kauffmann et al. (2003); Brinchmann et al. (2004); Tremonti et al. (2004)
Gaia Sources ("`gaia_source`")	astrometric	DR2	Vizier	Gaia Collaboration et al. (2018)

When either the host name or coordinate is known, finding the counterparts of the host in external catalogs becomes a trivial problem. Even some survey catalogs are not directly accessible using host names alone; the corresponding sky coordinates of these names are still easily obtainable from VACs. However, host names or coordinates are only known for a fraction of events. About half of all events, even having redshift or classification, do not have a host name or coordinate reported. Subject to the availability of host information, we use two different approaches to search for potential counterparts of hosts in external catalogs. We outline the two approaches here. The details are discussed in Appendix D.

3.1.1. Events with Known Hosts

Nearly half of the transients in our sample have host galaxies reported. Their host galaxies are usually identified by experienced human observers, whereas the involvement of automated programs is increasing in recent years (e.g., Gupta et al. 2016; Sako et al. 2018; Gagliano et al. 2021). Often host galaxy names are reported, with or without the corresponding coordinates. Occasionally only host coordinates are provided. Whenever possible, we use host names to access their properties in VACs to ensure reliable matching. Meanwhile, many survey catalogs are only accessible with sky coordinates. We resolve the existing host names in VACs to find their corresponding host coordinates. Even though some events already have host coordinates reported along with host names, to avoid possible errors, we would like to update their as-reported host coordinates with the name-resolved coordinates in VACs.

To ensure that the host names and coordinates we use are reliable, we conducted a systematic quality check of their original reference sources. Only host names and coordinates from reliable reference sources are used in this work. The procedure of quality checking and the selection of reference sources are described in Appendix C. For events with known host names, we first resolve their names in NED and SIMBAD to find the best-available host coordinates. The procedure is described in Appendix D.1. Once succeeded, we use the name-resolved host coordinates to access other catalogs and find host properties. We use the existing, as-reported host coordinates when no host name can be resolved, but host coordinates are given to search external catalogs. When searching external catalogs, we use a fixed radius of 15″ for the name-resolved host coordinates and 30″ for the as-reported coordinates. Such large search radii minimize the chance of missing the true host due to ambiguous or inaccurate coordinates, a common situation for irregular, disturbed, or well-resolved large host galaxies. Even using known host coordinates, the returned sources do not always correspond to our known host galaxies. Other non-host objects may also populate the field given by our relatively large search radii. Therefore, we always perform a local cross-matching to ensure that the host properties we compiled are complete and uncontaminated (Section 3.2). Finally, we use transient coordinates to access nearby sources

for events without any host coordinate, either name-resolved or as-reported.

3.1.2. Events without Host Information

For events without host galaxies reported, we aim to identify their best host candidates and then compile their properties in external catalogs. We search transient coordinates in VACs and survey catalogs with flexible, per-case search radii to enclose catalog sources that match the true hosts. Transient coordinates may have significant angular offsets to their hosts. Therefore, the search radius for each event must be adjusted to optimize the chance to enclose the true host. We determine the search radius primarily based on the transient redshifts. For events with redshifts reported, we use the angular distance corresponding to a projected distance of 45 kpc at the reported redshift to search catalog sources. When there is no redshift reported, we use a default search radius of $30''$. Specifically, for GRBs without known redshifts, we use 3 times the 90% error radius as the search radius. Due to this conservative search radius, we may miss their true hosts, but this also reduces misidentified hosts. We set the lower and upper limits of the search radius to $15''$ and $2'$, but for GRBs, the lower and upper limits are $5''$ and $15''$. These catalog sources are then cross-matched into spatially associated groups (i.e., host candidates) with compiled properties (Section 3.2).

3.2. Cross-matching Objects

To compile the properties of host galaxies in external catalogs, we need to ensure that only robustly detected catalog sources that spatially coincide with our best-available host coordinates are used. Meanwhile, when identifying new host galaxies, we also need to select a group of catalog sources in the field with a confirmed spatial association and readily compiled properties, rather than treating catalog sources as individual objects and choosing the best host candidates from them. We need to spatially cross-match sources we obtained in external catalogs and establish their correspondent relationship for both purposes.

Cross-matching more than 20 catalogs with a wide latitude of angular resolution, sensitivity, and a mixture of photometric techniques is challenging. Several factors make the problem complicated. The sensitivities (or depths) of sky surveys are vastly different, where sources detected in one survey could be absent in other catalogs. Meanwhile, there could also be significant field-to-field variations of source densities and survey coverages. Some sources in certain catalogs could be artifacts or duplicates of existing sources in the same catalog that should be removed. Finally, even true and unique sources could be foreground stars rather than distant galaxies that could host our transients. Therefore, when cross-matching catalog sources, the criteria should be optimized for every catalog and tuned in each field. Artifacts and duplicates of the existing sources should be excluded to ensure that the cross-matched catalog sources are genuine and unique. Foreground stars should be flagged whenever possible to facilitate the identification of new host candidates.

Catalog sources are cross-matched once their coordinates coincide within a certain threshold. Source coordinates bear random errors depending on the signal-to-noise ratio of detection and the angular resolution of the survey (optical and pixel sampling), along with systematic errors inherited from survey-wide astrometric calibration. For resolved sources, such as our host

Table 5
Astrometric Tolerances for External Catalogs

Catalog	Tolerance	Catalog	Tolerance
NED	$0''\!.937$	2MASS XSC	$0''\!.968$
SIMBAD	$1''\!.230$	2MASS PSC	$0''\!.752$
HyperLEDA	$1''\!.227$	UKIDSS LAS	$0''\!.474$
GALEX MIS	$1''\!.641$	VHS	$0''\!.904$
GALEX AIS	$1''\!.824$	AllWISE	$1''\!.211$
SDSS	$0''\!.456$	unWISE	$1''\!.336$
PS1	$1''\!.421$	NSA	$1''\!.433$
DES	$0''\!.227$	SuperCOSMOS	$0''\!.813$
DESI-LS	$1''\!.196$	Gaia	$0''\!.010$
SkyMapper	$1''\!.077$	MPA-JHU	$0''\!.010$
VST-ATLAS	$0''\!.293$		

Note. The per-catalog astrometric tolerances are estimated by maximizing the overall connectivity score, as described in Appendix F. Gaia and MPA-JHU catalogs have fixed astrometric tolerances of $0''\!.01$. These catalogs are matched passively with other catalogs.

galaxies, the measured positions also depend on the photometric technique used and the de-blending of detected sources. Irregular, disturbed, or well-resolved galaxies are particularly affected by this issue. The criterion for spatial coincidence is, therefore, essential for any cross-matching algorithm. The conventional way to cross-match two catalogs relies on a constant angular distance threshold, which is heuristically tuned to maximize the number of reliably matched pairs while not causing mismatches. One may also use the position errors of sources when provided and match sources in a probabilistic way. Matching multiple catalogs, however, is a substantially more challenging problem. For example, Pineau et al. (2017) use χ^2 test in combinatorial analysis to cross-match multiple catalogs, while Salvato et al. (2018) establish the correspondence of sources with Bayesian statistics. These algorithms either rely on detailed source properties or require extensive calibration with large samples and are thus unfeasible given our large number of external catalogs. Due to data availability issues, we do not attempt to elaborate on the detailed source properties (astrometric, geometric, or photometric) when cross-matching. Also, the algorithm must be tuned and optimized using existing fields only, rather than a significant portion of the full catalog.

We cross-match these catalogs simultaneously in a single run using connected components in undirected graphs, where connectivity (i.e., adjacency matrix) is determined by pairwise angular distances, predefined per-catalog astrometric tolerances, and per-field matching thresholds. Here, the astrometric tolerance of each catalog (Table 5) is a predefined global constant reflecting the average positional error of sources in this catalog, while the matching threshold in each field represents an *acceptance threshold* of cross-matching that can be fine-tuned locally to compensate the field-to-field variation of source density and catalog coverage. The algorithm is a simple generalization of the two–catalog matching case, which is outlined below and described as the pseudocode in Appendix E:

1. *Selecting catalog sources.* We first select true and unique sources in each catalog in the vicinity of each queried coordinate (or "field"). The detailed criteria are discussed in Section 4. Real sources with potential quality issues (e.g., saturated in the image, close to a bright star, or field edge) are preserved. We leave such quality control of measured properties for those who use our data.

2. *Calculating pairwise distances.* We project selected catalog sources in the field onto a local two-dimensional Cartesian frame, in which sky coordinates are converted to directional angular offsets to the queried coordinate. Then we calculate the pairwise angular distances of catalog sources, excluding pairs within the same catalog that are always supposed to be disconnected.

3. *Finding connected source pairs.* The pairwise angular distances are further normalized with their corresponding pairwise astrometric tolerances. Here the pairwise astrometric tolerance for any two catalogs is defined as the square roots of their quadratically summed, per-catalog astrometric tolerances. Source pairs with normalized angular distances below the per-field matching threshold are considered as *connected*.

4. *Finding groups of connected sources.* We finally split sources into interconnected groups, in which any two sources are either directly connected or can be indirectly connected via a *chain* of connected sources. Each interconnected group of catalog sources is considered a cross-matched group. Isolated sources that are not connected to other sources are also considered as single-source *cross-matched* groups.

Analogous to a conventional probabilistic two-catalog cross-matching problem, assuming that per-catalog astrometric tolerances are isotropic positional errors of sources, the normalized distance is related to the Mahalanobis distance of the centroid position distributions for a source pair. Not surprisingly, the values of per-catalog astrometric tolerances and per-field matching thresholds are critical to the accuracy of cross-matching. We choose per-catalog astrometric tolerances using a global optimization method, and the per-field matching threshold is tuned with a similar approach, which is discussed in Appendix F. To further assess the quality of cross-matching, we introduced several metrics for cross-matched groups, which are discussed in Appendix G.

Before finding the known host galaxies or best host candidates and compiling their properties, some issues and complications remain with these cross-matched groups. First, a substantial fraction of catalog sources are actually Milky Way stars, rather than distant galaxies that could be transient hosts. These foreground stars may become interlopers during host identification and contaminate the compiled host properties if not handled properly. The situation worsens for the transients at lower Galactic latitudes. Therefore, foreground stars must be properly labeled and excluded when necessary. Another issue is that some catalogs may contribute more than one source in a single cross-matched group. Indeed, we do not directly connect sources contributed by the same catalog, but they can still be *passively* connected via other sources in the group. This leads to the *confusion* of multiple-catalog sources, where only one (or rarely, none) of these sources can reasonably represent the measured properties of the group. When such confusion occurs, we need to select the right source (or *representative* source) for the properties of the object measured by this survey. The procedure is described in Appendix H. It is worth noting that, even though we distinguish stars from galaxies here, we still allow possible stellar objects to be identified as host galaxies, given the inevitable confusion of star/galaxy separation in external catalogs (Section 3.4).

3.3. Finding Sources Matching Known Hosts

For events with known hosts, we only need to identify their counterparts in external catalogs or, more accurately, the cross-matched group near each transient that contains these catalog sources. We put the name-resolved or as-provided host coordinates into the field as a virtual source and cross-match it with other catalog sources using zero per-catalog astrometric tolerance. The cross-matched group that contains this virtual source is considered as the counterpart of the known host.

When using name-resolved host coordinates, the coordinate always coincides with (and thus matches) a NED or SIMBAD source with the same object name. We then mark the group as a "confirmed-by-name" host. When using as-provided host coordinates, and when the coordinate that matches a group has *not* been identified as a star, we consider the group a "confirmed-by-coordinate" host. If any cross-matched source in the group is believed to be a star rather than a galaxy, we only consider the group a primary host candidate instead of a confirmed host.

Finally, if we cannot identify a "confirmed-by-name" or "confirmed-by-coordinate" host using host coordinates, or if only transient coordinates are used for source searching, we proceed to rank cross-matched groups by their likelihood to be the true host using the procedure described in the following subsection. This includes the case in which the as-reported host coordinates coincide with a star or even match nothing in the search radius.

To ensure that these known host coordinates have successfully matched the indicated galaxies, we perform a comprehensive visual inspection of these fields (Appendix J.1). Occasionally, if the host coordinate failed to match the center of the indicated galaxy or the most prominent component of an irregular host, we manually reassign the host to the correct cross-matched group. We focus on the results of cross-matching. Generally, we do not judge the correctness of transient-host association in our upstream data sources. However, if there is *clearly* a better choice of a host galaxy than the indicated one, we mark the galaxy as an "alternative host" and flag the case *without* reassigning the host. To summarize, in 91 cases, the input host coordinate missed the indicated galaxy; in 257 cases, we noticed better hosts than the indicated ones. For various reasons, like image quality issues, we are unable to inspect 106 cases. The majority of known and cross-matched hosts (17,556) passed the inspection without any issues.

3.4. Finding New Host Candidates

For events without known *and* confirmed hosts, we identify their best host candidates among cross-matched groups. Host galaxies reported in previous surveys are often visually identified by experienced observers in discovery or archival images. Such a manual or semiautomatic workflow may remain effective if transients, mainly SNe, are classified and reported at the current rate; however, this would become infeasible for future high-efficiency follow-up programs or large transient samples, as we discussed here. Meanwhile, for state-of-the-art and future time-domain sky surveys, the growing involvement of galaxy properties in real-time alert processing and follow-up scheduling also requires efficient and reliable methods for automated host identification.

Finding host galaxies is usually straightforward for human observers using images. However, it may not be an easy problem for automated algorithms when only cataloged source properties are available. Transients may occur anywhere within and even far outside the optical radii of their host galaxies. Therefore, the right host is not always obvious, and confusion of multiple possible hosts is sometimes inevitable, even for human observers. Only one (or even none) could be the true host among those cross-matched groups in each field, assuming that sources are properly matched. We need an effective and robust method to *rank* these cross-matched groups by their possibility to be the host. We formulate the problem as follows: each cross-matched group could be characterized with a set of numeric or binary-valued parameters. We aim to construct a continuous scalar function of these parameters, whose value indicates the possibility that a cross-matched group can be the true host. The function need not be a probability estimate in nature, as we only use it as a numeric score to rank the groups and identify the best host candidate of each event. Also, the function should differentiate multiple possible hosts effectively to avoid confusion, i.e., being reasonably sensitive and *monotonic*. We noticed that the decision functions inside the conventional binary classifiers might satisfy these requirements if the classifier is trained to separate cross-matched groups into true hosts and other non-host objects. We, therefore, train the binary classifiers to distinguish *known* hosts from other nearby non-host objects and then use the trained decision function to rank cross-matched groups of events *without* known hosts.

There are several important aspects of this machine-learning-based host ranking method: the construction of the training data set, the parameterization of cross-matched groups, the evaluation of performance, and finally, the choice of classifiers. Here we outline the overall procedure to construct ranking functions. Relevant details are discussed in Appendix I.

The training set lays the groundwork of our ranking functions. Aiming for a clean training set, we choose the known and properly cross-matched hosts that passed our visual inspection without quality issues or alternative hosts (Section 3.3, Appendix J). We access external catalogs and cross-match sources again, supposing that neither host name nor coordinate is available so that the training set best resembles the actual situation of host ranking. Cross-matched groups near these transients are labeled as either "true hosts" or "other objects" based on our existing knowledge about their true hosts. We include both "true hosts" and "other objects" in the training set, where the latter enable the ranking function to reject non-host objects near transient coordinates. The training set, as a result, is much larger than the actual number of transients used for training.

There is a wide range of options for the input variables of ranking functions, such as transient-host offsets, results of cross-matching, or detailed source properties. We call a particular combination of input variables a *feature set*. As the starting point, we construct a basic feature set using only those universally available parameters of each group, regardless of which catalogs have been cross-matched. We also expand the basic feature set into its redshift-dependent version using a few transient redshift-relevant parameters. Some more detailed (but not universally available) properties in external catalogs, including optical-infrared magnitude, angular size, photometric redshift, and other derived parameters, may further improve the performance of ranking functions. Assuming that some

particular wide-field surveys have the required sensitivity to detect the true hosts, we further expand the basic feature set with similar source properties measured in these catalogs. However, these feature sets are only applicable in the coverage of these surveys, and the choice of host is also limited to groups with required parameters available.

Since the ranking functions are trained using events with known hosts, we can evaluate their performances by checking if a ranking function can recover those hosts. We define the *accuracy* of a ranking function as the chance that the group of a known host will rank the highest among all cross-matched groups in its field. In other words, the accuracy here is the fraction of known hosts that have been successfully recovered by the ranking function, which differs from the accuracy commonly used in statistics, machine learning, or other research disciplines. We choose this particular performance metric because the goal of training differs from the actual way the trained model was applied. The classifiers are trained to distinguish individual "true hosts" from "other objects," without the contextual information about other groups in the field; while the trained decision functions are used to rank (instead of to classify) all cross-matched groups in the same field and identify the best one. A result-oriented metric may better reflect the expected outcome of the training process.

The decision functions of trained binary classifiers are the actual ranking functions we use. To choose the best classifier for the purpose, we compare several conventional algorithms in the training and testing process, including Logistic Regression, Support Vector Machine (SVM; Cortes & Vapnik 1995), Random Forest (RF; Ho 1995; Breiman 2001), AdaBoost (Freund & Schapire 1997), Stochastic Gradient Descent (SGD), and Multilayer Perceptron (MLP). We test their performance under different input feature sets, including the basic feature set and its redshift-dependent version, as well as other feature sets that rely on certain detailed source properties in external catalogs. These classifiers all reach above 95% accuracy. Logistic Regression classifier, in particular, achieves $97.3\% \pm 0.3\%$ and $97.5\% \pm 0.6\%$ accuracy using the basic feature set and its redshift-dependent version. It also stably outperforms other classifiers when using those derived feature sets with detailed source properties. Therefore, we use Logistic Regression as the default classifier. When transient redshift is available, we use the redshift-dependent version of the basic feature set; otherwise, we use the default, redshift-independent version. When applicable, ranking scores from other classifier and feature set combinations are also provided, but we do not use them to rank candidates directly.

As a final note, the highest-ranking group in a field is referred to as the primary candidate, and other groups are collectively considered secondary candidates. However, if the primary candidate is already assigned in a field, other groups are all secondary candidates. We list the primary candidate first in the data set, followed by secondary candidates in descending order of their ranking scores. The structure of the database is described in Section 6.

3.5. Quality Control of Newly Identified Hosts

We use trained ranking functions to identify the best host candidates for events without reported hosts. Although performing reasonably well in cross-validation, the ranking functions have not been tested for performance for newly identified hosts. Meanwhile, quality control flags and metrics

are also desired for the individual events to facilitate the use of the database. We provide two sets of quality control flags and metrics based on the visual inspection and mock sample tests.

3.5.1. Visual Inspection

We perform a comprehensive visual inspection of new hosts identified by us. The inspection is aimed to (1) evaluate the accuracy of ranking functions on these new hosts; (2) understand the failure modes of ranking functions; (3) provide quality flags for the use of their host properties; and (4) make corrections to misidentified hosts when possible.

We rely on visual inspection because the true hosts remain unknown at this point. Other transient-host catalogs may serve as the *ground truth* here, but these catalogs may also have misidentified hosts, leading to inaccurate estimates of accuracy. Indeed, visual inspection is a *subjective* evaluation of quality and performance, which depends on the designed workflow, image quality, and our discernment; while being an independent test, it reveals true performance and failure modes that are otherwise hard to characterize.

We have previously inspected the cross-matching of known hosts (Section 3.3, Appendix J.1). For the sample of new hosts identified in this work, we focus on whether the default ranking functions have identified the most likely host we noticed in image cutouts. We make corrections when hosts are likely misidentified. Also, when there are multiple possible hosts, or when we cannot identify the most likely host in image cutouts, we mark these indecisive cases with appropriate flags. The procedure and criteria of the visual inspection are discussed in Appendix J.2.

As a summary, without considering cases in which we are unsure about the most likely hosts, the default ranking functions achieved 97.0% overall accuracy, comparable to the results from cross-validation. Taking those indecisive cases into consideration, the accuracy tops at 97.3% under the optimistic assumption that indecisive cases are always correctly identified; but this accuracy may also drop to 89.1% in an improbable pessimistic situation that the ranking function failed in all indecisive cases.

3.5.2. Confidence Scores

As a complementary method for quality control, we introduce a numeric metric for the reliability of new hosts, namely the confidence score. The confidence score characterizes the degree to which a host candidate stands out among nearby non-host objects. True hosts are, in fact, outliers among those more abundant non-host objects in the field. They should have significantly higher-ranking scores than nearby non-host objects, assuming properly trained ranking functions. If the ranking score of a candidate is comparable to the best scores that non-host objects can reach by chance, then the candidate is, to a certain extent, indistinguishable from non-host objects. This makes the candidate a less reliable one. We use randomized mock transient samples and metrics for outlier detection to carry out the comparison here.

We feed a randomized mock transient sample into the workflow to construct baseline distributions, i.e., the distribution for the best-ranking score that non-host objects can get by chance. With the corresponding cumulative distributions, we can map the ranking score of a new candidate into a percentage score, which we refer to as the confidence score. For a well-

behaved ranking function, reliable candidates should have confidence scores close to 100%, because their ranking scores are significantly higher than non-host field objects; transients with multiple possible hosts may have more than one group with high confidence scores; finally, if no candidate reached a high confidence score, then all its candidates are likely indistinguishable from non-host field objects, and the event itself is likely hostless. Indeed, the baseline distributions depend on the transient redshift and catalog coverage and should be constructed separately for each ranking function. We discuss the details of implementation in Appendix K.

Nevertheless, the confidence scores only serve as a reference. They are still based on the trained ranking functions themselves, which do *not* directly indicate the correctness of transient-host association. Misidentified hosts could have high confidence scores, while genuine hosts may also have low confidence scores. Cross-checking with the visual inspection results, or performing some basic quality assessment, is encouraged when using confidence scores.

3.6. Accuracy of Ranking Functions

We use trained ranking functions to identify host candidates for events without reported hosts. The default ranking functions achieve above 97% accuracy from the cross-validation using known hosts, and our visual inspection reveals a similar empirical accuracy in new hosts identified in this work. To further understand the behavior of ranking functions and to characterize the quality of the data set in detail, here we analyze the dependence of accuracy on some key transient or host parameters, including transient redshift, transient type, transient-host angular offset, and host optical-near-infrared (NIR) magnitude.

For the comparison of accuracy across types, we group transients into broad classes using the hierarchy in Table 2. Nonsupernova transients, mainly GRBs, are grouped into "Other." Some unclassified transients, which enter our sample because either the redshift or host galaxy is known, are grouped into "Unknown." For optical-NIR magnitudes of hosts, we use r-, i-, or z-band magnitudes from DESI Legacy Surveys (LS) and PS1. Outside their coverage, SDSS, VST-ATLAS, or SkyMapper, are also used. For continuous variables, we divide events into bins of equal sample size along the axis of interest to characterize the dependency of accuracy.

3.6.1. Accuracy from Cross-validation

We do not use trained classifiers to rank cross-matched groups when the host is already known. Instead, we trust the reported host and directly use the host coordinate to match the right group. However, as the training set, we can use these known hosts to test the accuracy and analyze the behavior of our ranking functions.

We estimate the accuracy with standard tenfold cross-validation. We divide the sample into 10 subsamples, where each subsample contains about the same number of fields. During each test, we use one subsample for testing and the rest for training. We repeat the test 10 times and use each subsample once in turn for testing. The results are averaged over 10 tests to represent the performance. Besides estimating the overall accuracy, we also derive the dependence of accuracy on transient redshift, transient type, transient-host angular offset, and host optical-NIR magnitude by binning the
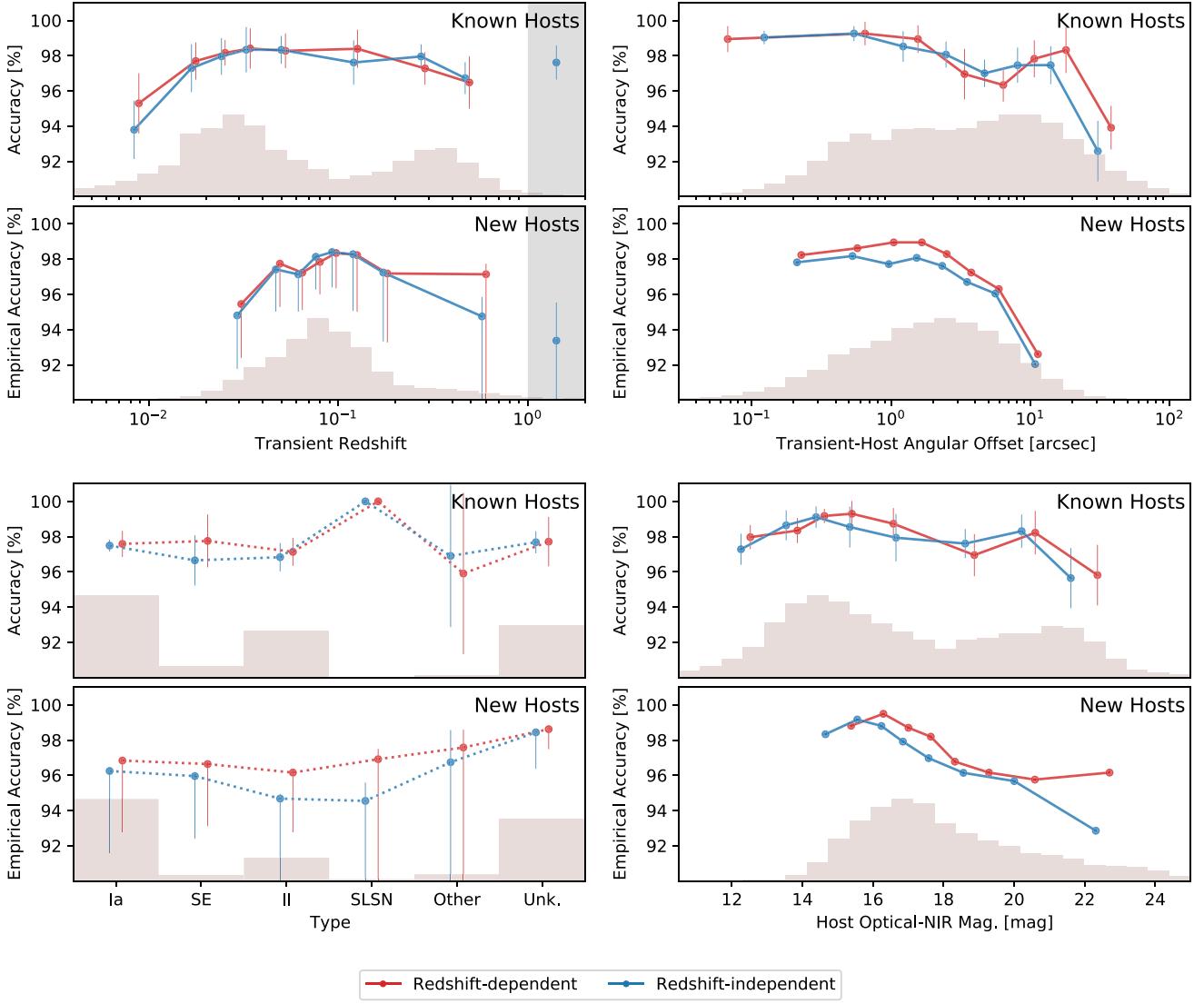
Figure 7. The accuracy of host cross-identification with our default ranking functions, using redshift-dependent variables (red) or not (blue). The four panels here show the dependence of accuracy on transient redshift, transient-host angular offset, transient type, and host optical-NIR magnitude. Except for the panel of transient type, we split the sample into bins of equal number counts along the axis of interest and plot the per-bin accuracy at the median value of the bin. We estimate the accuracy in two separate ways. With the sample of known hosts (i.e., the training set), we calculated the average accuracy using tenfold cross-validation, where the standard deviations are indicated as error bars. For the sample of new hosts, we estimated an *empirical* accuracy with a visual inspection, defined as the fraction of visually identified hosts that are also chosen by our algorithm. Vertical *error bars* outline the range of accuracy due to the uncertain cases (Appendix J.2). We plot the distribution of both samples as shaded areas in each subpanel. The densities are over the logarithmic axis for redshift and transient-host angular offset, so the histogram area corresponds to the sample size. For the dependence on redshift, we also plot the subset of points without redshift, in the gray shaded area of the first panel.

test sample, estimating the per-bin accuracy, and taking the average of 10 separate tests. For each test, to estimate the per-bin accuracy to the percent-level, each bin must contain at least a few hundred events, which limits the bin number along the axis.

We summarize the dependency of accuracy in the upper subpanels of Figure 7. We only show the accuracy curves for the two default ranking functions for clarity. Accuracy curves using other feature sets are summarized in Appendix I.3. The change of accuracy over the redshift range is at the level of a few percent, indicating the stable performance of the ranking functions. The accuracy is lower than average for low-redshift or high-redshift events. Also, at larger transient-host offsets or for fainter hosts, the accuracy is lower than average. The degradation of accuracy for low-redshift events is likely attributable to larger transient-host angular offsets. Toward

the higher redshifts, the properties of true hosts become less distinguishable from those ubiquitous faint galaxies, possibly leading to higher failure rates. Notably, even at the relatively small transient-host angular offsets, the accuracy is only at 98%. Failure happens here when a more distant object is chosen rather than the true host in close proximity. We do not see a significant variation of accuracy across transient types. The accuracy in "SLSN" and "Other" (mainly TDEs and GRBs) is noisy due to their limited sample size compared to the number of folds. Using transient redshift-relevant parameters marginally boosts the overall accuracy, and the improvement is clear at large transient-host offsets.

For the accuracy of other feature sets (Appendix I.3), we notice that the inclusion of some source properties can further improve the accuracy above our basic feature sets. Even only at the percent-level, the increase in accuracy indicates a

significant reduction of failure rates. However, this comes with the cost that some host galaxies may not have the required features and are thus ignored by these ranking functions.

3.6.2. Empirical Accuracy from Visual Inspection

We also estimate the accuracy of new hosts from the results of the visual inspection. Because the inspection relies on our discernment instead of some *ground truth*, we refer to this as *extitempirical accuracy*. In the subsample of events with unambiguous and clearly visible hosts, we use the fraction of successfully identified cases as the empirical accuracy. There are, however, many cases in which the most likely host cannot be determined visually, usually due to image quality issues, confusion of multiple nearby galaxies, and even the complete absence of any likely hosts. The uncertainty in accuracy due to these cases should be taken into consideration. Therefore, we further estimate the optimistic and pessimistic limits of empirical accuracy supposing that *all* these indecisive cases are either successfully identified or missed by the ranking functions (Section 3.5.1).

We focus on the dependence of the empirical accuracy on the key transient and host parameters, which is outlined in the lower subpanels in Figure 7. We find similar patterns of dependence as we have seen in known hosts. The accuracy does not change significantly along the axes. We find lower than average accuracy in low-redshift or high-redshift bins, larger transient-host angular offsets, and fainter hosts. Considering the loss of accuracy due to indecisive cases, even in the pessimistic situation, the accuracy of the low- to mid-redshift events remains around 95%. However, for high-redshift events, the large fraction of indecisive cases implies a likely lower accuracy than outlined here—the pessimistic lower limit of the last redshift bin is under 70%. For this reason, "SLSN" and "Other" (mainly GRBs), usually high-redshift events, have pessimistic lower limits around 80% and 50%, respectively. Using redshift as input may improve the results for high-redshift or fainter hosts among cases with identifiable hosts, although the improvement may not be significant considering the broad range of accuracy due to indecisive cases.

Finally, it shall be noted that, for indecisive cases, their redshift and type are intrinsic properties of the events, so that we can quote the optimistic and pessimistic limits for these two curves. However, the transient-host angular offset and host optical-NIR magnitude depend on the host and are unknown for indecisive cases. Therefore, we are unable to quote the optimistic and pessimistic limits for these two panels.

3.6.3. Typical Failure Modes

To understand the reason for which the ranking functions miss the true hosts, we rerun our redshift-independent default ranking function on the entire training set, i.e., the known and properly cross-matched hosts that passed our visual inspection, and then select cases in which the top-ranking groups are not the true hosts. On the other hand, from the visual inspection of new hosts, we also summarize cases where the ranking functions missed the most likely hosts as we noticed in the images. We analyze the failed cases on either side and find some common failure modes.

Figure 8 shows examples of failed cases, grouped into four common failure modes. First, in some fields, the true host could

be ambiguous, given the existing input parameters. Some of these cases can be resolved using more detailed source properties like shape parameters or galaxy redshifts, but some remain challenging even for human observers. Second, in some cases, the true host lies beyond the search radius, so another group inside the radius is mistakenly chosen as the host. These are usually low-redshift events with larger than typical transient-host angular offsets. Since we exclude such cases for the training set, the examples here are selected from new hosts. Failure like this can be reduced with larger search radii at the cost of higher computational and storage overloads. Third, some well-resolved nearby galaxies are split into multiple sources in certain survey catalogs. These parts or substructures of galaxies, when cross-matched into groups, become inter-lopers that confuse our ranking functions and reduce the accuracy (see also Appendix G). Fourth, the ranking scores can be undesirably boosted or penalized in certain situations, leading to the ignorance of a clear and unambiguous nearby host that other methods may not easily miss. Here the second and third failure modes are closely related to low-redshift events, which are less of a concern at higher redshifts. However, the other two failure modes are limits of the method itself and can only be resolved using better-tailored ranking functions.

We estimate the accuracy based on the assumption that the true host is always inside the search radius. In reality, there are occasional cases where the true host is beyond the radius, as shown in Figure 8. Most of the time, the true host is just outside the search radius. For the training set, we excluded 262 events due to the absence of the true host in the radius. This represents a small fraction compared to the resulted training set (17,421). For new hosts, at least 91 events are affected by this issue (Appendix J.2).

3.7. Notes on Undetected Hosts

Our training and testing process implicitly assumes that every transient should have a host. In reality, it is not uncommon to see transients without apparent host galaxies in discovery or archival images. We adhere to public, wide-field surveys to identify host galaxies, while many hosts, especially those of high-redshift transients (e.g., SLSNe, GRBs), may drop below the sensitivity limits of these wide-field surveys and are thus undetected. Intergalactic stars may also contribute to such hostless transients. Their "host galaxies" are certainly misidentified and should be excluded from any analysis.

We use visual inspection to flag possible hostless transients (Appendix J.2). The criteria we apply are rather conservative. We only rely on DESI-LS images or, occasionally, PS1 images without stacking artifacts, to identify them. We avoid labeling an event as hostless if its vicinity (angular separation $\lesssim 15''$) contains any possible faint galaxies. These faint and sometimes barely resolved galaxies are ubiquitous in survey images, and they are generally distant background galaxies per the photometric redshift catalog of Zhou et al. (2021). However, we cannot eliminate the possibility that some of them are actually real host galaxies at large projected distances compared to their sizes. We check image cutouts of the larger field of view for low-redshift cases to exclude possible hosts at extreme angular distances. Finally, if there is neither object in the image nor sources in survey catalogs, but the name of a nearby VAC source clearly indicates the host, we do not label the event as hostless. We assume that their hosts, if they exist,
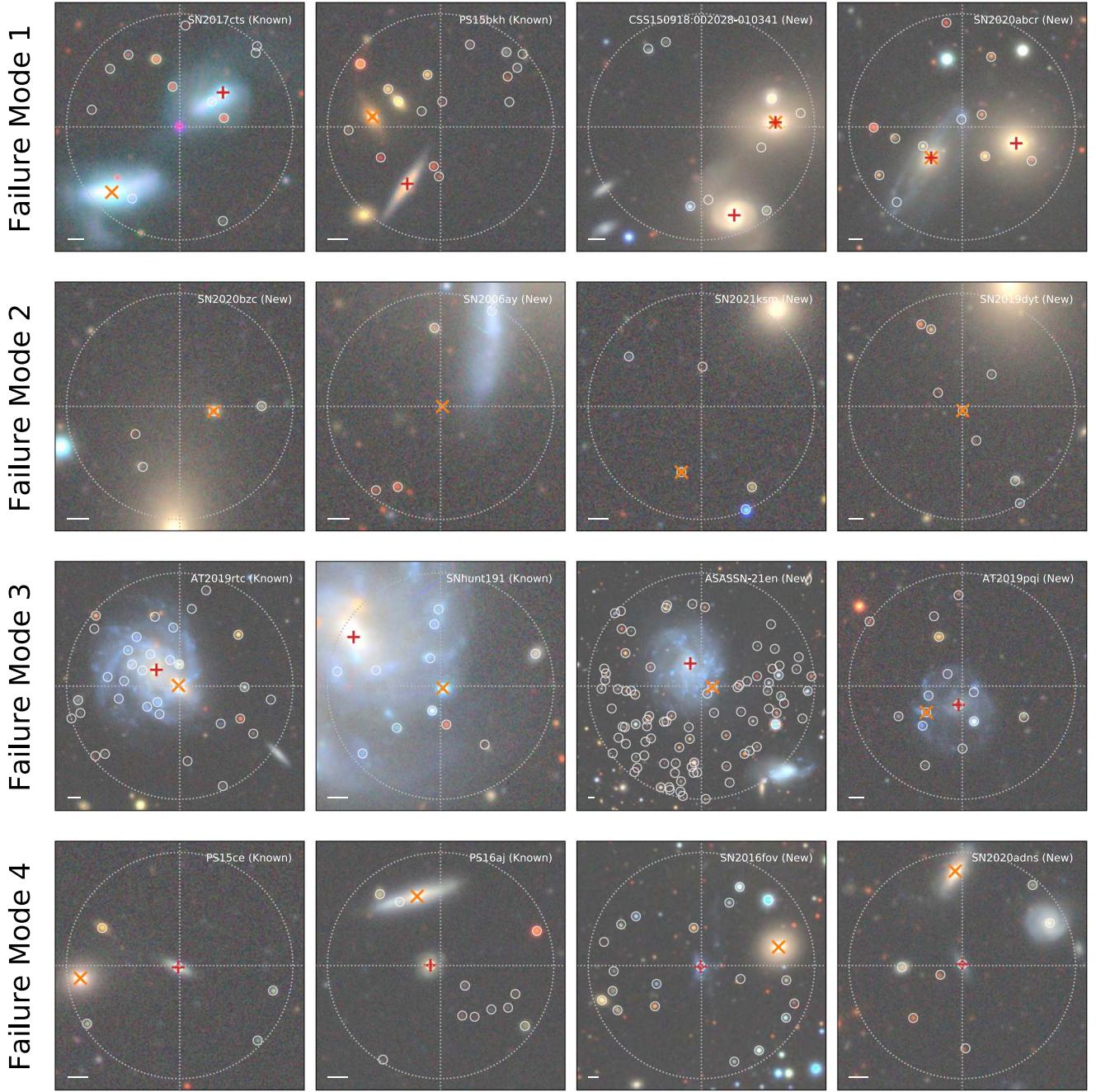
Figure 8. Examples of incorrectly identified hosts from cross-validation of known hosts and visual inspection of new hosts. Each panel shows a field centered at the transient event, with the event name and angular scale of $5''$ indicated at the corners. Inside the search radius of each event (dotted circle), the best host (either previously reported or visually identified) and algorithm-identified host are highlighted with red and orange crosses, and other cross-matched sources are marked with white circles. Each row here represents a common failure mode: (1) inevitable confusion of multiple likely host candidates in the field, in which the true host could be ambiguous even for visual identification; (2) the true host is located beyond the search radius due to an exceptionally large transient-host offset; (3) part of a well-resolved, nearby large galaxy is identified as the host, instead of its main component; and (4) another cross-matched object, either nearby or far in the background, is erroneously identified as the host at higher ranks. Events at lower redshifts are more susceptible to the second and third cases.

are low-mass galaxies with small sizes. If they are marginally detectable, they should be excluded by the criteria above. Our inspection leaves 408 likely hostless events at the sensitivity limits of our background images and external catalogs.

These 408 possible hostless transients include 176 GRBs, 158 SNe Ia, 30 SLSNe, 43 CC SNe that are not SLSNe, and a few more unclassified events. Since one event can have multiple types, the numbers above do not match the sample size. More than half of these possible hostless events have redshifts reported, with a median value of 0.4. GRBs, and

SLSNe dominate the high-redshift side of the population, and the low-redshift side ($z < 0.1$) consists exclusively of SNe (42 in total). Therefore, both low-redshift and high-redshift events may contribute to the population of hostless transients here.

However, the actual number of hostless transients could be underestimated here. Take the actually detected GRB hosts in Hjorth et al. (2012) for instance. At a limiting magnitude of $R \simeq 24$, typical of modern wide-field sky surveys, only 24% of their hosts are detectable; at a deeper limiting magnitude of $R \simeq 24.5$, more than half of their hosts remain undetected. Yet,

this does not include the upper limits reported in Hjorth et al. ([2012](#)). Our visually identified sample of hostless transients is certainly nonexhaustive.

Finally, the confidence score is an alternative, quantitative approach to identify hostless events. These events do not have hosts; as a result, their best-ranking candidates are nearby non-host objects, which are expected to follow the baseline distributions as we discussed in Section [3.5.2](#). We show the distribution of confidence scores for possible hostless transients in Appendix [K](#).

3.8. Notes on the Workflow

We discussed the cross-identification of host galaxies under two different circumstances: events with known hosts and events without hosts reported. The procedure to access external catalogs and cross-match catalog sources are similar, as outlined in Figure [6](#). In fact, we combine these two parallel workflows into a single pipeline. Accessing and compiling properties of known hosts is just a trivial case of the entire procedure, where the part of ranking host candidates can be skipped. The entire pipeline is summarized as pseudocode in Appendix [E](#).

In Figure [9](#), we show some typical host galaxies that went through our pipeline, including both known hosts and new hosts. These hosts are randomly selected near a few redshift values as examples. Even having more than 20 catalogs with nonnegligible catalog-to-catalog offsets of sky coordinates, our cross-matching algorithm works reasonably well for hosts across the entire redshift range.

Similar to the transient collection, we store the compiled properties of host galaxies in a separate data collection, indexed by the unique identifier of each event. For events with confirmed hosts, either by name or by coordinate, we store the properties of the cross-matched host and other cross-matched groups. When there is no confirmed host besides the properties of the primary candidate, we also store the properties of other cross-matched groups ("secondary candidates") in descending order by their ranking score in the field. Note that, when the as-reported host coordinates match a stellar object, the group is marked as a primary candidate instead of a "confirmed-by-coordinate" host (Section [3.3](#)). Otherwise, the group with the highest-ranking score becomes the default primary candidate.

We retrieve and compile host properties separately for each event, and consequently, there could be duplicates of hosts if several transients have been detected in the same galaxy. However, this is a relatively uncommon situation, as only 1886 known hosts in our database have more than one transient. Therefore, we do not index these cross-matched hosts uniquely to eliminate duplicates at this moment. Our statistics and analyses are also based on the idea of transient-host pairs rather than individual hosts.

3.9. Comparison with Similar Works

The technique to cross-identify host galaxies and the data products presented in this work is unique in many aspects. Here we compare our method and data set with similar works in the literature and discuss the advantages and drawbacks of the different approaches.[37]

---

[37] It has come to our attention that Aggarwal et al. ([2021](#)) have proposed a Bayesian method for host galaxy identification.

3.9.1. The Directional Light Radius Method

The Directional Light Radius (DLR) method (e.g., Sullivan et al. [2006](#); Gupta et al. [2016](#); Sako et al. [2018](#)), which relies on galaxy shapes and sizes to identify hosts, is probably the most widely used technique in previous works. The angular distances of potential host galaxies to a transient are normalized by their elliptical radii in the direction of the transient (i.e., the DLR). The resulted parameter (d_{DLR}) is a comparison of transient-galaxy angular distance and galaxy angular size, taking the axis ratio and position angle (inclination or projection effects) into consideration. Nearby galaxies with smaller d_{DLR} are more likely to be the host.[38] The DLR method implicitly assumes that transients in galaxies have a similar radial extent as the stellar light, which is reasonable for most stellar explosions. Calculating d_{DLR} requires no more than the basic shape and size parameters of galaxies, making this method easy to implement and thus popular for similar studies.

The method also has some clear drawbacks. First, shape and size parameters, such as moments of light, are method- and data-dependent. The resulted d_{DLR} are thus comparable only within the same survey. Such parameters are also noisy for low signal-to-noise detections or marginally resolved galaxies, leading to significant errors in the estimated d_{DLR}. Second, even assuming that transients following stellar light and galaxy isophotes are elliptical, the dramatic differences in radial light profiles (or concentration of light) remains a source of bias. For example, the effective radius (r_{50}) is commonly used to calculate d_{DLR}, but the fraction of enclosed light in units of r_{50} depends on the light profile, even for the simplest models. At four times the effective radius, the fraction of enclosed stellar light is 99.1% for an exponential disk, but for de Vaucouleurs' profile, the fraction is only 84.7%. The same d_{DLR} value, therefore, corresponds to a different fraction of enclosed light and hence the probability of being (or not being) the true host. Third, the comparison of d_{DLR} implicitly assumes a uniform transient rate and detection efficiency across galaxies, which may become an issue when there are multiple candidates with comparable d_{DLR}. Finally, for many photometric catalogs, there is a specific issue with nearby large galaxies. Most photometric pipelines are not optimized for galaxies of large angular sizes. Their resolved substructures, such as clumps, spiral arms, and even massive star clusters, can be broken into individual sources by profile-fitting or de-blending algorithms. Such substructures may impede host identifying, where the source with the smallest d_{DLR} could be a substructure instead of the galaxy core.

The DLR method is not immediately applicable to our entire transient sample due to the heterogeneity in transient types, redshift range, and catalog coverage. For example, shape and size parameters required by the DLR method are not always provided in external catalogs. Even when available, they are in diverse representations and are survey-dependent. However, for some specific catalogs, we implement the method. We test the accuracy when only d_{DLR} is used to rank host galaxies. We also include this d_{DLR} parameter in some feature sets and analyze the performance. Only using d_{DLR} to rank nearby groups, we can recover nearly 90% of known hosts. The implied 10% misassociation rate is higher than the estimate of

---

[38] In this work and in Gupta et al. ([2016](#)), d_{DLR} refers to the normalized transient-galaxy angular distance by DLR. In Gagliano et al. ([2021](#)), d_{DLR} refers to the DLR itself, and θ/d_{DLR} is the normalized angular distance (Equation (9)).
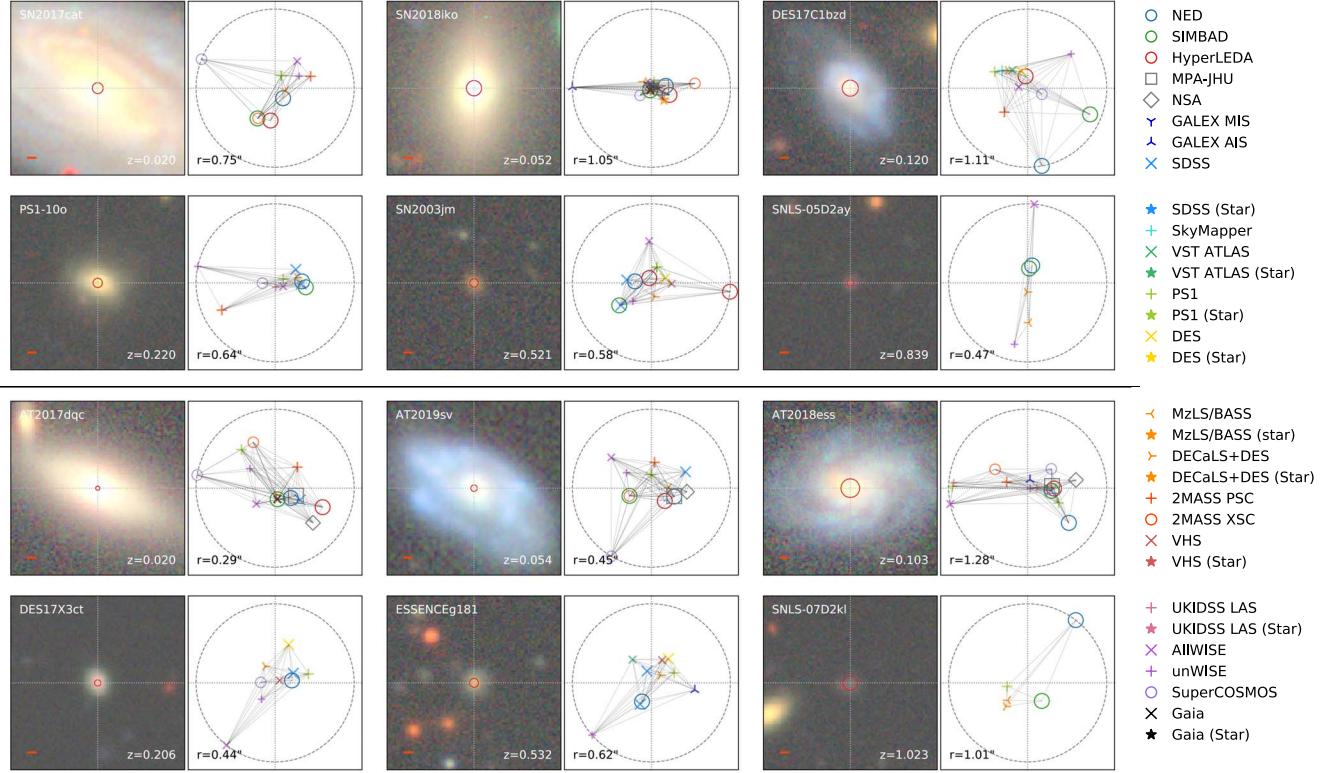
Figure 9. Examples of cross-matching catalog sources using astrometric tolerances and connectivity of objects. Each panel here shows the color composite image of a host galaxy on the left side, with transient name, redshift, and 1″ angular scale (red bar) indicated. The cross-matched catalog sources are plotted with various symbols on the right side, and a zoom-in view of the red circle (radius indicated) centered at the host. To cross-match objects, the distances of catalog sources are normalized by their pairwise astrometric tolerances and then compared to their thresholds of connectivity (Section 3.2). Connected pairs under this criteria are indicated as gray lines, and groups of interconnected catalog sources are further labeled as cross-matched objects. For clarity, we only zoom into the host here, while other catalog sources in the search radius are also cross-matched into groups in the same way. These panels are randomly selected near redshifts 0.02, 0.05, 0.1, 0.2, 0.5, and 1, where the upper six panels are known hosts, and the lower six panels are new hosts with our algorithm (Section 3.4). The cross-matching algorithm we developed is effective yet flexible for most hosts, while more complex cases are discussed in Appendix G.

5% in Gagliano et al. (2021) while close to the mock sample performance in Gupta et al. (2016). On the other hand, feature sets including d_{DLR} achieve higher accuracy than the default ranking functions. However, the improved performance may not be fully attributed to this single parameter.

3.9.2. The Gradient Ascent Method

The recent work of Gagliano et al. (2021) represents a more sophisticated genre of technique, where the gradient of surface brightness extracted from image cutouts, instead of cataloged source properties, is used to associate transients with potential hosts. They locate the peak of surface brightness (i.e., the core of the potential host) in the image cutout and use the cataloged source at the peak position for host properties. Finding the peak of light in an image is a nontrivial task in the presence of foreground stars and resolved galaxy substructures. They remove point sources, smooth out the fluctuations over the extended light component, and construct the gradient field for the light. After then, starting from the transient location, the algorithm ascends to the core of a nearby galaxy following this gradient field. This method works remarkably well for the low-redshift host galaxies with resolved substructures where the conventional DLR method could easily fail.

However, the technique requires direct analysis of image cutouts, potentially with higher computational costs. The quality and availability of image data could also be a constraint of its applicability. Also, the hyper-parameters in the entire workflow, such as the smoothing scale and the star/galaxy separation criteria, must be tuned for each survey. Finally, analogous to an optimization problem, the chance remains that the current solution (i.e., the peak of surface brightness) is only a *local minimum* instead of the global best solution, or the solution converges in the wrong direction.

3.9.3. The Ranking Function Method

In this work, we introduce a machine-learning-based host candidate ranking technique. We characterize potential hosts by various parameters, including basic parameters that are universally available for every cross-matched group and, optionally, detailed source properties in external catalogs. These parameters are then fed into a ranking function, which sorts these potential hosts by their estimated possibility to be the right host. The ranking function is trained using known, properly cross-matched, and visually inspected host galaxies in our data set, essentially turning the problem into a regular classification or regression problem.

Using a *trained* algorithm instead of a designed one based on certain assumptions is the primary difference of our approach compared to other methods. Those already identified hosts, including non-host objects in the same field, are used to train ranking functions and objectively evaluate their performances. A trained algorithm generally uses fewer assumptions and is thus less vulnerable to biases inherited from the imposed assumptions. This method is also extensible and flexible. Any

characteristics of potential hosts can be used as input, and many conventional classification and regression algorithms can act as the ranking function. Actually, this is a framework for a family of methods. However, the method relies on a well-constructed and sufficiently large training set, which may not be available for other similar problems. One may create a training set using mock transient-host pairs as done in Gupta et al. (2016), but the assumptions made when generating mock transient-host pairs could implicitly bias the results. We also assume that the cases to which the trained ranking function is applied are similar to those in the training data set. This is less of a concern since the new hosts have empirical accuracy comparable to the accuracy from cross-validation, but the transferability of the trained model should be tested in other similar applications.

We provide a matrix of trained ranking functions, but the default ranking functions we used deserve further discussion. To ensure the completeness of the input variables, the default ranking functions only take the basic characteristics of cross-matched groups that are independent of catalogs that have been matched. Detailed source properties in external catalogs, which proved to improve the accuracy, are not used. Therefore, the behavior of the default ranking functions might be partly driven by non-physical factors. We use a simple Logistic Regression classifier here. Such a simple linear model may suffer from underfitting, but the model is also insensitive to outliers in the training set and systematic variations in some variables. In other words, a simple linear algorithm guarantees the robustness of the trained model, but it may also lack the capability to differentiate multiple likely candidates.

3.9.4. Single-catalog versus Multiple-catalog Techniques

We choose hosts from cross-matched groups, where each group may contain sources from multiple external catalogs. This multi-catalog approach, driven by our motivation to provide rich host properties across multiple surveys, clearly differs from previous works. Our method and data set naturally provide better coverage of host properties over a wide range of wavelengths, robustly cross-matched across various catalogs, but the downsides are also clear. To begin with, cross-matching multiple catalogs, which itself is a challenging problem for galaxies with a wide range of angular sizes and redshifts, becomes an extra step. More importantly, due to the sensitivity limits and partial sky coverage of surveys, cross-identified hosts cannot have complete source properties across these 21 external catalogs. Consequently, there is a trade-off between the accuracy and applicability of the training functions if one plans to use detailed source properties as the input parameters. Adhering to those universally available basic features avoids the problem, but the performance could be driven by non-physical factors. Also, the compiled data set could be sparse, especially for high-redshift hosts or hosts beyond the coverage of major sky surveys. On the contrary, single-catalog techniques can fully utilize the existing source properties when finding host galaxies. Host properties compiled in this way are also consistently measured and are only subject to the selection function of one survey. However, the data provided are always limited without further cross-matching with other catalogs.

3.9.5. Comparison with the Gagliano et al. (2021) Data Set

Finally, besides the different methods to identify host galaxies as we discussed above, it is also worth comparing the publicly available data set presented here with the GHOST database maintained by Gagliano et al. (2021). First, our work includes more transient types, such as TDEs, GRBs, and other rare events, with fine-grained taxonomy, while Gagliano et al. (2021) focus mainly on spectroscopically confirmed SNe. Second, our transient sample includes 18,140 events with known and cross-matched hosts and 18,100 new hosts. The larger sample size compared to the GHOST database is mainly due to our different sample selection criteria and the inclusion of other transient types. Third, we choose to trust the reported hosts when available and only use the algorithm when there is no usable host information, while Gagliano et al. (2021) perform consistent catalog-based and image-based host matching for all their SNe. Fourth, the data set presented here includes cross-matched properties in 21 catalogs, while Gagliano et al. (2021) provide source properties mainly in the PS1 catalog, with limited data in other catalogs. We also include properties of secondary host candidates. Finally, besides the static version of the database in various formats, at the moment, we do not provide the visualization and data access tools as Gagliano et al. (2021) did. This could be a part of our future updates.

4. Host Data Compilation

The general guideline to choosing external catalogs is that the data should be available for a large fraction of our hosts, and in combination, these catalogs should cover the entire UV-optical-IR wavelength range. Following this principle, we select a wide variety of catalogs (Tables 3 and 4). Here, we discuss the basic characteristics, source selection, and star-galaxy separation criteria used for our external catalogs. We also briefly summarize the coverage of host properties in these catalogs. Note that the column names mentioned are as they appeared in our data sources.

4.1. GALEX

We choose the All-sky Imaging Survey (AIS) GR6/7 (Bianchi et al. 2011) and the Medium-depth Imaging Survey (MIS) GR5 (Bianchi et al. 2017) catalogs of the GALEX mission for UV photometry of host galaxies. They provide fixed circular apertures (8 and 17 pixels) and elliptical Kron magnitudes in the far-UV (FUV) and near-UV (NUV) bands, with related shape parameters for apertures and standard `SExtractor` flags in each band. The AIS has point-source-limiting magnitudes (5σ, AB) of 20.8 (NUV) and 19.9 (FUV), covering $2.6 \times 10^4 \deg^2$ (63% of the sky), while the MIS covers only 1000 $\deg^2$ but reaches deeper limiting magnitudes of 22.7 in NUV and 22.6 in FUV.

We do not impose source selection criteria for GALEX catalogs. Sources in both catalogs are always assumed to be valid objects. There are sources that are affected by dichroic mirror, detector window, or detector edge reflection artifacts. These sources, although they shall be excluded for photometric analysis as recommended by the references above, usually match true objects in deeper optical surveys. We, therefore, preserve these sources in our cross-matching process, and these potentially corrupted photometric measurements can be excluded later using corresponding artifact bits in `Nafl` and `Fafl` columns. Given the relatively low angular resolution of the GALEX survey, we do not use their star-galaxy separation parameters as given in the catalogs. Taking both AIS and MIS

into account, 9125 of our known hosts have GALEX magnitudes in at least one band (usually NUV), while 6791 have magnitudes in both GALEX bands. Meanwhile, 8161 new hosts have at least one band available, and 5993 have both bands available.

4.2. SDSS Photometric and Spectroscopic Catalogs

We also use the SDSS DR16 source catalog (Ahumada et al. 2020) for optical photometry of host galaxies. This data release covers $1.4 \times 10^4 \, \mathrm{deg}^2$ (34% the sky) in *ugriz* bands, reaching point-source sensitivity down to 22.2 mag in *g* and *r* (95% completeness, asinh magnitude). We accessed the photometric object table at the SDSS SkyServer.[39] The table provides complete point-spread function (PSF), profile-fitting (de Vaucouleurs, exponential and composite), fixed aperture, and Petrosian photometry of the detected objects. We exclude duplicate sources by selecting primary survey objects (mode=1). Stars are identified using their pipeline morphological classification (class) when cross-matching objects. This morphological indicator is based on the comparison of model-fitting magnitudes with PSF-fitting magnitudes. However, we only perform star-galaxy separation for sources brighter than $i = 20$ (asinh magnitude). Fainter than this magnitude, the purity of stars begins to drop.[40]

We also search for spectroscopic targets in the MPA-JHU catalogs (Kauffmann et al. 2003; Brinchmann et al. 2004; Tremonti et al. 2004). These catalogs are based on the spectra of SDSS DR8 (Aihara et al. 2011), which provides spectral properties (spectral indices, line equivalent widths, and fluxes) and derived physical parameters (star formation rates, stellar masses, and gas-phase metallicities) based on SDSS spectroscopy. We select objects that are classified as galaxies (SPECTROTYPE='GALAXY') with reliable measurements (RELIABLE=1). Spectra are also deduplicated using the list provided by the authors.

Within our transient-host pairs, 12,018 known hosts and 10,998 new hosts have SDSS photometry. Given that SDSS uses asinh magnitudes (Lupton et al. 1999), reported magnitudes close to the detection limit may have a very low signal-to-noise ratio. Requiring a minimal S/N of 5 in *g*, *r*, and *i* bands, 9845 known hosts and 9711 new hosts have reliable SDSS magnitudes. Further including the two shallower bands (*z*, *u*), the number count drops to 6575 and 6783 for known hosts and new hosts. Meanwhile, 3811 known hosts and 3709 new hosts have reliable spectroscopic measurements in the MPA-JHU catalogs.

4.3. DESI Legacy Imaging Survey

The DESI-LS (Dey et al. 2019) includes three photometric surveys that are complementary in the sky coverage and filter set: the Mayall z-band Legacy Survey (MzLS), which maps the high galactic altitude region in the northern hemisphere above decl. $= 30°$ in the *z* band; the Beijing-Arizona Sky Survey (BASS), which has a similar footprint above decl. $= 30°$ in the *g* and *r* bands; and the Dark Energy Camera Legacy Survey (DECaLS), which covers the equatorial region. They jointly cover $1.4 \times 10^4 \, \mathrm{deg}^2$ in the *g*, *r*, and *z* bands.

---

[39] skyserver.sdss.org/dr16/en/home.aspx
[40] classic.sdss.org/dr7/products/general/stargalsep.html

While conducted using different facilities, the reduced images of these surveys are processed using the same photometric pipeline (tractor; Lang et al. 2016), which creates catalogs by measuring source properties jointly over multiple images. This allows the incorporation of data from other surveys. Besides the existing unWISE stacked images (see Section 4.6), the latest LS DR8 also processed DES DR1 images, further extending the total sky coverage to $1.94 \times 10^4 \, \mathrm{deg}^2$ (*grz*; *W*1, *W*2), reaching decl. $= -60°$ in the southern Galactic cap region.

The tractor catalogs provide profile-fitting measurements of sources, where we take the geometric parameters related to de Vaucouleurs and exponential components, the fraction of de Vaucouleurs components, and the pipeline-chosen types and magnitudes. We search for primary objects within these three surveys (brick_primary=1), so that duplicates can be excluded. When cross-matching sources, stellar sources are identified using the morphological classification flag (type), where moderately bright sources ($g < 22$) with "PSF" type are considered as stellar sources. At $g = 22$ mag, about half of the extended sources are classified as "REX," a special type for *possibly* extended sources that cannot be robustly classified due to their low signal-to-noise ratios. There are also about an equal number of pointlike ("PSF") sources and "REX" sources near this magnitude. We conclude that the signal-to-noise of detection fainter than this limit is insufficient to make definitive source morphology classification, so we do not use source morphology classification beyond this limit. Besides source characteristics in tractor catalogs, we further include the photometric redshift estimated in Zhou et al. (2021).

Within our database, 14,576 known hosts and 13,958 new hosts have photometric measurements from these three surveys.

4.4. Pan-STARRS

We also retrieve photometric measurements of our host galaxies and candidates in the Pan-STARRS DR2 (PS1 DR2) stacked object table, which mainly includes the 3π Steradian Survey (Chambers et al. 2016). The 3π Steradian Survey imaged the entire northern sky above decl. $= -30°$ in *grizy*, reaching multi-epoch stacked limiting magnitudes (5σ, AB) of 21.4 (*y*) to 23.3 (*g*). The catalog provides band-wise Kron and PSF magnitudes with related geometric parameters. Given its wide coverage and high sensitivity, PS1 DR2 photometry is vastly available and relatively complete for our known hosts and new hosts.

We select primary detections (primaryDetection=1) of objects within StackObjectThin table, with valid *g*-band detection (gKronMag > -999). When cross-matching objects, bright objects that have a significant difference in point-source and Kron magnitudes ($i_{\mathrm{PSF}} < 20$ mag, $i_{\mathrm{PSF}} - i_{\mathrm{Kron}} > 0.05$ mag) are considered as pointlike sources. There are 14,446 known hosts and 13,797 new hosts that match PS1 DR2 sources.

4.5. 2MASS Point and Extended Source Catalogs

We use the 2MASS Point Source Catalog (PSC; Skrutskie et al. 2006) and the Extended Source Catalog (XSC; Jarrett et al. 2000) as our primary data source for NIR photometry. They both cover the entire sky in *J*, *H*, and K_s bands. The PSC measured fixed aperture magnitudes of over 5×10^8 presumably point sources, reaching limiting magnitudes (10σ, Vega) of $J = 15.8$ and $K_s = 14.3$. The XSC contains 1.6×10^6

sources, measured with isophotal, Kron, and extrapolated *total radius* apertures, reaching limiting magnitudes (10σ, Vega) of $J \simeq 14.7$ to $K_s \simeq 13.1$, depending on the surface brightness profile. We accessed both catalogs as a few percent for objects in the PSC are actually unresolved extragalactic sources (Rahman et al. 2016), and the PSC also has better depth.

We neither imposed selection criteria when searching for objects nor did we use point-source or extended source classifications to identify stellar and galaxies when cross-matching objects. Within our hosts, 10,707 known hosts and 8117 new hosts have matched PSC sources. Meanwhile, 9419 known hosts and 5529 new hosts have matched XSC sources. Most XSC objects also have detection in PSC, while 1368 known hosts and 2613 new hosts only cross-matched with PSC objects. However, as we have discussed, PSC is optimized for point-source photometry, and well-resolved galaxies might be split into multiple point sources in PSC. Therefore, we always use XSC photometry when available.

4.6. Mid-IR Photometry Based on WISE Data

Several different catalogs based on the images of the WISE survey (Wright et al. 2010) contribute to mid-infrared host magnitudes. WISE surveyed the entire sky in four mid-infrared passbands ($W1$–$W4$; 3.4, 4.6, 12, and 22 μm). Primarily we use the AllWISE Source Catalog (Cutri et al. 2014), which measures source magnitudes using co-added images of the original WISE mission and the extended, post-cryogenic *NEOWISE* phase (Mainzer et al. 2011). The sensitivity is higher at shorter wavelengths with more episodes of stacked imaging, reaching limiting magnitudes (95% completeness, point source, Vega) of 17.1 in $W1$ and 15.7 in $W2$. In $W3$ and $W4$ bands, the limiting magnitudes (95% completeness, point source, Vega) are 11.5 and 7.7, respectively.

We also included measurements in a few other WISE-based catalogs. Notably, Lang (2014) produced resolution-optimized co-adds of WISE images ("unWISE"), allowing more sources to be detected than the original AllWISE catalog. These co-added images are already used for the model-fitting photometry in DESI-LS (Dey et al. 2019; see Section 4.3). Beyond that, we also use the All-sky unWISE Catalog (Schlafly et al. 2019) as a complement to the official AllWISE catalog and `tractor`-measured mid-IR magnitudes in LS DR8. Primary sources (`primary>0`) are selected to avoid duplicates. Note that both AllWISE and unWISE catalogs measure objects as pointlike sources, and consequently, the magnitudes of extended sources would be inaccurate compared to model-fitting magnitudes. Well-resolved nearby galaxies could also be split into multiple sources in AllWISE and unWISE catalogs.

Of our host galaxies, 15,061 known hosts and 13,060 new hosts have measurements in the AllWISE catalog. Meanwhile 16,444 known hosts and 15208 new hosts matched objects in the unWISE catalog. Finally, DESI-LS catalogs perform forced photometry over unWISE images, and magnitudes are thus available for all optically detected objects, although the signal-to-noise ratio of detection may vary.

4.7. DES

For hosts in the southern hemisphere, the DES DR2 catalog is our primary source for deep optical photometry. The survey imaged 5000 $\deg^2$ of the southern Galactic cap region in $grizY$ bands, reaching co-added, point-source-limiting magnitudes

(10σ, AB) of $g = 24.33$ and $r = 24.08$ in its first major data release (Abbott et al. 2018). The catalog provides band-wise fixed circular aperture magnitudes and elliptical Kron magnitudes, with aperture geometric parameters, other `SExtractor` flags, and star-galaxy separation parameters in each band. We note that the catalog of DESI-LS already contains source properties in the footprint of DES. However, the *official* DES catalog contains full five-band data with source properties measured using an alternative photometry method. We, therefore, include DES in our database.

We searched objects in the co-added source catalog of DES DR2. Stars are identified using the criteria in Equation (4) (Sevilla-Noarbe et al. 2018), with an additional magnitude limit of $i < 21.5$ for better purity of stars. Within our transient-host pairs, 5549 known hosts and 3031 new hosts have DES sources matched. All of them have complete photometry in all five DES bands.

4.8. UKIDSS LAS

The Large Area Survey (LAS) of UKIDSS (Lawrence et al. 2007) is a near-infrared ($YJHK$) survey covering two continuous regions in the northern Galactic cap region and the southern side of SDSS Stripe 82. The 4000 $\deg^2$ survey area has been intensively imaged by several major optical surveys, making it an important data source for NIR photometry. Sources in the catalog have fixed aperture and Petrosian magnitudes (with aperture parameters) reported in each band, with a designed photometric depth (5σ point source, Vega) of 18.2 mag in K and 20.3 mag in J. Some sources have two separate epochs of J-band measurements.

We search primary objects (m = 1) in the main source table, where artifacts (cl = 0 or pn > 0.1) are rejected. Sources with high probability of being stars ($p^* > 0.95$) are marked as stars when cross-matching with other catalogs. We have valid UKIDSS LAS photometry in at least one band for 5036 known hosts and 2886 new hosts. Meanwhile, 3620 known hosts and 2258 new hosts have complete photometric detection in all four bands.

4.9. VHS

The VHS DR4 catalog (McMahon et al. 2013) is our primary source for deep NIR photometry in the southern hemisphere. This hemispheric survey imaged most of the southern sky in $YJHK_s$ bands, with comparable sensitivity to the UKIDSS LAS, reaching point-source limiting magnitudes (5σ Vega) of 20.2 in J and 18.1 in Ks. Fixed aperture magnitudes of three diameters ($2''$, $2.''8$, $5.''7$) and Petrosian magnitudes are reported for each band.

We searched for primary objects (PriOrSec=1 or priOrSec=FrameId) in the source table, where artifacts (Mclass=0 or pNoise > 0.1) are excluded. When cross-matching with other catalogs, objects with high probabilities of being stars ($p^* > 0.95$) are marked as stellar objects. Within our hosts, 5664 known hosts and 2318 new hosts have valid photometry in VHS.

4.10. VST-ATLAS

The survey footprint of the VST-ATLAS consists of two separate fields in the southern hemisphere that substantially overlap with DECaLS and DES, covering a total area of 4700 $\deg^2$ in $ugriz$ bands. The sensitivities are comparable to

the SDSS, reaching 23.2 mag in g and 22.6 mag in r (5σ median depth, AB). Sources have fixed aperture magnitudes in three diameters ($2''$, $2.''8$, $5.''7$,) and Petrosian magnitudes reported in each band. We use the VST-ATLAS as a complementary data source for optical photometry, especially in the u band that is not imaged by DECaLS or DES.

When accessing the catalog, we select primary objects and exclude duplicates of known sources (`PriOrSec=1` or `priOrSec=FrameId`). Possible noise or artifacts (`Mclass=0` or `pNoise>0.1`) are also excluded. When cross-matching with other catalogs, objects that are very likely to be stars by probability ($p^* > 0.95$) are considered as stars. Within our hosts, 1616 known hosts and 1436 new hosts have valid photometry in the VST-ATLAS.

4.11. SkyMapper

The SkyMapper Southern Survey (Keller et al. 2007) aims to image the entire southern hemisphere in $uvgriz$ bands with multi-epoch sensitivity similar to SDSS. The latest publicly available data release (DR2; Onken et al. 2019) we used here contains images of the snapshot-style shallow survey and the deeper main survey. The local sensitivity ranges from about 18 mag to 22 mag in g and r (10σ, point source, AB), depending on the completeness of the survey. We accessed the main table in which PSF and Petrosian magnitudes of sources are provided in each band, along with source geometric properties. Though not as deep as other optical catalogs here, its hemispheric coverage and u-band filter remains a great advantage over other surveys.

We searched for sources in the master table without a selection cut. Given that the angular resolution of SkyMapper is slightly lower than other optical surveys we use, we also do not use their star-galaxy separation parameters when cross-matching with other catalogs. In our catalog, 6896 known hosts and 4218 new hosts have valid photometric magnitudes in the SkyMapper catalog.

4.12. SuperCOSMOS

We include the All-sky galaxy catalog (Peacock et al. 2016) of SuperCOSMOS surveys (Hambly et al. 2001) as an auxiliary data source for optical photometry. The catalog is based on the digitized photographic plates with the UK Schmidt Camera, Palomar Schmidt Telescope, and ESO Schmidt Camera, providing calibrated B-, R-, and I-band photometry down to limiting magnitudes (AB) of $B_J \lesssim 21$, $R_F \lesssim 19.5$ and $I_N \lesssim 18.5$. Unlike other optical catalogs we use here, the magnitudes in this catalog are measured in the traditional Johnson–Cousins system, but a significant advantage is its complete coverage of the entire sky and its good sensitivity in the B band.

We obtained the static version of this catalog at the SuperCOSMOS Science Archive.[41] No specific selection cut was made when searching for objects. For our hosts, 12,307 known hosts and 12,038 new hosts have matched photometry in the SuperCOSMOS All-sky galaxy catalog.

4.13. NSA

The NASA-Sloan Atlas (NSA) is a catalog of local galaxies ($z \lesssim 0.15$) based on the SDSS and GALEX data sets. With improved image reduction, Blanton et al. (2011) and source extraction optimized for nearby large galaxies, and the catalog measured the detailed properties of about 0.6 million galaxies. Notably, Petrosian photometry with elliptical aperture, best-fitting Sérsic profile parameters with variable index, derived asymmetry and concentration indices, and K-corrected source characteristics are provided for objects, besides their classical SDSS-style circular Petrosian photometry and azimuthally averaged radial light profile.

We do not make specific selection cuts when searching for objects in this catalog, and all objects are considered to be galaxies. Within our database, 5638 known hosts and 4044 new hosts matched objects in this catalog.

4.14. HyperLEDA

HyperLEDA (Makarov et al. 2014) is another value-added catalog that compiles properties of galaxies across various surveys and reference sources. While HyperLEDA is not used in resolving host names, we accessed their table of homogenized astrophysical parameters to further enrich our available host properties. These properties include, but are not limited to, morphological T-type and detailed signatures (bars, rings, etc.), geometric properties, stellar and gas-phase kinematics, H I 21 cm and FIR fluxes, Mg II spectral indices, etc.

We only select objects that are classified as galaxies in HyperLEDA. Multiplicities of galaxies, like pairs, triplets, and groups, are not used as their sky coordinates could be quite uncertain. We have 12,728 known hosts and 8465 new hosts matched with HyperLEDA objects, but the availability of galaxy properties varies significantly from one object to another.

4.15. Gaia

We also searched for objects in the Gaia DR2 catalog (Gaia Collaboration et al. 2016, 2018). Gaia is a specialized astrometric mission for Galactic stars, which we do not use for host properties here. Instead, we use the astrometric properties of cross-matched Gaia sources to identify foreground stars against distant host candidates. Most galaxies, even in the magnitude limit of the Gaia instrument, are absent in the catalog, while the bright and compact galaxy nuclei are sometimes detected and cataloged.

Here we identify stars with the astrometric excess noise parameter (ϵ). This parameter, measured in the unit of milliarcseconds, characterizes the goodness-of-fit with their astrometric model, where values closer to zero indicate astrometrically well-behaved sources. Stars usually have lower ϵ compared to galaxies at similar magnitudes. This parameter has been used in previous works to separate stars and galaxies, while the detailed criteria vary. DESI-LS Pipeline labels sources above $G = 19$ mag with $\epsilon < \sqrt{10}$ as point sources.[42] This is independent of their source morphological classification. Considering the degradation of astrometric fits at fainter magnitudes, Koposov et al. (2017) select stars with a magnitude-dependent cut (Equation (1)) of $\log_{10}\epsilon < 0.15(G - 15) + 0.25$. Though known to be effective, the purity and accuracy of similar methods are yet to be tested and compared.
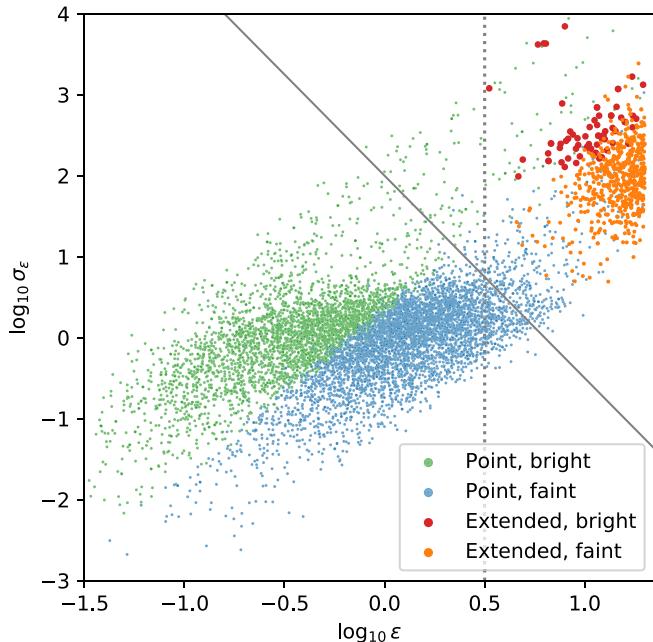
Figure 10. Star-galaxy separation using Gaia astrometric excess noise (ϵ) and its statistical significance (σ_ϵ), illustrated with a cross-matched example of Gaia and MzLS/BASS. Sources are classified into "point" (PSF) or "extended" (EXP, DEV, COMP, REX) by DESI-LS pipeline morphological types, and further separated into "bright" or "faint" groups by Gaia G-band magnitude of 19. DESI-LS use $G < 19$ and $\log_{10} \epsilon < 0.5$ (vertical dashed line) as their criterion for astrometric point sources, which is independent of their morphological classification. To improve the effectiveness at fainter magnitudes while ensuring the purity of point sources, we use a magnitude-independent selection cut (solid line) that aligns with the gap that separates pointlike and extended sources in this plane.

We use the following empirical criterion for stellar sources:

$$\log_{10} \epsilon < -0.4 \log_{10} \sigma_\epsilon + 0.8.$$

Rather than doing a simple magnitude-dependent cut on ϵ, we use the significance level of astrometric excess noise (σ_ϵ) in our criterion to optimize the efficiency at fainter magnitudes. As illustrated in Figure 10, there is a clear gap that separates pointlike and extended sources in $\log_{10} \epsilon$–$\log_{10} \sigma_\epsilon$ plane, which extends to fainter magnitudes. To ensure that our selection cut yields a complete sample of extended sources, or equivalently, a pure sample of point sources, we cut near the cloud of point sources. Note that objects with nonnegligible ϵ are not guaranteed to be galaxies, as unresolved binaries may also have positive ϵ. This is not against our goal to keep a complete sample of galaxies.

4.16. Extinction Corrections

Host properties like photometric magnitudes and intrinsic colors need the correction of foreground Galactic extinction to be fairly interpreted and compared. Accurate correction of Milky Way extinction requires high-resolution reddening maps, extinction curves, filter transmission profiles, and source spectral characteristics, which are beyond the scope of this paper. Without introducing extra complexities, we provide empirical extinction corrections of host magnitudes in each band.

The foreground reddening at the positions of host galaxies is estimated using the map of Schlegel et al. (1998), as

implemented in mwdust[43] (Bovy et al. 2016). Extinction correction in GALEX, SDSS, 2MASS, and WISE (*W1*, *W2*) bands are estimated using the empirical coefficients in Yuan et al. (2013). We also calculate another set of corrections for GALEX bands using Peek & Schiminovich (2013), given the vast range of estimated extinction coefficients for GALEX bands in the literature. These extinction corrections are provided separately in the database and are not directly applied to the measured magnitudes in each catalog.

5. Statistics of Transient-host Pairs

In this section, we summarize the basic statistics of our transient-host pairs, including both the confirmed host galaxies and primary host candidates. Secondary host candidates in each field are not included in the statistics of host properties here.

5.1. Angular and Physical Distances

The transient-host angular distance is expected to be comparable to the optical sizes of galaxies, except for nuclear events such as TDEs, and outlying events like Ca-rich gap transients (Kasliwal et al. 2012; Lunnan et al. 2017). Therefore, without considering the size evolution of galaxies, the angular transient-host distance should scale with the cosmological angular size-redshift relation. The distances of our transient-host pairs agree with this general trend (Figure 11). Toward higher redshifts, the angular distance decreases following the angular diameter distance until $z \sim 2$, where the angular scale peaks. Converted to the proper distance at their redshifts, the deprojected physical distance is similar across the entire redshift range, with very little or no redshift evolution. The average physical distance of transient-host pairs is 6.50 kpc, with twenty-fifth, fiftieth, and seventy-fifth percentiles of 1.66 kpc, 4.01 kpc, and 8.29 kpc, respectively. This also justifies our redshift-dependent search radius of 30 kpc when searching for objects in external catalogs for events without any host information reported.

As a side note, at intermediate to high redshift, an angular distance of 1″ projects to a few kiloparsecs, as indicated by the dashed line in the right panel of Figure 11. The deprojected physical distance here may have larger errors than for lower-redshift events. Also, our deprojected distance does not consider the ellipticity of host galaxies. A more comprehensive analysis of transient-host distances will be performed in our future work.

5.2. Redshift Coverage of Host Properties

For transients with host properties available in a certain external catalog, we inspect their redshift distributions (Figure 12). We convert redshifts to cosmological luminosity distance moduli for better illustration and comparison with host photometric magnitudes. The aforementioned double-peaked transient redshift distribution remains clear in the known hosts (top row, Figure 12), but the newly identified hosts do not have such a double-peaked distribution. The double-peaked transient redshift distribution is also imprinted onto other external catalogs.

For external catalogs, besides being shaped by the underlying transient sample, their redshift distributions also reflect their relative sensitivity or depth. External catalogs with better

---

[43] github.com/jobovy/mwdust
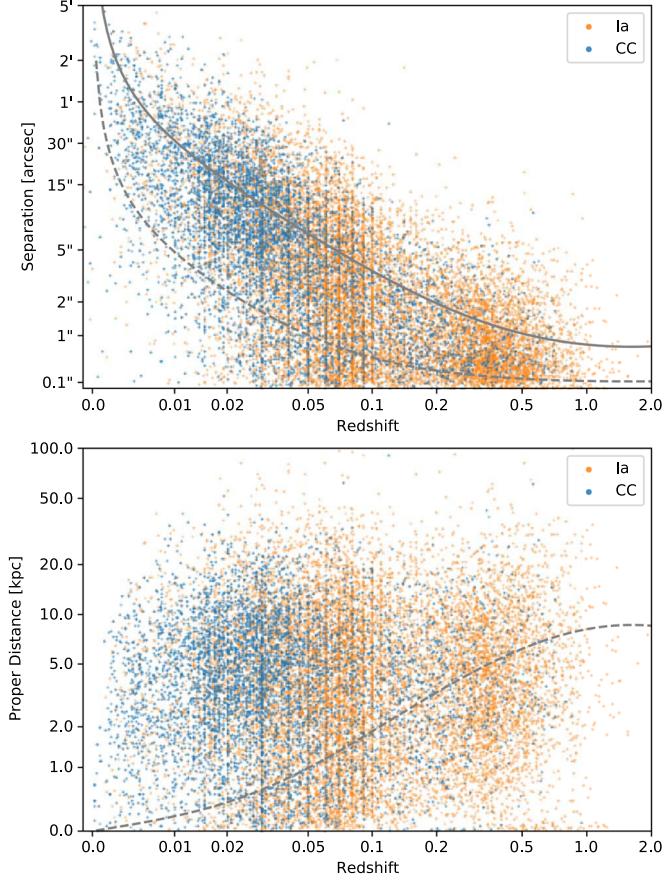
Figure 11. Top: the angular transient-host distances in our database, including both known host galaxies and newly identified hosts. Thermonuclear (Ia) and core-collapse (CC) supernovae are plotted separately with orange and blue dots. Angular distances matching 1 kpc and 6.5 kpc (average deprojected physical distance) are outlined in dashed and solid gray lines, respectively. Bottom: the deprojected physical distance of transient-host pairs (in kiloparsecs), with similar symbols as in the top panel. Physical distance matching $1''$ angular scale (typical position accuracy of galaxy coordinates) is indicated in gray dashed line. The angular distance mainly follows the cosmological angular diameter distance, where most events happen within 1 and 10 kpc. Converted to the physical distances, we find little or no redshift dependence of transient-host distance, with an average value of 6.5 kpc.

Figure 12. Redshift and luminosity distance distributions of transients with host properties available in each external catalog, shown as normalized densities with respect to distance modulus. The top two rows show the distribution for events with confirmed known hosts and newly identified candidates, while other rows show subsets of all hosts (confirmed and newly identified) with data in the external catalog available, as indicated at the left side. Here the distributions mainly reflect the depth of external catalogs, where deeper surveys like GALEX MIS and DES have more higher-redshift events than those shallow surveys (e.g., SkyMapper). The double-peaked redshift distribution of transients is clear in the known hosts.

sensitivity, such as GALEX MIS and DES, tend to have a higher fraction of high-redshift events. Conversely, external catalogs with relatively lower sensitivity, like 2MASS, GALEX AIS, and SkyMapper, have more events at lower redshifts. As a result of such vastly different catalog-wise redshift distributions, it could be nontrivial to find a proper set of external catalogs that provides adequate coverage in both redshift and photometric wavelength.

5.3. Coverage of Host Properties by Types

Here we summarize the fraction of hosts (both known and newly identified) that have valid catalog objects cross-matched in an external catalog, under each transient type label. Results are listed in Figure 13. In the first three rows, we also show the fraction of transients with redshift, with confirmed hosts or newly identified primary host candidates. Note that only types with at least five events are listed here. The coverage of host properties here also reflects the sky coverage and depth (or sensitivity) of these external catalogs. Full-sky and hemispheric

surveys, including WISE, GALEX, SkyMapper, PS1, and SuperCOSMOS, have better coverage of host properties in general. Meanwhile, some transient types have slightly better coverage of host properties than others, such as Ia-91bg and Ia-02cx. This could be a result of their redshift distributions, as events at lower redshifts may have more complete properties compiled in external catalogs. Secondary factors, like systematic differences of their host luminosities or SEDs, may also have a nonnegligible effect. Such nonuniform coverage of host
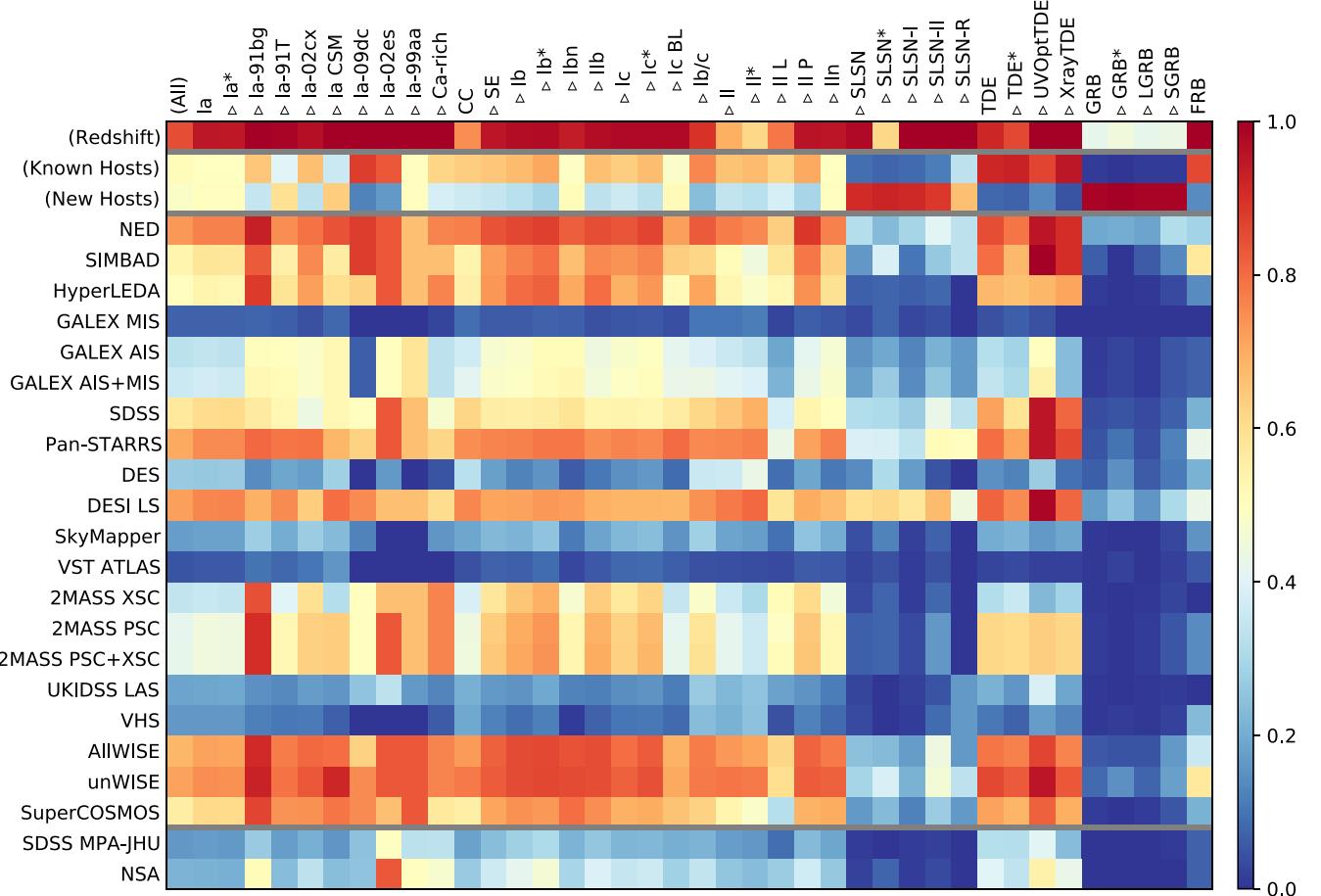
Figure 13. Fractional converge of host properties within each data set, for each major transient type. The fractions of events with valid redshift, known hosts, or newly identified hosts are also indicated in three separate columns. The fractional coverage accounted here primarily reflects the coverage of a catalog in sky area and depth. Wide-area sky surveys, including PS1, WISE, SDSS, and SuperCOSMOS, have generally better coverage over almost all transient types. The depth or sensitivity of a catalog may also have a nonnegligible effect on the availability. Furthermore, transient subtypes that require detailed follow-up to identify have relatively complete host properties as they are usually biased to lower redshifts.

properties is naturally expected due to the heterogeneity of our transient sample and external catalogs, where the cosmic volume is not coherently sampled on both sides and events are classified in different ways.

5.4. Host Photometric Magnitudes

The external catalogs we choose have substantial overlap in their nominal photometric bands. Here we compare photometric magnitude distributions of known hosts and newly identified primary candidates, grouped by their photometric bands (Figure 14). Even for the same photometric band, the filter profiles, photometeric techniques, and calibrations could be quite dissimilar across these surveys. We could compare the reported magnitudes in different surveys if they measure the same region of SED and the results are in the same magnitude system. For a fair comparison, we cut magnitudes to a signal-to-noise ratio of 5.

Besides the clearly different sample sizes and limiting magnitudes as we have discussed in Section 4, we also noticed that the double-peaked redshift distribution also imprints on the photometric magnitudes of a few bands, including *grizy* and *W1*, *W2*. More specifically, optical surveys with substantial coverage in the northern hemisphere, like SDSS, MzLS/BASS, and PS1, tend to have such features. Southern hemisphere

surveys, such as DES, VST-ATLAS, and SkyMapper, do not have a secondary peak. Even excluding SNe in SDSS Stripe 82, which contributed a significant fraction of the medium- to high-redshift events (Figure 4, 5), the second peak remains in *grz* bands. However, when excluding events above $z \simeq 0.1$, the second peak in *grz* bands disappears. Therefore, we expect that the double-peaked magnitude distribution is related to the double-peaked transient redshift distribution, where high-redshift events in the northern hemisphere (or northern Galactic cap) contribute to the fainter, secondary peak. The fainter hosts of low-redshift events do not lead to such double-peaked distributions.

5.5. Photometry Calibration across Catalogs

As we have discussed above, the external catalogs we choose overlap significantly in sky coverage and filter set. Many hosts, as a result, could have multiple magnitudes reported for the same (yet not exactly identical) band, contributed by different surveys, reductions, or measurements. These magnitudes might be inconsistent, i.e., their offsets are greater than their claimed random errors. There could be catalog-wise systematic errors, which are not properly accounted for in their reported magnitudes.
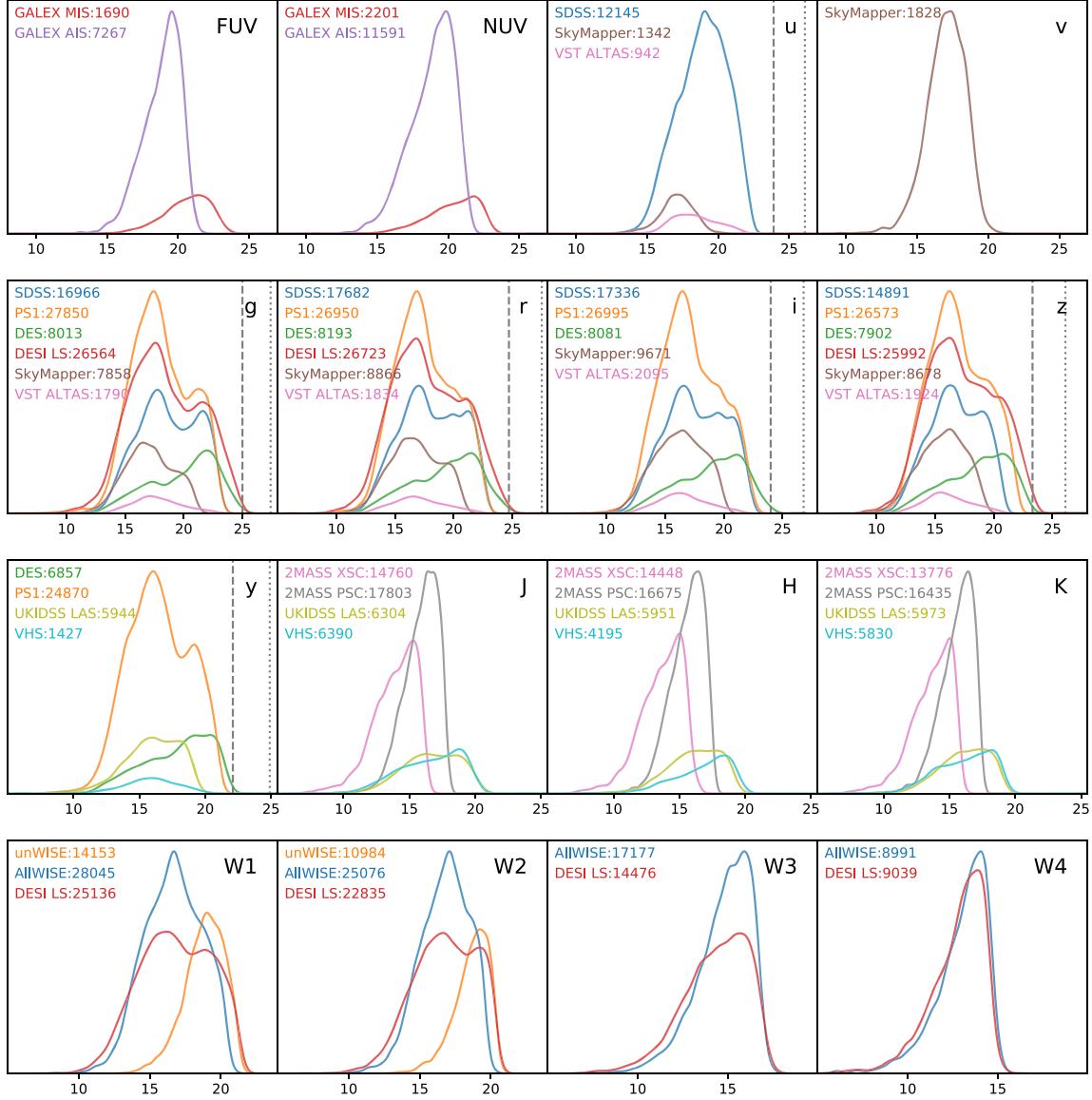
Figure 14. The photometric magnitude (AB) distribution of known hosts and newly identified hosts in our external catalogs, grouped by their nominal photometric bands. In each panel, the photometric band is indicated at the upper right corner, while the catalog and number count of valid values are listed in the legend. Histogram heights are scaled to the highest peak in each photometric band. We applied a signal-to-noise cut of 5 for a fair comparison. Vega magnitudes are converted to the AB system using the zero-points in Jarrett et al. (2011). Note that deep optical catalogs with substantial coverage in the northern hemisphere have prominent double-peaked magnitude distributions. These events are mostly at higher redshifts, rather than in fainter hosts at lower redshifts.

Many factors, like detailed filter profiles and photometric calibration, may contribute to the offsets here. Beyond that, perhaps the most important factor is the choice of photometric techniques. Modern surveys measure extended objects using either aperture photometry (e.g., Petrosian 1976; Kron 1980) or profile-fitting photometry (e.g., Lupton et al. 1999; Lang et al. 2016). Considering factors like the galaxy light profile, morphological uniformity, source angular size, and sky background level, none of these methods is fully immune to potential biases. Moreover, de-blending (or not) can also be a source of inconsistencies in magnitudes, especially for crowded fields or well-resolved galaxies. Therefore, careful selection and cross-calibration of external catalogs might be required to reduce biases when combining or comparing the photometric measurements for each band. Or at least, the systematic uncertainties should be estimated and taken into account for these measured magnitudes. This remains a part of our future work.

6. Data Format and Availability

This database is designed to be maintainable and updated regularly to incorporate recently detected and classified transient events, as well as newly released survey data sets for host galaxies. The changes in upstream data sources and the updates to this catalog should be traceable using a version control solution.

The data is publicly available as a document-based database and can be imported and accessed with standard database protocols. Finally, static, "clean" versions of the database are provided on a regular basis, in standard machine-readable formats like FITS and NumPy.

6.1. Database Structure

Most modern astronomical databases, particularly those used in sky surveys, are relational databases, or sometimes loosely called "SQL-like" databases. Records in a relational database are organized into separate tables with strict predefined formats ("data model" or "schema"), and records are interlinked by reference keys within, and across tables. However, the heterogeneity and complexity of our data bring challenges to conventional table-based relational database systems. Creating separate tables for every data source would be inevitable in this case, let alone accessing or updating tables via enormous reference keys and status flags.

We use Mongodb,[44] a document-oriented database solution, to store our data. Its design philosophy grants great flexibility in data structure and naturally fits our purpose. Such a database keeps collections of *documents*, where every document is an independent, dictionary-like (key-value style) data container with versatile, even nested or hierarchical, structures. For a single event, the query results and cross-identified hosts are independent of other events. Rather than distributing these records across multiple tables, it is more convenient to group them into one self-contained document per event. This is also easier to access and maintain compared to joining multiple tables by reference keys in SQL-like languages.

Our database contains multiple data collections: (1) a master collection of transient events ("events"), including their sky coordinates, redshifts, classification, and reported host galaxies, along with other basic characteristics; (2) a collection for query results in NED and SIMBAD ("vacs"), as the center coordinates and searching radii used in other catalogs relies on results returned from VACs; (3) multiple data collections for the query results in survey catalogs, grouped by their data sources ("vizier," "local" for local catalogs, "mast," "datalab," and "sdss" for SDSS SkyServer); (4) a collection for cross-identified hosts ("host_summary"), including the list of cross-matched objects, their properties in various catalogs, and their ranks when multiple candidates are present. Every unique event points to a document in each collection that shares a unique index.

Maintaining multiple collections is mainly for easier implementation and better maintainability. Finally, as an end-user product, we also created a cleaned-up version, where transient characteristics and properties of hosts or candidates are combined into a stand-alone collection, with user-friendly structures. Tables in standard FITS and NumPy formats are also released on a regular basis, but only the properties of confirmed hosts and primary candidates are included.

6.2. Maintenance of the Database

We plan to update the database on a regular basis to incorporate recent transient events in our data sources and the latest survey catalogs for host properties. This requires the following: (1) a mechanism to trace the change in upstream data sources, (2) bookkeeping of historical versions after a document is updated or removed, and (3) extendable catalog searching and object cross-matching codes to easily incorporate new survey catalogs.

To trace the change in the transient event data sources, we generate "version control codes" (VCCs) for each event, with

one VCC per data source. This is the MD5 hash value of the original record in formatted, plain text representation. Any change in the data source, even a very minor one, would result in a different VCC than the previous one, indicating that catalog queries and host cross-matching must be performed again for this event. Besides VCCs, we also set human-readable date stamps for each update. Moreover, events that no longer exist in any data source would be marked for removal.

To trace the changes within our database, when a document is updated or removed, we keep a snapshot of the previous version. These snapshots are stored in separate collections, indexed with the unique identifiers of transient events, and organized in chronological order for each event. This is done only for our transient, VAC, and host collections, as the contents of survey catalogs are static, and their versions are indicated by the generation of the data release.

Finally, to ensure that our database is extendable, we have developed standardized, general-purpose query scripts for external survey catalogs. New catalogs on Vizier, Data Lab, and MAST can be included using configuration files, without adding new code. We also create separate task scripts for our transient data sources yet share those commonly used routines (e.g., determining search radius) to retain the modularity of the program and the consistency of their behavior.

6.3. Data Access

We provide the cleaned-up, reorganized version of the host property collection in three formats for users with different purposes:

1. The human-readable, self-contained, plain text JSON format uses one file per event. JSON is a conventional data exchange format that stores data in nested dictionary-like (*key-value* style) or array-like structures. This is suitable for those interested in an individual or smaller sample of events. Also, when performance is not a critical factor of consideration and query operations are not required, JSON files can also be used instead of other formats. The format is also used by the OSC.
2. The full snapshot of the host property collection in BSON format, known as BSON, is a binary analog and extension of JSON format that is used internally in MongoDB for data storage. The snapshot contains one single large file that can be directly imported into a local MongoDB database. This version is suitable for those using a significant portion of the data set, or those who need query and aggregation capabilities.
3. The compiled tables of transient and host properties, in standard NumPy and FITS formats, is a format that only includes properties of known and newly identified hosts, without detailed data quality control flags and properties of other cross-matched objects that may contain the right host when our host identification fails. This is a convenient format when only properties of robustly identified and cross-matched hosts are concerned.

These data are available at doi:10.5281/zenodo.5568962. The data schema for the host collection is illustrated using an example in Appendix L.

7. Summary

We assemble a database for extragalactic transient events and the properties of their host galaxies. Transients including SNe, TDEs, GRBs, FRBs, and GW events are compiled from various data sources. Based on the availability of host names or coordinates in those data sources, we search the value-added catalogs and survey catalogs for photometric and spectroscopic properties of potential host galaxies. The returned objects and sources in those catalogs are then cross-matched to find host properties. For events without known hosts, we also rank cross-matched objects to identify their best host candidates and compile their properties.

We design detailed strategies for source queries in external catalogs to maximize the chance of finding the right host. We search sky coordinates of host galaxies whenever available, and we always obtain host coordinates using value-added catalogs if their names are provided. For transient events without host coordinates, we search their event coordinates. The search radius is then determined with a fixed projected distance at the transient redshift, or a default radius when transient redshift is unknown. The search radii of GRBs, as an exception, are partly based on their localization accuracy.

We also design custom procedures to improve the accuracy of object cross-matching and host candidate ranking when necessary. When cross-matching sources across catalogs, we identify groups of spatially associated sources by their pairwise angular distances, the empirically calibrated per-catalog astrometric tolerances, and the per-field matching thresholds. When required, we rank cross-matched groups to find new host candidates as per their angular distances to the reported transient coordinates, the results of source matching in external catalogs, the geometric properties of the cross-matched groups, their possibility of being foreground stars, and optionally, detailed source properties in cross-matched catalogs. Here stars are identified and excluded using relatively conservative photometric, morphological, or astrometric criteria. Primary host candidates are identified using a Logistic Regression-based ranking algorithm, trained with transient-host pairs with the known, properly cross-matched, and visually inspected host galaxies. The ranking algorithm reaches above 97% overall accuracy in our cross-validation using known hosts. Also, the visual inspection of newly identified hosts reveals a similar accuracy.

Aiming to optimize the coverage of host properties, we select source catalogs of wide-area sky surveys with complementary survey regions and filter sets. We search catalogs of GALEX and WISE missions for UV and mid-infrared photometry. Meanwhile, we use catalogs of various ground-based surveys for optical-IR properties, including 2MASS, VHS, and UKIDSS for near-infrared, as well as SDSS, DES, DECaLS, MzLS/BASS, PS1, VST-ATLAS, and SkyMapper for optical photometry. Extra properties are supplied by NSA, HyperLEDA, and MPA-JHU catalogs.

We provide our data in human-readable, plain text `JSON` files, `MongoDB` database snapshots, and standard `NumPy/FITS` tables. The data, including documentation and examples, can be accessed online at doi:10.5281/zenodo.5568962. Source code for this project is also available at https://github.com/shiaki/THEx-database.

By the sample size of transients, the coverage of transient types, the number of newly identified host candidates, and the availability of host galaxy properties, our database is the largest publicly available data set for transient host galaxies.

Appendix A
Acronyms and References for Galaxy Data Sources

We list the acronyms of survey programs and data sources in Table 6, along with their references.

Table 6
Acronyms and References for Missions, Surveys, and Online Services

Name	Acronym	References
Galaxy Evolution Explorer	GALEX	Martin et al. (2005)
Sloan Digital Sky Survey	SDSS	York et al. (2000)
Panoramic Survey Telescope and Rapid Response System	Pan-STARRS	Chambers et al. (2016)
Mayall z-band Legacy Survey	MzLS	Dey et al. (2019)
Beijing-Arizona Sky Survey	BASS	Zou et al. (2017)
Dark Energy Camera Legacy Survey	DECaLS	Dey et al. (2019)
Dark Energy Survey	DES	DES Collaboration et al. (2016)
VLT Survey Telescope ATLAS	VST-ATLAS	Shanks et al. (2015)
SkyMapper Southern Sky Survey	SkyMapper	Keller et al. (2007)
Two Micron All-sky Survey	2MASS	Skrutskie et al. (2006)
Vista Hemisphere Survey	VHS	McMahon et al. (2013)
UKIRT Infrared Deep Sky Survey	UKIDSS	Lawrence et al. (2007)
Wide-field Infrared Survey Explorer	WISE	Wright et al. (2010)
SuperCOSMOS Sky Surveys	SuperCOSMOS	Hambly et al. (2001)
NASA/IPAC Extragalactic Database	NED	Helou et al. (1991)
SIMBAD Astronomical Database	SIMBAD	Wenger et al. (2000)
HyperLeda Extragalactic database	HyperLEDA	Paturel et al. (2003)
NASA-Sloan Atlas	NSA	...
Gaia	...	Gaia Collaboration et al. (2016)

Note. References listed are project or mission description papers.

Appendix B
Standardization of Type Labels

Below we discuss our detailed procedure to standardize transient type labels. Here typewriter fonts indicate the actual standardized labels used in our database.

Homogenizing type labels. When more than one type label refers to essentially the same group of transients, we choose a single, unique label to designate all of them. To decouple type labels from yet-unclear physical scenarios or progenitor properties, we follow the rule that rare subtypes are named either after their commonly referred to prototype events or by their most distinctive observational signatures. For example, "Iax" and "Ia-02cx" are used interchangeably in the literature for a subgroup of faint-and-slow thermonuclear SNe, so we combine events under the "Iax" label into `Ia-02cx.` Similarly, the "Ia-02ic," "Ia-11kx," and "Ian" labels are merged into the more descriptive and commonly used `Ia-CSM.` Events labeled as "Ia-SC" (super-Chandrasekhar) and "Ia-06gz" are grouped into `Ia-09dc` because "Ia-SC" makes assumptions about the progenitor properties and "Ia-06gz" is less commonly used compared to `Ia-09dc.` This is already partly done within OSC, and we further extended their list of predefined synonyms.[45]

Handling hybrid, intermediate, or transitional types. A very small fraction of events are labeled as such composite types. The rule to relabel these composite types depends on the reason for the classification. For composite labels due to the mixed usage of criteria or techniques, we split the labels and set them separately. As an example, SNe classified as "IIn P" are labeled as *both* `IIn` and `II P` simultaneously, because "n" indicates a spectral signature while "P" indicates a light-curve shape. For unstandardized notations of later-defined subtypes in this style, we use the proper subtype labels instead. Events classified as "Ia/IIn," for example, will be labeled as `Ia CSM`

following this rule, rather than `Ia` *and* `IIn` as in the previous case. Finally, for composite types, due to their ambiguous photometric or spectroscopic signatures, we also split the labels. Meanwhile, we add an additional "`AMBIGUOUS`" flag to these events (see later). Events classified as "Ia-CSM/IIn" (not Ia/IIn) will receive both the "`Ia-CSM`" and "`IIn`" labels, with "`AMBIGUOUS`" flags. An exception to this rule is that Type Ib/c SNe are *not* labeled as both "`Ib`" and "`Ic`." Instead, we reserve a "`Ib/c`" label for these events.

Rebuilding the hierarchy of type labels. The type labels of each event are aggregated from multiple reference sources, often with variations in the granularity of their label system. To fit these type labels into a consistent and maximally compatible classification scheme, we assign each event its best-refined subtype (there can be multiple) *and* all their *physical* parent labels, so that each label in our database points to a complete sample of events, including its subtypes. The hierarchy of transient types is listed in Table 2. For example, we always add "`Ia`" labels to events with "`Ia-91T`," "`Ia-91bg`," and other similar SN Ia subtype labels. Events classified as "`Ic BL`" are also labeled as "`Ic`," "`SE`" (stripped envelope), and "`CC`" (core collapse). Note that SNe IIb are considered as a subtype of Ib, instead of Type II. These events are more related to stripped-envelope SNe (Ib, Ic; Filippenko et al. 1993). Similarly, earlier works classify Ca-rich transients as a subtype of SN Ib, but here we group these events under SN Ia, as they are likely from thermonuclear events (Perets et al. 2010). Furthermore, we set a separate branch for SLSNe ("`SLSN`"), due to their nonhomogeneous observational signatures and possible diversity of physical scenarios (Gal-Yam 2019).

Flagging ambiguous and conflicting cases. Besides curated type labels, we also set special quality control flags to indicate the status of label homogenization. For a single event, when its original list of type labels contains an ambiguous label that has been split into two labels with physically incompatible types, we mark the event as an ambiguous case ("`AMBIGUOUS`"). Meanwhile, when its original list of type labels contains any

---

[45] github.com/astrocatalogs/supernovae/blob/master/input/type-synonyms.json

two labels with incompatible types (i.e., inconsistent classification in different reference sources), we label the event as a conflicting case ("CONFLICT"). Here any two labels become incompatible when they are not on the same major branch of the classification tree (Ia, stripped-envelope "SE," II, and non-SN labels), as listed in Table 2.

Flagging peculiar events. Events of relatively rare subtypes are always labeled as peculiar ("Pec"), and "Pec" labels in original data sources are also preserved. Here we do not consider "Pec" as a subtype or an intermediate level in the hierarchy of classification. This special label is used as a flag, rather than a subtype. Therefore, "Ia Pec" events are always labeled as "Ia" and "Pec" simultaneously. Meanwhile, other rare and more specified subtypes of SNe Ia, such as Ia-91T and Ia-91bg, are also labeled as "Pec." There could be known subtypes that are only classified as peculiar events, without detailed subtype labels. We further consult the review paper of Taubenberger (2017) to assign the proper subtype labels for 60 peculiar SNe Ia in our sample.

Assigning type labels for GRBs, TDEs, and rare events. As discussed in Section 2.1.3, we always assign "GRB" labels for both long and short GRBs. Whenever possible, we use their T_{90} value (the time when 90% of photons reach the detector) to further classify them into short GRBs ("SGRB," $T_{90} < 2$ s) or long GRBs ("LGRB," $T_{90} \geqslant 2$ s). When there is an associated optical supernova detection (rather than the afterglow), we also preserve other type labels assigned to this event. Similarly, we assign the "TDE" labels for all TDEs in our combined list. For records in the review from French et al. (2020), we further assign "UVOptTDE," "XrayTDE," "LikelyXrayTDE," and "PossibleXrayTDE" per their wavelengths and the quality of detection. FRBs and GW events are labeled as "FRB" and "GW," respectively. Except for GW170817, other events in the Open Kilonova Catalog are only labeled as candidates ("PossibleKilonova").

Appendix C
Host Galaxy Reference Sources

The host galaxy names and coordinates in the OSC are collected from a wide variety of reference sources, including astronomical catalogs, individual papers, as well as supernova discoveries reported to the TNS, The ATel, and, historically, to the Central Bureau of Astronomical Telegrams that issues International Astronomical Union Circulars (IAUCs) and Central Bureau Electronic Telegrams (CBETs). We use their host names and coordinates to match sources in external catalogs and to train our ranking functions. The contribution and data quality of these reference sources, therefore, must be assessed beforehand. Here we focus on whether the provided host names or coordinates point to any possible host galaxies, while the reliability of transient-host association here (against other nearby galaxies) will be addressed later in Appendix J.1.

We collect the host names and coordinates reported in OSC along with their reference sources (catalogs, individual papers, TNS/ATel reports, and IAUC/CBET issues). For supernova

reports in TNS/ATel/IAUC/CBAT, we consider the entire platform or series, instead of individual indexed reports therein, as the data source. We resolve host names in NED and SIMBAD to find their corresponding sky coordinates. Coordinates in NED are preferred when names are recognized at both sides. To examine if these name-resolved or as-provided coordinates are valid coordinates of galaxies, we obtained DESI-LS image cutouts using the Sky Viewer.[46] These image stamps are centered at transient coordinates, with the box size covering four times the largest transient-host angular offset among all known host coordinates of the event. Name-resolved and as-reported host coordinates of this event are plotted onto the image, with anonymized names of reference sources. Images are also randomly shuffled to minimize consistency issues in this quick inspection.

We classify these host coordinates into four cases: (1) "Good," host coordinate is clearly centered on a possible host galaxy in the image; (2) "Off," host coordinate points to a possible host galaxy in the image but deviates from its center (or the most prominent component) by more than 2″, which *may* challenge the cross-matching algorithm; (3) "Missed," host coordinate points to a galaxy whose association with the transient is dubious, or no galaxy is visible at the indicated position due to erroneous coordinate or very faint host, and (4) "Unknown," inspection cannot be performed due to image issues. When the host is ambiguous in the image, we consider the coordinate "Good" or "Off" when it points to any possible galaxy, without determining if the galaxy appears to be the most likely host or not.

In Table 7, we summarize the number of host names and coordinates provided in each reference source ("Total") and the number of host names that are recognized in either NED or SIMBAD ("Resolved"). Due to the limited sky coverage of DESI-LS, we are only able to inspect a subset of host names and coordinates ("Inspected"). We evaluate the quality using the "Good" fraction and "Missed" fraction, with the total number of "Good," "Off," and "Missed" as the denominator. Most reference sources have "Good" fractions above 90%. Notably, the host coordinates in the Asiago Supernova Catalogue have a relatively lower "Good" fraction due to the truncation of R.A. (in "hh:mm:ss" format) at seconds, leading to an average position shift of several arcseconds. The issue is also propagated into the data set of Lennarz et al. (2012). Therefore, we ignore the host coordinates provided in these two reference sources. Furthermore, host coordinates from individual papers are also ignored for their low overall "Good" fraction.

Finally, it should be emphasized that the results shall not be considered as an examination of their original host galaxy assignments. The resolution and sensitivity of image cutouts, the procedure of data compilation in OSC and its upstream data sources, the accuracy of coordinates in NED and SIMBAD, and, fundamentally, the empirical and nonexhaustive nature of this inspection can all affect the "Good" fraction of these reference sources.

---

[46] https://www.legacysurvey.org/viewer

Table 7
Major Contributors of Host Information in the Open Supernova Catalog

Reference Source	Total	Resolved	Inspected	Good	Off	Missed	Unknown	Fraction	
								Good	Missed
Latest Supernovae[a]	8716	8654	6482	6020	186	68	208	0.960	0.011
Asiago Supernova Catalogue (Barbon et al. 1999)	6746	3461	2625	2437	60	13	115	0.971	0.005
Transient Name Server[b]	4822	4785	3731	3475	104	38	114	0.961	0.011
SDSS-II Supernova Survey Data Release (Sako et al. 2018)	4300	487	491	463	1	14	13	0.969	0.029
"A Unified Supernova Catalogue" (Lennarz et al. 2012)	3860	3393	2706	2519	47	15	125	0.976	0.006
The Pan-STARRS Survey for Transients[c] (Huber et al. 2015)	2770	2755	2236	2109	47	40	40	0.960	0.018
The Astronomer's Telegram[d]	2689	2552	1636	1515	48	27	46	0.953	0.017
International Astronomical Union Circulars	1785	1783	1406	1300	34	10	62	0.967	0.007
Redshift-independent Distances in NED (NED-D v13.1.0; Steer et al. 2017)	1566	1564	1432	1315	24	20	73	0.968	0.015
Central Bureau Electronic Telegrams[e]	1417	1409	1013	931	24	6	52	0.969	0.006
ASAS-SN Supernovae[f]	1110	1060	779	726	22	10	21	0.958	0.013
Gaia Photometric Science Alerts[g] (Hodgkin et al. 2021)	845	841	525	492	18	7	8	0.952	0.014
UC Berkeley Filippenko Group's Supernova Database[h]	656	655	520	484	9	2	25	0.978	0.004
Berkeley SN Ia Program Low-redshift SN Ia Sample (Silverman et al. 2012)	578	577	460	424	18	5	13	0.949	0.011
Catalina Real-time Transient Survey, Supernova Hunt[i]	393	367	296	279	6	6	5	0.959	0.021
…	266	255	200	187	3	4	6	0.964	0.021
CBAT Transient Objects Confirmation Page[j]	335	334	259	243	10	0	6	0.960	0.000
SDSS-II Photometrically Classified SN Ia (Campbell et al. 2013)	317	317	320	303	1	11	5	0.962	0.035
Cosmicflows-2 (Tully et al. 2013)	288	288	228	212	7	0	9	0.968	0.000
LOSS Stripped-envelope SN Sample (Shivvers et al. 2019)	218	218	167	155	2	0	10	0.987	0.000
LOSS SN Ia Light-curve Data Set (Ganeshalingam et al. 2010)	162	162	139	127	6	0	6	0.955	0.000
SDSS-II, SN Ia Host Properties (Gupta et al. 2011)	156	156	159	150	0	4	5	0.974	0.026
Jha et al. (2007)	131	131	104	95	6	1	2	0.931	0.010
Reindl et al. (2005)	111	111	92	83	6	1	2	0.922	0.011
Weyant et al. (2014)	100	100	73	63	2	0	8	0.969	0.000
Wang et al. (2006)	98	98	73	66	5	0	2	0.930	0.000
CfAIR2 SN Ia Near-infrared Light-curve Sample (Friedman et al. 2015)	95	93	71	64	2	1	4	0.955	0.015
CfA4 SN Ia Light-curve Sample (Hicken et al. 2012)	89	89	74	67	2	3	2	0.931	0.042
LOSS SN Ia Photometric Data Release (Stahl et al. 2019)	86	86	69	63	3	0	3	0.955	0.000
Parodi et al. (2000)	67	67	56	50	4	1	1	0.909	0.018
Rodríguez et al. (2014)	52	52	40	38	1	0	1	0.974	0.000
Galbany et al. (2016)	50	50	35	34	0	0	1	1.000	0.000
Other Reference Sources (five or more SNe)	1429	1422	1236	1072	37	49	78	0.926	0.042
Other Reference Sources (fewer than five SNe)	389	362	328	270	14	2	42	0.944	0.007
Asiago Supernova Catalogue (Barbon et al. 1999)	6491	...	5654	1712	1637	1901	404	0.326	0.362
"A Unified Supernova Catalogue" (Lennarz et al. 2012)	5509	...	4892	2448	346	1732	366	0.541	0.383
SDSS-II Supernova Survey Data Release (Sako et al. 2018)	3813	...	3757	3619	48	34	56	0.978	0.009
Pan-STARRS Photometrically Classified SNe (Jones et al. 2018)	1163	...	1198	1152	2	8	36	0.991	0.007
SDSS-II SN Ia Rates Sample (Smith et al. 2012)	341	...	334	321	1	9	3	0.970	0.027
Prieto et al. (2008)	254	...	258	249	3	0	6	0.988	0.000
Nearby Supernova Factory SN Ia Host Sample (Childress et al. 2013)	253	...	229	200	17	6	6	0.897	0.027
CANDELS SN Ia Rate Sample (Rodney et al. 2014)	63	...	78	58	1	0	19	0.983	0.000
Other Reference Sources	15	...	12	5	4	2	1	0.455	0.182

Notes. Table shows the statistics of reference sources for host names (before the horizontal line) and for host coordinates (after the horizontal line). Columns are as defined in Appendix C.
[a] https://www.rochesterastronomy.org/supernova.html
[b] https://www.wis-tns.org/
[c] https://star.pst.qub.ac.uk/ps1threepi/psdb/
[d] https://astronomerstelegram.org/
[e] http://www.cbat.eps.harvard.edu/cbet/RecentCBETs.html
[f] http://www.astronomy.ohio-state.edu/asassn/sn_list.html
[g] http://gsaweb.ast.cam.ac.uk/alerts/home
[h] http://heracles.astro.berkeley.edu/sndb/
[i] http://nesssi.cacr.caltech.edu/SNhunt/, which appeared as two separate reference sources in OSC (see the following line).
[j] http://www.cbat.eps.harvard.edu/unconf/tocp.html

Appendix D
Accessing External Catalogs

D.1. Resolving Host Names in Value-added Catalogs

Here we discuss the procedure to resolve existing host names and choose the best name-resolved coordinate for the host.

We choose NED and SIMBAD to resolve existing host names,[47] as they are the most commonly used astronomical databases and presumably have the best-available compilation of object names, sky coordinates, redshifts, and other basic data. These VACs indeed provide some measured and derived properties of hosts, but the data are usually incomplete or not up-to-date. Accessing other survey catalogs for host properties remains necessary to compile their properties. There are other online services like HyperLEDA that are capable of resolving object names, while scripted access is not supported yet.

We collect host names of each event from reliable reference sources (Appendix C). These host names are resolved individually and separately in NED and SIMBAD to obtain their corresponding coordinates and other basic data. We exclude resolved coordinates with large uncertainties (error circle above $2\rlap{.}''5$ in NED, precision grade of 5 or below in SIMBAD), and then choose the best resolved coordinate of each host name from the remaining ones. For each name, we use a coordinate in NED, unless the name is *only* resolved in SIMBAD. Host names without resolved and precise coordinates in either VAC are considered as unresolved.

When there are more than one resolved host names at this moment, we check their consistencies. Pairs of name-resolved coordinates with separations above $5''$ are considered inconsistent, which implies possible disagreement of transient-host association among reference sources. In the presence of any inconsistent pair, we skip the remaining steps and conclude without a best name-resolved coordinate.

Finally, when no host name is successfully resolved, we leave without a best name-resolved coordinate. When there is only one resolved name, we use it for the best name-resolved coordinate. When multiple names are resolved, we choose one for the best name-resolved coordinate. We prefer common galaxy names (e.g., NGC, IC, UGC, and PGC galaxies) over IAU-style source names (catalog name + sexagesimal coordinate), and anonymous galaxies ("Anon" + sexagesimal coordinate) have the lowest priority. We also prefer names resolved simultaneously in NED and SIMBAD if the rule above cannot decide the best host.

D.2. Searching Nearby Sources in External Catalogs

When searching host (or transient) coordinates in external catalogs for nearby sources, to optimize the chance of finding the right host in external catalogs, we set the sky coordinate and radius to search based on the following approach. If there is a best name-resolved host coordinate, we use this coordinate to search other catalogs with a search radius of $15''$. When resolving host names, we already have their basic properties in NED or SIMBAD, but we also perform a coordinate-based search again in the same catalog to access other nearby sources. If there is no name-resolved coordinate, but a host coordinate is directly provided in any of those upstream data sources, we use this as-reported host coordinate with a search radius of $30''$.

Finally, when neither name-resolved nor as-reported host coordinate is available, we use transient coordinates to search for nearby sources. The search radius in this situation is set based on the redshift of the event. For events with known redshifts, we set the search radius to match a projected distance of 45 kpc at the transient redshift. Otherwise, we use a default search radius of $30\rlap{.}''$ Specifically, for GRBs with neither a redshift estimate nor a later supernova detection, we use three times the 90% error radius as the search radius. When searching transient coordinates, we clip the search radius within $15''$ and $2'$ to increase the chance of enclosing the right host, without including too many irrelevant non-host objects. For GRBs with search radii set by error circles, we clip the search radii within $5''$ and $15''$. The workflow is also summarized in Figure 15.

When coordinates and search radii are determined, we search for nearby sources of transient events in external catalogs. Depending on their availability, our external catalogs can be classified into two categories, those accessible at online services (via web application programming interface or API, client-side packages, or other protocols) and those provided as static tables in text or binary formats.

For online catalogs, as per the service where they are hosted, we use the proper interface or tool to access their data. Catalogs hosted by the Astro Data Lab of NSF's National Optical-Infrared Astronomy Research Laboratory (formerly NOAO Data Lab) are accessed using their official `Python` client.[48] Meanwhile, catalogs hosted at VizieR[49] and SDSS SkyServer[50] are accessed via the third-party `astroquery` package (Ginsburg et al. 2019). Finally, one catalog hosted at MAST[51] is accessed directly using their web API. We record the detailed return statuses of queries when accessing online catalogs. Failed requests due to any error are flagged to be made again to ensure that our database includes everything available in those external catalogs.

Local catalogs, on the contrary, are provided as static binary or plain text tables and thus cannot be accessed using coordinates and search radii like the online services. We spatially index their catalog sources so that they can be queried with the same interface as we used for online services. Depending on the size of the catalog, we use two different solutions. Sources in smaller catalogs (below 10^7 records) are indexed with kd-tree (as implemented in `SciPy`) using their three-dimensional Cartesian coordinates on the unit celestial sphere. Regarding algorithm complexity, efficiency, and scalability, this is not the optimal solution, but it is quick and easy to deploy. Sources in larger catalogs (above 10^7 records) are indexed with `healpix` (via `healpy`), where sky coordinates are converted to one-dimensional pixel indices in which they reside. Searching for sources near a position becomes finding sources with certain pixel indices, which significantly improves performance. We index sources with different resolution levels, and when searching sources, the best spatial resolution is chosen to match the search radius so that the region is fully covered, and the speed is optimized at the same time. Sources inside the pixels of interest but beyond the search radius are excluded. To further improve performance, before indexing spatial coordinates using either method, we

---

[47] Accessed via `astroquery` (Ginsburg et al. 2019), github.com/astropy/astroquery

[48] github.com/noaodatalab/datalab
[49] vizier.u-strasbg.fr
[50] skyserver.sdss.org/dr16/en/home.aspx
[51] archive.stsci.edu
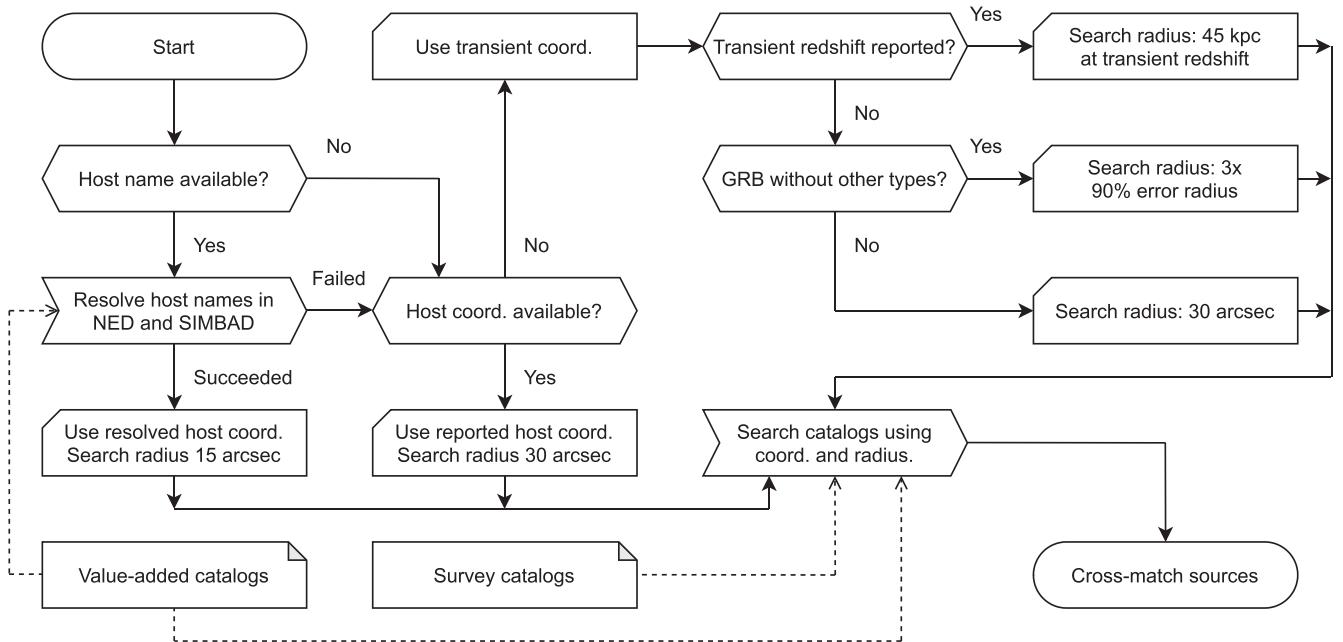
Figure 15. The procedure to access value-added catalogs and survey catalogs for host properties. For each event, we determine the coordinate to search (host or transient coordinate) and the search radius based on the availability of information in our transient data sources. Whenever possible, we search host coordinates resolved using VACs or as provided in the transient data set, instead of the transient coordinate. When searching a transient coordinate, to optimize the chance of covering the true host, we determine the search radius using transient redshift if available. In the absence of transient redshift, the error circle is also used for GRBs.

make quality control cuts and drop the unused columns to reduce the size of the catalog. Catalogs published in text formats are also converted to binary formats for better speed.

Having queried results in various external catalogs, we create local data collections in our database to store them. We do this for each service or interface via which the data are accessed (VizieR, Data Lab, local, etc.). In each data collection, besides adding lists of returned sources in external catalogs, we also add auxiliary information including the search radius and the source of searched coordinates (host coordinates obtained by resolving names, host coordinates as provided in data sources, or transient coordinates).

Appendix E
Algorithm for Cross-matching

We list our algorithm for cross-matching as pseudocode below.

Algorithm 1. Cross-identification of Host Galaxies

---

procedure CROSS_IDENTIFY s, k, is_star, catalog
   input: s: length-n array of real number pairs, projected Cartesian coordinates of sources;
          k: length-n array of real numbers, per-catalog astrometric tolerances for sources in s;
        is_star: length-n array of binary flags, whether the source is star-like (`true`) or not (`false`);
        catalog: length-n integer array, external catalog identifiers of sources.
  output: is_confirmed: binary flag, if primary_host is a confirmed host (`true`) or a primary candidate (`false`);
        primary_host: integer set, indices of sources for the confirmed host or primary candidate;
        secondary_hosts: list of integer sets, indices of sources in each secondary host candidate.

  is_confirmed ← false
  primary_host ← null
  secondary_hosts ← empty list

  for all s_i, $s_j \in S$ where $i < j$ **do**
    $d_{ij}, d_{ji} \leftarrow \|s_i - s_j\|/(k_i^2 + k_j^2)^{1/2}$ ▷ d: (n+1, n+1) matrix, normalized distance;
  for $s_i \in S$ **do**
    $d_{i,n+1}, d_{n+1,i} \leftarrow \|s_i\|/k_i$ ▷ normalized distance to the queried coordinate.
  for $t_i \in$ thres_axis **do**   ▷ thres_axis: array of real numbers, trial matching thresholds;
    $\mathcal{W}_i \leftarrow$ *connected_components* (d, t_i) ▷ $\mathcal{W}$: array of integer sets, possible matching configurations;
    $\mathcal{V}_i \leftarrow$ *count_valid_pairs* $(\mathcal{W}_i, d, t_i$, catalog)   ▷ $\mathcal{V}$: array of integers, number of valid pairs.

  W_{opt} := the unique element in $\mathcal{W}$, which maximizes $\mathcal{V}$ at minimal $|\log(\texttt{thres\_axis})|$ ▷ the optimal configuration;
  t_{opt} := the unique element in thres_axis corresponding to W_{opt} ▷ the optimal matching threshold;
  w_0 := the unique element in W_{opt}, where $(n + 1) \in w_0$ ▷ group matching the queried coordinate.

(Continued)

Q_i ← *quality_metrics* (w_i, s, d, t_{opt}) for w_i in W_{opt} ▷   Q: array of tuples, quality metrics of groups.

if queried host coordinate **and** *length* (w_0) > 1 **then**   ▷   host position is queried and properly matched.
 primary_host ←w_0
 if for all $p \in$ *representative_source* (w_0), is_star[p] = false **then**   ▷   host position matched a non-stellar object.
   is_confirmed ← true
 if using name-resolved coordinate **then**
   is_confirmed ← true

if is_confirmed = false **then**
 for $w_i \in W_{\text{opt}}$ **do**
   $v_i \leftarrow$ *representative_sources* (w_i)
   $\theta_i \leftarrow \overline{\|s_p\|}$ where $p \in w_i$ ▷ θ_i: mean angular distance to the queried coordinate.
   $f_i \leftarrow$ *ranking_function* (θ_i, v_i, catalog, is_star, Q_i,...)
 for w_i in W, sorted by f in descending order, **do**
   if primary_host = null **then**   ▷   transient coordinate is queried, or host coordinate unmatched.
    primary_host ←w_i
   else
     insert w_i to the end of secondary_hosts

function count_valid_pairs$W$, d, t, catalogs $\rightarrow V$
  input: W: list of integer sets, indices of sources in cross-matched groups;
       d: ($n + 1$, $n + 1$) matrix, normalized pairwise distances of sources;
       t: positive real number, matching threshold.
       catalogs: external catalog identifiers of sources;
  output: V: integer, the number of "valid pairs" under this configuration.

  $V \leftarrow 0$
  for $w_i \in W$ **do**
   forAll $p, q \in w_i$ where $p < q$ **do**
    if $d_{pq} < t$ **and** catalog [p] ≠ catalog [q] **then**
      $V \leftarrow V + 1$
function CONNECTED_COMPONENTSd, $t \rightarrow W$
  input: d: ($n + 1$, $n + 1$) matrix, normalized pairwise distances of sources;
       t: positive real number, matching threshold.
  output: W: list of integer sets, a unique partition of integer set $\{1,...,n + 1\}$ that cannot be further refined,
         where for any $w_i \in W$ and for any $i \in w_i$, $\{j; j \in Q, j \neq i$ and $d_{ij} < t\} \subseteq w_i$
  (abridged)

function REPRESENTATIVE_SOURCES(w) $\rightarrow v$
  input: w: set of integers, indices of sources in a cross-matched group.
  output: v: set of integers, a subset of w, in which each external catalog contributes at most one source.
  (abridged)

function RANKING_FUNCTION (θ, v, catalogs, is_star, Q), ... $\rightarrow f$
  input: θ: real number, average angular distance to the origin, i.e., the queried coordinate;
       v: set of integers, indices of representative sources in a cross-matched group;
       catalogs: array of integers, external catalog identifiers of sources;
       is_star: array of binary flags, indicator for stellar sources.
       ...:   other source properties.
  output: f: real number, the ranking score.
  (abridged)

function QUALITY_METRICS(w, s, d, t) $\rightarrow Q$
  input: w: set of integers, indices of representative sources in a cross-matched group;
       s: array of real number pairs, projected Cartesian coordinates of sources;
       d: ($n + 1$, $n + 1$) matrix, normalized pairwise distances of sources;
       t: real number, a cross-matching threshold.
  output: Q: 3-tuple, quality metrics of the cross-matched group.
  (abridged)

Appendix F
Finding Astrometric Tolerances and Matching Thresholds

Every unique object in the sky, in ideal situations, corresponds to a cross-matched group of catalog sources, in which each catalog contributes at most one source. Given a constant matching threshold, when per-catalog astrometric tolerances are excessively large, irrelevant nearby sources can be included in the cross-matched groups, and even nearby cross-matched groups can be *glued* together into one group; conversely, insufficient per-catalog astrometric tolerances, in this case, may leave some catalog sources unmatched, and even split a properly matched group into multiple smaller groups, leading to the incompleteness of compiled properties and the nonunique correspondence of cross-matched groups and true objects in the sky. The astrometric tolerance of each catalog should depend on its astrometry calibration and image resolution, as well as the angular size and signal-to-noise ratio of individual detected sources. Such information, however, is not always provided in external catalogs. We, therefore, determine the per-catalog astrometric tolerances in an empirical approach.

Finding the best per-catalog astrometric tolerances requires a performance metric, i.e., a numeric indicator for the goodness of cross-matching. We use the number of "valid pairs" cross-matched under a constant, unity per-field matching threshold as the performance metric (or "objective function") to be optimized. Here two catalog sources form a valid pair if (1) they are *directly* connected under the existing per-catalog astrometric tolerances and a constant, unity per-field matching threshold, and (2) both sources are uniquely matched in this group, without the confusion of multiple sources in the same catalog. Such a combinatorial objective function is tailored toward large cross-matched groups with the least confusion. We use the first criterion to increase the chance of having sources matched, while the second criterion penalizes confusion of multiple sources—inside a group, we only count valid pairs of uniquely matched sources. The number of sources and their maximal number of possible connections are always limited in a field. When properly matched, the same source can participate in more valid pairs and thus has a higher per-source contribution; meanwhile, a higher fraction of those possible connections would become real. From either perspective, counting the number of connections is a reasonable choice.

Maximizing the objective function is a computationally intensive optimization problem, where the value for each catalog is a free parameter to tune, and evaluating the objective function requires cross-matching all catalog sources in all existing fields for once. Even more complicated is that the objective function is a nondifferentiable and noncontinuous one, which may have a plethora of local extrema that challenges the conventional optimization algorithms. Given the unknown properties of this objective function, instead of using any local or global optimization algorithm, we search the best astrometric tolerances with a Monte Carlo approach. We replace the log-likelihood function in the Markov Chain Monte Carlo algorithm with our objective function so that astrometric tolerances with better overall performances (i.e., a higher number of valid pairs) are preferred during the sampling process. Given enough time, we could find a set of best global astrometric tolerances. It must be emphasized that the Monte Carlo sample, in this case, should not be interpreted as the *posterior* of the best-fitting parameters, as the objective function itself is not a likelihood or any other probability density in nature.

Converting the objective function to a *log-likelihood* function needs some scaling. Any monotonically increasing function, in principle, can be used here as a *scaling function*. However, if the gradient of the scaling function is too steep, then minor changes in the number of valid pairs may lead to a significant change in the density of the Monte Carlo sample. Given that the Monte Carlo sample traces linear density while the value of the log-likelihood function is supposed to be logarithmic, the sampling process can be thus extremely sensitive to local the maxima, leading to *trapped* chains, slow convergence, and an over-constrained, *spiky* Monte Carlo sample. On the contrary, if the gradient of the scaling function is too shallow, then the log-likelihood function cannot effectively perceive the improvement of the objective function, leading to an overly smooth Monte Carlo sample with a low contrast of density. We tested several possible scaling functions, and a simple power function with an index of 0.618 leads to the best balance of sampling efficiency and the contrast of sample density.

There are 30,000 fields in our sample. To evaluate the objective function, all these fields must be cross-matched again using the newly proposed astrometric tolerances. To make the problem computationally feasible, we precalculate the pairwise distance matrices and combine them into larger block diagonal matrices so that each evaluation of the objective function takes only a few seconds on a single core of a conventional desktop computer (Intel Core i7-9700K, 3.6 GHz), and sampling 200,000 steps using emcee (Foreman-Mackey et al. 2013) takes less than a day. We locate the best astrometric tolerances by finding the maximum of the marginalized Monte Carlo sample density using a Gaussian smoothing kernel of $0''.125$, after throwing away the first 20,000 points. The final sampled parameter space is illustrated in Figure 16, and the best-overall parameters are listed in Table 5.

When finding the astrometric tolerance using Monte Carlo sampling, we used a fixed, unity matching threshold for all of our fields. We consider the astrometric tolerance as a globally averaged property of a catalog, reflecting the typical scatter of source coordinates averaged over all existing fields. However, the globally averaged astrometric tolerances are not necessarily the optimal values for individual fields. There could be field-to-field variations of source density and catalog coverage. Meanwhile, the scatter of source coordinates for hosts may also depend on their redshifts. The same set of globally averaged astrometric tolerances could be overly large in certain fields but too small for some other fields, combining or splitting properly matched groups and leading to mismatching or incompleteness of host properties.

To enable some local fine-tuning of the cross-matching criteria, we allow the matching threshold to vary in each field. Tuning the matching threshold is equivalent to scaling the astrometric tolerances by a constant factor in this field. When cross-matching, we use the matching threshold that maximizes the number of valid pairs matched in this field, and the matching threshold is chosen from 31 logarithmically spaced values from 0.316 to 3.16. When several matching thresholds produce the same maximal number of valid pairs, we use the one that is closest to unity.
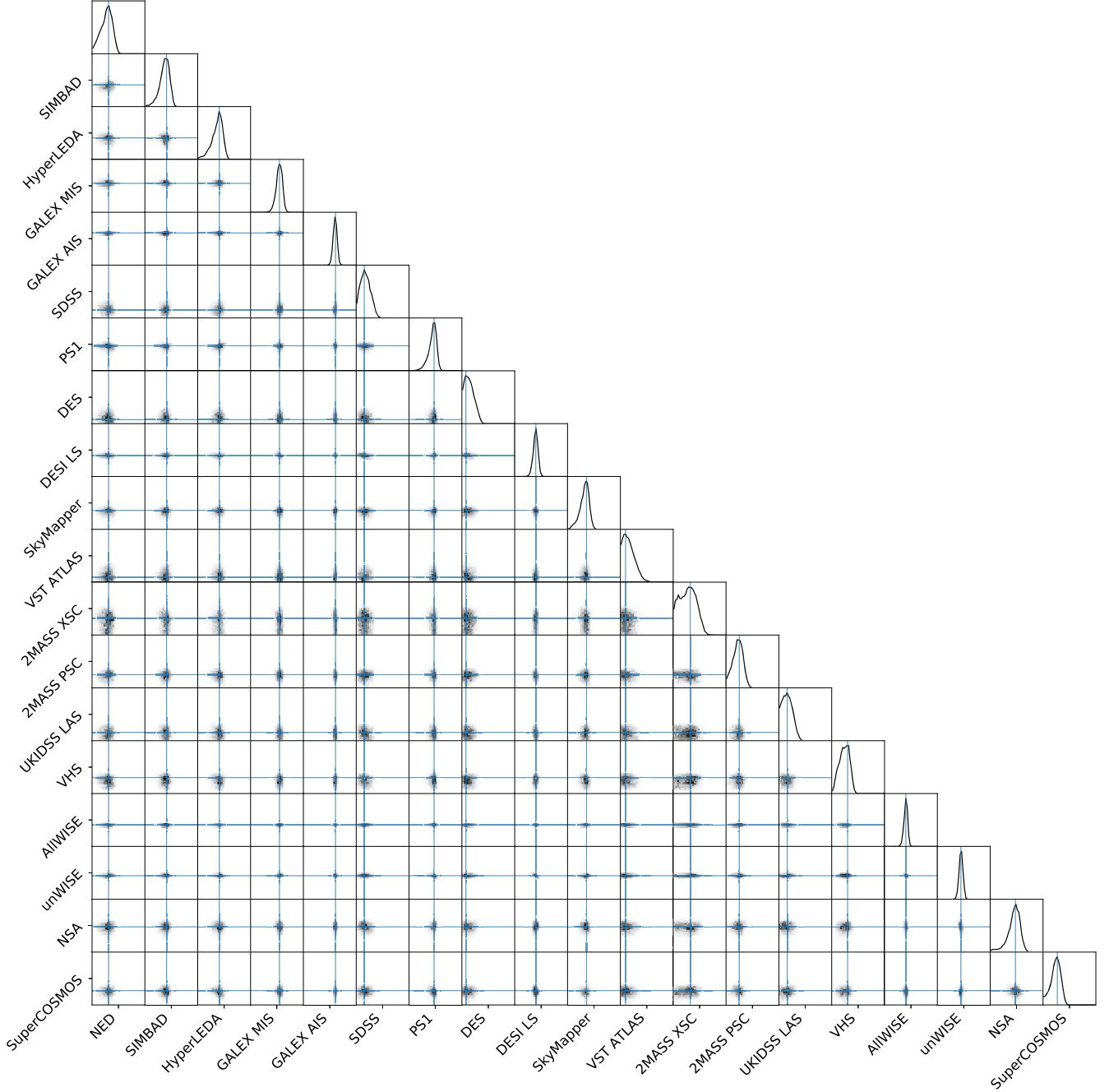
Figure 16. Finding the best astrometric tolerances for external catalogs using Monte Carlo sampling. The results here show a Monte Carlo sample of 200,000 points, optimized for an objective function with a constant power index of 0.618 for the scaling function, as discussed in Appendix F. We use a fixed axis range of $(0'', 3'')$ for all subpanels, including marginalized distributions and two-dimensional projections of the sample. The best-overall astrometric tolerances, determined by searching the peak of marginalized distribution using a Gaussian kernel of $0''.125$, are indicated with blue lines. The detailed values are listed in Table 5.

In Figure 17, we show the distribution of the fine-tuned per-field cross-matching threshold, along with the improvements in the number of valid pairs, compared to the number of pairs matched under a constant threshold of unity. For confirmed hosts and primary candidates, after the cross-matching threshold is tuned, there is about 10% to 20% increase in the number of valid pairs. We note that we allow the matching threshold to vary in each field; the default matching threshold of unity is still the best value in most fields.
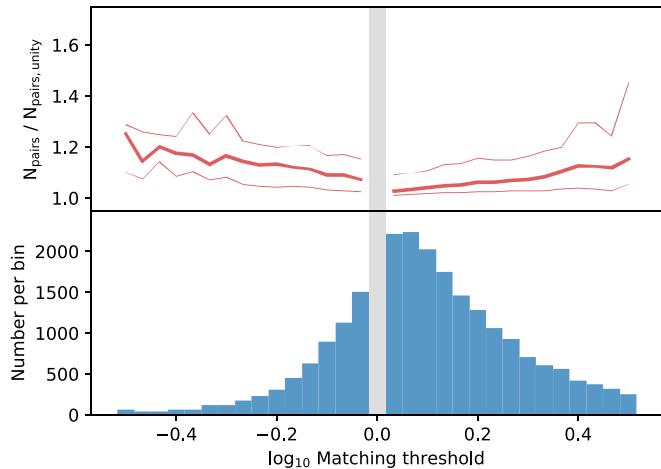
Figure 17. Top: the improvement of connectivity after tuning the matching threshold. We count the number of valid pairs (N_{pair}) using the adjusted matching threshold and compare it to the number of valid pairs matched under unity threshold ($N_{pair,unity}$). Their ratio represents the improvement of cross-matching after per-field fine-tuning of the matching threshold. The solid line shows the moving median of $N_{pair}/N_{pair,unity}$, while the thin lines show the corresponding 25%–75%. Bottom: histogram of the locally adjusted matching threshold. More than half of the fields are cross-matched using a default unity matching threshold (gray zone), so we only show the distribution for those with different matching thresholds.

Appendix G
Quality Metrics of Cross-matching

Cross-matching sources across multiple external catalogs is a nontrivial problem in many aspects. We use per-catalog astrometric tolerances and per-field matching thresholds to determine the connectivity of sources, which we have already discussed in Appendix H. To further ensure that our cross-matching is valid and the compiled host properties are accurate, we introduce three quality control metrics for cross-matched groups:

1. Mean offset (d) measures the average angular distance of sources to the group centroid, i.e., the equal-weighted mean position of sources in this group. This metric measures the size of a cross-matched group, where compact groups should have a lower mean offset.
2. Axis ratio (σ_a/σ_b) measures the *roundness* of a cross-matched group, calculated using the eigenvalues of a second-order moment matrix. This is the axis ratio

(semimajor to semiminor) of sky coordinate distribution, where lower values (closer to 1) indicate a more *isotropic* distribution. Along with this axis ratio, we also provide similar indicators, such as ellipticity and the absolute value of Pearson's r in this case.

3. Degree of connectivity (F) is the number of direct connections in a group divided by the maximal possible number of connections in a group with N sources. This parameter measures the compactness of a group, where a compact group is expected to have most sources interconnected, except for source pairs from the same catalog; meanwhile, a loose, wide-spanning group should have many source pairs unconnected, resulting in a lower value for this number.

For all of the cross-matched groups, including known hosts, newly identified host candidates, and other cross-matched groups, we always provide these three quality control metrics. Figure 18 shows the distribution of these quality control metrics for known and newly identified hosts. We also make sure that these parameters behave well in the limiting case of only a few sources matched, when some parameters may have singular values. For the connectivity, we use the number of direct matches divided by the number of maximal possible matches plus one to avoid division by zero for the single-source groups. For ellipticity, we assume that the single-source groups are close to groups with only two sources. Their axis ratio equals 0, ellipticity equals 1, and Pearson's r is also 1.

Besides these quality metrics, we also provide lists of catalogs that cover the field and that have been matched by the confirmed host or primary candidate of the field. For catalogs covering the field but not matched by the confirmed host or primary candidate, we also provide the nearest source, in these catalogs, to our confirmed host or primary candidate. We note that such kind of situation does not necessarily indicate quality issues in cross-matching, but this can be used as a diagnostic for potentially under matched (*splitted*) groups and to locate missed sources. Detailed statistics of cross-matching are listed in Table 8.

It is worth noting that, nearby large galaxies pose a significant challenge to our cross-matching algorithm, and their properties should be used with care. These well-resolved galaxies could have large catalog-to-catalog offsets in sky coordinates, and sometimes these large galaxies are even split into multiple sources in certain catalogs. We demonstrate relatively successful cases and failed cases in Figure 19.
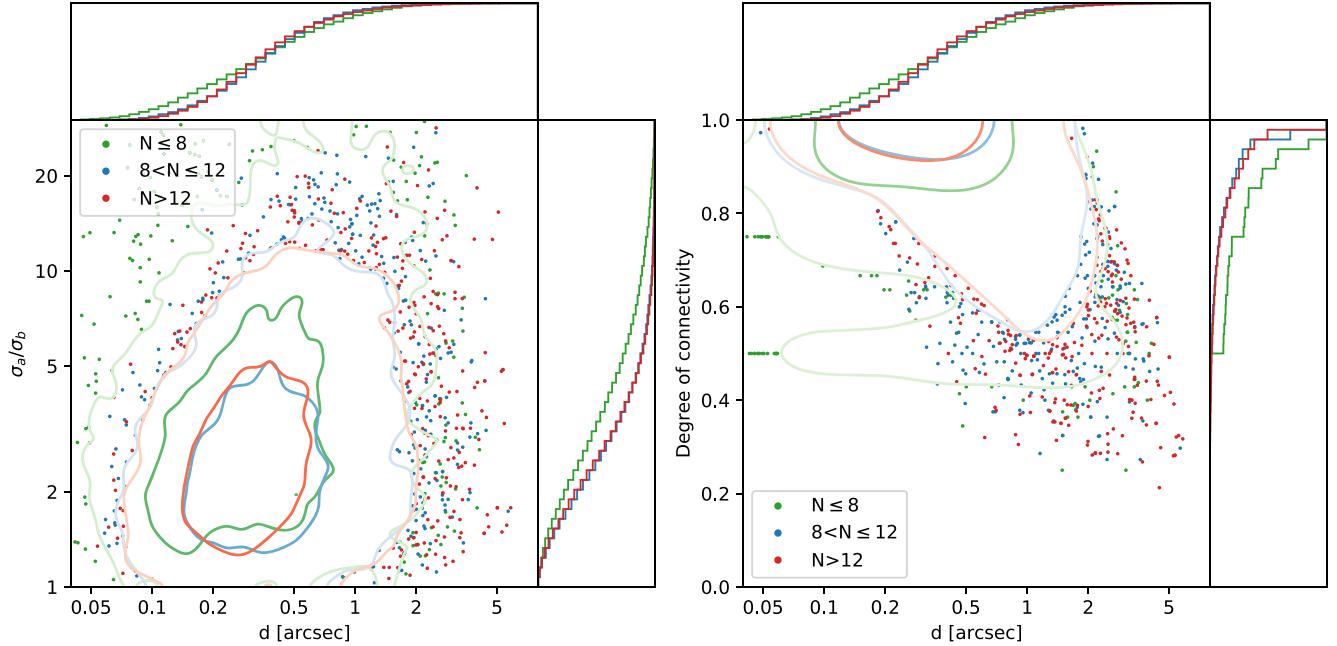
Figure 18. Three quality control metrics of cross-matching for known and newly identified hosts. For cross-matched objects in a group, we calculate their average distance to the group centroid (*mean offset*, d), the axis ratio (semimajor to minor) of covariance ellipse (σ_a/σ_b), and the number of directly connected pairs to the maximal number of allowed connections (i.e., *degree of connectivity*). We split our hosts into three subgroups of nearly equal sizes per their number of cross-matched catalog objects (N). The dark and light contours in each panel enclose 50% and 95% of the hosts in each subgroup, while hosts outside 95% contours are plotted as individual points. The top and right sides of each panel show the cumulative histograms of each metric for each subgroup. Properly matched groups are expected to be compact (low d) and isotropic (σ_a/σ_b near unity), in which catalog objects are adequately connected (high degree of connectivity). Cross-matched groups with extreme values, like large d and σ_a/σ_b, or very low degree of connectivity, should be used with care.

Table 8
Summary of Cross-matching

Catalog	Matched	Nearby		Confusion
		$<2''$	$<5''$	
NED	29070	18	239	733
SIMBAD	21005	38	275	184
HyperLEDA	21123	25	237	27
GALEX MIS	2991	18	251	2
GALEX AIS	16181	93	1170	48
SDSS	22945	33	306	504
PS1	28145	44	450	302
DES	8559	19	116	72
DESI-LS	28439	34	313	1538
SkyMapper	11070	31	249	101
VST-ATLAS	3045	5	32	508
2MASS XSC	14900	11	63	3
2MASS PSC	18745	31	211	389
UKIDSS LAS	8183	16	181	304
VHS	7968	24	185	122
AllWISE	28009	118	1175	102
unWISE	31530	150	1126	966
NSA	9634	14	161	1
SCOS	24265	87	597	375
Gaia	9608	23	750	174
MPA-JHU	7558	5	81	6

Note. *Matched* indicates the number of confirmed hosts and primary candidates with the catalog matched. *Nearby* shows the number of cases when a confirmed host or primary candidate did not match the source in a certain catalog, but there exists sources in this catalog within some angular distance. This need not be a problem of cross-matching. *Confusion* shows the number of confirmed hosts and primary candidates with multiple sources in this catalog that matched.
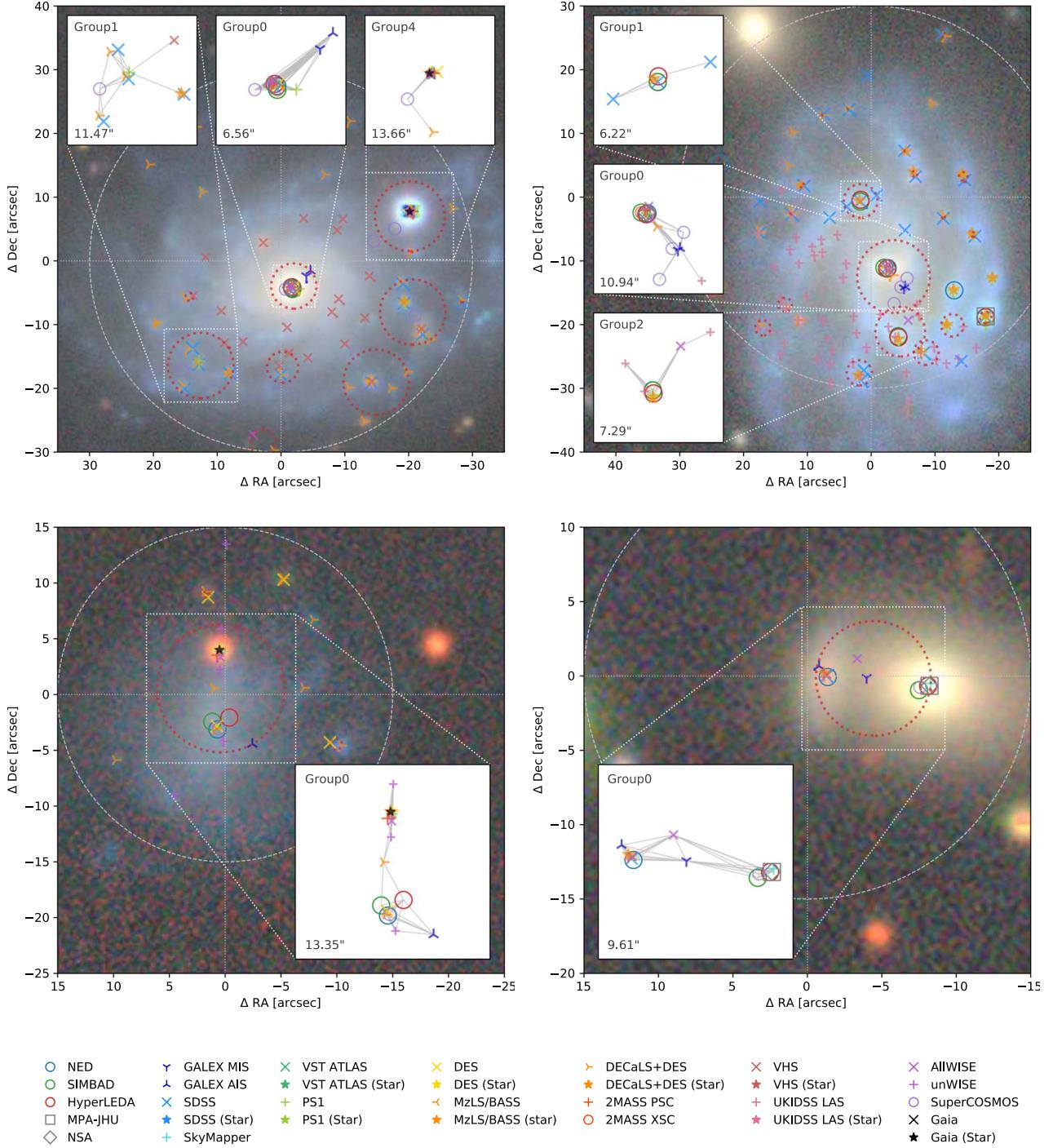
Figure 19. Top: cross-matching catalog sources in fields with nearby large galaxies. Two examples here are PS16fdq (left) and SNhunt124 (right). The spiral arms, bright clumps, and extended outskirts of nearby large galaxies are often broken into individual sources when the object detection algorithm is not tailored toward such situations. Nearby large galaxies with low surface brightness, flat light profile, or irregular morphology may also have significant catalog-to-catalog offsets in source coordinates. Consequently, their compiled host properties are subject to incompleteness or mismatching. Even the cross-matched substructures may interfere with the identification of new hosts. Bottom: cross-matching catalog sources with non-optimal astrometric tolerances or matching thresholds. Examples here are SN2017hdn (left) and SDSS-II SN 15822 (right). The cross-matched groups in both cases contain at least two distinct objects, which are *glued* into a single group either due to large scatter of source coordinates or by a few *bridging sources*. Such groups can be excluded by their extreme values of connectivity and mean offset. We illustrate the connectivity of catalog sources in selected groups with zoom-in panels, where the panel sizes are indicated at lower left corners. Red dashed circles indicate cross-matched groups with spanning distances above 1″.

Appendix H
Star/Galaxy Separation and Resolution of Confusion

Before locating the groups corresponding to our known hosts or ranking cross-matched groups to identify the best host candidates, we need to assign source properties to cross-matched groups. The procedure here includes a simple star/galaxy separation, as foreground stars should not be considered as host candidates, as well as the selection of source properties when a group includes multiple sources from the same catalog, i.e., resolving confusion of multiple matched sources.

To accurately identify foreground stars in each field, upon cross-matching catalog sources, we utilize source morphology indicators or star/galaxy separation parameters provided in these catalogs to flag potential stellar sources. More accurately, these indicators or parameters only separate point and extended sources. Some extragalactic objects, such as quasars and cores of extremely compact galaxies, could be classified as point sources. Conversely, some extended sources could be foreground sources like stellar multiplicities or nebulae. Here "stars" and "galaxies" are only used loosely to refer to these two types of sources. We choose rather strict criteria for stellar sources so that the identified subset of sources is pure and genuine galaxies would not be easily flagged as stellar sources. After flagging individual stellar sources in external catalogs, we check if our cross-matched groups contain stellar sources. Without the confusion of multiple sources, when any catalog source in a cross-matched group satisfies the criteria for stellar sources, we flag the entire group as a star.

Even with the optimized astrometric tolerances and matching thresholds, confusion or missing of sources could still occur in some fields, i.e., a certain catalog may contribute more than one valid source to a cross-matched group, or a catalog source that corresponds to the true host is not included in the group. Both cases are penalized by our target function of optimization as described in Appendix F and should have already been minimized. However, the first case, if present, could still lead to mismatching of host properties.

Whenever a cross-matched group contains more than one source from a certain catalog, we keep all sources but select only one of them to represent the measured properties of the cross-matched group within that catalog. For the photometric catalogs, the brightest source is selected as the representative source, and the deepest photometric band of each survey is used to rank confusing sources. For the other catalogs, the one closest to the group centroid, i.e., the average position of sources in this group, is selected to represent the properties of this group. In any case, we prefer non-stellar sources over stellar ones when finding the representatives within confusing sources.

Finally, in the case of a cross-matched group containing confusing sources in certain catalogs, and the group is not yet labeled as a star under stellar criteria for other catalogs without confusion, we select representative sources for catalogs with confusion first, and then label the entire group as a star only if any of these representative sources satisfies our stellar criteria. In other words, when there is a confusion of multiple sources in a catalog with stellar source criteria, and the group is not labeled as a star by other catalogs without confusion, it would only be labeled as a star when all confusing sources satisfy stellar criteria in at least one of the catalogs that has multiple confusing sources.

Appendix I
The Ranking Functions

I.1. Constructing the Training Set

To construct a training set for ranking functions, we choose transients that (1) have name-resolved or as-reported host coordinates in our trusted reference sources, (2) have valid sources in external catalogs cross-matched using host coordinates, and (3) have been visually inspected by us to have acceptable quality, i.e., "OK" cases without further flags,

comments, or alternative hosts (Appendix J.1). We search their transient coordinates again in external catalogs, with the search radius set as if neither host names nor host coordinates are known. We then cross-match the retrieved catalog sources in these fields with the same astrometric tolerances. Query results and cross-matched groups are stored in a separate "training" database, with an identical structure as the main database. Cross-matched groups closest to the known, visually confirmed hosts in the main database are labeled as "true hosts," while other groups in the field are labeled as "other objects." In the absence of a "true host," we exclude the field from the training set. About 0.3% of the "true hosts" have large angular offsets ($\geqslant 3''$) to their counterparts in the main database, which are usually caused by the change of cross-matching thresholds (and hence group members) under different search radii and field centers (Appendix F). We do not use these fields for the training set.

When assembling the training set, we do not include fields in which the true hosts are beyond the search radii. However, we keep track of these excluded fields to estimate their impact on the accuracy afterward.

I.2. Constructing the Feature Sets

To characterize cross-matched groups in each field, we start from the following parameters:

1. The angular distances of group centroids to the transient coordinate ($\delta\theta$, in arcseconds); besides linear angular distances, we also include $\mathrm{arcsinh}(\delta\theta/\mathrm{arcsec})$ and $1/(1 + \delta\theta/\mathrm{arcsec})$ in the list of parameters;
2. The projected physical distances (δr, in kiloparsecs), for events with known redshifts only; similarly, besides linear distances, we also include $\mathrm{arcsinh}(\delta r/15\,\mathrm{kpc})$ and $1/(1 + \delta r/15\,\mathrm{kpc})$ in the list of parameters;
3. Binary flags indicating if the cross-matched group contains source(s) from a specific catalog; every survey or value-added catalog corresponds to one flag;
4. Geometric signatures and quality metrics of cross-matched groups, including the mean offset (d), axes ratio (σ_b/σ_a) and the degree of connectivity (F) of each group; these parameters are discussed in detail in Appendix G;
5. *Stellarity* parameters, including a binary flag indicating if any representative source satisfies the criteria for stellar sources and the fraction of representative sources from catalogs with star-galaxy separation criteria that are likely stars;
6. Field-aware contextual parameters, including the quantile ranks within the field for (1) the number of catalogs matched, (2) the linear distance to transient coordinates, and (3) the connectivity parameter (F).

The list above includes 32 or 35 parameters (3 angular distances, 3 optional projected physical distances, 21 binary flags for external catalogs, 3 geometric and quality metrics, 2 parameters for *stellarity*, and 3 contextual parameters). We choose these parameters because this is a *complete* set of parameters, which is universally available for any cross-matched groups, even in sky areas that are not covered by certain surveys or for hosts that are beyond the sensitivity limits of certain surveys. Also, they do not rely on detailed galaxy properties. These two feature sets, which we name as `Basic8` (32 parameters) and `Basic8_z` (35 parameters, including

redshift-dependent projected distances), lay the groundwork for other input feature sets.

We also extend the "basic" feature sets above using other more detailed source properties. Since optical-NIR magnitudes are the most available kind of source properties in external catalogs, we create the `AnyMag/AnyMag_z` feature sets, which include existing parameters in `Basic8/Basic8_z` feature sets, with one more parameter, the optical-NIR magnitude of the group. This optical-NIR magnitude is selected in DESI-LS, PS1, DES, SDSS, VST-ATLAS, and SkyMapper catalogs, in the preferred order here. We use r-, i-, or z-band extended source magnitude in these catalogs, whichever is available first. Due to their dependence on source optical-NIR magnitudes, the feature sets also limit the choice of possible hosts to groups with any of these catalogs matched. Other groups would not be ranked and selected as the host candidate.

Furthermore, since PS1 and DESI-LS are the two most widely available external catalogs for our host candidates, we construct feature sets based on their measured source properties. The `PS1/PS1_z`feature sets, which are also based on the existing `Basic8/Basic8_z`feature sets, have the following additional parameters:

1. The Kron magnitudes (Kron 1980) of PS1 sources, in the r, i, or z bands, whichever is properly measured first in this preferred order;
2. The Kron radii in the same band as the magnitude above, which measure the angular sizes of objects using the first radial moments of their surface brightness profiles;
3. The ratio of $\delta\theta$ to the Kron radius above, which is analogous to d_{DLR} used in the DLR method, but without correcting for the axis ratio or inclination of galaxies.

Since the Kron radius here includes seeing contribution, the sources sizes are overestimated, especially for compact sources. This may reduce the distinguishing power of the ranking function in some situations.

We also constructed `LS/LS_z`feature sets, which include parameters in the existing `Basic8/Basic8_z` feature sets, with the following additional ones:

1. The profile-fitting magnitude of DESI-LS sources, preferably in the r band, but the z or $g$bands are also used in the absence of r-band data;
2. The measured half-light radius R_{50} and the calculated R_{90} (radius enclosing 90% of total flux) under the best-fitting light profile; for point sources, both radii are set to $0.1''$;
3. d_{DLR} calculated using R_{50} and R_{90}, considering the position angle and axis ratio of galaxies, following the definition in Sullivan et al. (2006);
4. The photometric redshifts z_{ph} of LS sources from Zhou et al. (2021), the squared difference of transient redshift and z_{ph}, and, finally,

$$\sqrt{(z - z_{\mathrm{ph}})^2 / \sigma_{z_{\mathrm{ph}}}^2 + 1} - 1 \qquad (1)$$

where $\sigma_{z_{\mathrm{ph}}}$ is the uncertainty of z_{ph} estimate.

Note that R_{90} of galaxies are derived from R_{50} using the ratio of R_{90} to R_{50} under the best-fitting light profile. For the exponential profile, we use $R_{90}/R_{50} \simeq 3.52$; while for de Vaucouleurs' profile, we use $R_{90}/R_{50} \simeq 5.58$. The position angles and axis ratios are converted from ellipticities in complex representation. When fitted with composite models

(exponential plus de Vaucouleurs'), the complex ellipticity, R_{50}, and R_{90} are the linear combination of the two components, weighted by their fractions in the total flux. The last three z_{ph}-related parameters are only used for transients with known redshifts (i.e., only in `LS_z` feature set). The function forms are chosen to separate true hosts (with consistent redshift) and other objects (with possible inconsistent redshift) easily using simple linear classifiers.

For testing and comparison, we also assemble feature sets that only use information relevant to PS1 or LS catalogs. We derived `PS1sub/PS1sub_z` feature sets from the existing `PS1/PS1_z` feature sets, in which parameters related to multi-catalog cross-matching and other cross-matched catalogs are ignored. Similarly, we assemble `LSsub/LSsub_z` feature sets using only the parameters available from LS sources alone, including z_{ph}-related parameters.

Feature sets using more detailed source properties in external catalogs are only available for some host candidates. We plot the fraction of availability in the bottom subpanels of Figure 20.

I.3. Testing the Classifiers

Having the training set ready, we trained several commonly used binary classifiers, including Logistic Regression, SVM (with regularization parameter $C = 0.25$), RF (with a maximal depth of 5), AdaBoost, SGD, and MLP, using the default hyper-parameters in `scikit-learn` (v0.22.1) unless noted.

We train ranking functions with every possible combination of classifiers and input feature sets. When using feature sets with transient redshift-dependent parameters (i.e., feature sets with "_z" suffix), we limit our training and testing to transients with known redshifts; otherwise, the entire training set is used. If the feature set relies on source properties in a particular external catalog, we further limit the training set to transients in the coverage of that survey, whose "true hosts" have a source in that catalog matched. Even for transients in the coverage of the survey, the "true hosts" may not have the required parameter available. We show the fraction of availability in Figure 20.

Furthermore, we create two special ranking functions for comparison; one selects the nearest group to the transient coordinate ("Naive Nearest"), and the other selects the group with the minimal d_{DLR} (based on R_{50}) in the field ("Simple DLR"). They directly use $-\delta\theta$ or d_{DLR} as the ranking scores, where the negative signs select groups with the lowest $\delta\theta$ or d_{DLR}. The latter ranking function is only limited to transients and host candidates with DESI-LS parameters.

We assess the performance of these trained ranking functions by the chance that the true host has the highest-ranking score in its field, which we loosely refer to as the accuracy. The accuracy is averaged over a standard tenfold cross-validation (Section 3.6.1), where training and test data sets are split based on fields rather than individual cross-matched groups. Besides the overall accuracy, we also analyzed the dependence of the accuracy on transient redshift, transient-host angular offset, host optical-NIR magnitude, and transient type. The results are estimated in bins of equal sample size (except for transient type), and the accuracy in each bin is averaged over tenfold tests. The results are summarized in Figure 20.

We summarize the overall performance of ranking functions in Table 9. With the "basic" feature sets, the accuracy is above 96% for all classifiers, in which the Logistic Regression
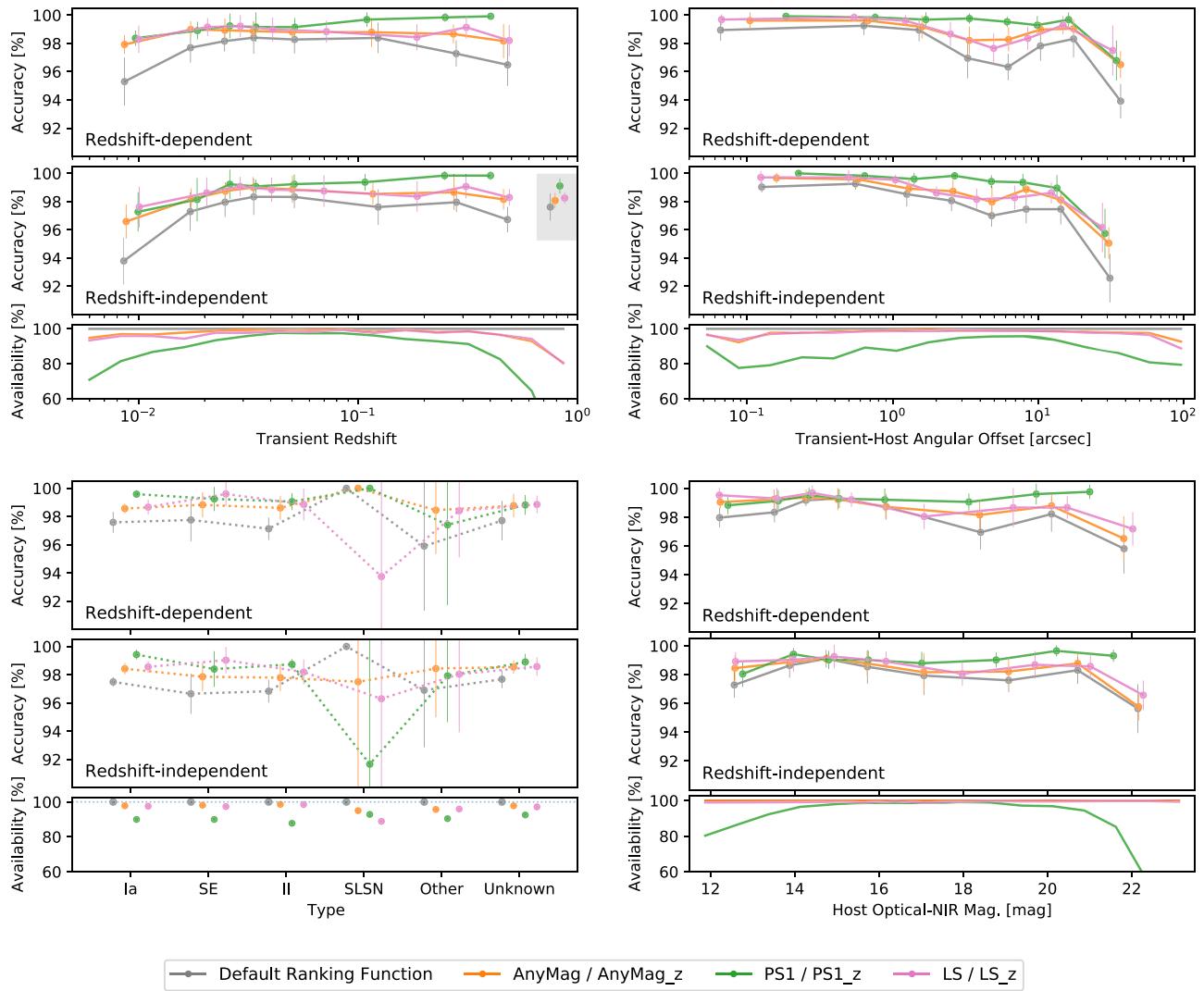
Figure 20. Accuracy of the ranking function estimated using the training sample, averaged over standard tenfold cross-validation. We show the dependence of accuracy on transient redshift, transient-host angular offset, transient type, and host optical-NIR magnitudes in four separate panels. Except for the panel of transient type, we divide the sample into equal-sized bins over the axis of interest and plot the per-bin accuracy at the median value of each bin. Each curve shows the accuracy of a feature set. For clarity, the accuracy curves of redshift-dependent feature sets (with "_z" suffix) and redshift-independent feature sets are grouped into two separate subpanels. Some feature sets rely on certain source properties that are not always available for those known hosts. We estimate the accuracy using the subset of known hosts with the relevant features available and outline the fraction of true hosts with relevant parameters available in the bottom subpanels. Specifically, in the panel of transient redshift, we plot the accuracy of events without redshift in the gray shaded area, if applicable.

classifier achieves $97.5\% \pm 0.6\%$ accuracy using the transient redshift-related parameters or $97.3\% \pm 3\%$ without using redshift-related parameters. Therefore, we use Logistic Regression and "basic" feature sets as our default ranking functions. When transient redshift is available, we use the redshift-dependent feature set (`Basic_z`); otherwise, we use the redshift-independent version (`Basic`).

We calculate the ranking scores of other feature sets in our released data set, but we do not use the scores to rank cross-matched groups. However, it is worth comparing their performances with the default ranking functions. We see a clear increase in accuracy when source properties are included (Table 9, Figure 20). For example, adding any optical-NIR magnitude to the existing redshift-dependent "basic" feature set improves the accuracy to 98.6%. The inclusion of source properties in PS1 and LS can even push the accuracy further to 99.3% and 98.8%, respectively. For their redshift-independent versions, the accuracy is similar, if not marginally lower. Even

a percent-level increase of accuracy in this range indicates a significant reduction of failure rates compared to the default ranking function. Therefore, the inclusion of detailed source properties brings a substantial performance improvement.

The improvements in accuracy, however, come with the cost that true hosts without the required extra parameters measured are ignored. For example, many true hosts do not have PS1-relevant parameters measured, especially for fainter or brighter hosts close to the sensitivity and saturation limits (Figure 20, "Availability" subpanels). Even PS1-based feature sets outperform our default ranking functions ("basic" feature sets); this may not benefit host matching in reality. On the contrary, feature sets with DESI-LS-relevant parameters are available for most of the true hosts. *Assuming* those true hosts, if above the sensitivity of DESI-LS, will be detected and cataloged, the improved accuracy would facilitate host matching of transient surveys.

Table 9
Average Accuracy of Ranking Functions from Cross-validation

Name	Basic(_z)	AnyMag(_z)	PS1(_z)	LS(_z)	PS1sub(_z)	LSsub(_z)
	With redshift-dependent parameters					
Logistic	97.5 ± 0.6	98.6 ± 0.3	99.3 ± 0.1	98.8 ± 0.4	90.7 ± 1.1	97.6 ± 0.4
SVM	97.4 ± 0.7	98.6 ± 0.3	99.3 ± 0.1	98.7 ± 0.4	90.3 ± 1.1	97.5 ± 0.4
RF	97.3 ± 0.6	98.4 ± 0.3	99.3 ± 0.1	98.6 ± 0.5	87.6 ± 1.5	96.1 ± 0.5
AdaBoost	97.0 ± 0.4	98.4 ± 0.2	99.0 ± 0.4	98.6 ± 0.3	89.8 ± 1.0	97.1 ± 0.4
SGD	96.9 ± 0.6	98.1 ± 0.2	99.2 ± 0.2	98.5 ± 0.4	91.0 ± 1.6	96.6 ± 0.4
MLP	96.8 ± 0.6	97.9 ± 0.3	99.1 ± 0.2	98.3 ± 0.6	89.9 ± 1.3	94.5 ± 0.8
	Only redshift-independent parameters					
Logistic	97.3 ± 0.3	98.3 ± 0.3	99.0 ± 0.3	98.5 ± 0.3	89.0 ± 0.8	95.7 ± 0.5
SVM	97.4 ± 0.4	98.2 ± 0.3	99.0 ± 0.3	98.5 ± 0.3	88.2 ± 0.7	95.6 ± 0.6
RF	97.1 ± 0.3	98.1 ± 0.4	99.0 ± 0.3	98.3 ± 0.3	88.5 ± 0.8	95.0 ± 0.5
AdaBoost	96.7 ± 0.4	98.0 ± 0.3	98.9 ± 0.3	98.3 ± 0.3	90.4 ± 0.8	95.6 ± 0.4
SGD	96.9 ± 0.5	97.8 ± 0.4	99.0 ± 0.3	98.2 ± 0.3	88.7 ± 0.7	94.8 ± 0.6
MLP	96.4 ± 0.4	97.6 ± 0.4	99.0 ± 0.3	97.9 ± 0.5	89.1 ± 0.8	93.6 ± 0.8
	Other methods					
Naive Nearest	53.2 ± 1.0	⋯	⋯	⋯	⋯	⋯
Simple DLR	⋯	⋯	⋯	⋯	⋯	89.8 ± 0.4

Note. Each row represents the accuracy of a classifier using the input feature set indicated in each column. The upper and lower parts show the results for redshift-dependent feature sets (with "_z" suffix) and redshift-independent feature sets. The last two rows show the accuracy of two special ranking functions for comparison.

We also notice that when using DESI-LS-relevant parameters alone (LSsub/LSsub_z), one can reach comparable accuracy to our default ranking functions (i.e., "basic" feature sets). Also, LSsub/LSsub_z feature sets have a reasonable fraction of availability in the survey footprint (Figure 20), clearly higher than the coverage of PS1-relevant parameters. This implies that the training framework would also work for single-catalog host matching if the catalog and input parameters are properly chosen. On the contrary, PS1-relevant parameters improve the accuracy of the default feature sets, but using these parameters alone leads to low accuracy of around 90%.

Finally, ranking cross-matched groups by d_{DLR}, the accuracy is about 90%, lower than the DLR-only accuracy in Gagliano et al. (2021), and close to the DLR-only mock sample performance of Gupta et al. (2016). Choosing the nearest group to the transient position as the host, only about half of the true hosts are successfully recovered. The ranking function method we present here significantly outperforms these methods.

Appendix J
Visual Inspection of Cross-identified Host Galaxies

We rely on existing name-resolved or as-reported host coordinates to create training sets for ranking functions, where the quality of input data may affect the training results. Meanwhile, the trained ranking functions, although achieving good accuracy in known transient-host pairs, have not been tested for performance on newly identified hosts. A comprehensive visual inspection of cross-identified hosts, in either case, can serve as a subjective yet independent test of data quality.

To inspect the cross-identified hosts, we obtained image cutouts of their fields, including DESI-LS grz color composite images from the Sky Viewer[52] and PS1 gri color composite

and g-band FITS images from the Hierarchical Progressive Survey (HiPS) data sets hosted at Centre de Données astronomiques de Strasbourg (CDS),[53] as well as DSS2 color composite and red-band FITS images from the HiPS data sets at CDS. Image cutouts are centered at the queried coordinates (i.e., known host coordinates or transient coordinates). The box size is set to 2.5 times the queried radius, rounded up to the nearest multiplicity of 15. Preferably we use DESI-LS images for visual inspection for their better sensitivity over PS1 images. DSS2 images, on the contrary, are only used when neither DESI-LS images nor PS1 images are available due to their lower sensitivities and resolutions compared to modern digital sky surveys.

We developed specialized software to facilitate the inspection, which overlays symbols and markers of transient coordinates, cross-matched groups, possible stellar objects, and projected distance scales (5, 10, and 20 kpc) onto the background image. We indicate the cross-matched catalogs of each group using color bands on these symbols. Furthermore, for groups with DESI-LS sources cross-matched, we also use upward or downward arrows to indicate if transient redshift is below or above the 95% confidence interval of their photometric redshift estimates in Zhou et al. (2021).

We conducted the visual inspection separately for the samples of known hosts and new hosts. We summarize the workflow in Figure 21. Below we describe the detailed procedure and the results.

J.1. Inspection of Known Hosts

We group cross-identified hosts with known coordinates into three cases: "OK," "Failed," or "Unclear." We focus on whether the known host coordinate successfully matches the galaxy it indicates.

First, we check the image cutout and cross-matched groups to find the galaxy indicated by the host coordinate. We consider
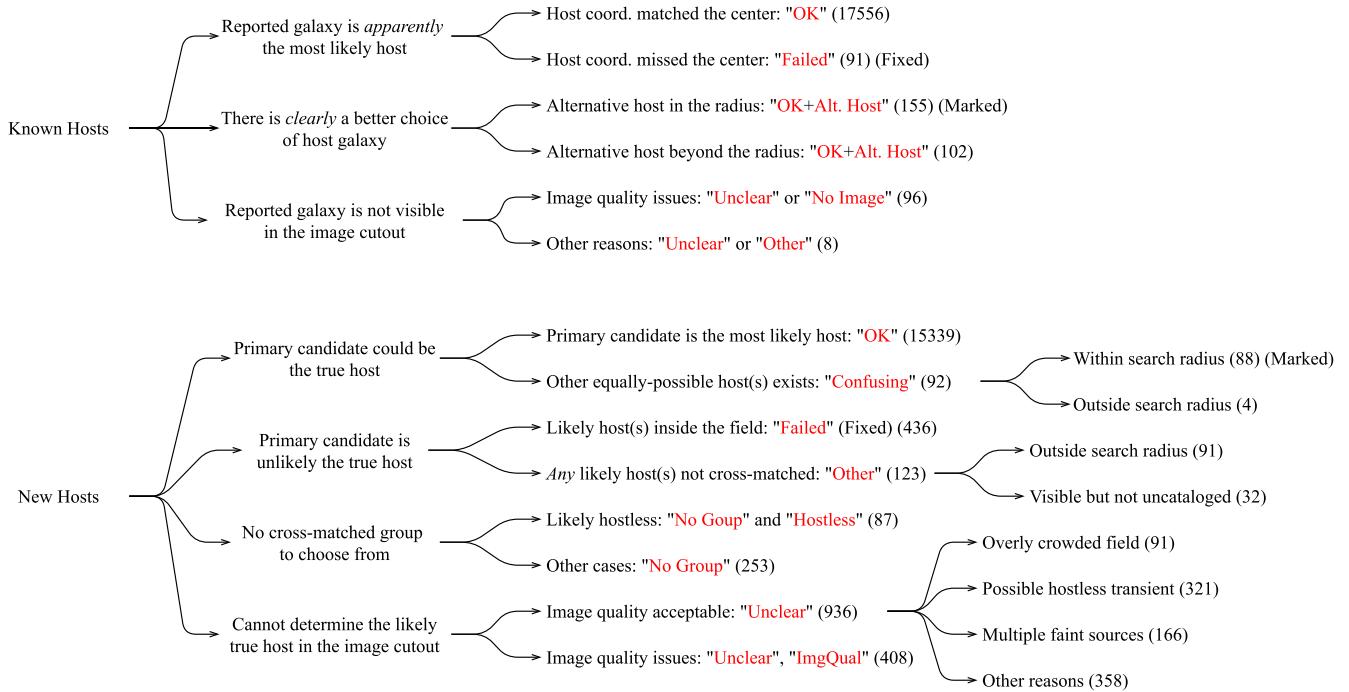
Figure 21. The decision tree of visual inspection for transients with known and properly cross-matched hosts (Appendix J.1) and transients with newly identified hosts (Appendix J.2). For transients with known hosts, we assign "OK" or "Failed" based on the results of cross-matching. If there is clearly a better host than the reported one, we also flag the case and mark the "alternative host" in the image. If we cannot determine if the indicated host is successfully cross-matched and if any better host exists, we consider this an "Unclear" case and set flags for the detailed reasons. For events with new hosts, if the algorithm chose the unique best host in the images, we consider the case "OK." If the algorithm chose one likely host while there are other equally possible hosts, we consider this a "Confusing" case. If the algorithm missed the best host in the radius, we consider the case a "Failed" one and then manually reassign the host if we can. However, if the best host is outside the radius, we group the case under "Other." If no group has been cross-matched near the event, the case falls into "No Group." Finally, similar to the case of known hosts, if we cannot determine the best host from the image, we consider the case "Unclear" with appropriate flags indicating the detailed reason.

the case "OK" if the known host coordinate matched the core of the galaxy or the most prominent *main component* of an irregular galaxy. If the known host coordinate failed to match the core or the main component of the galaxy, we consider this a "Failed" case. Such a situation occurs when the reported coordinate has large offsets from the actual central component cross-matched from multiple external catalogs (e.g., irregular galaxies, nearby large galaxies). We then mark the correct group manually. Host properties, in this case, will be updated with our manual correction.

Generally, we do not judge the correctness of reported hosts during this inspection. When there are more than one equally possible host as we see, and the input host coordinate points to any of them, we accept the galaxy as the true host. Occasionally, the known host coordinate points to a galaxy, but there is an *obviously* better choice of the host, not just equally possible, from our visual inspection. We consider the case "OK" if the known host coordinate matched the galaxy it indicates, but at the same time, we set an "alternative host" flag to indicate that we prefer another host. We manually mark the preferred host when it has been cross-matched in external catalogs. If the preferred host lies beyond the search radius and is therefore not cross-matched, we set an "alternative host beyond radius" flag to indicate the situation and then mark the position of the alternative host (R.A., decl. offsets from mouse click) if its core is visible.

In some cases, we cannot distinguish the indicated galaxy in the image, and hence the robustness of cross-matching and the existence (or not) of any better host. We consider these cases as "Unclear." Fields can be marked "Unclear" for various reasons.

Still, image sensitivity or resolution issues, for which we set the "image quality issue" flag, is a primary cause. High-redshift hosts outside the coverage of DESI-LS and PS1 sometimes fall in this category.

Furthermore, if the transient position is likely inaccurate, due to either an equinox conversion issue of historical coordinates or a possible mistake in upstream data sources, we set an additional "possible error in metadata" flag. If matched by name-resolved coordinates, their host properties remain usable, but transient-host offsets could be wrong. Other uncommon cases are grouped under "Other" with appropriate descriptive flags.

J.2. Inspection of Newly Identified Hosts

Newly identified hosts are grouped into four major cases: OK, Failed, Confusing, or Unclear. Cases including "Other" and "No Group" are also used when applicable.

First, we visually identify the galaxy that appears to be the most likely host in the field. Then we reveal the host identified by the default ranking function. Revealing the choice of the algorithm later may reduce the confirmation bias in this inspection. If the default ranking function chose the most likely host as we see in the image, we consider the case "OK." Similar to the inspection of known hosts, we require that the center or the *main component* of the galaxy is selected. Otherwise, this would be considered as a "Failed" case, with manual reassignment of the host.

Occasionally, there could be more than one visually identified, equally possible host (for example, the transient is
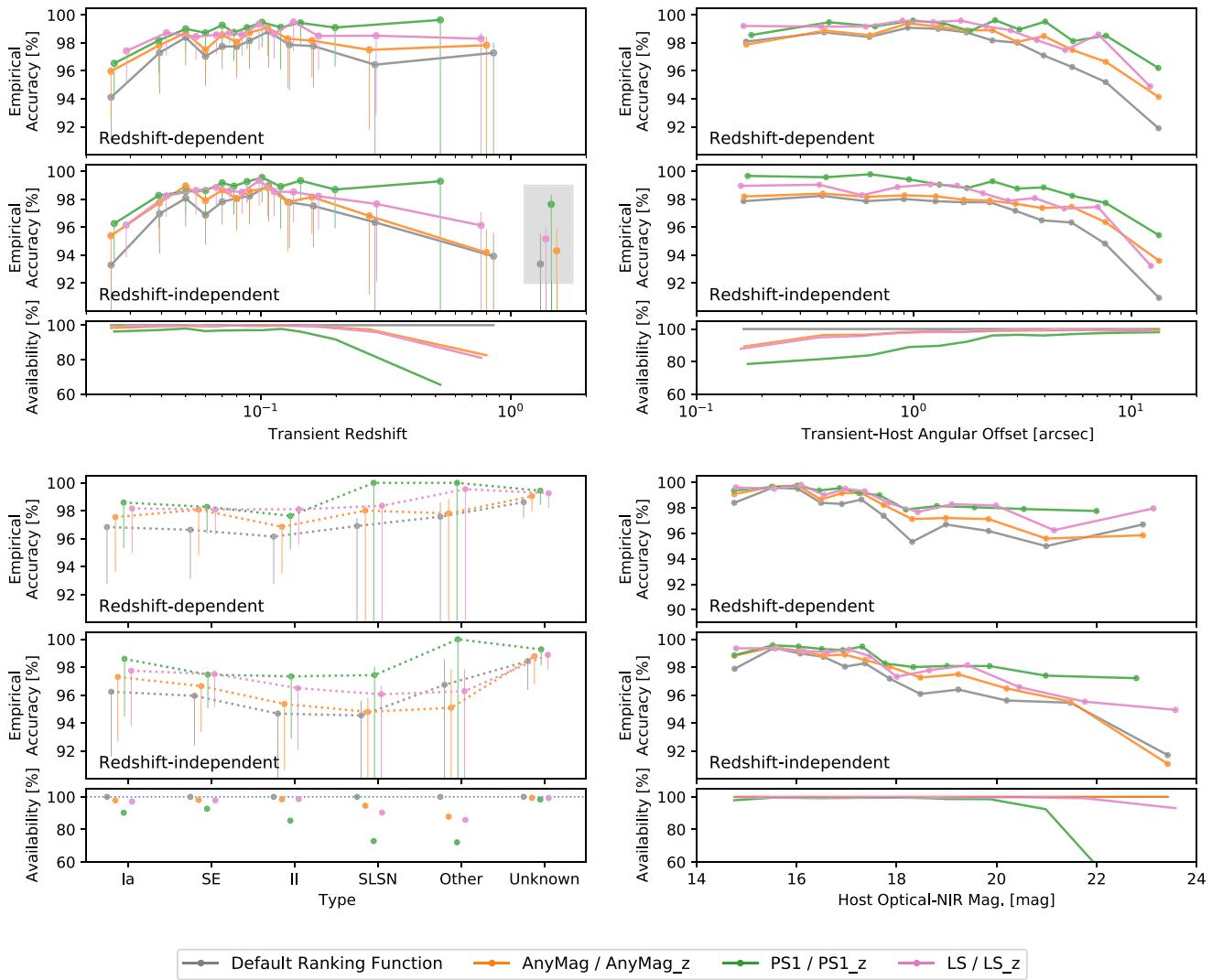
Figure 22. The empirical accuracy of new hosts from visual inspection. Here the accuracy is the fraction of correctly identified hosts in cases where the host, either correctly identified or not, is clearly visible and unambiguous. We show the dependence of accuracy on transient redshift, transient-host angular distance, transient type, and host optical-NIR magnitude in four separate panels. Except for the panel of transient type, we divide the sample into bins of equal sample size over the axis of interest and plot the per-bin accuracy and mean value there. Each colored accuracy curve represents the performance of an input feature set using the default Logistic Regression classifier. Redshift-dependent and redshift-independent feature sets are grouped into two subpanels for clarity. We calculate the accuracy for feature sets using detailed source properties with the subset of events whose best hosts (either correctly identified or not) have the catalogs cross-matched. Since these properties are not universally available, we further plot the fraction of best hosts with the required features available in the bottom subpanels to indicate the possible loss of accuracy. Furthermore, as transient type and redshift are independent of the best host, we also estimated the uncertainty of accuracy contributed by indecisive cases ("Unclear" and "Confusing" cases). The optimistic and pessimistic limits of accuracy are indicated as "error bars" of these points. Similar to Figure 7, for events without redshift, we plot their accuracy in the rightmost gray shaded area of the redshift panel.

in a group or cluster environment but not clearly associated with a specific member), and the default ranking function could choose *any* of them, so we consider the case a "Confusing" one and then manually mark other possible hosts. If any possible host is beyond the search radius and is, therefore, not cross-matched into a group, we set a "host beyond radius" flag and then mark the galaxy (R.A., decl. offsets) in the image when possible.

If the ranking function missed the most likely host, or rarely, all possible hosts when there are multiple, we consider this a "Failed" case and then manually mark the most likely host(s). This also includes the situation in which the ranking function chose some substructure of the most likely host rather than the main component. However, if the most likely host lies beyond

the search radius, or if it is marginally visible but not cataloged and cross-matched, we group the case into "Other" instead of "Failed," because this is a failure in the source accessing and cross-matching process, rather than the ranking function itself.

When there is no cross-matched group for the ranking function to choose from, a possible situation for high-redshift transients such as GRBs, we mark the case with "No Group." If the sensitivity of the background image is fairly good (DESI-LS image, or PS1 image without stacking artifacts) and there is no possible host near the transient position, we further flag the case as "hostless."

For any reason, if the most likely host is not clearly visible in the image, we consider the case as "Unclear." This includes, but is not limited to, the following situations as we flagged: (1)

49

Table 10
Empirical Accuracy of Ranking Functions

Name	Basic(_z)	AnyMag(_z)	PS1(_z)	LS(_z)
	With redshift-dependent parameters			
Accuracy	97.4	98.0	98.8	98.6
Upper Limit	97.5	98.1	98.9	98.7
Lower Limit	92.5	93.5	95.6	95.0
	Only redshift-independent parameters			
Accuracy	96.8	97.5	98.7	97.9
Upper Limit	97.0	97.6	98.7	98.0
Lower Limit	88.9	90.7	93.9	93.2

Note. The empirical accuracy is estimated for the default ranking functions (Appendix I). We use the fraction of clearly visible and unambiguous hosts that have been successfully recovered as the accuracy here. The upper and lower limits of accuracy due to indecisive cases are also quoted.

the transient is likely hostless ("hostless") in images of fairly good sensitivity (see also 3.7); (2) the field is overly crowded ("crowded," e.g., low-galactic latitude regions), hindering the visual identification of the host; (3) there are multiple faint sources, likely background galaxies, in the vicinity ("multiple faint sources"), in which no one appears to be a better host than others, yet possible association with the transient cannot be fully excluded; (4) low-resolution, low-sensitivity, or corrupted images ("image quality issue"). When possible, we also comment with the detailed reason if the case does not fit in these typical reasons.

J.3. Empirical Accuracy from Visual Inspection

We summarize the results of visual inspection in Table 10, Figure 7, and Figure 22.

We use *empirical* accuracy (Section 3.6.2) to evaluate the performance of ranking functions in new hosts. The empirical accuracy here refers to the fraction of clearly visible and unambiguous hosts that the ranking function has successfully recovered. Assuming that the best host is inside the search radius, our default ranking functions achieve 97.4% when redshift-dependent parameters are used for ranking, and 96.8% when only redshift-independent parameters are used (Table 10).

There are a considerable fraction of indecisive cases ("Unclear," "Confusing") in which the best host is unidentifiable, ambiguous, or even absent. The empirical accuracy must take this potential source of uncertainty into account. We quote the range of empirical accuracy by making assumptions about the performance of ranking functions in these indecisive cases. Assuming that these indecisive cases are all successfully identified, we may estimate an optimistic upper limit of accuracy; on the other hand, in the improbable worst situation where all their hosts are misidentified, we can estimate a pessimistic lower limit of accuracy. The upper and lower limits of accuracy are also listed in Table 10. Considering the uncertainty due to these indecisive cases, the accuracy of the redshift-dependent default ranking function ranges from an optimistic case of 97.5% to an unlikely pessimistic case of 92.5%. For the redshift-independent default ranking function, the limits are 97.0% and 88.9%, respectively. These indecisive cases are more relevant to high-redshift events (Figure 7, Figure 22) and are less of a concern for mid- and low-redshift transients.

We checked the accuracy of the default ranking functions, while the accuracy of other ranking functions can be derived once the visually identified most likely host is known in each field. We noticed that the inclusion of detailed source properties in PS1 or DESI-LS catalogs could clearly improve the results, as we pointed out earlier in Section I.3. The improved accuracy has similar dependence on key transient and host parameters as the default ranking functions due to the overlap in their input parameters or training sets. Any percent-level improvement of accuracy in this range implies a significant reduction of misassociation rates compared to the default ranking functions. However, the improvement comes with the cost that the algorithm could ignore some true hosts if they do not have the required parameters measured. For example, PS1/PS1_z feature sets, which rely on source properties in PS1 catalog, have close to 99% accuracy. However, the sensitivity of the PS1 catalog may not detect some fainter hosts, which leads to a decreased feature availability toward higher redshift. Higher-redshift transients will likely not benefit from the improved accuracy here.

Appendix K
Confidence Scores of Host Candidates

To evaluate the reliability of individual newly identified hosts, besides visual inspection, we also derived a separate set of numeric metrics for each cross-matched group, which we refer to as the *confidence score*.

The confidence score is a metric for outliers or anomalies, which distinguish likely host galaxies from the bulk of nearby non-host objects by comparing their ranking scores. We create a reference sample of nearby non-host objects by generating 16,384 mock transients with randomized positions (i.e., without association with any galaxies). Galaxies and stars near these mock transients are always nearby non-host objects rather than their "true hosts." The highest-ranking "host candidates" of these mock transients then outline the possible range of ranking scores, represented as a baseline distribution, that a nearby non-host object may reach by chance under the best cases. For a true transient, if the ranking score of a host candidate is at the higher-value tail of the baseline distribution, then this candidate is likely an outlier among nearby non-host objects, and we can be more confident about this new candidate. Conversely, if the ranking score of this host candidate is inside or even below the range of the baseline distribution, then this candidate could be a nearby non-host object that ranked high in the field simply by chance. When no host candidate in a field achieves an outstanding ranking score compared to the baseline distribution, the transient itself could be a hostless one.

We use the same source searching and cross-matching workflow for these mock transients under a constant search radius of $45''$, where the ranking scores are calculated using the same trained ranking functions as we used for those true transients. For each feature set and classifier combination, we construct the baseline distribution using the highest-ranking score of each mock transient. The confidence score is defined as the percentile rank of a new ranking score in the baseline distribution of the same feature set and classifier. A value closer to 100% indicates a more outstanding and possibly more reliable candidate. Clearly, if the ranking scores depend on the transient redshift and catalog coverage, then the baseline distributions should also be adjusted for these factors. To compensate the redshift dependence of some feature sets (with
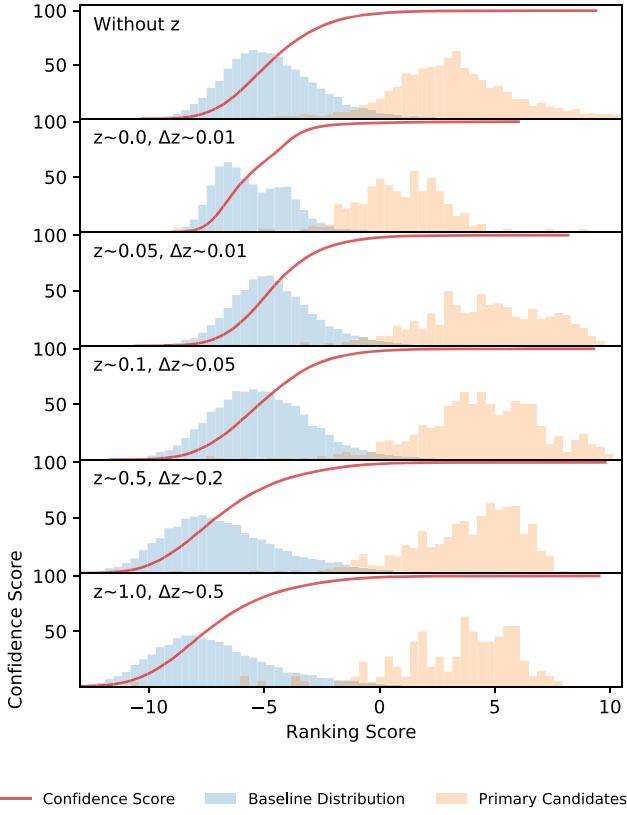
Figure 23. Examples of estimating confidence scores from baseline distributions. We compare the ranking scores of new host candidates (orange) to a baseline distribution (blue), i.e., the highest-ranking scores that nearby *non-host* objects can achieve by chance in a randomized mock transient sample. We use the cumulative percentage curve of the baseline distribution (red) to quote the ranking scores of new host candidates. True hosts, ideally, should have ranking scores at the positive tail of the baseline distribution. Their confidence scores are closer to 100%. Here we show the case for the redshift-independent default ranking function (top panel) and the redshift-dependent default ranking function in several redshift intervals (centered at z, with bin width Δz).

"_z" suffix), we generate a series of baseline distributions by setting these mock transients at some fixed redshifts on a predefined axis so that we can choose the nearest baseline distribution on the axis later. To ensure similar catalog coverage of mock transients as our existing true transients, we generate mock transients following the density of true transients on the celestial sphere, represented in a `healpix` grid smoothed by a Gaussian kernel of FWHM $\sim$1 degree. Furthermore, we also calculate local confidence scores using only the nearest 1024 mock transients, rather than the entire sample of 16384 mock transients, for baseline distributions. Despite being noisy due to smaller sample sizes, this localized version of confidence scores may better offset the dependence of baseline distributions on the coverage of external catalogs.

In Figure 23, we show examples of baseline distributions, ranking scores of newly identified hosts, and their derived confidence scores with our default ranking functions. For the redshift-dependent ranking function, we also show the results in several redshift bins. The baseline distributions are clearly redshift-dependent. Toward higher redshift, although angular distance-related parameters may still distinguish true hosts from other objects, other properties of hosts become similar to those ubiquitous galaxies in the field. However, the ranking
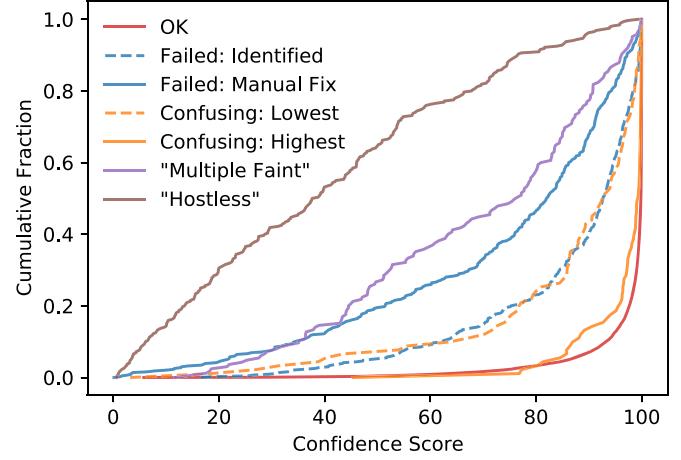

Figure 24. The cumulative distribution of confidence scores for some host candidates identified during our visual inspection of new hosts (Appendix J). Successfully identified hosts ("OK") usually have very high confidence scores. However, higher confidence scores do not guarantee more reliable hosts; misidentified hosts ("Failed: Identified") still have higher confidence scores than the manually reassigned hosts in the same field ("Failed: Manual Fix"). When there are multiple possible hosts, the other equally possible host ("Confusing: Lowest") also has high confidence scores, although not as high as the *best* host selected by the algorithm in the same field ("Confusing: Highest"). We see much lower confidence scores for the selected "host" when the vicinity contains multiple faint sources, yet no one stands out to be a clear and unambiguous host ("Multiple Faint"). Finally, the *best candidate* of possible hostless transients ("Hostless") has much lower confidence scores than other candidates. They basically follow the baseline distribution. Results here are global confidence scores using the default ranking function in each field.

scores of true hosts remain well above the ranges of baseline distributions, and the behavior of confidence scores remains stable. In Figure 24, we show the cumulative distributions of confidence scores for a few kinds of host candidates identified during our visual inspection. Successfully identified new hosts ("OK"), as expected, have confidence scores close to 100%. Misidentified hosts ("Failed: Identified"), as a comparison, have lower confidence scores than successful cases. Higher confidence scores usually, but not always, mean better candidates. The manually reassigned hosts ("Failed: Manual Fix") of "Failed" cases have lower confidence scores than galaxies misidentified in the same field ("Failed: Identified"). Similarly, in "Confusing" cases, the equally possible hosts ("Confusing: Lowest") have lower confidence scores than the ones chosen by the ranking functions in the same field ("Confusing: Highest"). There are "Unclear" cases with multiple faint sources in the vicinity of transient coordinates, in which no one appears to be a clearly visible and unambiguous host, while any transient-host association cannot be fully excluded. Their highest-ranking "hosts" ("Multiple Faint") have much lower confidence scores, although the distribution overlaps with manually reassigned hosts ("Failed: Manual Fix"). Finally, the primary candidates of possible hostless transients ("Hostless") have very low confidence scores, following the baseline distributions.

As a summary, the confidence score here serves as a quantitative metric for the reliability of host candidates, which clearly differentiates robust hosts from those less reliable ones. However, the confidence score is based on the same training framework as our ranking algorithm, which does *not* directly indicate the correctness of individual host identifications. Cross-checking with visual inspection is encouraged.

Appendix L
Database Schema

Here we show the schema of host collection using an example with comments. The output is abbreviated for clarity.

```
# transient info
"event_name": "SN2011fe", # preferred name
"event_alias": ["SN2011fe", "PTF11kly", ...], # other alternative names
"event_ra": 210.7737, # RA of event (degree)
"event_dec": 54.2736, # Dec of event (degree)
"event_z": 0.0008, # reported transient or host redshift
"event_dxy": [-59.8553, -270.9878], # event position in the field center (arcsec)
"revised_type": ["_SN", "Ia", ...], # curated type labels

# as-reported host info
"reported_host": {
 "name": ["M101", "NGC 5457"], # reported host name(s)
 "ra": 210.8042, # reported host RA (degree)
 "dec": 54.3492, # reported host Dec (degree)
},
```

```
# summary of host name resolving,
"resolved_host_coord" : {
 "valid_names" : ["M101", "NGC 5457"], # list of valid host names
 "resolved_names" : { # resolved host coord. of valid names
 # ** key: input host name; value: results of name resolving
 "M101" : {
 # selected coord. for this name
 "vac" : "NED", # selected reference for host coord
 "ra" : "14:03:12.5448", # resolved RA
 "dec" : "+54:20:56.22", # resolved Dec
 "ra_deg" : 210.8022, # resolved RA (degree)
 "dec_deg" : 54.3489 # resolved Dec (degree)
 "src" : "MESSIER 101/PREC:1.20e-01,1.20e-01", # resolved name and coord. precision

 "SIMBAD" : { # results from SIMBAD
 "src" : "M 101/QUAL:C", "ra" : "14:03:12.583", "dec" : "+54:20:55.50",
 "ra_deg" : 210.8024, "dec_deg" : 54.3487, "raw_rec" : [...] # raw results in SIMBAD
 },
 "NED" : {...}, # results from NED
 },
 "NGC 5457": {...},
 ...
}
```

```
# consistency check of name-resolved coord.
"resolved_coord_inconsistent" : False, # flag for any inconsistent pairs
"inconsistent_pairs" : [...], # list of inconsistent pairs, if any
# ** list of tuple: (name, the other host name, distance in arcsecond)

# best name-resolved coord., same structure as values under "resolved_names"
"resolved_coord_best" : {
"vac" : "...", "ra" : "14:03:12.5448", "dec" : "+54:20:56.22",
"ra_deg" : 210.8022, "dec_deg" : 54.3489, "src" : "MESSIER 101/PREC:1.20e-01,1.20e-01",
"SIMBAD" : {...}, "NED" : {...},
}
}
```

```
# flags for input host coord.
"host_coord_known" : True, # host coord known or not
"host_coord_matched" : True, # host coord cross-matched or not
"host_coord_type" : "name_resolved", # type of host coord
# "name_resolved" or "reported_host_coord"

# queried coordinates and projection
"field_coord": {
"ra": 210.8022, # field center RA (degree)
```

(Continued)

```
"dec": 54.3489, # field center Dec (degree)
# field centers at the queried coordinate (host or transient).

"coord_src": "name_resolved", # source of queried coordinate
# "name_resolved", "reported_host_coord", "event_coord"

"coord_comm" : ["NED", "MESSIER 101", ...], # flags for provided coordinates

"radius": 15.0, # search radius (arcsec
"radius_src" : "name_resolved_default", # source of search radius
# "name_resolved_default", "reported_host_coord_default",
# "redshift", "grb_err_circle", "event_coord_default"

# unit vectors in Cartesian frame
"cvec": [-0.5006, -0.2984, 0.8125], # queried coordinate, or field center
"vn_ra": [0.5120, -0.8589, 0.0000], # unit vector in RA direction
"vn_dec": [0.6979, 0.4161, 0.5828], # unit vector in Dec direction
},
```

```
# cross-matched groups.
"N_groups": 2, # Number of cross-matched groups
"groups": [
# ** list of dictionaries, one for each cross-matched group
# confirmed host or primary candidate comes the first, followed by other groups
{
# ** key: catalog name or ancillary info (begin with underscore)
"TwoMASSXSC": { # sources matched in 2MASS XSC catalog (example)
"srcs": [
# ** list of dictionaries, one for each source
# the representative comes the first, followed by others, if any
{
# ** key: name of source property; value: measured value
"2MASX": "14031258+5420555"
"J.ext": 6.517,
"H.ext": 5.805,
...
},
...
],
"_confusion": False # flag for confusion in this catalog
},
"MPAJHU": { # sources matched in MPA-JHU catalog (example)
"srcs": [{...}, ...], # detailed source properties
"_confusion": False,
},
... # other catalogs, and cross-matched sources
```

```
"_xid_flag": "confirmed", # type of cross-matched group
# "confirmed", "primary", "secondary", "event", "other"

# summary of cross-matched sources
"_group_uid" : "2ebb53f01d24345f...", # unique group id from cross-mateched sources
"_includes_queried_coord" : True, # queried coord cross-matched this group or not
"_group_srcid" : {
# ** key: catalog; value: list of source names
"HyperLEDA" : ["NGC5457"],
"TwoMASSXSC" : ["14031258+5420555"],
...
}
"_confusion": True, # flag for confusion in this group
"_confusion_cats": ["LS8pz", "Gaia2"], # list of catalogs with confusion

# source positions
"_avr_radec": [210.8023, 54.3489], # mean RA/Dec of sources (degree)
"_avr_dxy": [0.1064, -0.0938], # mean offsets w.r.t. the field center (arcsec)
```

(Continued)

```
"_std_dxy": [0.3033, 0.3943], # std. dev. of offsets (arcsec)
"_cov_dxy": [[0.0991, 0.0429], [0.0429, 0.1674]], # covariance of offsets (arcsec$^2$)
```

```
# quality metrics for cross-matching
"_shape_r": 0.3332, # Pearson's r
"_shape_q": 0.6455, # q=sigma_b / sigma_a
"_shape_p": 0.2154, # p=(1—q) / (1 + q)
"_shape_e": 0.7637, # "eccentricity"
"_avr_dist": 0.4288, # "average distance" or "mean offset" (arcsec)
"_N_max_conn": 87.0, # maximal possible number of connections
"_N_conn": 82.0, # actual number of connections
"_F_conn": 0.9318, # degree of connectivity

# star-galaxy separation
"_is_stellar": False, # flag, any representative source is a star
"_stellar_srcs": [# stellar sources
# ** list of (catalog, index) pairs; index in the "srcs" list of each catalog.
["LS8n", 2], ...
],
"_stellar_frac_all" : 0.0, # fraction of sources marked as stars
"_stellar_frac_repr" : 0.0, # fraction of representative sources marked as stars
# denominator does not include sources in catalogs without star-galaxy separation metrics

# summary of host candidate ranking
"_rank_score" : { # ranking scores
# ** key: feature set or ancillar info.
"Basic8Z" : {
# ** key: classifier; value: results
"Logistic_v3" : {
"score" : -0.6959, # raw ranking score
"rank" : 0, # rank in the field
"cs_global" : 98.3947, # global confidence score
"cs_local" : 98.3398, # local confidence score
},
"LSVM_v3": {...}, "RF_v3": {...}, # other classifiers and results
"_X" : [...], # raw input parameters
}
"LSZ" : {...}, "PS1Z": {...}, ... # other feature sets and results
```

```
"_default" : { # results from default ranking function
"score" : -0.6959, # raw ranking score
"rank" : 0, # rank in the field
"features" : "Basic8Z", # feature set used
"ranker" : "Logistic_v3", # classifier used
"cs_global" : 98.3947, # global confidence score
"cs_local" : 98.3398, # local confidence score
},
"_cscore_local" : { # ancillary info. for local confidence score
"avr_dist" : 11.5686, # average distance to mock transients (degrees)
"max_dist" : 18.8843, # maximal distance to mock transients (degrees)
"N" : 1024 # number of mock transients used
}
},
"_rank_features" : { # input parameters for ranking functions
# ** key: group features; value: list of values
"object_count" : [...], "angular_distance" : [...], "ls8_size_dRe" : [...], ...
}
}, ...
],
```

```
# coordinate of confirmed host or primary candidate, after cross-identification
"host_ra": 210.8024, # cross-matched host RA (degree)
"host_dec": 54.3487, # cross-matched host Dec (degree)
"host_dist": 84.6571, # distance to transient position (arcsec)
```

```
"host_dxy": [0.1064, -0.0938], # position w.r.t. field center (arcsec)
"host_offset": [59.8684, 270.7856], # position w.r.t. transient (arcsec)
"host_srcid" : ["NED", "object_name", "MESSIER 101"], # source id of cross-matched host
# ** (catalog, column name, source name)

# results from visual inspection, if applicable (copied from another event for example)
"vis_insp" : {
# flags and comments
"status" : "Other", # status: "OK", "Failed", "Confusing", "Unclear", "Other"
"flags" : ["H", ...], #
"comment" : "possible_host_beyond_radius prefer_alternative_host",
# other flags and comments besides the "status" above

# manual reassignment
"reassigned" : { # reassigned host, if within the radius
# ** key: host name ("G0", "G1", "G2", ...)
"G0": {
# ** key: index of cross-matched group
"1": [
1, # index of selected group under "groups"
[-5.4725, 9.2440], # marked position in the image, RA Dec offset in arcsec
[-6.0513, 8.9101] # '_avr_dxy' of selected group
], ...
},
# each host may contain multiple cross-matched groups
},
"pin_pos" : [(-38.8769, -30.1731), ...], # manually marked host(s), if outside the radius
# ** list of (delta_RA, delta_Dec) pairs, w.r.t. field center, in arcseconds

# summary of visual inspection
"resolved_by" : "use_existing", # source of compiled host data
"case_kind" : ["known_host", "ok", ...], # automatically generated descriptive flags
"case_code" : "D1" # automatically generated case code
}
```

```
# cross-matching threshold
"xid_thres": {
"N_cps": 3, # number of groups under optimal threshold,
"thres_axis": [...], # trial values of matching threshold,
"conn_score": [...], # number of valid pairs under each trial value
"thres": 1.0 # the optimal matching threshold
},

# coverage of catalogs.
"coverage": {
"field": ["NED", "MPAJHU", ...], # names of catalogs covering this field
"primary": ["NED", "MPAJHU", ...], # catalogs matched in the primary group
"primary_unmatched": { # catalogs not matched in the primary group
# key: catalog name, value: list, [group_id, source_id, dist]
"GALEXAIS67" : [
1, # zero-indexed id of the group including source in this catalog
0, # index of the source in that group
2.9453 # distance of primary group to that source (arcsec)
],
... # ... other unmatched catalogs
}
}
```

```
# excluded sources in the field
"excluded_sources": {
# ** key: catalog or ancillary info (begin with underscore)
"PS1v2": { # excluded sources in PS1 DR2
"N_src": 3, # N of sources in this catalog
"N_excl": 3, # N excluded
"excl_srcs": [# info of excluded source,
# ** list of dictionaries, one for each excluded source
{
```

<div align="center">(Continued)</div>

```
"src": {...}, # source info
"reason": "FILTER FUNC" # reason, "FILTER FUNC" or "BEYOND RADIUS"
}, ...
]
}, ...
"_N_src": 21, # Number of sources in the field
"_N_excl": 3, # N of excluded sources.
},

"last_update": "2021-03-06T02:04:36.303Z", # time stamp for last update
"vcc": { # version control code for each data source
# ** key: data source ("tde", "grb", "rare"); value: version control code
"sn": "d147bb8861a7e838..."
},
"rand_id": 0.1157, # random number within (0, 1)
"_id": "vtdrakibitozjnpg", # global unique id of this event
```

ORCID iDs

Yu-Jing Qin ⓘ https://orcid.org/0000-0003-3658-6026
Ann Zabludoff ⓘ https://orcid.org/0000-0001-6047-8469
Marina Kisley ⓘ https://orcid.org/0000-0003-4573-8095
Iair Arcavi ⓘ https://orcid.org/0000-0001-7090-4898
Kobus Barnard ⓘ https://orcid.org/0000-0002-8568-9518
Peter Behroozi ⓘ https://orcid.org/0000-0002-2517-6446
K. Decker French ⓘ https://orcid.org/0000-0002-4235-7337
Curtis McCully ⓘ https://orcid.org/0000-0001-5807-7893
Nirav Merchant ⓘ https://orcid.org/0000-0002-1988-8180

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, ApJS, 239, 18
Ackermann, M., Ajello, M., Asano, K., et al. 2013, ApJS, 209, 11
Aggarwal, K., Budavári, T., Deller, A. T., et al. 2021, ApJ, 911, 95
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29
Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31
Auchettl, K., Guillochon, J., & Ramirez-Ruiz, E. 2017, ApJ, 838, 149
Bannister, K. W., Deller, A. T., Phillips, C., et al. 2019, Sci, 365, 565
Barbon, R., Buondí, V., Cappellaro, E., & Turatto, M. 1999, A&AS, 139, 531
Barthelmy, S. D., Gehrels, N., Paciesas, W., et al. 2009, GRB Coordinates
    Network Circular, 10251
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Berger, E. 2009, ApJ, 690, 231
Bhandari, S., Sadler, E. M., Prochaska, J. X., et al. 2020, ApJL, 895, L37
Bianchi, L., Herald, J., Efremova, B., et al. 2011, Ap&SS, 335, 161
Bianchi, L., Shiao, B., & Thilker, D. 2017, ApJS, 230, 24
Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011,
    AJ, 142, 31
Bošnjak, Ž., Götz, D., Bouchet, L., Schanne, S., & Cordier, B. 2014, A&A,
    561, A25
Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016,
    ApJ, 818, 130
Breiman, L. 2001, Mach. Learn., 45, 5
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Campbell, H., D'Andrea, C. B., Nichol, R. C., et al. 2013, ApJ, 763, 88
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
Chatterjee, S., Law, C. J., Wharton, R. S., et al. 2017, Natur, 541, 58
Childress, M., Aldering, G., Antilogus, P., et al. 2013, ApJ, 770, 107
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2014, yCat, II/328
DES Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Filippenko, A. V., Matheson, T., & Ho, L. C. 1993, ApJL, 415, L103

Foley, R. J., & Mandel, K. 2013, ApJ, 778, 167
Fong, W., Berger, E., & Fox, D. B. 2010, ApJ, 708, 9
Fong, W., Berger, E., Metzger, B. D., et al. 2014, ApJ, 780, 118
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP,
    125, 306
Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., et al. 2021, AJ, 161, 242
French, K. D., Arcavi, I., & Zabludoff, A. 2016, ApJL, 818, L21
French, K. D., Wevers, T., Law-Smith, J., Graur, O., & Zabludoff, A. I. 2020,
    SSRv, 216, 32
French, K. D., & Zabludoff, A. I. 2018, ApJ, 868, 99
Freund, Y., & Schapire, R. E. 1997, J. Comput. Syst. Sci., 55, 119
Friedman, A. S., Wood-Vasey, W. M., Marion, G. H., et al. 2015, ApJS, 220, 9
Frieman, J. A., Bassett, B., Becker, A., et al. 2008, AJ, 135, 338
Gagliano, A., Narayan, G., Engel, A., & Carrasco Kind, M. 2021, ApJ, 908, 170
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gal-Yam, A. 2019, ARA&A, 57, 305
Gal-Yam, A., Mazzali, P. A., Manulis, I., & Bishop, D. 2013, PASP, 125, 749
Galbany, L., Hamuy, M., Phillips, M. M., et al. 2016, AJ, 151, 33
Ganeshalingam, M., Li, W., Filippenko, A. V., et al. 2010, ApJS, 190, 418
Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, AJ, 157, 98
Graur, O., Bianco, F. B., Huang, S., et al. 2017, ApJ, 837, 120
Graur, O., French, K. D., Zahid, H. J., et al. 2018, ApJ, 853, 39
Guillochon, J., Parrent, J., Kelley, L. Z., & Margutti, R. 2017, ApJ, 835, 64
Gupta, R. R., D'Andrea, C. B., Sako, M., et al. 2011, ApJ, 740, 92
Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. 2016, AJ, 152, 154
Hambly, N. C., MacGillivray, H. T., Read, M. A., et al. 2001, MNRAS,
    326, 1279
Heintz, K. E., Prochaska, J. X., Simha, S., et al. 2020, ApJ, 903, 152
Helou, G., Madore, B. F., Schmitz, M., et al. 1991, in Databases and On-line
    Data in Astronomy, ed. M. A. Albrecht & D. Egret (Berlin: Springer), 89
Hicken, M., Challis, P., Kirshner, R. P., et al. 2012, ApJS, 200, 12
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hjorth, J., Malesani, D., Jakobsson, P., et al. 2012, ApJ, 756, 187
Ho, T. K. 1995, in Proc. 3rd Int. Conf. on Document Analysis and Recognition
    (Los Alamitos, CA: IEEE Computer Society), 278
Hodgkin, S. T., Harrison, D. L., Breedt, E., et al. 2021, A&A, 652, A76
Huber, M., Chambers, K. C., Flewelling, H., et al. 2015, ATel, 7153, 1
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jarrett, T. H., Chester, T., Cutri, R., et al. 2000, AJ, 119, 2498
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, ApJ, 735, 112
Jha, S., Riess, A. G., & Kirshner, R. P. 2007, ApJ, 659, 122
Jin, Z.-P., Hotokezaka, K., Li, X., et al. 2016, NatCo, 7, 12898
Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2018, ApJ, 857, 51
Kasliwal, M. M., Kulkarni, S. R., Gal-Yam, A., et al. 2012, ApJ, 755, 161
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Keane, E. F., Johnston, S., Bhandari, S., et al. 2016, Natur, 530, 453
Keller, S. C., Schmidt, B. P., Bessell, M. S., et al. 2007, PASA, 24, 1
Kessler, R., Bassett, B., Belov, P., et al. 2010, PASP, 122, 1415
Koposov, S. E., Belokurov, V., & Torrealba, G. 2017, MNRAS, 470, 2702
Kron, R. G. 1980, ApJS, 43, 305
Lang, D. 2014, AJ, 147, 108

Lang, D., Hogg, D. W., & Mykytyn, D. 2016, The Tractor: Probabilistic astronomical source detection and measurement, Astrophysics Source Code Library, ascl:1604.008

Law, C. J., Butler, B. J., Prochaska, J. X., et al. 2020, ApJ, 899, 161

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Leaman, J., Li, W., Chornock, R., & Filippenko, A. V. 2011, MNRAS, 412, 1419

Lennarz, D., Altmann, D., & Wiebusch, C. 2012, A&A, 538, A120

Li, W., Chornock, R., Leaman, J., et al. 2011, MNRAS, 412, 1473

Lien, A., Sakamoto, T., Barthelmy, S. D., et al. 2016, ApJ, 829, 7

Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31

Lunnan, R., Kasliwal, M. M., Cao, Y., et al. 2017, ApJ, 836, 60

Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406

Lyman, J. D., Levan, A. J., Tanvir, N. R., et al. 2017, MNRAS, 467, 1795

Macquart, J. P., Prochaska, J. X., McQuinn, M., et al. 2020, Natur, 581, 391

Mahony, E. K., Ekers, R. D., Macquart, J.-P., et al. 2018, ApJL, 867, L10

Mainzer, A., Grav, T., Bauer, J., et al. 2011, ApJ, 743, 156

Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, A&A, 570, A13

Marcote, B., Nimmo, K., Hessels, J. W. T., et al. 2020, Natur, 577, 190

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1

McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, Msngr, 154, 35

Mereghetti, S., Götz, D., Borkowski, J., Walter, R., & Pedersen, H. 2003, A&A, 411, L291

Miknaitis, G., Pignata, G., Rest, A., et al. 2007, ApJ, 666, 674

Möller, A., Peloton, J., Ishida, E. E. O., et al. 2021, MNRAS, 501, 3272

Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, ApJS, 236, 9

Oemler, A. J., & Tinsley, B. M. 1979, AJ, 84, 985

Onken, C. A., Wolf, C., Bessell, M. S., et al. 2019, PASA, 36, e033

Parodi, B. R., Saha, A., Sandage, A., & Tammann, G. A. 2000, ApJ, 540, 634

Paturel, G., Petit, C., Prugniel, P., et al. 2003, A&A, 412, 45

Peacock, J. A., Hambly, N. C., Bilicki, M., et al. 2016, MNRAS, 462, 2085

Peek, J. E. G., & Schiminovich, D. 2013, ApJ, 771, 68

Perets, H. B., Gal-Yam, A., Mazzali, P. A., et al. 2010, Natur, 465, 322

Perley, D. A., Metzger, B. D., Granot, J., et al. 2009, ApJ, 696, 1871

Petroff, E., Barr, E. D., Jameson, A., et al. 2016, PASA, 33, e045

Petrosian, V. 1976, ApJL, 210, L53

Pineau, F. X., Derriere, S., Motch, C., et al. 2017, A&A, 597, A89

Prieto, J. L., Stanek, K. Z., & Beacom, J. F. 2008, ApJ, 673, 999

Prochaska, J. X., Macquart, J.-P., McQuinn, M., et al. 2019, Sci, 366, 231

Rahman, M., Ménard, B., & Scranton, R. 2016, MNRAS, 457, 3912

Ravi, V., Catha, M., D'Addario, L., et al. 2019, Natur, 572, 352

Reindl, B., Tammann, G. A., Sandage, A., & Saha, A. 2005, ApJ, 624, 532

Rodney, S. A., Riess, A. G., Strolger, L.-G., et al. 2014, AJ, 148, 13

Rodríguez, Ó., Clocchiatti, A., & Hamuy, M. 2014, AJ, 148, 107

Sako, M., Bassett, B., Becker, A. C., et al. 2018, PASP, 130, 064002

Salvato, M., Buchner, J., Budavári, T., et al. 2018, MNRAS, 473, 4937

Savaglio, S. 2006, NJPh, 8, 195

Savaglio, S., Glazebrook, K., & Le Borgne, D. 2006, in AIP Conf. Proc. 836, Gamma-Ray Bursts in the Swift Era, ed. S. S. Holt, N. Gehrels, & J. A. Nousek (Melville, NY: AIP), 540

Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019, ApJS, 240, 30

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., et al. 1998, ApJ, 507, 46

Sevilla-Noarbe, I., Hoyle, B., Marchã, M. J., et al. 2018, MNRAS, 481, 5451

Shanks, T., Metcalfe, N., Chehade, B., et al. 2015, MNRAS, 451, 4238

Shivvers, I., Modjaz, M., Zheng, W., et al. 2017, PASP, 129, 054201

Shivvers, I., Filippenko, A. V., Silverman, J. M., et al. 2019, MNRAS, 482, 1545

Silverman, J. M., Nugent, P., Gal-Yam, A., et al. 2016, in IAU Proc. 29B, Astronomy in Focus, as presented at the IAU XXIX General Assembly, 2015 (Cambridge: Cambridge Univ. Press), 237

Silverman, J. M., Foley, R. J., Filippenko, A. V., et al. 2012, MNRAS, 425, 1789

Silverman, J. M., Nugent, P. E., Gal-Yam, A., et al. 2013, ApJS, 207, 3

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Smartt, S. J., Valenti, S., Fraser, M., et al. 2015, A&A, 579, A40

Smith, K. W., Williams, R. D., Young, D. R., et al. 2019, RNAAS, 3, 26

Smith, M., Nichol, R. C., Dilday, B., et al. 2012, ApJ, 755, 61

Smith, N., Li, W., Filippenko, A. V., & Chornock, R. 2011, MNRAS, 412, 1522

Stahl, B. E., Zheng, W., de Jaeger, T., et al. 2019, MNRAS, 490, 3882

Steer, I., Madore, B. F., Mazzarella, J. M., et al. 2017, AJ, 153, 37

Stone, N. C., Generozov, A., Vasiliev, E., & Metzger, B. D. 2018, MNRAS, 480, 5060

Sullivan, M., Le Borgne, D., Pritchet, C. J., et al. 2006, ApJ, 648, 868

Tammann, G. A. 1978, MmSAI, 49, 315

Taubenberger, S. 2017, in Handbook of Supernovae, ed. A. W. Alsabti & P. Murdin (Cham: Springer International), 317

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

Troja, E., Ryan, G., Piro, L., et al. 2018, NatCo, 9, 4089

Tully, R. B., Courtois, H. M., Dolphin, A. E., et al. 2013, AJ, 146, 86

van den Bergh, S. 1959, AnAp, 22, 123

van Velzen, S., Gezari, S., Hammerstein, E., et al. 2021, ApJ, 908, 4

Vianello, G., Götz, D., & Mereghetti, S. 2009, A&A, 495, 1005

Wang, X., Wang, L., Pain, R., Zhou, X., & Li, Z. 2006, ApJ, 645, 488

Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9

Weyant, A., Wood-Vasey, W. M., Allen, L., et al. 2014, ApJ, 784, 105

Williams, P. K. G., & Berger, E. 2016, ApJL, 821, L22

Wiseman, P., Smith, M., Childress, M., et al. 2020, MNRAS, 495, 4040

Wolf, C., Onken, C. A., Luvaul, L. C., et al. 2018, PASA, 35, e010

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Yang, B., Jin, Z.-P., Li, X., et al. 2015, NatCo, 6, 7323

Yaron, O., & Gal-Yam, A. 2012, PASP, 124, 668

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

Yuan, H. B., Liu, X. W., & Xiang, M. S. 2013, MNRAS, 430, 2188

Zhang, F.-W., Shao, L., Yan, J.-Z., & Wei, D.-M. 2012, ApJ, 750, 88

Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2021, MNRAS, 501, 3309

Zou, H., Zhou, X., Fan, X., et al. 2017, PASP, 129, 064101