5-3-2017

A New All-sky Catalogue of Candidate Protoplanetary Disks from Aggregated Optical and Infrared Surveys

Daniel Horenstein
Georgia State University

A NEW ALL-SKY CATALOGUE OF CANDIDATE PROTOPLANETARY DISKS FROM AGGREGATED OPTICAL AND INFRARED SURVEYS

by

DANIEL HORENSTEIN

Under the Direction of Sébastien Lépine, PhD

ABSTRACT

To guide research into formation scenarios for giant planets, I present a catalogue of 199,391 sources with candidate protoplanetary disks. These candidates are identified through an iterative color selection process in a 5,460-dimensional color-color space. As part of this process, photometry and SIMBAD classifications are aggregated from multiple large surveys, collectively covering 14 optical and infrared bands from 0.35 – 22 microns. An estimated false positive rate of 36.1% makes this catalogue a suitably sturdy foundation for a wide range of subsequent studies. After a discussion of the context surrounding this research, the forthcoming journal article describing this catalogue is reprinted in its entirety, and commentary is offered on the results.

INDEX WORDS: protoplanetary disks — stars: pre-main sequence — surveys

A NEW ALL-SKY CATALOGUE OF CANDIDATE PROTOPLANETARY DISKS FROM AGGREGATED OPTICAL AND INFRARED SURVEYS

by

DANIEL HORENSTEIN

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2017

A NEW ALL-SKY CATALOGUE OF CANDIDATE PROTOPLANETARY DISKS FROM AGGREGATED OPTICAL AND INFRARED SURVEYS

by

DANIEL HORENSTEIN

Committee Chair:	Sébastien Lépine
Committee:	Fabien Baron
	Russel White

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2017

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Preface

"How do you build a planet?" It's one of those maddeningly simple classroom questions that astronomy has yet to answer definitively. Without a clear picture of how the most massive planet in our Solar System formed — much less Jupiter's cousins, the other giant planets — a key genealogical link in our astronomical family history remains obscured.

While two competing theories have been proposed (see Section 2.2) and there may even be a consensus leaning towards one model over the other, the fact remains that neither is grounded in extensive, empirical evidence. Indeed, each theory disagrees in some way with prior observations (Section 2.2), implying that neither, on its own, provides a full account of giant planet formation.

With the results presented here, I hope to lay the foundation for resolving this ambiguity. I present a catalogue of positions and aggregated wide-band photometry for 199,391 sources with candidate protoplanetary disks. An estimated false positive rate of 36.1% implies that approximately 127,000 of these sources may harbor unidentified but genuine protoplanetary disks. The full catalogue will be made freely available online.

This catalogue represents a significant step forward for research potential in planetary formation. Although exceptionally detailed, our knowledge of the Solar System is limited to one — and only one — planetary system. This is hardly an adequate basis for generalization, especially considering the more than four billion years intervening between the epoch of planetary formation and the resulting Solar System that exists today. The number

of protoplanetary disks identified around other stars is steadily increasing, partially thanks to more powerful infrared and sub-millimeter instrumentation such as WISE and ALMA, but use of these instruments is competitive and limited, and proposed observations must be justified using the similarly limited scope of existing protoplanetary disk confirmations. This enforced specificity of target selection risks pigeonholing any investigations of planetary formation that would hope to generalize their conclusions.

The catalogue described here is broad enough to provide a significantly more representative sample of protoplanetary disks from which research projects can begin, even surveys with the relatively straightforward (but time-consuming) goal of confirming the candidates presented here. Furthermore, this breadth permits the imposition of exacting or unusual criteria when selecting targets for studies examining planetary formation theories, astrobiology, or other aspects of planetary science. By focusing on protoplanetary disks, research can move beyond the inductive regime of observing fully formed planets, shifting instead to observing planetary systems directly during their formative stages.

The following chapter, Chapter 2, includes the journal article submitted for publication on March 11, 2017. The article is reproduced in its entirety, with the following modifications and additions:

1. The formatting has been changed in order to conform to Georgia State University publication guidelines.

2. Figure and table captions were written to conform to style guidelines for AAS journals, and in some cases, this formatting conflicts with GSU style requirements. In each of

these cases, rather than compromise the utility or legibility of the associated figure or table, an abbreviated caption has been substituted, with the original caption included as a chapter endnote.

3. Some minor elements of style have been changed in order to improve readability.

4. No major changes have been made to the content of the journal article, but in light of the fact that the article is still undergoing peer review, the forthcoming article should be regarded as the authoritative reference on this research. As a result of the review process, changes and additions in content may be present in the journal article.

5. Stemming from thoughtful comments and discussions with my esteemed colleagues, a brief section of commentary has been added in Chapter 3.

6. Appendix A provides visual evidence for the efficiency of each of the Round 1 color-color cuts, as detailed in Section 2.4 and listed in Table 2.4. These figures extend the summary illustration from Figure 2.2.

7. Similarly, Appendix B illustrates the efficiency of each of the Round 2 color-color cuts, as detailed in Section 2.5, listed in Table 2.6, and summarized in Figure 2.4. Note that due to formatting constraints, there are two figures for each cut: one showing known young stellar objects; the other, known objects of other types.

8. A number of known young stellar objects in known star-forming regions are successfully recovered by the color selection process detailed in Sections 2.3 through 2.5 — an

important benchmark that demonstrates the effectiveness of this method for identifying young stellar objects. Appendix C highlights some of the known star-forming regions identified in Figure 2.5.

In addition to the multiple research avenues that this catalogue opens, additional opportunities exist for education and public outreach. For students at all levels of science education, public catalogues such as this one provide a ready opportunity to teach transferrable coding skills in conjunction with reinforcing astronomical concepts. Building on my legacy of volunteering at over 30 public outreach events during my time at GSU, including making lasting contributions to Hard Labor Creek and Urban Life Observatories and initiating multiple new, off-campus events related to White House Astronomy Night, the story of how our Solar System formed is a captivating subject for all audiences. This story is our story.

This is a beginning.

– *Daniel Horenstein*

March 2017

CHAPTER 2

Journal Article

2.1 Abstract

We present a catalogue of 199,391 point sources with optical and infrared colors that are consistent with stars hosting protoplanetary disks. To anchor our selection process, a list of known protoplanetary disks is compiled from the literature, and lists of control field sources are selected from locations presumed to have little ongoing star formation. Optical and infrared magnitudes from multiple photometric surveys, covering up to 14 different bands, are used to define regions that reliably distinguish stars with known disks from other control field objects in a 5,460-dimensional color-color space. These color-color cuts are applied in an all-sky search of point sources in the AllWISE catalogue. SIMBAD classifications and aggregated magnitudes are used to define additional color-color cuts that efficiently distinguish known young stellar objects from sources of various other types. These additional cuts are applied to all targets either absent from SIMBAD or without conclusive SIMBAD object types to form the new catalogue of 199,391 likely young stellar objects. An estimated false positive rate of 36.1% implies the detection of approximately 127,000 heretofore unidentified protoplanetary disks. The positions of these candidates on the sky are largely consistent with a spatial distribution in the young Galactic disk. A number of nearby star-forming regions are successfully recovered through this process, and they appear to include many sources not previously reported to be young stellar objects.

2.2 Introduction

While multiple mechanisms of planetary formation have been posited and seem to work well for the Solar System, there remains a dearth of observational constraints on the broader viability of these processes. Indeed, many recent developments in planetary dynamics have forced drastic revisions of concepts previously accepted as canonical: the discovery of Hot Jupiters (Mayor & Queloz 1995); evidence for planetary migration (Lin et al. 1996); and the existence of satellites around Kuiper Belt objects (Brown et al. 2006).

The collisions and conglomeration of rocky particles to form planetesimals and eventually larger planets is generally accepted as the dominant plausible mechanism in the formation of rocky planets. The processes behind the formation of gas giants are less clear (Bodenheimer & Lin 2002; Matsuo et al. 2007). It is possible that similar "core accretion" processes form the solid cores of giant planets, which at larger orbital radii (tens of AU) become massive enough to accumulate thick, gaseous envelopes from the pre-main sequence circumstellar material (Pollack et al. 1996). Alternatively, it is also possible that dynamic processes cause overdensities of gas to form first, into which the rocky particles fall as a result of this "disk instability" (Boss 1997). Each explanation has its own strengths and weaknesses, but neither has extensive observational evidence for or against it (Helled et al. 2014).

This tension cannot be resolved sufficiently when the bulk of data concerning planetary formation is confined to the single data point of the Solar System, although a repository of examples of younger planetary systems is slowly growing. A necessary prerequisite to

making large-scale, observationally constrained, statistical arguments concerning planetary formation processes is the collection of a census of protoplanetary disks.

Compiling a large sample of this type is no small task. A protoplanetary disk will only be present after a protostellar molecular cloud has collapsed into a disk but before the protostar has reached the zero-age main sequence. While the lifetimes of planets may be long, on the order of several billion years, the protoplanetary disk phase is expected to be much shorter, on the order of several million years (Yasui et al. 2014; Williams & Cieza 2011). This comparative brevity is compounded by the difficulty that the confirmation of protoplanetary disks is usually based on infrared excess shown in the spectral index between 2 and 25 microns (Greene et al. 1994; Williams & Cieza 2011), which requires either spectroscopy or multi-band photometry. These observationally intensive methods have thus far limited most identifications of protoplanetary disks — or of their hosts, Class II young stellar objects (YSOs) — to targeted surveys, as opposed to all-sky searches, and to the relatively small number of photometric wavebands accessible to any single observatory.

No longer must this be the case. All-sky photometry of point sources in multiple wavebands has become available over the past two decades and is still being collected. This paper showcases the power of aggregating large, publicly available catalogues by devising a new method to identify a substantial number of candidate protoplanetary disks with optical and infrared photometry. Section 2.3 describes the means by which photometric data were gathered for a sample of known protoplanetary disks and a control group of field sources, and Section 2.4 details the first stage of defining colors that uniquely characterize protoplan-

etary disks. These approaches are extended in Section 2.5 to include an all-sky search of the AllWISE catalogue and refinements to the color selection criteria. The final catalogue of 199,391 candidate protoplanetary disks, with an estimated false positive rate of only 36.1%, is presented in Section 2.6 and contextualized against recent literature in Section 2.7.

2.3 Data Acquisition: Known Protoplanetary Disks and Control Field Sources

To begin analyzing the colors of young stellar objects (YSOs) with confirmed protoplanetary disks, a literature search was performed to construct a catalogue of stars known to have such disks. The papers from which targets were selected are listed in Table 2.1, along with any additional criteria that were applied when selecting targets from a particular paper. No magnitude or flux data were collected at this stage; only target coordinates or names. In the cases of targets where no coordinates were provided, SIMBAD was used to resolve the provided names. In cases with no provided coordinates and no SIMBAD record, coordinates were inferred from the target name itself when possible; a 2MASS catalogue identifier, for instance, is comprised of the right ascension and declination of the source. Claimed disk detections with no provided coordinates, no SIMBAD records, and unhelpfully opaque names were not selected. Duplicate targets, with nearly identical coordinates present in more than one of these papers, were merged into single records.

The coordinates of these 1,126 stars with previously known protoplanetary disks are plotted in Figure 2.1. This list of stars with known disks is not claimed to be exhaustive, but it does show the distribution expected for YSOs: there are many sources at low Galactic

Table 2.1. Literature Sources of Stars with Known Protoplanetary Disks

Paper	Location and Criteria
Acke et al. (2004)	Table 1
Alonso-Albi et al. (2009)	Table 1; disk detected
Evans, II et al. (2003)	CLOUDS, OFF-CLOUD, CORES, & STARS; candidate YSOs; $-0.3 > \alpha \geq -1.6$
Luhman et al. (2010)	Table 7; SED Class II
Mannings & Sargent (1997)	Table 1
Mannings & Sargent (2000)	Table 1; thermal continuum emission detected
Rebull et al. (2010)	Table 4 Table 8; Mm = "M"
Scholz et al. (2006)	Table 1

Note. — Papers from which the names or coordinates of previously confirmed or (claimed to be) probable protoplanetary disks were collected. This is not a comprehensive list of all stars with confirmed protoplanetary disks, but it is claimed to be broad enough to be representative; see the distribution in Figure 2.1 and the discussion thereof in Figure 2.3. If additional criteria were applied when collecting data from a given paper, these constraints are also provided. The criterion on α for an accreting circumstellar disk comes from Greene et al. (1994).

latitudes, across a range of Galactic longitudes, with fewer sources at higher latitudes (Portegies Zwart et al. 2010). The Gould Belt, which is posited to contain most nearby star-forming regions (Ward-Thompson et al. 2007), is somewhat visible as a chevron with a peak in Galactic latitude at longitude $l = 0°$ and a latitudinal trough at longitude $l = \pm 180°$. The clumpy concentrations of most stars with known protoplanetary disks within small areas on the sky

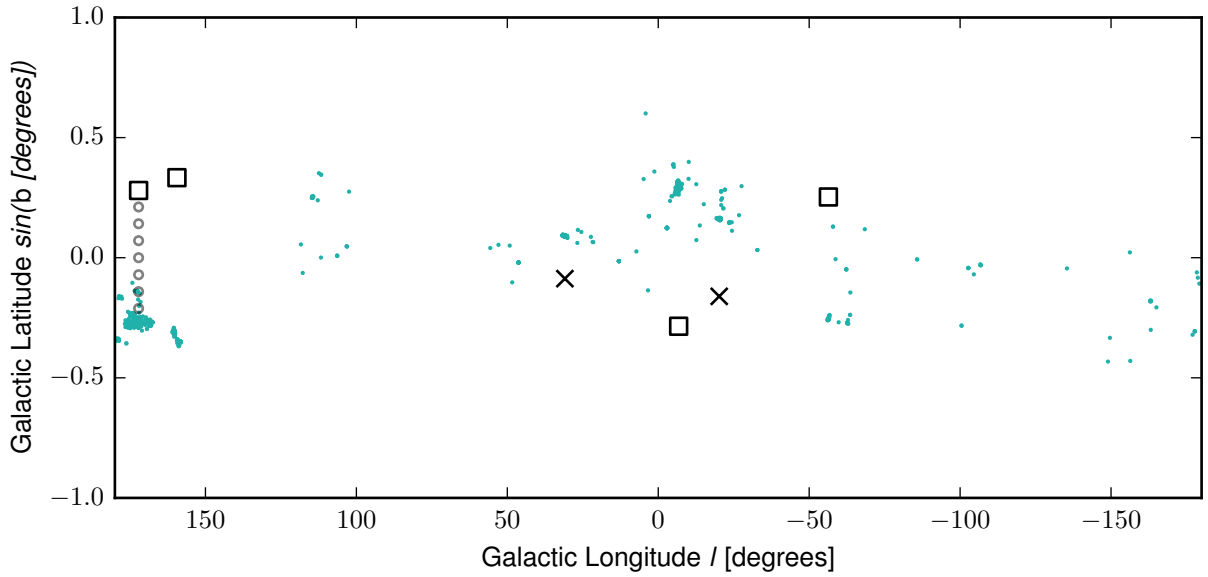

Figure 2.1 — Locations of stars with known protoplanetary disks (turquoise), with control field regions marked with squares. Few protoplanetary disks are expected at these control field locations (a prediction supported later by the distribution of known YSOs in Figure 2.3). Symbol size is not representative of the extent of these regions. Control field regions were selected to be at opposite Galactic latitudes from the six densest loci of known protoplanetary disks, although the two control field regions closest to the Galactic plane (X symbols) were discarded in order to mitigate possible reddening due to interstellar dust. These considerations for extrinsic reddening are discussed at the end of Section 2.3. Between the largest clump of known disks and the control field opposite it, seven intermediate fields (gray circles) were selected to investigate the effect of reddening as a function of Galactic latitude. Sources from these intermediate fields were not used in the subsequent color selection process.

is also expected, as disks are often confirmed by obtaining a target's spectral index between 2 and 25 microns (Williams & Cieza 2011; Oliveira et al. 2009; Furlan et al. 2009), and this spectroscopy or multi-band photometry can be an observationally intensive endeavor, one that is typically limited to fields-of-view already suspected to harbor star-forming regions.

A control sample of sources not expected to show significant ongoing star or planetary system formation was assembled in order to determine the typical colors of normal field stars (as opposed to YSOs). This reference sample is necessary for the later step of isolating

the characteristic colors of stars with known protoplanetary disks. After identifying the approximate centroids of the regions on the sky with the greatest number densities of known disks, six locations from which to draw control field sources were chosen at points of equal Galactic longitude and opposite Galactic latitude to the six corresponding centroids. For each pair of regions — a centroid of stars with known disks and its corresponding control field locus — the requirement for equal *absolute* Galactic latitude ensures a similar stellar number density in both regions. Due to the inclination of the Gould belt relative to the Galactic plane, negating the Galactic latitude sign of each centroid of known disks provides a control field locus "mirrored" across the Galactic plane, a locus that is not in the Gould Belt and hence presumably displays little ongoing star formation.

At each of these six control field locations, which are marked in Figure 2.1, the coordinates of all records in the 2MASS Point Source Catalog (Cutri et al. 2003) within a 0.5° radius were retrieved. 2MASS was chosen for this purpose because stars with protoplanetary disks are expected to display color excesses in mid-infrared and far-infrared regimes (Williams & Cieza 2011); as such, choosing field sources from this near-infrared spectral region is a compromise expected to capture both main sequence stars that emit primarily at optical wavelengths and stars with protoplanetary disks that are (relatively) bright at longer infrared wavelengths.

Magnitudes were retrieved from each of the photometric surveys listed in Table 2.2 for each of the stars with known protoplanetary disks and each of the sources from the control fields. Cross-matching was accomplished by using the X-Match service available from CDS, which returns all sources from a specified VizieR catalogue within 5" of each coordinate

Table 2.2. Photometric Surveys Queried

Survey	Filters	Criteria	Reference
SDSS (DR9)	u', g', r', i', z'	cl = 6; Q = 2, 3, or 0	Ahn et al. (2012)
APASS (DR9)	B, g', V, r', i'	···	Henden et al. (2016)
Tycho-2	B, V	···	Hog et al. (2000)
2MASS (PSC)	J, H, K	Qflg = A, B, C, or D	Cutri et al. (2003)
AllWISE	$W1$, $W2$, $W3$, $W4$	qph = A, B, or C	Cutri et al. (2014)

Note. — Surveys from which photometric magnitudes were collected for all sources presented in this work. The filters used by each survey are presented, as are the quality criteria applied before accepting a magnitude in the data aggregation process detailed in Section 2.3. For more details about the quality flags for each survey, including whether the flag describes an individual magnitude or the overall source record, consult the VizieR page or the provided reference for that survey.

pair in a user-supplied list. When aggregating the data returned by these cross-matches, only the single result closest to each queried object was accepted. Not all objects in the known protoplanetary disk and control field source lists were found to have corresponding records in each of the five surveys queried. For each query result matched successfully on the basis of its spatial coordinates, each individual magnitude from the result was only accepted upon meeting additional photometric quality criteria unique to each survey, as specified in Table 2.2. As such, photometry is incomplete for some stars with known protoplanetary disks and some control field sources across the 14 wavebands available; this incompleteness is accounted for throughout the subsequent data analysis.

Due to varying angular resolutions among the surveys, it is possible that a single source from a queried survey could be matched to more than one object in the lists of stars with

known protoplanetary disks and control field sources. These potential duplicates are not excluded, on the grounds that not enough information exists to match the query result source to a specific known protoplanetary disk or control field source (or other source not on either list) definitively. It is assumed that any spurious matches will not have the requisite colors necessary to survive the color selection process detailed below.

The same data aggregation process, with the same surveys, was used for all other objects presented later in this work.

For one of the control fields, seven intermediate fields were queried every 4° of Galactic latitude (at constant Galactic longitude) between the control field location and the corresponding locus of stars with known protoplanetary disks in order to investigate the effect of reddening from interstellar dust. The locations of these intermediate fields are indicated in Figure 2.1. This reddening is expected to be more severe at lower Galactic latitudes. Based on the shift of the visually apparent main sequence in a few color-color plots for these intermediate fields, reddening was noticeable only at the lowest latitudes: $\Delta(W1-W2)$, for instance, was approximately 0.1 near the Galactic plane, and $\Delta(g'-i')$ was as much as 0.3, but these effects diminished within 4° of Galactic latitude and were no longer visible outside of 8°. Erring on the side of caution to eliminate control fields that could have ambiguous sources of reddening, the two control fields within 10° of the Galactic equator were excluded from further consideration. In the few color-color plots examined, the rejection of these two fields proved successful at more clearly separating the control field sources from the known protoplanetary disks — a separation that should be expected when choosing control field

regions with little ongoing star formation. The seven intermediate fields were not included in further comparisons involving field sources.

2.4 Color Selection: Known Protoplanetary Disks and Control Field Sources

Defining inclusion or exclusion regions in color-color spaces has previously been used to great effect in separating accreting YSOs from the broader populations of targets in which they are contained (Rebull et al. 2007; Harvey et al. 2007; Chapman et al. 2007). More recently, this technique has been extended to include survey aggregation for a few specific regions on the sky (Carlson et al. 2012) and to search for more evolved circumstellar debris disks (Binks & Jeffries 2016) by leveraging prior color criteria derived from observational and theoretical studies.

In surveys conducted with a single instrument and a limited number of wavebands, all possible color-color plots could be examined visually. With the 14 aggregated magnitudes described above, however, each source occupies a point in a 105-dimensional color and magnitude space, assuming such a source has a high-quality magnitude measurement in each of the 14 available bands (which many do not). Two of these colors or magnitudes will be required for the production of a color-color, color-magnitude, or magnitude-magnitude plot. It is therefore impractical to render each of the 5,460 possible color-color plots (including color-magnitude plots and magnitude-magnitude plots, without loss of generality) and attempt to define regions by eye that separate known protoplanetary disks from control field sources. The vast majority of these plots are unlikely to show any salient information —

using a blue color such as $u' - g'$, for example, is unlikely to be relevant for circumstellar disks that are characterized by infrared excess — but addressing this concern directly would require the introduction of assumptions about the colors and underlying physics of stars with protoplanetary disks. One goal of this work is to take a strictly empirical approach and mitigate any physical assumptions.

Furthermore, defining an exclusion region, or "cut," in one of these dimensions may then reveal further cuts that can be made in other dimensions. To resolve these issues in a pragmatic manner, a two-step iterative process was applied. In the first step, the Kolmogorov-Smirnov (KS) statistic was calculated in each of the 105 color or magnitude dimensions in order to quantify the degree to which the distribution of stars with known protoplanetary disks was statistically different from the distribution of control field sources in that dimension. While the actual value of the statistic may be of questionable utility in determining colors of interest — if the KS test is required to determine that the two populations may be distinct, it is unlikely that a cut will be able to be made by eye — the relative values of the test statistic among the various color and magnitude dimensions will suggest which color-color plots will be most worthwhile to render and to examine for useful cuts, with higher values likely proving more fruitful. The goal in making these cuts was to eliminate as many control field sources as possible while retaining as many known protoplanetary disks as possible. Following this second step of making one or more cuts in a single color-color space, the KS statistics for the surviving sources were recalculated in the same 105 color and magnitude dimensions and used to make additional cuts, after which

the process was repeated as often as necessary to minimize the KS statistics in all color and magnitude dimensions.

For this particular sample of stars with known protoplanetary disks and sources from control fields, an excerpt of the initial KS statistics, prior to making any color-color cuts, is shown in the left two columns of Table 2.3. The values given here are the test statistics themselves, not probabilities, which would require the assumption of an underlying statistical distribution. These values have been corrected for varying sample size, with

$$Z_n \equiv D_{KS}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where D_{KS} is defined to be the greatest difference between the cumulative distribution functions of two samples of sizes n_1 and n_2 (Metchev & Grindlay 2002). This correction is necessary because sources with incomplete or low-quality photometry in one or both of the requisite wavebands for a particular color or magnitude are not included in the calculation of the corresponding KS statistic. As noted above, the values themselves are not important: consider only how the values are ordered and by how much one value differs from the others around it. The initial statistics suggest that the known protoplanetary disks and control field sources would show the largest separation in the ($W2$, $W1 - W2$)-plane, which is rendered in the top two rows of Figure 2.2. The first (leftmost) column shows the two populations prior to making any cuts. As predicted, the two populations show a significant separation, and three color-magnitude cuts were made by eye to delineate a boundary between the two populations, as plotted in the second (middle) column of Figure 2.2 and listed in Table 2.4.

Table 2.3. KS Statistics Excerpt, Round 1

Initial		After Cut 3		After Cut 9	
Dimension	Z_n	Dimension	Z_n	Dimension	Z_n
$W1-W2$	27.787	H	26.045	J	17.343
$W2$	27.588	K	25.842	H	5.620
$J-W2$	26.712	J	23.810	K	4.959
$H-W2$	26.678	$J-K$	20.686	$r'-J$	4.018
$K-W2$	26.537	$J-H$	20.526	$i'-J$	3.946
$W1$	25.956	⋮	⋮	$g'-J$	3.847
$J-W1$	25.747			$r'-i'$	3.647
$H-W1$	25.394	$W1-W4$	3.882	⋮	⋮
$J-K$	25.211	$W2$	3.537		
K	24.530	B	3.515	$g'-H$	2.822
$H-K$	24.030	⋮	⋮	$J-K$	2.685
$K-W1$	23.544			$J-H$	2.667
$W3$	23.149	$B-W1$	1.709	⋮	⋮
$J-H$	23.101	$W1-W2$	1.705		
H	22.996	$u'-r'$	1.593	$W1-W4$	1.032
J	19.786	⋮	⋮	$K-W4$	0.902
⋮	⋮			$W2-W4$	0.857
		$u'-g'$	1.035	$W3-W4$	0.847
$u'-g'$	3.146	$J-W2$	0.813	$u'-B$	—
u'	2.903	$u'-B$	—	$u'-g'$	—
$u'-B$	2.630	$u'-V$	—	$u'-V$	—
$u'-W4$	2.295	⋮	⋮	⋮	⋮
$z-W4$	1.972				

Note. — An excerpt of the Kolmogorov-Smirnov (KS) statistics comparing the distributions of stars with known protoplanetary disks and control field sources in 105 colors and magnitudes. The left, center, and right two-column pairs show KS statistics: prior to defining any inclusion regions, or "cuts;" recalculated for surviving sources after making the first three cuts listed in Table 2.4; and after making all nine color-magnitude cuts, respectively.[(1)]

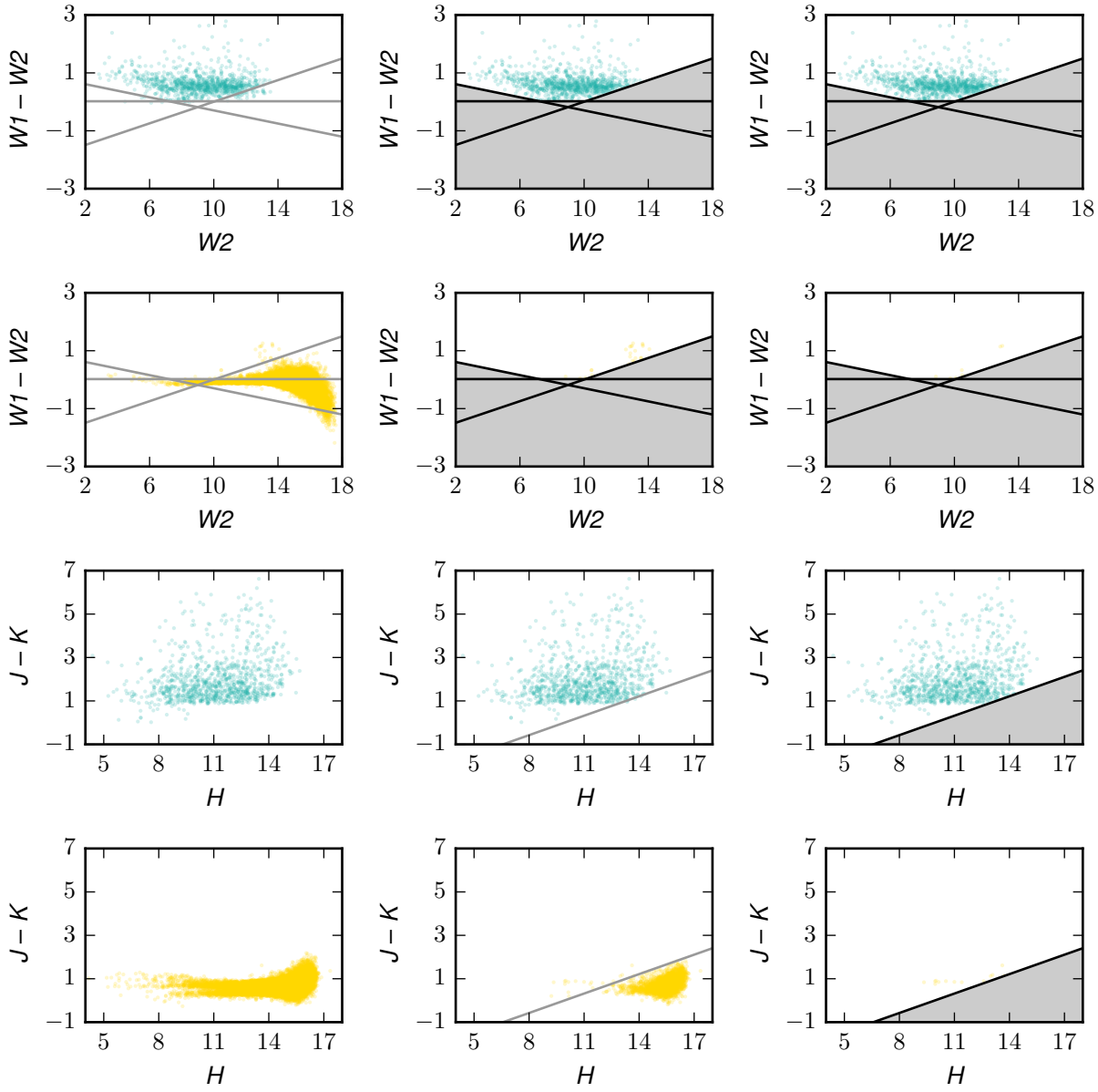

Figure 2.2 — Colors of stars with known protoplanetary disks and sources from control fields through the color selection process from Section 2.4. Turquoise points represent known protoplanetary disks; gold points, sources from control fields presumed to have little ongoing star formation. The first, second, and third columns show sources: prior to any color-color cuts; after the first three cuts from Table 2.4; and after all nine color-magnitude cuts, respectively. Each cut is previewed in gray in its applicable color-magnitude space, then shown as a black line and gray exclusion region. A source must have reasonably good photometry (see Table 2.2) in all requisite bands for a color-magnitude space in order to be plotted and affected by any cuts in that space. All plotted points are 80% transparent.[(2)]

Table 2.4. Color-Color Cuts, Round 1

Criterion
$W1 - W2 > 0.02$
$W1 - W2 > 0.186667 \cdot W2 - 1.863337$
$W1 - W2 > -0.113333 \cdot W2 + 0.836665$
$J - K > 0.298333 \cdot H - 2.963329$
$J - H > 0.483333 \cdot H - 6.083329$
$J - W1 > 1.25 \cdot J - 19$
$J - K > 0.625 \cdot J - 8.5$
$H - K > 0.04 \cdot J - 0.3$
$J - K > 0.148125 \cdot J - 1.2925$

Note. — Inclusion criteria applied to identify the unique region of multidimensional color-color space occupied by stars with known protoplanetary disks (and hence cutting out sources from control fields, as described in Section 2.4). For all cases in this table, the included regions are defined with lower-dimensional color or color-magnitude cuts. The cuts are listed from top to bottom in the order in which they were made.

KS statistics were then recalculated, as shown in the center two columns of Table 2.3, for all sources that do not fail any of the first three cuts. (This "not failing" criterion for cuts with incomplete photometry is discussed in more detail below.) After the first three cuts, the distribution of the remaining stars with known protoplanetary disks and that of the remaining sources from control fields were no longer significantly different in $W1 - W2$ or in $W2$, as reflected in the much smaller values of Z_n. Several color and magnitude dimensions did not

contain enough surviving sources with good-quality photometry to recalculate a revised KS statistic. Excluding magnitude-magnitude spaces from consideration as analogous to a single color space — a cut available in the (H, K)-plane, for instance, should also be apparent in the $H - K$ color alone — the (H, $J - K$)-plane was selected as the color-magnitude space with the next greatest potential to separate the known protoplanetary disks from the control field sources. This space is rendered in the bottom two rows of Figure 2.2. A fourth cut was identified in the (H, $J-K$)-plane, as listed in Table 2.4. This two-step process of generating KS statistics, after taking into account any prior color-color cuts, and then making additional cuts by eye was repeated until the surviving populations of known protoplanetary disks and control field sources showed no significant separation in any of the 105 colors or magnitudes (with the possible exception of J, as discussed in more detail below).

A summary of this first round of color-color cuts is presented in Table 2.4 in the order in which they were made. No source was excluded solely on the basis of incomplete or poor-quality photometry: a source must have reasonably good-quality photometry (see Table 2.2) in all requisite wavebands in order to fail a cut. These represent "strict" cuts that do not retain all stars with known protoplanetary disks but rather balance including as many of these stars as possible while simultaneously excluding as many control field sources as possible. This approach is justified: some stars claimed to have protoplanetary disks may be incorrectly classified or anomalous, and some control field sources may by chance be stars with protoplanetary disks. Of the 1,126 confirmed protoplanetary disks, 1,072 (95.2%) are recovered after this first round of color-color cuts; of 37,049 control field sources, only 739

(2.0%) remain. As expected for YSOs characterized by infrared excesses, the most efficient cuts used only the near- and mid-infrared magnitudes from 2MASS and WISE.

The last (rightmost) column of the color-magnitude plots presented in Figure 2.2 reflects all nine color-color cuts made thus far, as listed in Table 2.4. KS statistics accounting for these nine cuts are given in the two right-hand columns of Table 2.3. Visual inspection of the one remaining suggested cut, in the $(J,\ r' - J)$-plane, showed that a "strict" cut in this color-magnitude space would have been of negligible benefit, affecting only 18 control field sources that had not already been cut. Such a cut would also have introduced of significant complexity by involving an optical magnitude (r') into the second round of color selection, as described in the next section.

While the use of a proposed two-dimensional analogue to the KS test was investigated, Figure 2.2 highlights that it is, in fact, the researcher who requires that the data be two-dimensional for the production of a color-color plot; indeed, it can be seen that shifting the population of stars with known protoplanetary disks along a single color or magnitude axis will make it either more or less distinct from the population of sources from control fields. Although this suggests that one-dimensional color histograms could be sufficient for making cuts, taking such an approach would sacrifice a significant amount of potential by restricting cuts to the 105-dimensional color and magnitude space instead of the 5,460-dimensional color-color, color-magnitude, and magnitude-magnitude space.

Pure magnitude cuts were avoided in order to prevent the introduction of an explicit distance bias, although implicit distance cutoffs remain due to the saturation and magnitude

completeness limits of the surveys used. Indeed, many selection effects are likely to be present in the magnitude data, and hence also in the sample of sources remaining after color-color cuts are made, due to the amalgamation of all selection effects present in the surveys. This paper does not attempt to mitigate these selection effects in order to make pronouncements concerning all protoplanetary disks in existence, but rather to work with the data available in order to highlight the protoplanetary disks that should be observable today with current instrumentation.

2.5 Data Acquisition and Color Selection: Known YSOs and Known Non-YSOs

In the second phase of our selection process, the color-color cuts defined above are applied to an all-sky search of the AllWISE catalogue. Conveniently, AllWISE contains photometry for each source not only from the WISE survey itself ($W1$, $W2$, $W3$, $W4$) but also from the nearest 2MASS point source (J, H, K), if one exists within a 3" search radius from the WISE detection (with caveats; see Cutri et al. (2014)). An SQL query was designed to return all AllWISE targets that:

1. are not extended sources;
2. have good-quality photometry in $W1$ and $W2$, the two WISE bands relevant to the nine cuts specified in Table 2.4; and
3. do not fail any of these nine cuts.

The query was structured in this way to account for incomplete photometry: while a WISE source without a 2MASS association would fail any cut that requires J and $W1$

magnitudes, for example, such a source will not necessarily fail the SQL query. In such a case, there is not enough information to exclude the possibility that the source has similar colors to stars with known protoplanetary disks, provided that the source satisfies other selection criteria. That said, the requirement for good $W1$ and $W2$ photometry has the effect of selecting only the most reliable WISE detections from which to launch the second stage of color selection. The impact on the false negative rate from this decision to select only high-quality WISE detections is discussed in Section 2.6.

623,867 AllWISE sources were returned that satisfied the three criteria above. The SIMBAD classifications of these objects were retrieved, using the closest SIMBAD match within a 5" search radius. 71,214 (11.4%) of the sources in this subset of AllWISE were found to have SIMBAD records. Of these matches, 6,997 (9.8% of matches) are listed as having previously been identified as confirmed or candidate young stellar objects (YSOs), using at least one of the following SIMBAD object abbreviations: "Y*O," "Y*?," "TT*," "TT?," "HH," "Or*," "FU*," "Ae*," and "Ae?." 24,540 objects (34.5% of matches) have "ambiguous" SIMBAD classifications, such as the generic "*" with no other object types, which are insufficient for including or excluding the presence of a protoplanetary disk. 39,677 objects (55.7% of matches) are classified as non-YSOs, based on all other SIMBAD object types. This leaves 577,193 new potential candidate YSOs (92.5% of sources in this AllWISE subset), including those with ambiguous SIMBAD classifications.

The coordinates of all sources in the AllWISE query results are plotted in Figure 2.3, color-coded by SIMBAD type. A distinctly non-isotropic distribution already begins to

emerge even at this intermediate stage of the color selection process, with clear over-densities of potential candidate YSOs near the Galactic plane and in the direction of the Galactic center. Because the Galactic bulge is known to be comprised primarily of old stars (Portegies Zwart et al. 2010), this concentration near the bulge is likely due not to YSOs but rather to old stars with exceptionally red colors, such as red giants and supergiants — note the blatant clump of known non-YSO contaminants in the same direction. These intrinsically red stars represent a population of contaminants that ideally should be eliminated with a second round of color-color cuts. Survey artifact arcs are also visible in the distribution of potential candidate YSOs and should likewise be eliminated with further cuts. The heightened density of potential candidate YSOs in the same regions as groups of known YSOs near the Galactic anti-center is an encouraging sign that the initial phase of the color selection process does show a preference for YSOs and will be successful at identifying new objects of similar types. In contrast to the highly localized distribution of known YSOs recovered through these initial cuts, the generally uniform distribution of known non-YSOs further supports the idea that color-color cuts will be more effective as distinguishing criteria for candidate YSOs than spatial cuts would be.

To make this secondary set of color-color cuts, the same 14 magnitudes as listed in Table 2.2 were retrieved for all sources returned by the AllWISE SQL query — YSOs, potential candidate YSOs, and non-YSOs — and the same iterative two-step process was used to distinguish the known YSOs from the known non-YSOs in 5,460-dimensional color-color space. The first two columns of Figure 2.4 and the left four columns of Table 2.5 demonstrate

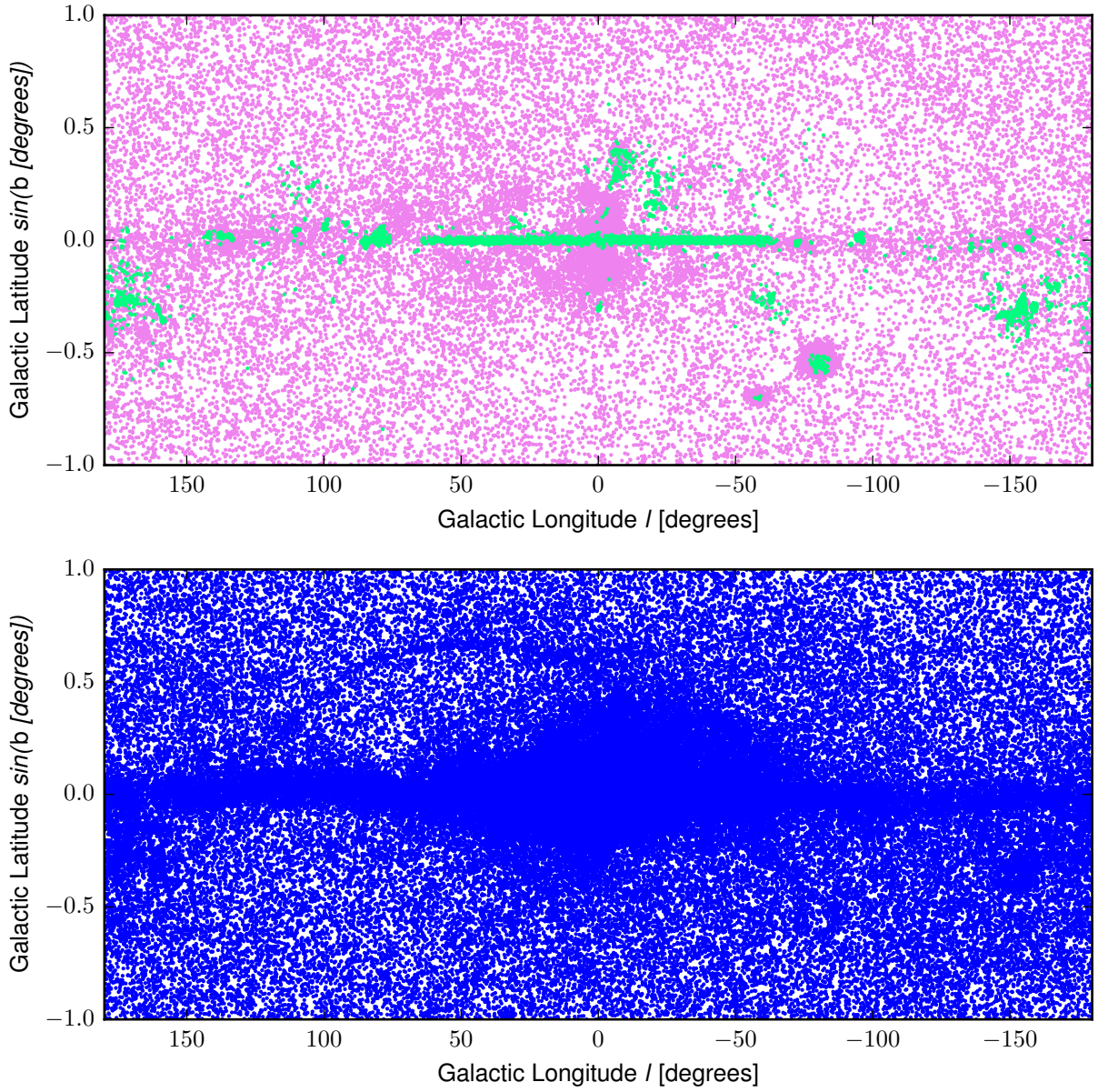

Figure 2.3 — AllWISE sources that satisfy Round 1 of color-color cuts, as specified in Table 2.4. Details on the AllWISE query and on SIMBAD classification are provided in Section 2.5. The top panel shows sources with unambiguous SIMBAD types: mint points represent previously identified YSOs, while fuchsia points represent objects previously identified as non-YSOs. These non-YSOs are contaminants that would lessen the confidence in our color selection process for YSOs — confidence that is restored by the second round of color-color cuts, as described in Section 2.5 and revisited for comparison in Figure 2.5. The bottom panel shows sources (blue) without SIMBAD records or with ambiguous SIMBAD types.[3]

the effect of the first set of color-color cuts made in this second round of refining the unique colors of known YSOs. These are "strict" cuts, as detailed in Section 2.4. Table 2.6 presents a summary of all 13 cuts made in this second round, in the order in which they were made, and the third column of Figure 2.4 reflects the application of all of these cuts, as do the two right-hand columns of Table 2.5.

Changing the focus from stars with known protoplanetary disks in the first round of color selection to known YSOs in the second round is not a trivial shift. Stars with protoplanetary disks are a subset of YSOs, and not all YSOs are expected to have observable disks. To assuage this concern of specificity, an additional criterion was imposed on the potential candidate YSOs: that a candidate source must pass at least one of the nine color-color cuts from Table 2.4, which were tailored specifically to characterize stars with known protoplanetary disks. This restriction eliminates potential candidate YSOs that meet only the more lenient criterion of "not failing" any of the cuts. (The same requirement was made of known YSOs and non-YSOs for comparison; see Section 2.6 below). This additional criterion suggests — but does not guarantee for any given source — that the candidate YSOs presented in the final catalogue adhere not only to colors broadly associated with YSOs but also to colors particularly associated with stars hosting protoplanetary disks.

2.6 Final Catalogue

Applying this second round of color-color cuts to the potential candidate YSOs from AllWISE leaves 199,391 high-confidence sources with candidate protoplanetary disks satisfying all of

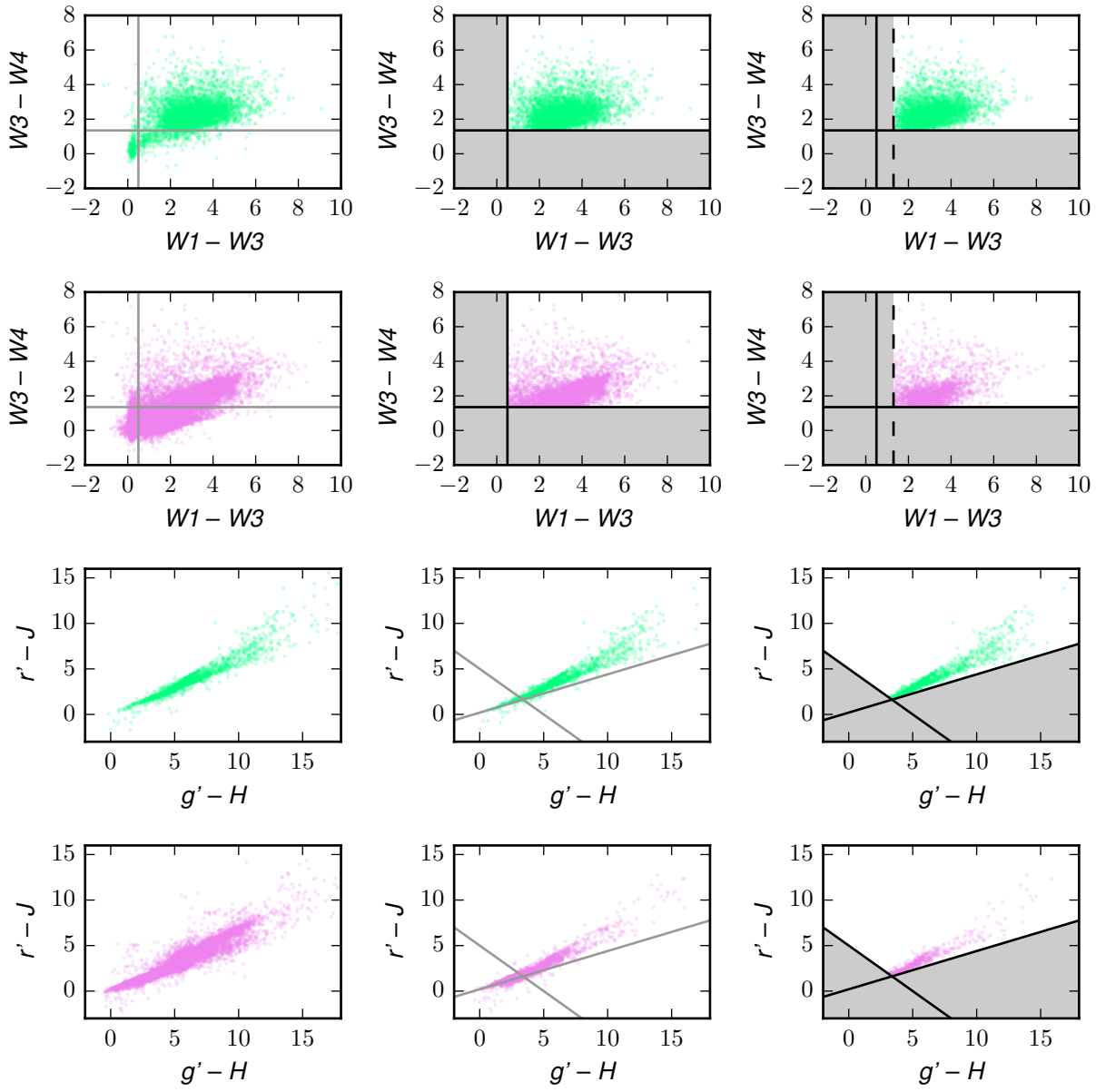

Figure 2.4 — Colors of known YSOs and known non-YSOs from AllWISE, according to SIMBAD type, through the color selection process. Mint points represent known YSOs; fuchsia points, non-YSOs. The first, second, and third columns show sources: prior to any new color-color cuts; after the first two cuts from Table 2.6; and after all 13 additional color-color cuts. Each cut is previewed in gray in its applicable color-magnitude space, then shown as a black line and gray exclusion region. The sixth cut, a stricter one in $W1-W3$ than the second cut is, is shown as a dashed line. See the plotting style notes for Figure 2.2.[(4)]

Table 2.5. KS Statistics Excerpt, Round 2

Initial		After Cut 2		After Cut 13	
Dimension	Z_n	Dimension	Z_n	Dimension	Z_n
$W3-W4$	47.097	$g'-J$	15.136	$W1$	8.389
$W2-W4$	46.591	$r'-J$	15.060	K	7.395
$W1-W4$	45.794	$g'-H$	14.851	$W2$	6.407
$W1-W3$	44.342	$i'-J$	14.703	$W3-W4$	5.932
$K-W4$	44.234	$i'-H$	14.536	$W2-W4$	5.723
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u'-r'$	8.504	$u'-W4$	8.424	$W4$	3.893
$g'-H$	7.976	$W3-W4$	8.325	$g'-H$	3.879
$B-g'$	7.871	$B-V$	8.057	$g'-J$	3.827
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u'-z'$	5.279	$K-W3$	7.254	$g'-z'$	3.400
$r'-J$	4.774	$W1-W3$	7.230	$r'-J$	3.398
$g'-i'$	4.574	$W2-W3$	6.996	$r'-i'$	3.304
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$V-z'$	3.425	$H-W1$	3.274	$V-z'$	1.009
$B-W2$	3.351	$V-W4$	3.265	$B-g'$	0.966
$V-K$	2.305	$H-K$	3.038	$V-i'$	0.930
$B-W1$	2.197	$B-W4$	2.135	$u'-i'$	0.733
B	1.968	B	1.543	$B-i'$	0.709

Note. — An excerpt of the Kolmogorov-Smirnov (KS) statistics comparing the distributions of known YSOs and known non-YSOs in 105 colors and magnitudes. Following Table 2.3, the left, center, and right two-column pairs show KS statistics: prior to making any new color-color cuts; after making the first two cuts listed in Table 2.6; and after making all of these Round 2 cuts, respectively.[(5)]

Table 2.6. Color-Color Cuts, Round 2

Criterion
$W3 - W4 > 1.35$
$[W1 - W3 > 0.5]$
$r' - J > 0.42 \cdot (g' - H) + 0.2$
$r' - J > -1 \cdot (g' - H) + 5$
$W3 - W4 > -1 \cdot K + 9$
$W1 - W3 > 1.3$
$W3 - W4 > -0.375 \cdot W1 + 4.75$
$i' - z' > -0.789474 \cdot (g' - J) + 2.921054$
$K - W3 < 3.278689 \cdot (J - H) + 1.868852$
$i' - z' > 1.272727 \cdot W1 - 16.363634$
$i' - J > 1.008 \cdot K - 11.828$
$g' - W2 > 2.8$
$g' - K > 2.811494 \cdot K - 32.753904$

Note. — Inclusion criteria applied to identify the unique region of multidimensional color-color space occupied by known YSOs (and hence cutting out objects of other known, unambiguous SIMBAD classifications, as described in Section 2.5). In some cases, the included region is defined with lower-dimensional color or color-magnitude cuts. The cuts are listed from top to bottom in the order in which they were made. Any target subjected to this round of cuts must not have failed any of the color-color cuts from Round 1, as given in Table 2.4. Note that there is effectively only one cut in the $W1 - W3$ color; as this is an iterative process, it was found that a stricter cut than the original one could be made.

the criteria introduced above. An excerpt of this final catalogue is shown in Table 2.7. The full catalogue of sources with candidate protoplanetary disks is available with the online version of this journal article. The coordinates of all sources in the final catalogue are shown in the bottom panel of Figure 2.5; the top panel shows the coordinates of known YSOs and non-YSOs from AllWISE that also survive the color selection process, so that their distributions may be compared. Contrasting Figure 2.5 with Figure 2.3 illustrates the effectiveness and necessity of the second round of color-color cuts. As with the first round of cuts, not all of the known YSOs are recovered, and not all of the known non-YSOs are eliminated, though these goals are largely achieved. Of 6,997 known YSOs retrieved in the AllWISE query, 5,279 (75.4%) are recovered after the second round of color-color cuts, while of the 39,677 known non-YSOs retrieved, only 2,978 (7.5%) remain.

Based on the objects of unambiguous SIMBAD types that are recovered from the AllWISE query results after the second round of color-color cuts, a false positive rate of 36.1% is estimated, implying that up to two-thirds of the 199,391 new candidates, or approximately 127,000 sources, may represent stars with genuine protoplanetary disks. The percentage of control field sources that survive the first round of cuts, prior to the AllWISE and SIMBAD queries, is not factored into this false positive estimate: control field sources that are in AllWISE and are not by chance stars with protoplanetary disks will not have been classified as YSOs and may have already been classified as non-YSOs. In the former case, the would-be false positives will have been caught by the SIMBAD query; in the latter case, they will have been subsumed into the false positive rate from non-YSOs as quoted above. Errors in

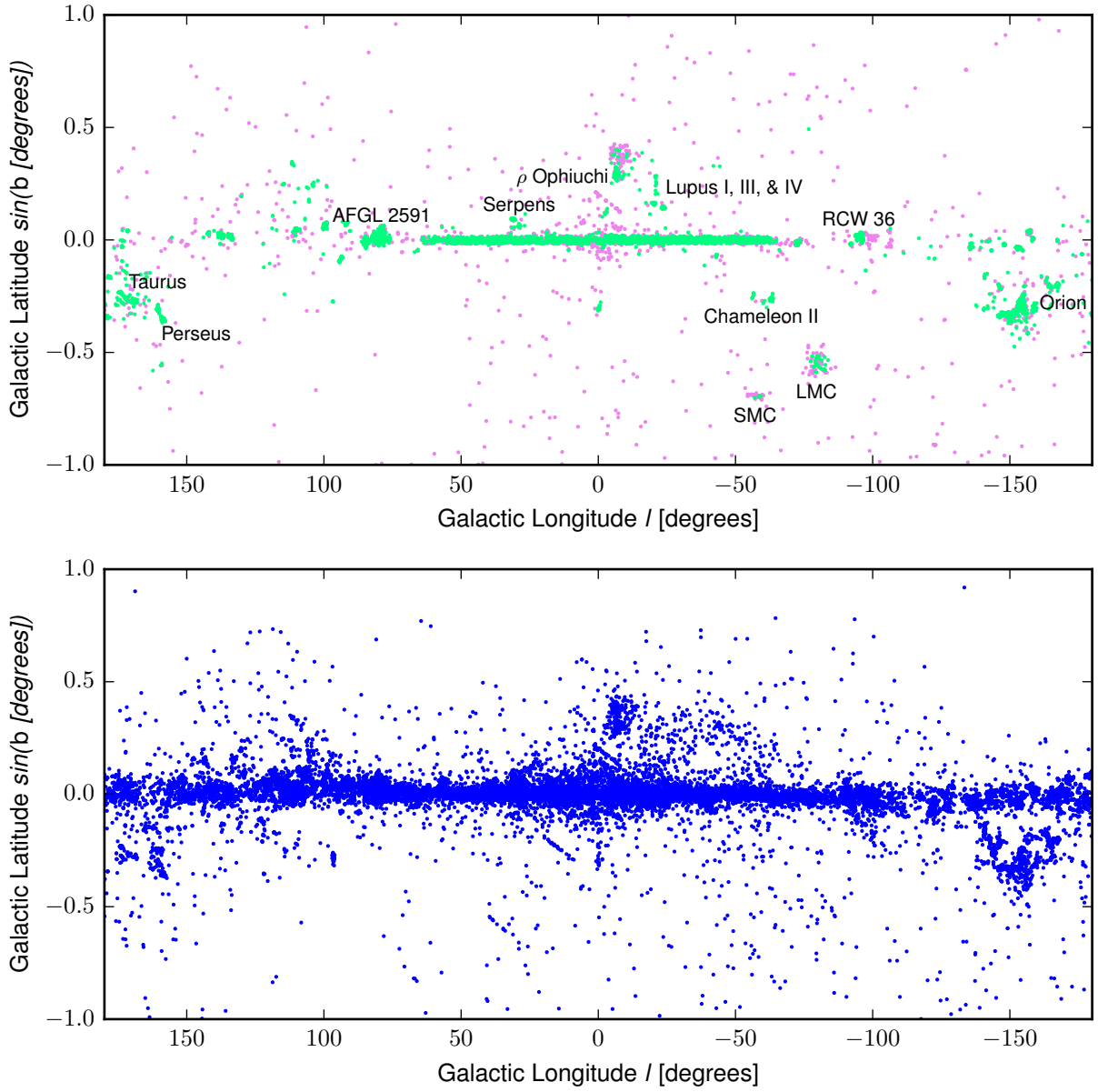

Figure 2.5 — AllWISE sources that satisfy Rounds 1 and 2 of color-color cuts, including the requirement to pass at least one of the Round 1 cuts. As in Figure 2.3, the top panel shows sources with unambiguous SIMBAD types (YSOs in mint and non-YSOs in fuchsia). The bottom panel shows sources (blue) with candidate protoplanetary disks, as presented in the final catalogue. Multiple previously known star-forming regions are successfully recovered by this process and are labeled in the top panel. Numerical designations in the Lupus and Chameleon complexes follow the designations in Evans, II et al. (2003). Some features presumed to be survey artifacts are still visible in the set of candidates, but the overall effect of these artifacts appears to be small.[6]

Table 2.7. Final Catalogue Excerpt

ID (AllWISE)	g'	r'	i'	J	H	K	$W1$	$W2$	$W3$	$W4$
J000006.45+675349.7	15.731	14.491	13.786	11.856	11.096	10.525	9.627	9.159	6.953	4.498
J000103.31+470122.7	14.417	13.333	12.787	11.004	10.166	9.630	9.087	8.685	6.572	4.655
J000124.69+674050.0	14.536	13.589	12.991	11.264	10.908	···	10.059	9.688	8.137	6.189
J000126.82+673749.9	16.833	15.121	14.048	11.441	10.689	10.341	10.031	9.948	8.593	···
J000139.32+644136.4	16.347	15.216	14.484	13.259	12.501	12.027	10.546	10.060	7.843	5.778
J000139.64-194949.7	···	···	···	17.146	···	14.791	14.426	13.509	10.591	8.571
J000229.33+672049.9	16.874	14.867	13.686	11.260	10.457	9.948	9.317	8.938	7.121	2.778
J000710.77+672122.4	14.966	13.417	12.553	10.642	10.174	9.945	9.695	9.673	7.671	4.692
J000749.49+610314.1	15.130	14.594	14.314	12.677	11.738	10.890	9.854	9.345	7.187	5.238
J000928.31+674803.5	16.102	14.684	13.423	11.416	10.779	10.507	10.380	10.254	8.835	6.370
J001138.58+615357.4	15.328	14.444	13.905	12.281	11.839	11.469	10.728	10.168	8.290	6.722
J001230.40+653305.2	14.550	13.245	12.536	10.340	9.347	8.428	7.086	6.420	4.287	1.994
J001245.74+591932.8	17.304	16.351	15.832	13.737	12.702	11.805	11.076	10.228	6.575	4.257
J001956.46+525100.1	15.144	14.264	13.815	12.233	11.465	10.892	10.077	9.571	7.534	4.050
J002056.08+525120.9	16.190	15.152	14.725	13.114	12.329	11.842	10.976	10.494	8.359	5.715
J002114.42+663528.2	12.403	11.246	10.491	8.908	8.563	8.393	8.273	8.234	6.773	3.238
J002704.12+653348.5	16.396	14.978	14.130	11.748	10.665	9.919	9.055	8.377	6.457	4.262
J002843.45+652507.3	15.807	14.432	13.691	11.519	10.643	10.000	9.136	8.566	6.130	4.102
J003000.90+652553.2	15.323	14.167	13.545	11.794	10.991	10.392	9.463	8.942	7.295	5.631
J003035.47+652958.1	16.346	15.045	14.225	12.423	11.319	10.528	9.665	9.238	7.057	5.010

Note. — An excerpt of columns and rows available in the final catalogue of 199,391 sources with candidate protoplanetary disks. Because all sources in the final catalogue must be present in AllWISE, per the SQL query from Section 2.5, the WISE identifiers provide convenient names for the candidates presented here.[(7)]

SIMBAD types are possible, but they are presumed to be marginal and not to be biased towards either YSO or non-YSO misclassification.

Possible contaminants contributing to this false positive rate include quasars at high Galactic latitudes and objects severely affected by interstellar reddening at low Galactic latitudes. If a spatial cut were to be made at this final step that excluded likely extragalactic sources outside of Galactic latitudes $-30° < b < 30°$, the false positive rate would improve to 34.2% (from 5,243 known YSOs and 2,726 known non-YSOs recovered). This would imply that of 199,137 remaining new candidates, 131,000 would be real protoplanetary disks — but such a search would no longer be all-sky.

That said, the distribution presented in Figure 2.5 presents several visual features that encourage confidence in this catalogue of new candidate protoplanetary disks. Note the near-complete elimination of the Galactic bulge, which is known to be comprised primarily of old stars (Portegies Zwart et al. 2010), from Figure 2.3 to Figure 2.5. Despite this elimination of the bulge, note the retention of sources in the young Galactic disk across all longitudes, even in the direction of the Galactic bulge. Additionally, note the elimination of most high-latitude sources: these objects are presumed to be spatially removed from the dynamic environment of the Galactic disk, where most star formation occurs (Portegies Zwart et al. 2010).

A surprisingly high false negative rate of 66.8% is estimated for this new catalogue. The bulk of these false negatives are not a result of either round of color-color cuts: 4.8% of confirmed protoplanetary disks are wrongly excluded by the first round of cuts; 53.7% of the

remaining confirmed disks are not present in the AllWISE SQL query results; and 24.6% of known YSOs from the AllWISE query results are excluded by the second round of cuts. No known YSOs are excluded by the final criterion of passing at least one of the first round cuts. The false negative rate from the color-color cuts (both rounds) alone is therefore 28.2%, which is comparable to the false positive rate. A detailed accounting of how the first round of color-color cuts affects all 1,126 sources in the initial sample of stars with confirmed protoplanetary disks is as follows:

- 54 (4.8%) do not survive the first round of color-color cuts;
- 7 (0.6%) do not have counterpart sources in AllWISE;
- 569 (50.6%) fail the AllWISE SQL query for one or more of the following reasons:
  - 127 (11.3%) are flagged in AllWISE as being extended sources;
  - 0 (0.0%) do not have at least one measured $W1$ or $W2$ magnitude, regardless of measurement quality;
  - 17 (1.5%) do not have at least one J, H, or K magnitude from an associated 2MASS source in AllWISE;
  - 72 (6.4%) fail one or more of the first round cuts, allowing for incomplete photometry when necessary;
  - 6 (0.5%) do not have good AllWISE photometric quality flags in $W1$ and/or $W2$;

  - 509 (45.2%) do not have good AllWISE contamination flags in $W1$ and/or $W2$, primarily due to diffraction spikes or halos from other bright sources; and

- 496 (44.0%) are successfully recovered by the AllWISE SQL query.

Figure 2.6 (in concert with the lower panel of Figure 2.5) illuminates a partial explanation of the high false negative rate and of the identification of previously unreported disks in known star-forming regions: the stars harboring these new candidate protoplanetary disks tend to be fainter sources in crowded areas, meaning that they will have been de-prioritized in targeted searches (in favor of brighter objects) and overlooked in automated surveys because of potential field contamination (with most automated surveys being optimized for comparatively sparser fields). While the new sources with candidate protoplanetary disks occupy similar color and magnitude ranges as stars with known protoplanetary disks do — as expected for a classification process based on selecting characteristic colors — fainter objects are much better represented among the new candidates. These fainter sources may be more distant or lower-mass stars (or lower-mass disks) than the previously confirmed protoplanetary disk host stars are. Indeed, the similar ranges of colors and magnitudes covered by the new candidates and the stars with known protoplanetary disks emphasize the importance of identifying these fainter sources that should be observable with current instrumentation but may not have stood out as worthwhile targets before.

One possible approach for mitigating the false negative rate would be to relax the contamination threshold for $W1$ and $W2$, but this change would likely incur the cost of a significantly higher false positive rate. As an example, consider the case of removing any

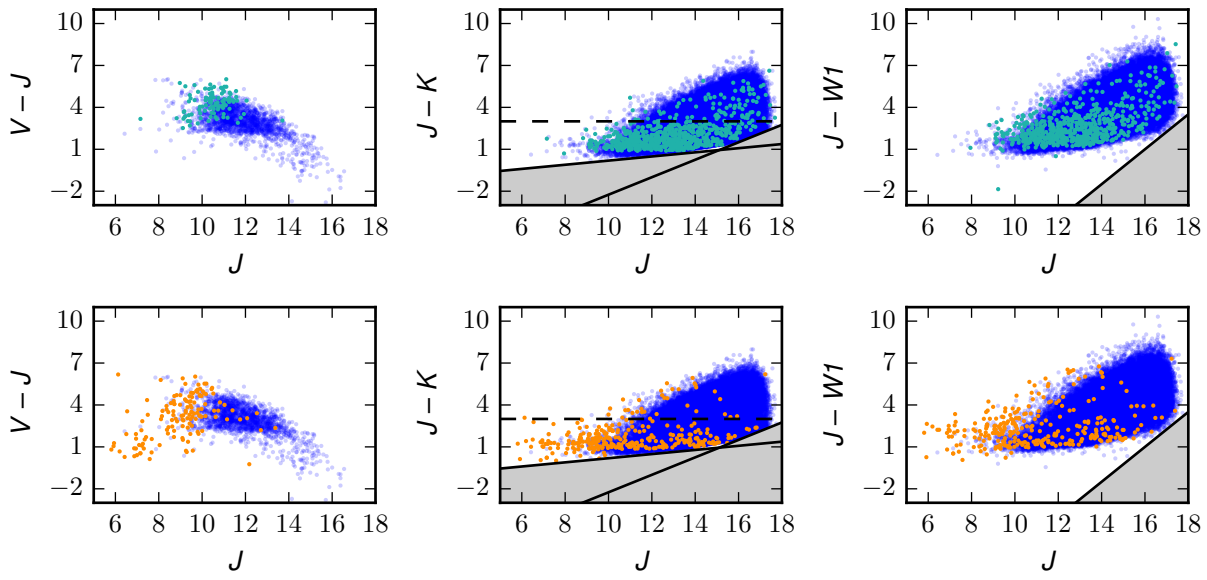

Figure 2.6 — Colors of stars with known protoplanetary disks, from the literature sources given in Table 2.1, and of sources with candidate protoplanetary disks in the final catalogue, as described in Section 2.6. Blue points mark these new candidates. Turquoise points show sources with known protoplanetary disks that would survive the color selection process, including the color-color cuts listed in Tables 2.4 and 2.6, the photometric quality criteria of the SQL query as described in Section 2.5, and the requirement to pass at least one of the Round 1 cuts. Tangerine points show known disks that would not survive these color-color cuts. All three populations occupy similar ranges of J magnitudes and $V - J$, $J - K$, and $J - W1$ colors, but fainter objects ($J > 10$) are much better represented among the new candidates. This skewed representation is expected: confirming circumstellar disks is an observationally intensive endeavor best suited to bright objects, as described in Section 2.2, and brighter objects are more likely to saturate in all-sky surveys such as AllWISE. Any cuts (from Round 1 or 2) made in these color-magnitude spaces are marked with solid black lines and gray exclusion regions. The dashed black line represents a hypothetical cut made at $J-K = 3$ to mitigate possible interstellar (as opposed to circumstellar) dust reddening above this line, as discussed in Section 2.6. Not all known protoplanetary disks and new candidates are shown: as in Figures 2.2 and 2.4 above, a source must have reasonably good photometry (see Table 2.2 for criteria) in both of the magnitudes used for each color-magnitude space in order to be plotted.

constraint on the contamination flag from the AllWISE SQL query. This revised all-sky query returns 1,621,218 results, compared to 623,867 AllWISE point sources from the query as described in Section 2.5. Under the assumptions that:

1. a similar proportion of these sources would be potential candidate YSOs (92.5%), on the basis of either not being present in SIMBAD or having ambiguous SIMBAD classifications,

2. the second round of color-color cuts would be equally as effective, with 75.4% of known YSOs (on the basis of SIMBAD types) and 34.5% of potential candidates surviving, and

3. the false positive and false negative rates given above for the final catalogue are reasonably accurate,

the alternative catalogue would have 518,000 candidates, with projected false positive and false negative rates of 54.7% and 38.8%, respectively. Compared to the catalogue presented in this paper — with 199,391 candidates and false positive and false negative rates of 36.1% and 66.8%, respectively — removing the contamination criterion would net 160% more candidate protoplanetary disks and decrease the false negative rate by 42%, at the expense of increasing the false positive rate by an untenable 52%. While these projections are only very rough estimates, the projected false positive rate of over 50% from removing the contamination criterion demonstrates the motivation for selecting the candidate YSOs of the highest quality,

acknowledging the high false negative rate and leaving the presumably missed detections to be the subject of later surveys.

The focus here on the best-quality candidates is intentional. With a relatively low false positive rate, only a handful of candidates would be needed in order to have a good chance of confirming multiple new protoplanetary disks with a modest observing campaign. While the false positive rate is high enough to be a non-trivial caveat, the on-sky distribution in Figure 2.5 would allow an observer to avoid suspicious candidates if so inclined, such as those buried in the Galactic plane or near visually apparent survey artifacts. At the same time, the catalogue is broad enough to allow an observer to impose unconventional or exacting criteria related to a specific program of study and still have a number of candidate protoplanetary disks available as acceptable targets.

As noted for the known protoplanetary disks in Figure 2.1, the clumping of the known YSOs is expected, due to the observationally intensive nature of confirming protoplanetary disks with small, targeted campaigns; indeed, a number of previously identified star-forming regions are recovered by the method presented here. These regions are labeled in Figure 2.5. Some of these regions, such as the Taurus Molecular Cloud and Chameleon II, appear to have been reasonably well-surveyed thus far: most of the objects recovered by the color selection process presented here have already been confirmed as YSOs, and our survey reveals relatively few new candidate YSOs in these areas. Other regions, such as Perseus and Orion, appear to have many more new candidate YSOs; these new targets present immediate opportunities for follow-up observation.

The clumping of candidates near the Galactic plane is expected as well, due to optical extinction sporadically obscuring regions and interstellar reddening artificially adding potential false positives. YSOs, particularly those with protoplanetary disks, are intrinsically dusty objects that are usually embedded in inherently dusty star-forming regions. Indeed, the increasing, linear correlation of $J - K$ as a function of J for the candidates presented in Figure 2.6 suggests that some candidates do suffer from interstellar reddening — though this does not preclude these candidates also showing reddening from dusty circumstellar disks, as evidenced by the good (albeit sparser) overlap with known protoplanetary disks in this color-magnitude space. This ambiguous reddening would be easy for a user of this catalogue to mitigate if so inclined, such as by making an additional cut to require J and K magnitude measurements of reasonable quality and exclude sources with $J - K > 3$. Making this cut would leave 31,393 sources with candidate protoplanetary disks, and the revised false positive rate of 33.0% would imply that 21,000 of these candidates would have genuine protoplanetary disks.

In addition to using this new catalogue for statistical analyses of planetary formation processes, studies of the spatial distribution of nearby dust could be an interesting application of this dataset, especially if a significant number of candidates have distances in the forthcoming Pan-STARRS or GAIA data. Such distances could also enable the calculation of a de-reddened, mid-IR spectral index from the available magnitudes without the need for further observations, which would permit the immediate confirmation (or rejection) of some of the sources with candidate protoplanetary disks.

An attempt was made to retrieve trigonometric parallaxes from Hipparcos (ESA 1997) for all candidates and known YSOs, but the overlap with the Hipparcos catalogue is poor: 0 known YSOs and 2 candidates. One possible explanation for this lack of overlap could be the saturation limits imposed on the quality criteria flags of large surveys, whereas Hipparcos focused primarily on bright targets (Perryman et al. 1997). Overlap with the first data release from GAIA (GAIA Collaboration et al. 2016) is similarly unsatisfactory — 3 known YSOs and 29 protoplanetary disk candidates with measured trigonometric parallaxes — but this coverage is expected to improve with future GAIA data releases that cover fainter objects.

As noted above, this work does not attempt to correct any biases present in the surveys used to aggregate the magnitudes. While these amalgamated biases are important to be aware of before making any fundamental, statistical proclamations, they also highlight a strength of this method: the candidate protoplanetary disks presented here are ripe for confirmation today, using only currently (or previously) available instrumentation.

Furthermore, the method presented in this work may itself be considered iterative: in a few years' time, once more protoplanetary disks have been confirmed, this method may be repeated in order to refine the list of sources with candidate protoplanetary disks. It is therefore critical for all observers to publish the results of any targeted investigations of specific protoplanetary disk candidates. Both detections and non-detections are useful for isolating the characteristic colors of stars hosting protoplanetary disks, and unpublished observations of promising candidates that do not turn out to have protoplanetary disks do not need to be duplicated. The catalogue that is presented here as a springboard will

enable the astronomical community to focus its resources more efficiently when investigating circumstellar disk dynamics and planetary formation processes. The authors hope that the community will take up these follow-up confirmation studies with vigor and that a subsequent reevaluation of this catalogue will occur.

2.7 Comparison with Recent Literature

Marton et al. (2016) recently presented a different method for finding characteristic colors of a sample of previously confirmed Class II YSOs (those hosting protoplanetary disks) and applying such criteria to all-sky photometric surveys. Their alternative methodology provides a useful counterpoint to the work presented here, and examining the differences in our two sets of results is worthwhile. The intersection of our two sets of sources with candidate protoplanetary disks is shown in Figure 2.7. This intersection is surprisingly small, representing only about 5% of either catalogue. Of the 133,980 candidate YSOs presented by Marton et al. (2016) and the 199,391 candidates presented here, only 8,652 sources are common to both catalogues, leaving 190,739 candidates in our new catalogue that are not selected by Marton et al. (2016) and 125,328 sources from Marton et al. (2016) that we do not recover as having candidate protoplanetary disks.

Details of how support vector machines (SVM), a machine learning technique, may be applied to YSO identification are provided by Marton et al. (2016). One immediate difference is that the SVM technique as implemented requires good photometry in all four WISE bands, whereas our technique can handle incomplete photometry in one or more of the 14

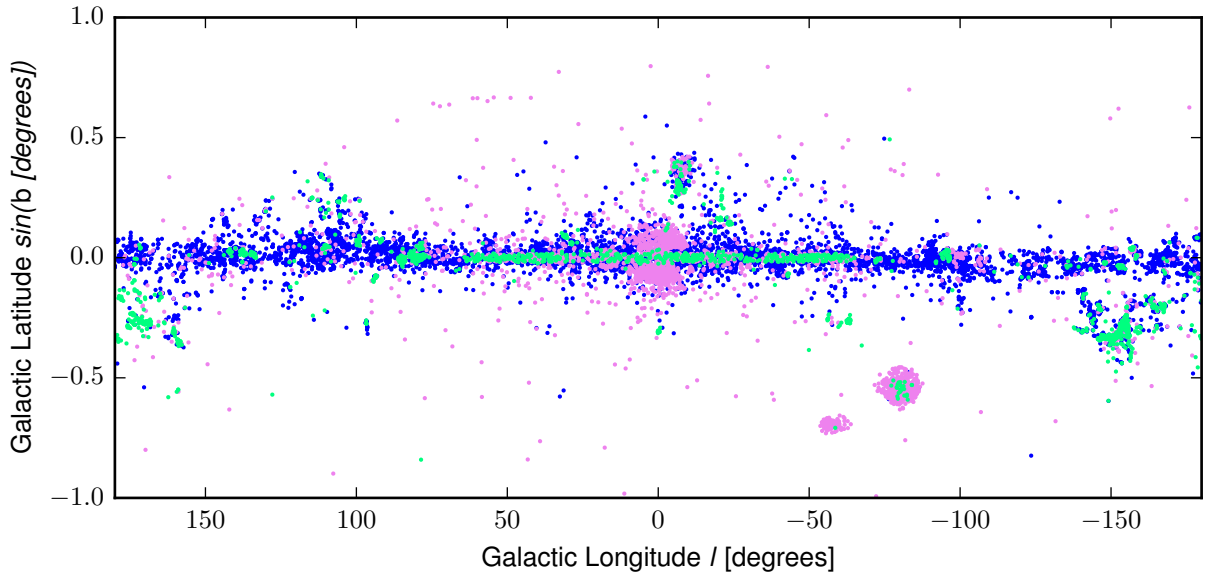

Figure 2.7 — Overlap between the sources with candidate protoplanetary disks presented in this work, as shown in Figure 2.5, and the candidates presented in Marton et al. (2016). Blue points represent candidates present in both catalogues; mint and fuchsia points represent candidates presented in Marton et al. (2016) that we classify as previously known YSOs or non-YSOs, respectively, on the basis of SIMBAD records. Of the 133,980 candidates presented in Marton et al. (2016), this work classifies 1,669 sources (1.2%) as known YSOs, 3,635 sources (2.7%) as known non-YSOs, and 8,652 sources (6.5%) as having candidate protoplanetary disks.

wavebands that we consider. Given the questionable quality of W4 photometry for all except the brightest (but non-saturated) stars, this requirement or, alternatively, flexibility could lead to significant differences in selecting a sample of stars with acceptable photometry, identifying characteristic YSO colors, and applying these criteria on an all-sky scale.

Differences in the on-sky spatial distributions of our two catalogues of sources with candidate protoplanetary disks are also immediately apparent (see Marton et al. (2016), Figure 6). The Marton et al. (2016) selection retains many targets in the direction of the Galactic bulge,

which is known to contain primarily old stars (Portegies Zwart et al. 2010) and as such is not expected to undergo significant star formation in the present epoch.

Based on the numbers of various SIMBAD types recovered from the Marton et al. (2016) selection process as provided in their supplementary data, we calculate a 56.5% false positive rate using our methodology above. We note the discrepancy between this false positive rate and the significantly smaller rates presented by Marton et al. (2016) (1% - 10%, per their Section 4.1). Their false positive calculations are detailed extensively; we only re-calculate it using our methodology in order to make a direct comparison between our two catalogues. Additionally, we note that some of the Marton et al. (2016) candidates are on our lists of known YSOs and non-YSOs, as shown in Figure 2.7; this divergence may be due to our choice to use all published SIMBAD types for an object, not only its primary classification.

SVM is a relatively recent development compared to color-color cuts, but the novelty in our approach lies in the aggregation of a large set of multi-band photometry and in making cuts iteratively, not in the concept of making the color-color cuts themselves. It is unclear why the Marton et al. (2016) catalogue of candidate YSOs differs to such an extent from the catalogue presented here. One issue with the SVM method is that the specific classification criteria are not provided, and hence the selection process employed by Marton et al. (2016) cannot be replicated unless a version of their code is made available. In addition to constructing our training samples differently, our approach leaves a higher level of control over the classification criteria up to the individuals making the color-color cuts — with

the resulting cuts clearly stated (Tables 2.4 and 2.6) and easily adoptable for use by the community at large.

2.8 Conclusions

A list of candidate protoplanetary disks extracted from the AllWISE infrared photometric catalogue is presented. From a sample of known protoplanetary disks, a series of nine color-color cuts was made to isolate the stars with known protoplanetary disks from a sample of control field sources (assumed to have little ongoing star formation) in a 5,460-dimensional color-color space. These exclusion criteria were applied in a query of point sources in the AllWISE catalogue, and SIMBAD object types were retrieved for the results. Thirteen further color-color cuts were made to isolate known young stellar objects from known non-young stellar objects, and these cuts were applied to the AllWISE results that were either not present in SIMBAD or had ambiguous classifications. The resulting catalogue of 199,391 candidate young stellar objects has an estimated false positive rate of 36.1%, suggesting the potential to follow up and confirm approximately 127,000 previously unidentified protoplanetary disks.

2.9 Acknowledgements

This research made use of the VizieR catalogue access tool, the SIMBAD database, and the cross-match service provided by CDS, Strasbourg, France.

This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013).

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is `http://www.sdss3.org/`. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

This research was made possible through the use of the AAVSO Photometric All-Sky Survey (APASS), funded by the Robert Martin Ayers Sciences Fund.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology. WISE is funded by the National Aeronautics and Space Administration.

This research made use of the NASA / IPAC Infrared Science Archive, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Notes

[1] Original caption, Table 2.3: "An excerpt of the Kolmogorov-Smirnov (KS) statistics calculated to compare the distributions of stars with known protoplanetary disks and control field sources in 105 colors and magnitudes. A higher value implies a greater separation between the population of known protoplanetary disks and that of control field sources in that color or magnitude dimension. KS statistics prior to defining any inclusion regions, or "cuts," are given in the left two columns of the table. The center two columns show the KS statistics recalculated for the surviving sources after making the first three color-magnitude cuts, as detailed in Section 2.4 and listed in Table 2.4. Note that the Z_n values are the test statistics themselves, not probabilities derived from the test statistics. Variations in sample sizes are taken into account. The ordering of the values, rather than the values themselves, is used to determine which of the 5,460 available color-color and color-magnitude plots will show the most separation between known protoplanetary disks and control field sources. In some instances, an insufficient number of sources with reasonable-quality photometry (see Table 2.2) survive to recalculate a particular KS statistic after making a new cut; each of these cases, such as $u' - B$ after the third cut, is marked with a dash. KS statistics recalculated after making all nine color-magnitude cuts listed in Table 2.4 are shown in the two right-hand columns. Note that the distributions of known protoplanetary disks and control field sources now show significant separation in none of the color or magnitude dimensions, with the possible exception of J, as addressed in the text."

[2] Original caption, Figure 2.2: "Colors of stars with known protoplanetary disks and sources from control fields through the color selection process from Section 2.4. Turquoise points in the first and third rows represent known protoplanetary disks. Gold points in the second and fourth rows represent sources from control field regions presumed to have little ongoing star formation. The first column shows all sources, prior to any color-color cuts. The second column shows sources remaining after the first three cuts from Table 2.4, all of which are in the ($W2$, $W1 - W2$)-plane; these cuts are marked with solid black lines, with the excluded regions shaded in gray. After three cuts, 82.2% of control field sources are eliminated, while 96.2% of known protoplanetary disks are retained. Not all sources are shown: a source must have reasonably

good photometry (see Table 2.2) in $W1$ and $W2$ in order to be plotted in the top two rows and to be affected by any of the first three cuts. Although the first cuts almost completely eliminate the control field sources with decent $W1$ and $W2$ photometry, the intersecting (H, $J-K$)-plane — which requires good photometry in J, H, and K but makes no such requirement of $W1$ or $W2$ — illustrates the necessity of further cuts and the motivation for making such cuts iteratively. KS statistics from Table 2.3 suggest that the next most productive cut will be made in this plane. The gray line in this latter color-magnitude dimension previews where the fourth cut will be made. The third column shows sources remaining after all nine color-magnitude cuts from Table 2.4. After nine cuts, 98.0% of control field sources are eliminated, while 95.2% of known protoplanetary disks are retained. All plotted points are 80% transparent in order to highlight the densest regions of sources in these color-magnitude spaces."

[3] Original caption, Figure 2.3: "AllWISE sources that satisfy Round 1 of color-color cuts, as specified in Table 2.4. Details on the AllWISE query and on SIMBAD classification are provided in Section 2.5. The top panel shows sources with unambiguous SIMBAD types: mint points represent previously identified YSOs, while fuchsia points represent objects previously identified as non-YSOs. These non-YSOs are contaminants that remain after the first round of color-color cuts (Section 2.4) and are the motivation behind designing a second round of cuts (Section 2.5), the results of which are shown in Figure 2.5. The bottom panel shows sources (blue) without SIMBAD records or with ambiguous SIMBAD types. This latter sample of sources that have not been identified definitively as YSOs or as non-YSOs is almost certainly a mixture of both types of objects."

[4] Original caption, Figure 2.4: "Colors of known YSOs and known non-YSOs from AllWISE, according to SIMBAD type, through the color selection process. Mint points in the first and third rows represent known YSOs; fuchsia points in the second and fourth rows represent known non-YSOs. The first column shows all sources, prior to any new color-color cuts, though all sources retrieved in the AllWISE query must not fail any cut from Table 2.4. The second column shows sources remaining after the first two cuts from Table 2.6, which are marked with solid black lines and gray exclusion regions. After two cuts, 81.9% of known non-YSOs are eliminated, while 87.0% of known YSOs are retained. One noteworthy effect of these

two cuts is the removal of probable saturation artifacts near $W1 - W3 = 0$ and $W3 - W4 = 1$. Similarly to Figure 2.2, a source must have reasonably good photometry in $W1$, $W3$, and $W4$ in order to be plotted in the top two rows and to be affected by either of these first two new cuts. No source is cut solely on the basis of missing or poor-quality photometry, as is especially relevant in $W4$. The gray lines in the bottom two rows preview the third and fourth new cuts as suggested in Table 2.5, both of which are in the ($g' - H$, $r' - J$)-plane. This intersecting plane requires good photometry in g', r', J, and H but places no restrictions on $W1$, $W3$, or $W4$. The third column shows sources remaining after all 13 color-color cuts from Table 2.6. Due to the iterative nature of this color selection process, it was found that a stricter cut than the original one could be made in the $W1 - W3$ color, as noted for the second row of Table 2.6; this sixth cut is marked with a dashed black line. After 13 cuts, 92.5% of known non-YSOs are eliminated, while 75.4% of known YSOs are retained. All plotted points are 80% transparent."

[5] Original caption, Table 2.5: "An excerpt of the Kolmogorov-Smirnov (KS) statistics comparing the distributions of known YSOs and known non-YSOs in 105 colors and magnitudes. Following Table 2.3, the Z_n values are test statistics, not probabilities, and a higher value implies a greater separation between the distributions of known YSOs and known non-YSOs, although it is the ordering of these values, not the values themselves, that are used to determine which of the 5,460 available color-color and color-magnitude plots will present the best opportunities for making cuts. In the left, center, and right sets of two columns, KS statistics are shown prior to making any new color-color cuts; after making the first two cuts listed in Table 2.6; and after making all of these Round 2 cuts, respectively. Note that any object subjected to this color selection process must not have failed any of the nine cuts specified in Table 2.4, as detailed in Sections 2.4 and 2.5. After all 13 Round 2 cuts, the distributions of known YSOs and non-YSOs do not show significant separation in any of the color or magnitude dimensions."

[6] Original caption, Figure 2.5: "AllWISE sources that satisfy Rounds 1 and 2 of color-color cuts, including the *a posteriori* requirement to pass at least one of the Round 1 cuts, which were tailored specifically to known protoplanetary disks, as opposed to being tailored more generally to known YSOs. As in Figure 2.3, the top panel shows sources with unambiguous SIMBAD types (YSOs in mint and non-YSOs in fuchsia), while

the bottom panel shows sources (blue) without SIMBAD records or with ambiguous SIMBAD types. These latter points are the sources with candidate protoplanetary disks. An overall false positive rate of 36.1% is estimated from the retention ratio of known non-YSOs to all sources with unambiguous SIMBAD types, implying that of the 199,391 new candidates presented, approximately 127,000 are sources with genuine protoplanetary disks. Multiple previously known star-forming regions are successfully recovered by this process and are labeled in the top panel. Numerical designations in the Lupus and Chameleon complexes follow the designations in Evans, II et al. (2003). Some features presumed to be survey artifacts are still visible in the set of candidates, such as the diagonal, linear pattern near Galactic longitudes $40° > l > -10°$, but the overall effect of these artifacts appears to be small. Other large patterns of candidates — such as the loop structures near ($l = 120°$, $b = 12°$) and ($l = -40°$, $b = 12°$) — may be real, based on the known YSOs in those regions. Seemingly sparse regions in the Galactic disk may be obscured due to intervening dust clouds."

[7] Original caption, Table 2.7: "An excerpt of columns and rows available in the final catalogue of 199,391 sources with candidate protoplanetary disks. Because all sources in the final catalogue must be present in AllWISE, per the SQL query from Section 2.5, the WISE identifiers provide convenient names for the candidates presented here. In addition to being present in AllWISE, each of the sources chosen for this excerpt also has a corresponding record in SDSS or APASS (or both). The following columns include any magnitudes aggregated via the data acquisition process described in Section 2.3. All magnitudes used in each round of color-color cuts, as listed in Tables 2.4 and 2.6, are included in this excerpt. These 10 filters, as listed by survey in Table 2.2, are arranged in order of increasing central wavelength (from the bluest, g', to the reddest, $W4$). When two of the surveys provided high-quality photometry in the same waveband for a given source, an average of the two magnitudes was taken. Due to photometric quality constraints and other limitations specific to each survey, as described in Table 2.2 and the references therein, not all sources in the final catalogue will have magnitude measurements in all wavebands."

CHAPTER 3

Commentary

As a result of many thoughtful conversations with advisors and colleagues, a brief commentary to provide clarification and further insight into some parts of the journal article (Chapter 2) seems constructive. Again, as noted in the preface (Chapter 1), publication of the journal article is still pending resolution of the peer review process, and the final, published version may include significant revisions from the material presented here.

In addition to searching for stars hosting protoplanetary disks specifically, this research also provides a survey of young stellar objects (YSOs) generally. Such a survey is important because it should not be biased towards the brightest YSOs or towards YSOs in known star-forming regions, both of which have guided previous surveys. (See, for instance, the "Cores to Disks" survey described in Evans, II et al. (2003), which studied YSOs extensively — but only for a few known star-forming regions, because the *Spitzer* observatory was not designed for all-sky campaigns.) The research presented here attempts to perform a more systematic search for YSOs. Many of the targets remaining after the color selection process are faint objects (see Figure 2.6), which are likely to be more common than brighter objects are. Importantly, once distances from GAIA are published for more of the new candidate YSOs, this catalogue should be able to identify star-forming regions spatially, thereby defining not only the rough boundaries of these regions but also revealing any isolated YSOs that are no longer embedded in their traditional environments. This broad relevance to YSOs and their distributions in space, brightness, and color suggest significantly more avenues of research

that may be pursued with targets selected from this catalogue, rather than focusing narrowly on protoplanetary disks.

One reason why this catalogue is hoped to be widely applicable is the breadth of stars hosting known protoplanetary disks. Indeed, different papers from Table 2.1 use different techniques to identify disks, meaning that these disks may include a diverse set of objects and "known to have a disk" is not a standardized criterion across all papers listed. Mannings & Sargent (1997) and Mannings & Sargent (2000), for instance, adopt criteria related to millimeter emission rather than infrared excess, and Scholz et al. (2006) goes even farther afield by investigating millimeter emission around brown dwarfs instead of protostellar objects. While having a broad sample from which to characterize a range of protoplanetary disks is advantageous when searching for previously unidentified disks, this approach brings its own caveats, such as the possible inclusion of debris disks in the catalogue presented here. In contrast to younger, warmer protoplanetary disks, in which the circumstellar dust is heated by processes related to the formation of the protostar, debris disks are older and colder, containing dust produced by the collisions of planets or other large objects — but a debris disk around a more massive star may be heated by the star itself to temperatures comparable to those of protoplanetary disks. Such degeneracy in observed properties is likely unavoidable in a large survey without follow-up observations of individual targets, but users of the catalogue should be aware of this caveat.

The central wavelengths of the filters used in the data aggregation process (see Table 2.2) are listed for reference in Table 3.1. As noted in Section 2.3, not all targets (stars

Table 3.1. Central Wavelengths of Filters

Filter	Wavelength (microns) by Survey				
	SDSS	APASS	Tycho-2	2MASS	AllWISE
u'	0.354	···	···	···	···
B	···	0.442	0.422	···	···
g'	0.475	0.475	···	···	···
V	···	0.540	0.527	···	···
r'	0.622	0.622	···	···	···
i'	0.763	0.763	···	···	···
z'	0.905	···	···	···	···
J	···	···	···	1.25	···
H	···	···	···	1.65	···
K	···	···	···	2.15	···
$W1$	···	···	···	···	3.4
$W2$	···	···	···	···	4.6
$W3$	···	···	···	···	12
$W4$	···	···	···	···	22

Note. — Central wavelengths of the filters listed in Table 2.2. As described in Section2.3, when a single target had reasonable-quality photometry in the same filter from two different surveys, an average of the two measured magnitudes was taken for that filter for that source.

with known disks, control field sources, known YSOs in AllWISE, known non-YSOs in AllWISE, and potential candidate YSOs in AllWISE, including the final set of candidates) have reasonable-quality photometry (see Table 2.2) in all wavebands queried; completeness by filter is quantified in Table 3.2. Coverage in the optical bands is noted to be remarkably poor, suggesting a possible area for improvement when this research is refined.

Table 3.2. Photometric Completeness

Filter	Known Protoplanetary Disks			Control Field Sources		Known YSOs		Known Non-YSOs		Candidate YSOs	
	Initial	Final	Post-SQL	Initial	Final	Initial	Final	Initial	Final	Initial	Final
$n_{objects}$	1,126	1,072	496	37,049	739	6,997	5,279	39,677	2,978	577,193	199,391
u'	19.9%	19.6%	25.0%	0.7%	0.1%	14.4%	13.9%	21.3%	8.4%	6.3%	4.2%
B	22.4%	22.7%	26.0%	19.6%	4.2%	16.9%	9.1%	47.2%	10.7%	11.3%	1.6%
g'	38.5%	38.5%	46.0%	21.0%	3.9%	28.2%	21.1%	54.3%	18.1%	16.3%	5.9%
V	22.6%	22.9%	26.2%	19.7%	3.9%	17.1%	9.3%	47.7%	10.8%	11.4%	1.6%
r'	38.2%	38.3%	45.8%	21.0%	4.1%	28.2%	21.1%	54.2%	18.2%	16.3%	5.9%
i'	38.1%	38.4%	46.4%	19.3%	3.8%	28.4%	21.3%	53.5%	18.3%	16.5%	6.0%
z'	19.9%	19.6%	25.0%	0.7%	0.1%	14.4%	13.9%	21.3%	8.4%	6.3%	4.2%
J	89.5%	89.3%	96.0%	98.9%	87.6%	78.3%	73.3%	95.1%	70.4%	77.5%	52.1%
H	93.0%	93.0%	97.4%	94.0%	9.5%	89.5%	87.2%	96.5%	84.4%	90.8%	84.7%
K	95.2%	95.2%	98.4%	78.4%	7.6%	95.9%	95.4%	97.2%	91.9%	93.1%	91.1%
$W1$	88.4%	88.1%	100.0%	84.1%	1.9%	99.9%	100.0%	99.1%	99.9%	99.9%	100.0%
$W2$	88.2%	87.9%	100.0%	82.3%	0.9%	100.0%	100.0%	99.9%	99.9%	99.9%	100.0%
$W3$	84.6%	85.7%	96.8%	13.6%	1.1%	90.3%	87.2%	98.7%	83.7%	78.5%	38.2%
$W4$	77.5%	78.9%	87.1%	2.8%	0.8%	82.0%	83.0%	74.7%	78.7%	57.1%	41.0%

Note. — Photometric completeness by filter for all wavebands listed in Table 2.2. In addition to magnitude limits, the imposition of photometric quality criteria may also contribute to incompleteness. In the case of stars with known protoplanetary disks, the distinction between "Final" and "Post-SQL" is that the former stage is assessed after the first round of color-color cuts (at the end of Section 2.4), while the latter stage is assessed after the SQL query of AllWISE (at the beginning of Section 2.5). Completeness estimates at both stages are provided here in light of the discussion of the false negative rate in Section 2.6.

A potential benefit of excluding the two control fields closest to the Galactic plane (see Section 2.3) is that red giants and supergiants may be overrepresented away from the plane and hence will be better characterized for elimination, thus leaving more (genuine) young red stars and fewer (contaminating) old red stars. Three optical colors are illustrated in Figure 3.1 as a way of comparing the spectral types of stars harboring known protoplanetary disks to those of the control field sources used in the color selection process of Section 2.4. The colors of the new candidate YSOs are also shown. These color distributions suggest that although the optical colors of the new candidate YSOs seem to match the colors of the sample of stars with known protoplanetary disks, with some contamination that skews the distribution towards that of the control field sources, the sample of control field sources may indeed represent a different population of stars. That said, it may not be surprising that more evolved, main sequence stars are either hotter (having reached the main sequence) or bluer (with less reddening from circumstellar dust) than YSOs are.

An optical-infrared color is provided in Figure 3.1 for reference, since most of the color-color cuts deal with infrared colors. The colors of the new candidate YSOs again match the colors of stars with known protoplanetary disks reasonably well in this fourth color, with some contamination skewing the candidates' distribution towards the presumably main sequence colors of the control field sources, but this time a signature of reddening is also apparent as a secondary peak. These reddened sources may represent disks around hotter YSOs, since most of the sources in the sample of stars with known protoplanetary disks tend to be cooler stars, or this peak could demonstrate contamination from interstellar (as opposed

to circumstellar) dust reddening. The possibility of having included candidate YSOs that are not actually young stars but instead reflect interstellar dust reddening is of particular concern closest to the Galactic disk, where most of the new candidates are located. Further study is needed to provide clarification on these points.

It should be noted that the KS test as described in Section 2.4 is the classical, univariate two-sample KS test and is not used to make any color-color cuts. Each axis on a color-color plot represents a single color (or magnitude, using the broader definition of "color-color" from Section 2.4), and the KS test is used to quantify the difference in each of these individual colors. Two of the colors showing the most separation between the training sample and the control sample (stars with known protoplanetary disks and control field sources, respectively, in Section 2.4; known YSOs and known non-YSOs, respectively, in Section 2.5) are used to construct a color-color plot, at which point one or more color-color cuts is made by eye. Of the remaining sources, univariate KS statistics are recalculated, and further cuts in other color-color spaces may be made according to the iterative process described in Section 2.4.

Making color-color cuts by eye is not claimed to be the most sophisticated method available for solving statistical classification problems. Support vector machines, for example, as used by Marton et al. (2016) may marginally improve the accuracy of each cut, but machine learning techniques carry their own pitfalls. Of particular note is that machine learning techniques often employ abstract, complex parameterizations of classification criteria that are not "user-friendly," in the sense that the classification scheme is difficult to replicate without a copy of the parameters originally used in the selection process. The simpler "by-eye"

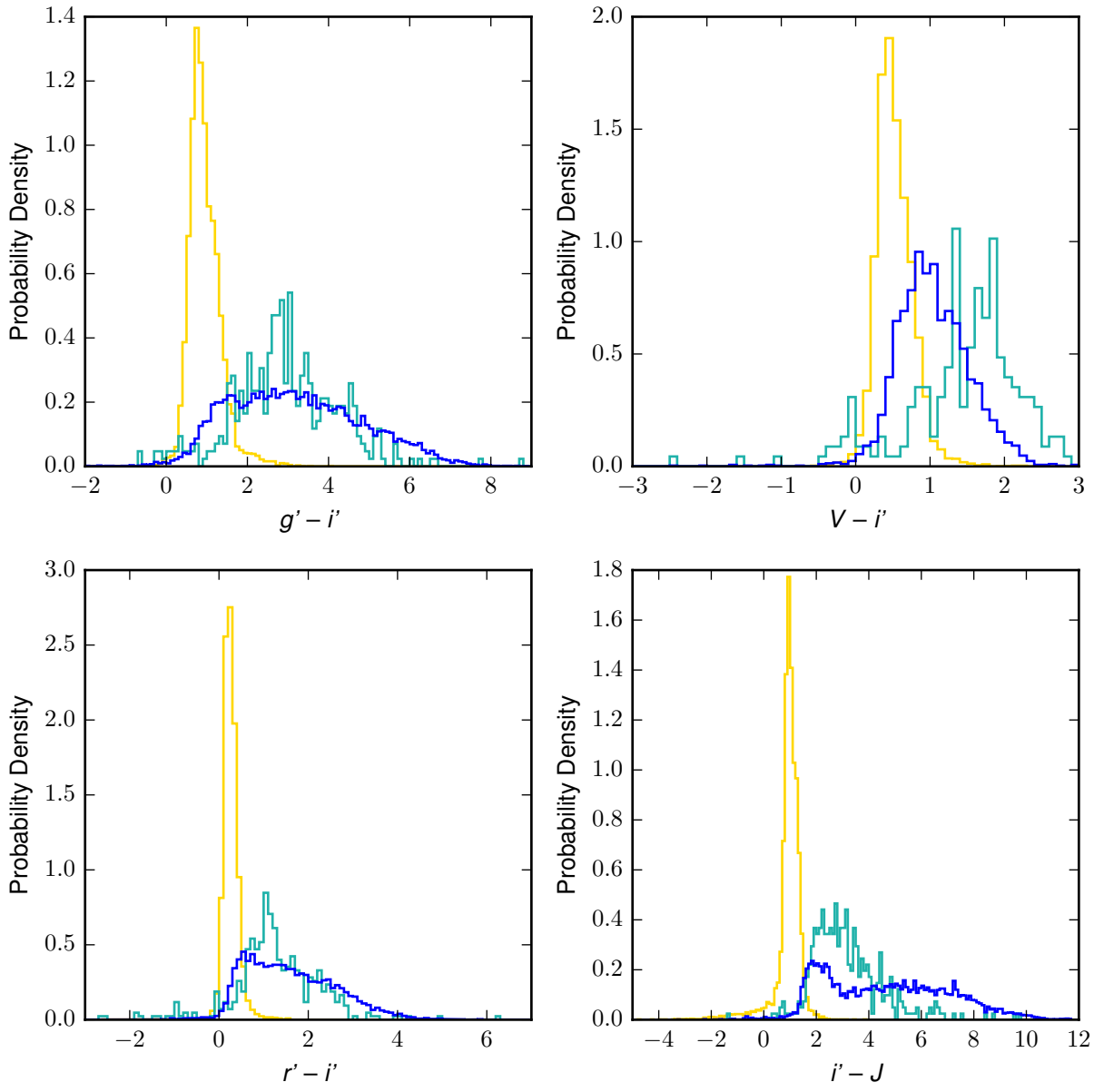

Figure 3.1 — Selected colors of stars hosting known protoplanetary disks (turquoise), control field sources (gold), and new candidate YSOs (blue). While the distributions of colors of new candidate YSOs seem to match those of the sample of stars with known protoplanetary disks reasonably well, there may be contamination that skews the distribution peak towards bluer colors and creates a secondary red peak in $i' - J$, although there may be other explanations for these features. The control field sources seem to represent a distinct population from the stars with known protoplanetary disks in each of the optical colors, which could be a sign that the spectral type distributions in the control fields are distinct from those of the stars with known protoplanetary disks.

method of making color-color cuts as presented here (Section 2.4) is already very effective at distinguishing training samples from control samples (Figures 2.2 and 2.4 and Appendices A and B), and the cuts themselves (Tables 2.4 and 2.6) are simple and efficient enough that they can be readily replicated by other investigators. More sophisticated methods of color selection may be worth considering in future work, but the method presented here carries its own merits.

While additional study will also be required to de-redden sources, possibly using Planck dust column maps (which are already published) and GAIA distances (only a few of which have been published at the time of writing), a preliminary analysis of spectral indices without taking reddening into account may nevertheless prove illuminating. Spectral indices for stars harboring known protoplanetary disks, known YSOs, known non-YSOs, and new candidate YSOs are shown in Figure 3.2. For any source in these categories with at least two reasonable-quality magnitudes (see Table 2.2) in the six filters within the 2 – 25 micron range (Greene et al. 1994; Williams & Cieza 2011) — H, K, $W1$, $W2$, $W3$, and $W4$ — the spectral index was calculated with the standard formula

$$\alpha \equiv \frac{d \log(\lambda F_\lambda)}{d \log \lambda}$$

where λ is the central wavelength of each filter and F_λ is the monochromatic flux (relative to the zero-magnitude flux) integrated across the bandwidth of that filter. This expression represents the slope of a best-fit line in log-log space. Fitting was accomplished with the linear regression function in the standard Python statistics package. Dashed lines demarcating

the classical boundaries for Class II YSOs ($-1.6 < \alpha < -0.3$) and flat-spectrum YSOs ($-0.3 < \alpha < 0.3$) are shown for reference. As expected, almost all of the stars with known protoplanetary disks fall neatly within the Class II bounds, and most of the known YSOs fall within the extended bounds (including flat-spectrum sources). Most known non-YSOs do not display this infrared excess. Most of the new candidate YSOs fall within the extended bounds, although the peak is slightly shifted towards sources without infrared excess. This may be due to the inclusion in the sample of stars with known protoplanetary disks of sources whose disks were detected at millimeter wavelengths instead of in the infrared. The tails on the distributions may be due to errors in magnitudes, especially because many sources are relatively faint (see Figure 2.6). Mitigating these tails may be one way to improve the fail positive rate of the catalogue.

As intended, this work lays a solid foundation for a variety of follow-up studies to test the accuracy of the catalogue presented here and to examine evolutionary models of protoplanetary disks specifically and YSOs generally. The aforementioned discussions with advisors and colleagues have proved invaluable towards this end, and it is clear that this line of inquiry into YSO development has a "bright" future ahead — magnitude limits notwithstanding.

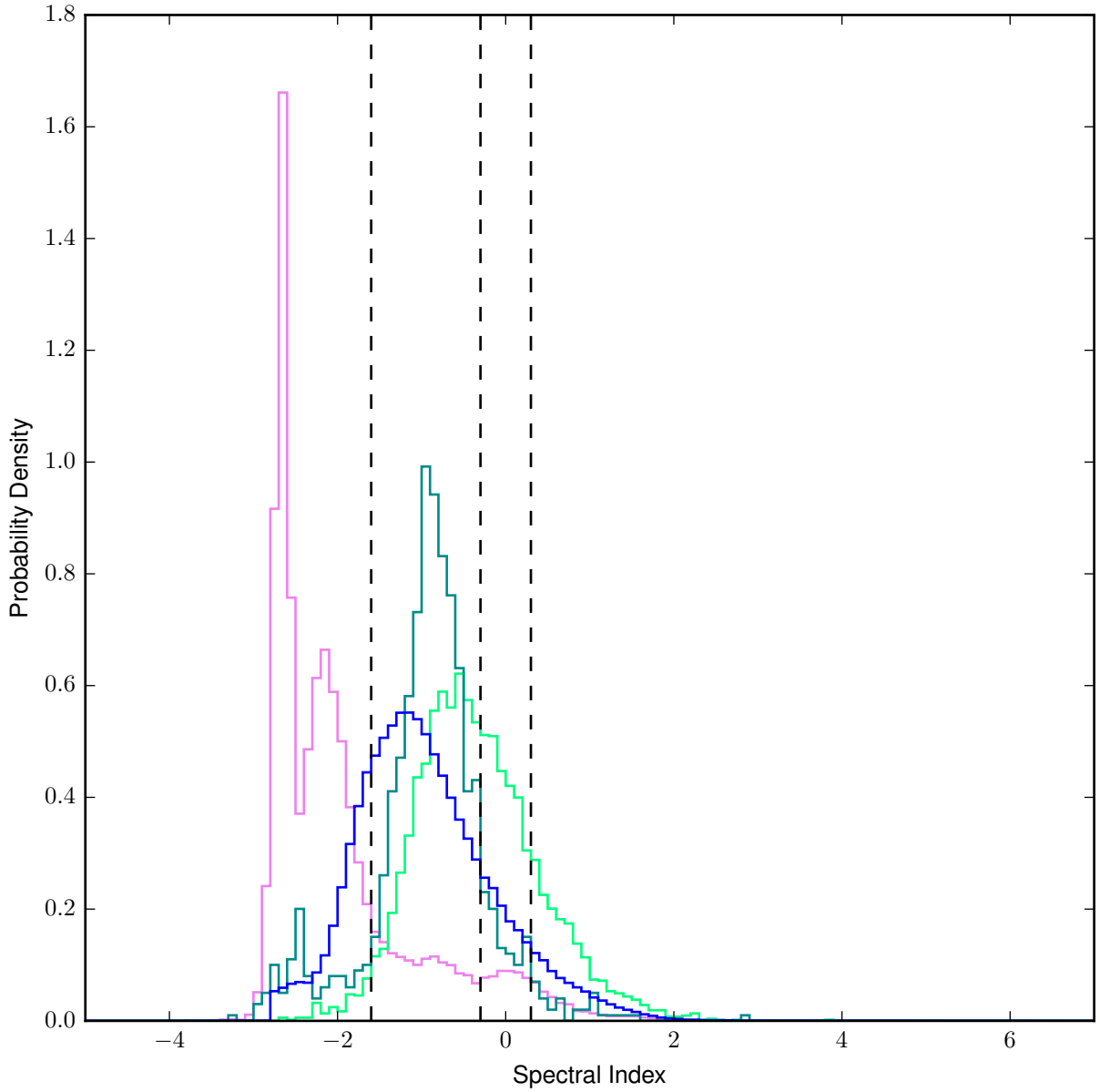

Figure 3.2 — Spectral indices of stars with known protoplanetary disks (turquoise), known YSOs (mint), known non-YSOs (fuchsia), and new candidate YSOs (blue), without attempting to correct for interstellar reddening. Similarly to Figure 3.1, the integral of each probability density function is equal to 1, and each bin has a width of 0.1, which is why maxima may exceed a density of 1. Dashed lines show the classical boundaries of Class II YSOs ($-1.6 < \alpha < -0.3$) and flat-spectrum YSOs ($-0.3 < \alpha < 0.3$). As expected, most stars with known protoplanetary disks and known YSOs fall within these bounds, while most known non-YSOs do not. Distribution tails may be due to magnitude measurement errors, as many of these sources are relatively faint (Figure 2.6).

REFERENCES

Acke, B., van den Ancker, M. E., Dullemond, C. P., van Boekel, R., & Waters, L. B. F. M. 2004, A&A, 422, 621

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21

Alonso-Albi, T., Fuente, A., Bachiller, R., et al. 2009, A&A, 497, 117

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Binks, A. S., & Jeffries, R. D. 2016

Bodenheimer, P., & Lin, D. N. C. 2002, AREPS, 30, 113

Boss, A. P. 1997, Science, 276, 1836

Brown, M. E., van Dam, M. A., & Bouchez, A. H. 2006, ApJ, 639, L43

Carlson, L. R., Sewiło, M., Meixner, M., Romita, K. A., & Lawton, B. 2012, A&A, 542, A66

Chapman, N. L., Lai, S.-P., Mundy, L. G., et al. 2007, ApJ, 667, 288

Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, NASA/IPAC Infrared Science Archive

Cutri, R. M., et al. 2014

ESA. 1997, ESA, SP-1200

Evans, II, N. J., Allen, L. E., Blake, G. A., et al. 2003, PASP, 115, 965

Furlan, E., Watson, D. M., McClure, M. K., et al. 2009, ApJ, 703, 1964

GAIA Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, Gaia

Greene, T. P., Wilking, B. A., André, P., Young, E. T., & Lada, C. J. 1994, ApJ, 434, 614

Harvey, P. M., Rebull, L. M., Brooke, T., et al. 2007, ApJ, 663, 1139

Helled, R., Bodenheimer, P., Podolak, M., et al. 2014, Protostars and Planets VI, 643

Henden, A. A., Templeton, M., Terrell, D., et al. 2016, AAVSO

Hog, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

Lin, D. N. C., Bodenheimer, P., & Richardson, D. C. 1996, Nature, 380, 6575

Luhman, K. L., Allen, P. R., Espaillat, C., Hartmann, L., & Calvet, N. 2010, ApJS, 186, 111

Mannings, V., & Sargent, A. 1997, ApJ, 490, 792

——. 2000, ApJ, 529, 391

Marton, G., Tóth, L. V., Paladini, R., et al. 2016, MNRAS, 458, 3479

Matsuo, T., Shibai, H., Ootsubo, T., & Tamura, M. 2007, ApJ, 662, 1282

Mayor, M., & Queloz, D. 1995, Nature, 378, 6555

Metchev, S. A., & Grindlay, J. E. 2002, MNRAS, 335, 73

Oliveira, I., Pontoppidan, K. L., Merín, B., et al. 2009, ApJ, 714, 778

Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49

Pollack, J. B., Hubickyj, O., Bodenheimer, P., et al. 1996, Icarus, 124, 62

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

Rebull, L. M., Padgett, D. L., McCabe, C.-E., et al. 2010, ApJS, 186, 259

Rebull, L. M., Stapelfeldt, K. R., Evans, II, N. J., et al. 2007, ApJS, 171, 447

Scholz, A., Jayawardhana, R., & Wood, K. 2006, ApJ, 645, 1498

Ward-Thompson, D., Di Francesco, J., Hatchell, J., et al. 2007, PASP, 119, 855

Williams, J. P., & Cieza, L. A. 2011, ARA&A, 49, 67

Yasui, C., Kobayashi, N., Tokunaga, A. T., & Saito, M. 2014, MNRAS, 442, 2543

APPENDICES

A Color-Color Cuts, Round 1, Complete

Each step of Round 1 of color-color cuts — as described in Section 2.4, summarized visually in Figure 2.2, and listed in Table 2.4 — is illustrated here. Each cut is previewed with a solid gray line, then depicted with a solid black line and gray exclusion region. Following Figures 2.1 and 2.2, stars hosting known protoplanetary disks are shown in turquoise, while control field sources are shown in gold. All plotted points are 80% transparent unless there are fewer than 100 points plotted in that specific color-color space. Each figure in this appendix contains the same color-color plots in the same order. Details on source selection are given in Section 2.3. To quantify the efficiency of each successive cut, each figure caption lists the percentages of known protoplantary disks retained and control field sources eliminated. Ideally, both of these quantities should be maximized in order to find a set of color-color cuts that reliably distinguishes stars with known protoplanetary disks from contaminant sources.

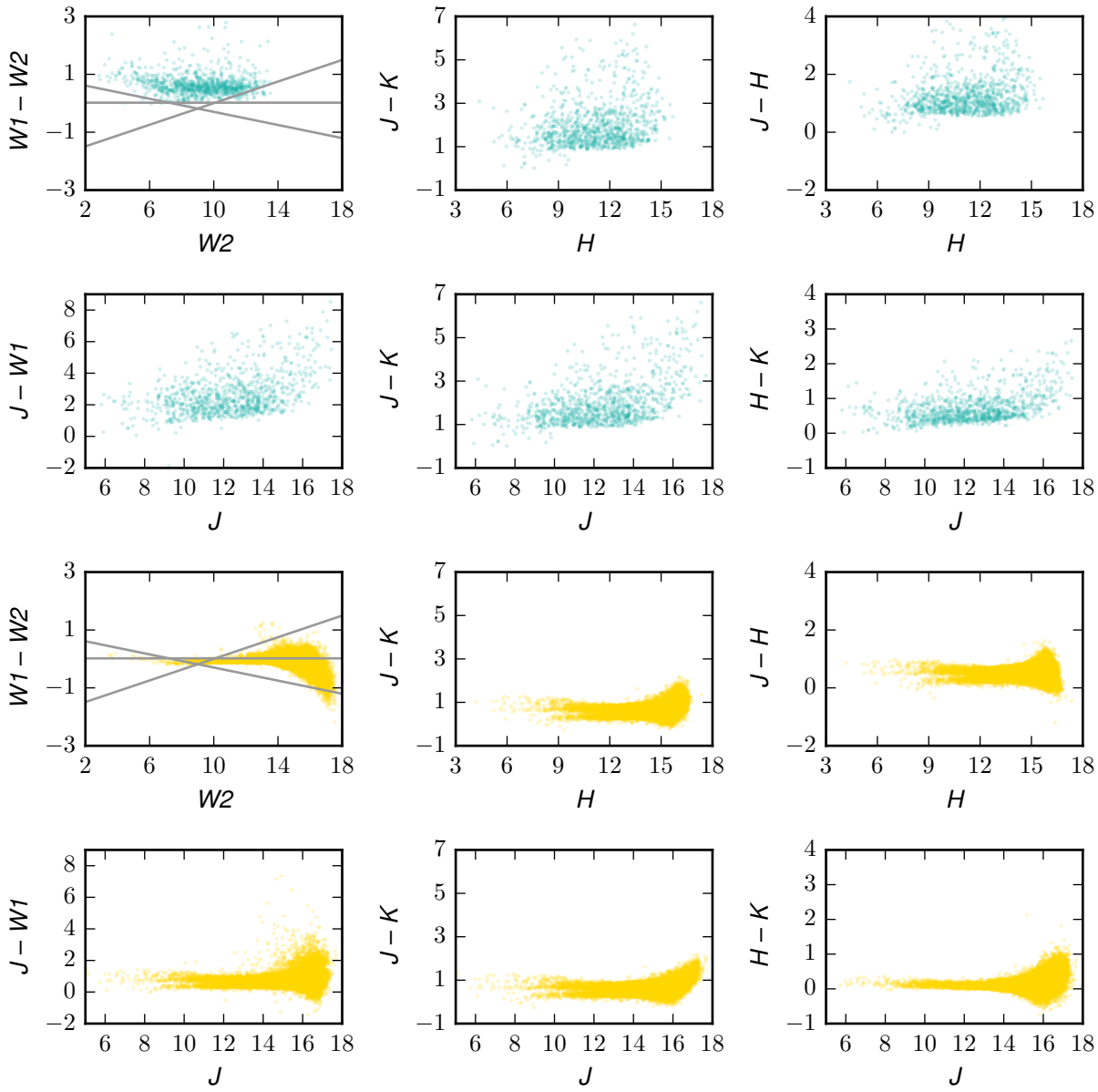

Figure A.1 — Initially, prior to any Round 1 color-color cuts: 1,126 (100%) stars with known protoplanetary disks (turquoise) are retained, while 0 (0%) of 37,049 control field sources (gold) are eliminated (unsurprisingly).

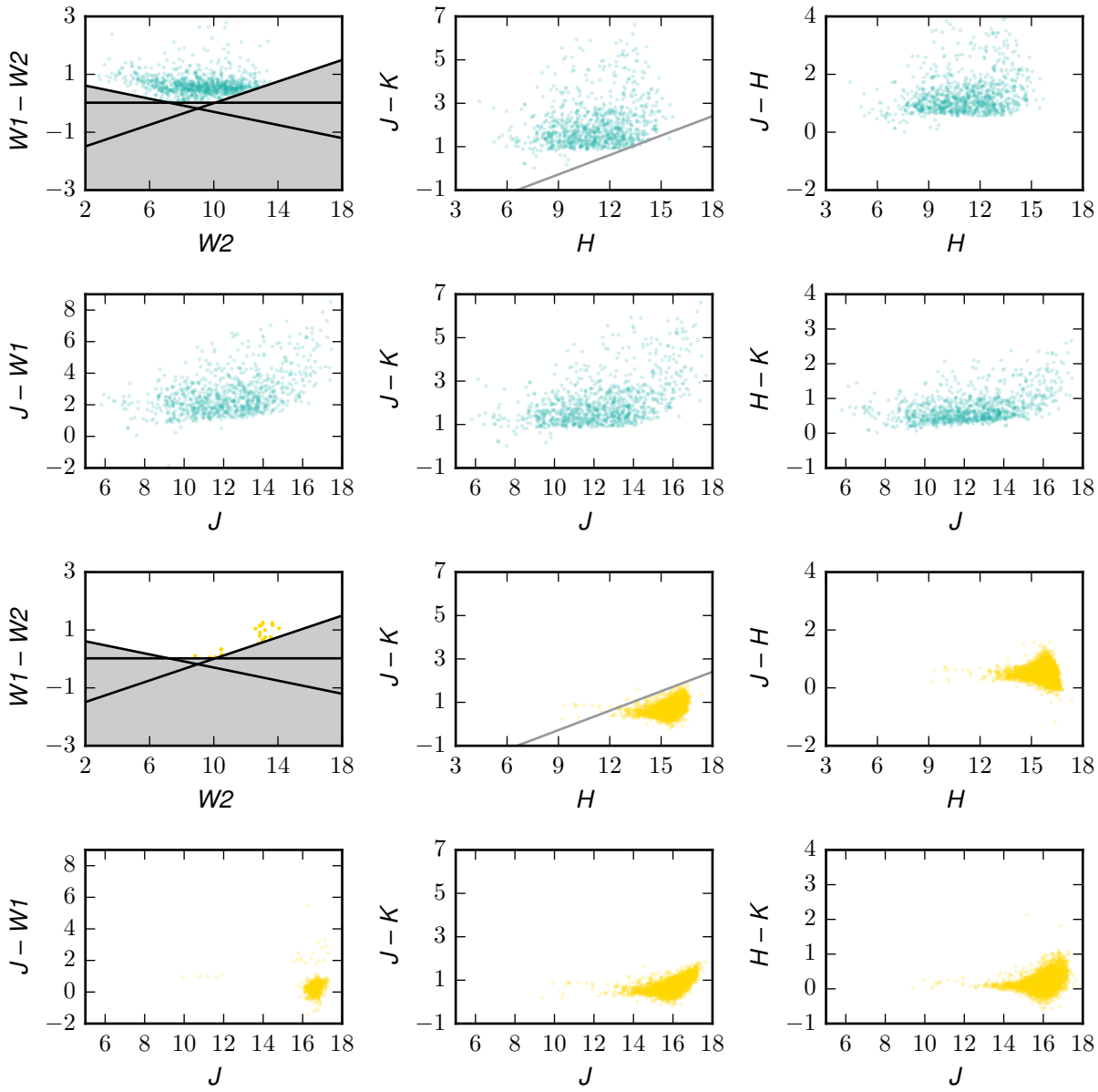

Figure A.2 — After the first three color-color cuts of Round 1: 1,083 (96.2%) stars with known protoplanetary disks (turquoise) are retained, while 30,464 (82.2%) control field sources (gold) are eliminated.

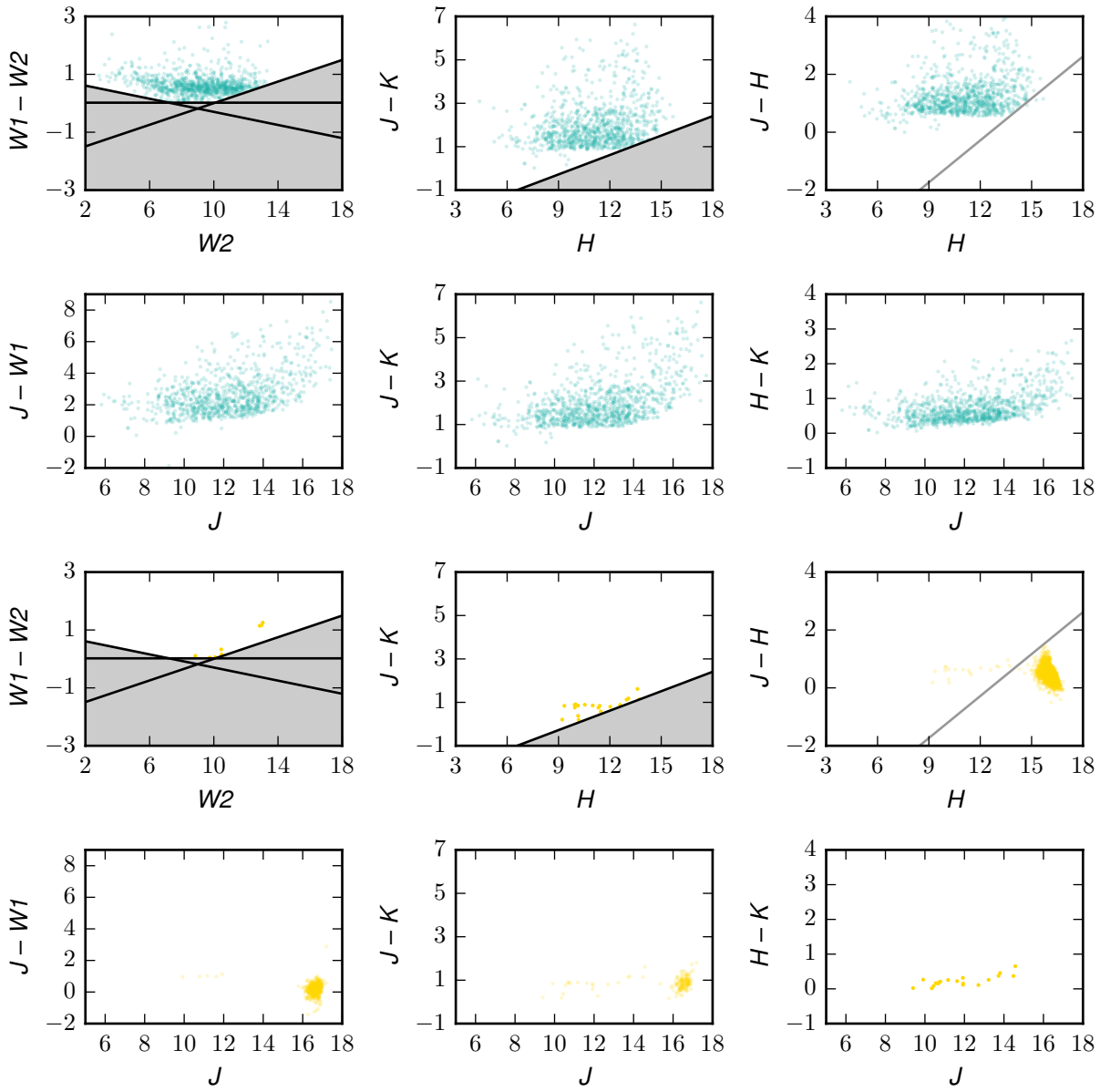

Figure A.3 — After the fourth color-color cut of Round 1: 1,077 (95.6%) stars with known protoplanetary disks (turquoise) are retained, while 33,990 (91.7%) control field sources (gold) are eliminated.

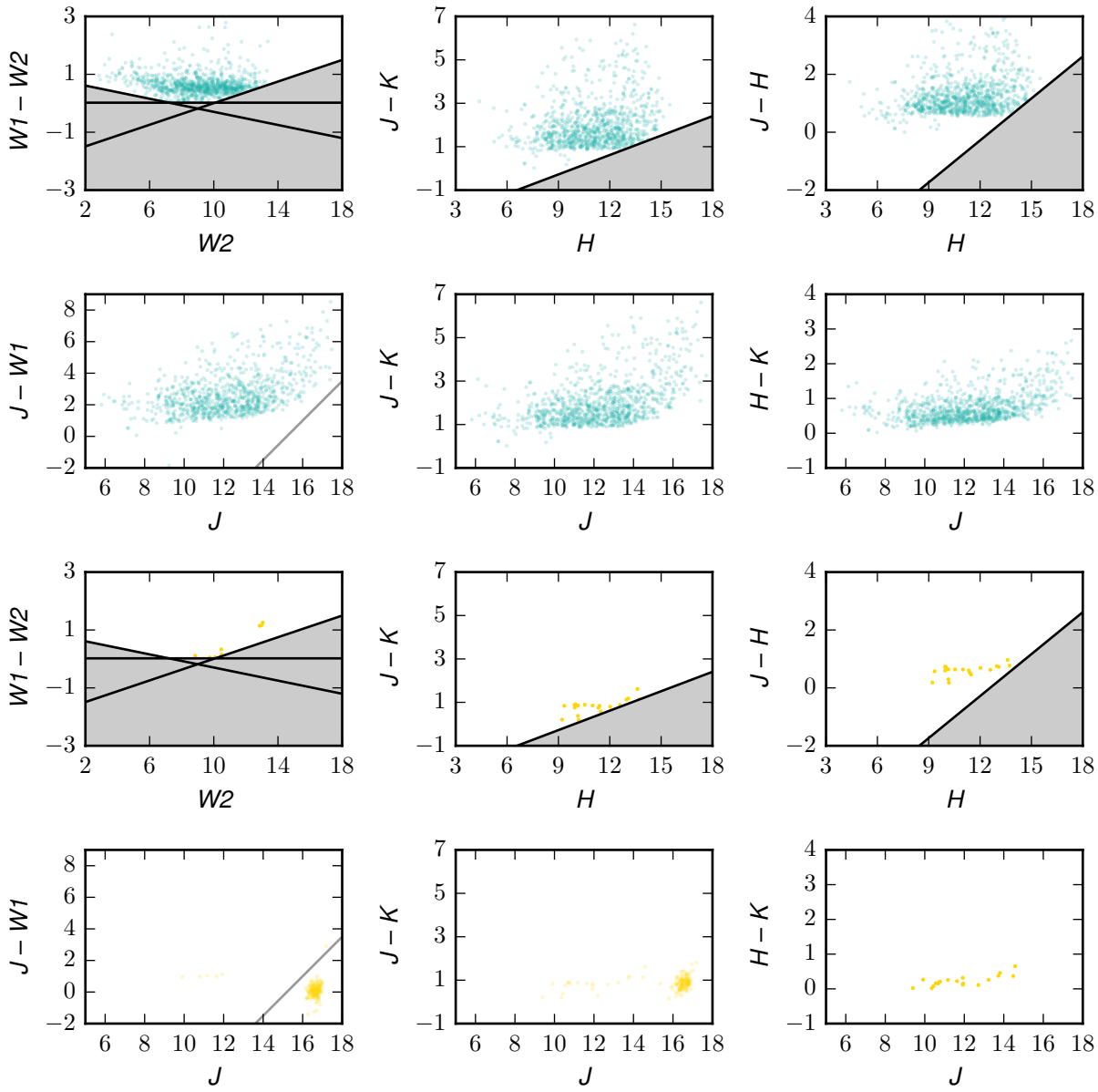

Figure A.4 — After the fifth color-color cut of Round 1: 1,074 (95.4%) stars with known protoplanetary disks (turquoise) are retained, while 36,019 (97.2%) control field sources (gold) are eliminated.

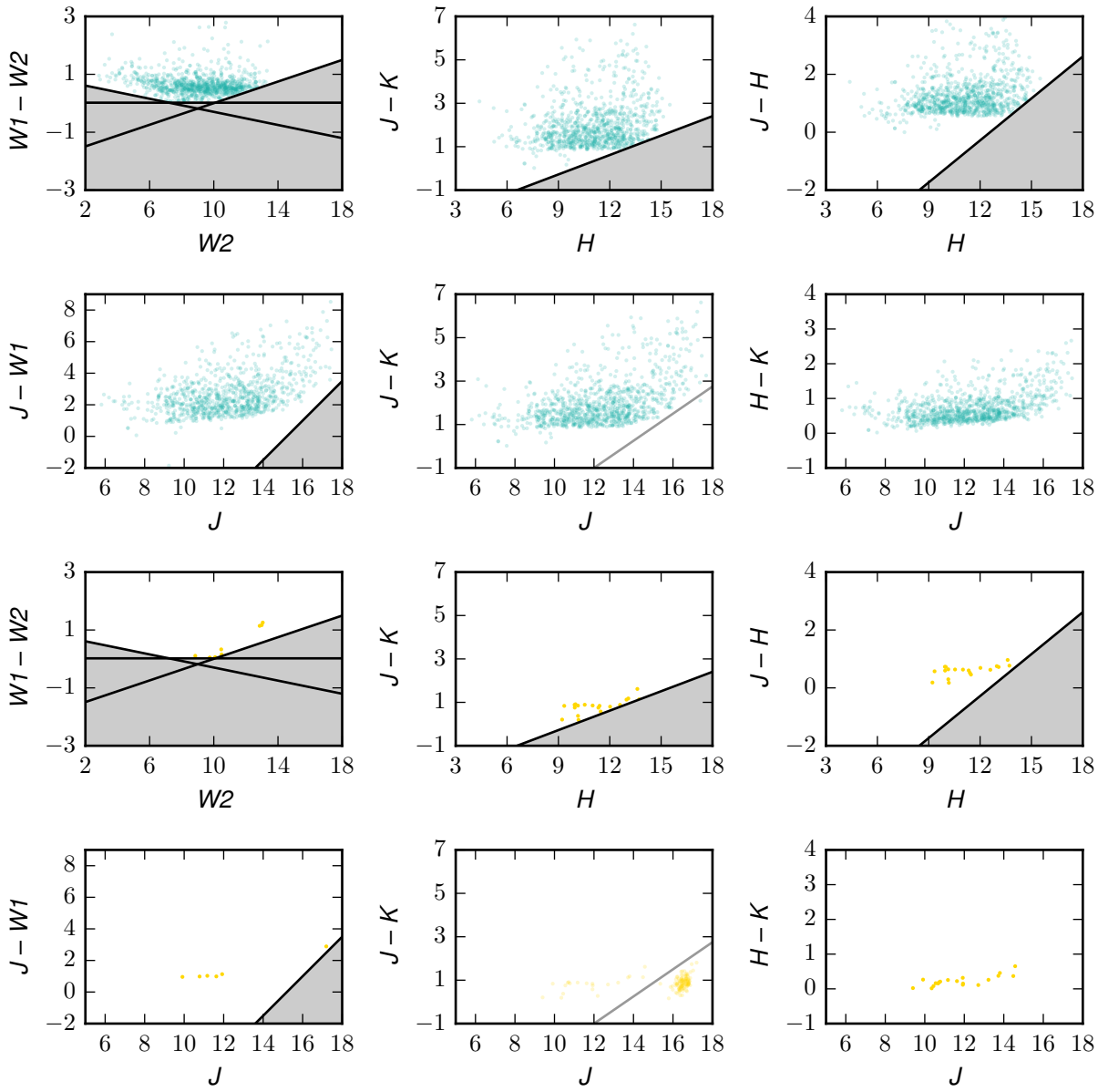

Figure A.5 — After the sixth color-color cut of Round 1: 1,074 (95.4%) stars with known protoplanetary disks (turquoise) are retained, while 36,192 (97.7%) control field sources (gold) are eliminated.

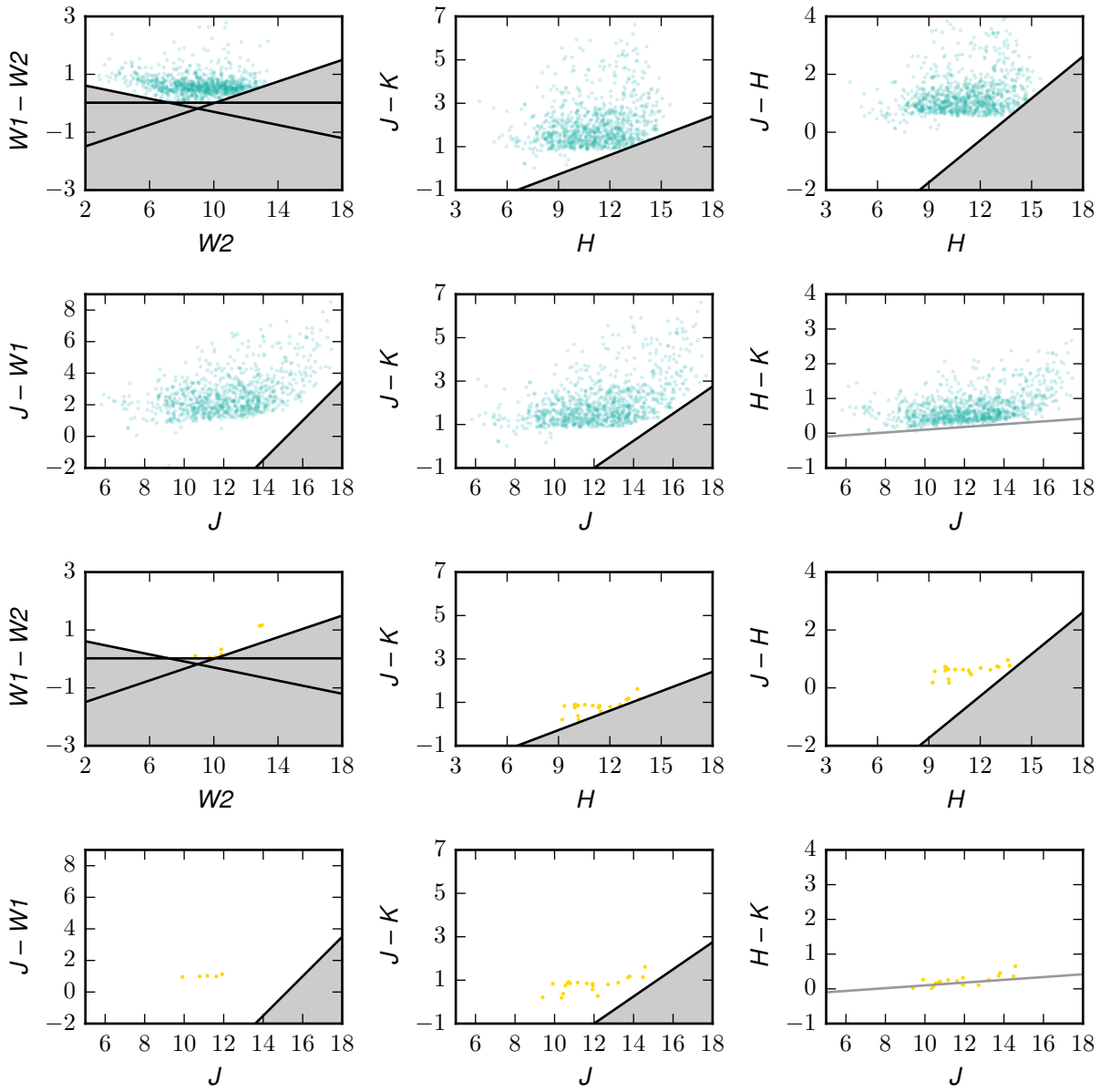

Figure A.6 — After the seventh color-color cut of Round 1: 1,073 (95.3%) stars with known protoplanetary disks (turquoise) are retained, while 36,303 (98.0%) control field sources (gold) are eliminated.

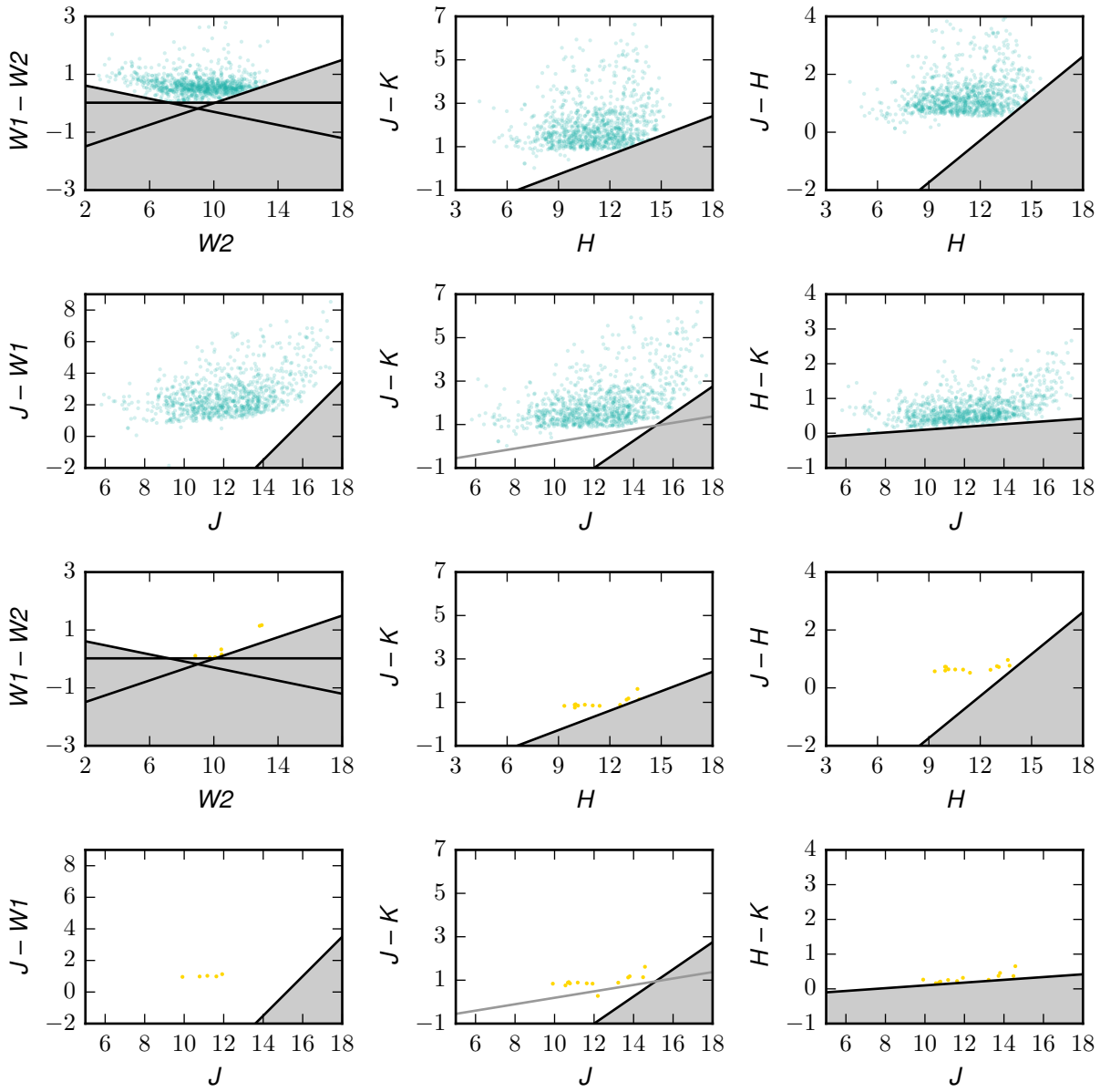

Figure A.7 — After the eighth color-color cut of Round 1: 1,072 (95.2%) stars with known protoplanetary disks (turquoise) are retained, while 36,309 (98.0%) control field sources (gold) are eliminated.

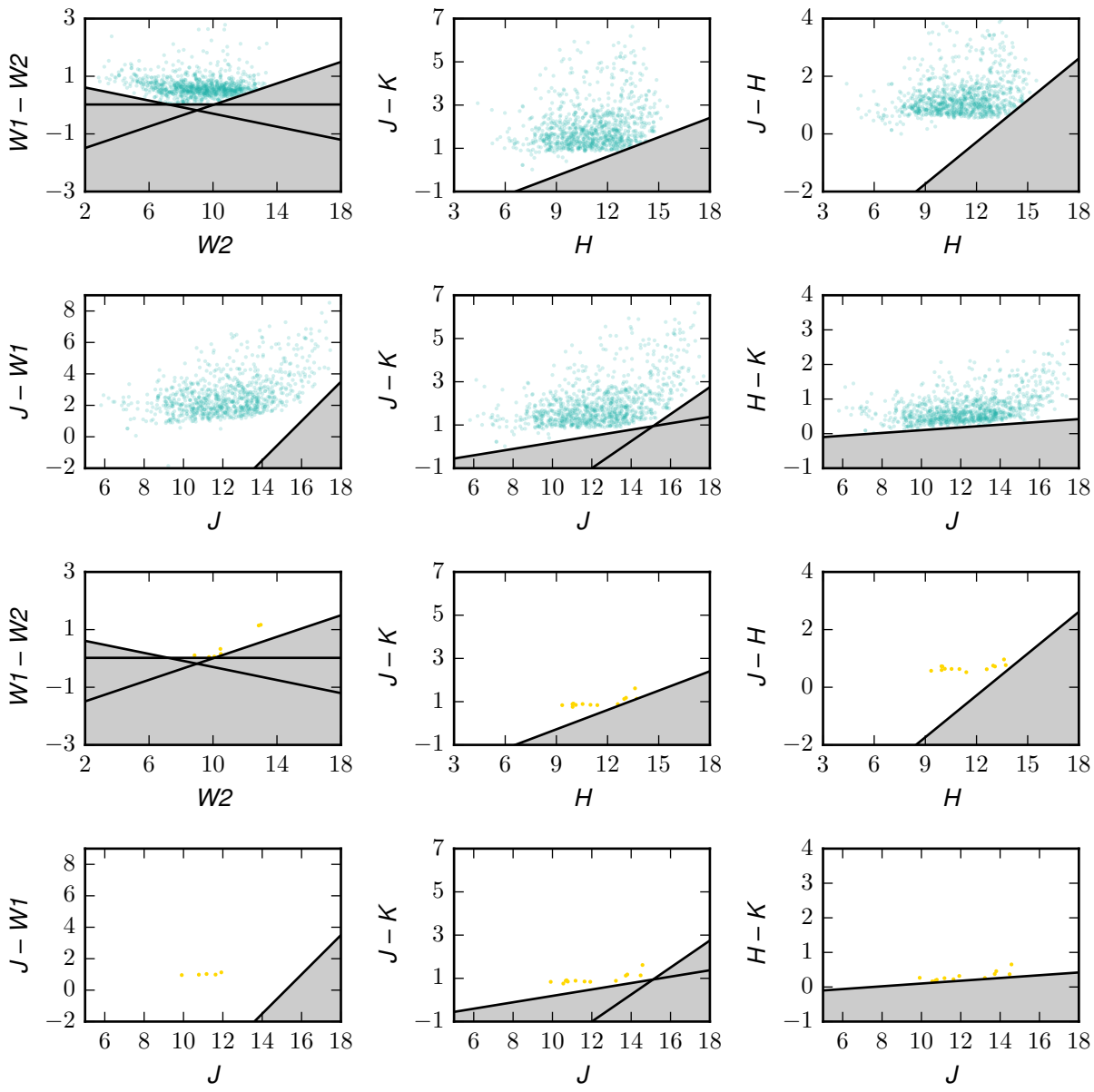

Figure A.8 — After the ninth and final color-color cut of Round 1: 1,072 (95.2%) stars with known protoplanetary disks (turquoise) are retained, while 36,310 (98.0%) control field sources (gold) are eliminated.

B Color-Color Cuts, Round 2, Complete

Each step of Round 2 of color-color cuts — as described in Section 2.5, summarized visually in Figure 2.4, and listed in Table 2.6 — is illustrated here. With one exception, each cut is previewed with a solid gray line, then depicted with a solid black line and gray exclusion region. (When a second cut is made in the $W1 - W3$ color, this cut is depicted with a dashed line.) Following Figures 2.3, 2.4, and 2.5, stars hosting known young stellar objects (YSOs) are shown in mint, while known non-YSOs are shown in fuchsia. All plotted points are 80% transparent unless there are fewer than 100 points plotted in that specific color-color space. Each figure in this appendix contains the same color-color plots in the same order, alternating between YSOs and non-YSOs. Details on source selection and SIMBAD classification are given in Section 2.5. Recall that all sources considered in this round must not have failed any of the cuts from Round 1. To quantify the efficiency of each successive cut in this second round, each figure caption lists the percentages of YSOs retained and non-YSOs eliminated. Ideally, both of these quantities should be maximized in order to find a set of color-color cuts that reliably distinguishes young stellar objects from contaminant sources.

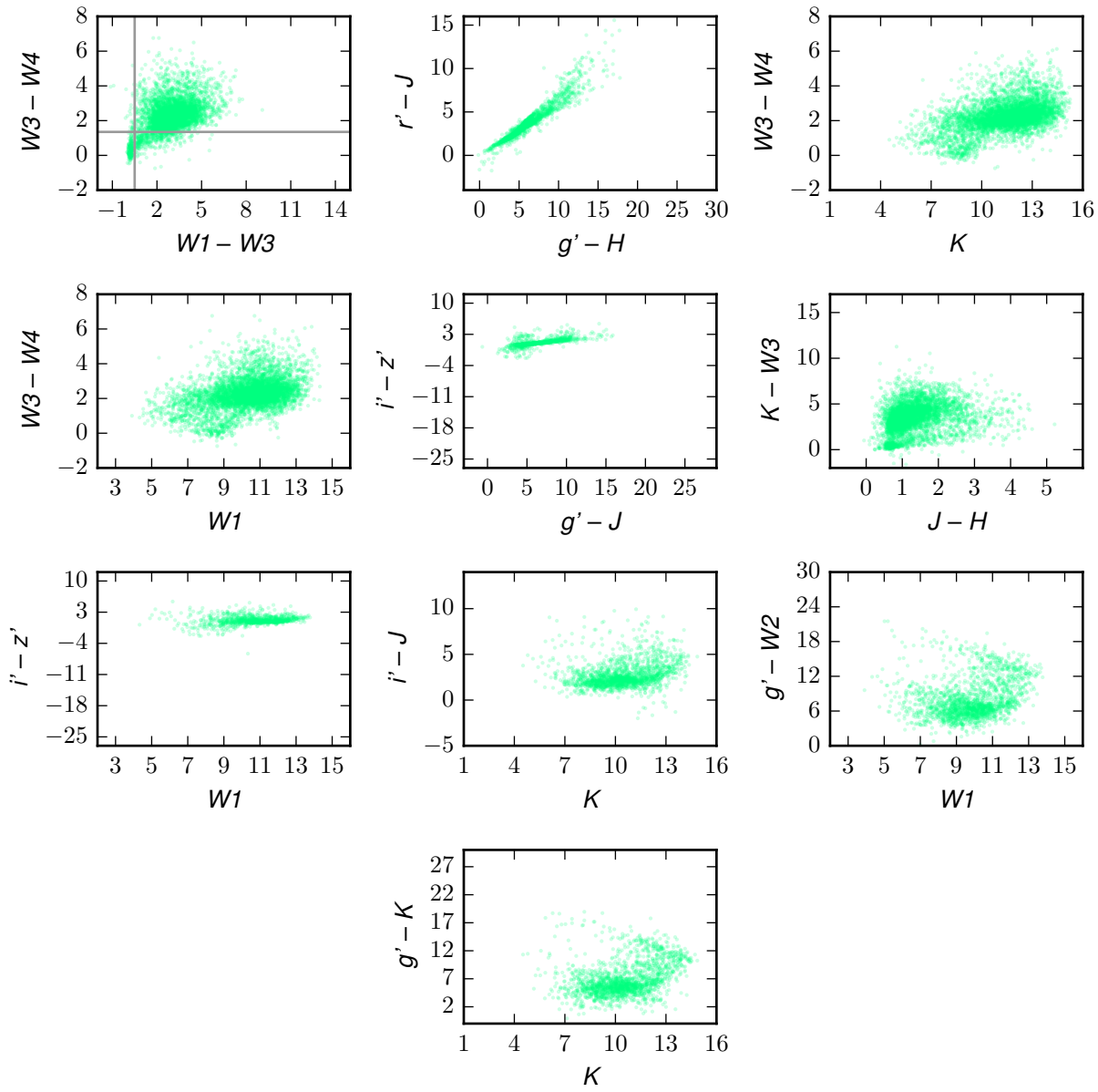

Figure B.1 — Initially, prior to any Round 2 color-color cuts: 6,997 (100%) known YSOs (mint) are retained, while 0 (0%) of 39,677 known non-YSOs (fuchsia, next figure) are eliminated (unsurprisingly).

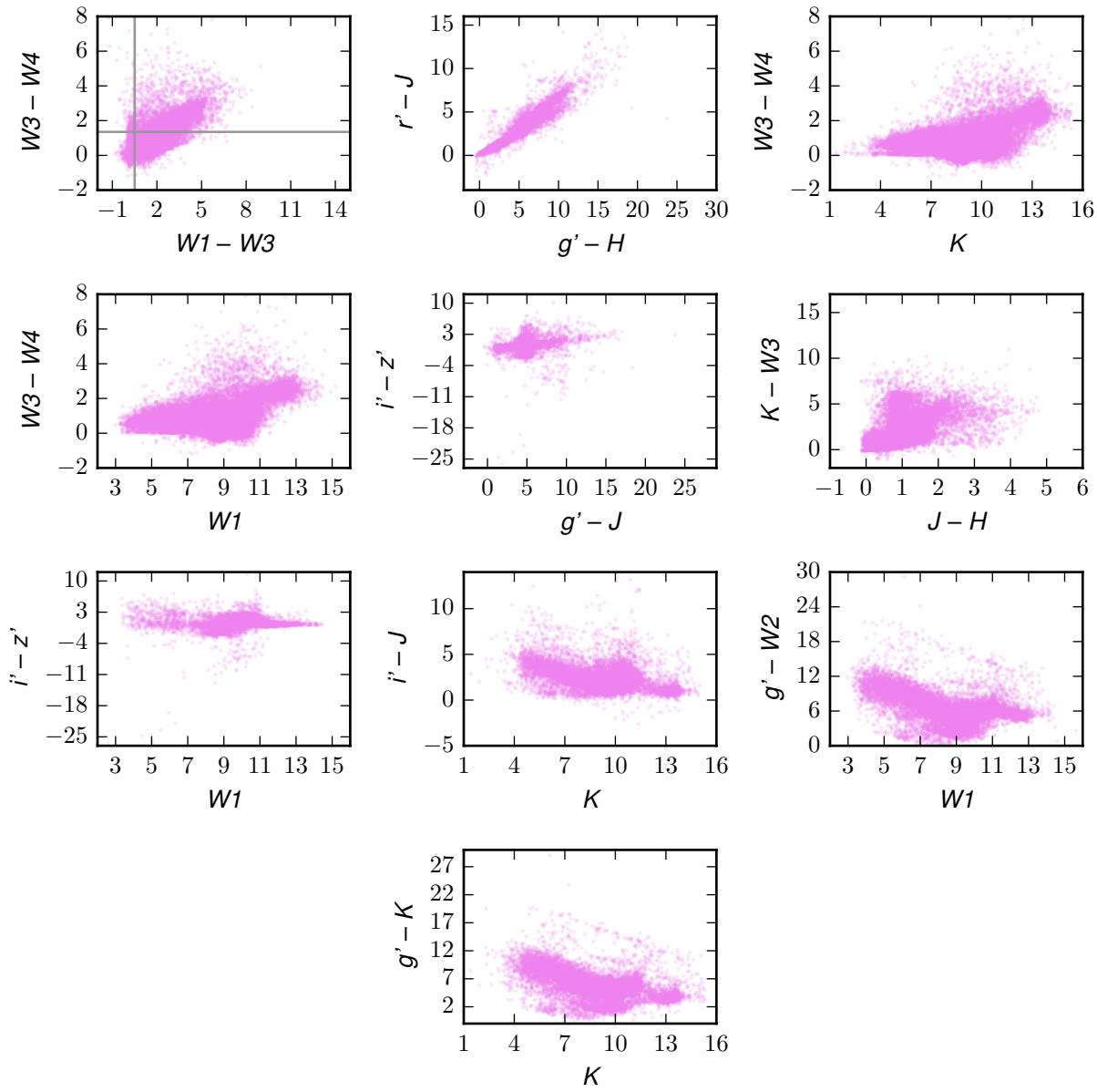

Figure B.2 — Initially, prior to any Round 2 color-color cuts: 6,997 (100%) known YSOs (mint, previous figure) are retained, while 0 (0%) of 39,677 known non-YSOs (fuchsia) are eliminated (unsurprisingly).

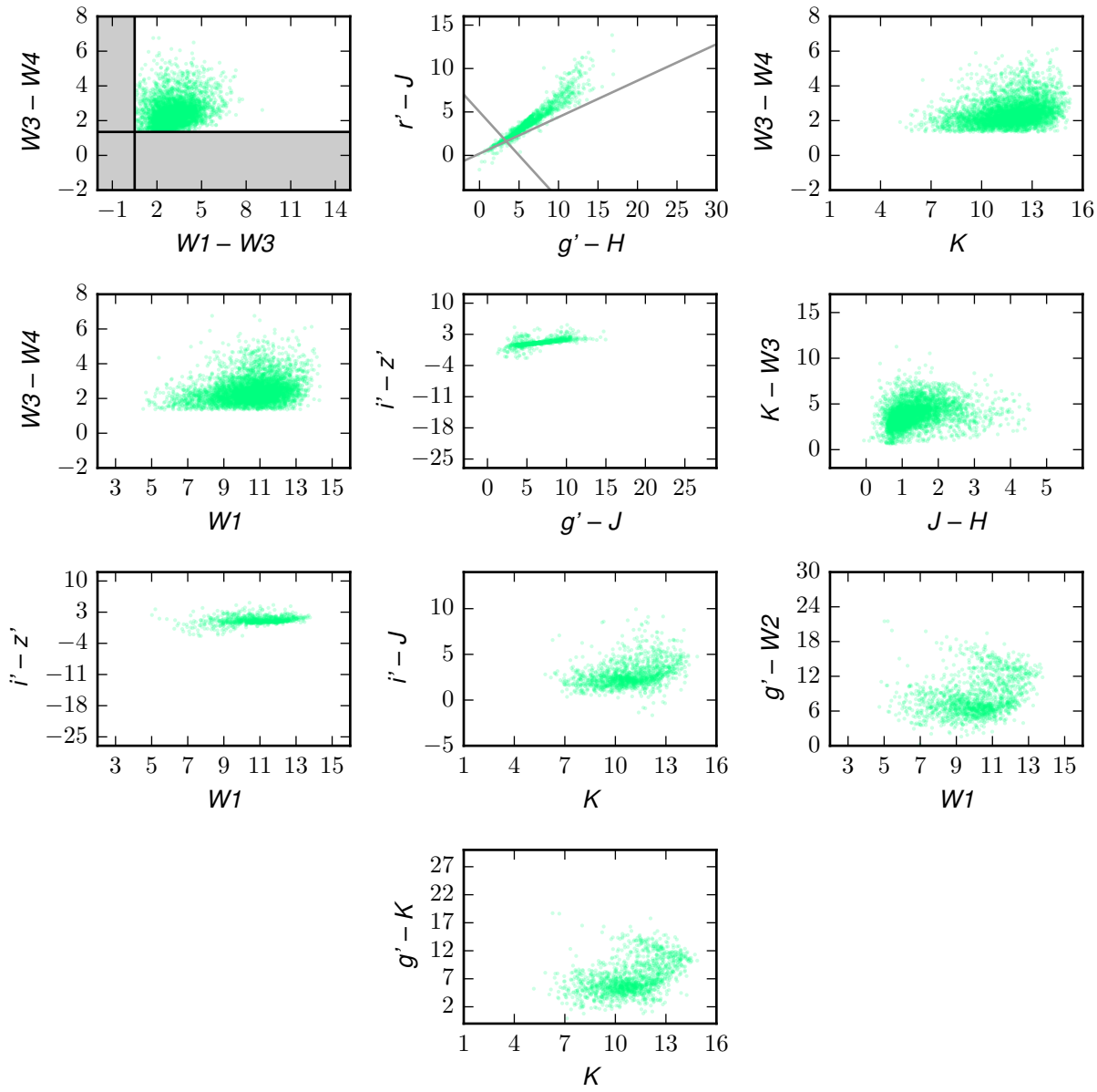

Figure B.3 — After the first and second color-color cuts of Round 2: 6,084 (87.0%) known YSOs (mint) are retained, while 32,486 (81.9%) known non-YSOs (fuchsia, next figure) are eliminated.

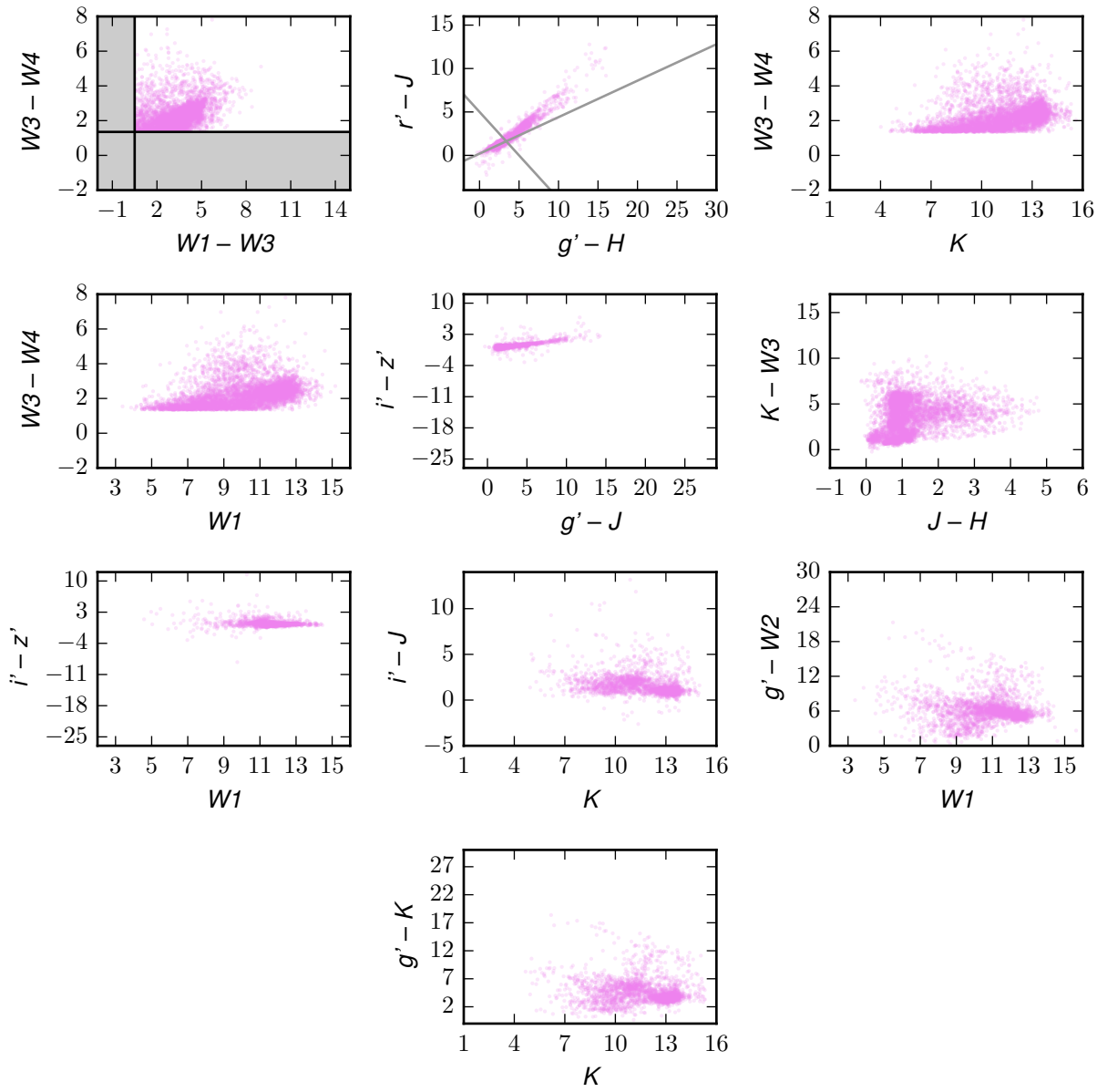

Figure B.4 — After the first and second color-color cuts of Round 2: 6,084 (87.0%) known YSOs (mint, previous figure) are retained, while 32,486 (81.9%) known non-YSOs (fuchsia) are eliminated.

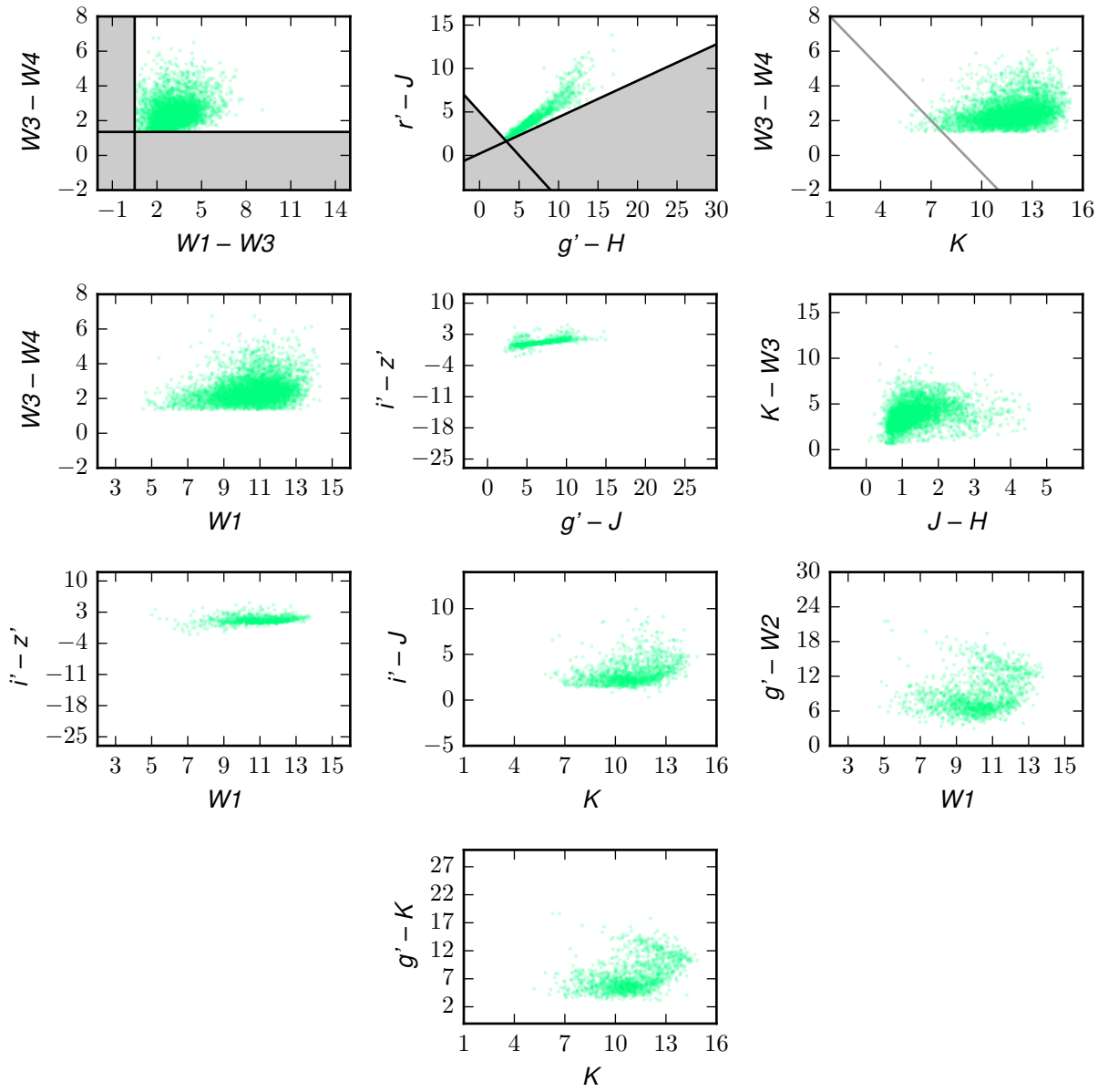

Figure B.5 — After the third and fourth color-color cuts of Round 2: 5,938 (84.9%) known YSOs (mint) are retained, while 33,824 (85.2%) known non-YSOs (fuchsia, next figure) are eliminated.

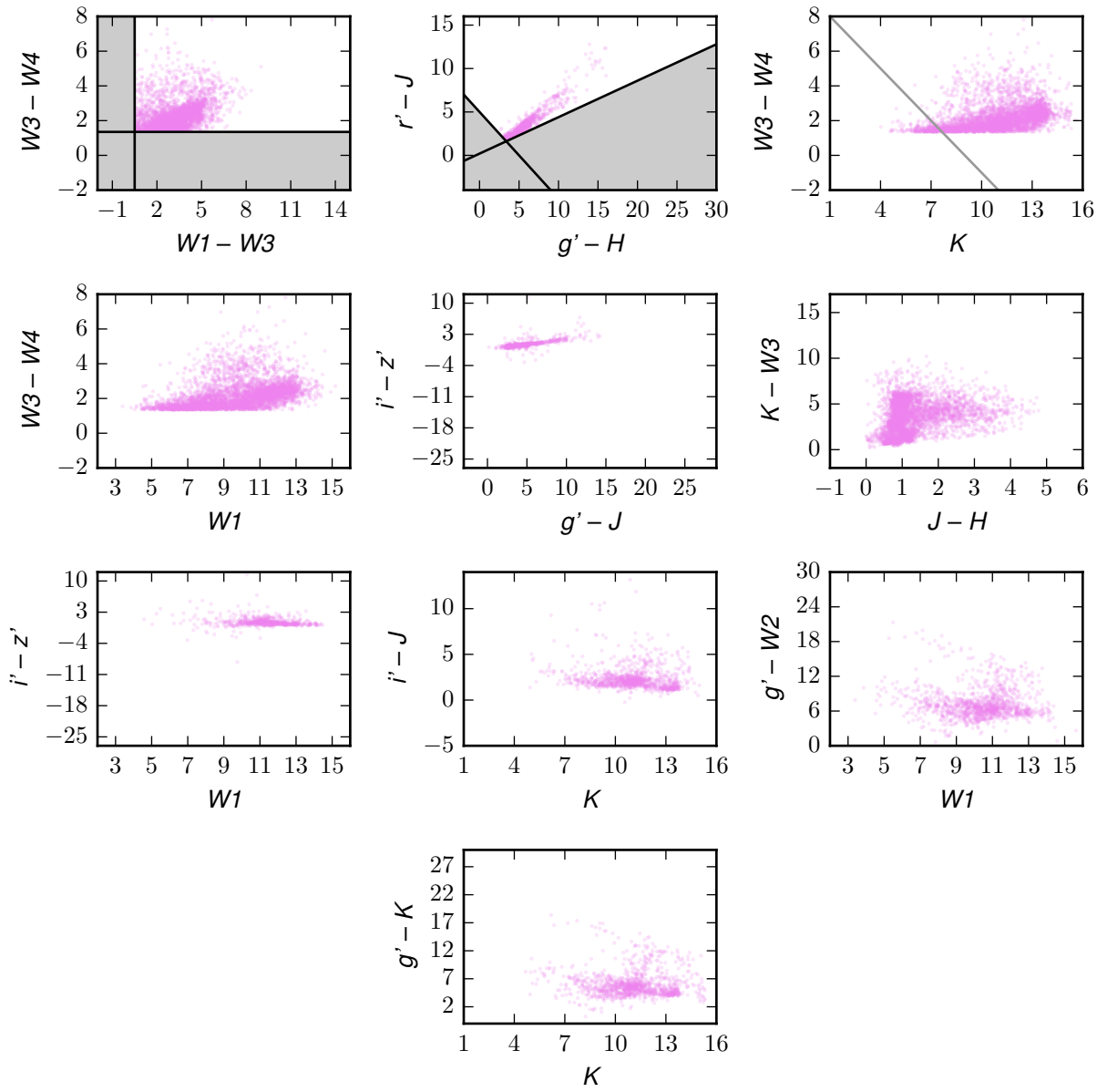

Figure B.6 — After the third and fourth color-color cuts of Round 2: 5,938 (84.9%) known YSOs (mint, previous figure) are retained, while 33,824 (85.2%) known non-YSOs (fuchsia) are eliminated.

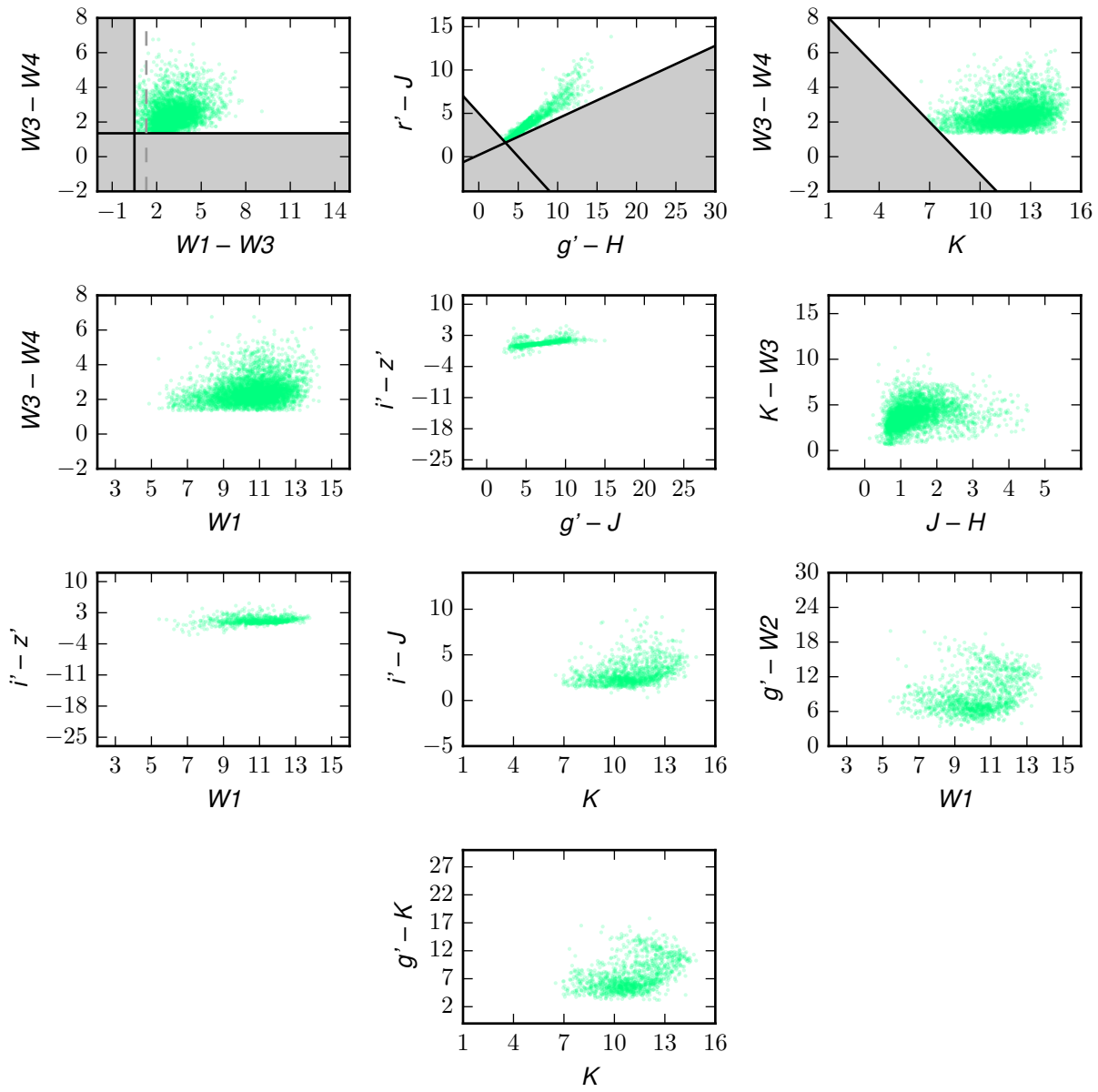

Figure B.7 — After the fifth color-color cut of Round 2: 5,908 (84.4%) known YSOs (mint) are retained, while 34,077 (85.9%) known non-YSOs (fuchsia, next figure) are eliminated. Note that the next cut is the second to be made in the $W1 - W3$ color and is delineated with a dashed line.

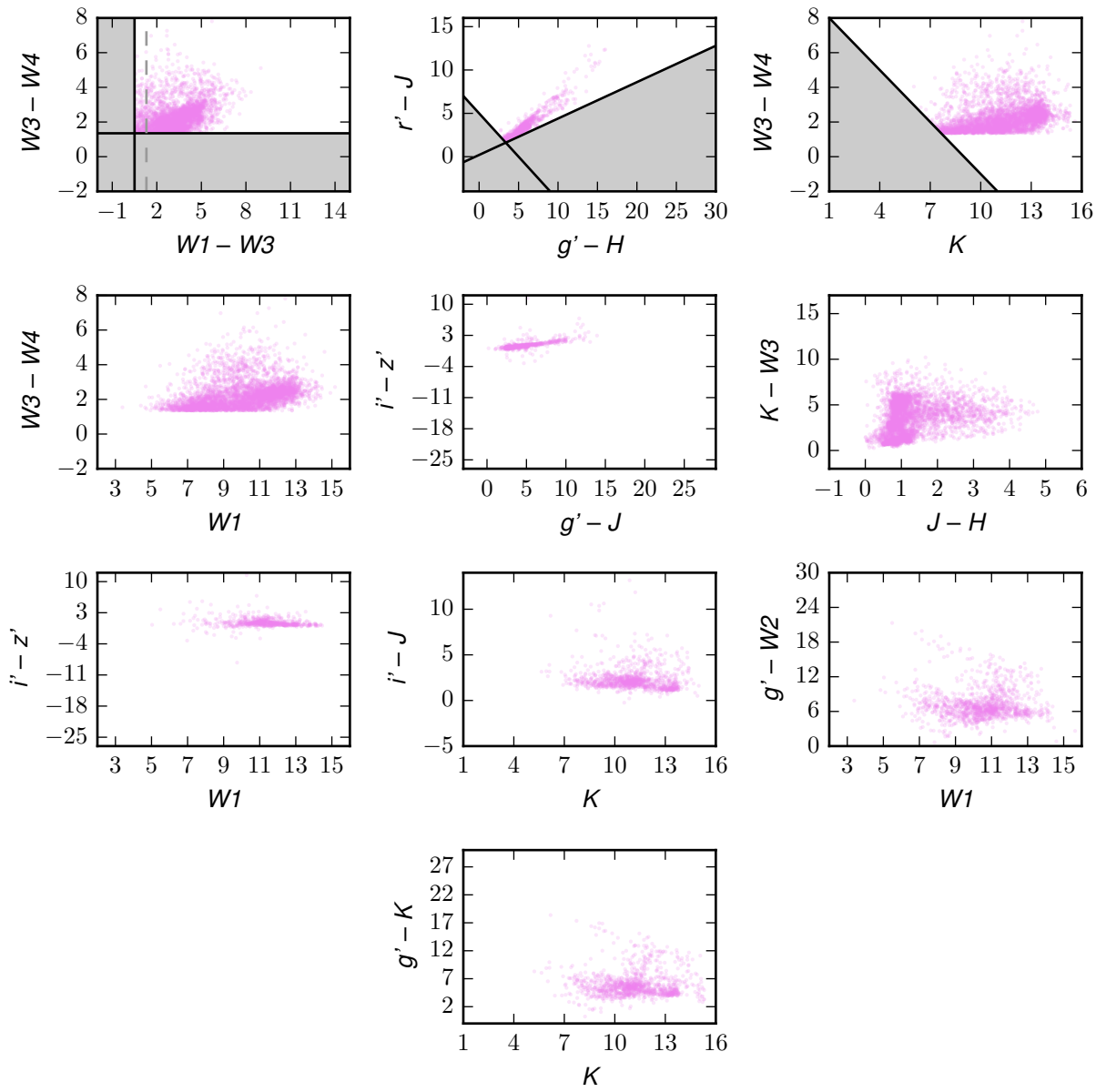

Figure B.8 — After the fifth color-color cut of Round 2: 5,908 (84.4%) known YSOs (mint, previous figure) are retained, while 34,077 (85.9%) known non-YSOs (fuchsia) are eliminated. Note that the next cut is the second to be made in the $W1 - W3$ color and is delineated with a dashed line.

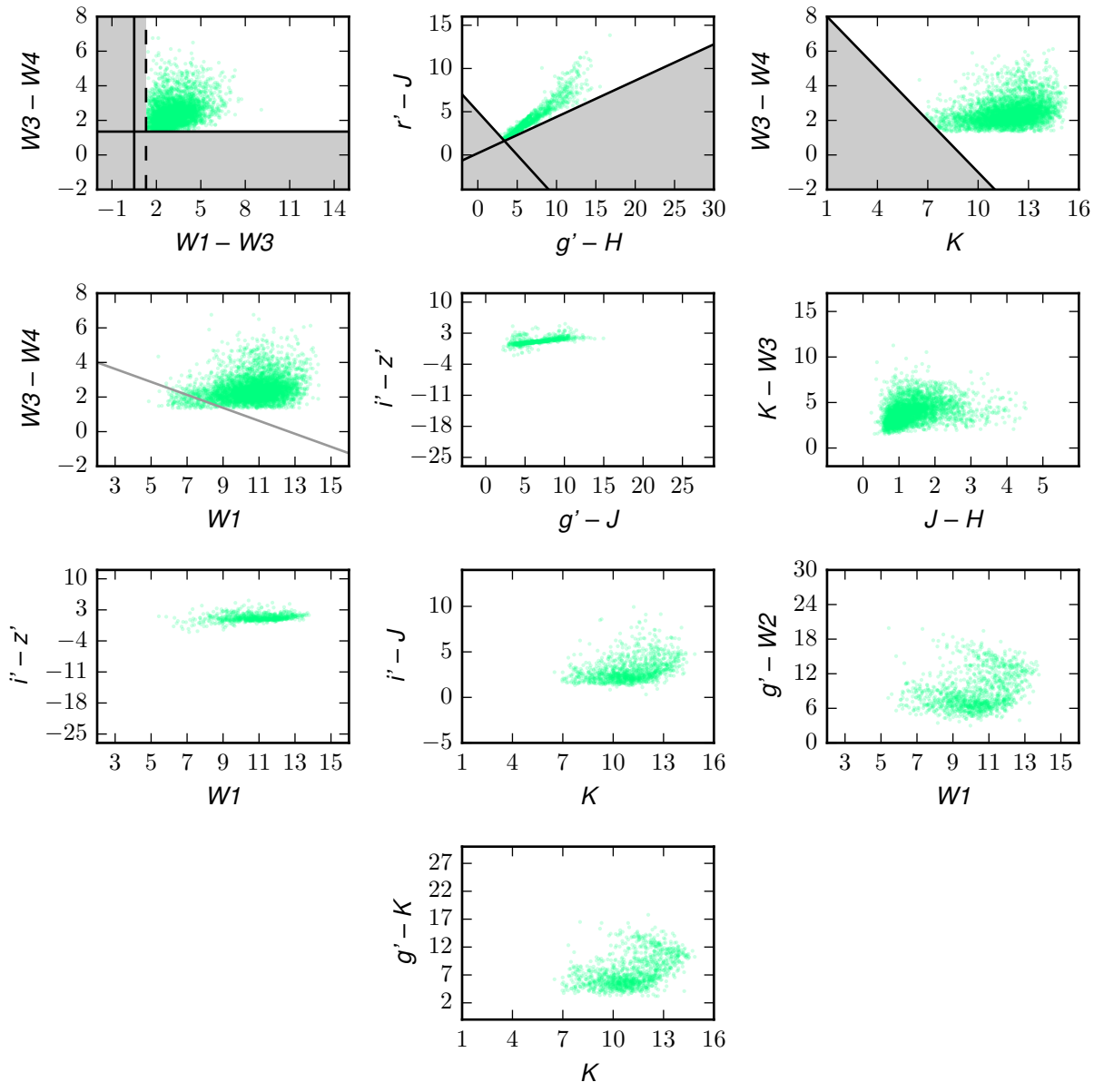

Figure B.9 — After the sixth color-color cut of Round 2: 5,676 (81.1%) known YSOs (mint) are retained, while 35,040 (88.3%) known non-YSOs (fuchsia, next figure) are eliminated.

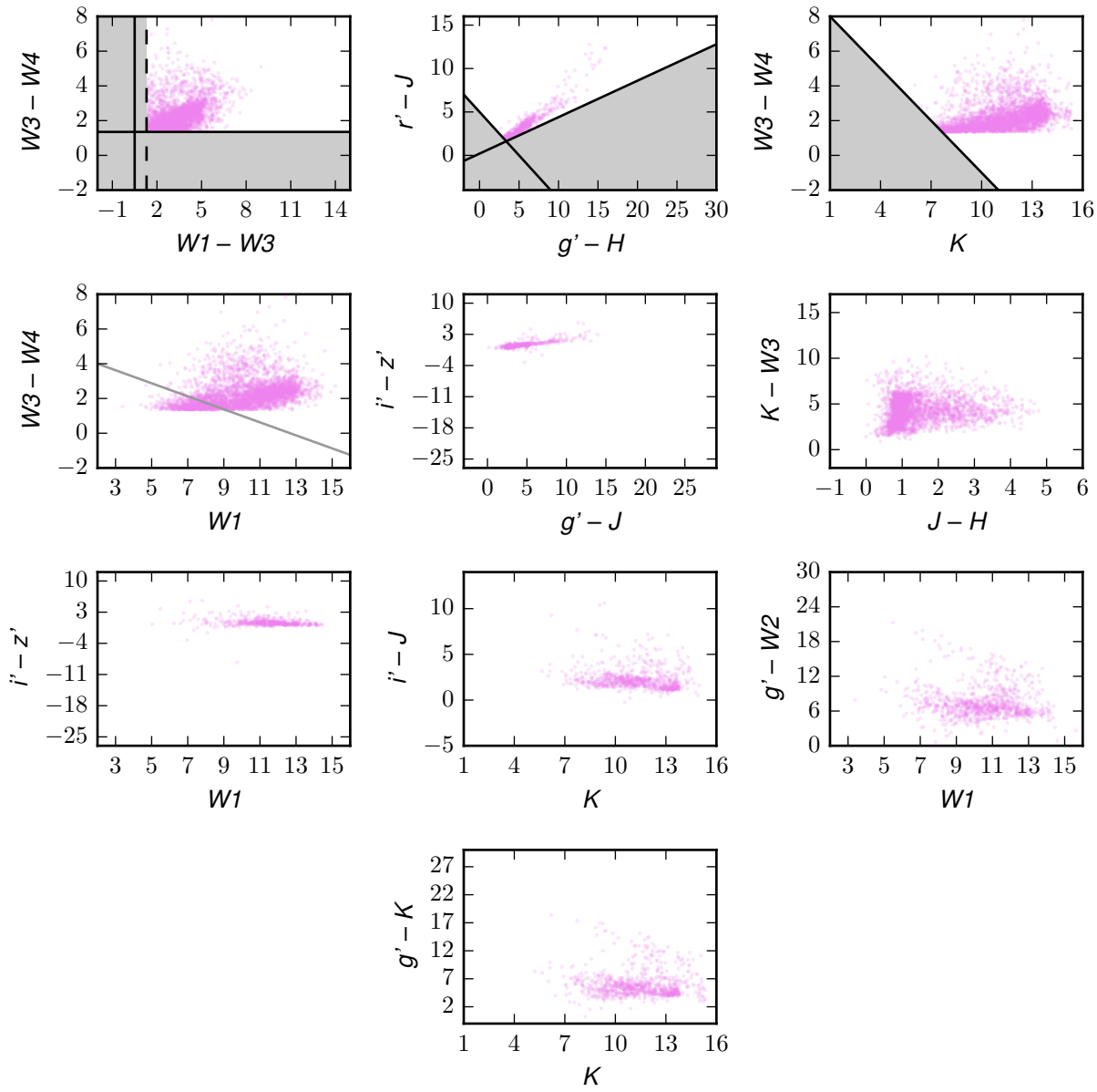

Figure B.10 — After the sixth color-color cut of Round 2: 5,676 (81.1%) known YSOs (mint, previous figure) are retained, while 35,040 (88.3%) known non-YSOs (fuchsia) are eliminated.

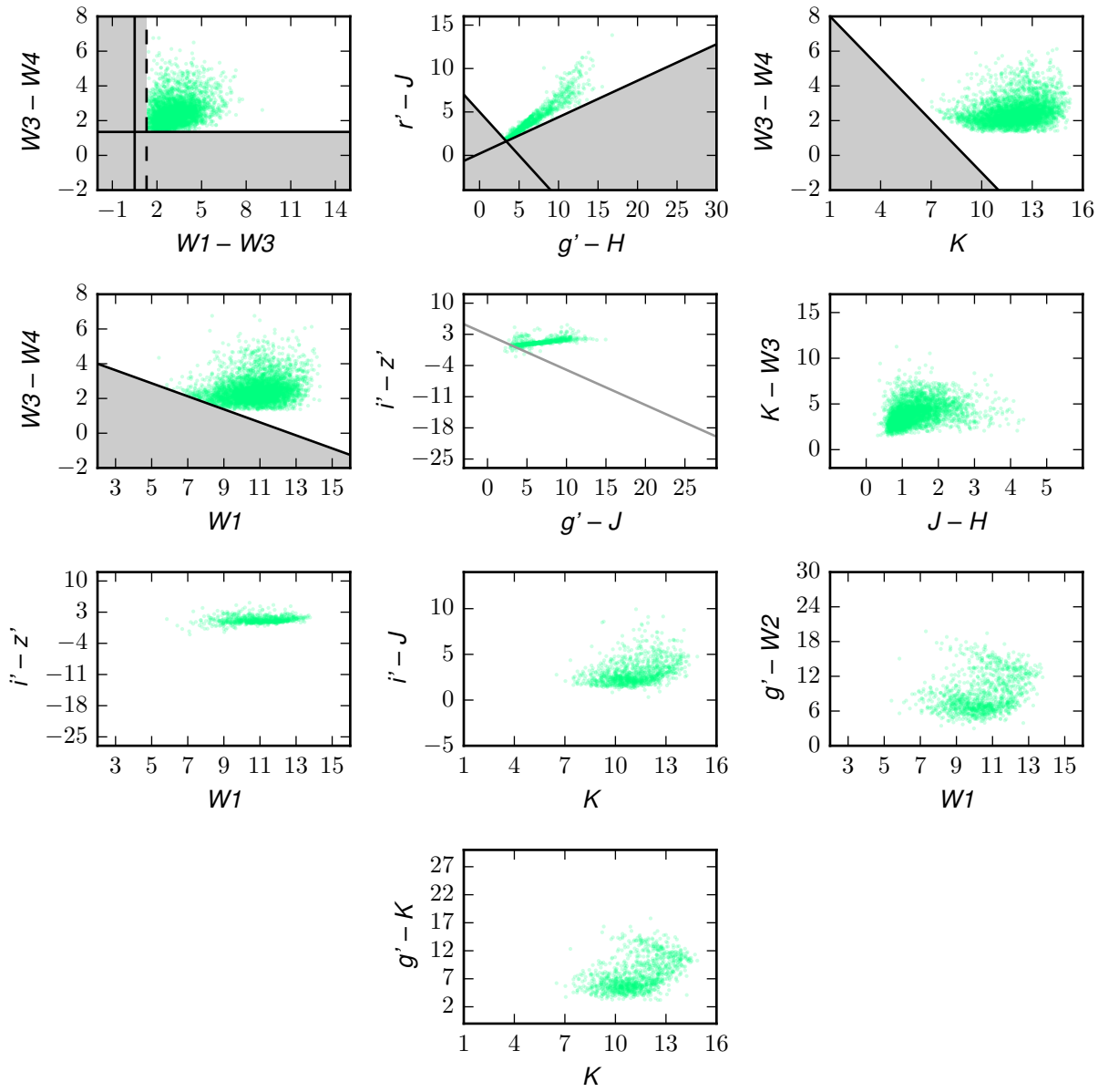

Figure B.11 — After the seventh color-color cut of Round 2: 5,514 (78.8%) known YSOs (mint) are retained, while 35,762 (90.1%) known non-YSOs (fuchsia, next figure) are eliminated.

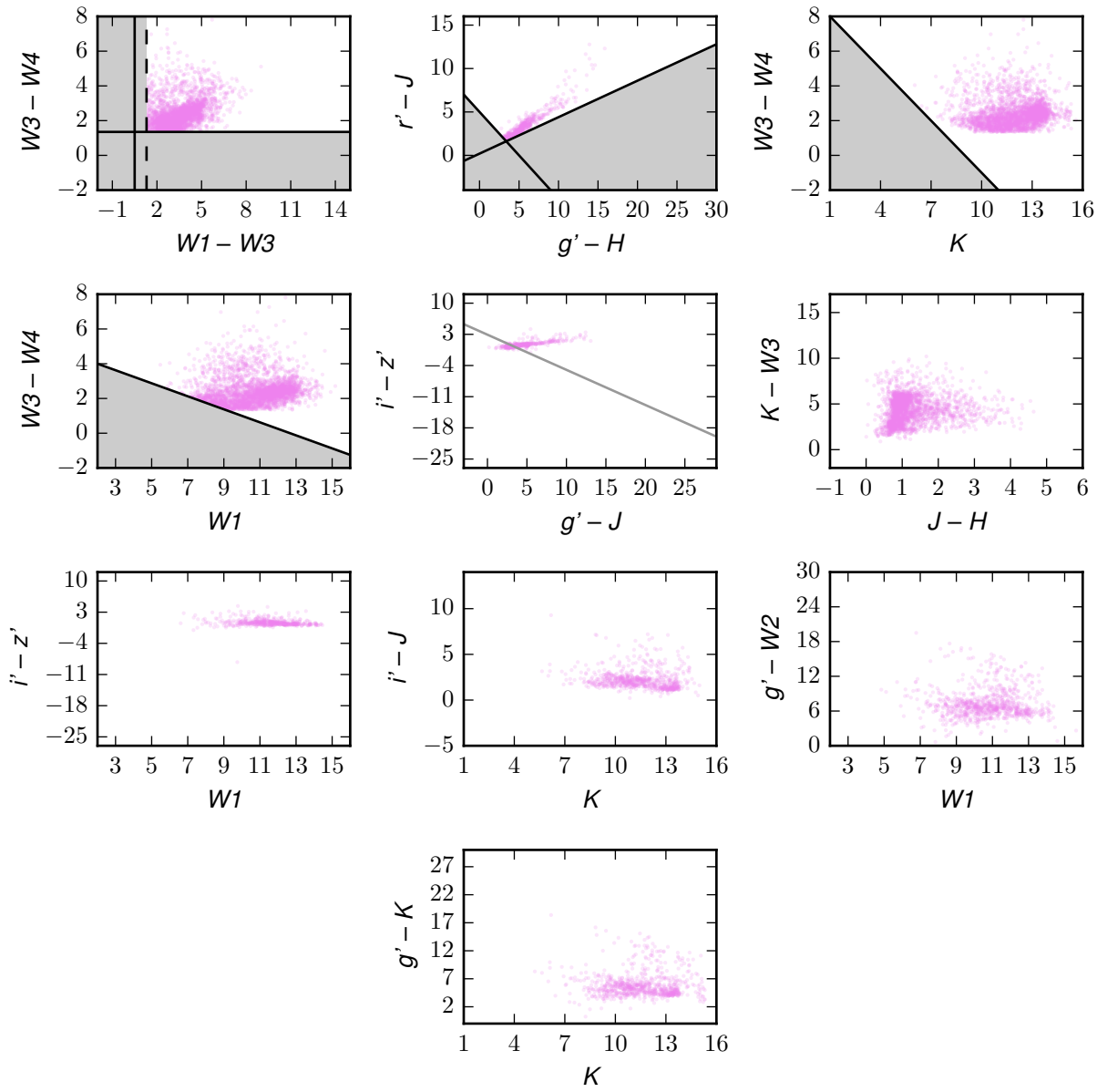

Figure B.12 — After the seventh color-color cut of Round 2: 5,514 (78.8%) known YSOs (mint, previous figure) are retained, while 35,762 (90.1%) known non-YSOs (fuchsia) are eliminated.

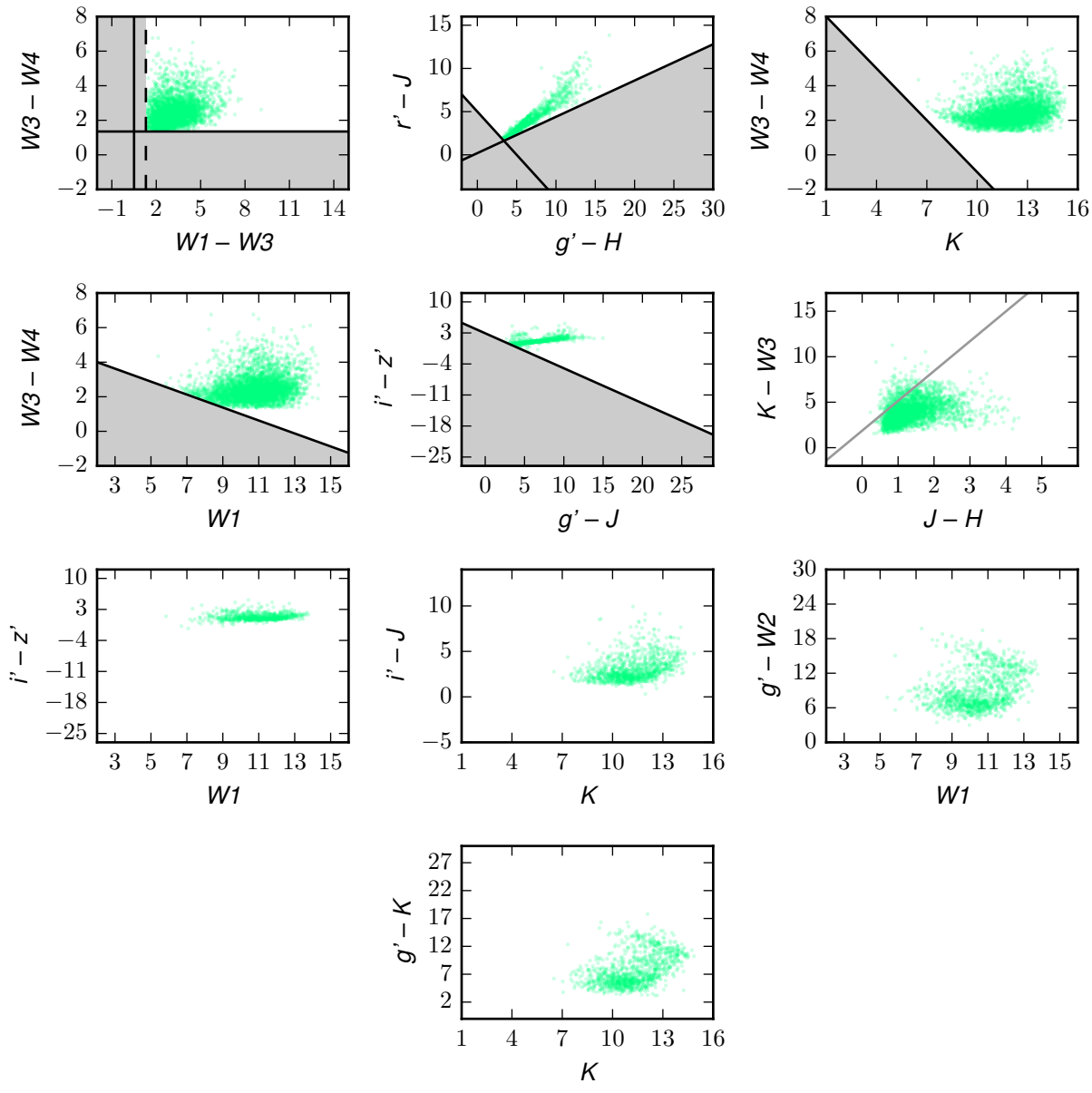

Figure B.13 — After the eighth color-color cut of Round 2: 5,498 (78.6%) known YSOs (mint) are retained, while 35,943 (90.6%) known non-YSOs (fuchsia, next figure) are eliminated.

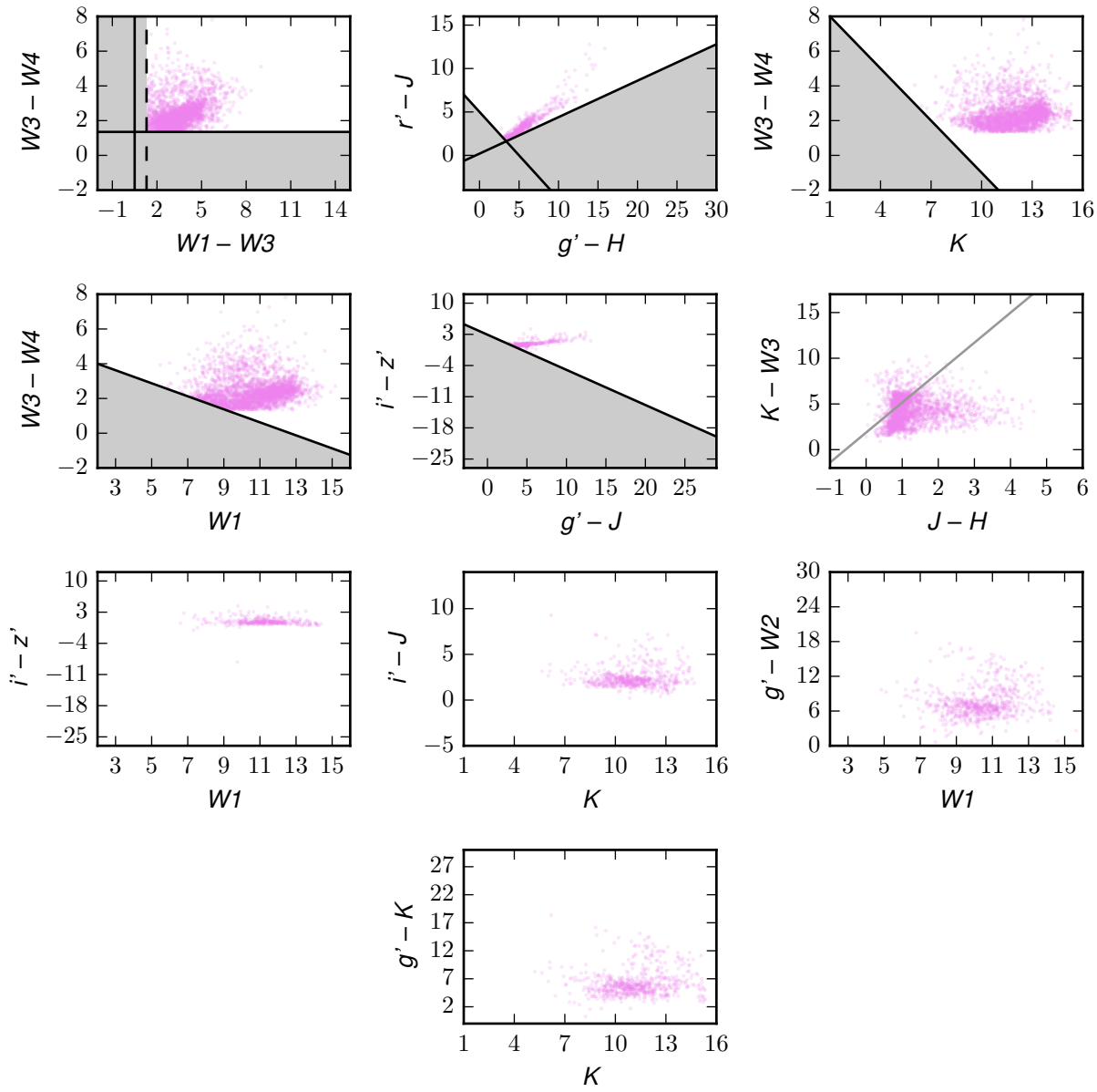

Figure B.14 — After the eighth color-color cut of Round 2: 5,498 (78.6%) known YSOs (mint, previous figure) are retained, while 35,943 (90.6%) known non-YSOs (fuchsia) are eliminated.

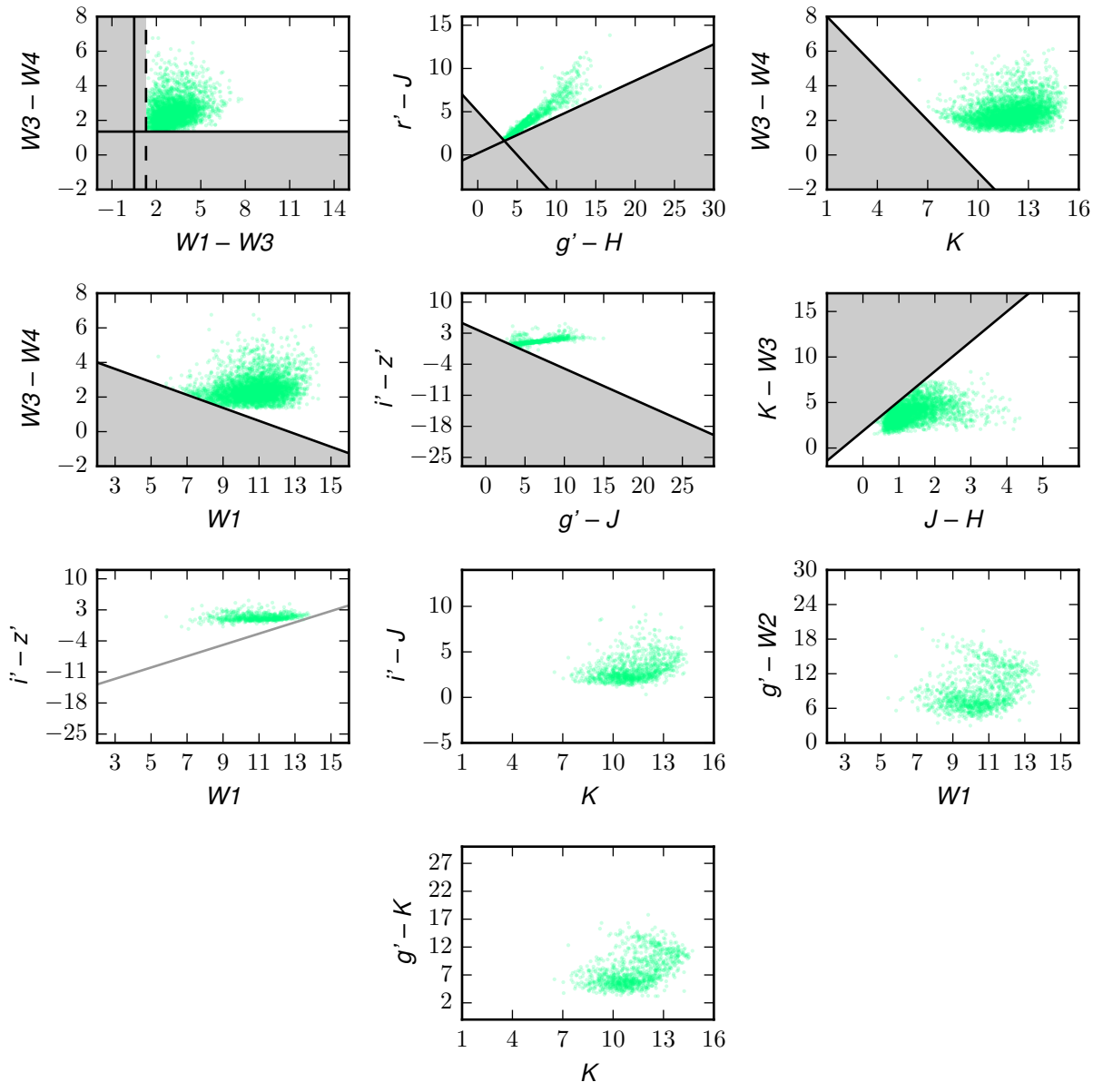

Figure B.15 — After the ninth color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint) are retained, while 36,641 (92.3%) known non-YSOs (fuchsia, next figure) are eliminated.

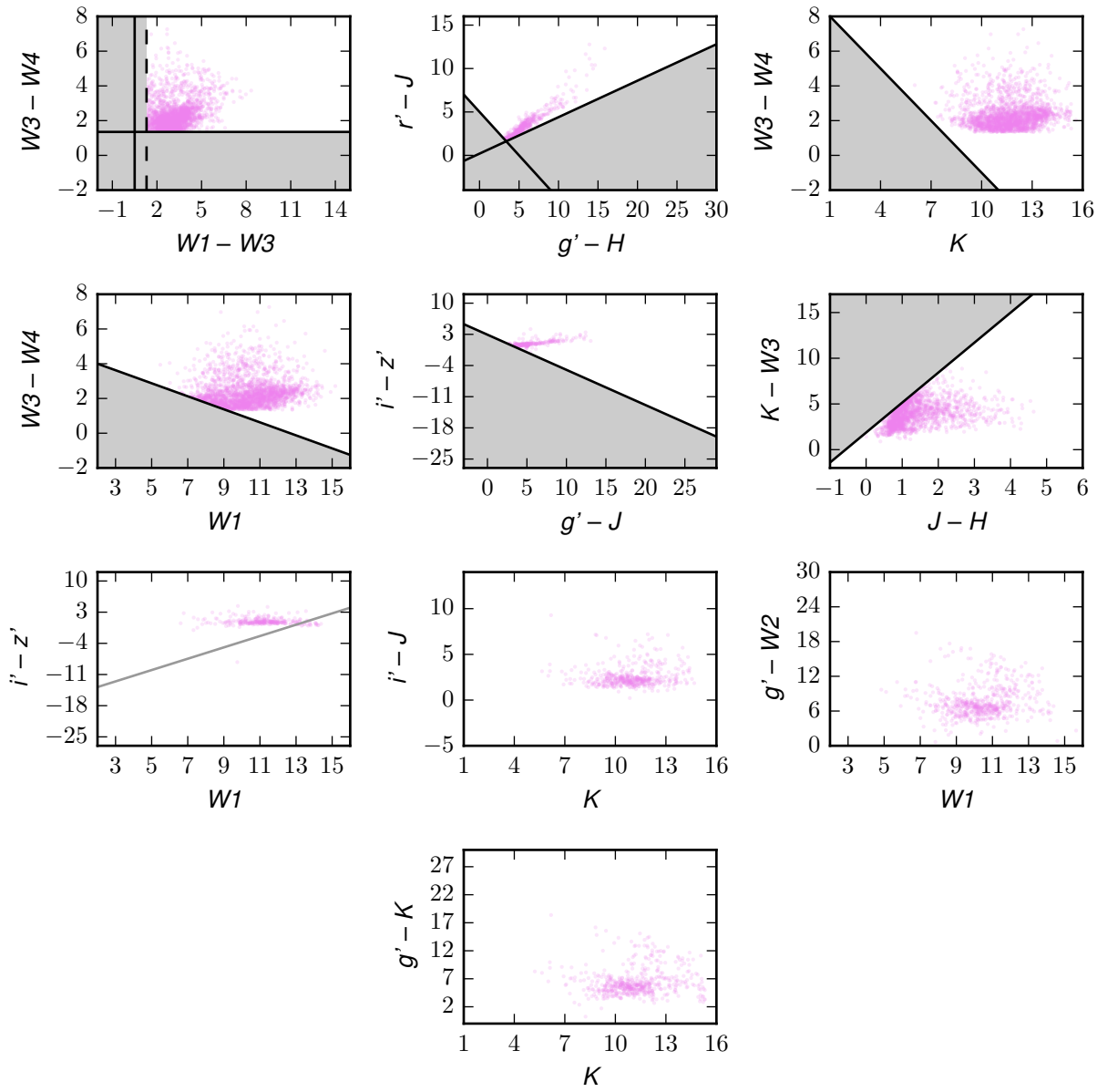

Figure B.16 — After the ninth color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint, previous figure) are retained, while 36,641 (92.3%) known non-YSOs (fuchsia) are eliminated.

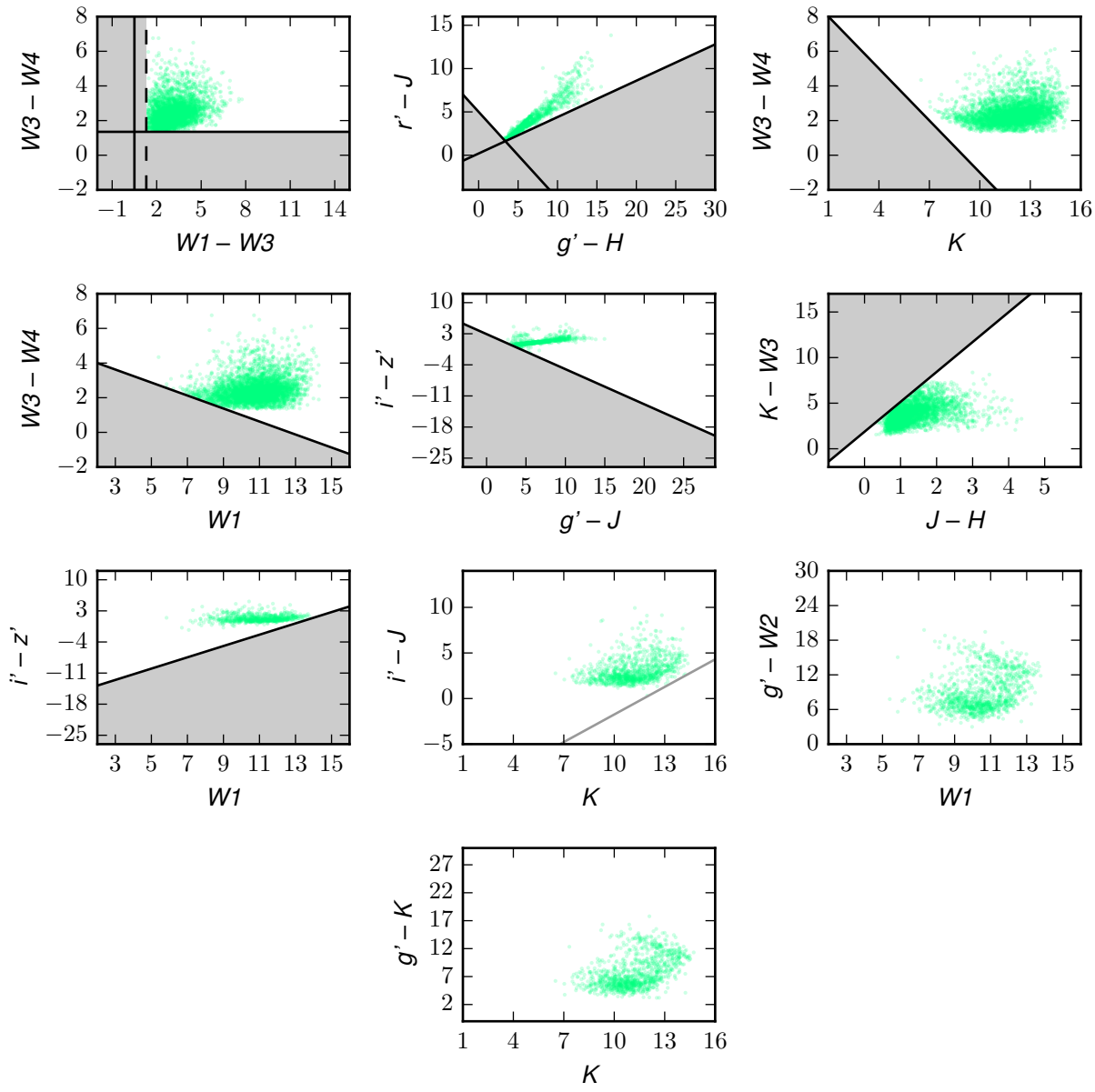

Figure B.17 — After the tenth color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint) are retained, while 36,664 (92.4%) known non-YSOs (fuchsia, next figure) are eliminated.

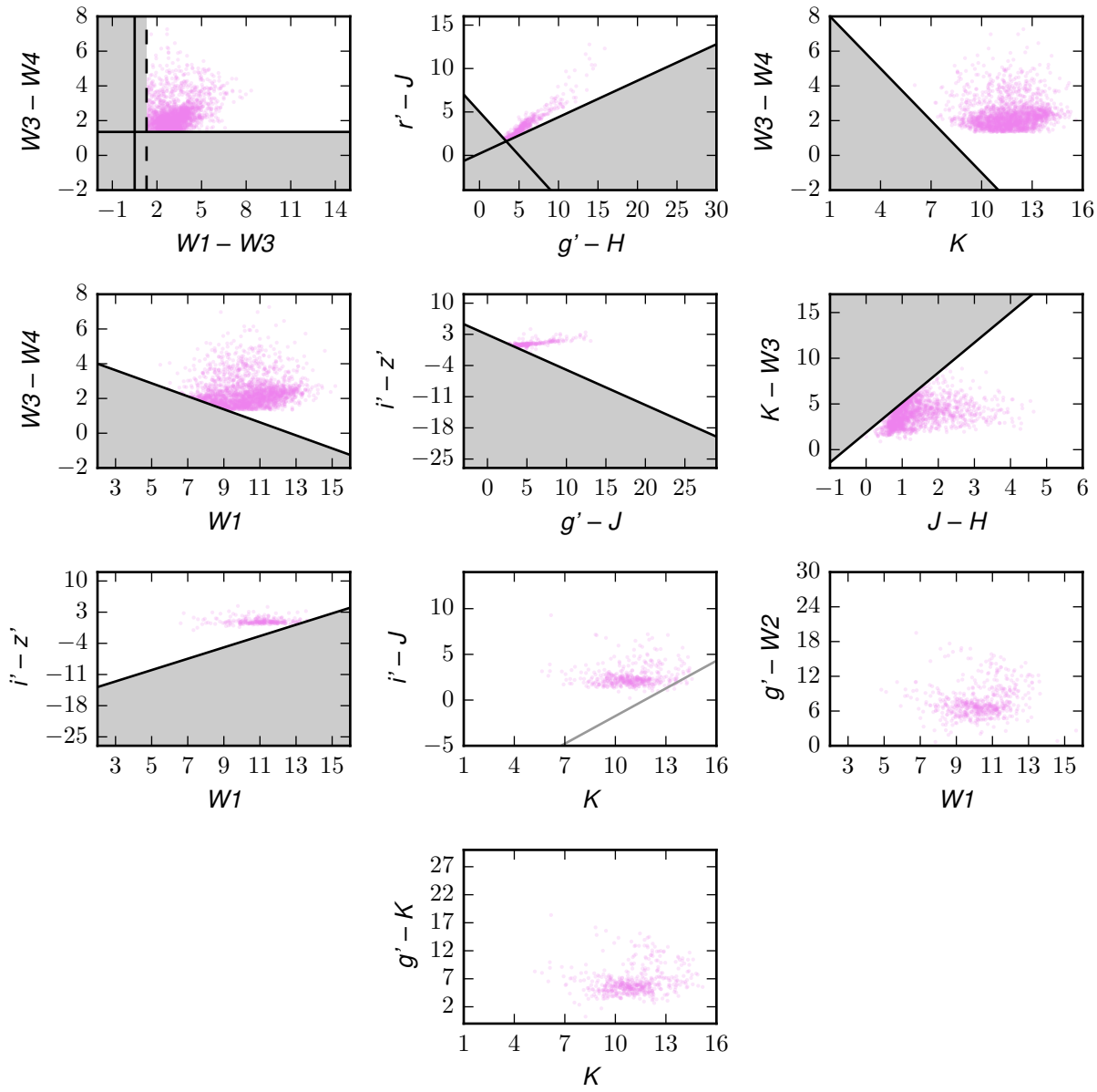

Figure B.18 — After the tenth color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint, previous figure) are retained, while 36,664 (92.4%) known non-YSOs (fuchsia) are eliminated.

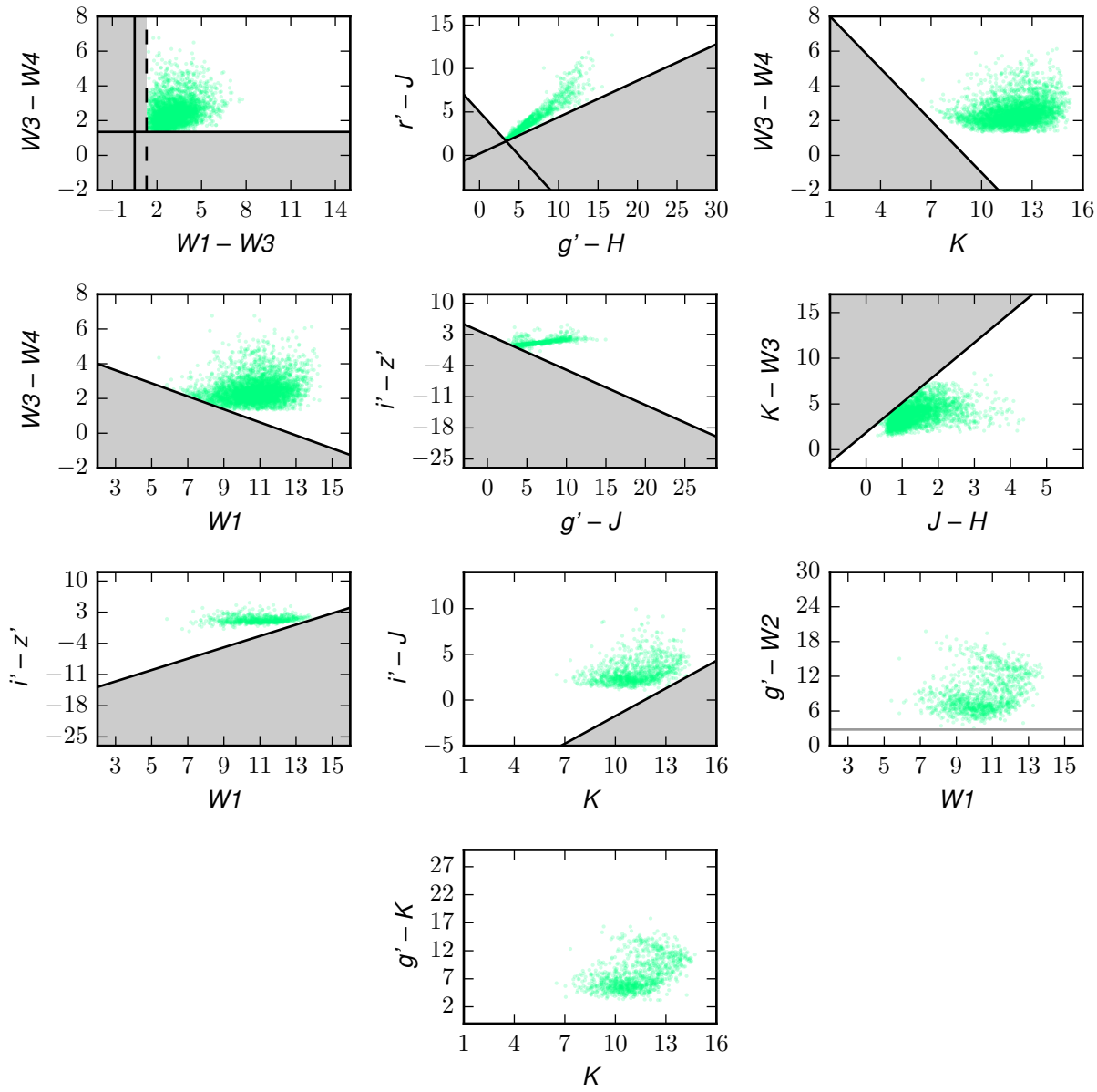

Figure B.19 — After the eleventh color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint) are retained, while 36,675 (92.4%) known non-YSOs (fuchsia, next figure) are eliminated.

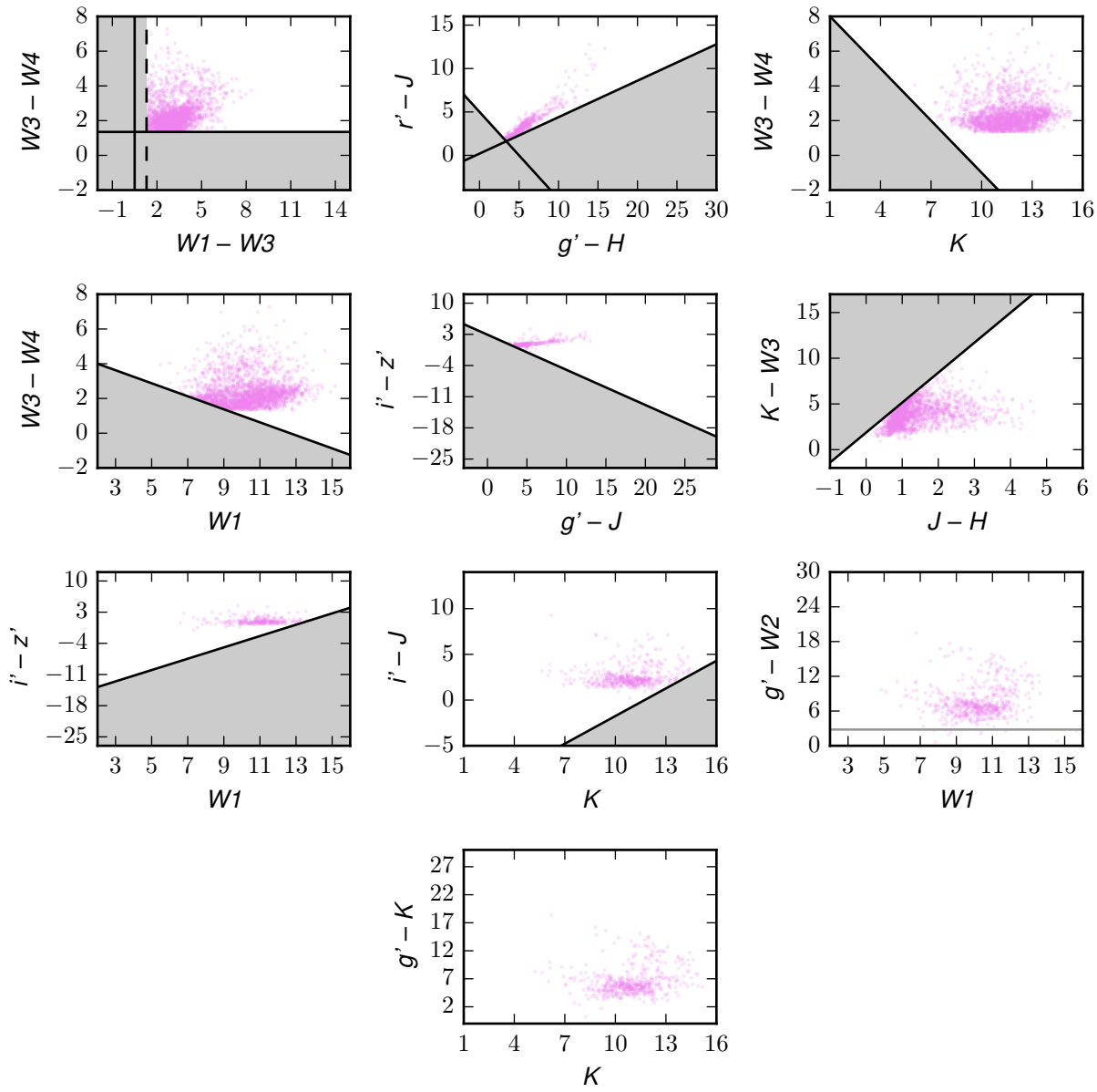

Figure B.20 — After the eleventh color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint, previous figure) are retained, while 36,675 (92.4%) known non-YSOs (fuchsia) are eliminated.

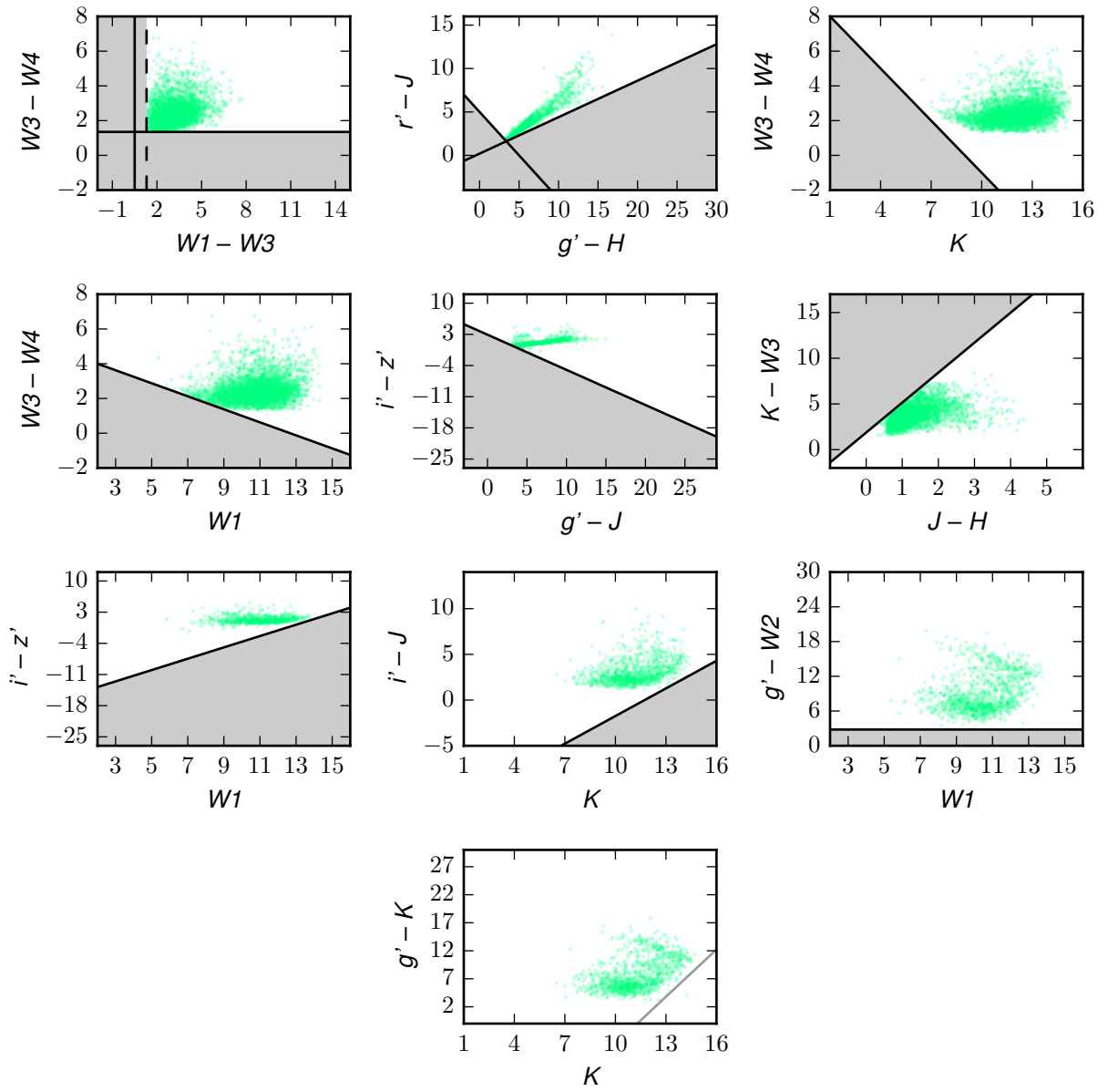

Figure B.21 — After the twelfth color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint) are retained, while 36,685 (92.5%) known non-YSOs (fuchsia, next figure) are eliminated.

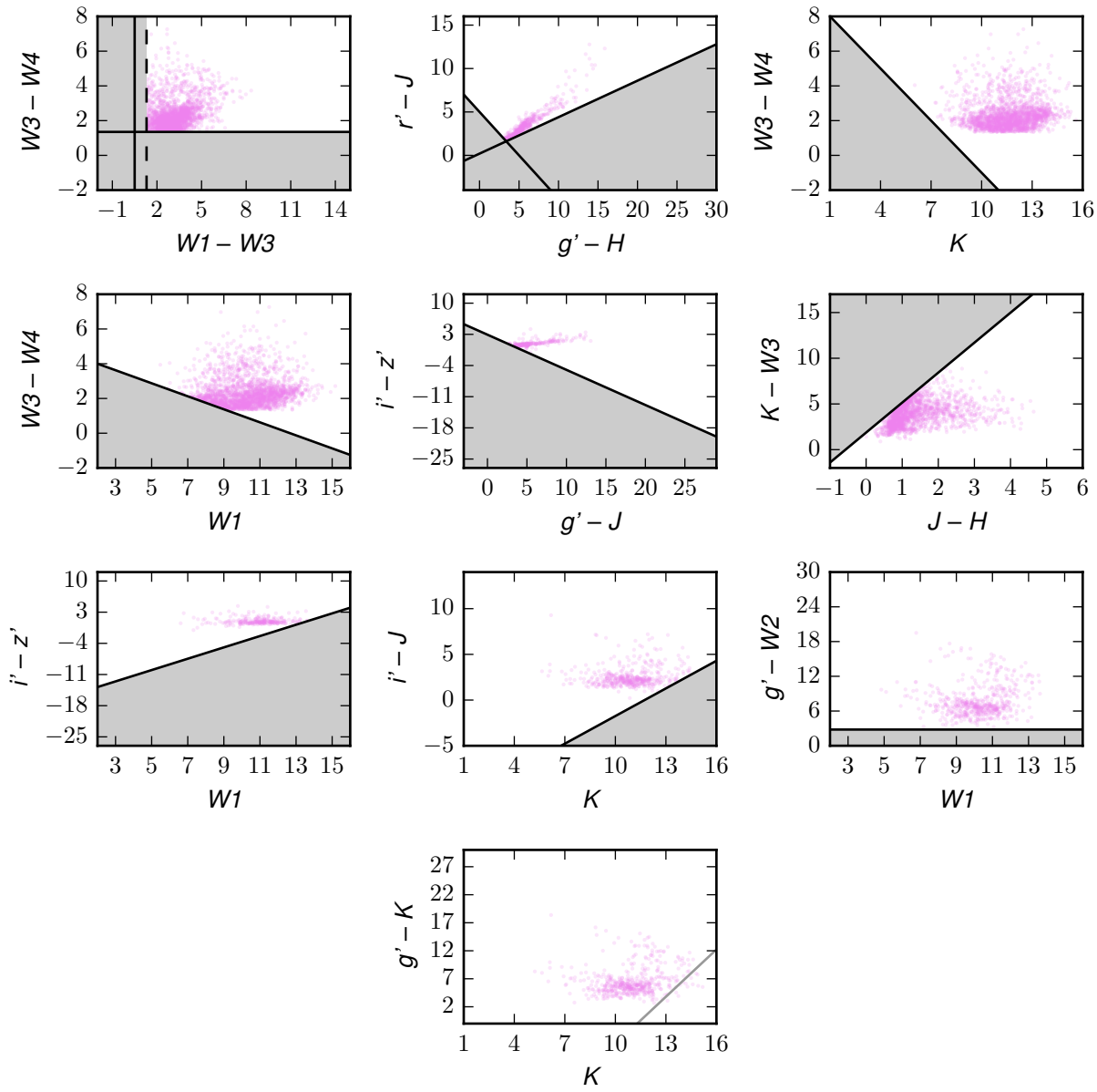

Figure B.22 — After the twelfth color-color cut of Round 2: 5,280 (75.5%) known YSOs (mint, previous figure) are retained, while 36,685 (92.5%) known non-YSOs (fuchsia) are eliminated.

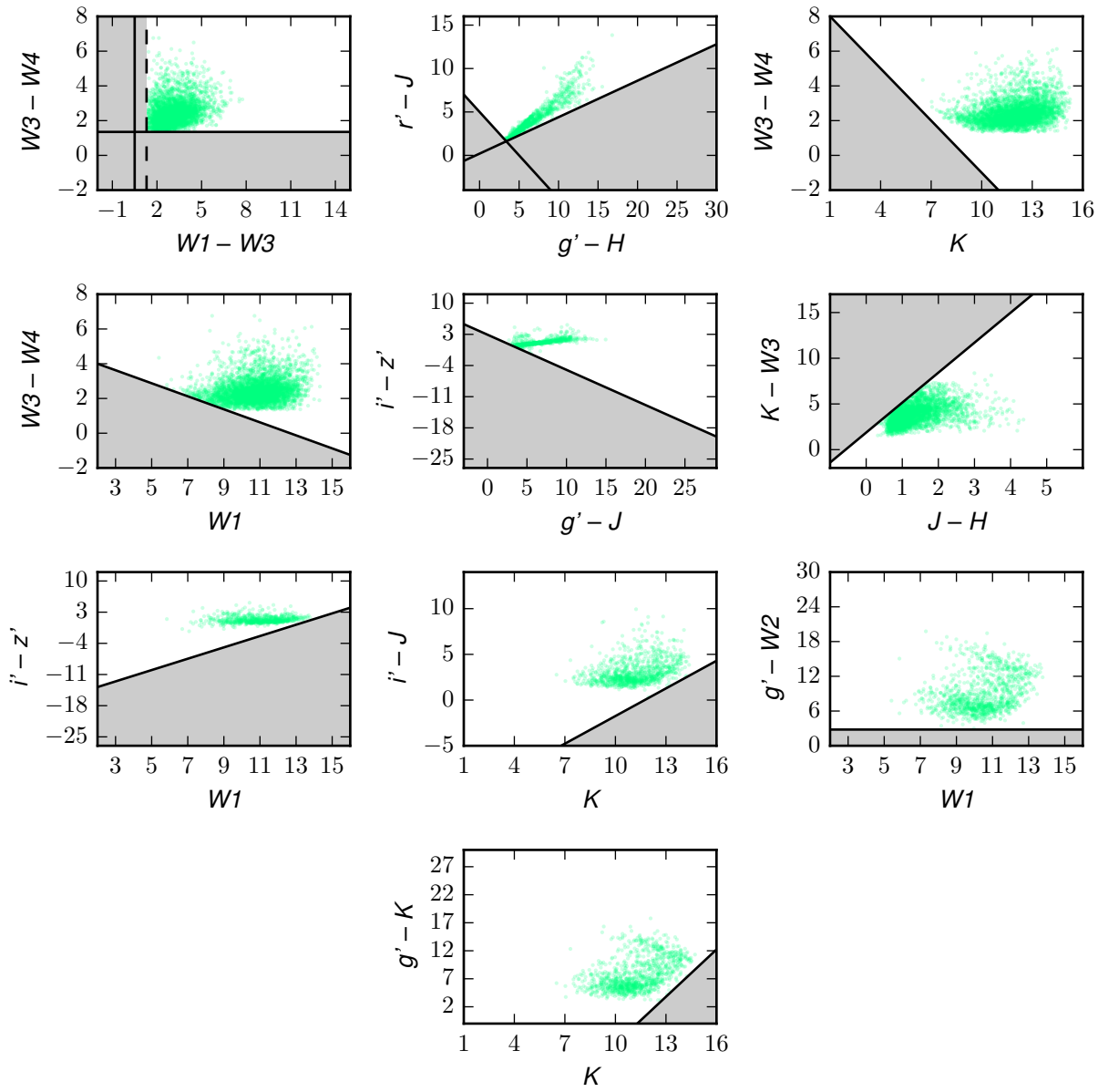

Figure B.23 — After the thirteenth and final color-color cut of Round 2: 5,279 (75.4%) known YSOs (mint) are retained, while 36,698 (92.5%) known non-YSOs (fuchsia, next figure) are eliminated.

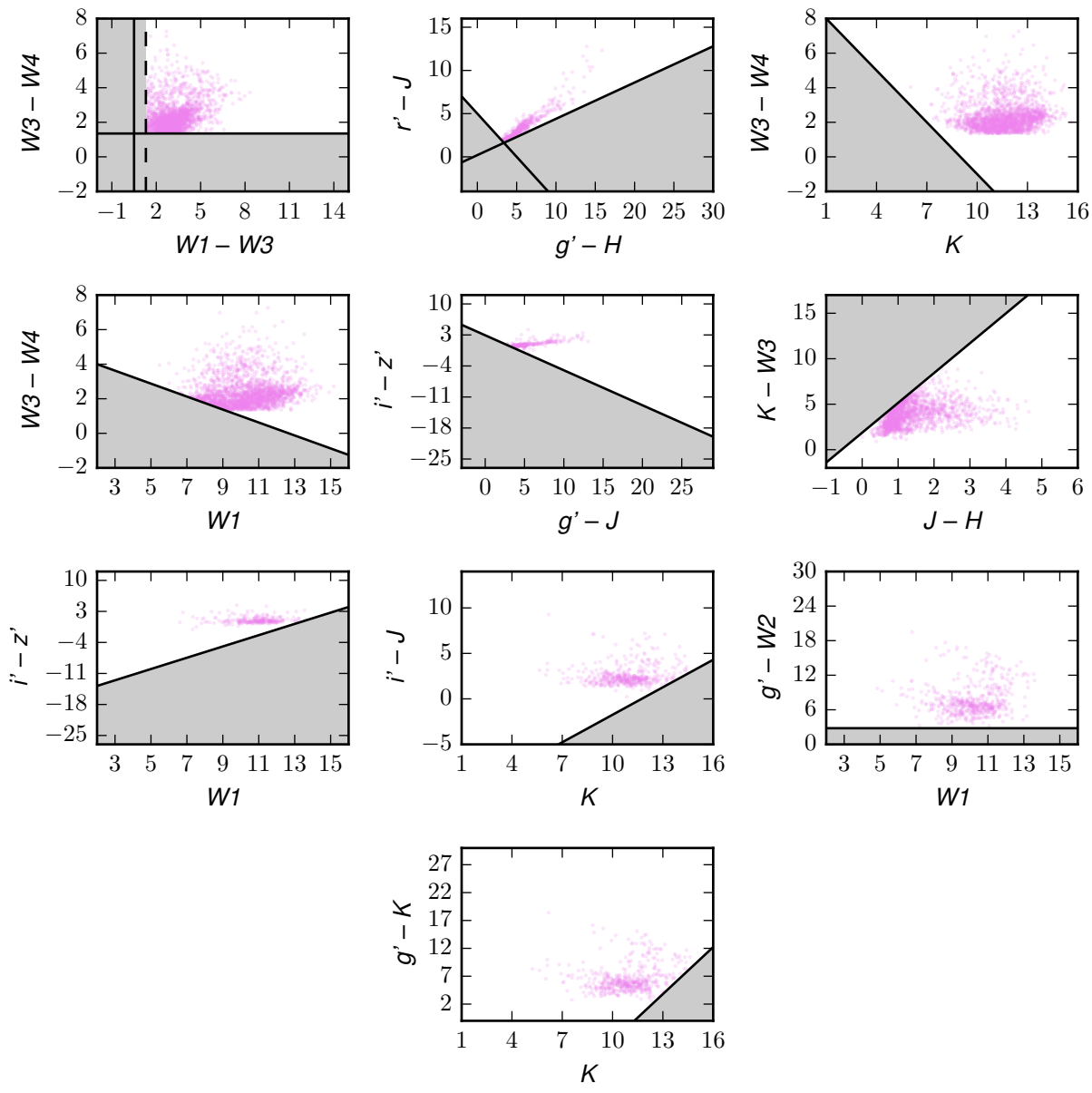

Figure B.24 — After the thirteenth and final color-color cut of Round 2: 5,279 (75.4%) known YSOs (mint, previous figure) are retained, while 36,698 (92.5%) known non-YSOs (fuchsia) are eliminated.

C Known Star-Forming Regions

Previously known star-forming regions, as labeled in Figure 2.5, are highlighted here. Following Figure 2.5, the coordinates of new candidate protoplanetary disks are shown in blue; known young stellar objects (YSOs) recovered by the color selection process as detailed in Sections 2.3 through 2.5 are shown in mint; and known non-YSOs erroneously recovered through the color selection process (false positives) are shown in fuchsia. The LMC and SMC — while containing some known YSOs — show few new candidates and are not highlighted. In contrast to prior figures, the points here are marginally larger and are 100% opaque, in order to make single sources more visible at these more magnified, and hence more sparsely populated, spatial scales.

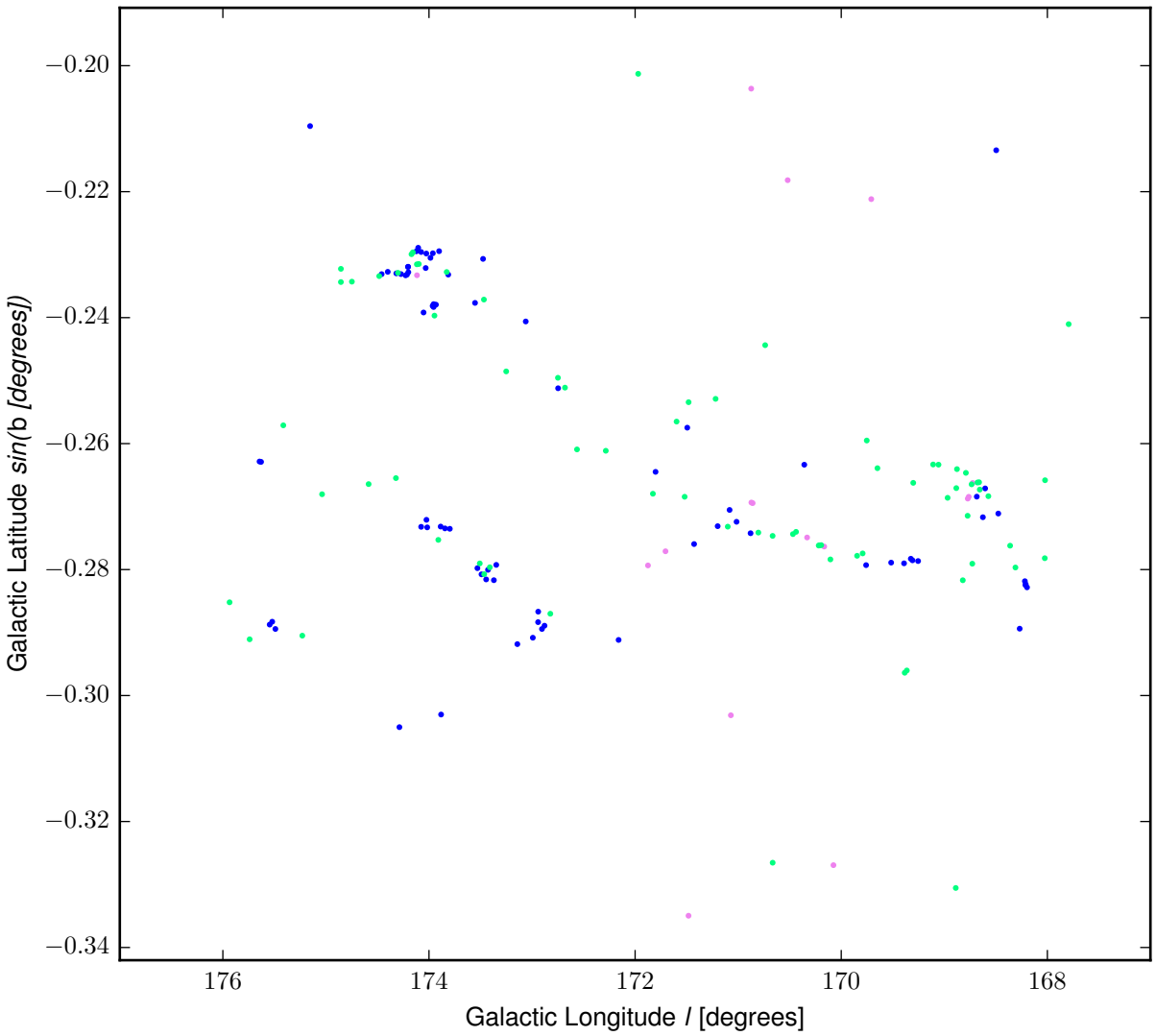

Figure C.1 — Known YSOs (mint) in the Taurus star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.

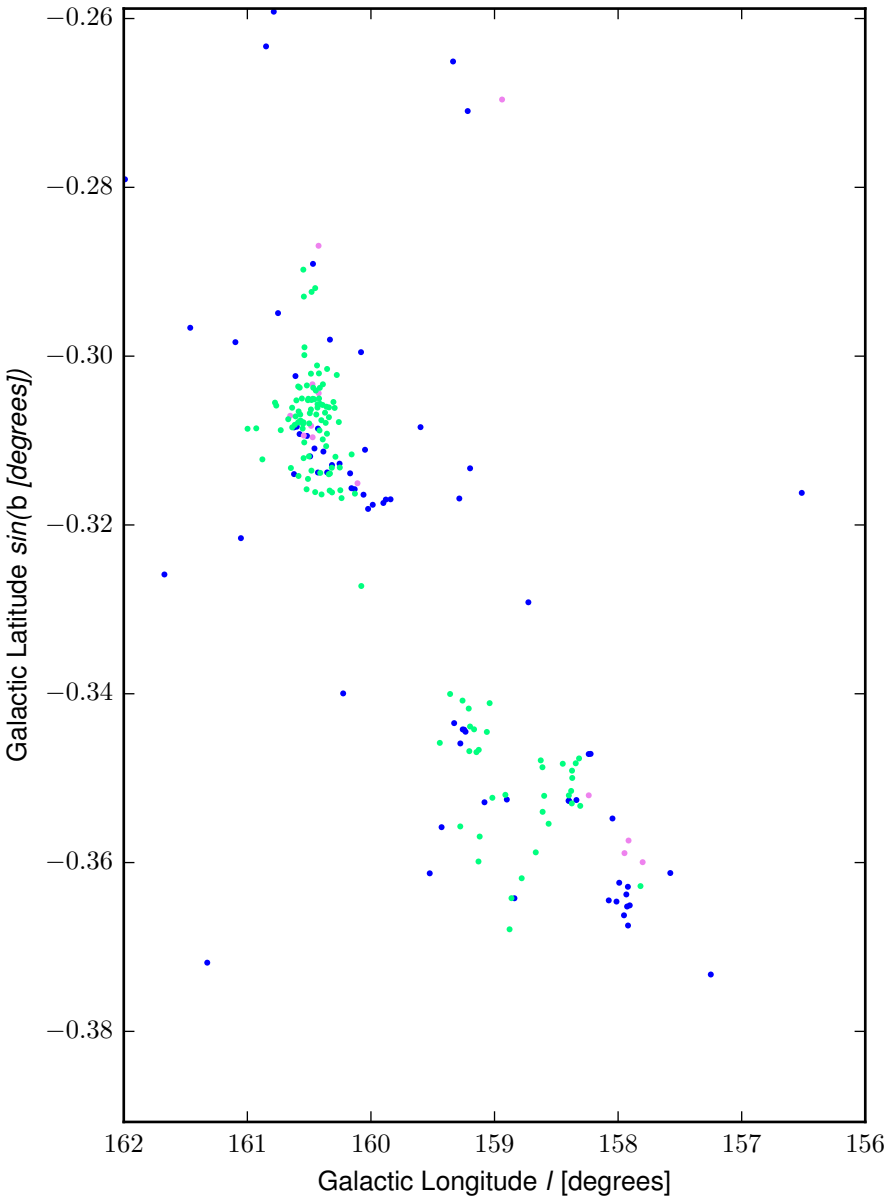

Figure C.2 — Known YSOs (mint) in the Perseus star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.

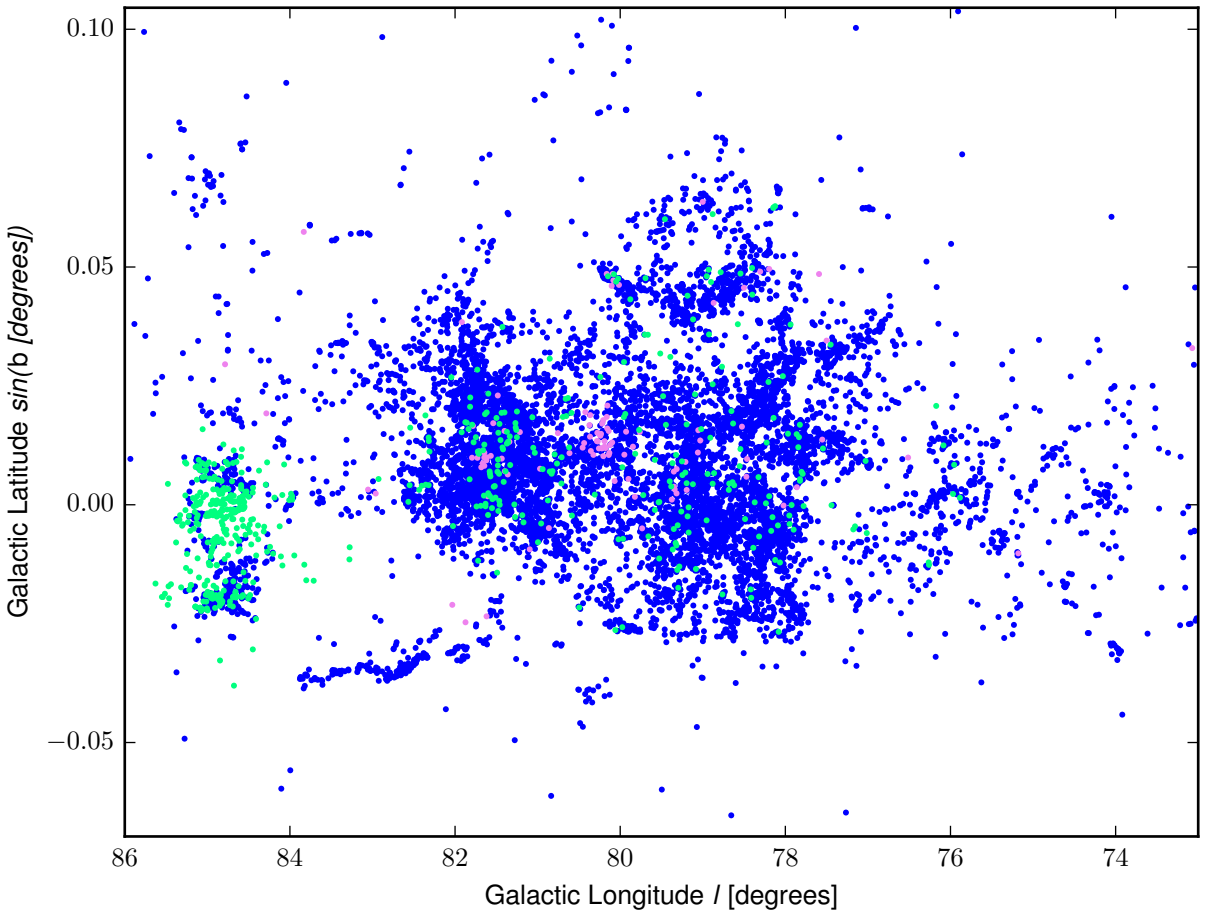

Figure C.3 — Known YSOs (mint) in the AFGL 2591 star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.

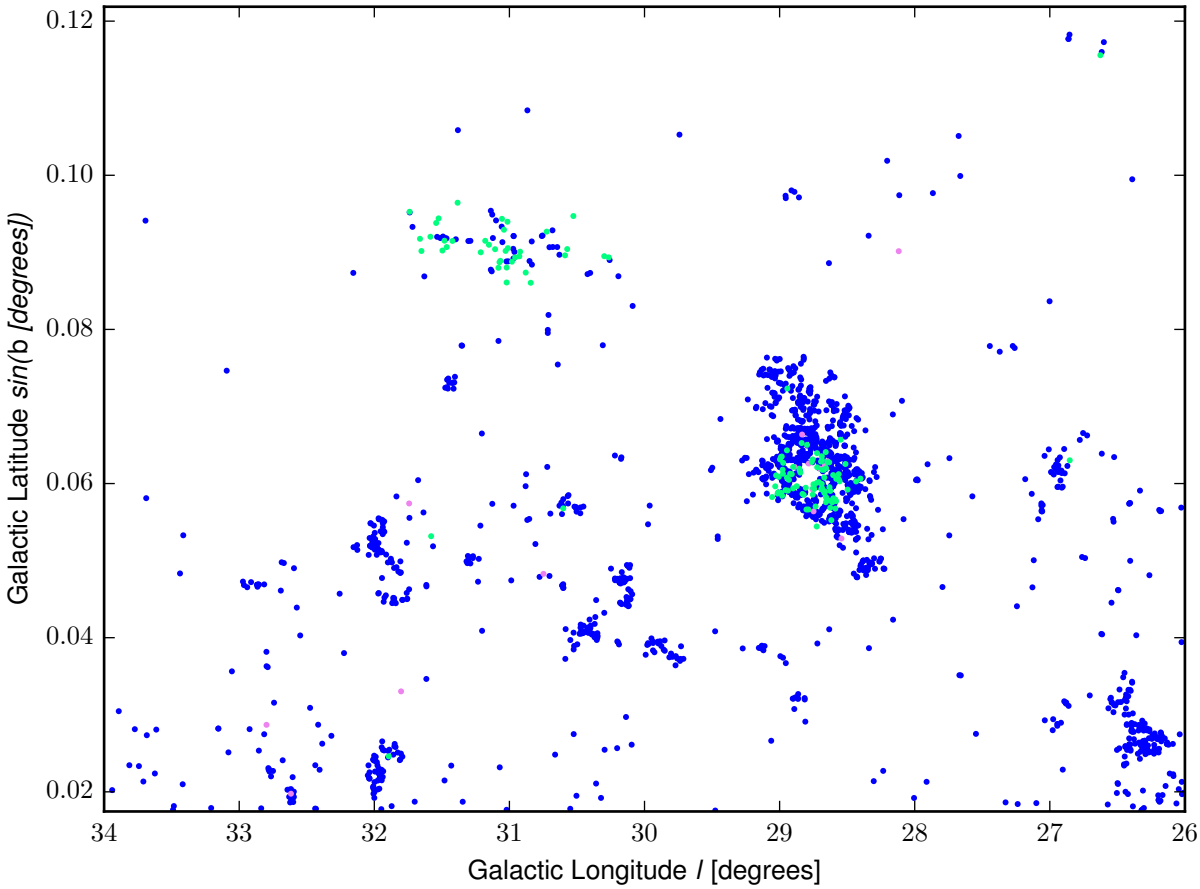

Figure C.4 — Known YSOs (mint) in the Serpens star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.

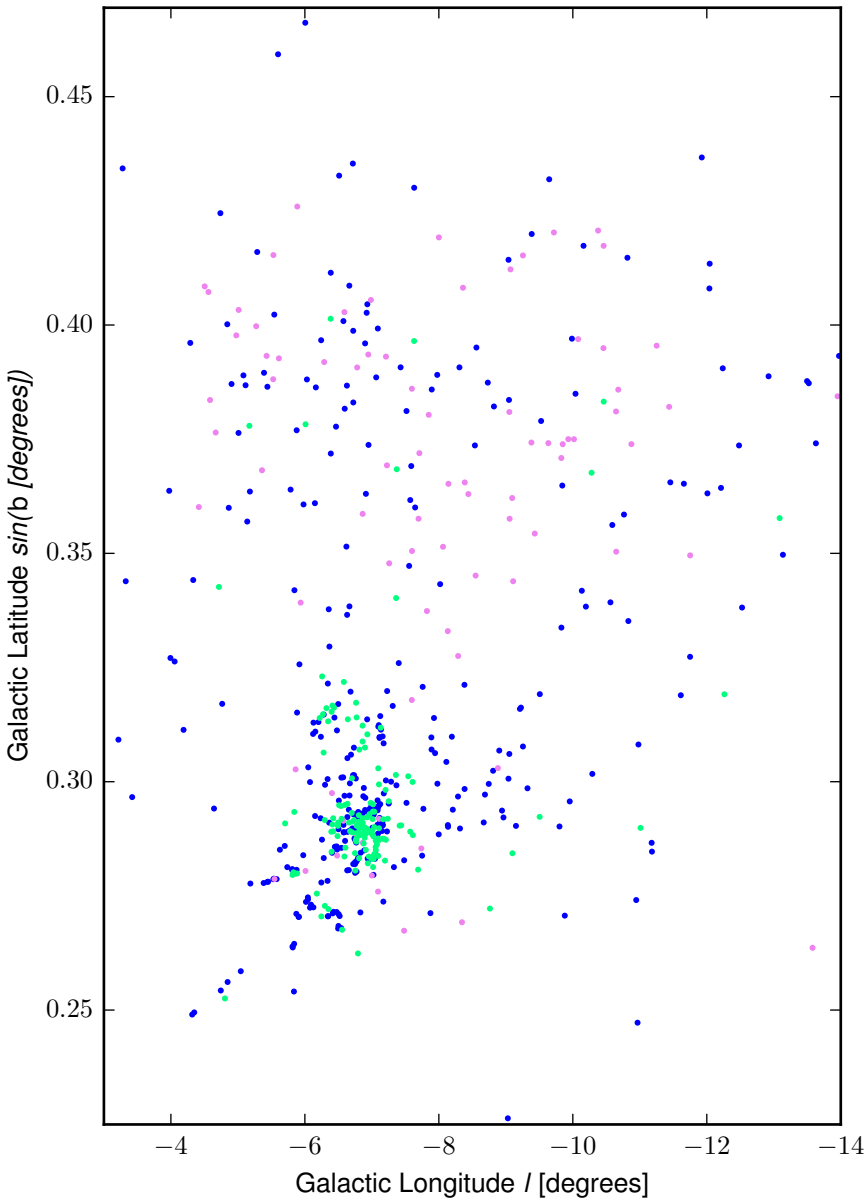

Figure C.5 — Known YSOs (mint) in the ρ Ophiuchi star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.

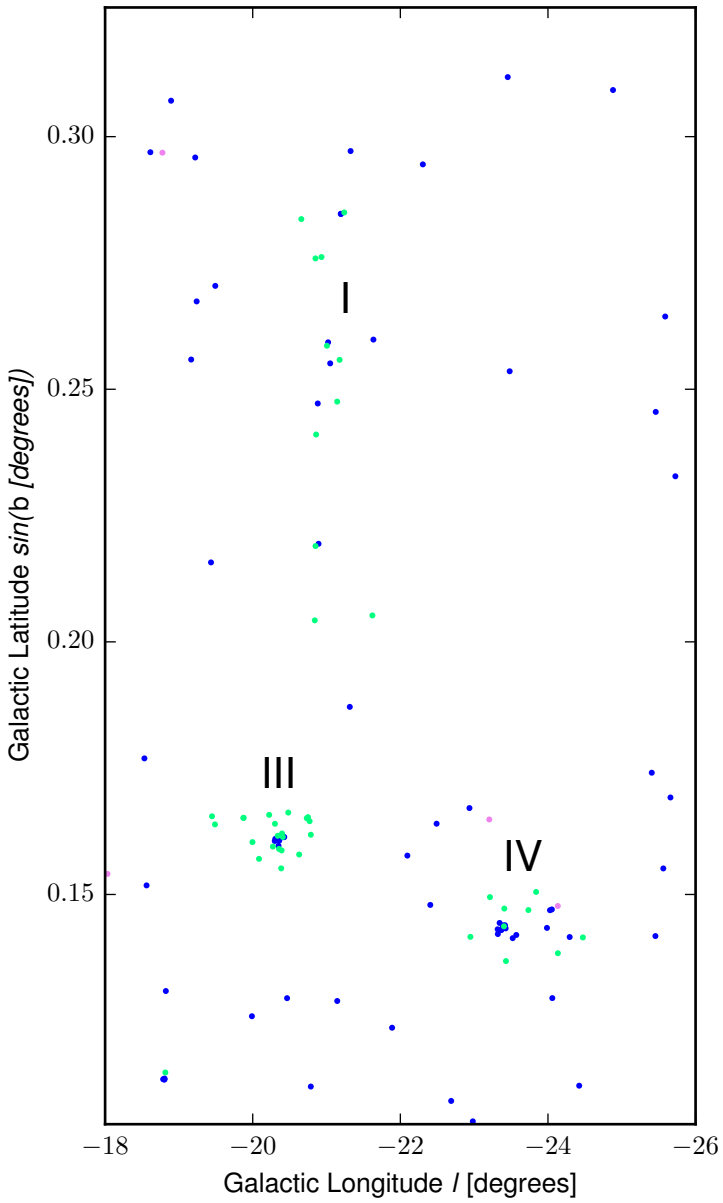

Figure C.6 — Known YSOs (mint) in the Lupus star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown. Numerical designations follow Evans, II et al. (2003).

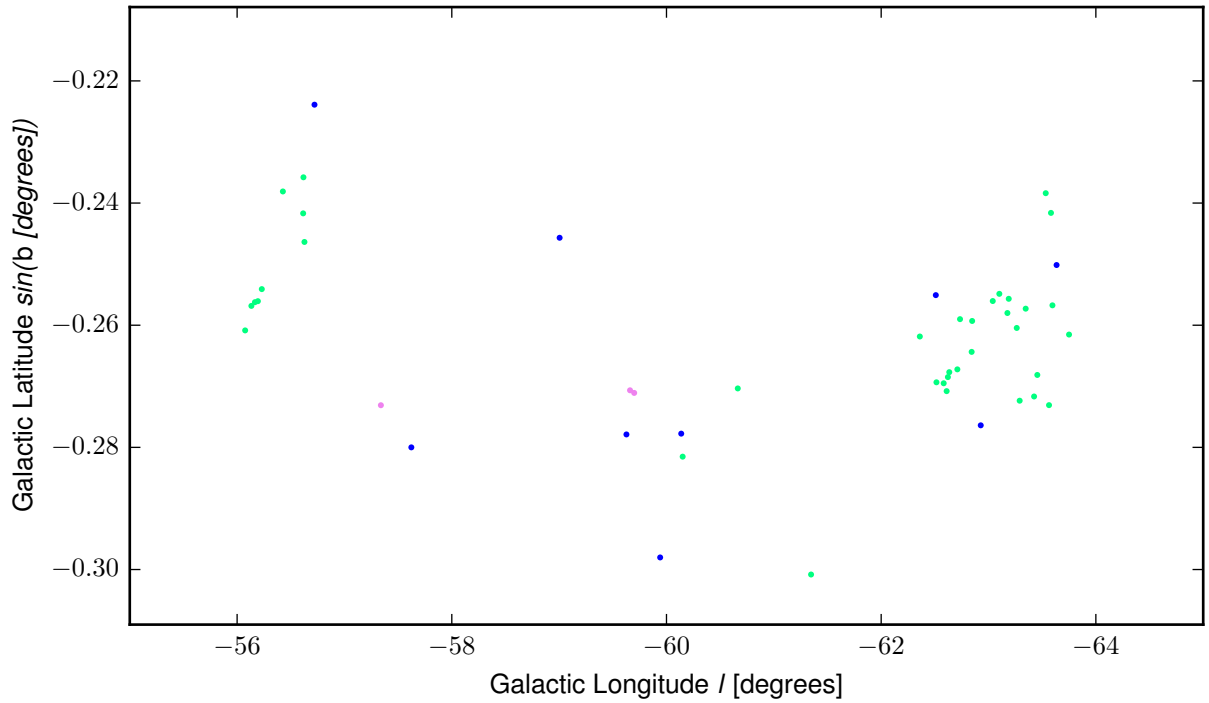

Figure C.7 — Known YSOs (mint) in the Chameleon star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown. The numerical designation follows Evans, II et al. (2003).

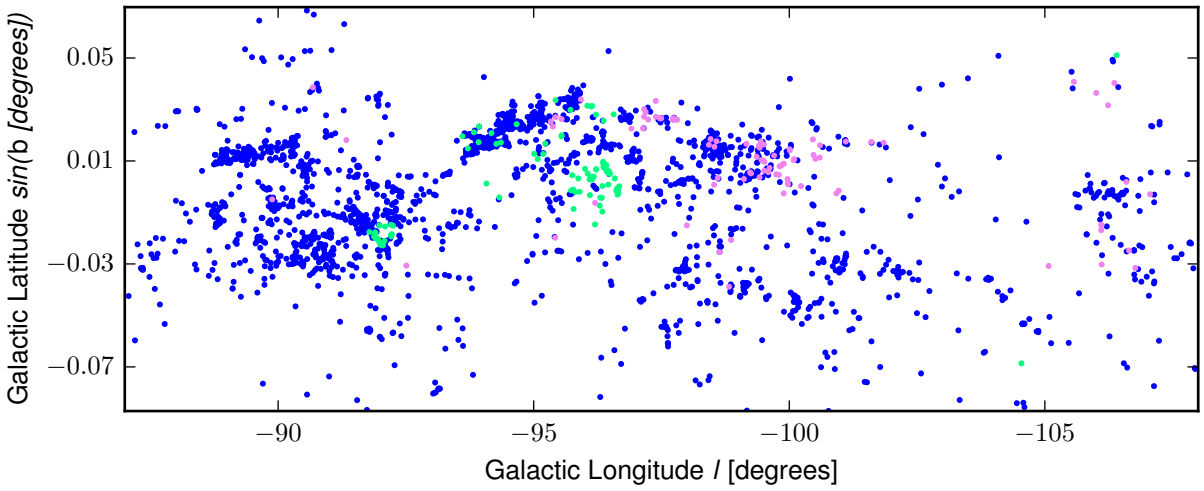

Figure C.8 — Known YSOs (mint) in the RCW 36 star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.

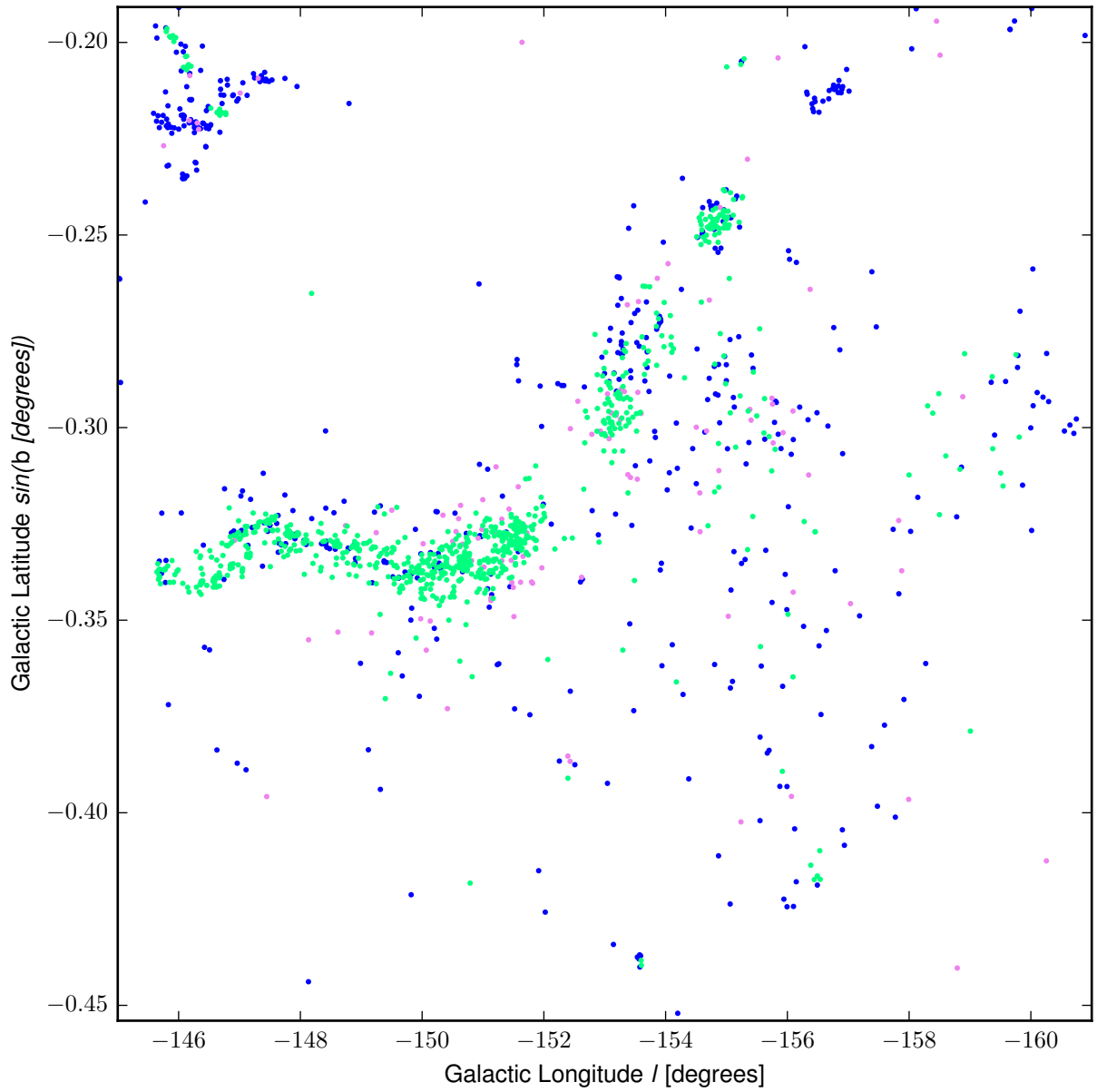

Figure C.9 — Known YSOs (mint) in the Orion star-forming region are successfully recovered through the color selection process, as well as known non-YSOs (fuchsia) erroneously recovered as false positives. New candidate protoplanetary disks (blue) are also shown.