

Alan Lin
BiGP
Prof. Banta
27 May 2021

Use of Monocle algorithm on MS/MS data to improve peptide identification

Abstract

In the world of bottom-up proteomics, using accurate monoisotopic peaks is important to identify peptides in a sample. Not only does the number of peptides identified decrease when the wrong monoisotopic mass is used, the accuracy of peptide matches also decreases. To combat this, Monocle was developed to take into account both small frame shifts and large isotopic shifts in the mass spectrum data in order to identify and re-assign the correct monoisotopic peak to the peptide. To test Monocle's effectiveness, two well-established peptide search algorithms, Comet and MaxQuant, were loaded with raw and corrected (by Monocle) mass spectrometry data, in order to compare the total peptide spectral matches (PSMs), the total unique peptides, and total proteins found by both algorithms. Here we show that Monocle correction of monoisotopic peaks increases the amount of peptides identified and the quality of peptide matches. The number of total peptides, unique sequences, and total proteins identified by Comet benefited from Monocle correction. No comparison can be made for MaxQuant because the software did not accept the file format of the data with correction. The accuracy of peptide matches also increased, indicated by a higher proportion of peptide matches sharing the same scan number and peptide sequence. These results demonstrate that Monocle is beneficial to bottom-up proteomics by increasing the number of peptides found, thereby increasing confidence of identifying the right protein. Monocle also seems to help bridge the gap between Comet and MaxQuant by helping Comet

match each MS/MS scan to the same peptide sequence as MaxQuant did. These results should translate to the numerous other search algorithms that proteomics have developed.

Introduction

Mass spectrometry works by breaking down molecules in a sample into isotopes, which can then be measured by its mass per charge, often by the speed of flight. A detector keeps track of the time the isotope takes to reach it: isotopes with higher mass achieve slower speeds, and therefore take more time to reach the end. Amino acids are composed of several key elements like carbon, hydrogen, oxygen and nitrogen. Each element and its isotopes have a unique mass, so the identification of a monoisotopic peak can really narrow down the possible peptides. By using Monocle to assign the correct monoisotopic peak, peptide matches should be more accurate. At the same time, changing the monoisotopic peak can also increase the number of peptides found associated with a certain protein, thereby increasing the confidence of protein identification. Not only does Monocle assign a different peak (large isotopic shift), it can also adjust for small mass errors to improve the scores of PSMs.

In proteomics, proteins of interest are commonly digested by the protease trypsin, into fragments that can then be “weighed” by mass spectrometry. After the production of mass spectrum plots which indicate the elemental (or isotopic) signature of the sample, searching for peptides that match the predicted mass is only part of the investigation. There is a method called target/decoy searching, which involves having decoy sequences in the database that can then be used to correctly identify wrong matches. This also allows us to filter through all total peptide matches for peptide matches that don’t fall under this “decoy” group. Decoys are considered as false matches, which brings in a term called false discovery rate (FDR). FDR refers to the ratio

of false peptide spectral matches (PSMs) and the total number of PSMs above a certain threshold, such as 1% FDR. To control search quality, peptide matches can be sorted by their scores and filtered based on a certain threshold. Thus, filtering plays a role in refining the accuracy of search outputs.

The bulk of this investigation is not biological in terms of experimental design, although applications of mass spectrometry can certainly be useful in studies such as colorectal cancer - more on that in future experiments. Instead, this paper investigates and reports on the total number of peptides, total number of unique peptides, and total numbers of proteins found by each search algorithm. Comparisons are drawn between data that has monoisotopic peak correction by Monocle and raw data. At the same time, search results can be compared across search algorithms, with a focus on the percentage of shared peptide matches - these occur when both algorithms find the same peptide sequence for a given scan number. Each file is filled with thousands of scans worth of data, so both Comet and MaxQuant keep track of what peptide is associated with each scan. The scan number and the associated sequence can then be compared between the two algorithms, allowing us to sort out Comet-unique searches, MaxQuant-unique searches, and mutual searches. Monocle plays a possible role in bridging the gap between search algorithms by providing more accurate monoisotopic masses to use in peptide identification.

Materials and Methods

This project uses Comet and MaxQuant search algorithms to compare results with and without Monocle correction of monoisotopic peaks. Raw data/mzxml files are downloaded from <https://www.ebi.ac.uk/pride/archive/projects/PXD019311>, in which there are two groups of data: label-free and TMT. A human proteome dataset is taken from Uniprot, and is the reviewed

version instead of all total proteins: <https://www.uniprot.org/proteomes/UP000005640>. A yeast proteome dataset can also be found at: <https://www.uniprot.org/proteomes/UP000002311>, but one was provided by the University of Washington server. Comet searches are done with a proteomics-dedicated server at University of Washington while MaxQuant searches are done locally with a downloaded program from

http://coxdocs.org/doku.php?id=maxquant:common:download_and_installation.

Comet searches begin by uploading mzxml/raw files onto the UW server database, creating a parameter file using a reviewed fasta file of the human proteome/yeast proteome to search against, another parameter file for linear discriminant analysis, and then finally running it using those parameters. For yeast searches, the procedure is repeated using a yeast proteome file. Afterwards, the search data is filtered through to a target FDR(false discovery rate) of 2% to identify peptides, giving us data on the total peptides, unique peptides, and the total number of proteins. These results will then be put through Python to create visualizations comparing searches using Monocle-corrected data and raw data.

MaxQuant requires loading raw data files from the aforementioned site and a reviewed fasta file of the human/yeast proteome. The search should yield similar results as Comet and will be used in Python to create visuals, as well. As MaxQuant refuses to cooperate with mzxml files created from Monocle, this study will compare Comet searches with/without Monocle correction to a generalized MaxQuant search with no Monocle correction.

Results and Discussion

For HeLa (Henrietta Lacks) label-free runs with differing fragmentation energy schemes (cid35, hcd28, hcd35), Monocle correction consistently increases the total number of peptides found through Comet (Figure 1). The same result holds true for the total number of proteins found through Comet (Figure 2). Lastly, Monocle correction increases the number of unique peptides found through Comet (Figure 3). Each fragmentation energy scheme had duplicate runs, so bars indicate the average of all runs, while the black error bars indicate two standard deviations away from the mean. These three figures all indicate the usefulness of Monocle: by increasing the number of peptides found, the pool of possible unique peptides and proteins also grows, thereby allowing for more accurate identification of peptides/proteins.

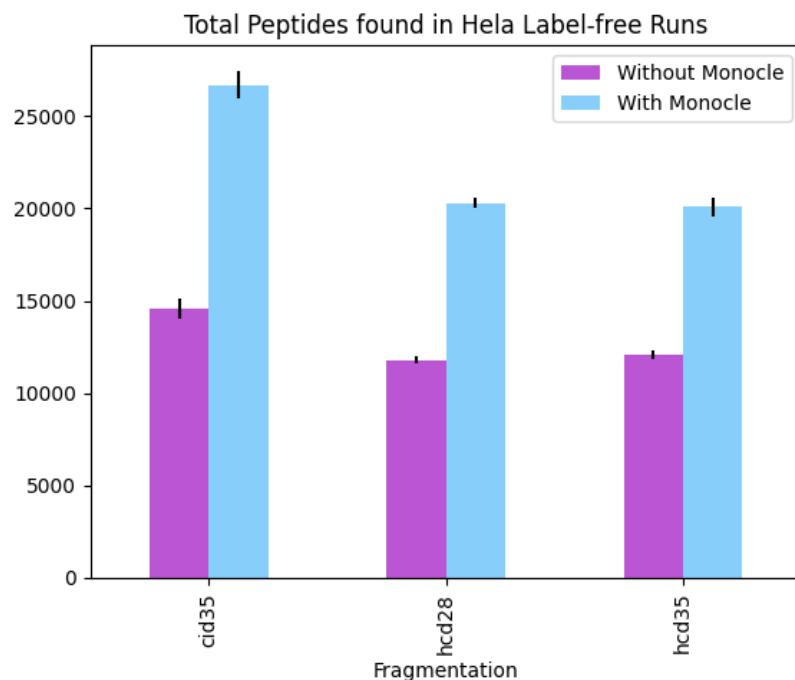


Figure 1: Across the board, Monocle increases the total peptides found by a large amount when using the Comet search algorithm.

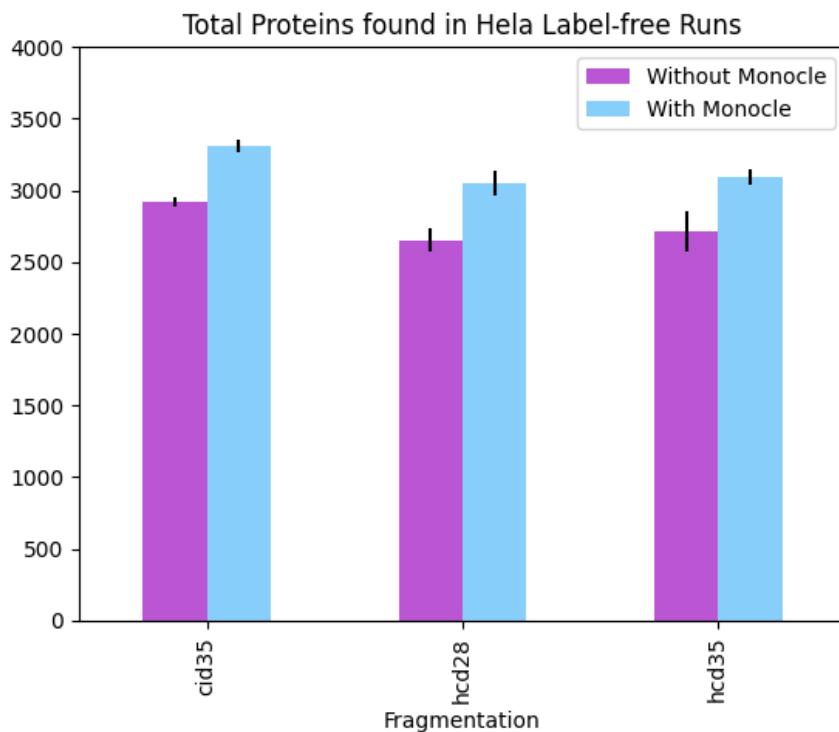


Figure 2: Comet search results using raw and Monocle-corrected data. More proteins were identified using Monocle correction.

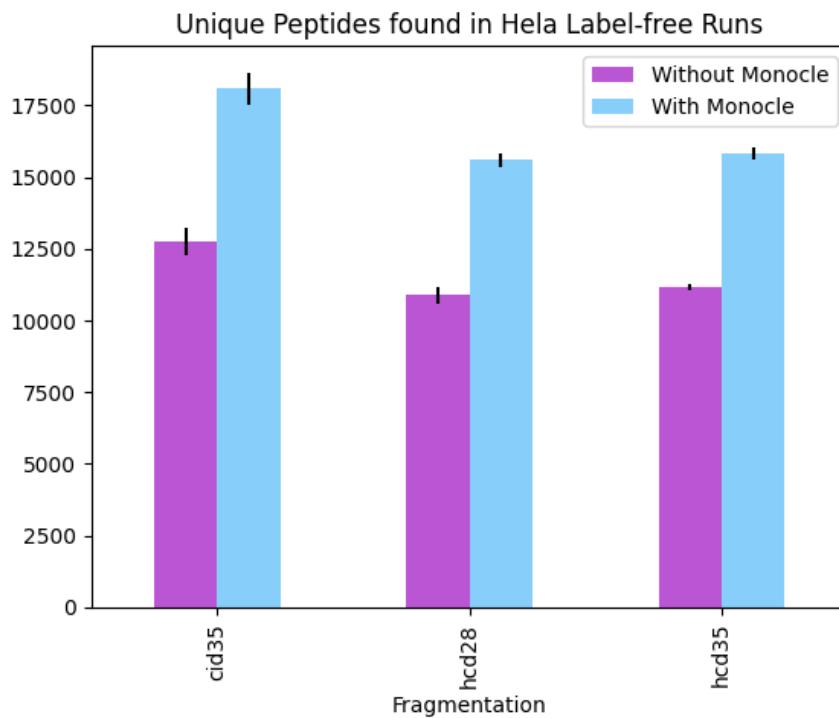


Figure 3: Comet search results using raw and Monocle-corrected data. Monocle correction yields a greater number of unique peptides found through Comet.

Label-free yeast runs have a similar result as label-free hela runs. Monocle correction consistently increased the number of peptides found (Figure 4), number of proteins identified (Figure 5), and the number of unique peptides found (Figure 6). As is the case with hela runs, bars represent the average of duplicate runs, with error bars representing two standard deviations away from the mean. For Comet searches using Monocle data, both runs using hcd28 fragmentation energy schemes resulted in the same number of peptides found, so there was no standard deviation to report (Figure 4).

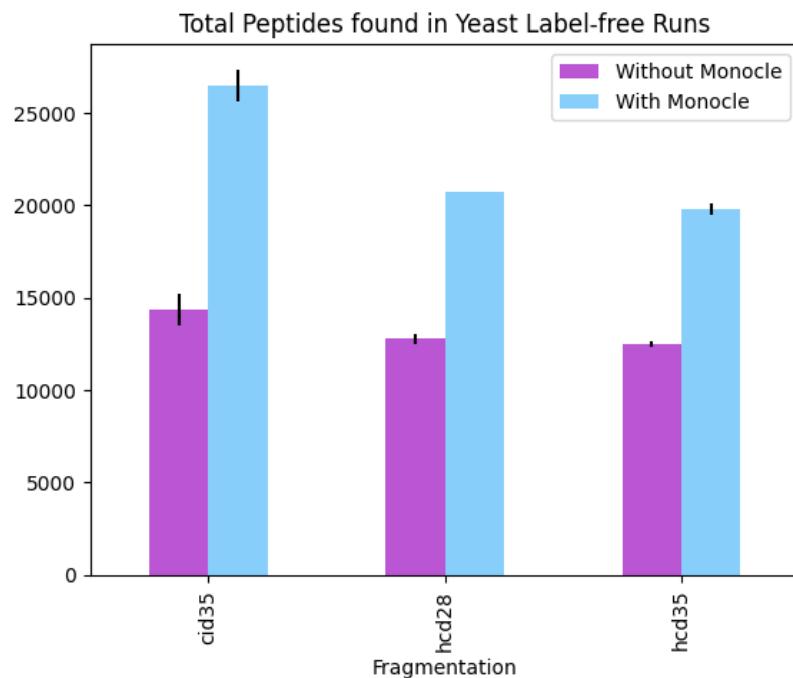


Figure 4: Comet search algorithm finds more peptide matches when using Monocle-corrected data compared to raw data.

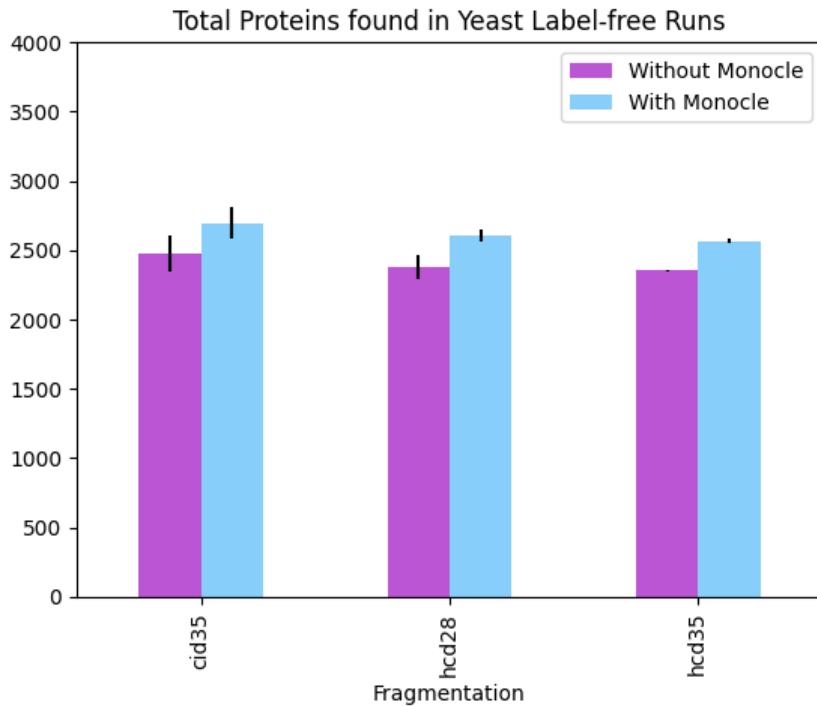


Figure 5: Comet search yields more proteins when using Monocle corrected data

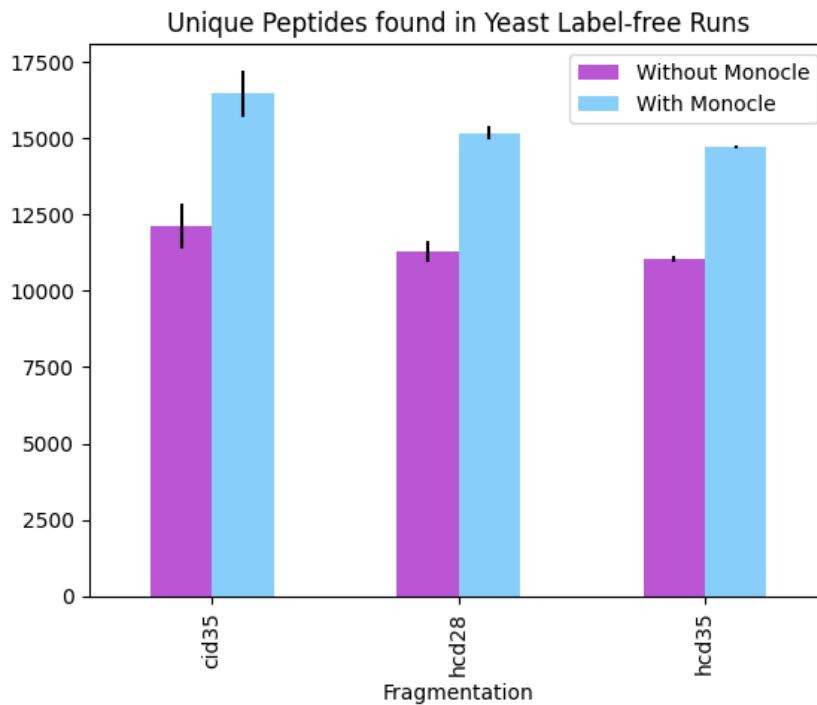


Figure 6: An increased number of unique peptides found by Comet algorithm when using Monocle corrected data.

This version of MaxQuant had difficulty with reading mzxml files (corrected data), so comparisons between raw and Monocle corrected data for the number of total and unique peptides found by MaxQuant cannot be drawn. Regardless, MaxQuant is finding over 16,000 total peptides on average for hela runs (Figure 7), which is similar to the amount of peptides that Comet finds (Figure 1). The average number of unique peptides from MaxQuant for hela runs (Figure 8) are also similar to the number of unique peptides found by Comet using raw data (Figure 3). A very similar story can be said about yeast runs (Figures 9 and 10). Even without the data to support it, it is highly probable that Monocle correction would allow MaxQuant to identify more total and unique peptides.

Average of total peptides found by MaxQuant in duplicate HeLa label-free runs

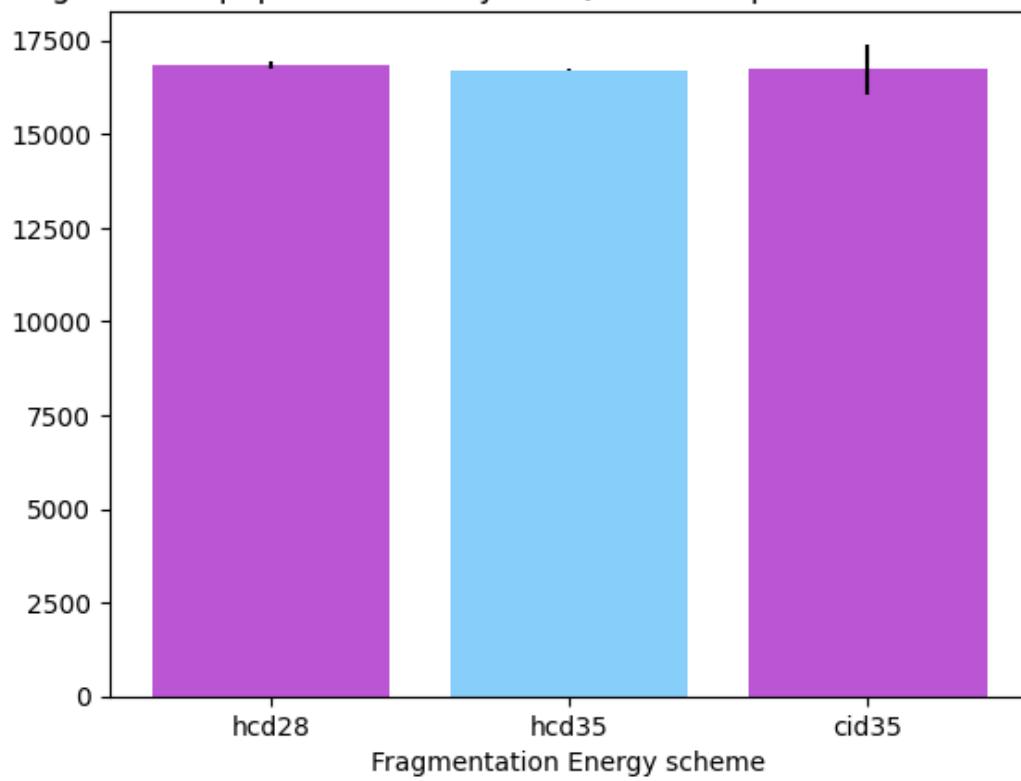


Figure 7: Total peptides found (average of duplicate HeLa runs) with error bars two standard deviations away from the mean. Uncorrected data used in MaxQuant search.

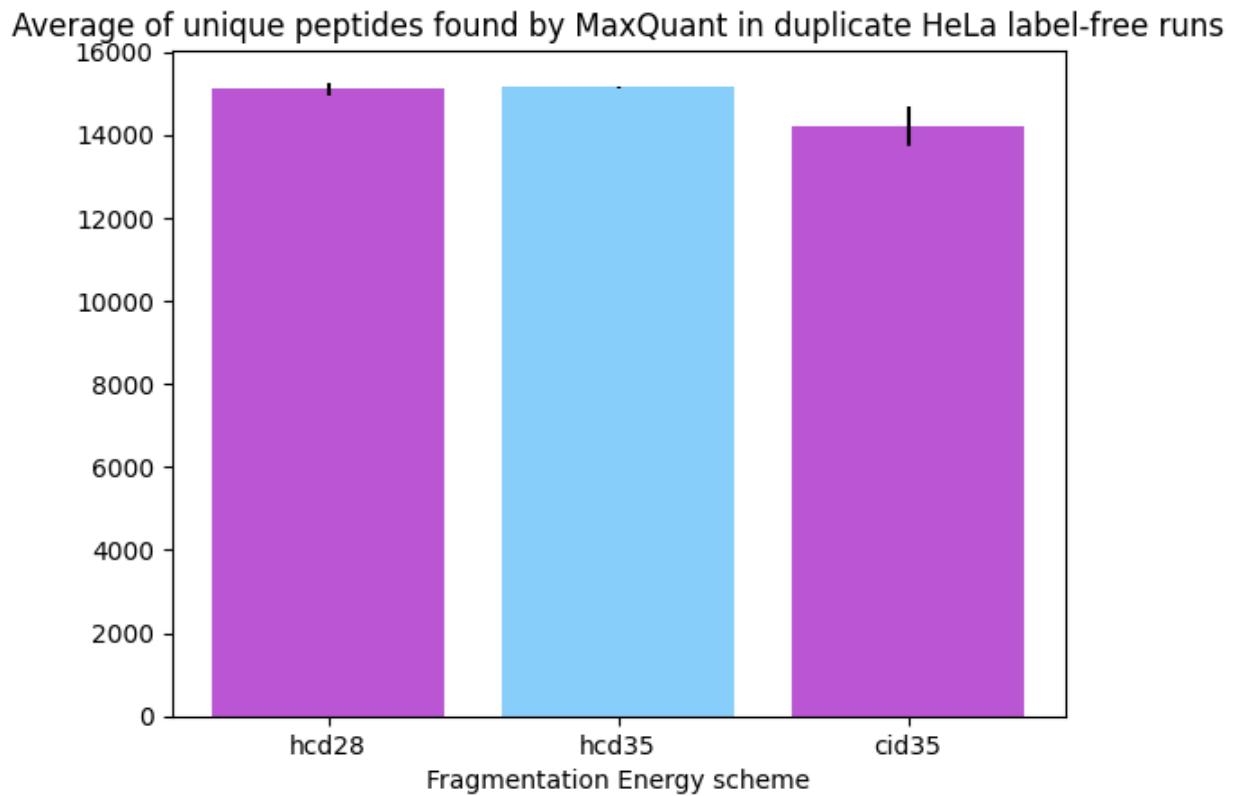


Figure 8: Total unique peptides found (average of duplicate HeLa runs) with error bars two standard deviations away from the mean. Uncorrected data used in MaxQuant search.

Average of total peptides found by MaxQuant in duplicate yeast label-free runs

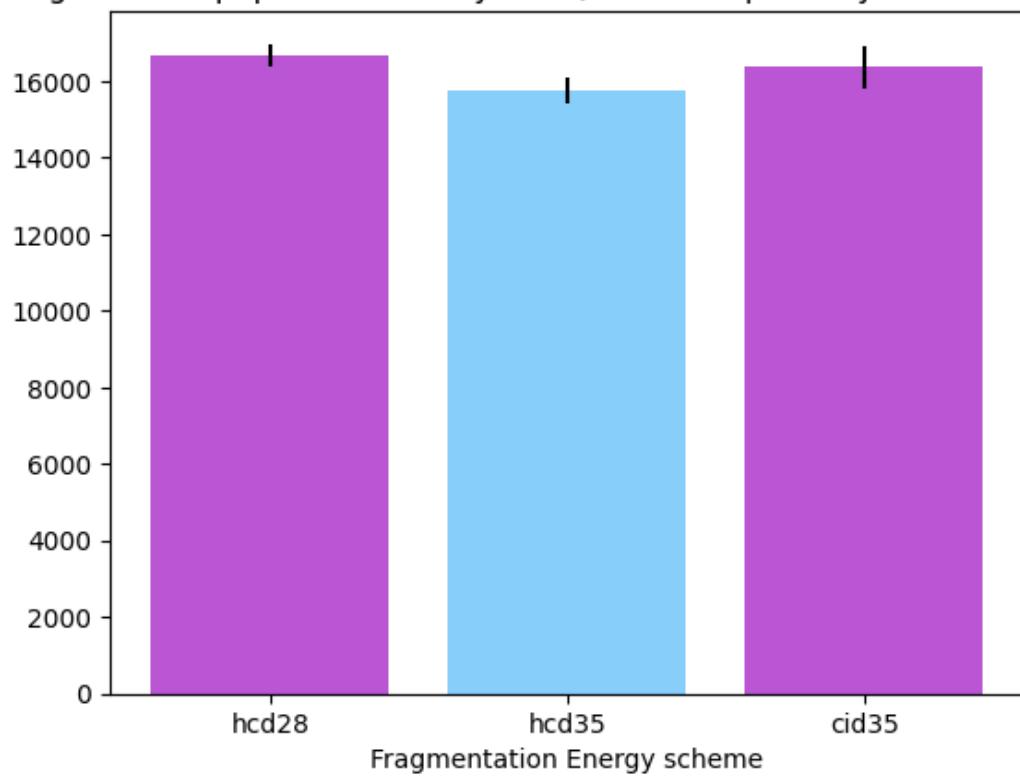


Figure 9: Total peptides found (average of duplicate yeast runs) with error bars two standard deviations away from the mean. Uncorrected data used in MaxQuant search.

Average of unique peptides found by MaxQuant in duplicate yeast label-free runs

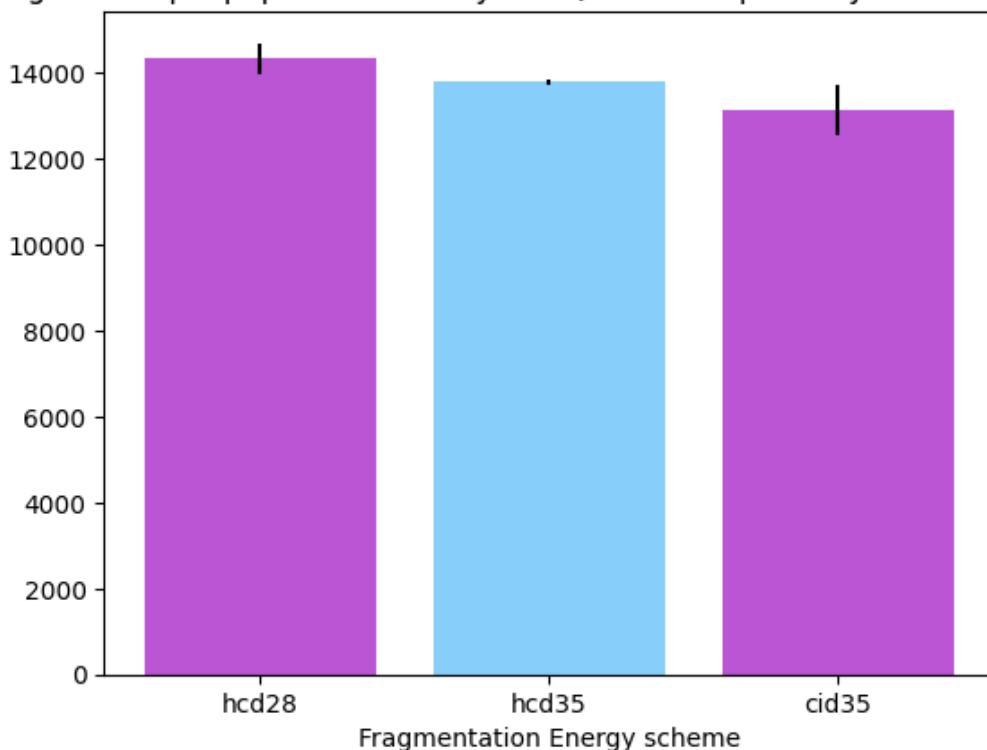
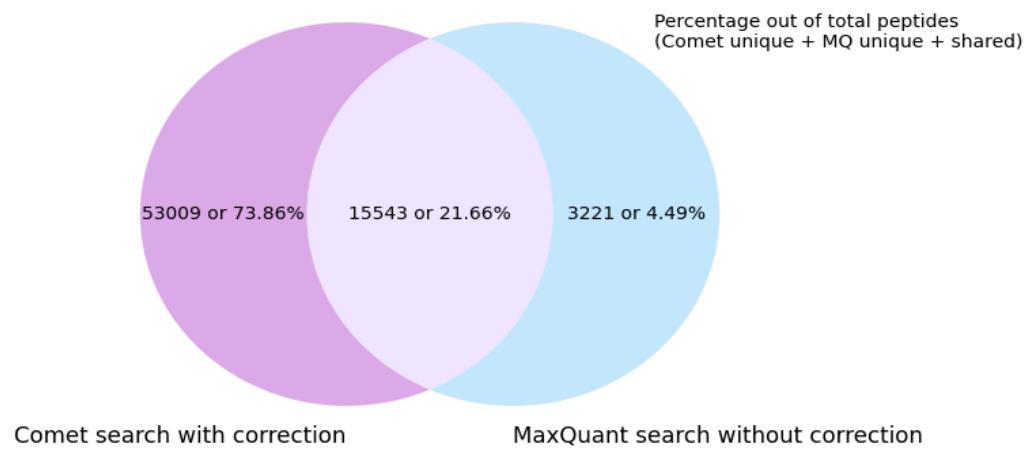


Figure 10: Total unique peptides found (average of duplicate yeast runs) with error bars two standard deviations away from the mean. Uncorrected data used in MaxQuant search.

The next comparisons to be made are between Comet and Maxquant. Peptide sequences from both search algorithms are sorted into three groups: unique to Comet, unique to Maxquant, and shared between both. In addition, Comet searches were put through a protein filter of 1% false discovery rate in order to accurately identify between true and false peptide spectral matches (PSMs). At the same time, searches from MaxQuant can also be filtered to the same FDR of 1 percent. As a result, there are unfiltered data and filtered data to compare.

Figures 10-13 provide Venn diagrams of g06608 hela label-free with cid35 fragmentation energy scheme. Filtering the peptide matches to 1% FDR in both search algorithms has a significant impact on the percentage of shared PSM matches between Comet and MaxQuant, especially for data that has been corrected by Monocle. While the total number of PSMs drops after filtering, the percentage of matches increases from 21.66 percent to 45.97 percent. It becomes unclear whether or not Monocle correction is responsible for this change, because Figure 8 finds a similar result, with the percentage of shared PSMs jumping from 14.67 percent to 48.89 percent for data that is uncorrected by Monocle. Figure 9 tells us that Monocle correction doesn't have a big impact on the percentage of shared PSMs between Comet and MaxQuant for filtered data, as 48.89 percent and 45.97 percent are rather close to each other. On the other hand, unfiltered data seems to benefit slightly from Monocle correction, with the percentage of shared PSMs jumping from 14.67 percent to 21.66 percent. Therefore, we cannot conclude that Monocle correction is significantly increasing the percentage of shared PSMs between Comet and MaxQuant for this run.

Unfiltered Peptide Spectral Matches for g06606 label-free hela run - cid35



Filtered Peptide Spectral Matches for g06606 label-free hela run - cid35

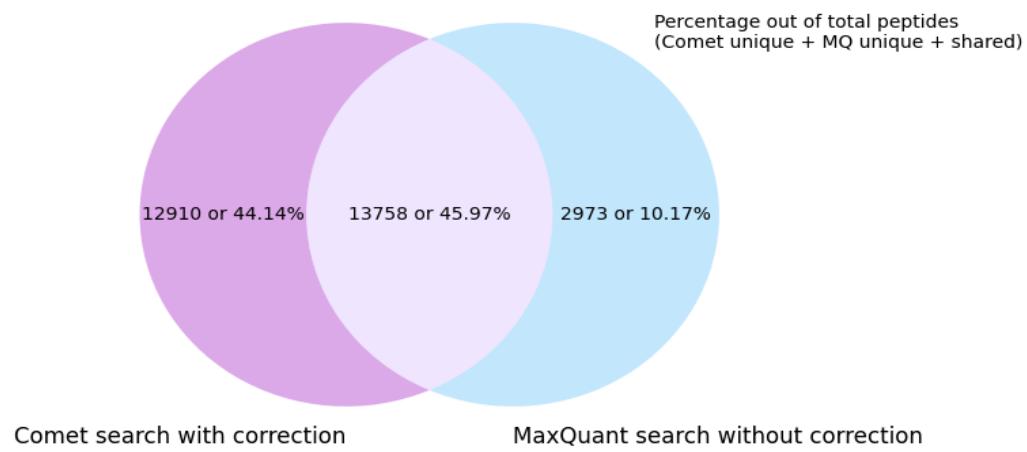
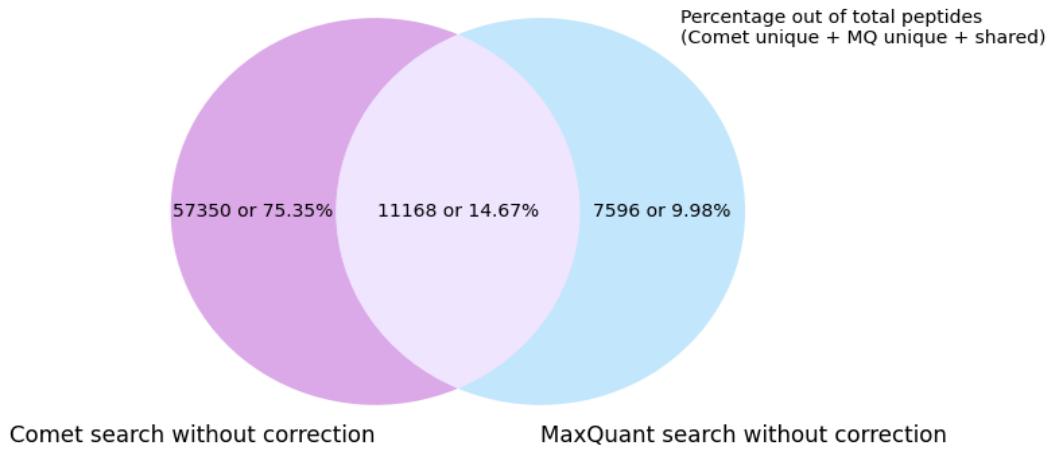


Figure 10: Unfiltered vs filtered PSMs from Comet and MQ using data that has been corrected by Monocle for g06606 hela label-free run with cid35 fragmentation energy scheme.

Unfiltered Peptide Spectral Matches for g06606 label-free hela run - cid35



Filtered Peptide Spectral Matches for g06606 label-free hela run - cid35

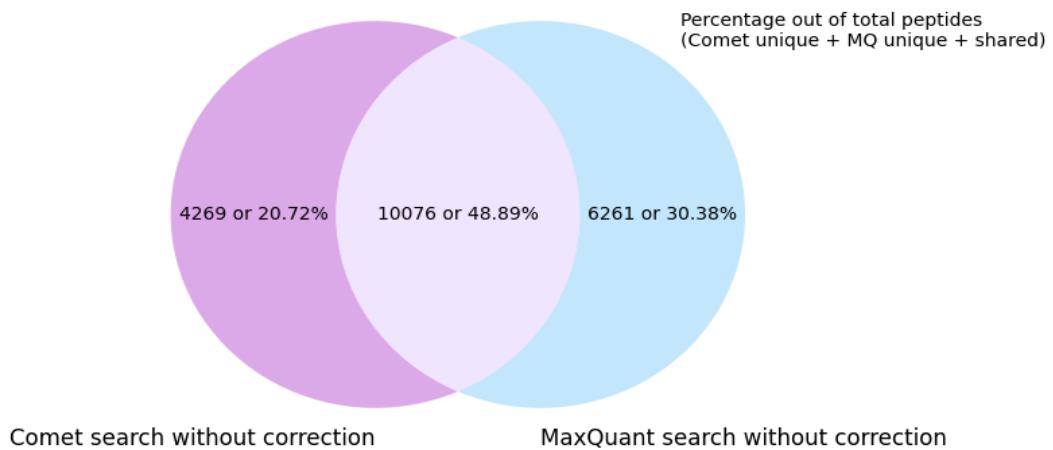
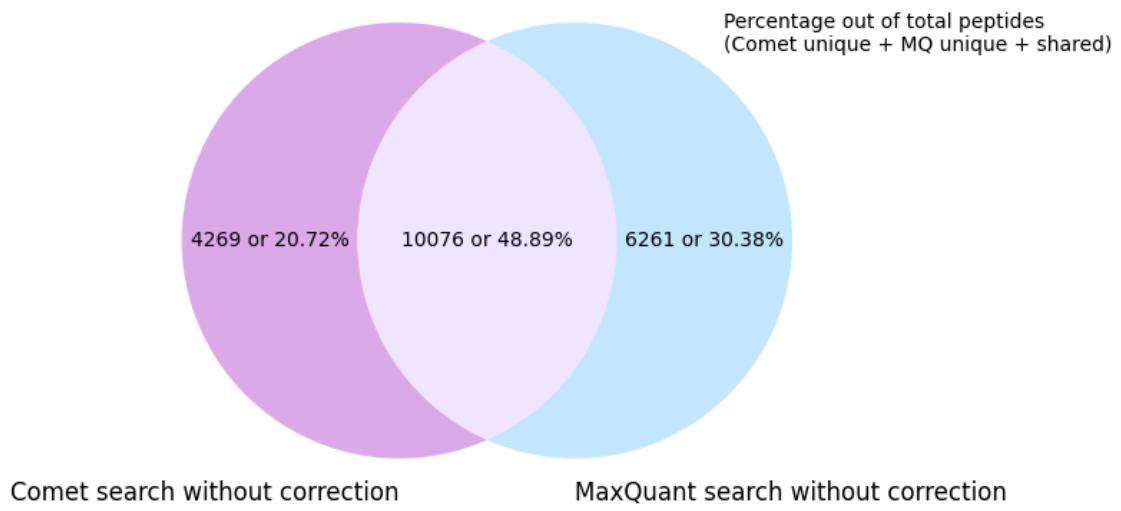


Figure 11: Unfiltered vs filtered PSMs from Comet and MQ using raw data for g06606 hela label-free run with cid35 fragmentation energy scheme.

Filtered Peptide Spectral Matches for g06606 label-free hela run - cid35



Filtered Peptide Spectral Matches for g06606 label-free hela run - cid35

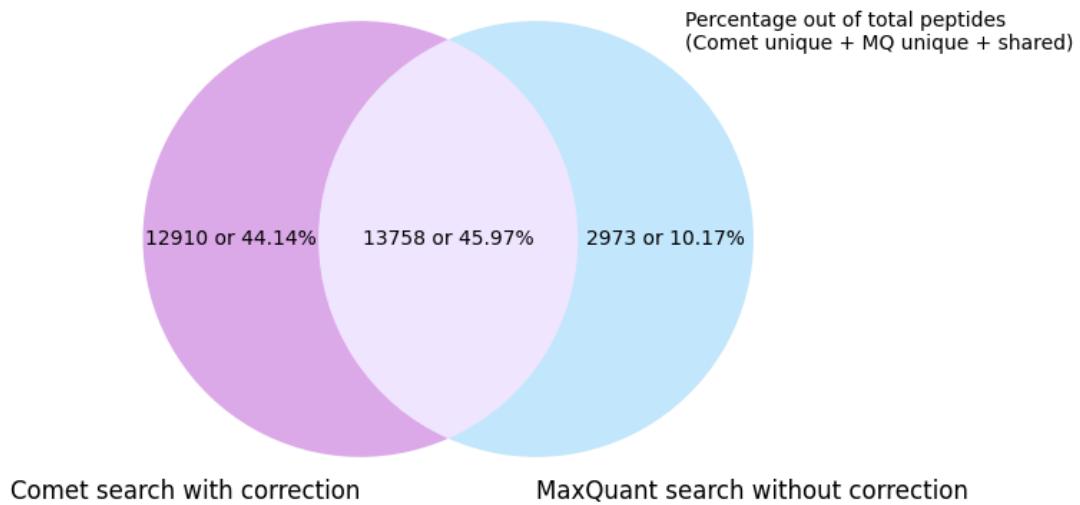
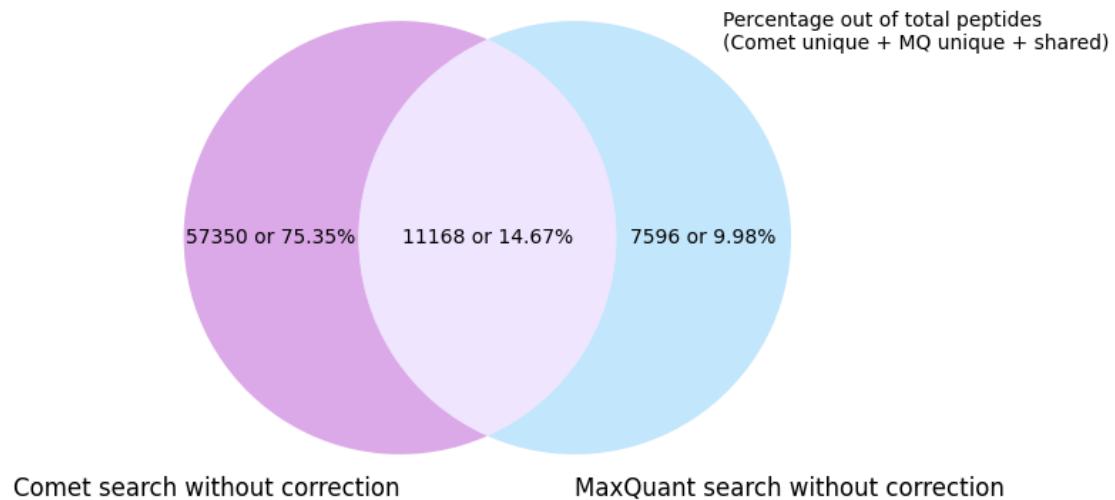


Figure 12: Raw vs. Correction comparison using filtered PSMs from Comet and MQ for g06606 hela label-free run with cid35 fragmentation energy scheme.

Unfiltered Peptide Spectral Matches for g06606 label-free hela run - cid35



Unfiltered Peptide Spectral Matches for g06606 label-free hela run - cid35

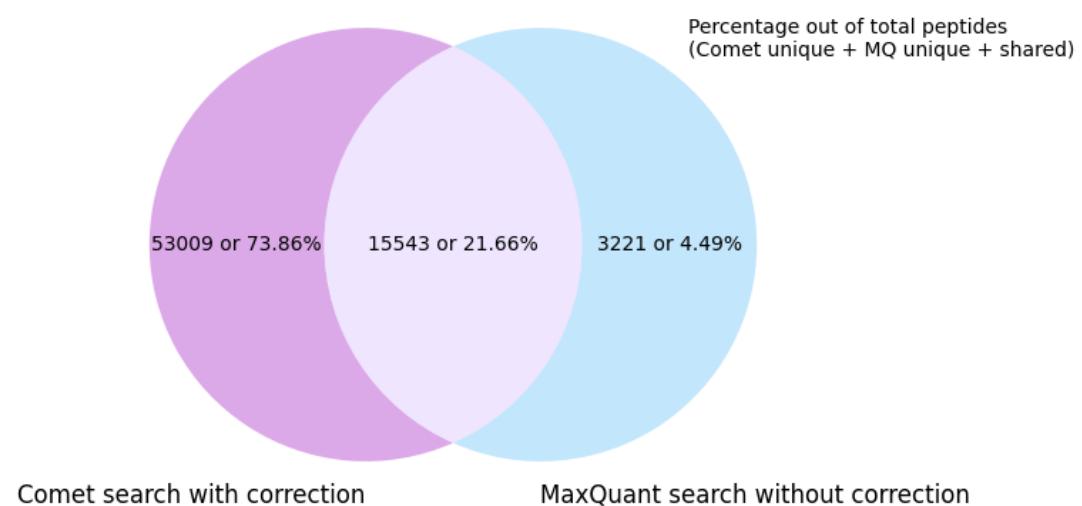


Figure 13: Raw vs. Correction comparison using unfiltered PSMs from Comet and MQ for g06606 hela label-free run with cid35 fragmentation energy scheme.

Figures 14-17 are for the run labelled as g06607 hela label-free with cid35 fragmentation energy scheme. Filtering the PSMs to 1% appears to increase the percentage of shared peptides between Comet and Maxquant for both raw and uncorrected data (Figures 14 and 15). Figures 16 and 17 tell a similar story as the previous run g06607. Again, it can't be concluded that Monocle correction is significantly affecting the shared PSM percentage between MQ and Comet for this run. As for the last cid35 run, g06608, a similar result can be said (Appendix Figures 1-4).

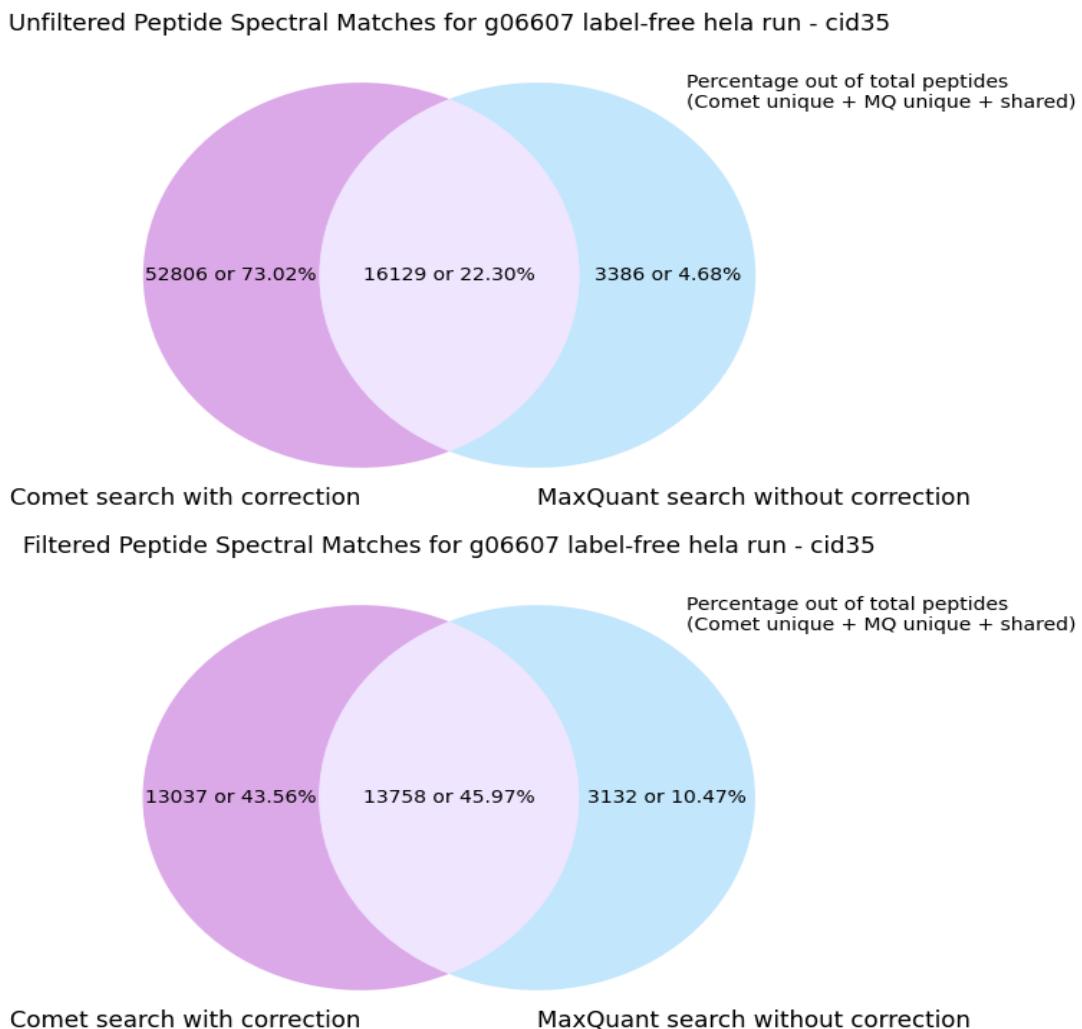
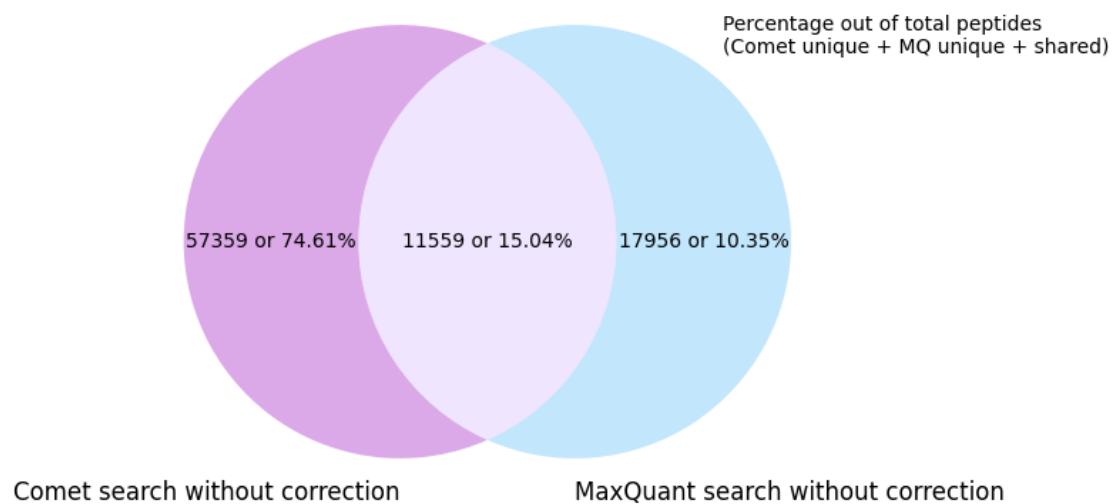


Figure 14: Unfiltered vs filtered PSMs from Comet and MQ using data that has been corrected by Monocle for g06607 hela label-free run with cid35 fragmentation energy scheme.

Unfiltered Peptide Spectral Matches for g06607 label-free hela run - cid35



Filtered Peptide Spectral Matches for g06607 label-free hela run - cid35

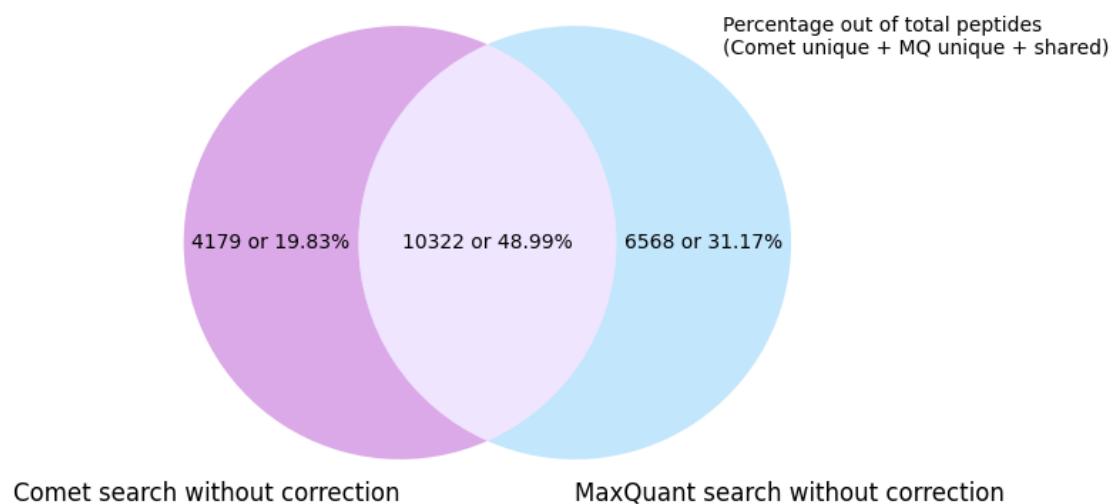
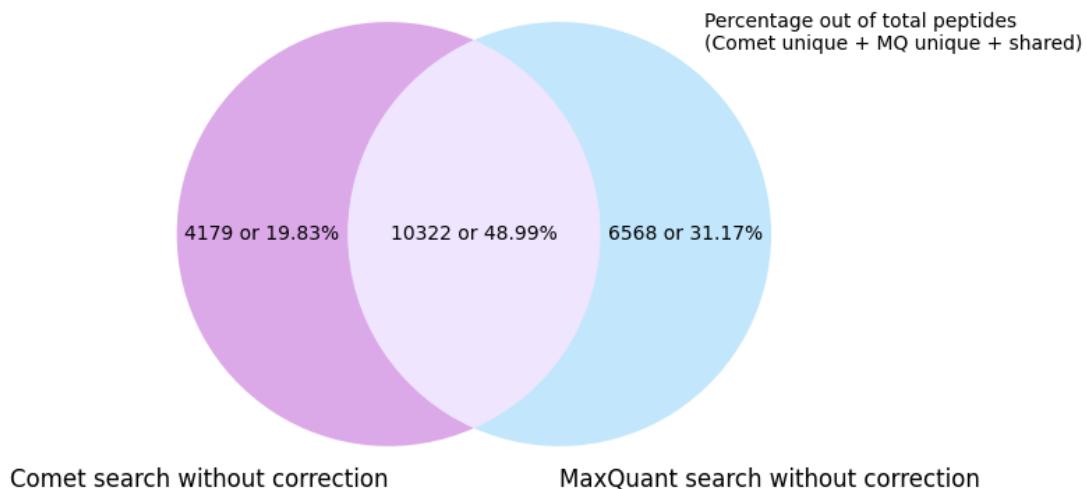


Figure 15: Unfiltered vs filtered PSMs from Comet and MQ using raw data for g06607 hela label-free run with cid35 fragmentation energy scheme.

Filtered Peptide Spectral Matches for g06607 label-free hela run - cid35



Filtered Peptide Spectral Matches for g06607 label-free hela run - cid35

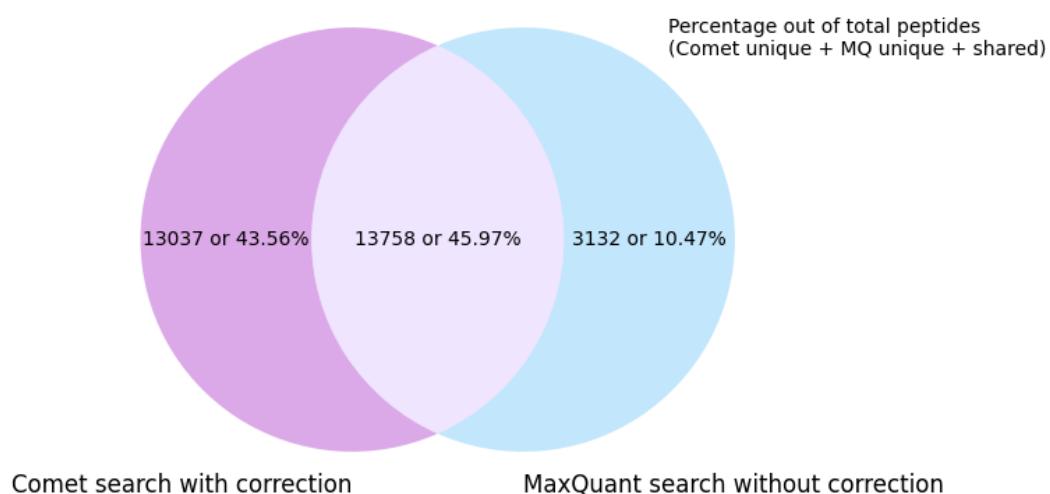
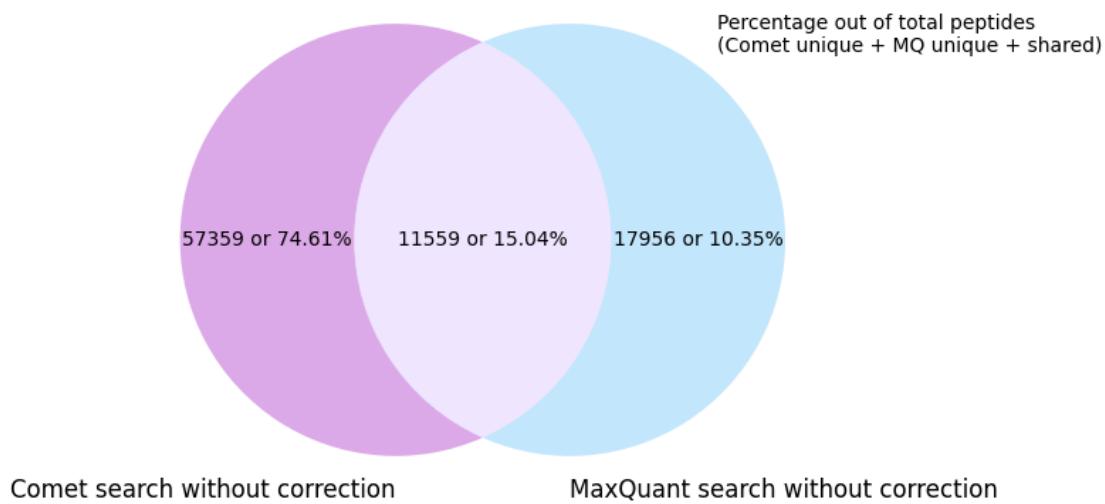


Figure 16: Raw vs. Correction comparison using filtered PSMs from Comet and MQ for g06607 hela label-free run with cid35 fragmentation energy scheme.

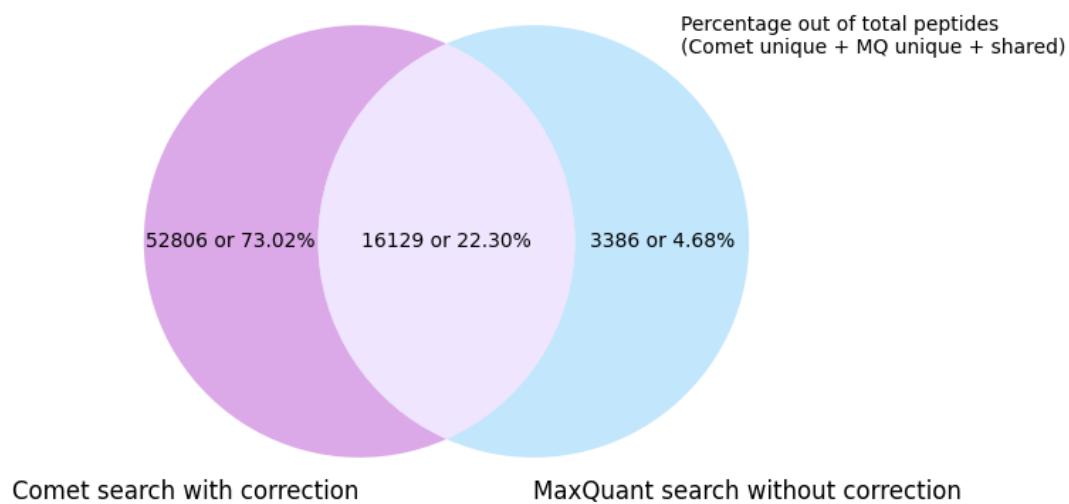
Unfiltered Peptide Spectral Matches for g06607 label-free hela run - cid35



Comet search without correction

MaxQuant search without correction

Unfiltered Peptide Spectral Matches for g06607 label-free hela run - cid35



Comet search with correction

MaxQuant search without correction

Figure 18: Raw vs. Correction comparison using unfiltered PSMs from Comet and MQ for g06606 hela label-free run with cid35 fragmentation energy scheme.

For HeLa runs using higher-energy collisional dissociation (g06629, g06630, g06637, g06638), Monocle correction along with filtering seems to positively impact the number and percentage of shared PSMs between Comet and MaxQuant (Appendix Figures 5-20).

Yeast runs with cid35 fragmentation energy schemes have similar results as the hela runs with cid35 fragmentation energy schemes (Appendix Figures 21-32). While the number of shared PSMs does increase, the overall percentage does not consistently increase with Monocle correction. Similarly to HeLa runs, shared peptide percentages in yeast runs using higher-energy collisional dissociation (g06627 -- Appendix Figures 33-36, g06628 -- Appendix Figures 37-40, g06634 -- Appendix Figures 41-44, g06635 -- Appendix Figures 45-48) appeared to benefit with Monocle correction. For both hela and yeast runs, it is difficult to conclude what factor is causing this improvement: do different fragmentation energy schemes allow for more accurate ion dissociation or is it the filtering process?

Higher-energy collisional dissociation (hcd) differs from collision-induced dissociation (cid) because peptide fragments that would normally be stable after cleavage by trypsin can be fragmented even further. Thus, there is a possibility that the total number of peptides identified increased from more peptide fragments. To investigate this, there is an output from Comet that could quantify the fraction of ions that match for each fragmentation energy scheme. It could potentially determine any differences that arise from using a different fragmentation energy scheme. On the other hand, MaxQuant also attempts to correct the monoisotopic peaks by estimating the mass, through very small ppm (parts per million) shifts in mass. As a result, it seems that MaxQuant and Monocle + Comet reach a similar conclusion for more PSMs.

Overall, Monocle proves to be beneficial for bottom-up proteomics. By increasing the number of peptides identified, not only can we be more confident in high-scoring peptide spectral matches, we can also be more confident in protein identification. Misassignment of monoisotopic peaks can have negative cascading effects on the accuracy of mass spectrometry

results. By correcting the mass used for the peptide investigation, any mass spectrum data would likely benefit from undergoing Monocle correction.

Future Investigations

Immediate changes: This investigation would benefit, quite obviously, from loading Monocle-corrected mass spectrum data into MaxQuant. The current version would crash whenever it tries to process an mzxml file. Therefore, it may be worth downloading an older version to see if this bug is present only in the current edition - it seems that mzxml file acceptance is a new feature. Furthermore, by loading corrected data into MaxQuant, it becomes possible to compare MaxQuant's own correction algorithm with Monocle's correction. MaxQuant performs two rounds of searching: the first round allows MaxQuant to adjust the mass by a very small frameshift (ppm). This results in a "corrected" version after the second round. By measuring the total peptides found and total unique peptides found with correction, it would be possible to determine whether or not Monocle's correction is superior to MaxQuant's own corrections. At the same time, comparisons can be drawn about the scoring that each algorithm applies to each scan. Do PSMs from Comet + Monocle have higher scores than PSMs from MaxQuant? This would affect the amount of mutual PSMs because the score determines what peptide sequence each scan number is associated with.

Short introduction: In regards to a more biological setting, proteomics and mass spectrometry can be adapted for investigations relating to colon cancer. Colon cancer is one of several leading causes of death for diagnosed cancer patients and continues to be prevalent due to a multitude of factors, such as being hereditary, a poor diet and lifestyle, and random mutations affecting the

suppression of tumor genes, etc. Despite large amounts of research and introduction of novel therapies, it remains difficult to treat and understand the complex systems underlying colon cancer. Most efforts are directed towards discovering cancer genes, often through genome-wide association studies (GWAS), which identify single-nucleotide polymorphisms (SNPs) associated with cancer characteristics. After identifying candidate genes, the focus turns towards placing the gene into a known pathway, such as cell proliferation, inflammation, immune response, etc.

Many genes “communicate” with one another through proteins, which act as transcription factors to regulate transcription. These proteins can deviate from their normal structure because mutations change their tertiary/quaternary structure or their levels can be over/under-expressed in the body. These differences in protein structure and level can be detected by mass spectrometry. By intercepting these cellular messages, identifying proteins as a potential biomarker could make it possible to diagnose colon cancer at its early stages. In addition, biomarkers can be identified for prognostic purposes or even monitor the progress of new treatments.

Another interesting factor for biomarker identification is the presence of proteins with post-translational modifications (PTMs). Some proteins aren’t mature directly after translation and must undergo some form of modification like phosphorylation and methylation in order to properly control cell signalling and metabolic pathways. PTMs could add yet another layer of complexity to understanding colorectal cancer. The identification of proteins that require phosphorylation and their targets can aid in the development of drugs with kinase inhibitors. PTMs can be identified through mass spectrometry as well, by accounting for any mass changes from the protein modification, which is somewhat similar to the correction that already occurs with Monocle.

Proposition: There have been multiple studies using mass spectrometry under the basis of identifying protein biomarkers in colorectal tumors, which are easy to obtain compared to other forms of cancer. Rather than going through the experimental procedures of obtaining tumor samples and digesting them with trypsin, it could be beneficial to take their original data and run it through Monocle. For example, Atak, et al., (2018) uses quantitative mass spectrometry to identify nine different proteins as potential biomarkers for colon adenocarcinomas. In their procedure, they also attempted to increase confidence in their protein IDs from Spectrum Mill by cross-checking in another database, the Trans-Proteomic Pipeline (TPP). I believe that Monocle correction can achieve a similar effect on their MS/MS data. If proteins that are identified with Monocle match up with the original data and TPP, then further investigations related to tracing these proteins during diagnosis can rest assured that they are not wasting their time on false biomarkers. On the other hand, if any proteins differ from the list produced by Spectrum Mill and TPP, then the opportunity arises to investigate the legitimacy of removing the protein from being included as a potential biomarker of colon adenocarcinoma.

Another study, Hung et al. (2009), used samples from Apc-mutant mice with tumors to obtain MS/MS spectra that were loaded into X!Comet (which seems to be the open-source Comet that was used for my primary investigation) in order to search against the proteome library. They followed a similar procedure of filtering the identified proteins through a false discovery rate in order to clean up any mismatches between scans and peptide sequences before retaining proteins that passed all filtering steps. Their investigation resulted in the identification of 51 proteins with elevated levels in plasma and were upregulated in cancerous tissue. Again, Monocle should provide more confidence in the results of their study, as long as the corrected data produces similar (or even better, the same!) peptide/protein identifications as their original data does.

Xue et al. (2009) was interested in investigating biomarkers related to colon cancer metastasis. Their team used two well-known colon cancer cell lines, SW480 and SW620, the first of which is derived from the primary site of colon cancer and the second is derived from a metastatic site in the lymph nodes of the same patient. Their data uses MS/MS spectra to search against the International Protein Index for the human proteome. Using Monocle to correct any misassigned monoisotopic peaks at this stage should prove beneficial, as it becomes possible to compare the results of the corrected search with their original findings.

More studies that can potentially be improved by Monocle implementation can be found in a review article by Martínez-Aguilar et al (2013), where they compile a large number of studies that apply quantitative mass spectrometry analysis to biomarker identification in colorectal cancer.

References:

Apurva Atak, et al., [Quantitative mass spectrometry analysis reveals a panel of nine proteins as diagnostic markers for colon adenocarcinomas](#), Oncotarget 13530–13544 (2018)

Hung, et al., [Comprehensive Proteome Analysis of an Apc Mouse Model Uncovers Proteins Associated with Intestinal Tumorigenesis](#), *Cancer Prevention Research* Vol. 2 Issue 3, DOI: 10.1158/1940-6207.CAPR-08-0153 (2009)

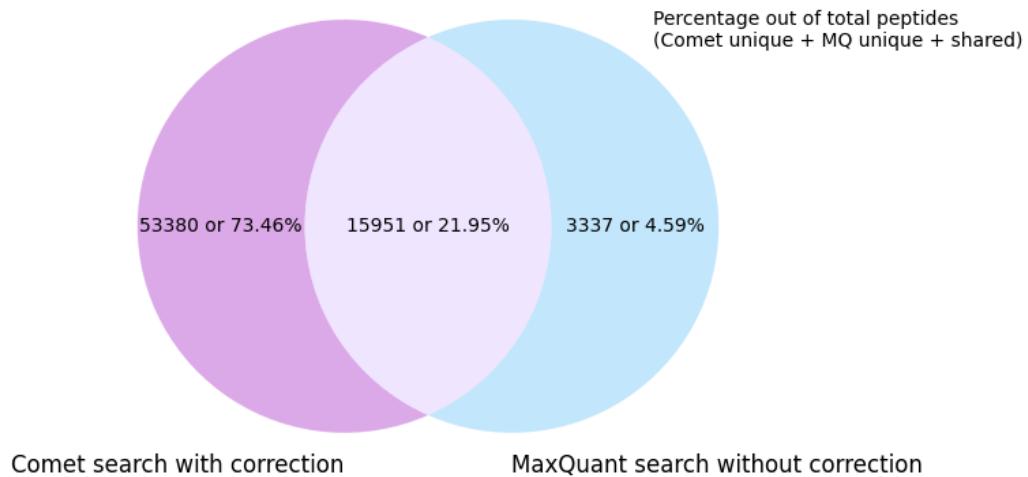
Xue, et al., [Identification of Serum Biomarkers for Colorectal Cancer Metastasis Using a Differential Secretome Approach](#), *Journal of Proteome Research* 2010, 9, 1, 545–555 (Published 2009)

Martinez-Aguilar, et al., [Quantitative mass spectrometry for colorectal cancer proteomics](#), *Proteomics Clinical Applications* Vol. 7 Issues 1-2 pg. 42-54 (2013)

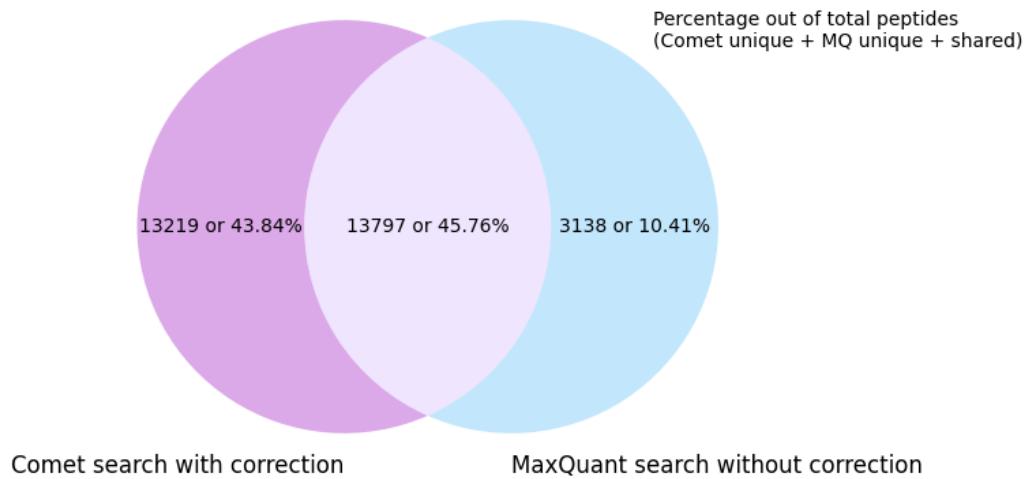
Appendix:

HELA FIGURES

Unfiltered Peptide Spectral Matches for g06608 label-free hela run - cid35



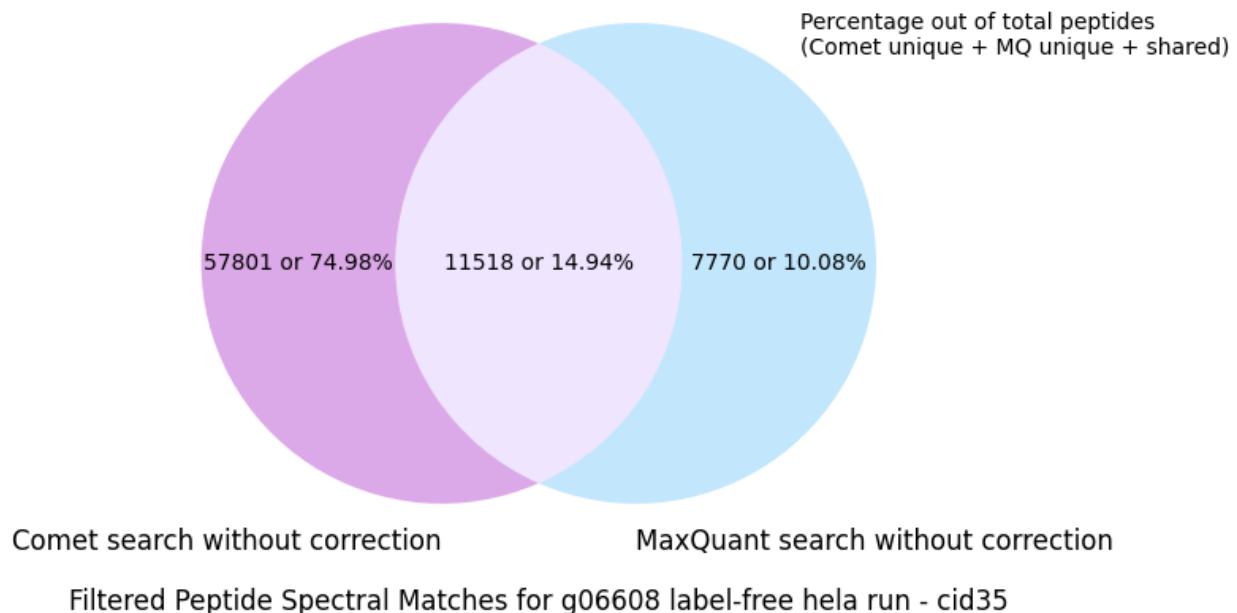
Filtered Peptide Spectral Matches for g06608 label-free hela run - cid35



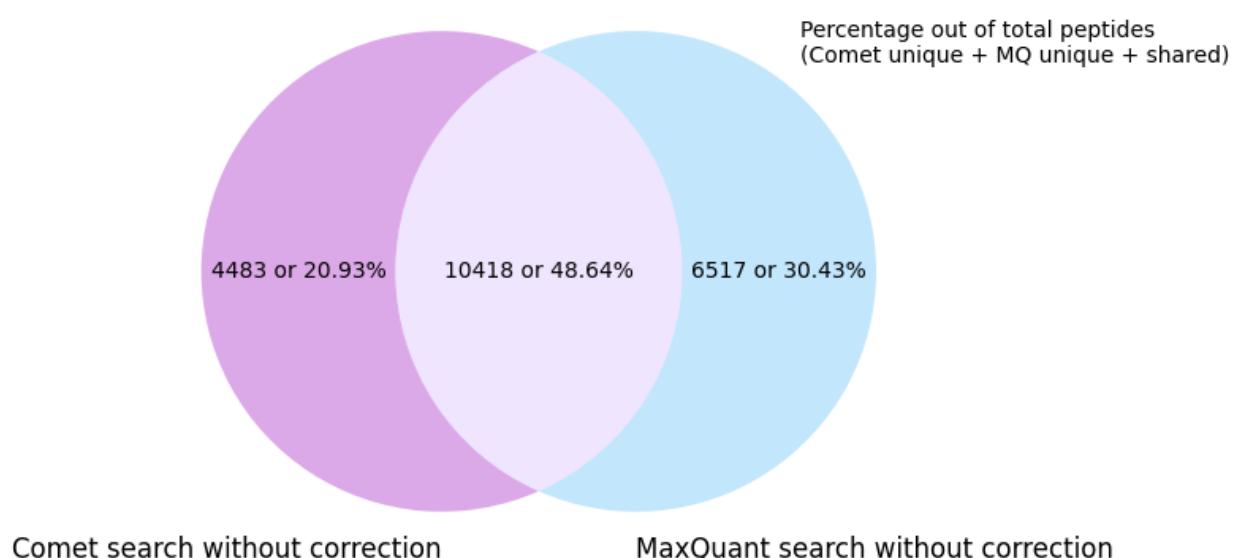
Appendix Figure 1: Unfiltered vs Filtered search results with Monocle correction for g06608.

Increased percentage of shared peptides with filtering.

Unfiltered Peptide Spectral Matches for g06608 label-free hela run - cid35



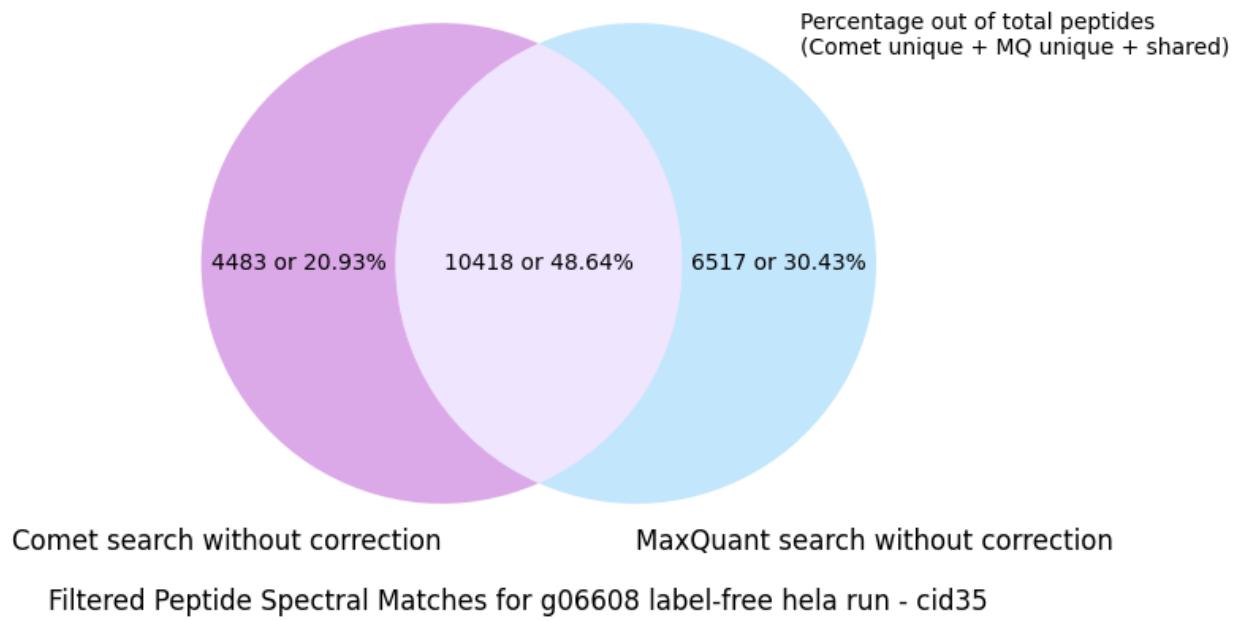
Filtered Peptide Spectral Matches for g06608 label-free hela run - cid35



Appendix Figure 2: Unfiltered vs Filtered search results without Monocle correction for g06608.

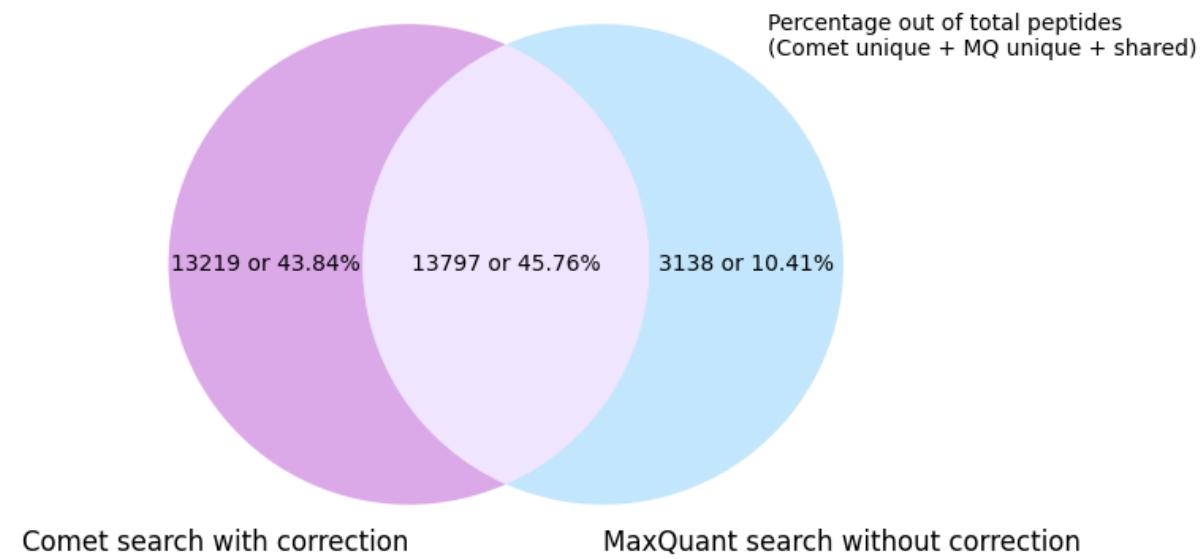
Increased percentage of shared peptides with filtering.

Filtered Peptide Spectral Matches for g06608 label-free hela run - cid35



Comet search without correction MaxQuant search without correction

Filtered Peptide Spectral Matches for g06608 label-free hela run - cid35



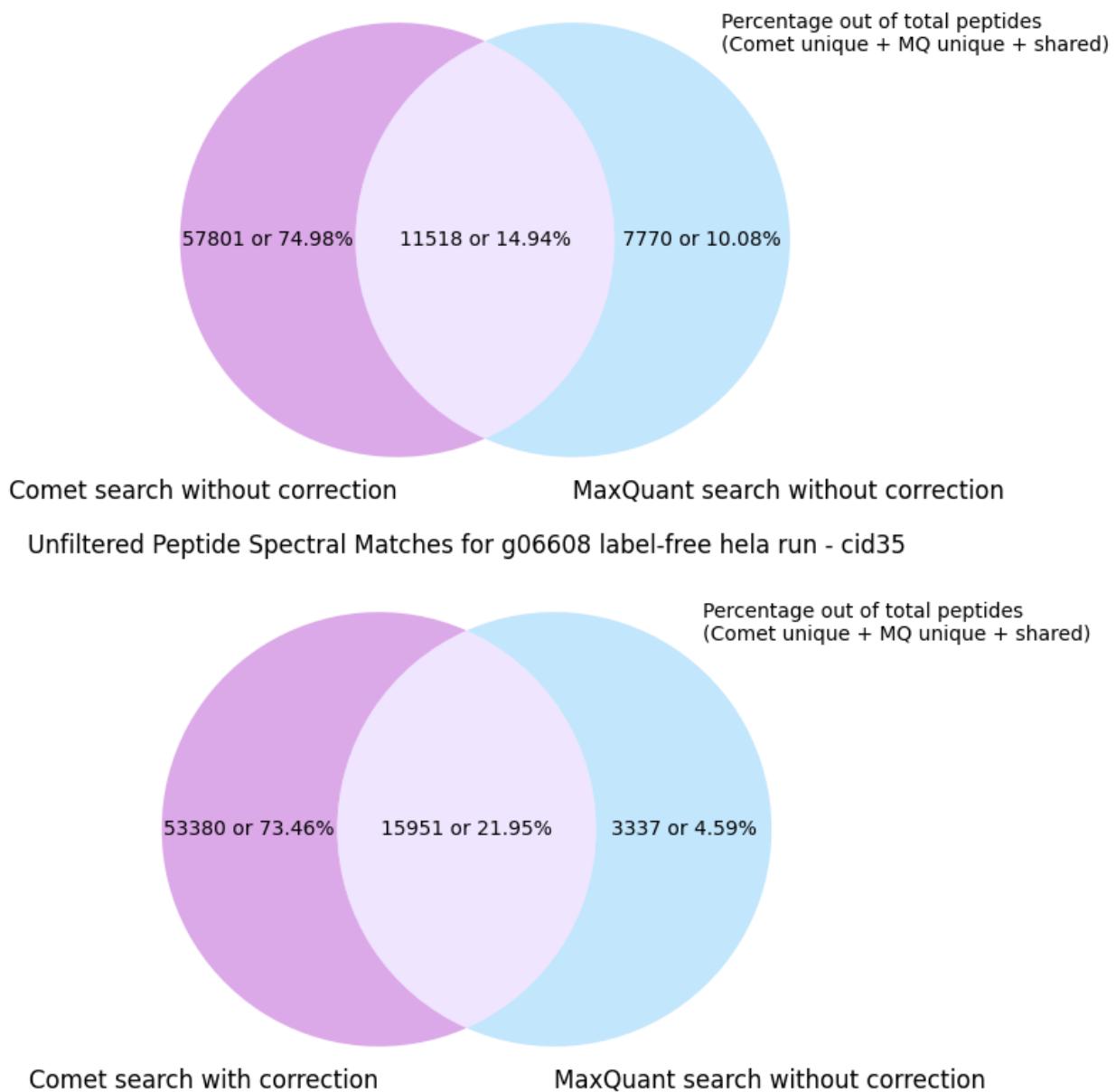
Comet search with correction

MaxQuant search without correction

Appendix Figure 3: Filtered search results with or without Monocle correction for g06608.

Increased number but not percentage of shared peptides with filtering.

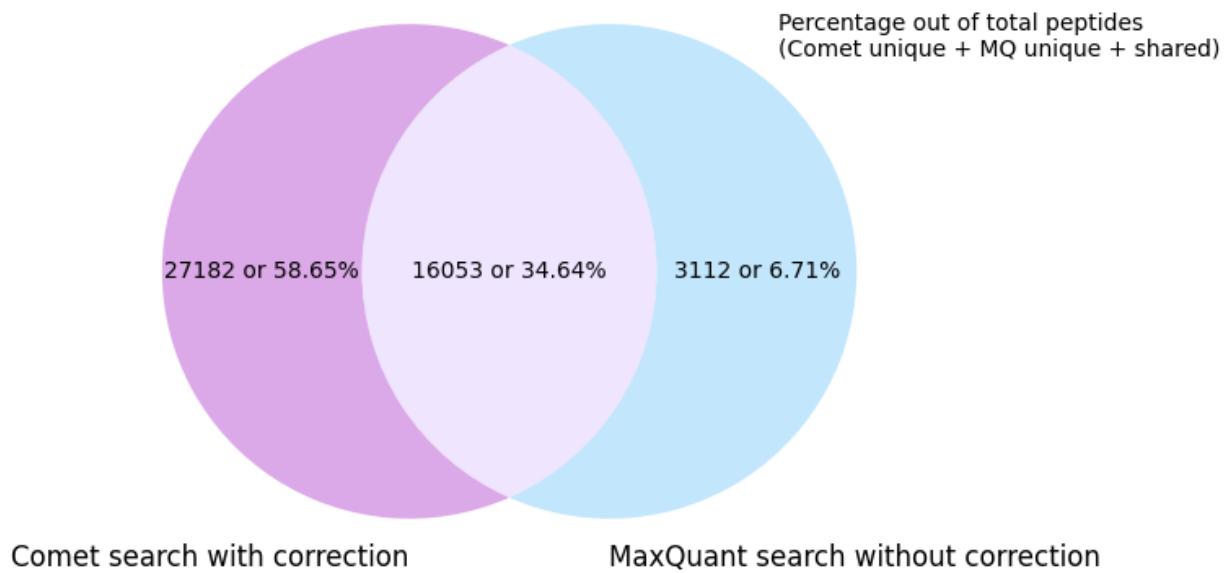
Unfiltered Peptide Spectral Matches for g06608 label-free hela run - cid35



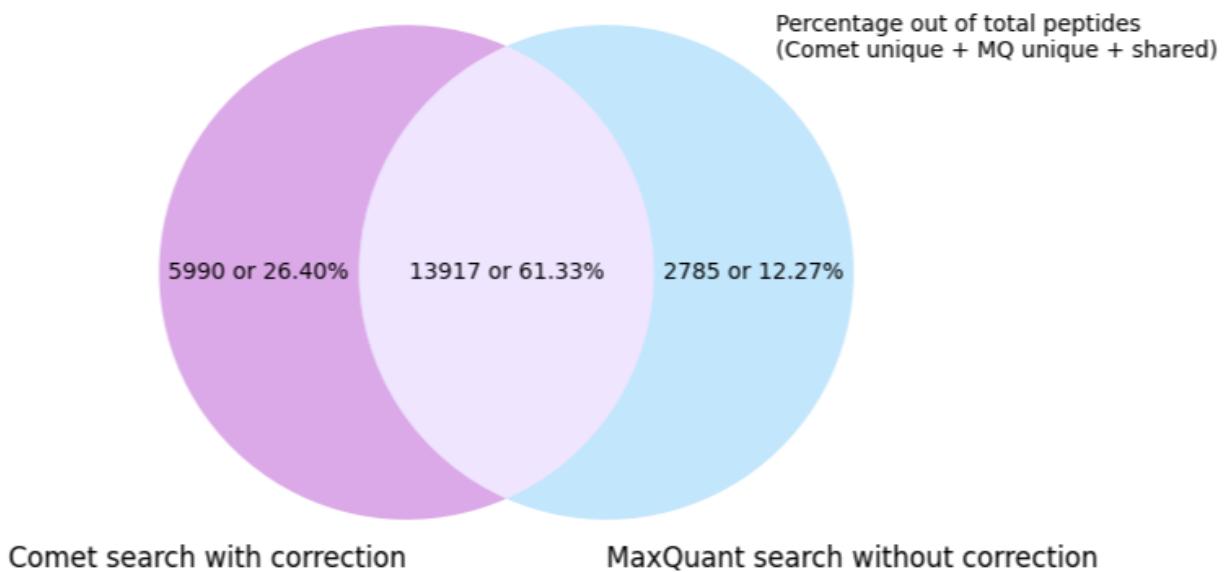
Appendix Figure 4: Unfiltered search results with or without Monocle correction for g06608.

Increased number and percentage of shared peptides with correction.

Unfiltered Peptide Spectral Matches for g06629 label-free hela run - hcd35



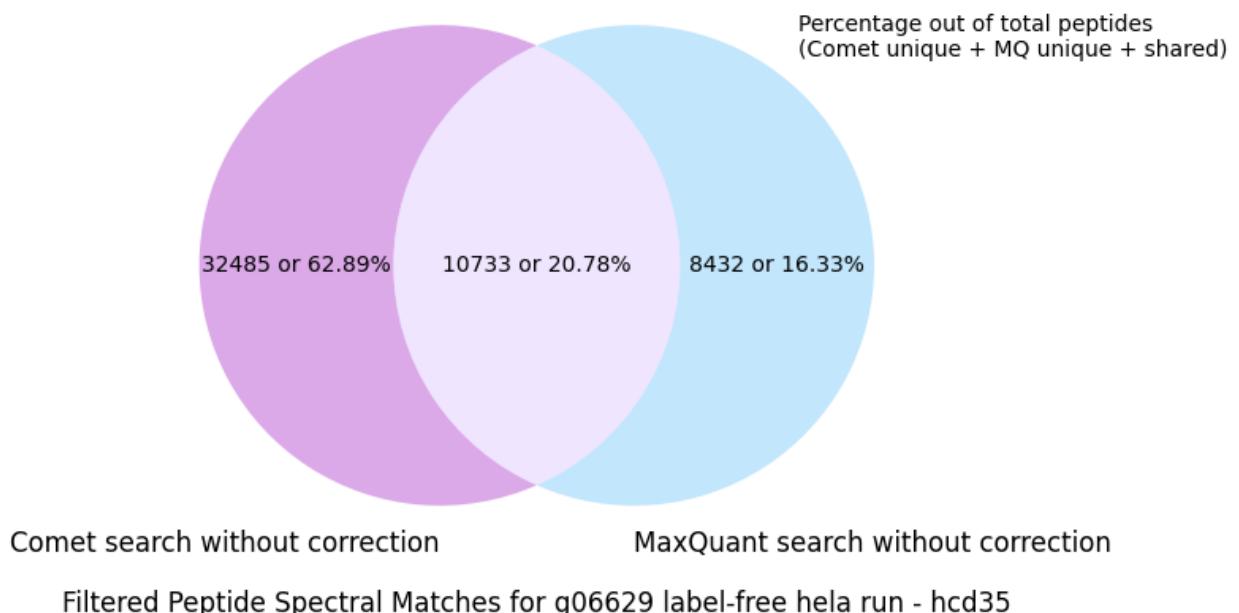
Filtered Peptide Spectral Matches for g06629 label-free hela run - hcd35



Appendix Figure 5: Unfiltered vs Filtered search results with Monocle correction for g06629.

Increased percentage of shared peptides with filtering and correction.

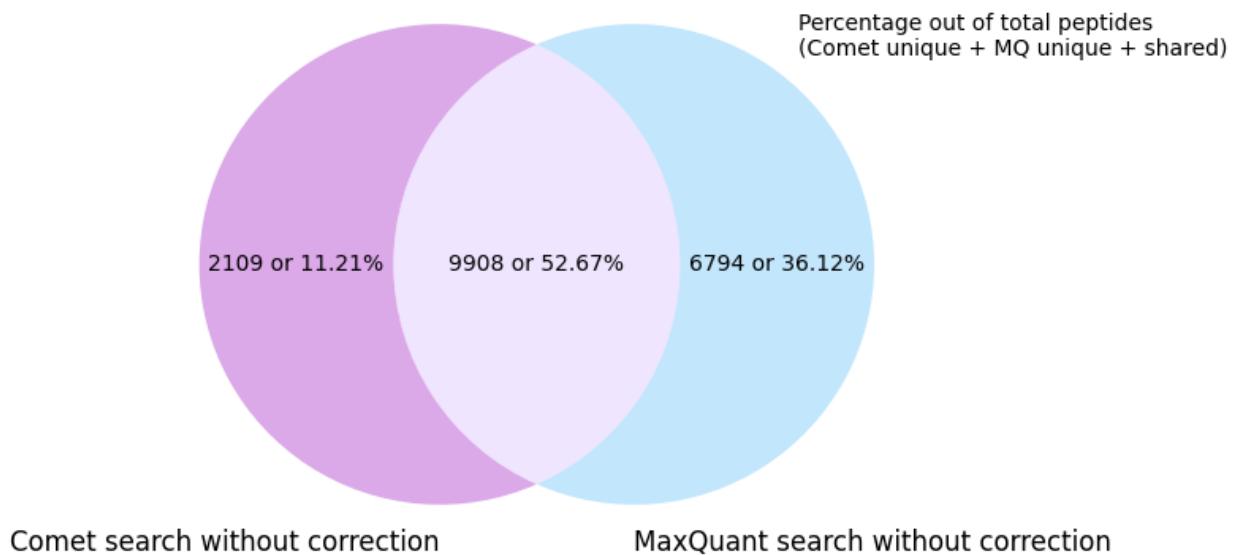
Unfiltered Peptide Spectral Matches for g06629 label-free hela run - hcd35



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06629 label-free hela run - hcd35



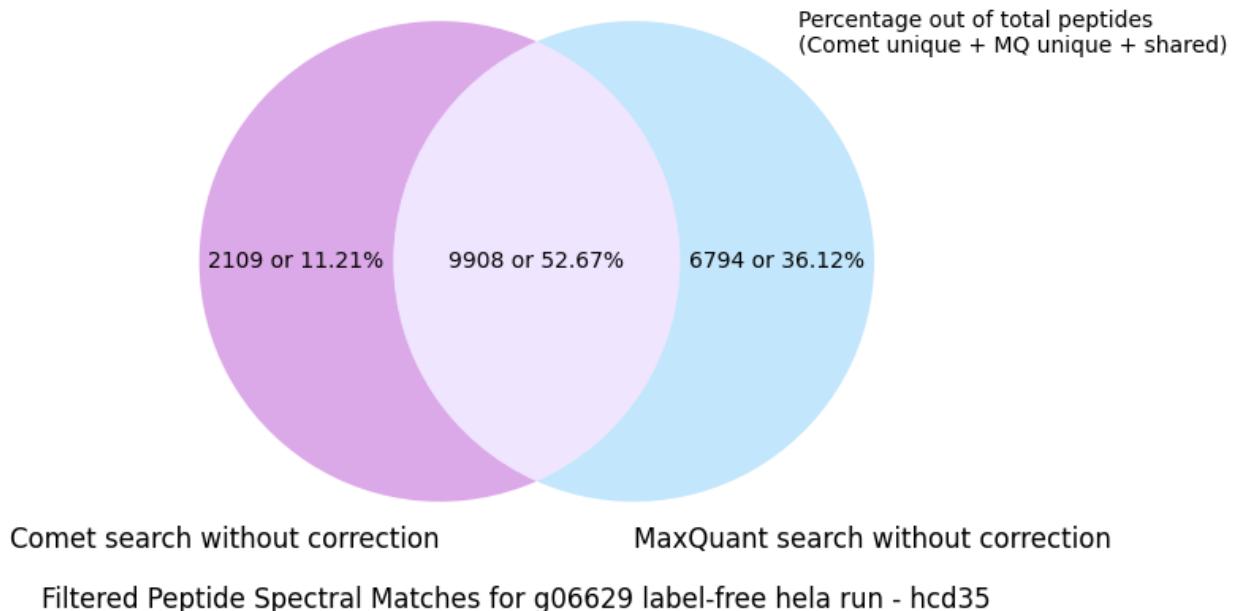
Comet search without correction

MaxQuant search without correction

Appendix Figure 6: Unfiltered vs Filtered search results without Monocle correction for g06629.

Increased percentage of shared peptides with filtering.

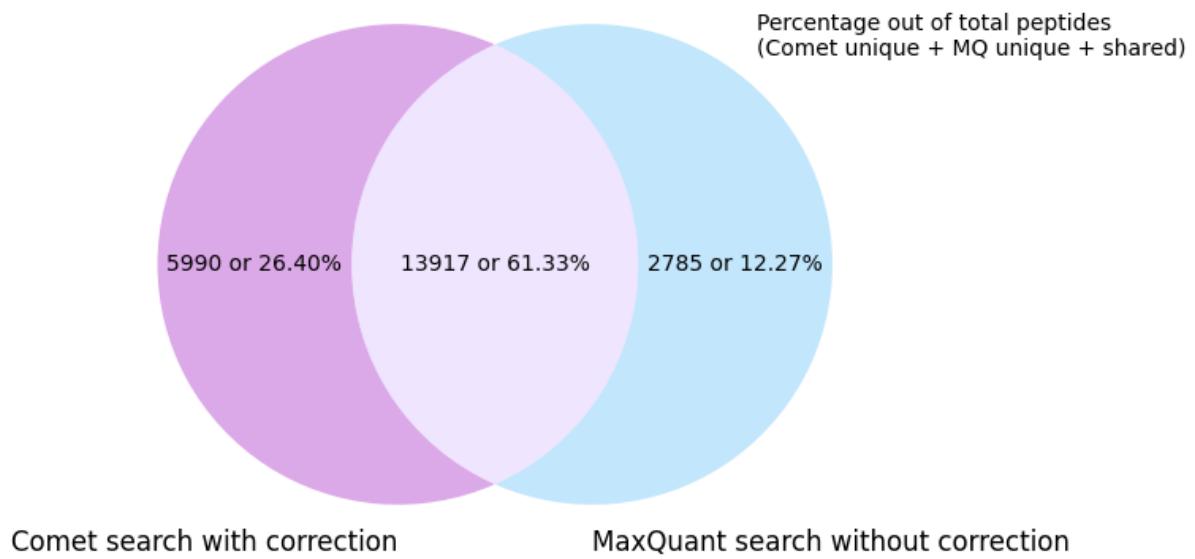
Filtered Peptide Spectral Matches for g06629 label-free hela run - hcd35



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06629 label-free hela run - hcd35



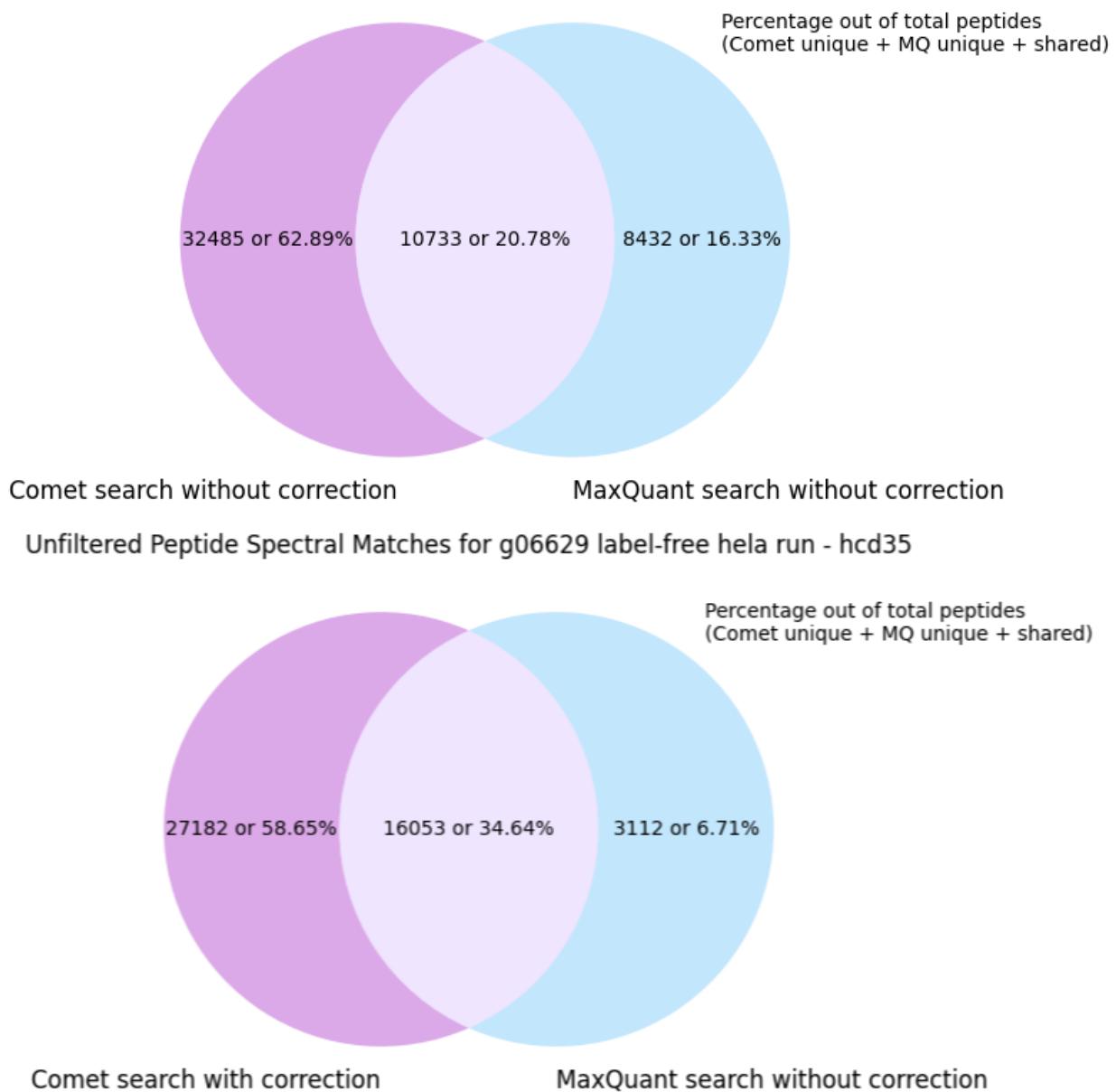
Comet search with correction

MaxQuant search without correction

Appendix Figure 7: Filtered search results with or without Monocle correction for g06629.

Increased number and percentage of shared peptides with filtering and correction.

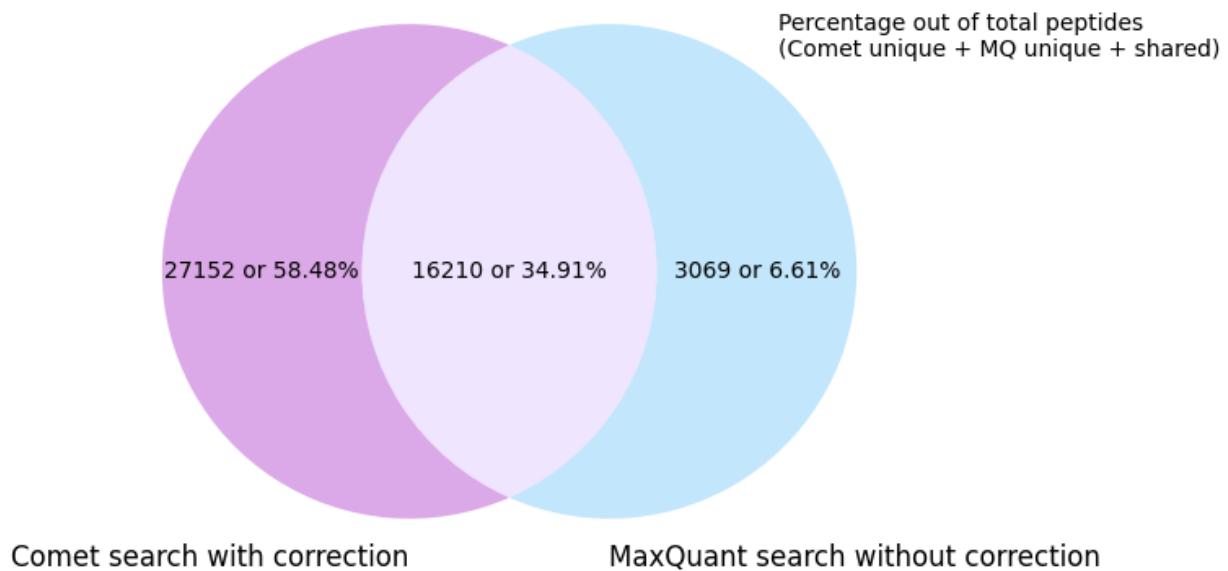
Unfiltered Peptide Spectral Matches for g06629 label-free hela run - hcd35



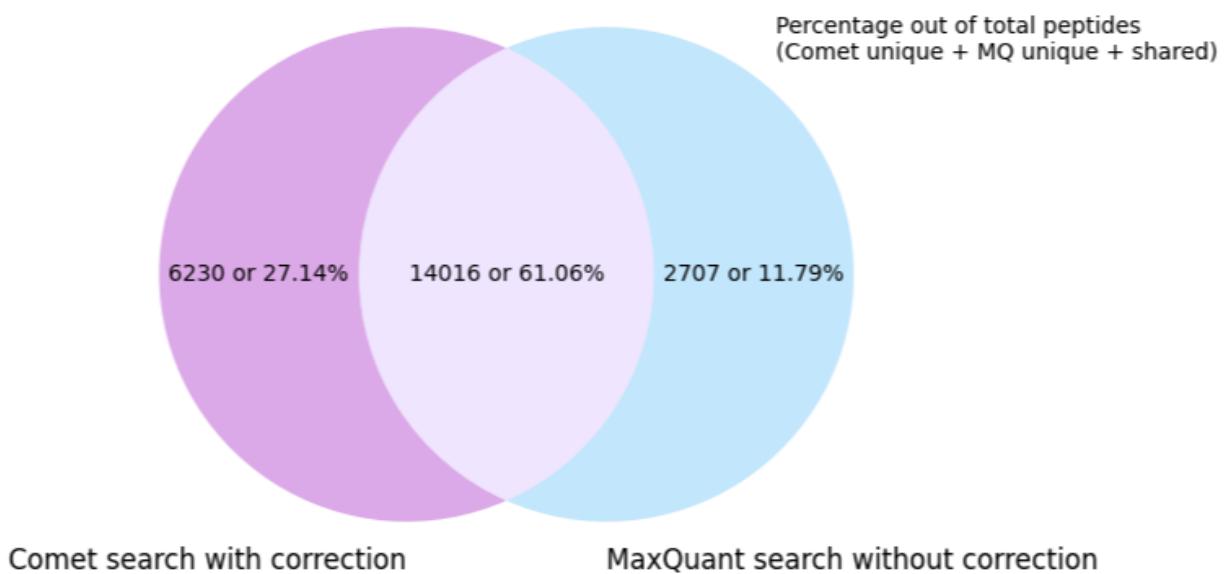
Appendix Figure 8: Unfiltered search results with or without Monocle correction for g06629.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06630 label-free hela run - hcd35



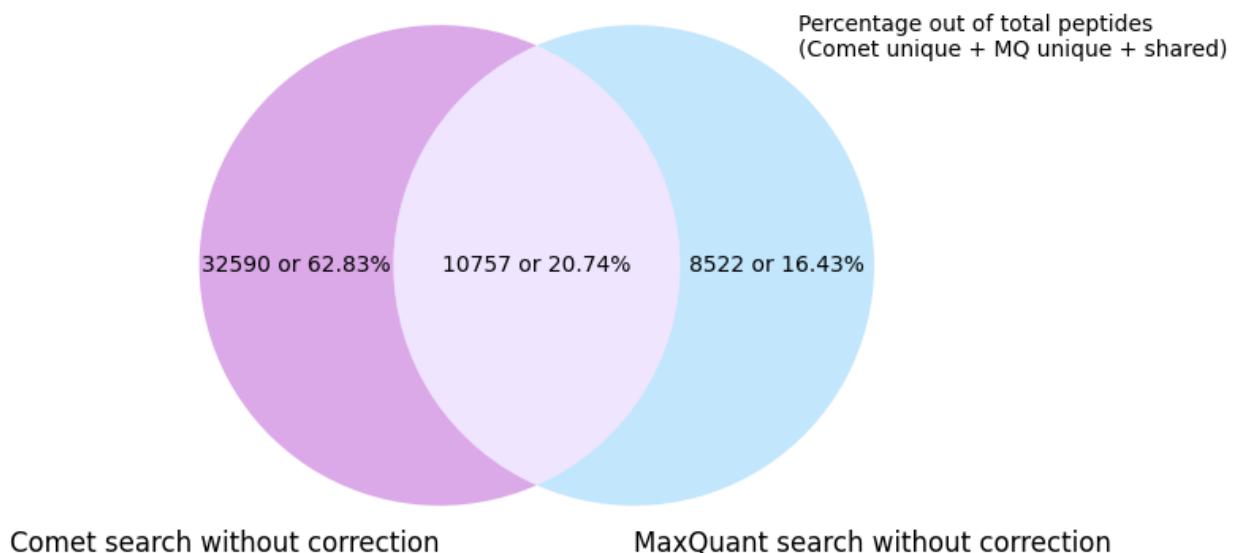
Filtered Peptide Spectral Matches for g06630 label-free hela run - hcd35



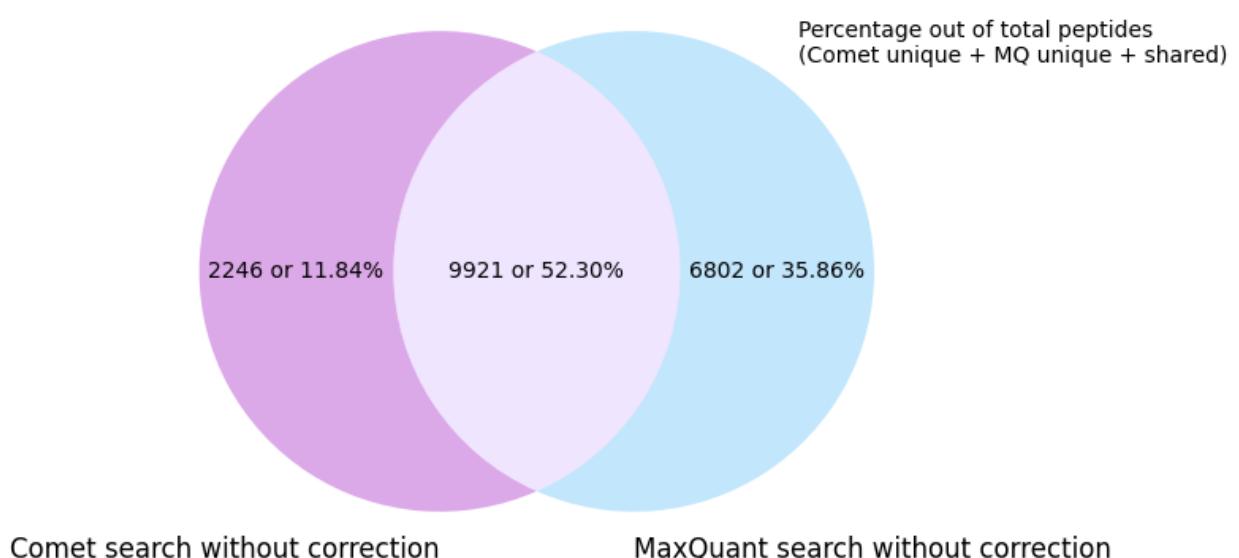
Appendix Figure 9: Unfiltered vs Filtered search results with Monocle correction for g06630.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06630 label-free hela run - hcd35

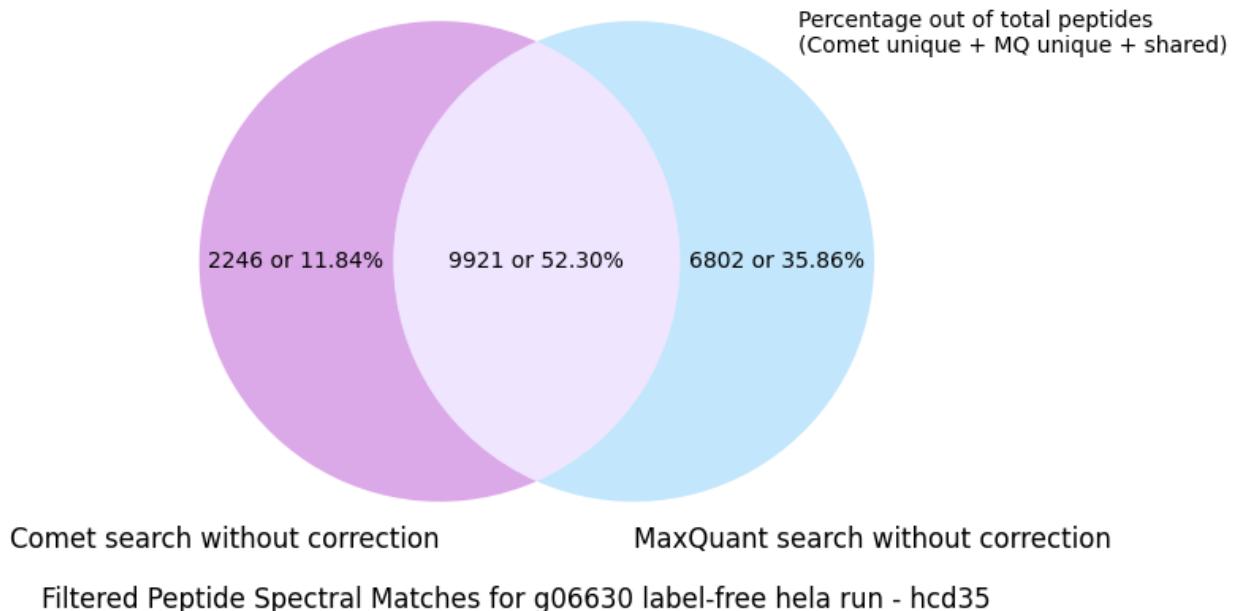


Filtered Peptide Spectral Matches for g06630 label-free hela run - hcd35

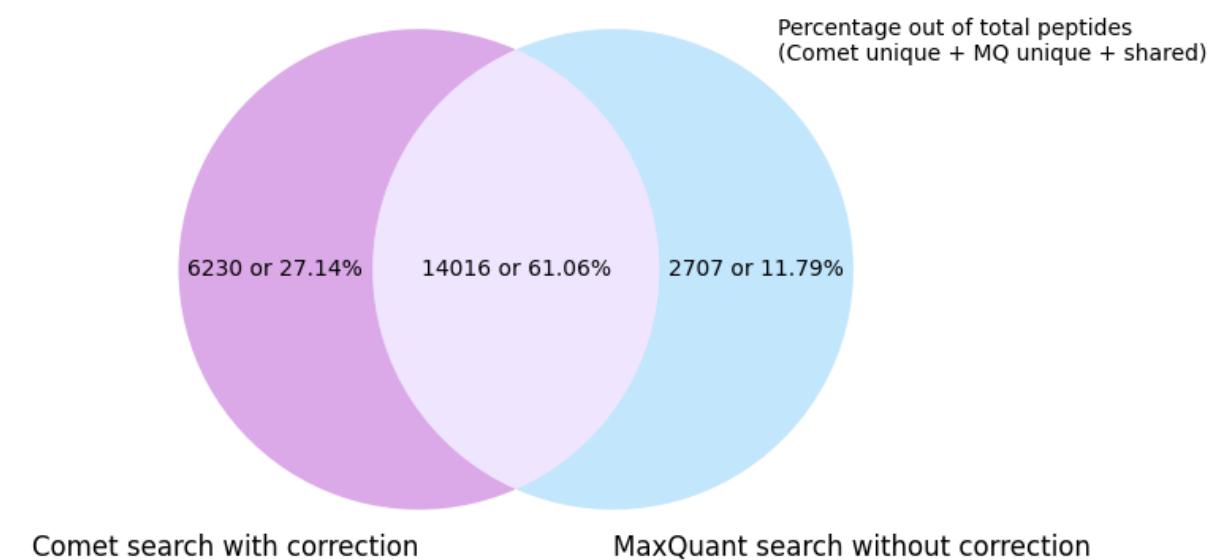


Appendix Figure 10: Unfiltered vs Filtered search results without Monocle correction for g06630. Increased percentage of shared peptides with filtering.

Filtered Peptide Spectral Matches for g06630 label-free hela run - hcd35



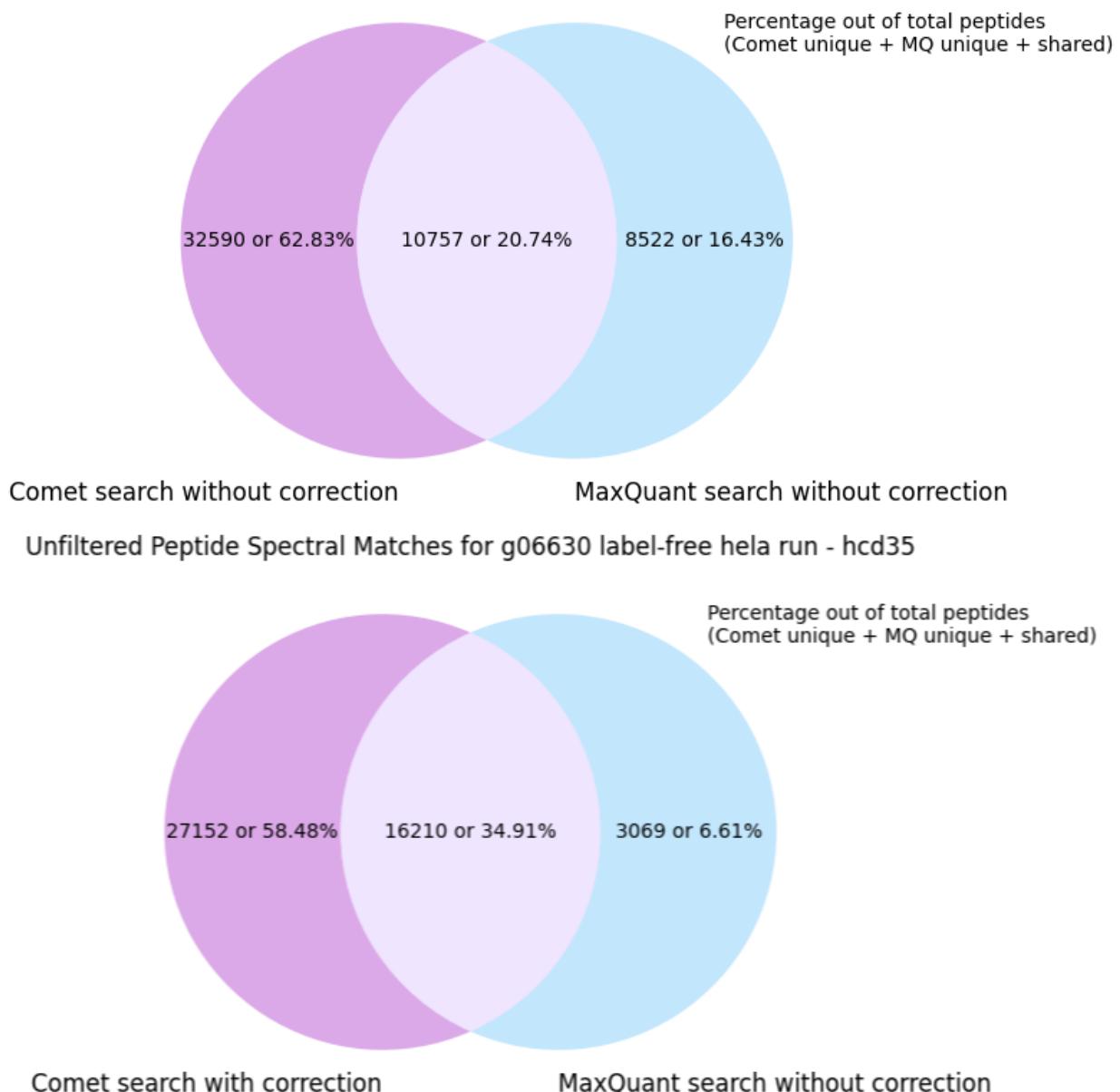
Filtered Peptide Spectral Matches for g06630 label-free hela run - hcd35



Appendix Figure 11: Filtered search results with or without Monocle correction for g06630.

Increased number and percentage of shared peptides with filtering and correction.

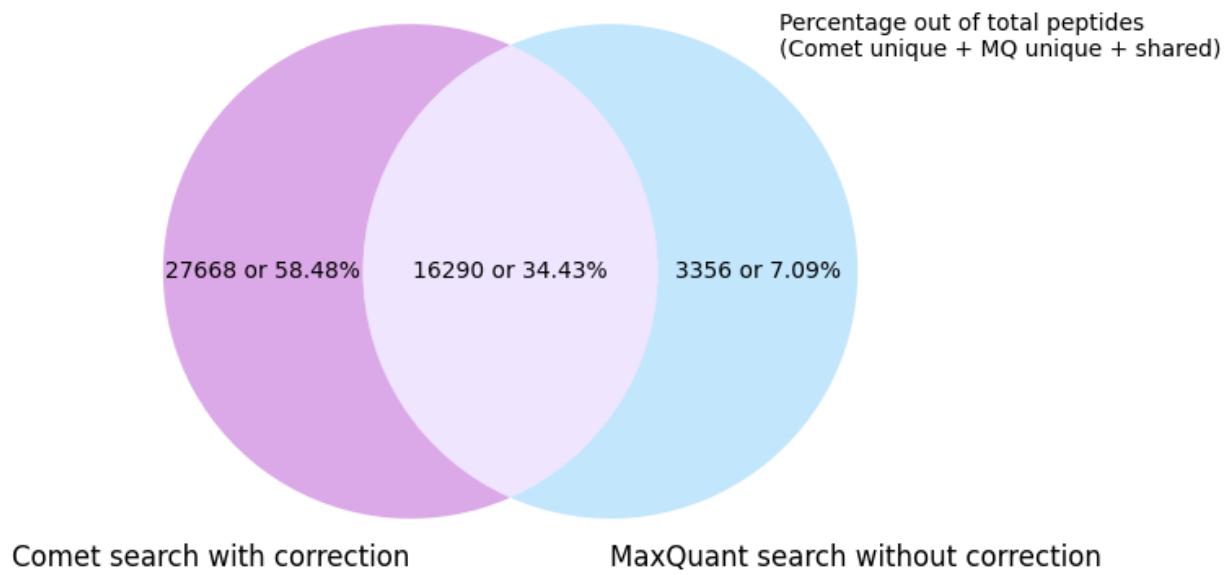
Unfiltered Peptide Spectral Matches for g06630 label-free hela run - hcd35



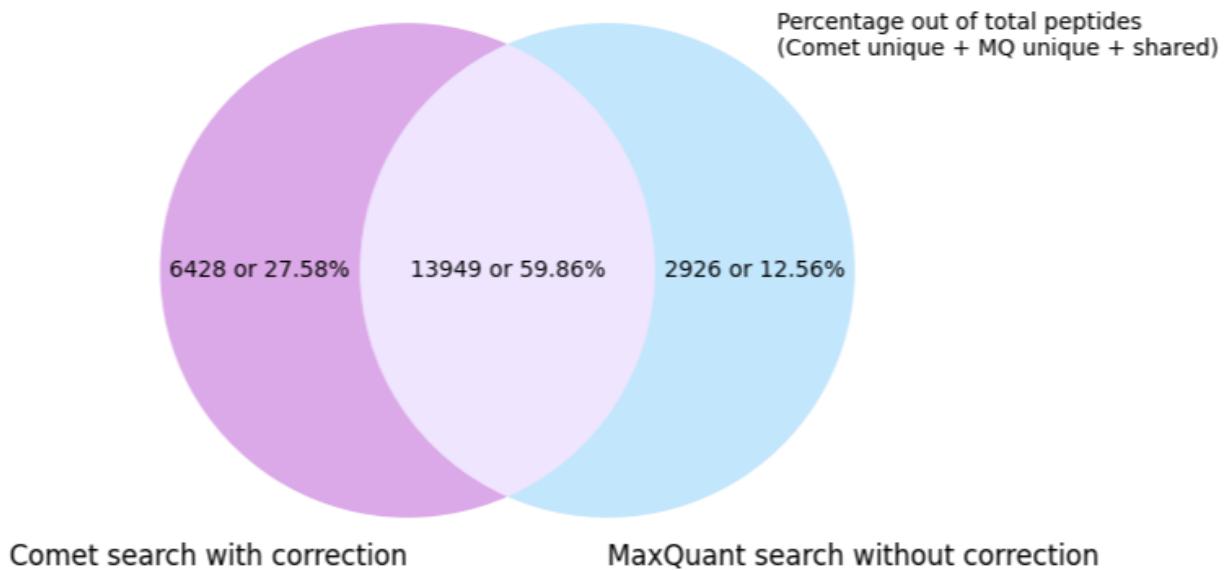
Appendix Figure 12: Unfiltered search results with or without Monocle correction for g06630.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06637 label-free hela run - hcd28



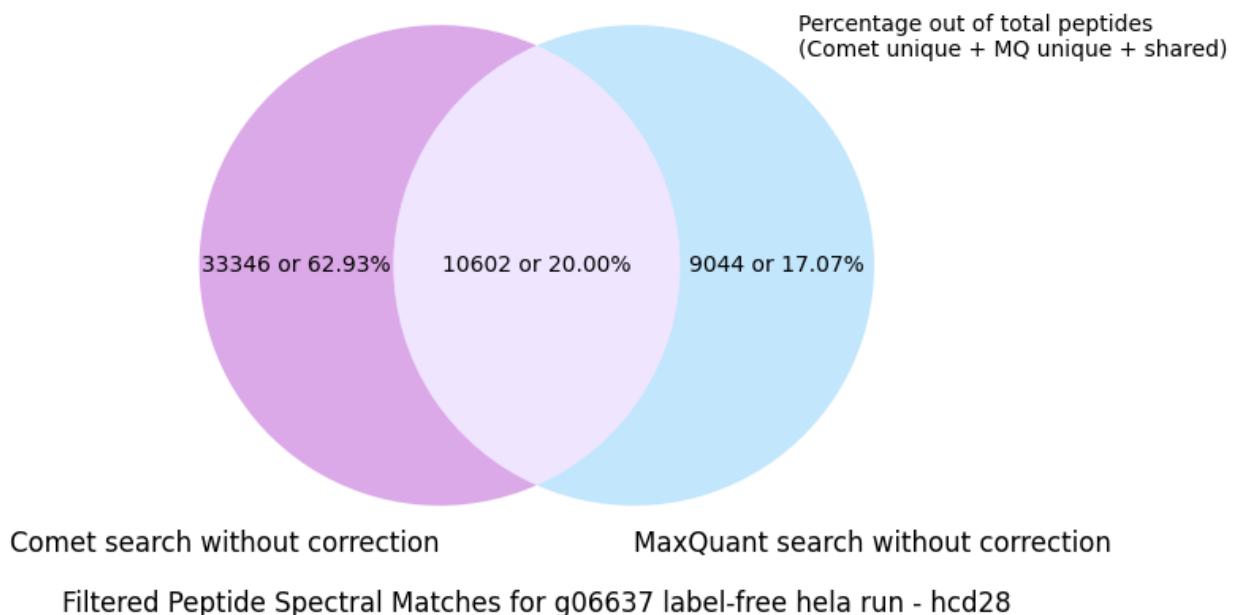
Filtered Peptide Spectral Matches for g06637 label-free hela run - hcd28



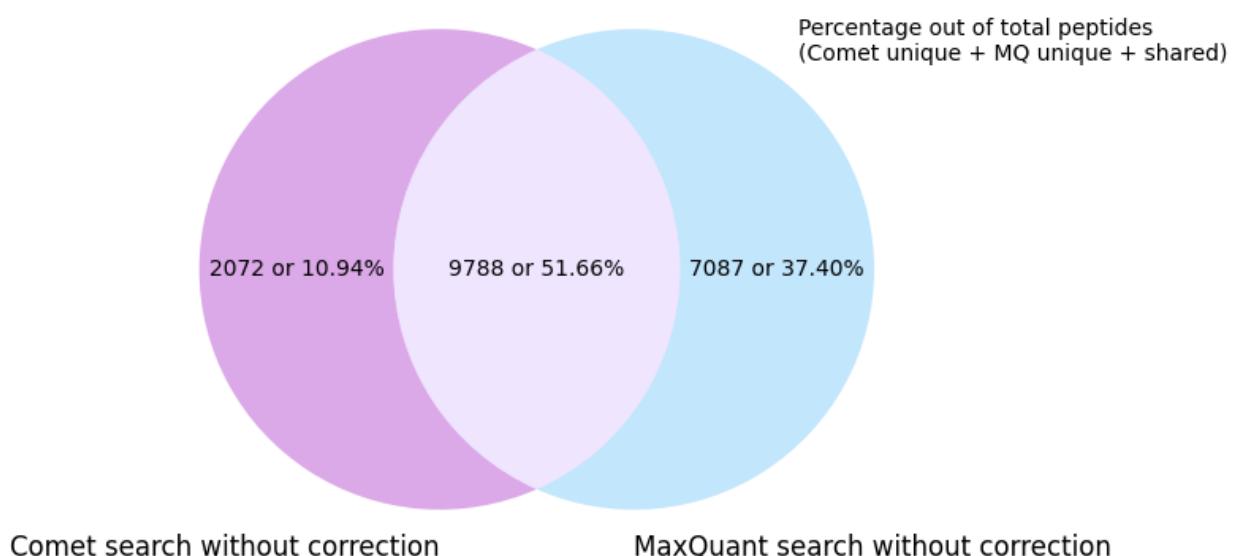
Appendix Figure 13: Unfiltered vs Filtered search results with Monocle correction for g06637.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06637 label-free hela run - hcd28

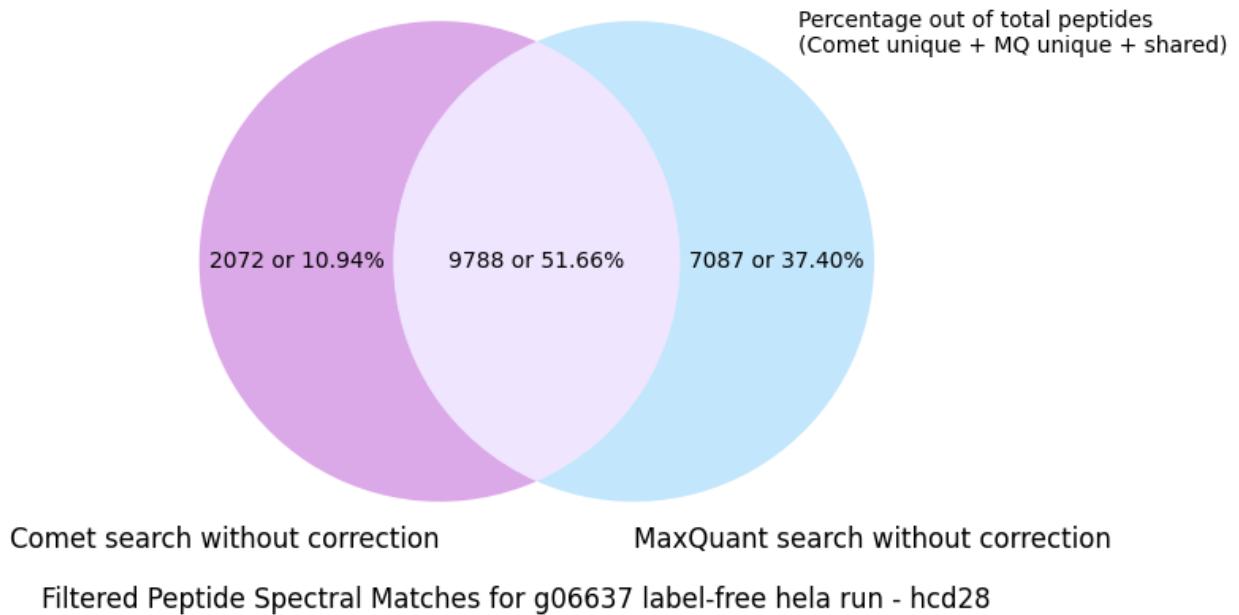


Filtered Peptide Spectral Matches for g06637 label-free hela run - hcd28



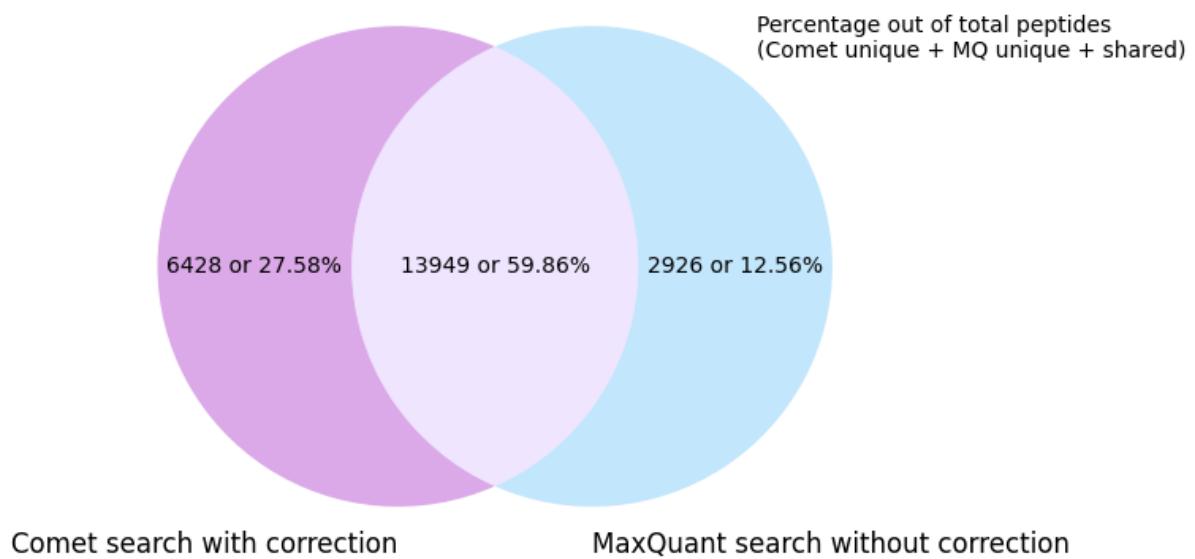
Appendix Figure 14: Unfiltered vs Filtered search results without Monocle correction for g06637. Increased percentage of shared peptides with filtering.

Filtered Peptide Spectral Matches for g06637 label-free hela run - hcd28



Comet search without correction MaxQuant search without correction

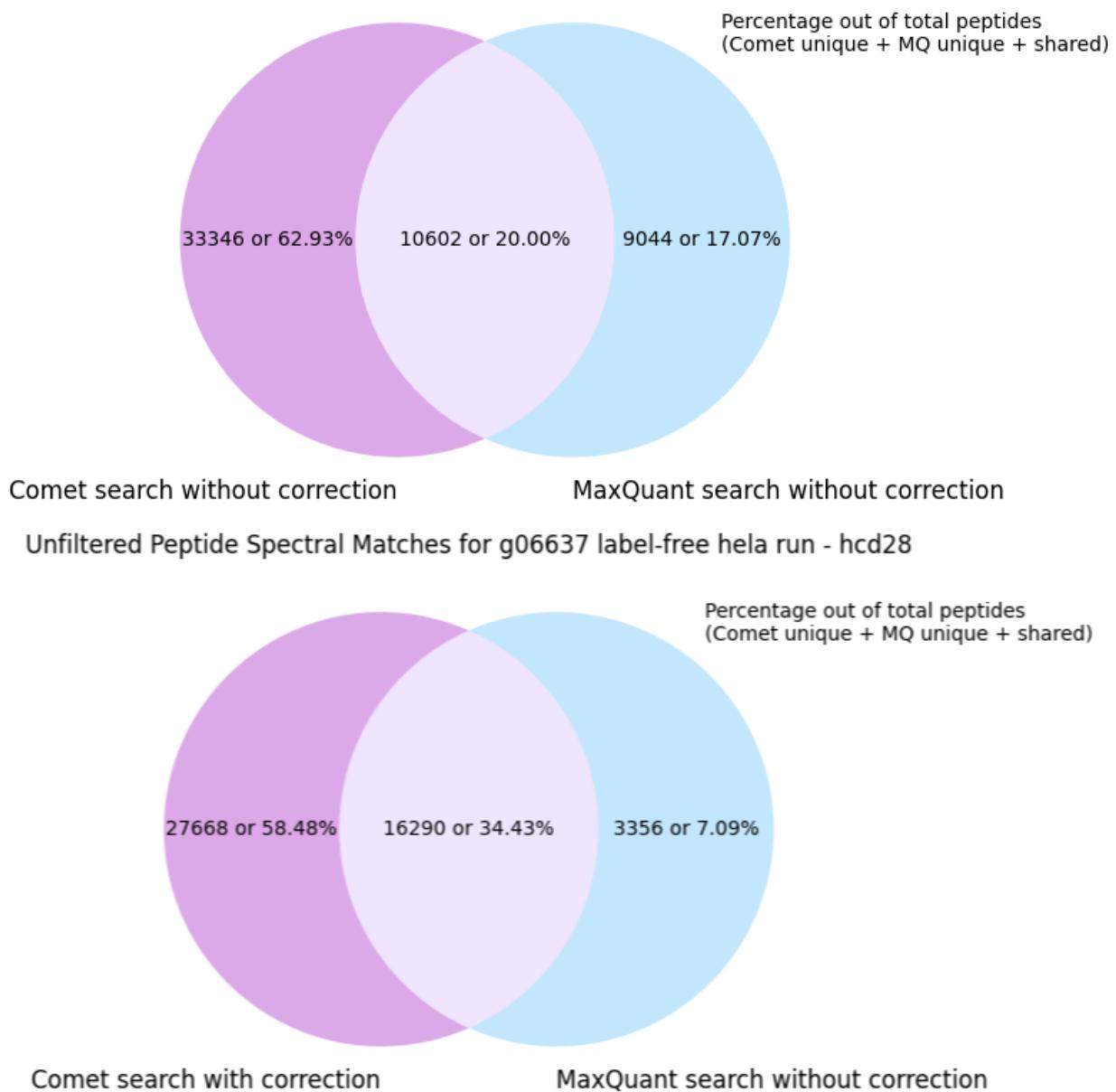
Filtered Peptide Spectral Matches for g06637 label-free hela run - hcd28



Appendix Figure 15: Filtered search results with or without Monocle correction for g06637.

Increased number and percentage of shared peptides with filtering and correction.

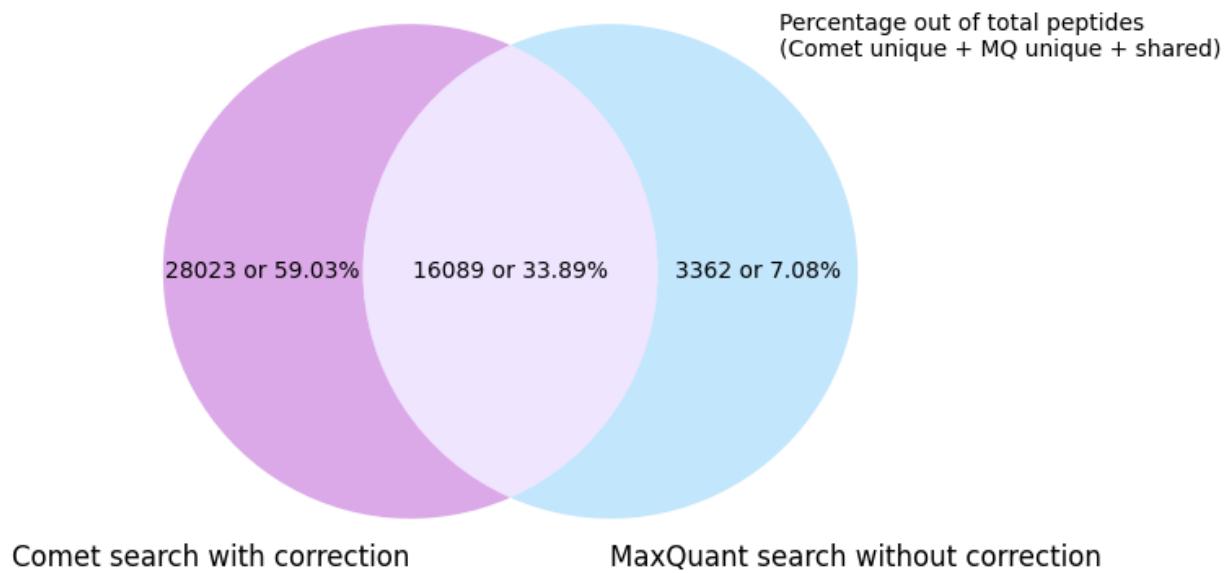
Unfiltered Peptide Spectral Matches for g06637 label-free hela run - hcd28



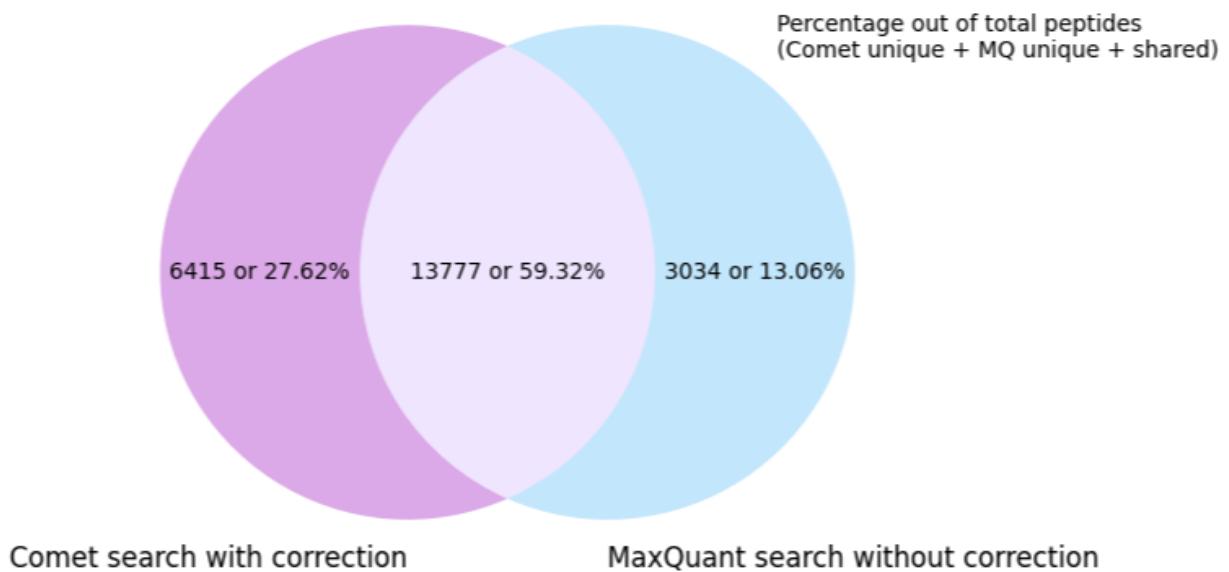
Appendix Figure 16: Unfiltered search results with or without Monocle correction for g06637.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06638 label-free hela run - hcd28



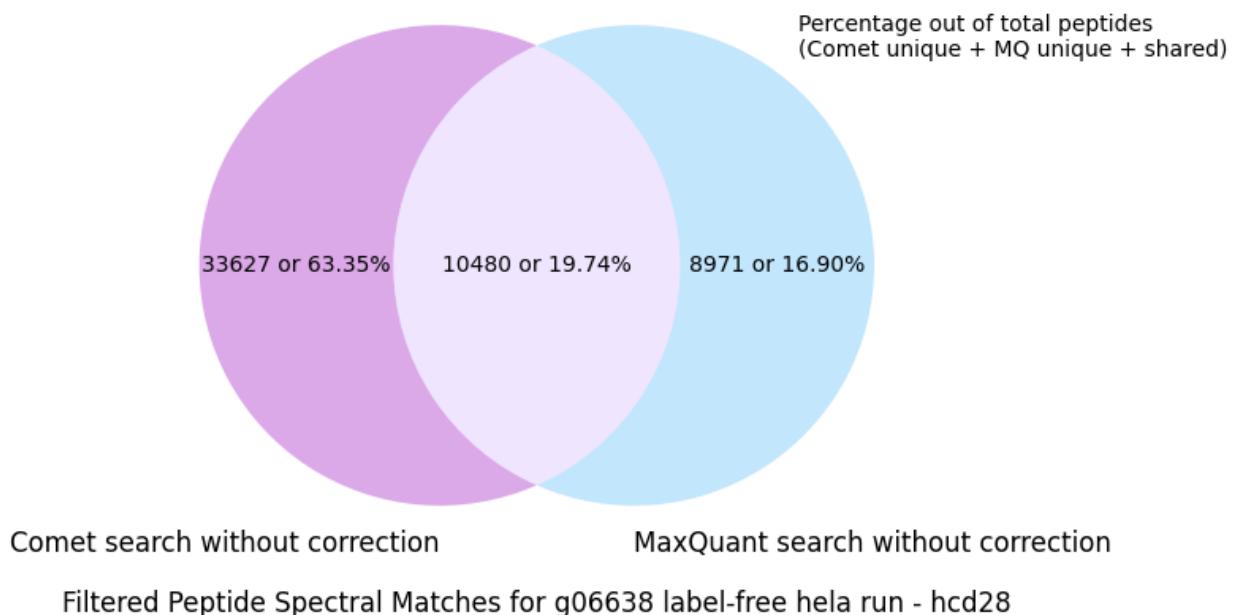
Filtered Peptide Spectral Matches for g06638 label-free hela run - hcd28



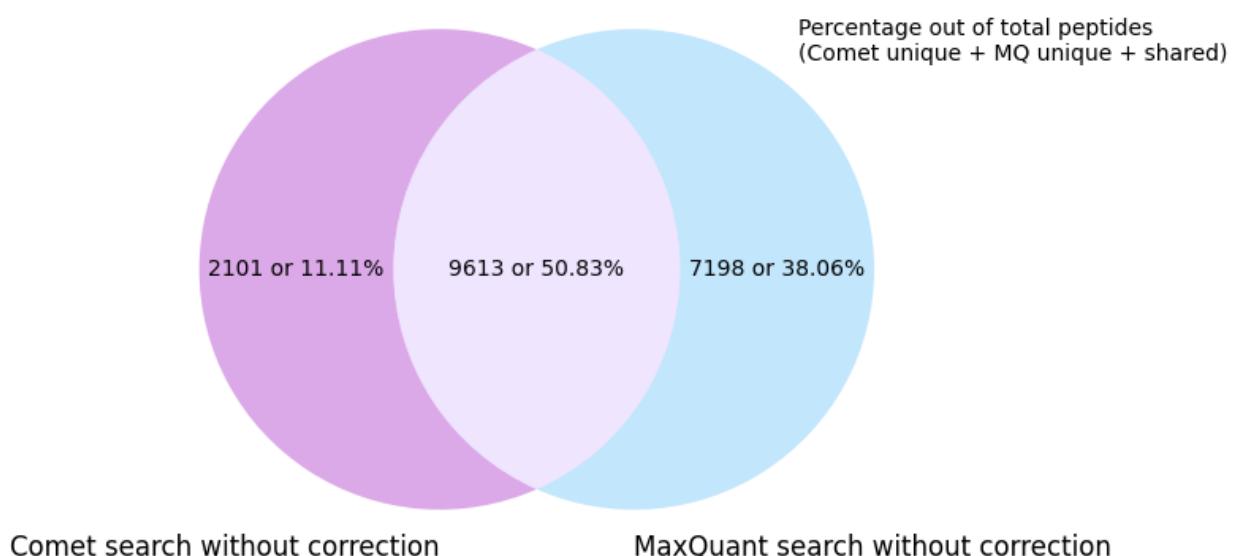
Appendix Figure 17: Unfiltered vs Filtered search results with Monocle correction for g06638.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06638 label-free hela run - hcd28

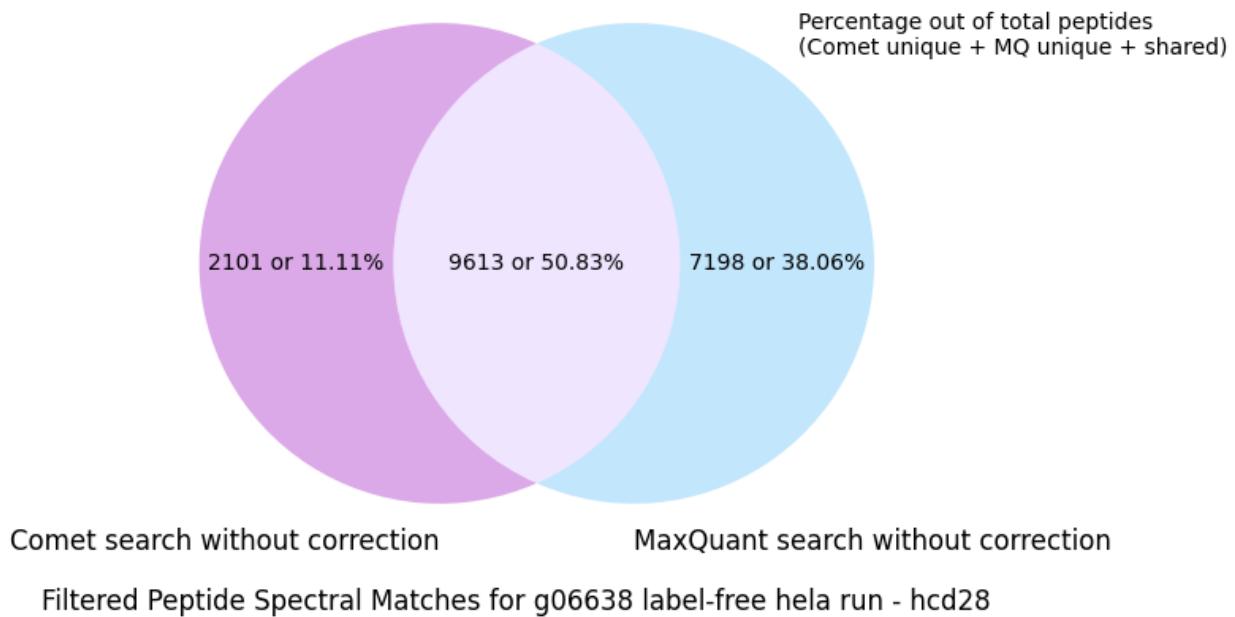


Filtered Peptide Spectral Matches for g06638 label-free hela run - hcd28



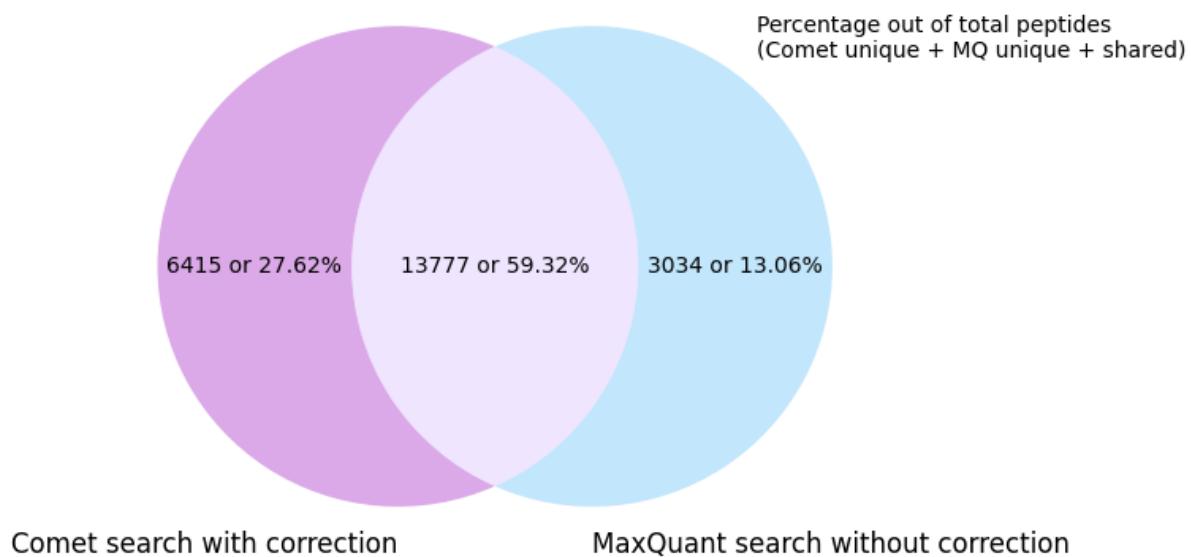
Appendix Figure 18: Unfiltered vs Filtered search results without Monocle correction for g06638. Increased percentage of shared peptides with filtering.

Filtered Peptide Spectral Matches for g06638 label-free hela run - hcd28



Comet search without correction MaxQuant search without correction

Filtered Peptide Spectral Matches for g06638 label-free hela run - hcd28



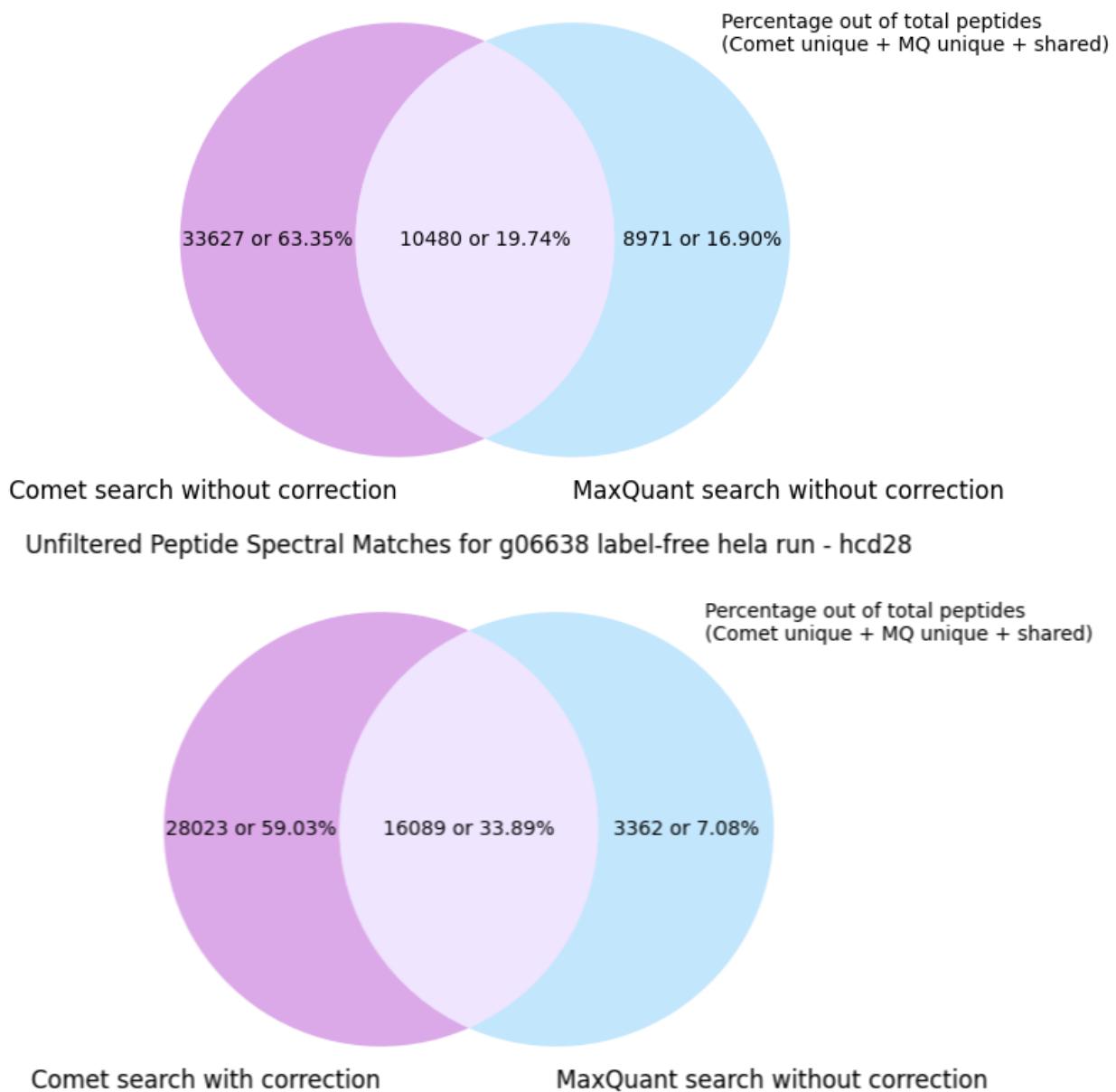
Comet search with correction

MaxQuant search without correction

Appendix Figure 19: Filtered search results with or without Monocle correction for g06638.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06638 label-free hela run - hcd28

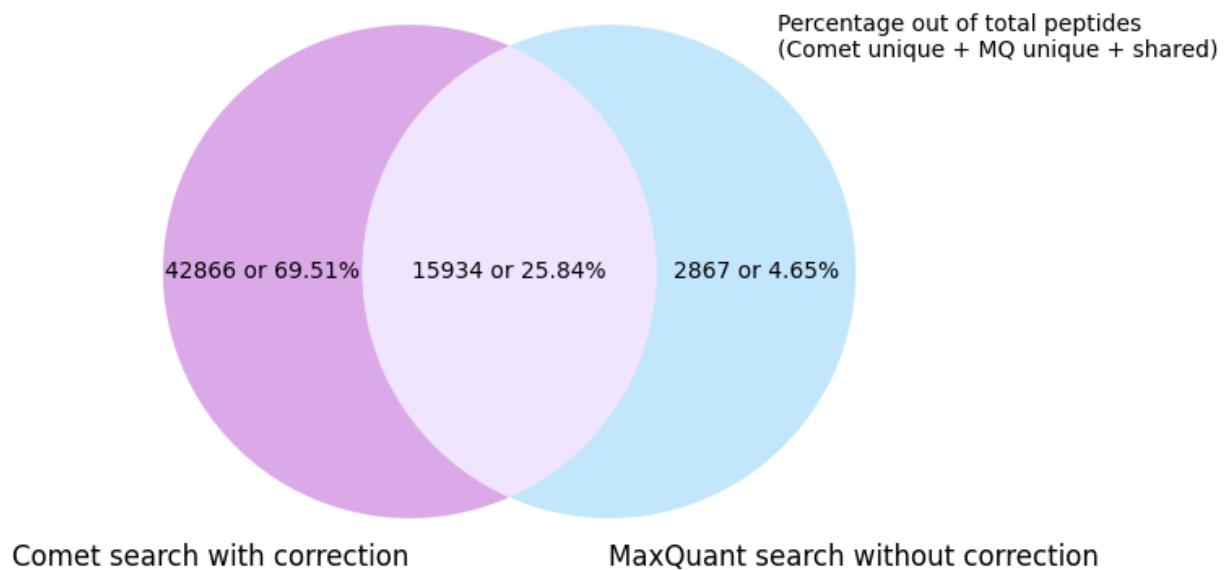


Appendix Figure 20: Unfiltered search results with or without Monocle correction for g06638.

Increased number and percentage of shared peptides with filtering and correction.

YEAST FIGURES

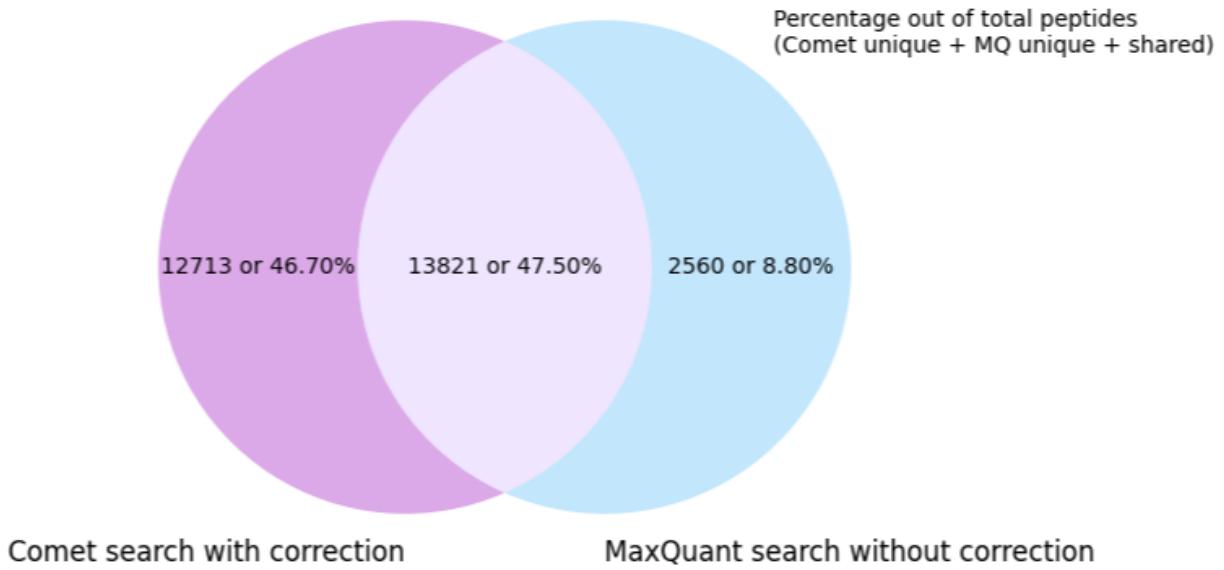
Unfiltered Peptide Spectral Matches for g06610 label-free yeast run - cid35



Comet search with correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06610 label-free yeast run - cid35



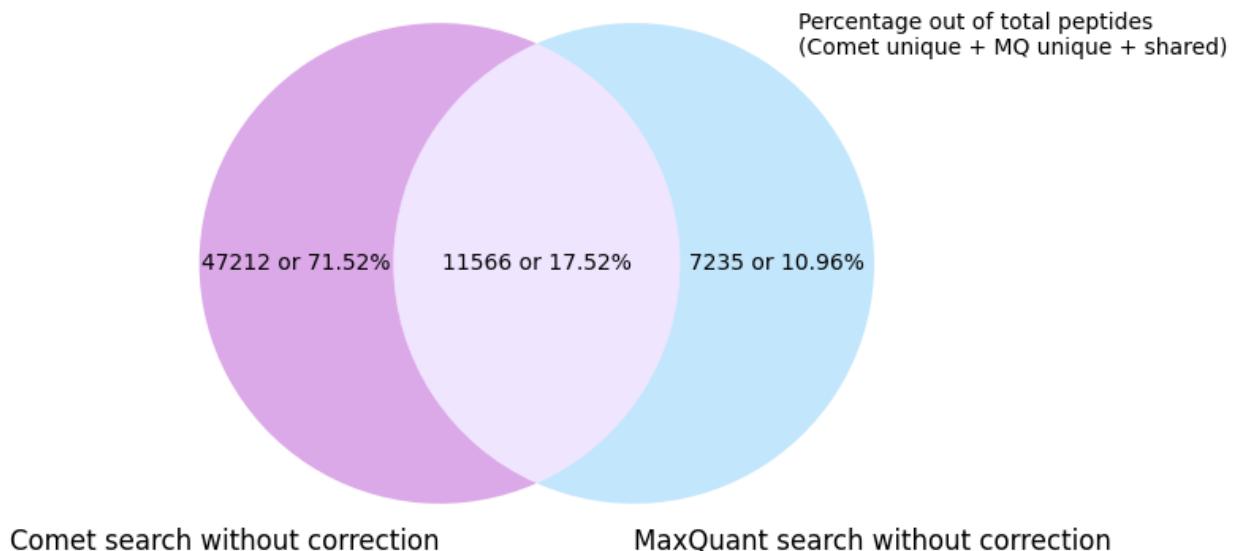
Comet search with correction

MaxQuant search without correction

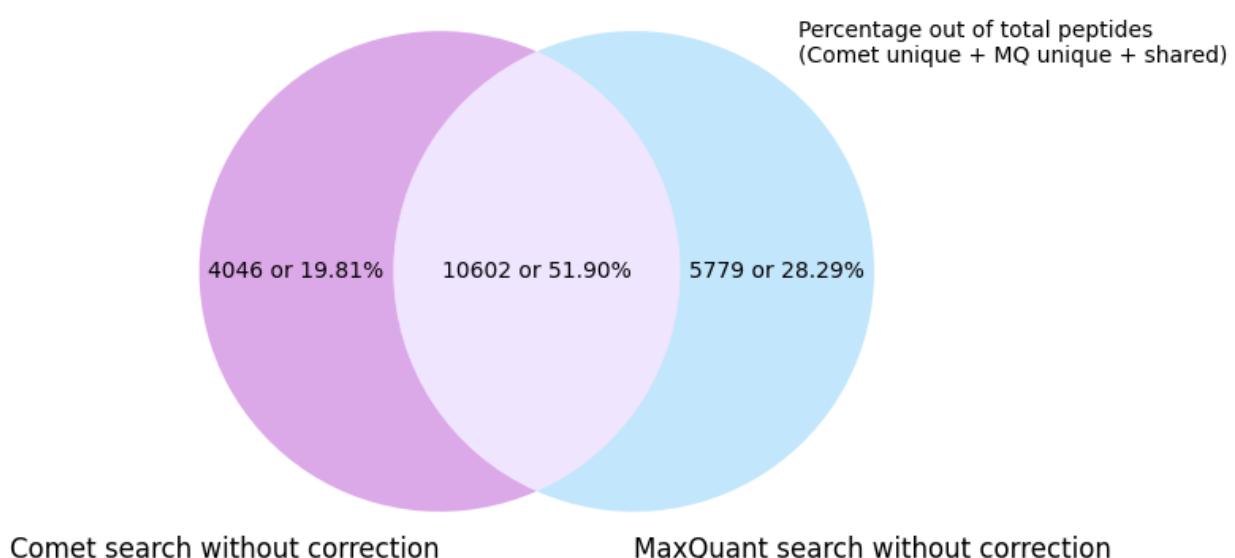
Appendix Figure 21: Unfiltered vs Filtered search results with Monocle correction for g06610.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06610 label-free yeast run - cid35

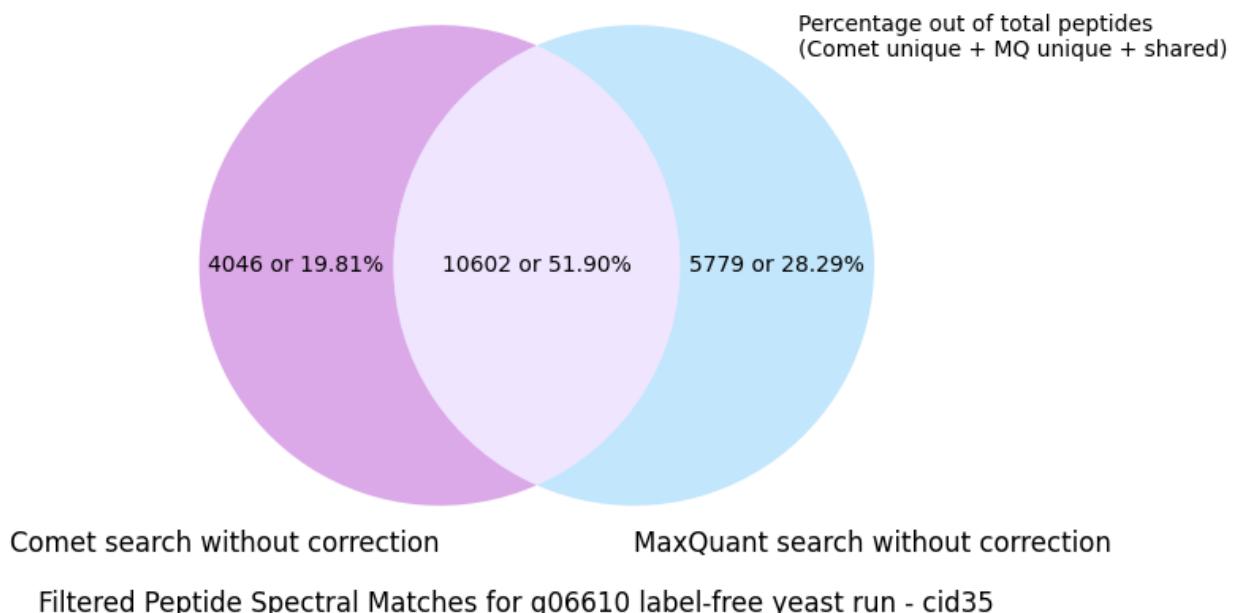


Filtered Peptide Spectral Matches for g06610 label-free yeast run - cid35



Appendix Figure 22: Unfiltered vs Filtered search results without Monocle correction for g06610. Increased percentage of shared peptides with filtering.

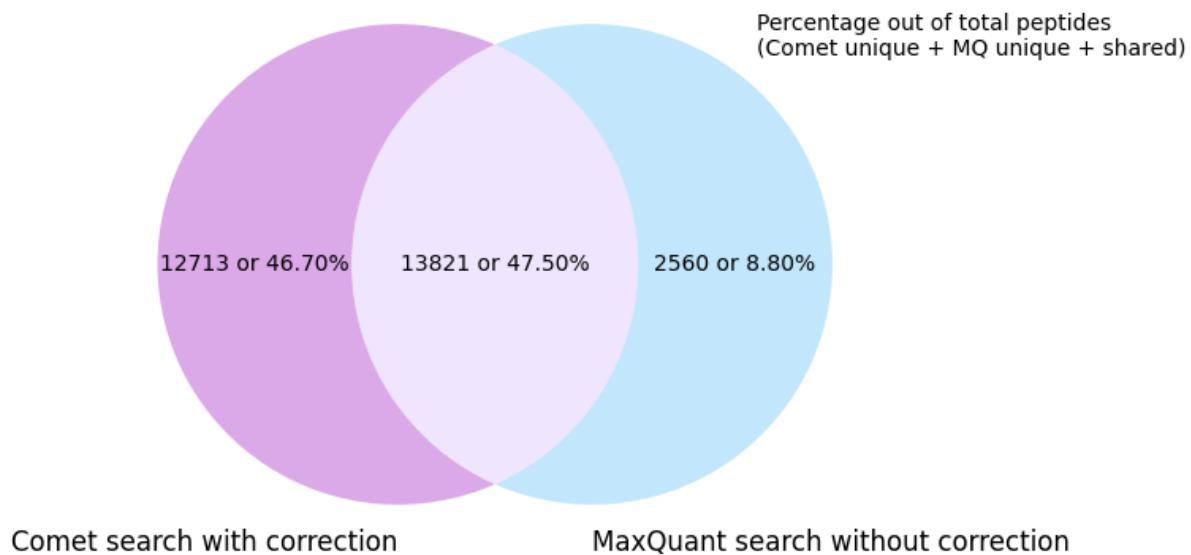
Filtered Peptide Spectral Matches for g06610 label-free yeast run - cid35



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06610 label-free yeast run - cid35



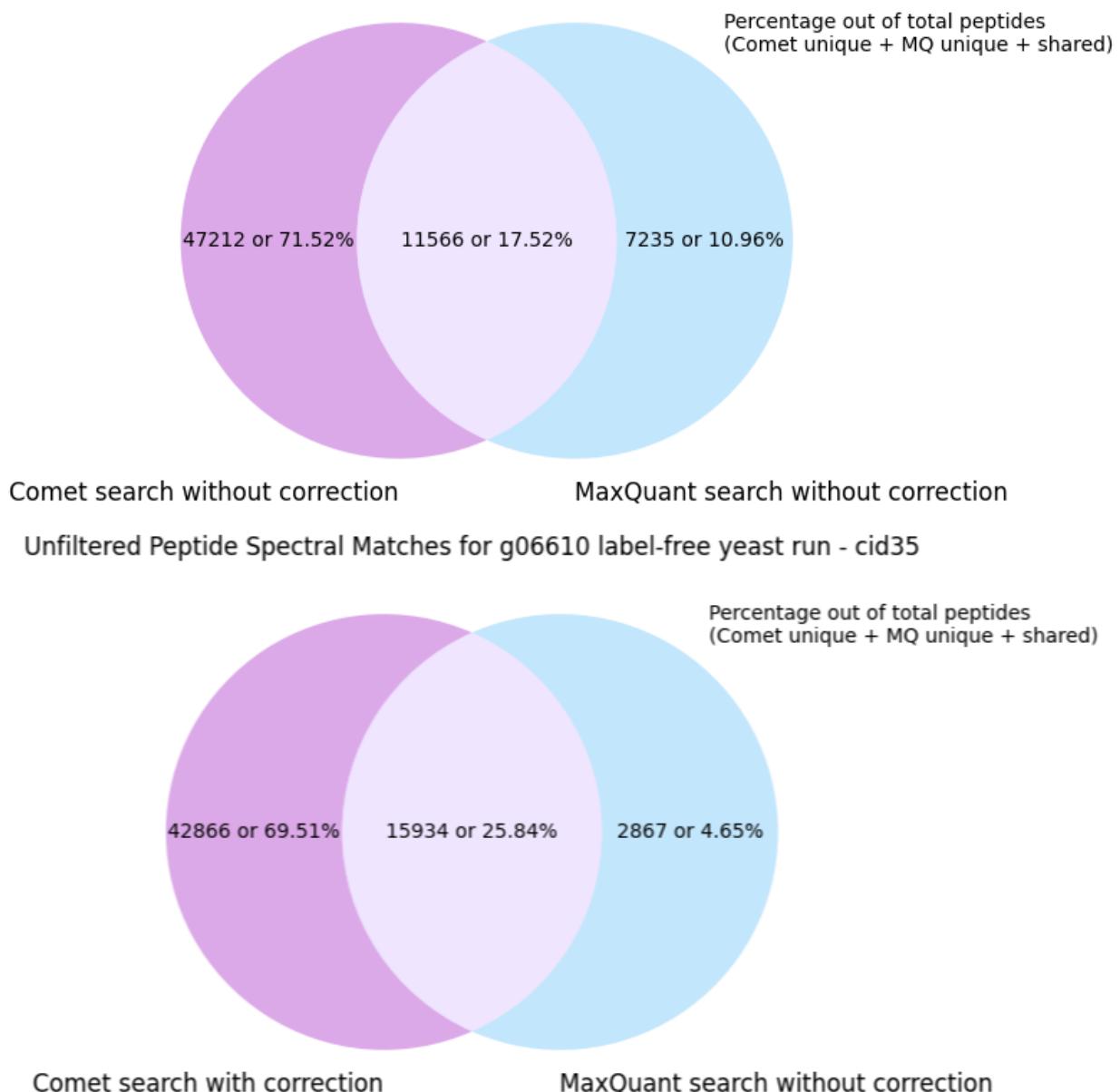
Comet search with correction

MaxQuant search without correction

Appendix Figure 23: Filtered search results with or without Monocle correction for g06610.

Increased number but decreased percentage of shared peptides with filtering and correction.

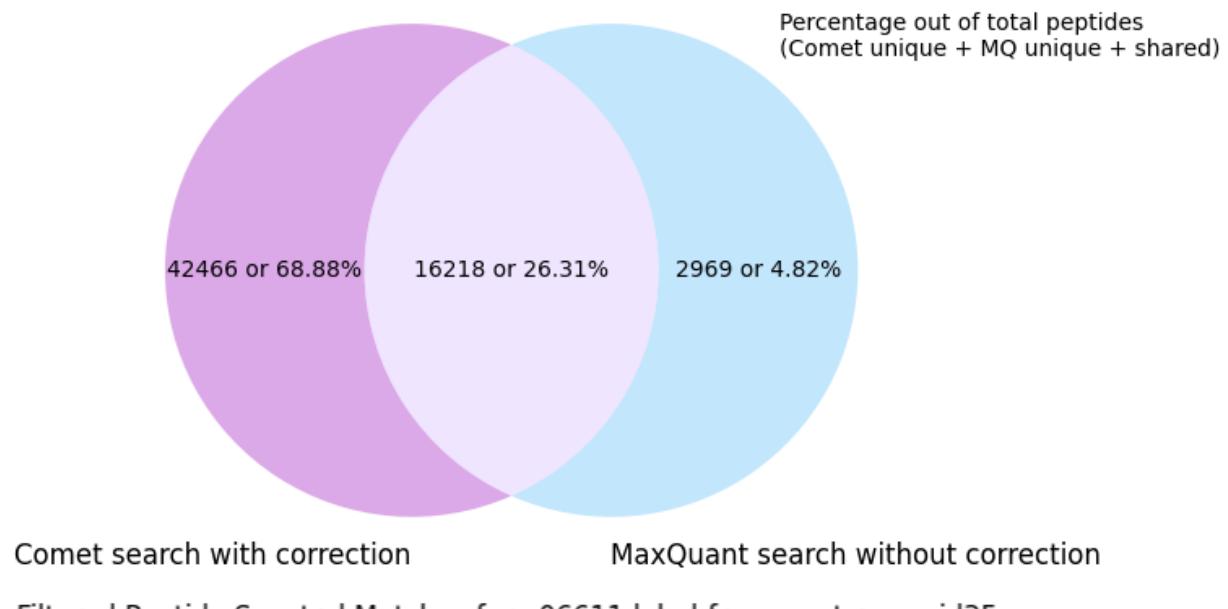
Unfiltered Peptide Spectral Matches for g06610 label-free yeast run - cid35



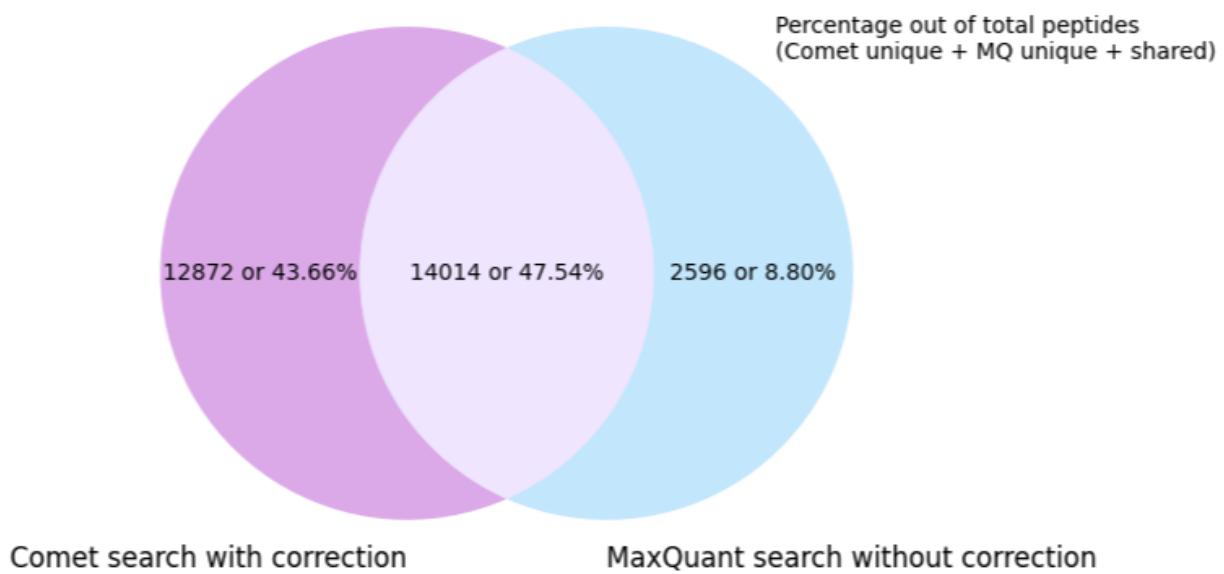
Appendix Figure 24: Unfiltered search results with or without Monocle correction for g06610.

Increased number and percentage of shared peptides with filtering.

Unfiltered Peptide Spectral Matches for g06611 label-free yeast run - cid35



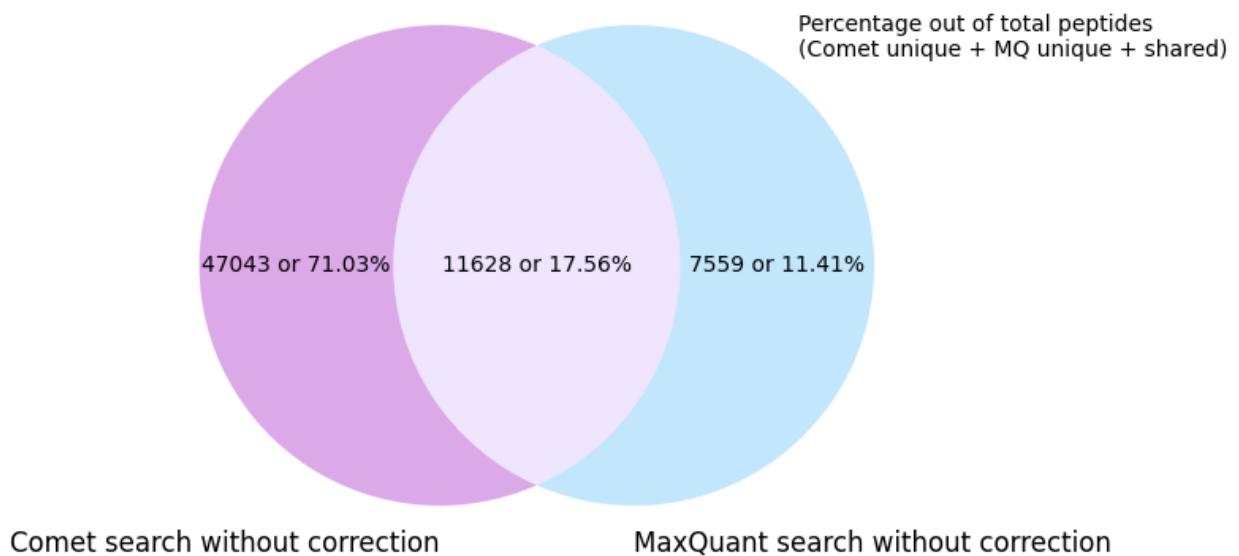
Filtered Peptide Spectral Matches for g06611 label-free yeast run - cid35



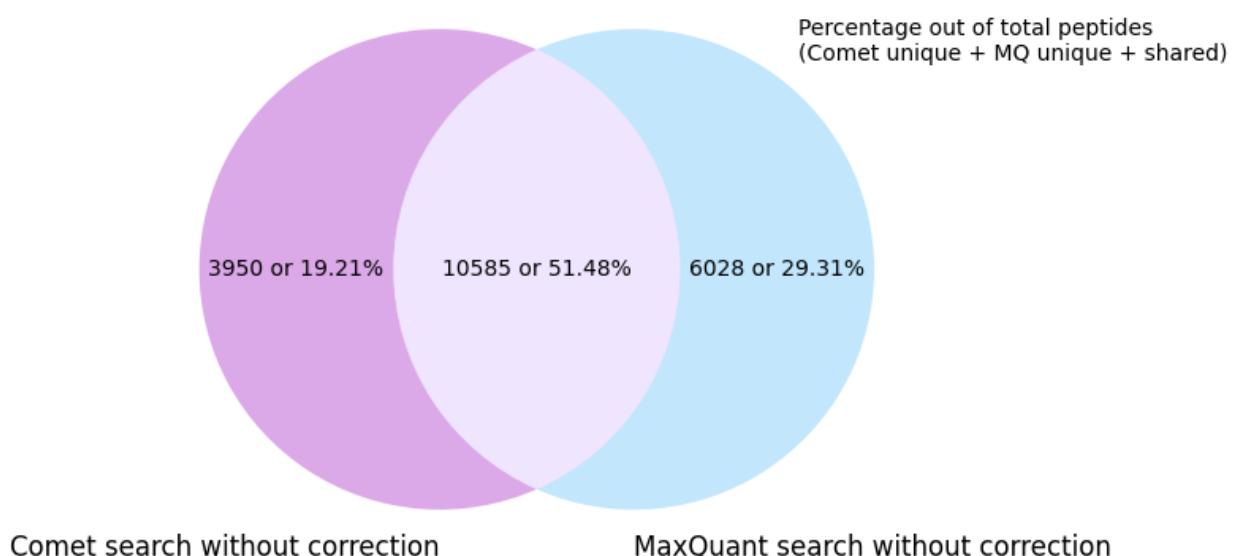
Appendix Figure 25: Unfiltered vs Filtered search results with Monocle correction for g06611.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06611 label-free yeast run - cid35

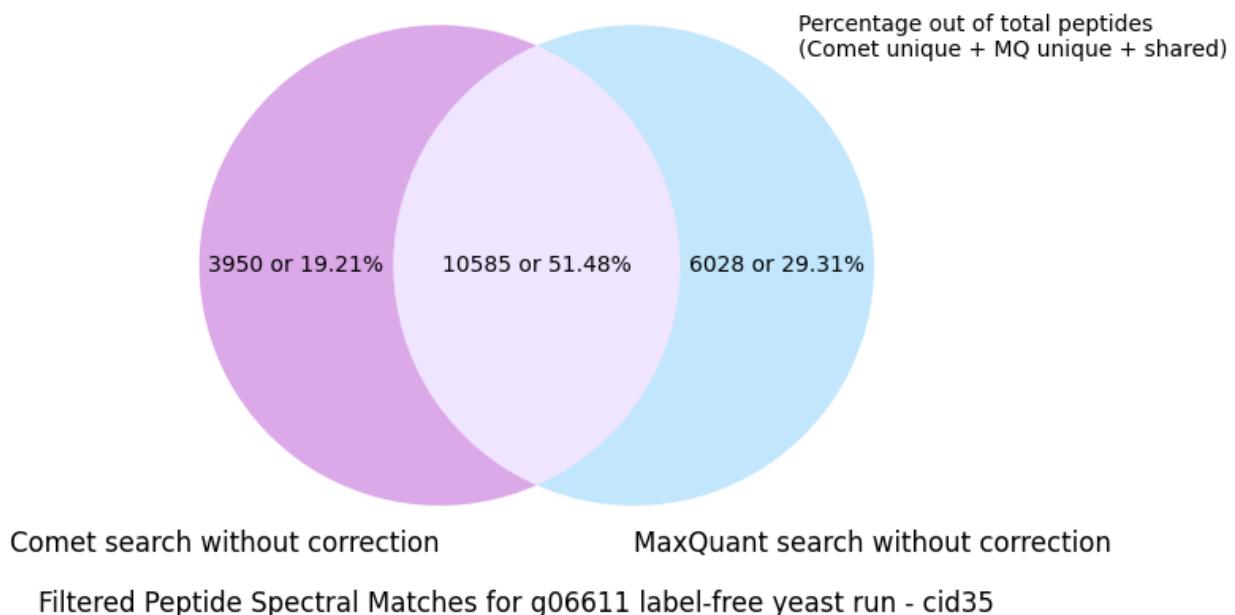


Filtered Peptide Spectral Matches for g06611 label-free yeast run - cid35



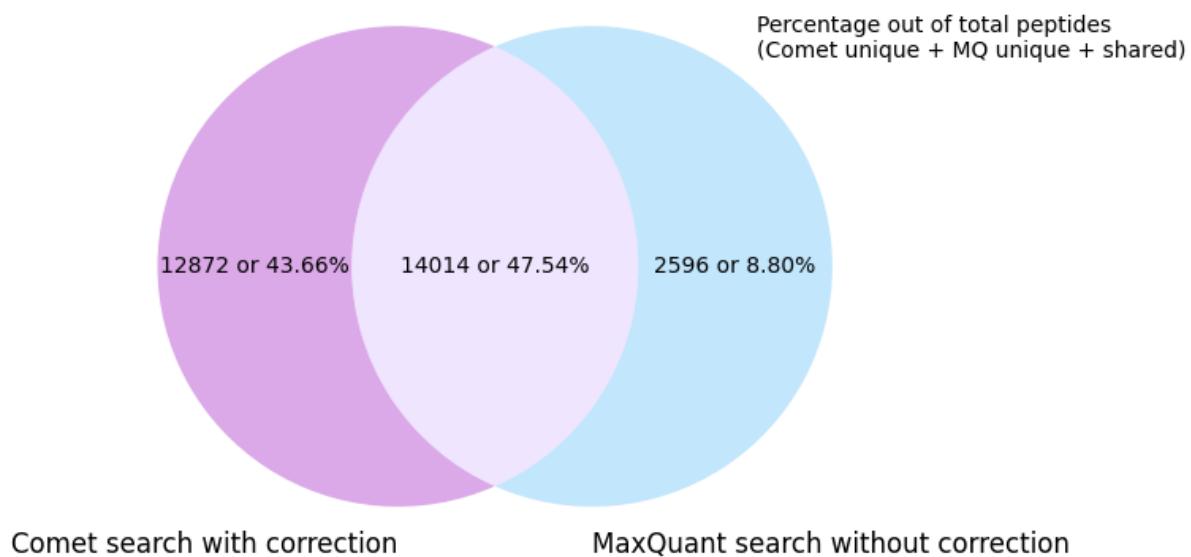
Appendix Figure 26: Unfiltered vs Filtered search results without Monocle correction for g06611. Increased percentage of shared peptides with filtering and correction.

Filtered Peptide Spectral Matches for g06611 label-free yeast run - cid35



Comet search without correction MaxQuant search without correction

Filtered Peptide Spectral Matches for g06611 label-free yeast run - cid35

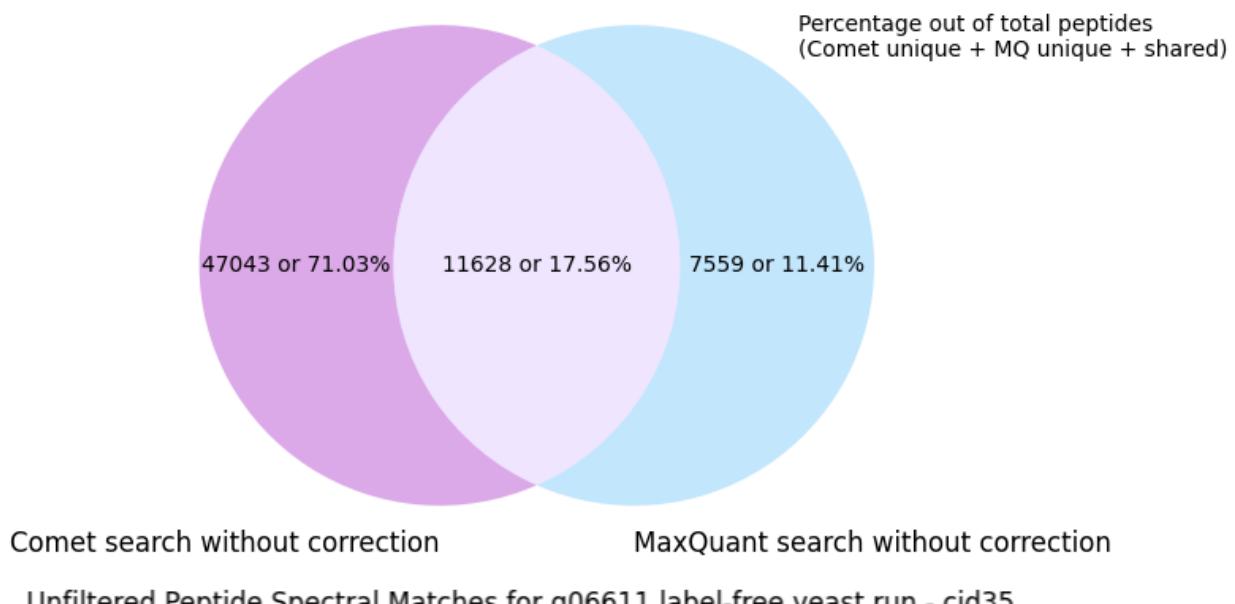


Comet search with correction MaxQuant search without correction

Appendix Figure 27: Filtered search results with or without Monocle correction for g06611.

Increased number but decreased percentage of shared peptides with filtering and correction.

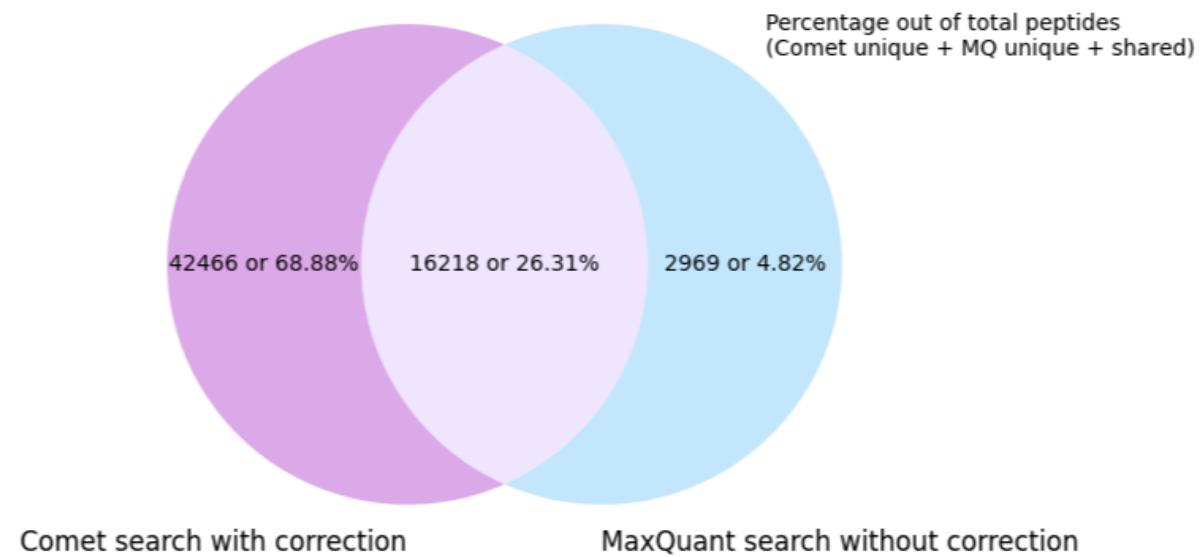
Unfiltered Peptide Spectral Matches for g06611 label-free yeast run - cid35



Comet search without correction

MaxQuant search without correction

Unfiltered Peptide Spectral Matches for g06611 label-free yeast run - cid35

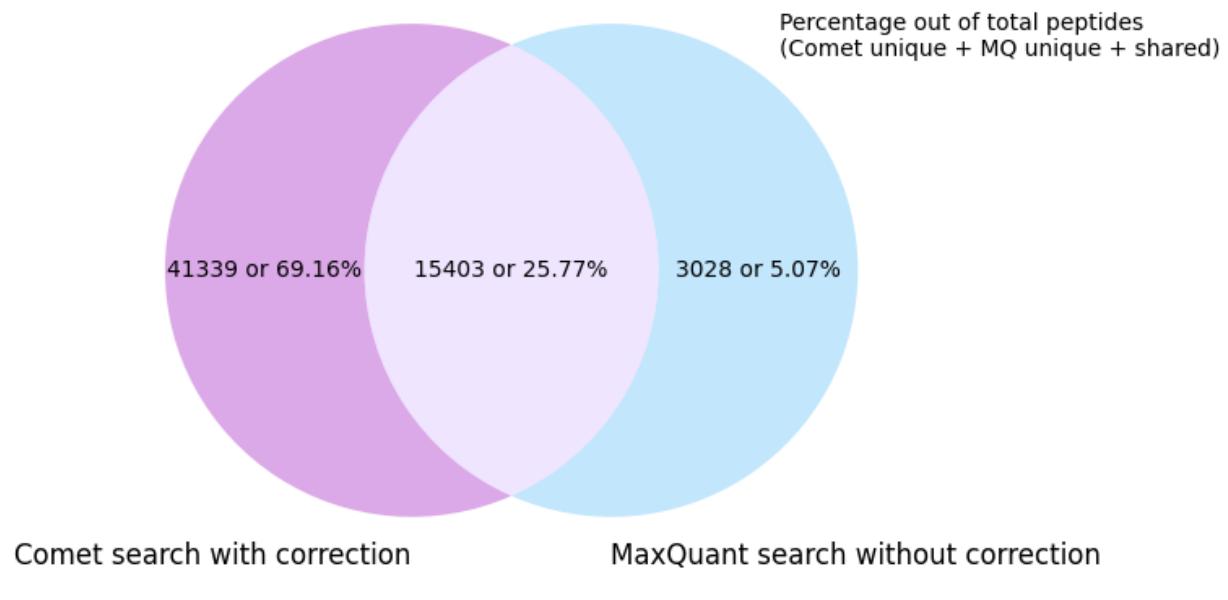


Comet search with correction

MaxQuant search without correction

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06612 label-free yeast run - cid35

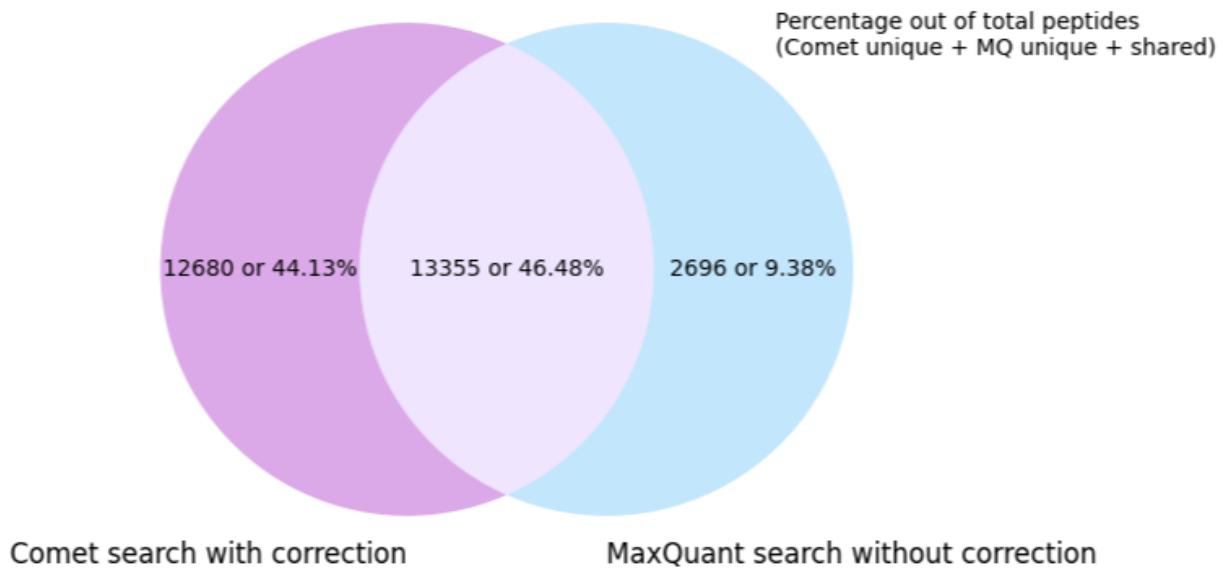


Comet search with correction

MaxQuant search without correction

Percentage out of total peptides
(Comet unique + MQ unique + shared)

Filtered Peptide Spectral Matches for g06612 label-free yeast run - cid35



Comet search with correction

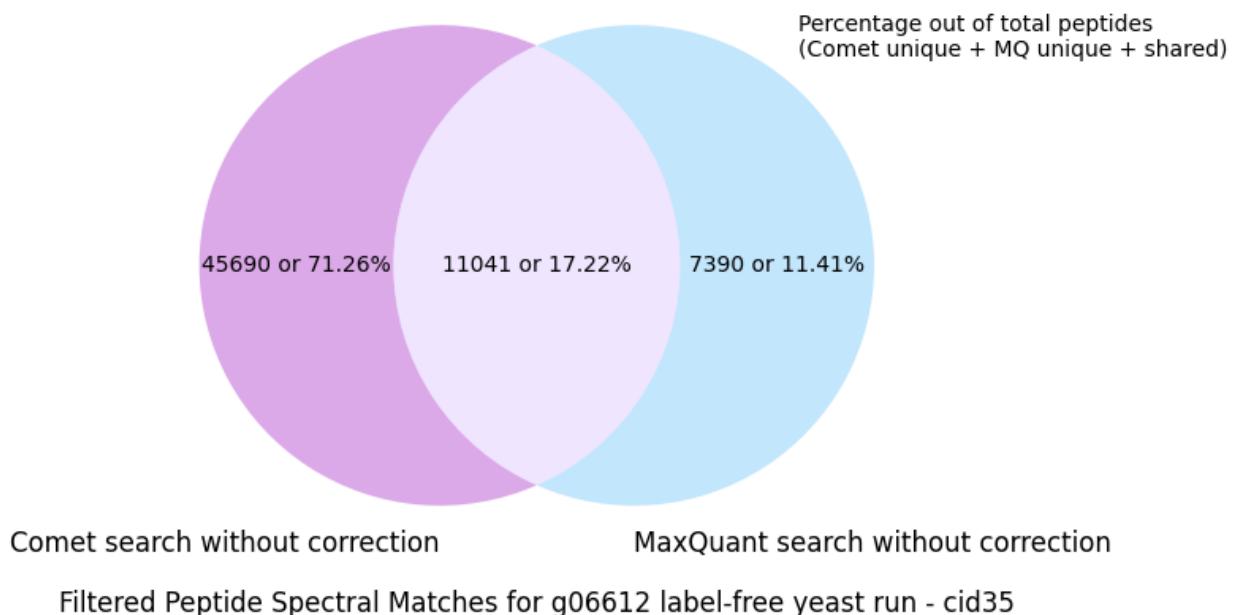
MaxQuant search without correction

Percentage out of total peptides
(Comet unique + MQ unique + shared)

Appendix Figure 29: Unfiltered vs Filtered search results with Monocle correction for g06612.

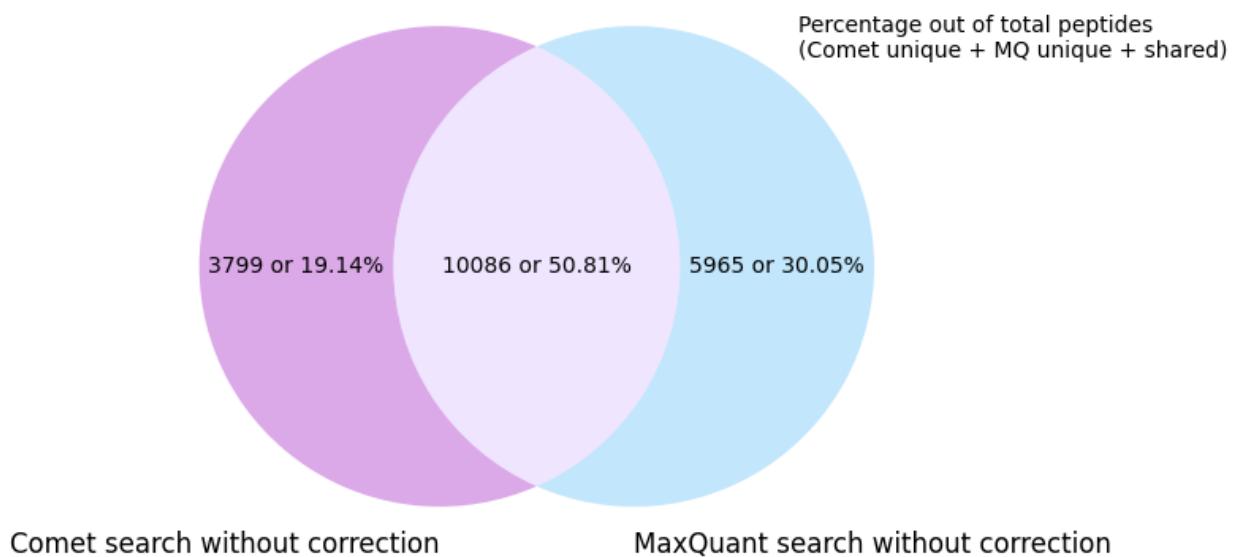
Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06612 label-free yeast run - cid35



Comet search without correction MaxQuant search without correction

Filtered Peptide Spectral Matches for g06612 label-free yeast run - cid35

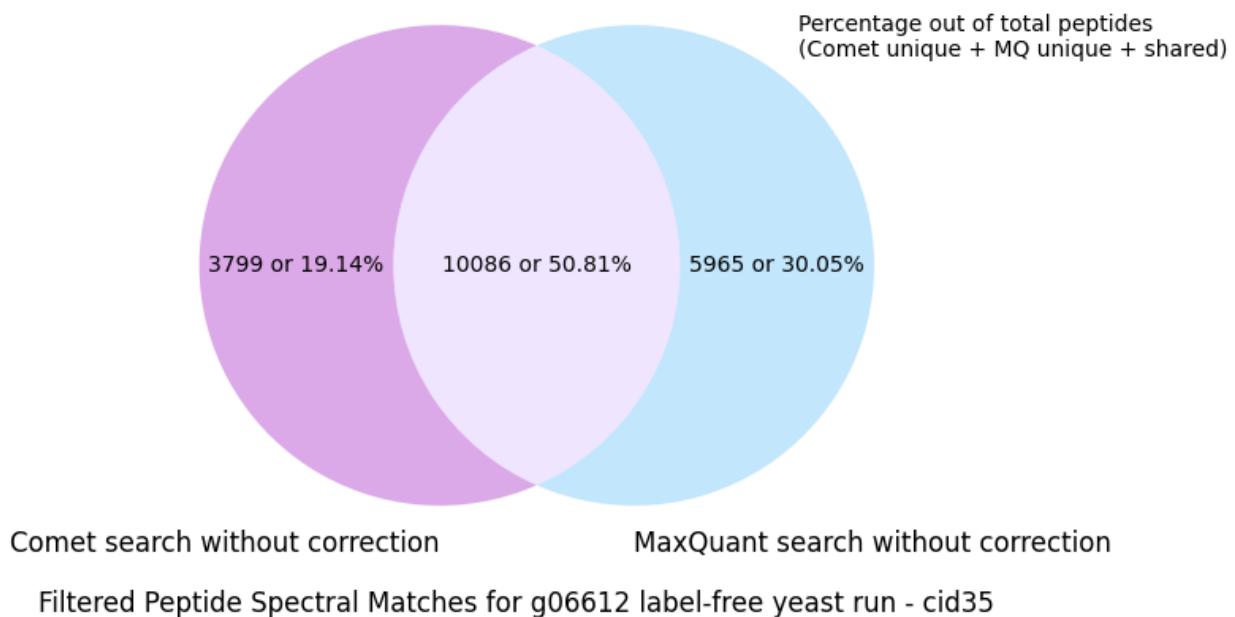


Comet search without correction

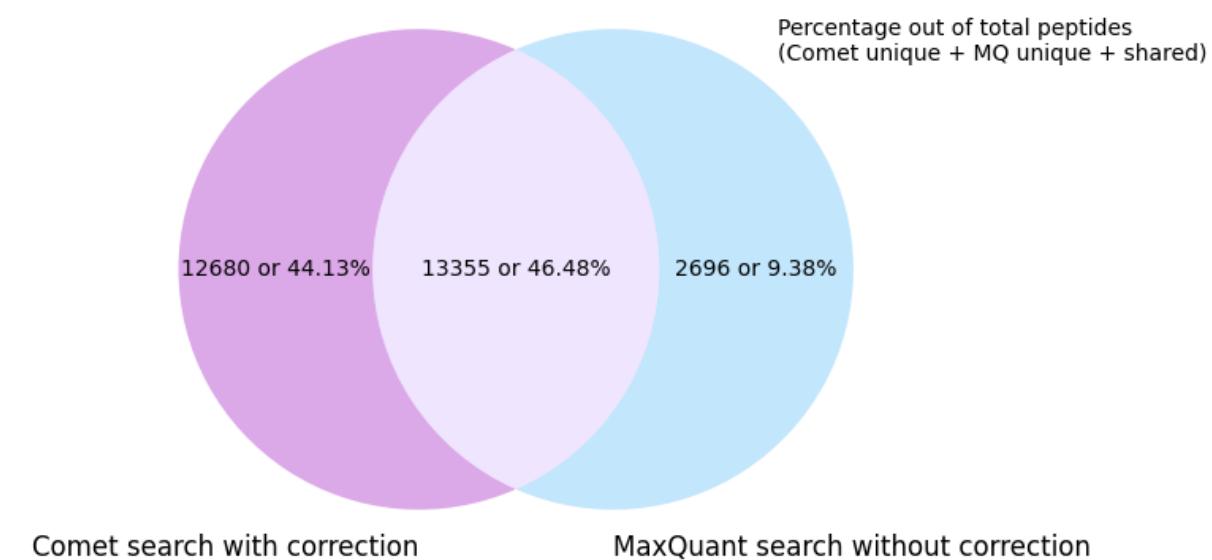
MaxQuant search without correction

Appendix Figure 30: Unfiltered vs Filtered search results without Monocle correction for g06612. Increased percentage of shared peptides with filtering.

Filtered Peptide Spectral Matches for g06612 label-free yeast run - cid35



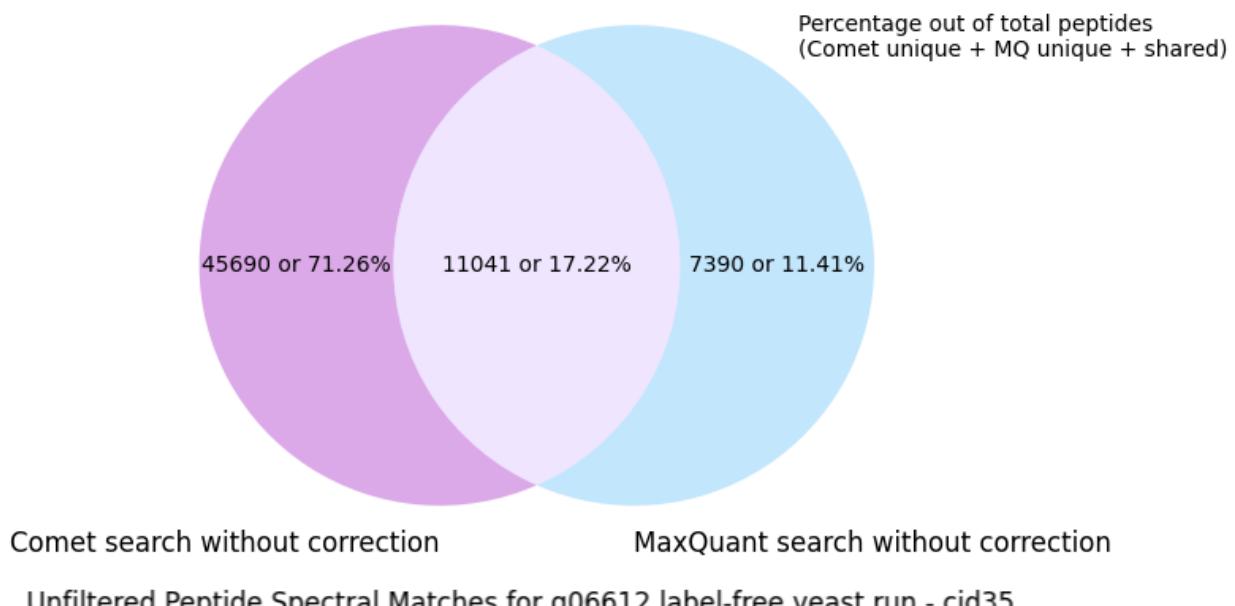
Filtered Peptide Spectral Matches for g06612 label-free yeast run - cid35



Appendix Figure 31: Filtered search results with or without Monocle correction for g06612.

Increased number but decreased percentage of shared peptides with filtering and correction.

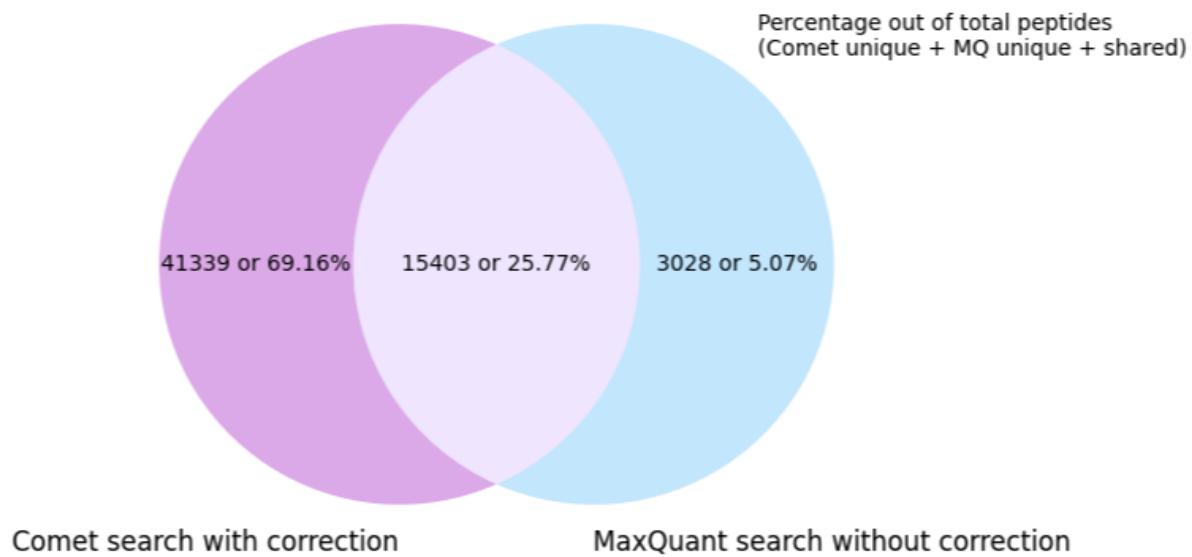
Unfiltered Peptide Spectral Matches for g06612 label-free yeast run - cid35



Comet search without correction

MaxQuant search without correction

Unfiltered Peptide Spectral Matches for g06612 label-free yeast run - cid35



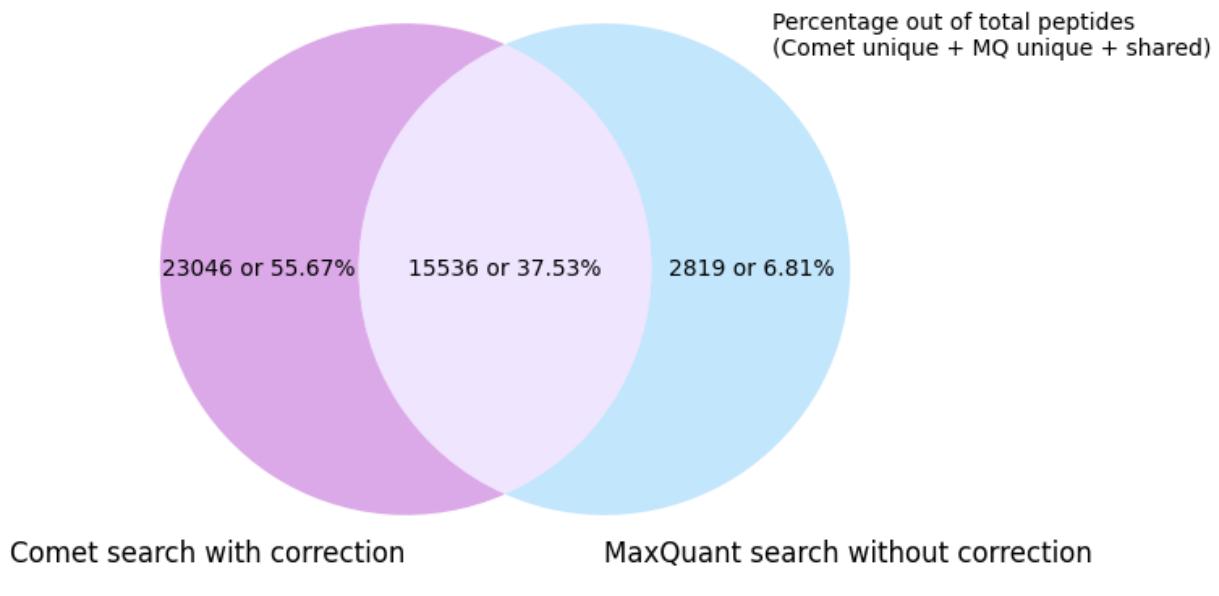
Comet search with correction

MaxQuant search without correction

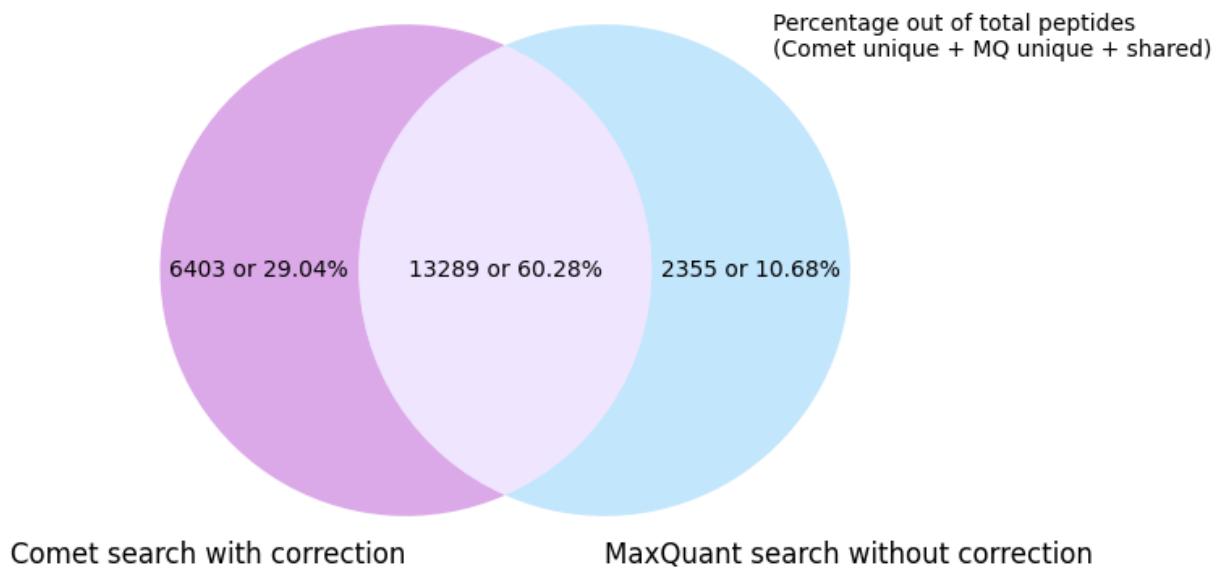
Appendix Figure 32: Unfiltered search results with or without Monocle correction for g06612.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



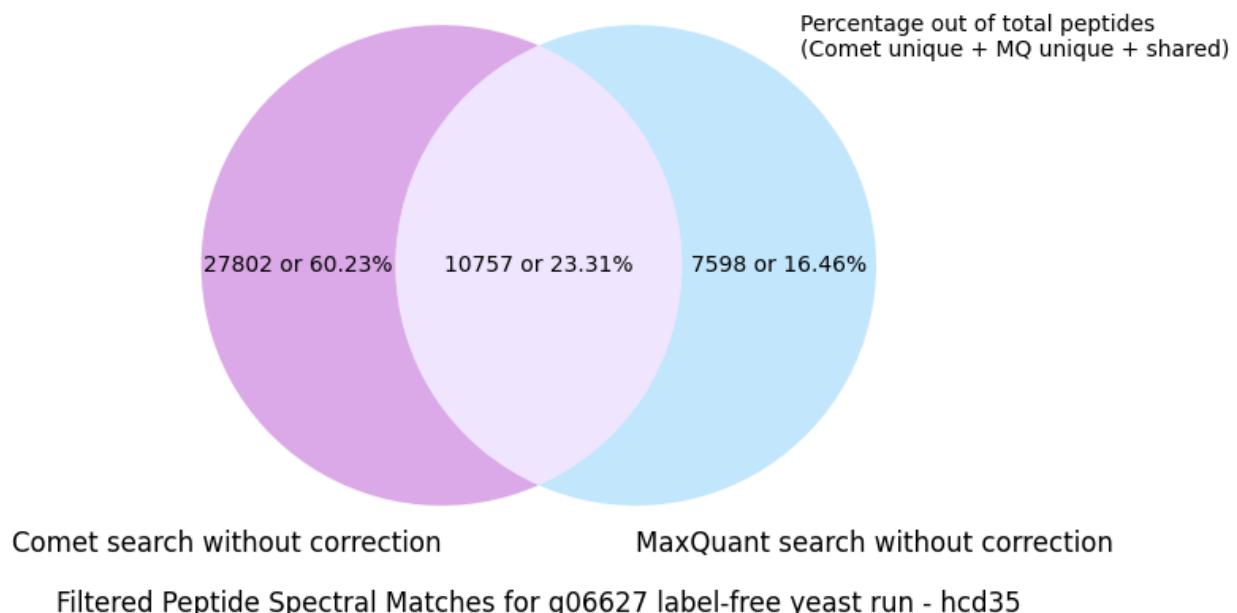
Filtered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



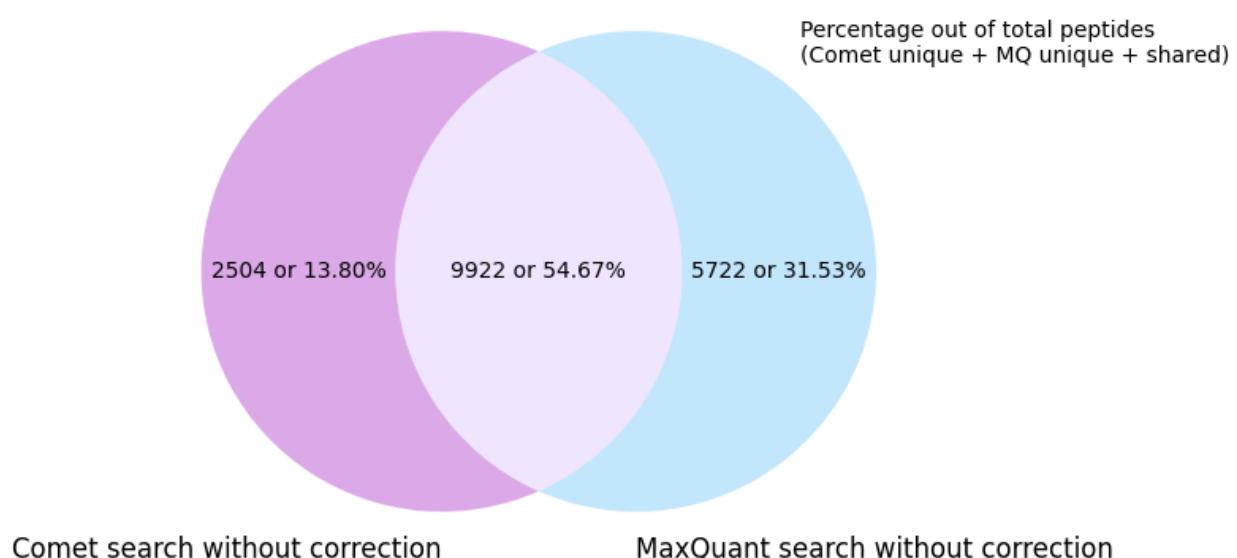
Appendix Figure 33: Unfiltered vs Filtered search results with Monocle correction for g06627.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06627 label-free yeast run - hcd35

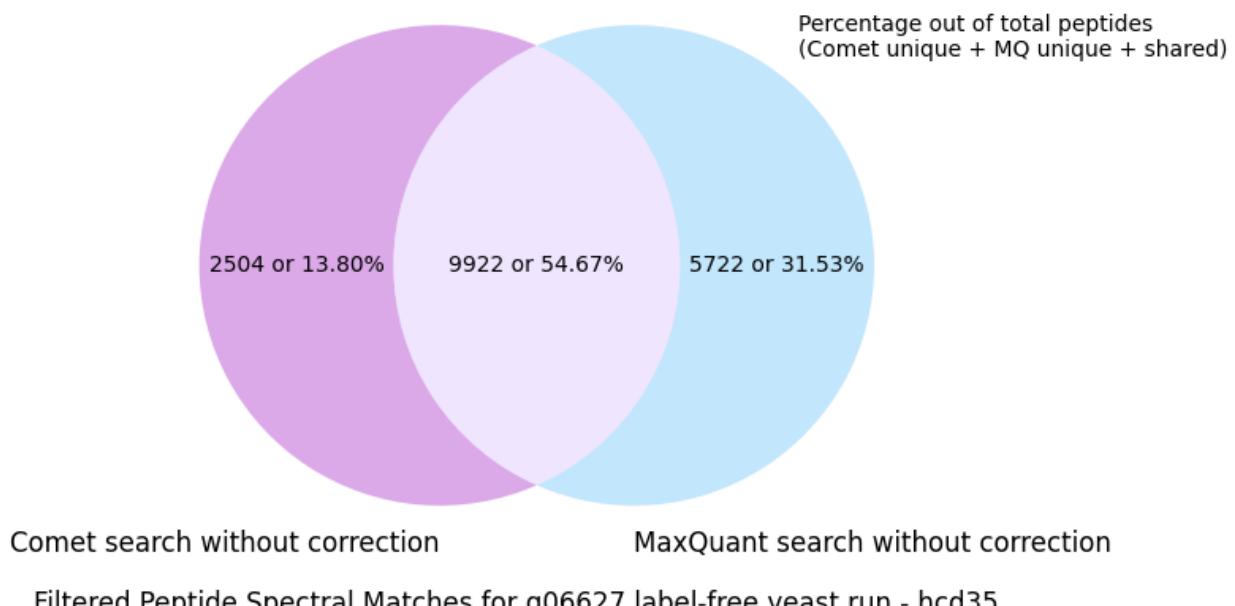


Filtered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



Appendix Figure 34: Unfiltered vs Filtered search results without Monocle correction for g06627. Increased percentage of shared peptides with filtering.

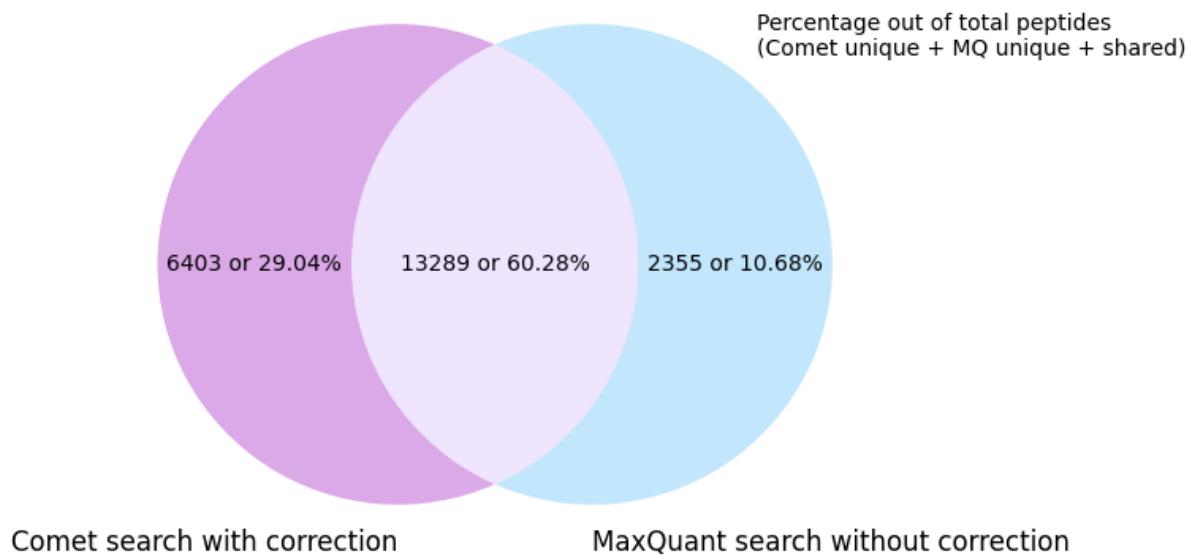
Filtered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



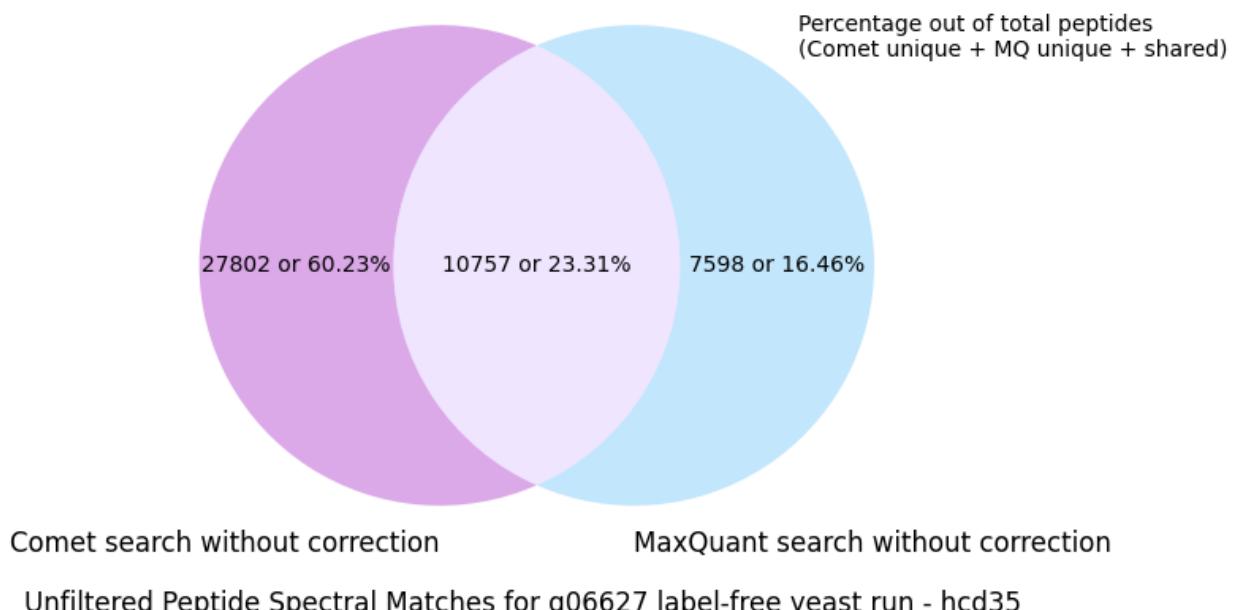
Comet search with correction

MaxQuant search without correction

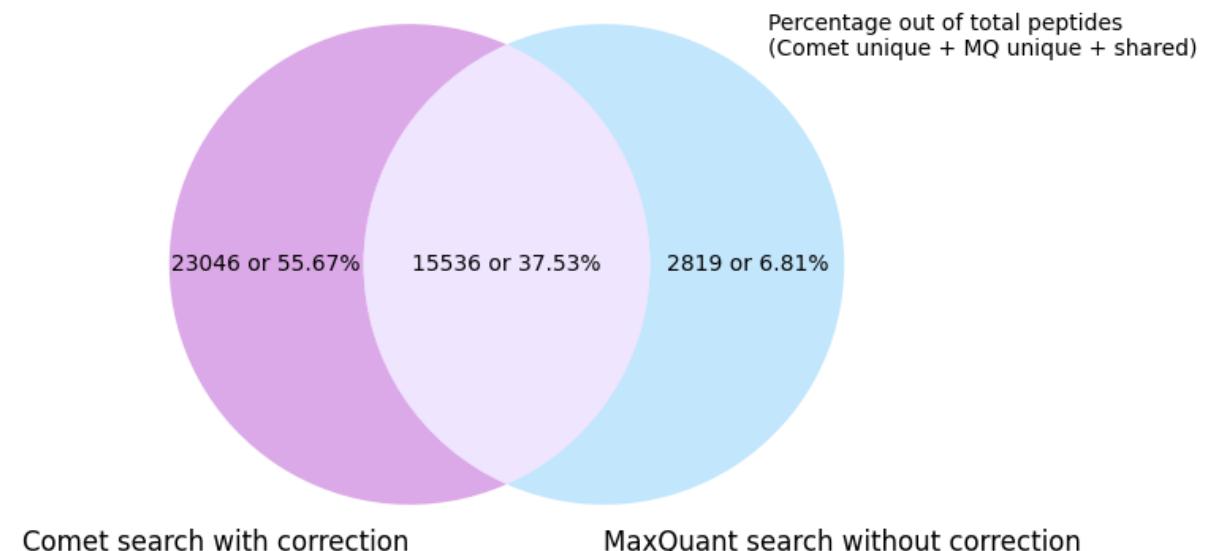
Appendix Figure 35: Filtered search results with or without Monocle correction for g06627.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



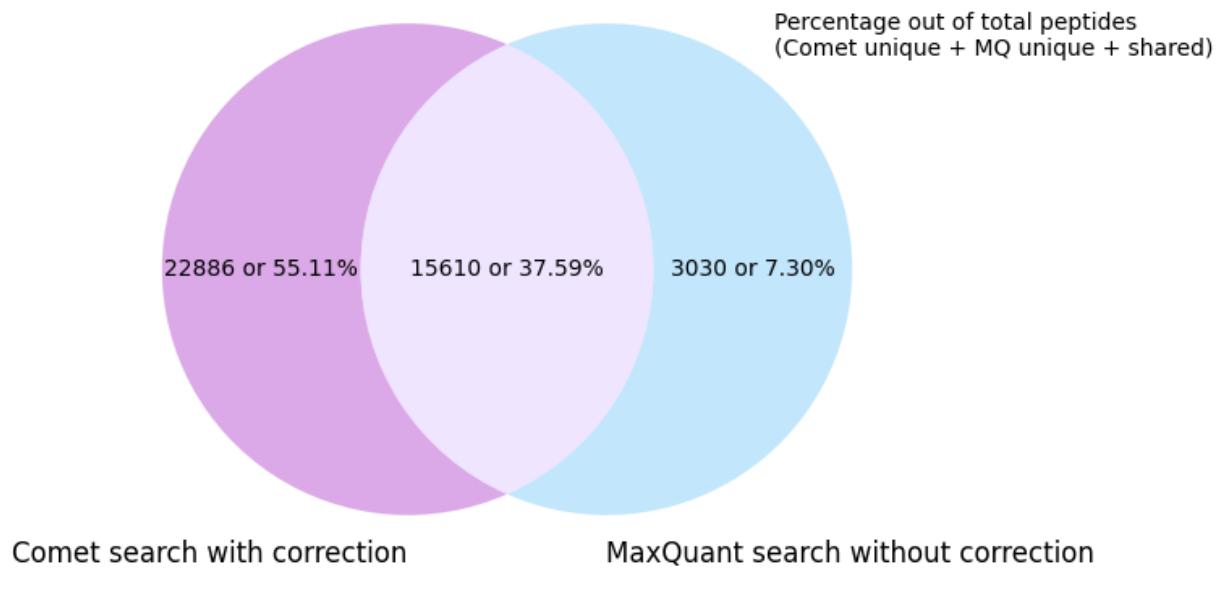
Unfiltered Peptide Spectral Matches for g06627 label-free yeast run - hcd35



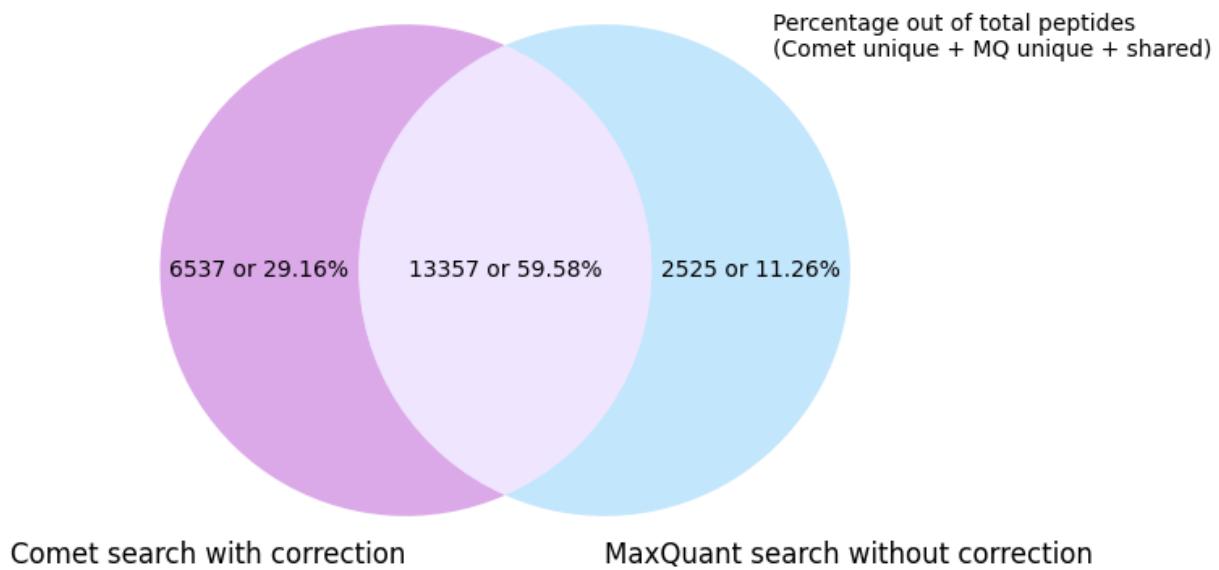
Appendix Figure 36: Unfiltered search results with or without Monocle correction for g06627.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



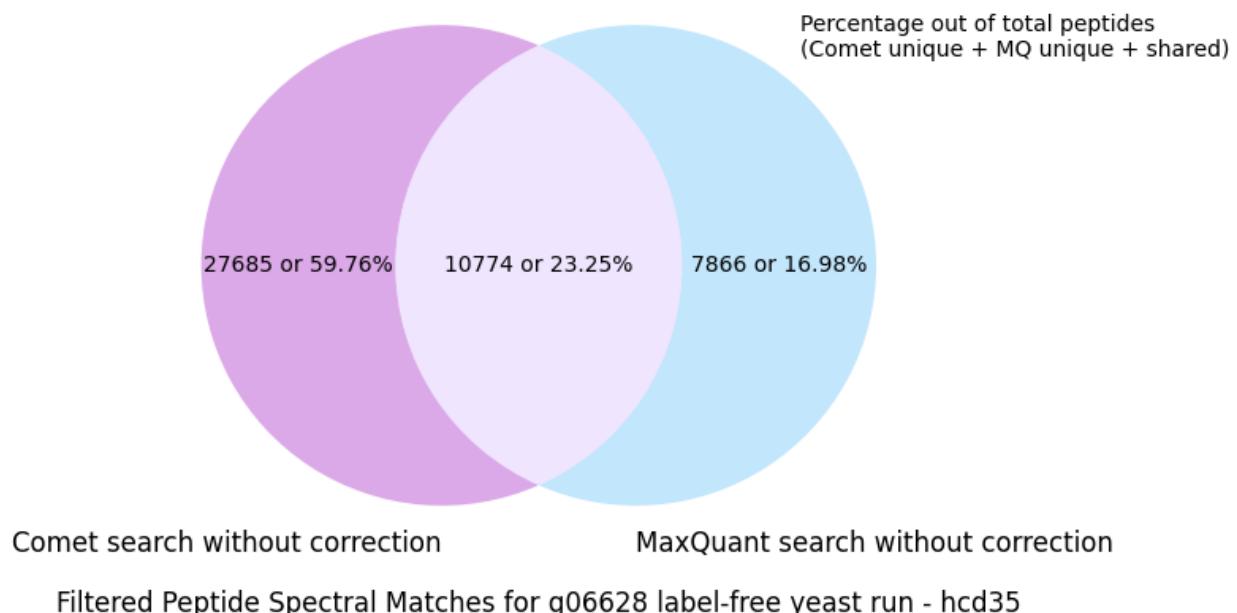
Filtered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



Appendix Figure 37: Unfiltered vs Filtered search results with Monocle correction for g06628.

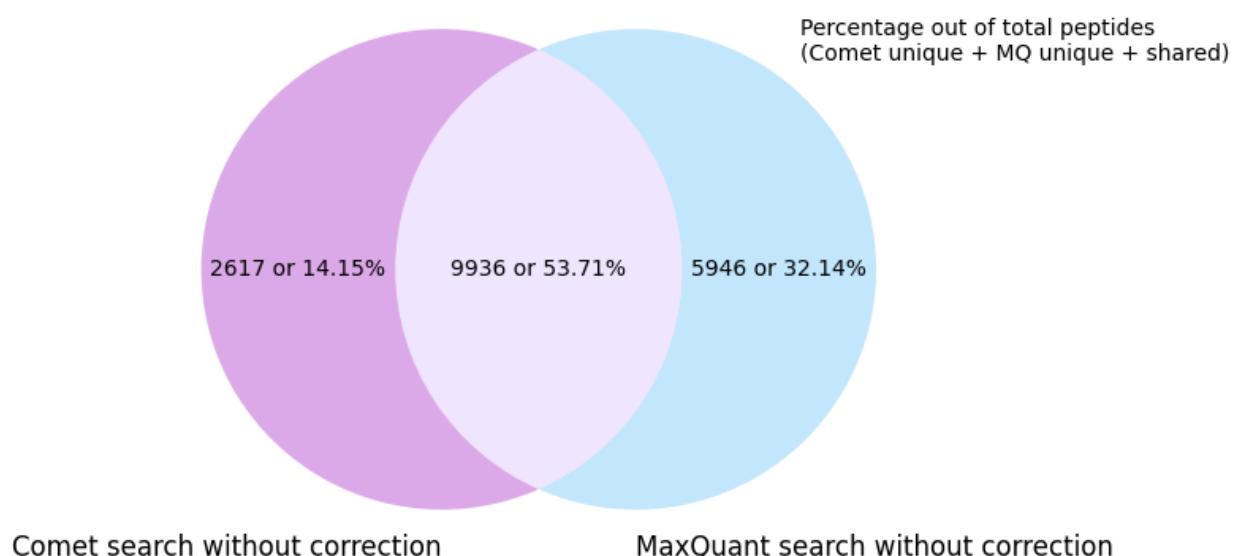
Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



Comet search without correction MaxQuant search without correction

Filtered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



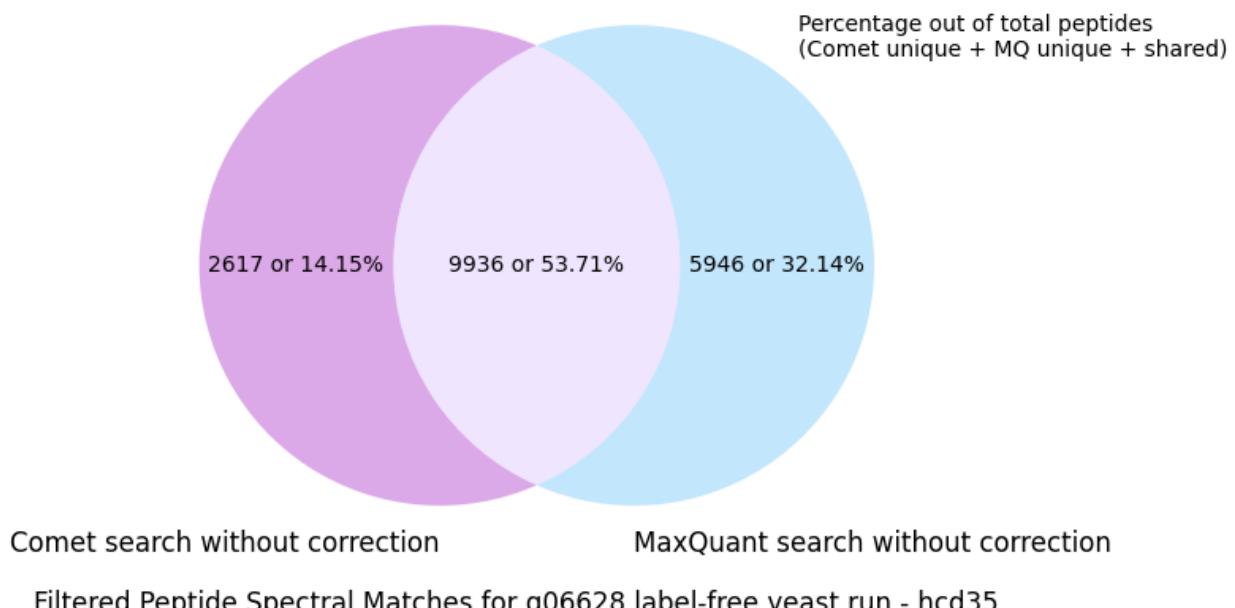
Comet search without correction

MaxQuant search without correction

Appendix figure 38: Unfiltered vs Filtered search results without Monocle correction for g06628.

Increased percentage of shared peptides with filtering.

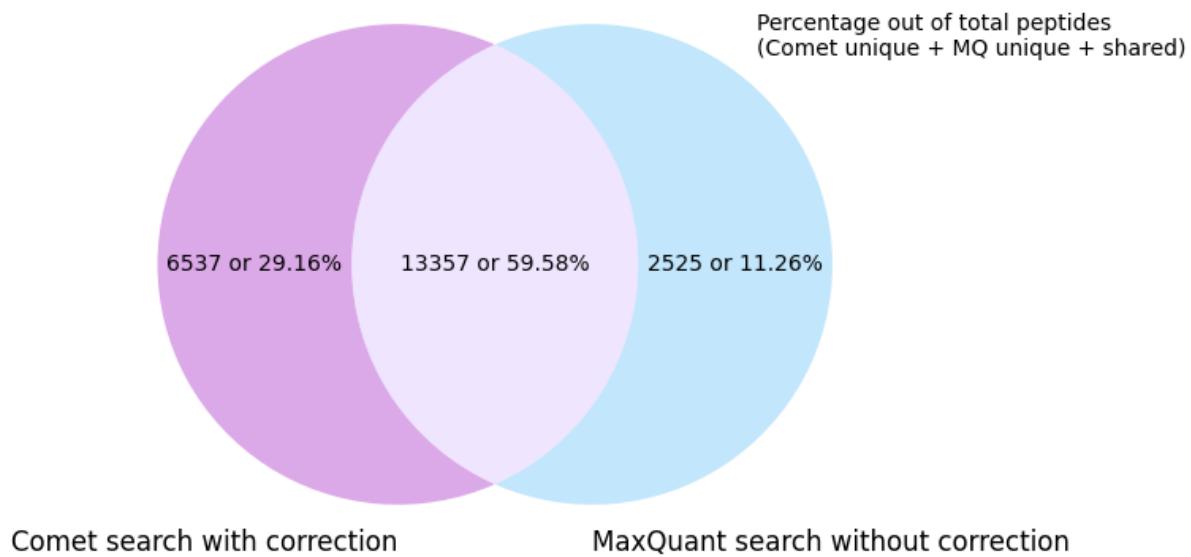
Filtered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



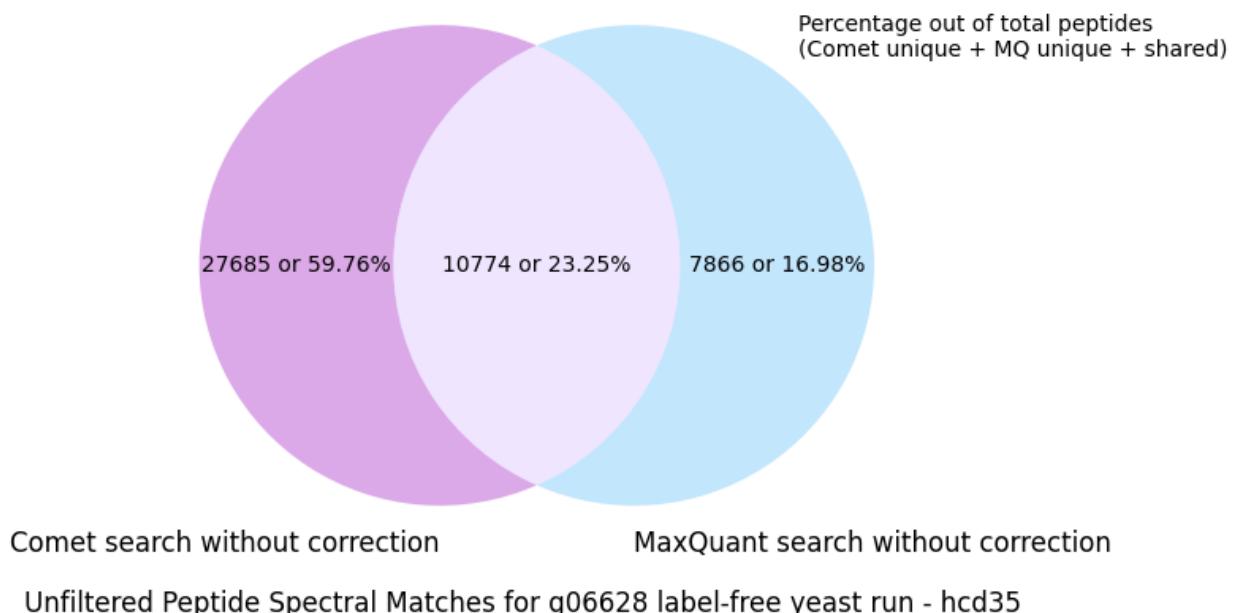
Comet search with correction

MaxQuant search without correction

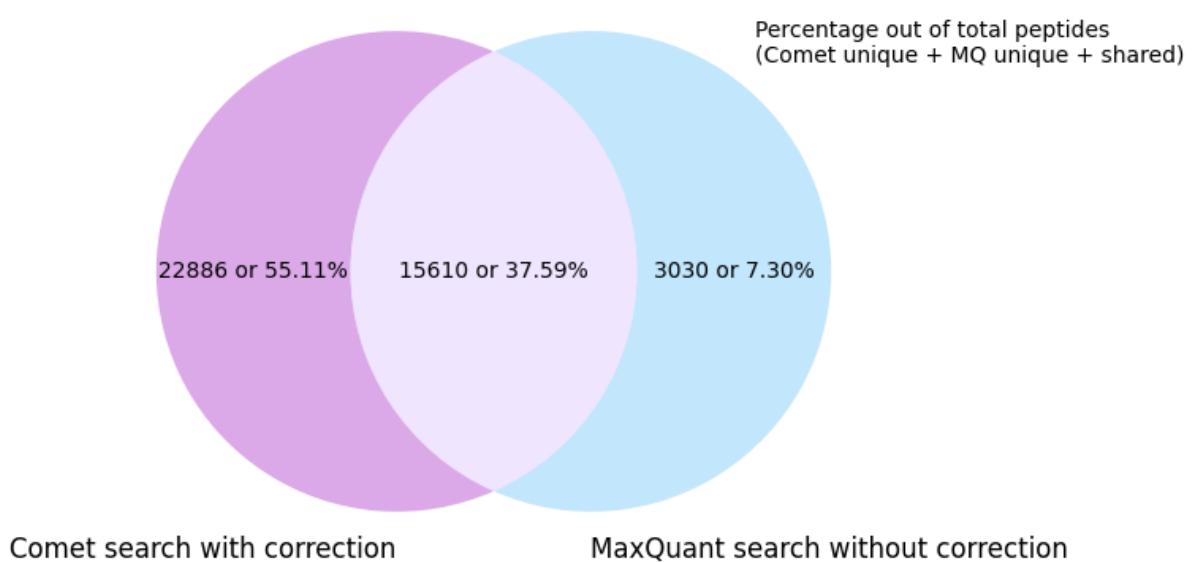
Appendix Figure 39: Filtered search results with or without Monocle correction for g06628.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



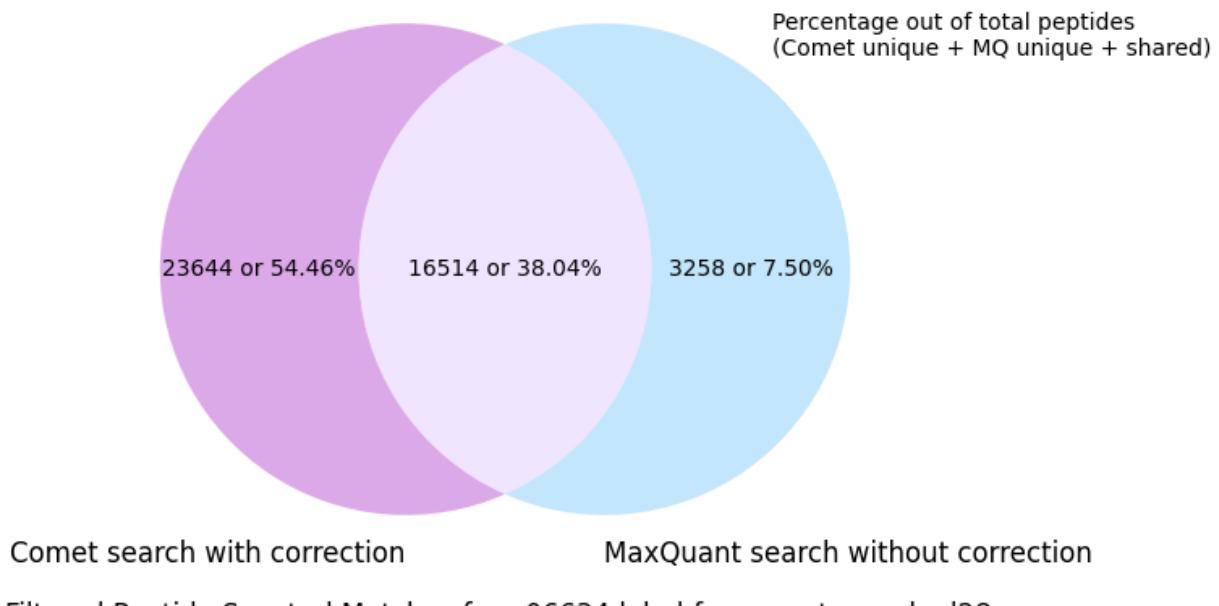
Unfiltered Peptide Spectral Matches for g06628 label-free yeast run - hcd35



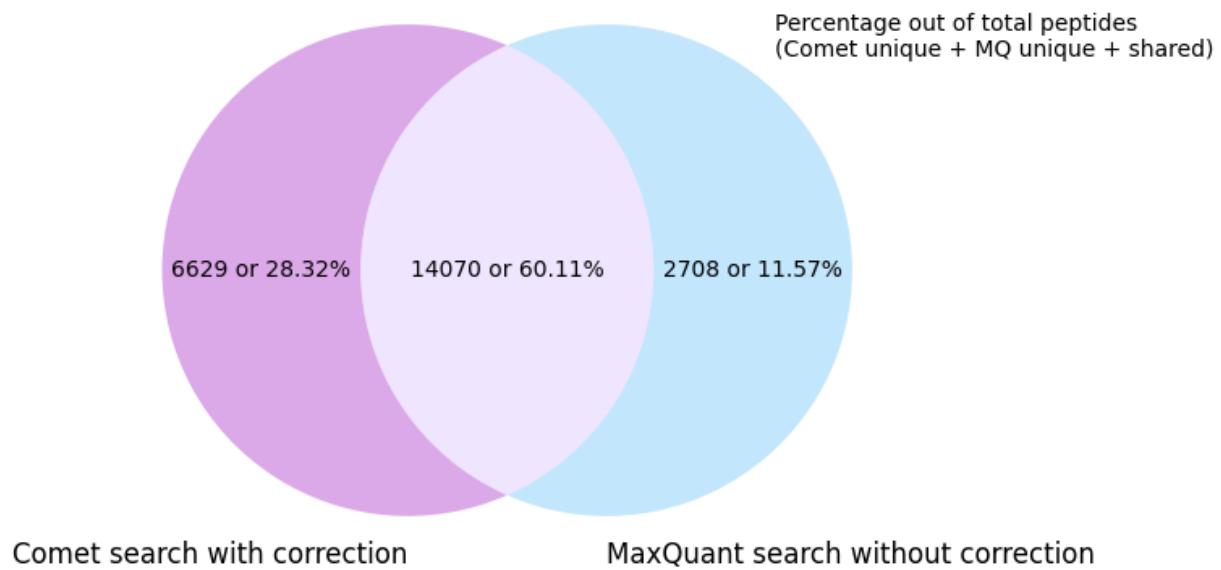
Appendix Figure 40: Unfiltered search results with or without Monocle correction for g06628.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



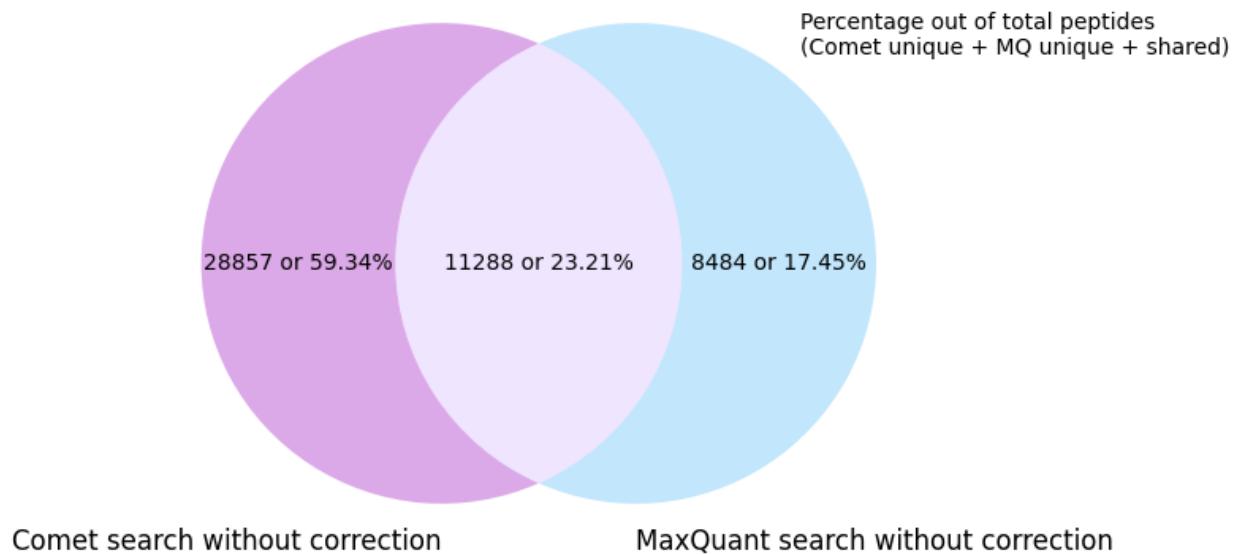
Filtered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



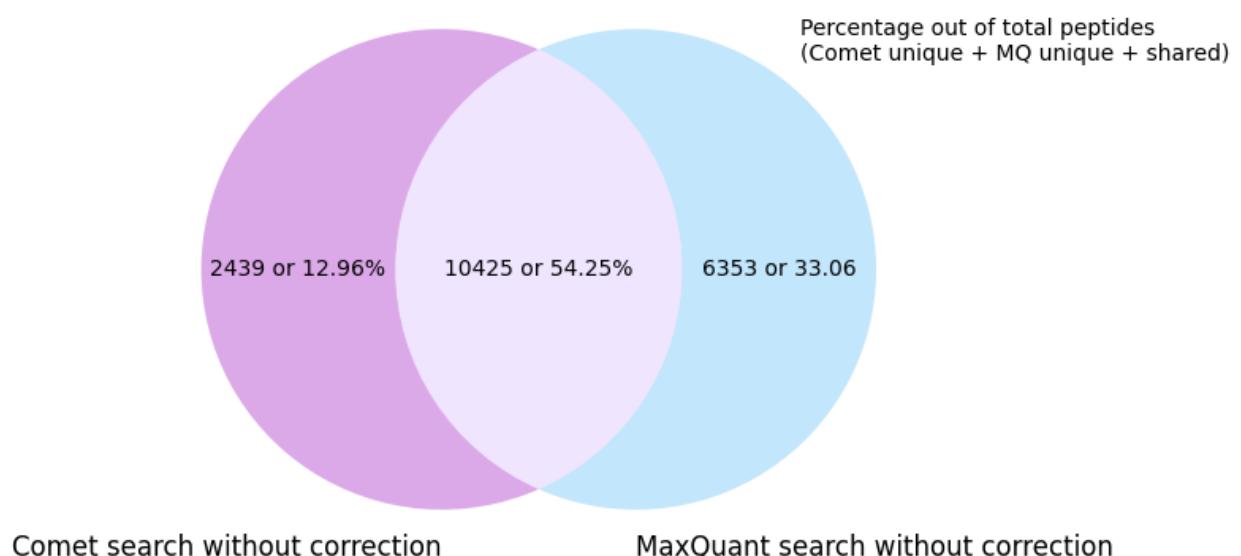
Appendix Figure 41: Unfiltered vs Filtered search results with Monocle correction for g06634.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06634 label-free yeast run - hcd28

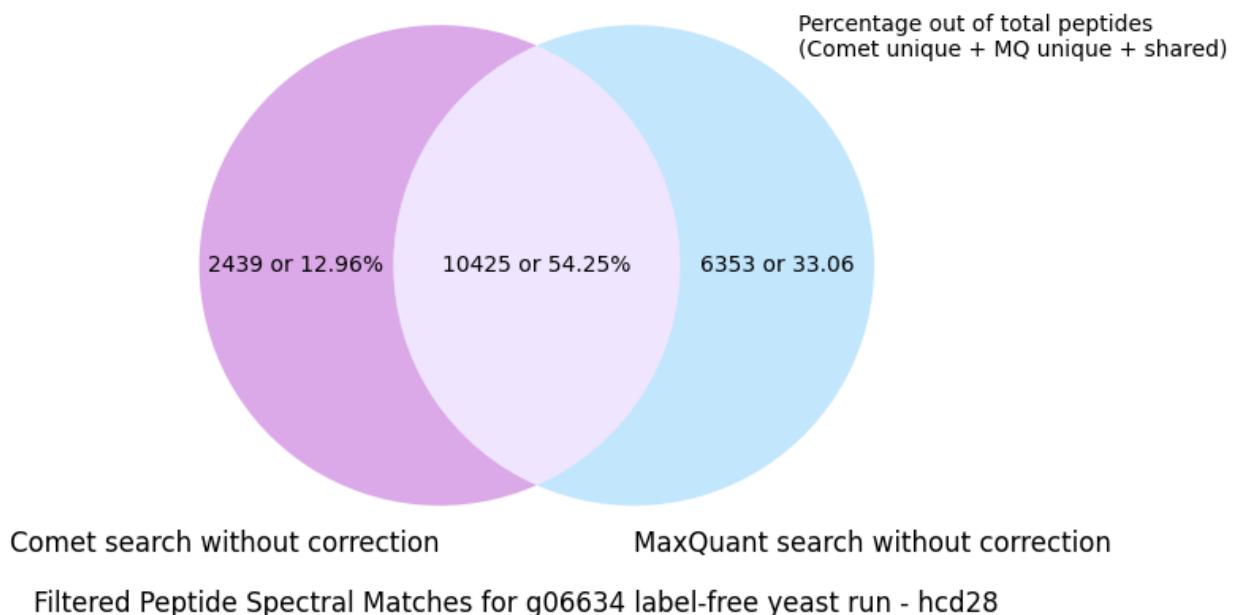


Filtered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



Appendix Figure 42: Unfiltered vs Filtered search results without Monocle correction for g06634. Increased percentage of shared peptides with filtering.

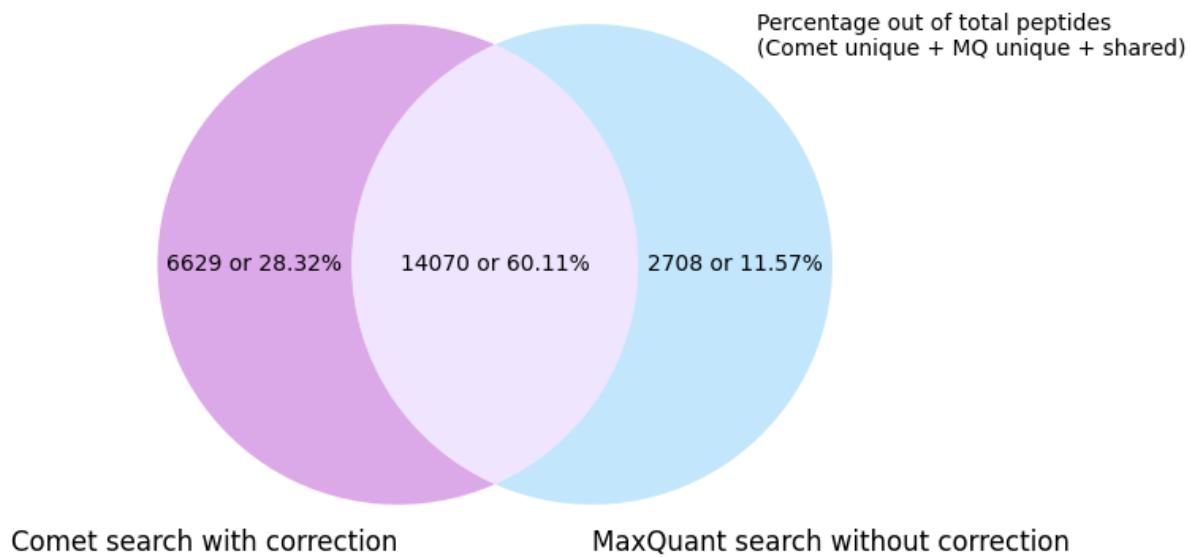
Filtered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



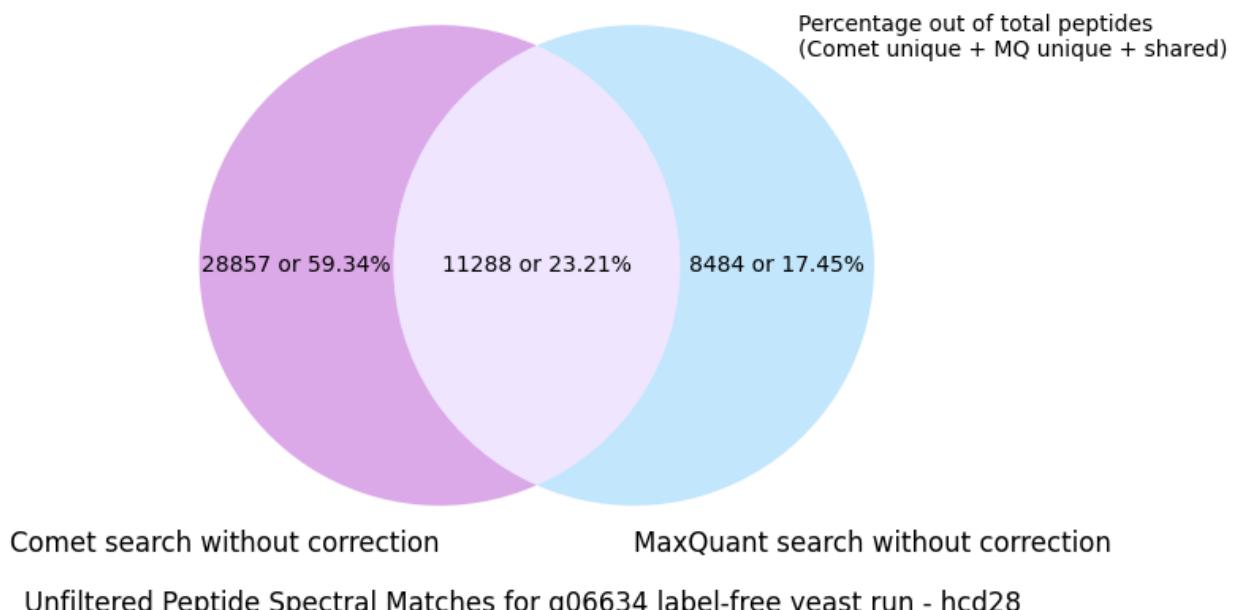
Comet search with correction

MaxQuant search without correction

Appendix Figure 43: Filtered search results with or without Monocle correction for g06634.

Increased number and percentage of shared peptides with filtering and correction.

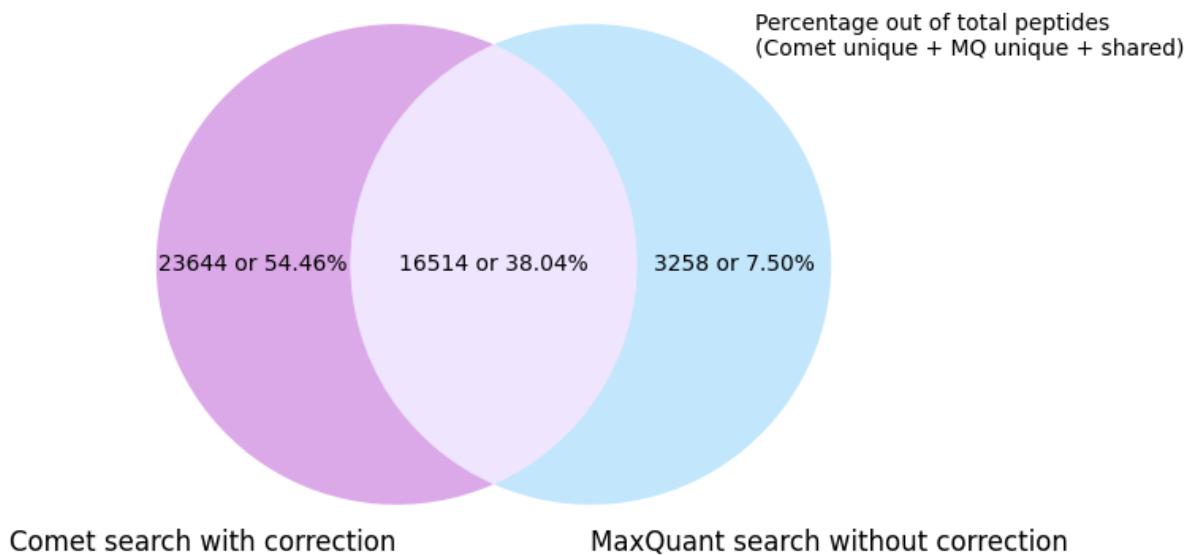
Unfiltered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



Comet search without correction

MaxQuant search without correction

Unfiltered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



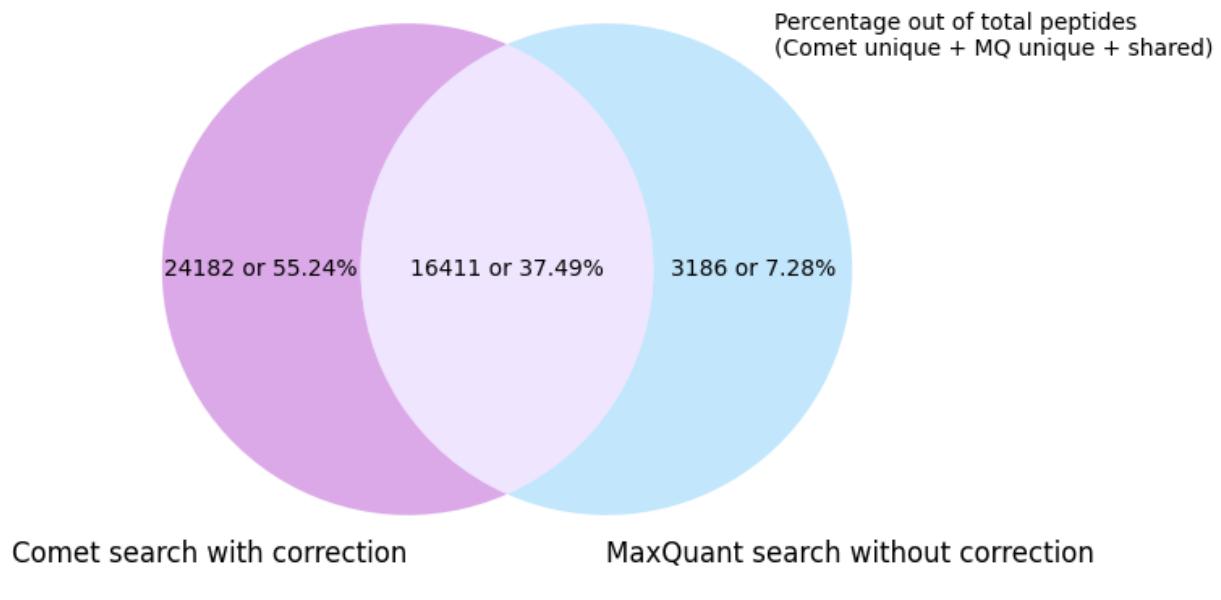
Comet search with correction

MaxQuant search without correction

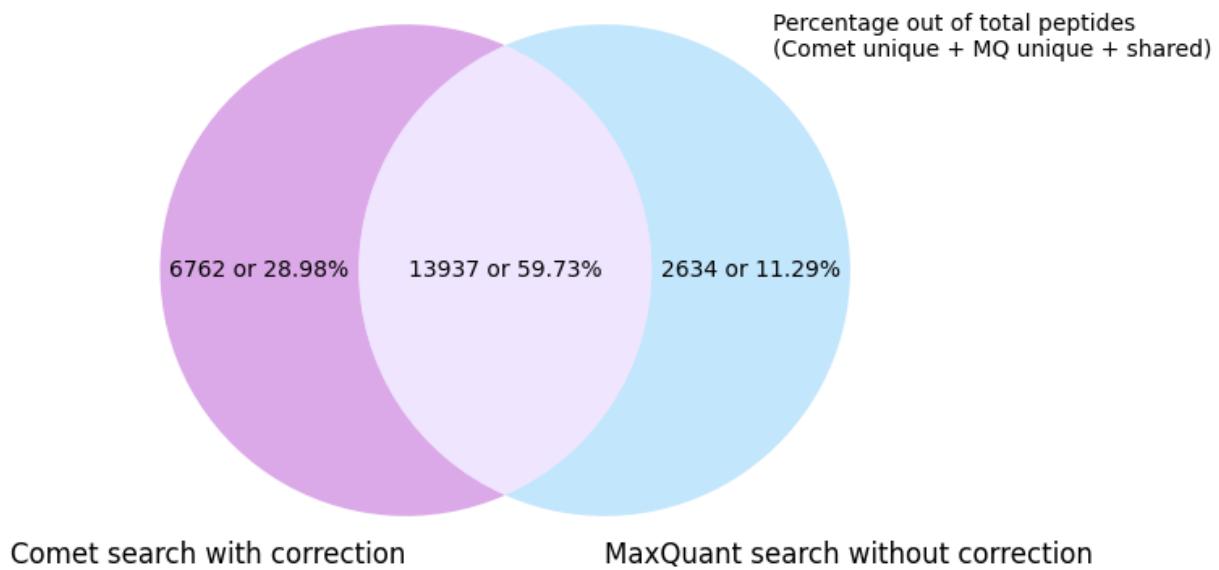
Appendix Figure 44: Unfiltered search results with or without Monocle correction for g06634.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06635 label-free yeast run - hcd28



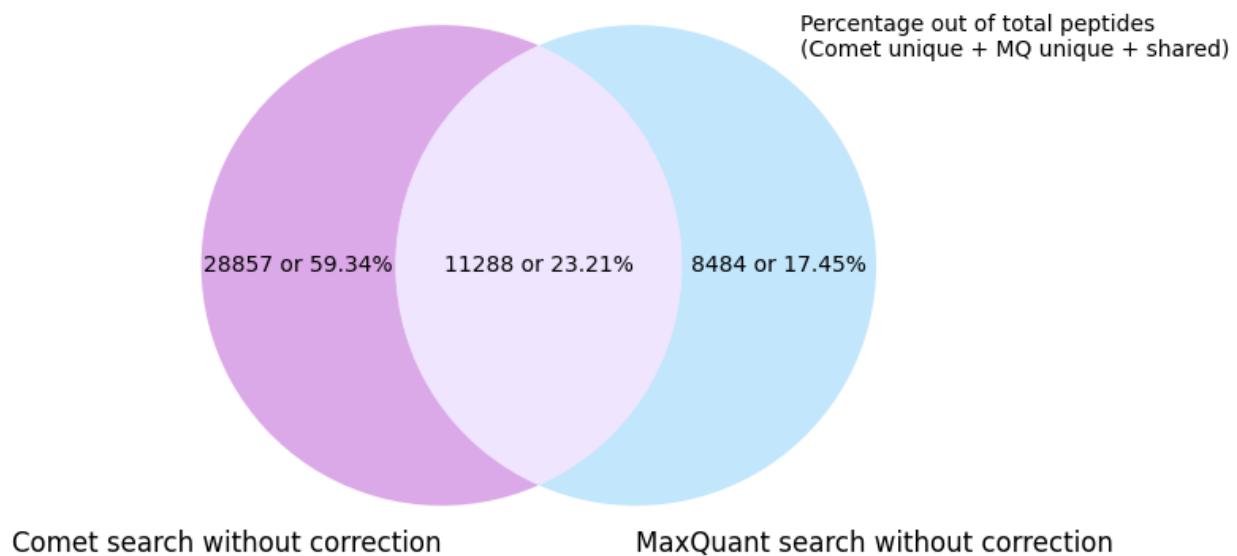
Filtered Peptide Spectral Matches for g06635 label-free yeast run - hcd28



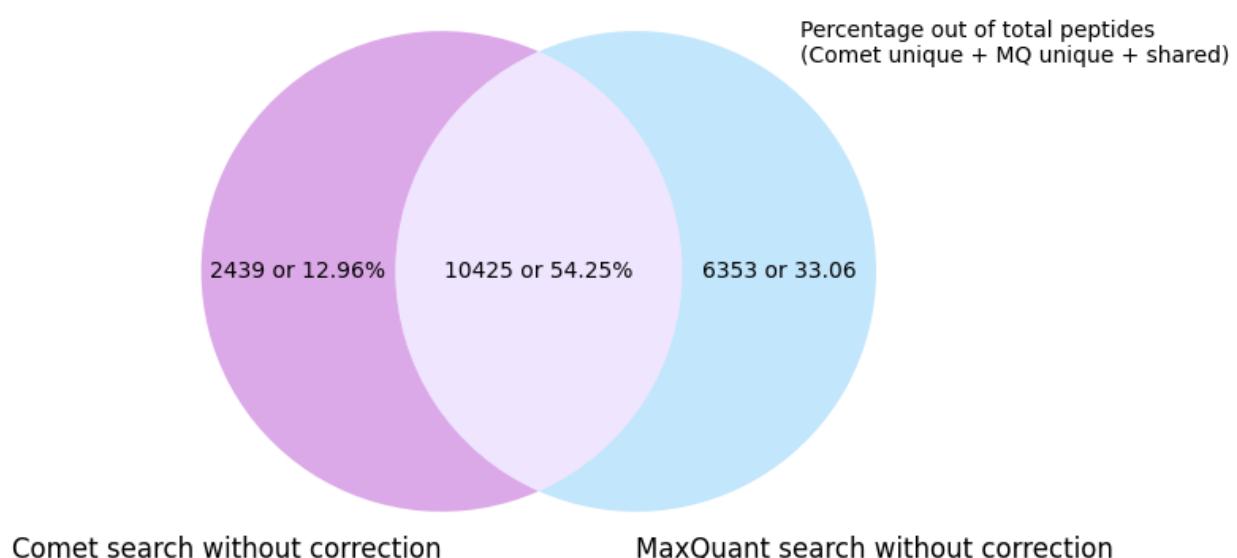
Appendix Figure 45: Unfiltered vs Filtered search results with Monocle correction for g06635.

Increased percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06634 label-free yeast run - hcd28

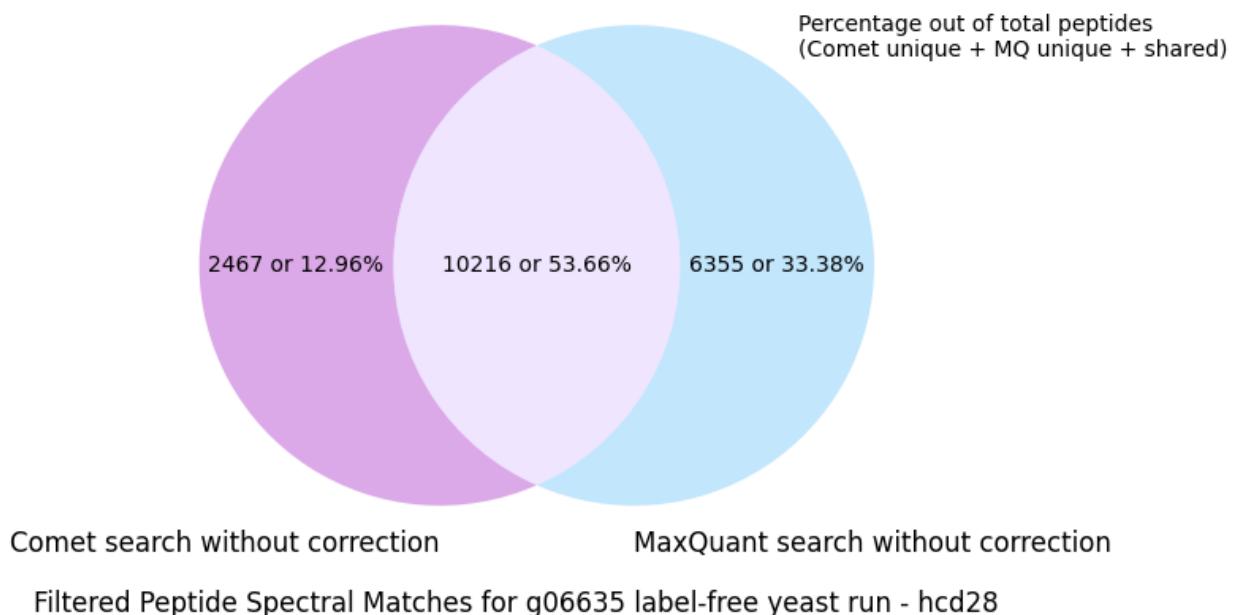


Filtered Peptide Spectral Matches for g06634 label-free yeast run - hcd28



Appendix Figure 46: Unfiltered vs Filtered search results without Monocle correction for g06635. Increased percentage of shared peptides with filtering.

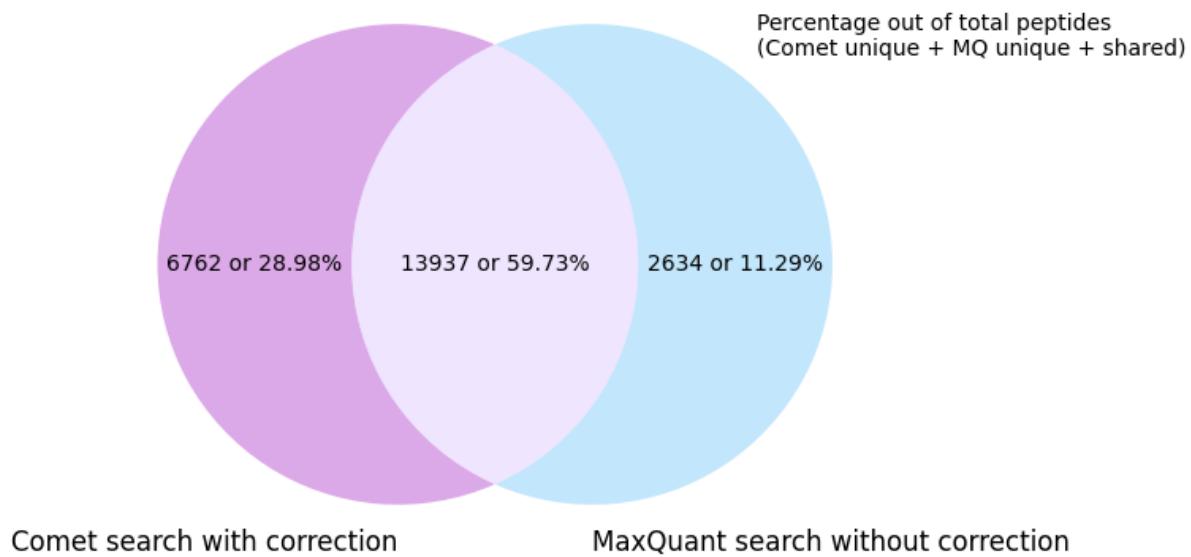
Filtered Peptide Spectral Matches for g06635 label-free yeast run - hcd28



Comet search without correction

MaxQuant search without correction

Filtered Peptide Spectral Matches for g06635 label-free yeast run - hcd28



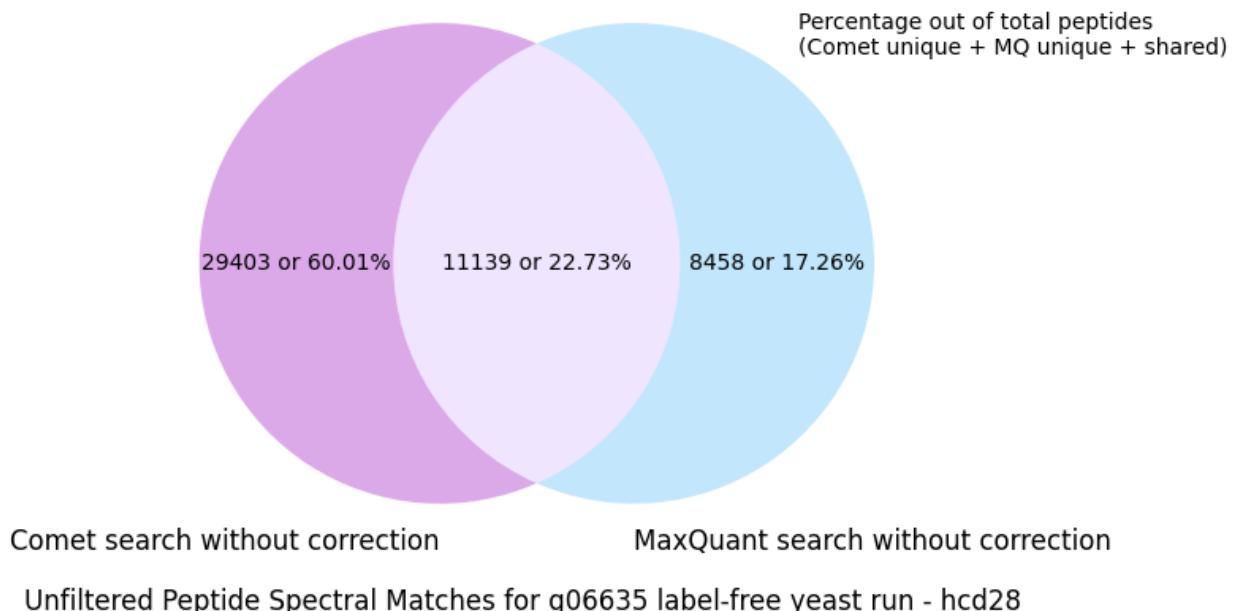
Comet search with correction

MaxQuant search without correction

Appendix Figure 47: Filtered search results with or without Monocle correction for g06635.

Increased number and percentage of shared peptides with filtering and correction.

Unfiltered Peptide Spectral Matches for g06635 label-free yeast run - hcd28

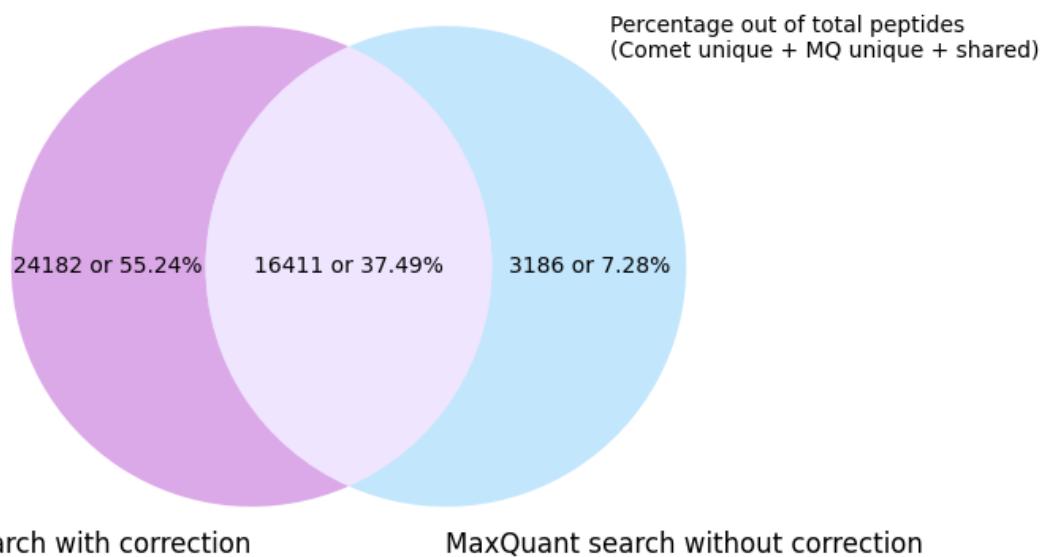


Comet search without correction MaxQuant search without correction

Comet search without correction

MaxQuant search without correction

Unfiltered Peptide Spectral Matches for g06635 label-free yeast run - hcd28



Comet search with correction

MaxQuant search without correction

Appendix Figure 48: Unfiltered search results with or without Monocle correction for g06635.

Increased number and percentage of shared peptides with filtering and correction.