



Comprehensive Mapping and Dynamics of Site-Specific Prolyl-Hydroxylation, Lysyl-Hydroxylation and Lysyl O-Glycosylation of Collagens Deposited in ECM During Zebrafish Heart Regeneration

Vivek Sarohi^{1,2}, Shriya Srivastava¹ and Trayambak Basak^{1,2*}

¹School of Biosciences and Bioengineering (BSBE), Indian Institute of Technology (IIT)- Mandi, Mandi, India, ²BioX Center, IIT-Mandi, Mandi, India

OPEN ACCESS

Edited by:

Qi Wu,
Aarhus University, Denmark

Reviewed by:

Luigi Scietti,
European Institute of Oncology (IEO),
Italy

Alexandra Naba,
University of Illinois at Chicago,
United States

*Correspondence:

Trayambak Basak
trayambak@iitmandi.ac.in

Specialty section:

This article was submitted to
Molecular Diagnostics and
Therapeutics,
a section of the journal
Frontiers in Molecular Biosciences

Received: 09 March 2022

Accepted: 09 May 2022

Published: 16 June 2022

Citation:

Sarohi V, Srivastava S and Basak T
(2022) Comprehensive Mapping and
Dynamics of Site-Specific Prolyl-
Hydroxylation, Lysyl-Hydroxylation and
Lysyl O-Glycosylation of Collagens
Deposited in ECM During Zebrafish
Heart Regeneration.
Front. Mol. Biosci. 9:892763.
doi: 10.3389/fmolb.2022.892763

Cardiac fibrosis-mediated heart failure (HF) is one of the major forms of end-stage cardiovascular diseases (CVDs). Cardiac fibrosis is an adaptive response of the myocardium upon any insult/injury. Excessive deposition of collagen molecules in the extracellular matrix (ECM) is the hallmark of fibrosis. This fibrotic response initially protects the myocardium from ventricular rupture. Although in mammals this fibrotic response progresses towards scar-tissue formation leading to HF, some fishes and urodeles have mastered the art of cardiac regeneration following injury-mediated fibrotic response. Zebrafish have a unique capability to regenerate the myocardium after post-amputation injury. Following post-amputation, the ECM of the zebrafish heart undergoes extensive remodeling and deposition of collagen. Being the most abundant protein of ECM, collagen plays important role in the assembly and cell-matrix interactions. However, the mechanism of ECM remodeling is not well understood. Collagen molecules undergo heavy post-translational modifications (PTMs) mainly hydroxylation of proline, lysine, and glycosylation of lysine during biosynthesis. The critical roles of these PTMs are emerging in several diseases, embryonic development, cell behavior regulation, and cell-matrix interactions. The site-specific identification of these collagen PTMs in zebrafish heart ECM is not known. As these highly modified peptides are not amenable to mass spectrometry (MS), the site-specific identification of these collagen PTMs is challenging. Here, we have implemented our in-house proteomics analytical pipeline to analyze two ECM proteomics datasets (PXD011627, PXD010092) of the zebrafish heart during regeneration (post-amputation). We report the first comprehensive site-specific collagen PTM map of zebrafish heart ECM. We have identified a total of 36 collagen chains (19 are reported for the first time here) harboring a total of 95 prolyl-3-hydroxylation, 108 hydroxylysine, 29 galactosyl-hydroxylysine, and 128 glucosylgalactosyl-hydroxylysine sites. Furthermore, we comprehensively map the three chains (COL1A1a, COL1A1b,

and COL1A2) of collagen I, the most abundant protein in zebrafish heart ECM. We achieved more than 95% sequence coverage for all the three chains of collagen I. Our analysis also revealed the dynamics of prolyl-3-hydroxylation occupancy oscillations during heart regeneration at these sites. Moreover, quantitative site-specific analysis of lysine-O-glycosylation microheterogeneity during heart regeneration revealed a significant ($p < 0.05$) elevation of site-specific (K^{1017}) glucosylgalactosyl-hydroxylysine on the col1a1a chain. Taken together, these site-specific PTM maps and the dynamic changes of site-specific collagen PTMs in ECM during heart regeneration will open up new avenues to decode ECM remodeling and may lay the foundation to tinker the cardiac regeneration process with new approaches.

Keywords: collagens, post-translational modifications, O-glycosylation, mass-spectrometry, microheterogeneity, extra-cellular matrix, cardiac regeneration

1 INTRODUCTION

Cardiac fibrosis leading to heart failure has remained one of the major causes of death worldwide (González et al., 2018). Extracellular matrix (ECM) remodeling in the cardiac tissues is one of the main mechanisms of fibrosis development to an injury/insult in the myocardium (Li et al., 2018). Excessive deposition of collagen molecules in the ECM during cardiac fibrosis has remained the hallmark of fibrosis (Cowling et al., 2019). Initially, this fibrotic response helps in preventing the rupture of the ventricular wall of the heart (Cowling et al., 2019). However, in the mammalian hearts, this persistent fibrotic response progresses towards scar tissue formation, tissue-stiffening and ultimately resulting in heart failure (González et al., 2018). Contrastingly, in nature, some non-mammalian vertebrates such as zebrafish and urodele amphibians (Cutie and Huang, 2021) have mastered a unique exercise that allows their cardiomyocytes to undergo restricted dedifferentiation achieving cardiac regeneration. This exceptional property of adult zebrafish permits heart regeneration in case of a severe tissue insult or experimental amputation (Cutie and Huang, 2021). This unique ability makes Zebrafish an appropriate model system to study complex cellular and molecular processes underlying cardiac regeneration (Bootorabi et al., 2017). Under chronic/acute injury conditions, excess collagens and other proteins get deposited in the myocardium ECM inducing fibrosis (Travers et al., 2016) and eventually causing systolic and diastolic dysfunction. In the case of zebrafish, the fibrosis regression that has been documented as a transient phase occurs through fibroblast inactivation (Sánchez-Iranzo et al., 2018). The crucial understanding of participation and contribution of zebrafish heart ECM remains majorly unexplored. The ECM has recently gained attention to probe a better insight into the mechanistic changes during heart regeneration in zebrafish. In fact, the zebrafish heart ECM has been shown to have the potency to induce mammalian heart regeneration (Chen et al., 2016). Although the role of a few ECM components such as hyaluronic acid has been reported to play a major role in the epithelial-to-mesenchymal transition during regeneration (Missinato et al., 2015), the roles of major components of ECM such as collagens, glycoproteins, and proteoglycans are not well understood.

Collagens and other cardiac ECM proteins are not just responsible for maintaining the architecture of the tissue but also play a vital role in mechanical and biochemical interactions that determine different cell behavior and functions (Frangogiannis, 2019).

In the ECM, collagens are the most abundant protein constituent (Di Lullo et al., 2002). Nauroy *et al.* have predicted 58 collagen genes contributing towards the zebrafish matrisome (Nauroy et al., 2018). Collagens are responsible for maintaining the structure and functions of cardiac tissues. These are also crucial in wound healing process (Willyard, 2018). Recently, ablation of COL1A2 in the zebrafish heart has been shown to impair cardiomyocyte regeneration (Sánchez-Iranzo et al., 2018). Other collagen chains such as COL7A11 and COL8A2 present in zebrafish heart ECM have been found upregulated during heart regeneration (Sánchez-Iranzo et al., 2018). Collagens are heavily post-translationally modified triple-helical molecules that form the interstitial fibers of the myocardium. Collagen post-translational modifications (PTMs) play a significant role in the function and stability of the triple helix. The proline residues are the most abundant amino acids found in collagen chains (Myllyharju, 2005). Prolines present in collagen chains are 4-hydroxylated (4-HyP) generally in the “Y” position of the “Gly-Xaa-Yaa” motif, providing the structural mobility for triple helix formation. In addition to this, there are proline residues also present in the “X” position of the “Gly-Xaa-HyP” motif that are commonly 3-hydroxylated (3-HyP) in nature. However, the occurrence of 3-HyP is rarer compared to 4-HyP (Weis et al., 2010). These prolyl hydroxylations are catalyzed by a specific class of prolyl hydroxylases (Hudson and Eyre, 2013; Merl-Pham et al., 2019; Salo and Myllyharju, 2021). Recently, it has been documented that genetic deletion of prolyl-3-hydroxylase 2 (P3H2) in mice evidenced that abolishment of prolyl-3-hydroxylations in basement membrane collagen could be involved in regulating platelet aggregation and development of eye tissues (Pokidysheva et al., 2014; Hudson et al., 2015). Osteogenesis Imperfecta is one of the reported diseases resulting from a lack of 3-hydroxyproline in COL1A1 (P^{1164}) (Morello et al., 2006; Cabral et al., 2007).

Apart from proline hydroxylation, lysine hydroxylation, and glycosylation are also critically important for the proper

functioning of collagens. In collagen chains, lysine residues occurring within the Gly-Xaa-Lys motif are likely to be hydroxylated by different family members of the lysyl hydroxylase; lysyl hydroxylase 1 (LH1), LH2, and LH3. Once hydroxylated, these hydroxylysines (HyK) can further be O-linked glycosylated by the addition of monosaccharides and/or disaccharides by specific glycosyltransferases (Kivirikko et al., 1973; Risteli and Kivirikko, 1974; Schegg et al., 2009; Scietti et al., 2018). Lysine (found in the Gly-Xaa-Lys motif) could be present in four different forms in collagen molecules deposited in the ECM. Unmodified lysine (K), hydroxylysine (HyK), galactosyl-hydroxylysine (G-HyK), and glucosyl-galactosyl-hydroxylysine (GG-HyK) occurring on a single site of the collagen chain. This site-specific variation of lysine residues is termed lysine microheterogeneity. Altered levels of lysine microheterogeneity in collagens lead to dysfunction in the tissues. Musculoskeletal defects (Geister et al., 2019), connective tissue disorder (Salo et al., 2008), and cerebral small vessel disease (Miyatake et al., 2018) are reported to have altered hydroxylysine and lysyl O-glycosylation levels in collagens. These studies indicate the significant role of collagen PTMs in maintaining ECM and tissue homeostasis. ECM homeostasis is remodeled upon injury to the myocardium and thereby ECM remodeling holds the key to the regeneration of amputated zebrafish hearts. Collagens are the most abundant components of ECM and play a significant role in the regeneration of zebrafish hearts. Various functions of collagen are dependent on site-specific PTMs (Sipilä et al., 2007; Jürgensen et al., 2011; Stawikowski et al., 2014). More importantly, these site-specific collagen PTMs could be tissue-specific which demands rigorous exploration in this field.

In the context of zebrafish cardiac tissue regeneration, the dynamics of the site-specific collagen PTM network are still unexplored. Therefore, to comprehensively map the site-specific collagen PTMs constituting the myocardium ECM and to understand the dynamics of collagen PTMs during the regeneration process, we found an appropriate proteomics study performed by Garcia-Puig et al. (2019) on zebrafish heart regeneration. To bridge this gap we downloaded this publicly available data set (PXD011627 and PXD010092) and generated a comprehensive map of site-specific collagen PTMs in adult zebrafish hearts. Furthermore, we quantitated the occupancy levels of collagen PTMs during the regeneration process. Understanding the dynamics of site-specific collagen PTMs in cardiac extracellular matrix during heart regeneration will open up new avenues to decode ECM remodeling and may lay the foundation to tinker the cardiac regeneration process with new approaches.

2 METHODS

2.1 Mass Spectrometry Data Source

In this study, two publicly available datasets were utilized. The first dataset with identifier “PXD011627” and the second dataset with identifier “PXD010092” were both submitted by Garcia-Puig et al. (2019) in ProteomeXchange. In this study, Garcia-Puig et al.

established the zebrafish myocardium ECM enrichment protocol by decellularizing zebrafish ventricular tissue samples (Garcia-Puig et al., 2019). Moreover, this study assessed the dynamic changes in the ECM proteome of regenerating zebrafish hearts using an amputation model. The samples of regenerating heart ventricles were taken at 7-, 14- and 30-days post-amputation (DPA). Ventricular decellularization was done using SDS and Triton-X. The sham model was used as a control in their study.

In their study, MS data acquisition was done differently for the two datasets. Easy-nLC 1,000 (Thermo) was used for liquid chromatography separation. Nanoflex (Thermo) was used as an ESI source and LC-MS/MS was done on Q-exactive HF mass spectrometer (Thermo) with HCD fragmentation and the data-dependent acquisition opted for dataset “PXD011627”. For dataset “PXD010092” a nano HPLC system (Proxeon) coupled with Maxis Impact (Bruker) Q-TOF mass spectrometer was used and each sample of regenerating and sham model heart ventricle was analyzed using LC-MS in duplicates. Furthermore, healthy human heart sample (ECM) mass-spectrometry data from Barallobre-Barreiro et al. (PXD028908) (Barallobre-Barreiro et al., 2021), and mice heart sample (ECM) mass-spectrometry data from Padmanabhan et al. (PXD002488) (Padmanabhan Iyer et al., 2016) was reanalyzed for the site-specific collagen PTM identification.

2.2 Raw Data Description

The publicly available datasets “PXD011627” and “PXD010092” were downloaded from ProteomeXchange. The complete *.d folders for “PXD010092” were shared by Prof. Angel Raya’s research group. Dataset “PXD011627” contains 9 (.raw) files and Dataset “PXD010092” contains 8 (.d) files, corresponding to two files (biological replicate) each for sham, 7-, 14- and 30-days post-amputation samples. Dataset “PXD010092” was not compatible with search engines because (.d) files are not readily accepted by many search engines. To overcome this, (.d) files were converted into. mgf and. mzML format using MSConvert with turbocharger filter. These converted files were further used as inputs by search engines MyriMatch (Tabb et al., 2007) and MSFragger (Kong et al., 2017), respectively. We also downloaded 60 (.raw) files from Barallobre-Barreiro et al. (PXD028908) and 6 (.raw) files from Padmanabhan et al. (PXD002488).

2.3 Database Search for the Identification of Collagen Chains and PTMs (Hydroxylation of Proline and Lysine, O-Glycosylation of Lysines) Using MyriMatch

Two different search engines were employed to reanalyze these datasets to identify collagen PTMs from the zebrafish myocardium ECM. All the database searches were performed on the high-performance computing (HPC) facility at the Indian Institute of Technology (IIT)- Mandi. First, a general database search was performed on raw MS data (*.raw files) of dataset “PXD011627” with a complete *Danio rerio* uniprot database having 62,593 entries, downloaded on 20 October 2020. For the PXD028908 dataset, a general database search was performed on raw MS data (*.raw files) with a complete *Homo sapiens* uniprot database having 20,396 entries,

downloaded on 16 February 2021. For the PXD002488 dataset, a general database search was performed on raw MS data (*.raw files) with a complete *Mus musculus* uniprot database having 17,090 entries, downloaded on 04 December 2021. Precursor ion tolerance was set at 10 ppm and fragment ion m/z tolerance was allowed up to 20 ppm for zebrafish and human data. However, for the mice dataset, a fragment ion tolerance of 50 ppm was used. Carbamidomethylation (+57.0236) on cysteine was used as static modification and methionine oxidation (+15.9949) and hydroxyproline (+15.9949) were used as dynamic modifications with up to 4 maximum dynamic modifications per peptide. A maximum of 2 missed cleavages were allowed for fully tryptic digestion. MyriMatch searches were performed to generate *.pepXML files for respective datasets. These *.pepXML files were further grouped for PSM matches, peptide, and protein group identification using IDPicker with <1% FDR. After a general database search, the identified list of ECM proteins from zebrafish and humans was exported from IDPicker as a *.FASTA database. This *.FASTA database (separately for zebrafish, humans, and mice) was used as a subset database to perform an in-depth collagen PTM search. In the subset database search, up to 4 missed cleavages for fully tryptic digestion were allowed with a similar precursor (10 ppm) and fragment ion tolerance (20 ppm) tolerance. For mice MS data, a fragment ion tolerance of 50 ppm was used. As mentioned previously, (Basak et al., 2016; Merl-Pham et al., 2019), a “Gly-Xaa-Yaa” based motif-specific PTM search strategy was employed using the MyriMatch motif-specific module for the identification of collagen PTMs. Carbamidomethylation on cysteine was used as static modification and methionine oxidation, hydroxyproline (P! +15.9949), hydroxylysine (GXX! +15.994,916), galactosyl-hydroxylysine (GXX! 178.047738) and glucosylgalactosyl-hydroxylysine (GXX! 340.100,562) were used as dynamic modifications with up to 10 maximum dynamic modifications per peptide. IDPicker was used for grouping the *.pepXML output file by controlling FDR at a 1% level for PSMs, peptides, and proteins. 3-hydroxyproline modifications were only considered in case a proline residue was found to be hydroxylated at the “X” position of a “G-Xaa-HyP” motif in the collagen chains. Furthermore, pLabel was used for manual inspection, analysis, and validation of subset database search PSMs for assigning a specific collagen PTM containing peptide.

2.4 Database Search for the Identification of Collagen Chains and PTMs (Hydroxylation of Proline and Lysine, O-Glycosylation of Hydroxylysine Sites) Using MSFragger*

mzML files were used for MSFragger mediated database search. In MSFragger based pipeline, the decoy database was generated by a philosopher from the same uniprot database (*Danio rerio*, 62,593 entries, downloaded on 20 October 2020) as in the case of the MyriMatch general search. Precursor tolerance and fragment ion tolerance were kept at 50 and 25 ppm, respectively. PSM, peptide, and protein level FDR were kept at < 1%, with a maximum of three missed cleavages of fully tryptic peptides. Carbamidomethylation (+57.0236) on cysteine was used as static modification and methionine oxidation (+15.9949), hydroxyproline (+15.9949), N-terminal acetylation (+42.010565), hydroxylysine (+15.994,916), galactosyl-hydroxylysine (+178.047738) and

glucosylgalactosyl-hydroxylysine (+340.100,562) were used as dynamic modifications with up to 5 maximum dynamic modification per peptide. IDPicker was used for grouping the *.pepXML output file generated from MSFragger searches by controlling FDR at <1% level for PSMs, peptides, and proteins. Similar to MyriMatch-based workflow, pLabel was used for manual inspection and validation for assigning a specific PSM to identify site-specific collagen PTM containing peptides.

2.5 Relative Abundances of Collagen Chains in the Heart ECM During Regeneration

Spectral counts (referred to the total number of peptides detected at ~1% FDR per protein group) from the results of the MyriMatch database search were exported from IDPicker. These spectral counts were normalized with the total spectral counts of each raw MS file ID. The normalized spectral counts corresponding to unique protein groups were further used to quantitate the relative abundance of collagens in wild-type and regenerating zebrafish hearts. Additionally, the same normalized spectral counts data for collagens from MyriMatch database search results were used to generate a heatmap of relative collagen expression during regeneration. Collagen chains identified with ≥ 3 normalized spectral counts were only considered for the heatmap representation.

2.6 Quantitation of Occupancy Level of Collagen PTM Sites Using Skyline

Database search results (*.pepXML files) generated by either MyriMatch and/or MSFragger were parsed through PeptideProphet (TPP pipeline module) for importing the probability scores (0-1) (Pedrioli, 2010). After parsing the *.pepXML files through PeptideProphet, these *.pepXML files were utilized to build a spectral library (.blib) using open-source Skyline (MacLean et al., 2010). A spectral library was used in Skyline for extracting the raw abundance intensity of unmodified and different forms of modified collagen peptides. Occupancy calculation of prolyl hydroxylation sites and microheterogeneity (unmodified, hydroxylated, O-glycosylation of lysines) of lysine sites were performed using MS¹ based targeted extraction pipeline as described previously (Merl-Pham et al., 2019).

2.7 Statistical Analysis

GraphPad prism was used for statistical analysis for calculating the dynamics of different collagen site-specific PTM level quantitation. ANOVA was performed and a *p* value < 0.05 was considered statistically significant.

3 RESULTS AND DISCUSSION

Myocardium extracellular matrix (ECM) consisting of collagens as a major component, is sought to be one of the key molecules involved in the remodeling process during zebrafish heart regeneration (Sánchez-Iranzo et al., 2018). Collagen and its types are typically triple-helical, large protein molecules

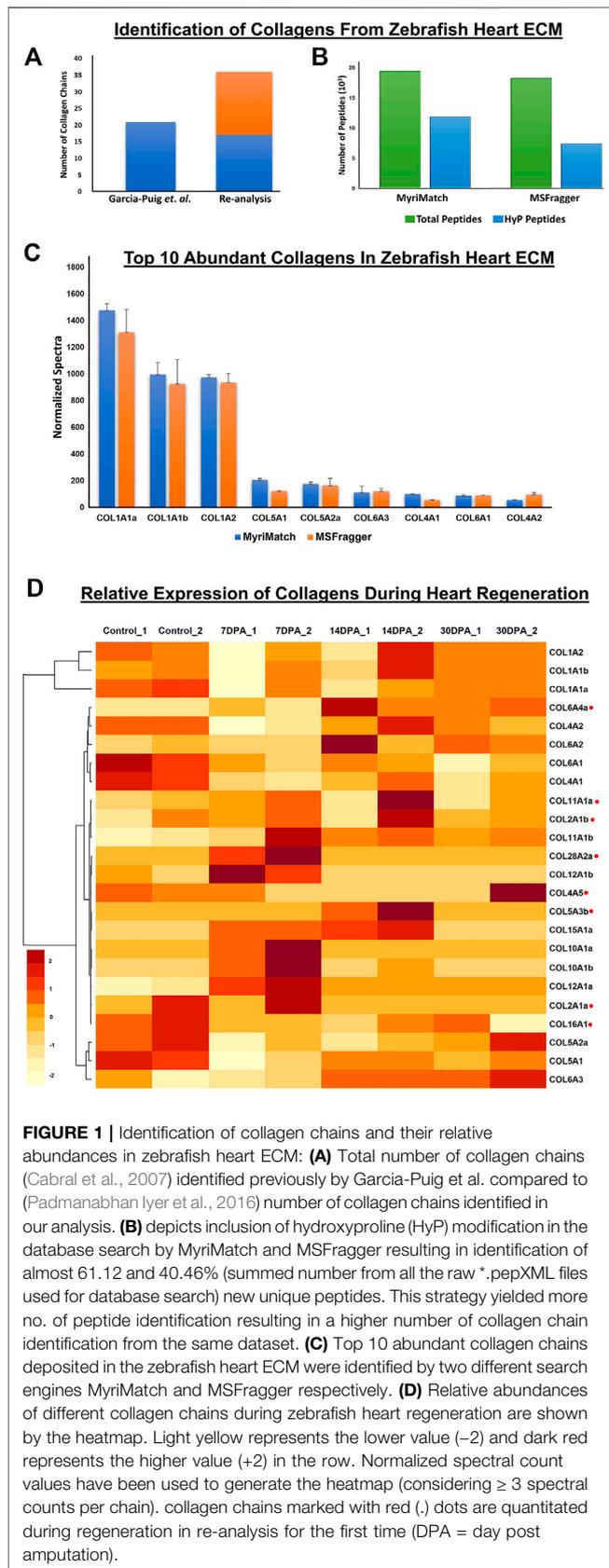


FIGURE 1 | Identification of collagen chains and their relative abundances in zebrafish heart ECM: **(A)** Total number of collagen chains (Cabral et al., 2007) identified previously by Garcia-Puig et al. compared to (Padmanabhan Iyer et al., 2016) number of collagen chains identified in our analysis. **(B)** depicts inclusion of hydroxyproline (HyP) modification in the database search by MyriMatch and MSFragger resulting in identification of almost 61.12 and 40.46% (summed number from all the raw *.pepXML files used for database search) new unique peptides. This strategy yielded more no. of peptide identification resulting in a higher number of collagen chain identification from the same dataset. **(C)** Top 10 abundant collagen chains deposited in the zebrafish heart ECM were identified by two different search engines MyriMatch and MSFragger respectively. **(D)** Relative abundances of different collagen chains during zebrafish heart regeneration are shown by the heatmap. Light yellow represents the lower value (-2) and dark red represents the higher value (+2) in the row. Normalized spectral count values have been used to generate the heatmap (considering ≥ 3 spectral counts per chain). collagen chains marked with red (.) dots are quantitated during regeneration in re-analysis for the first time (DPA = day post amputation).

secreted in the extracellular space and get assembled to form fibrils in the interstitium of the tissue responsible for structural support, cell-ECM interactions, and cell behavior. (Basak et al., 2016).

3.1 Identification of Collagen Chains, and Their Abundances in Wild Type Zebrafish Heart ECM-

To identify and quantitate the abundance of collagen chains present in wild-type zebrafish heart ECM, a dual database search engine-based (MyriMatch and MSFragger) comprehensive strategy was implemented. MyriMatch (Tabb et al., 2007) and MSFragger (Kong et al., 2017) were utilized in this study for in-depth identification of collagen chains present in the myocardium ECM of zebrafish from the publicly available datasets (PXD011627, PXD010092) as mentioned earlier. This analytical pipeline identified a total of 36 collagen chains present in the ECM of the zebrafish heart (**Figure 1A**; **Supplementary Table S1**). Out of these 36 collagen chains, 17 are commonly identified and 19 collagen chains are newly identified in the ECM of zebrafish heart (**Figure 1A**, **Supplementary Table S1**). Previously, Garcia-Puig et al. identified a total of 21 collagen chains from the enriched ECM of zebrafish hearts using this dataset (Garcia-Puig et al., 2019). Since collagens are heavily modified with prolyl hydroxylation, a standard proteomic data analysis pipeline will not be able to identify hydroxylated peptides of collagen chains present in the raw data (Merl-Pham et al., 2019). However, in our analytical pipeline, we included prolyl-hydroxylation as a dynamic modification in the database search yielding a higher number of collagen chain identification. Database searches yielded almost 19,456 and 18,274 total peptide ids from MyriMatch and MSFragger respectively. As expected, almost 61 and 40% of these identified peptides contained prolyl-hydroxylation (**Figure 1B**, **Supplementary Table S2**). Thus, it substantiates the use of prolyl-hydroxylation as an important “dynamic modification” in order to explore ECM mass-spectrometry data to increase the identification of collagen chains along with their increased sequence coverage. Normalized spectral counts were used to calculate the abundance of the top 10 collagen chains present in the ECM of wild-type zebrafish hearts (**Figure 1C**). Both MyriMatch and MSFragger unambiguously yielded COL1A1a, COL1A1b, and COL1A2 chains of collagen 1 to be the highest abundant collagen chains present in the cardiac ECM of zebrafish. These three chains form the triple-helical collagen 1 protomers (Gistelink et al., 2016), the most abundant protein present in the extracellular matrix of wild-type zebrafish hearts. Apart from Collagen 1, COL5A1, COL5A2a, COL6A3, COL4A1, COL6A1, and COL4A2 chains were found to be in the 10 most abundant collagen chains present in the ECM of WT zebrafish heart. These abundance plots yielded from two different search engines were in agreement highlighting the robustness of the analysis.

3.2 Relative Abundances of Collagen Chains in the Heart ECM During Regeneration–

The cardiac regeneration process involves ECM remodeling. Being the most abundant component of ECM (Di Lullo et al., 2002), levels of collagens are also altered during regeneration (Garcia-Puig et al., 2019) (Sánchez-Iranzo et al., 2018). As expected, normalized spectral counts of both the search engines show that Collagen 1 (COL1A1a, COL1A1b, and COL1A2) is most abundant in zebrafish heart ECM during 7, 14, and 30 days post-amputation (**Supplementary Figure S1**). Here, we specifically focused to showcase the relative changes of different collagen chains present in the ECM during regeneration (**Figure 1D**, **Supplementary Table S3**). Out of the total 36 collagen chains identified, relative abundances were calculated for 24 collagen chains summarized in the heatmap presented in **Figure 1D** (raw data for individual chains are provided in **Supplementary Table S3**). Interestingly, the level of total collagen 1 (summed abundances of all three chains of the triple-helix) decreased (to 0.68 fold, 32% decrease in the level, not significant) at 7 days post-amputation, was slightly increased at 14 days post-amputation (0.86 fold compared to control, 14% decrease in level compared to sham control model) and it also further increased (0.93 fold compared to sham model (control), 6.2% decrease in level compared to sham model) at 30 days post-amputation. Although these changes were not found to be statistically significant, the expression trend corroborated the previous analysis (Garcia-Puig et al., 2019). A similar expression was found for COL5A1 and COL5A2a during the regeneration process. Levels of these 2 collagen chains were first decreased at 7 days post-amputation and then increased at 14 and 30 days post-amputation (Garcia-Puig et al., 2019). The expression of COL4A2, COL5A1, and COL5A2a in our analysis is also similar to the analysis done by Garcia-Puig et al. (2019) in which they found a decrease in the level of COL4A2, COL5A1, and COL5A2a on 7 days post-amputation indicating a feedback post-transcriptional regulation. Furthermore in this analysis, we quantitated the relative abundances of 8 new collagen (COL6A4, COL11A1, COL2A1b, COL28A2a, COL4A5, COL5A3b, COL2A1a, COL1A1) (**Figure 1D**, marked with red dots) chains present in the myocardium ECM during heart regeneration. The oscillatory dynamics of abundances of these new collagen chains are depicted in the heatmap (**Figure 1D**, **Supplementary Table S3**). These expressions of collagen chains highlight the role of collagen in the extracellular remodeling during regeneration.

3.3 Global Characterization and Identification of Novel Site-specific Collagen Post-translational Modifications (PTMs) in Zebrafish Myocardium ECM

Collagen PTMs such as O-glycosylation (on hydroxylysine residues) and hydroxylation (on lysine and proline residues) have been shown to be crucial for embryonic development, assembly of ECM fibrils, and cell-matrix interactions (Pokidysheva et al., 2014; Hudson et al., 2015). Collagen

molecules get heavily post-translationally modified during the biosynthesis of new chains in the endoplasmic reticulum (Hennet, 2019). Recently, the use of a hydroxyproline-based search strategy using mass-spectrometry data yielded increased coverage for collagen chains (Merl-Pham et al., 2019; Basak et al., 2016; Shao et al., 2020). However, it is challenging to identify the site-specific O-glycosylation sites present in the big collagen chains (Basak et al., 2016). The role of collagen PTMs during tissue regeneration is completely unexplored. Classical amino acid analysis-based approaches could throw insights regarding the composition (including modified amino acids) of these collagen triple helices. However, site-specific identification of these collagen PTMs (hydroxylation, O-glycosylation)–can only be achieved by high-resolution mass spectrometry. So, this study primarily focused on the identification of site-specific collagen PTMs present in the ECM of the zebrafish heart. Furthermore, the dynamics of these PTMs in collagen I (most abundant collagen in heart ECM, a heteromeric trimer constituting three different chains COL1A1a, COL1A1b, and COL1A2) is being showcased during the regeneration of zebrafish heart. A total of 36 collagen chains were identified from the zebrafish heart ECM in this analysis as mentioned previously. Identification of collagen PTMs from a global MS analysis, without biochemical purification, has remained challenging. We had previously developed a MyriMatch-based workflow to identify site-specific collagen PTMs from crude ECM proteome preparations using high-resolution mass spectrometry (Basak et al., 2016). Here, our *inhouse* collagen PTM identification pipeline (**Figure 2**) with a dual search engine-based strategy maximized the identification of prolyl-hydroxylation, lysyl-hydroxylation, and O-glycosylation sites present in 23 collagen chains from zebrafish heart ECM (**Table 1**). We detected a total of 95 3-hydroxyproline, 108 hydroxylysine, 29 galactosyl-hydroxylysine, and 128 glucosylgalactosyl-hydroxylysine sites in 23 collagen chains (**Table 1**) present in the zebrafish heart ECM. This is the first catalog of site-specific collagen PTMs identified in the ECM of the zebrafish heart. Notably, site-specific collagen PTMs were identified in the highest abundant COL1A1a, along with COL11A1a and COL22A1 spanning across three orders of magnitude (log scale) (see **Table 1**). This highlights the sensitivity and the depth of dynamic range of our analytical pipeline to explore the PTMs of collagens present in the ECM of zebrafish heart. Collagen I seemed to have a higher number of prolyl-3-hydroxylations and lysyl-hydroxylation sites. Collagen 5 is a minor constituent of fibril assembly (Wenstrup et al., 2004). However, this analysis revealed 12 (P⁶⁸⁶, P⁷⁴⁶, P⁸²⁴, P¹¹¹², P¹¹¹⁵, P¹¹⁶³, P¹¹⁶⁶, P¹¹⁹⁰, P¹¹⁹³, P¹²⁵³, P¹⁴²¹, P¹⁴²⁴) and nine sites (P²⁴¹, P²⁷⁷, P⁵³², P⁶⁶¹, P⁸⁶⁸, P⁹⁰¹, P⁹³⁷, P⁹⁶⁷, P¹¹⁰⁵) of 3-HyP in COL5A1 and COL5A2a, respectively. COL5A2a was then further mapped with 11 sites (K³²⁹, K³³⁸, K⁵⁰⁰, K⁵⁵⁷, K⁵⁷², K⁷⁶¹, K⁷⁶¹, K⁷⁹⁴, K⁶³⁸, K⁸⁰³, K⁸⁸⁷) of glucosyl-galactosyl hydroxylysines. Only one galactosyl-hydroxylysine site (K¹³⁴⁴) was identified in COL5A1. As previously shown in the zebrafish genome (from fin tissue) analysis (Duran et al., 2015), no type III collagen chain is detected in our analysis from the zebrafish cardiac ECM. The absence of type III collagen is a unique feature of zebrafish heart

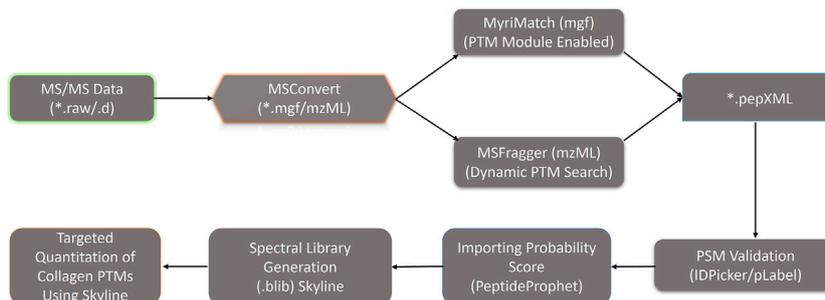


FIGURE 2 | Optimized dual database search engine-based MS analysis pipeline for the global identification and quantitation of site-specific collagen PTMs from zebrafish heart ECM. Thermo. raw or Bruker. d MS/MS files were initially converted to. mgf and. mzML files (by MSConvert) respectively and searched with MyriMatch and MSFragger to identify the collagen present in the zebrafish heart ECM. For MyriMatch, the subset of identified proteins was used as a second database to perform a PTM module enabled search defining specific sequence motifs for the site-specific identification of collagen PTMs in zebrafish heart ECM. For MSFragger, the PTM searches were conducted directly with the entire zebrafish database. From MyriMatch and MSFragger *.pep.XML output files containing each peptide spectrum match (PSM) were further parsed by PeptideProphet to compute the probability score (0, 1). The *.pep.XML output file parsed by PeptideProphet was further imported into Skyline along with all the raw MS/MS files in to generate the spectral library (.bib). This spectral library (.bib) in Skyline was used for the targeted MS¹-based extraction of all the PTM modified and unmodified collagen peptide species for each specific site. The area of MS¹ area for each peptide for different samples was computed from Skyline.

ECM and warrants further scope to investigate the role of collagen III during regeneration. Four different chains of collagen 6 (COL6A1, COL6A2, COL6A3, COL6A4a) were found to harbor O-glycosylation but only one prolyl-3-hydroxylation site (3-HyP⁴⁷⁰ in COL6A2) (Table 1). Although basement membrane collagen IV is not a major constituent of heart ECM; in the case of collagen IV alpha 1 only seven sites (K⁴⁶³, K⁴⁶⁶, K⁸⁸², K⁹⁰⁹, K¹¹⁴⁹, K¹¹⁷⁹, K¹¹⁸²) of glucosyl-galactosyl-hydroxylation were identified. Surprisingly, 24 glucosyl-galactosyl-hydroxylation sites were identified in COL4A2. A total of 4 (P²⁰¹, P²⁰⁴, P²⁹⁴, P²⁹⁷, P¹³³⁸) and 3 (P³³⁹, P⁵⁵⁵, P⁶¹³) prolyl-3-hydroxylation sites were identified in COL4A1 and COL4A2, respectively. It appeared that COL4A2 is way more glycosylated than COL4A1 in the zebrafish cardiac ECM compared to mice (Basak et al., 2016). Interestingly, we have also identified the COL4A5 chain with a common hydroxylysine site harboring microheterogeneity of galactosyl-hydroxylysine and glucosyl-galactosyl-hydroxylysine (K²²⁵). Additionally, many other sites of prolyl-3-hydroxylations, lysyl-hydroxylation and O-glycosylation sites of lysine have been identified in various other collagen types (COL11A1b, COL7A1, COL22A1, etc.; see Table 1) highlighting the advantageous nature of these analyses increasing the depth of site-specific PTM identification coverage in the ECM. Taken together, these results showcased the comprehensive mapping and site-specific identification of many collagens present in the zebrafish heart ECM.

3.4 Characterization of PTMs in Collagen I From the Zebrafish Heart ECM

Collagen I is the highest abundant ECM protein present in the zebrafish hearts (Gistelincx et al., 2016). The composition of the heterotrimer of the collagen I chain forming a protomer (triple helices) is different compared to mammals (Gistelincx

et al., 2016). In mammals, two alpha 1 chain and one alpha 2 chain form the protomer. However, in zebrafish, three different genes (COL1A1a, COL1A1b, and COL1A2) code for three different collagen I chains (collagen alpha 1a, collagen 1 alpha 1b (alpha 3), and collagen 1 alpha 2), respectively. High expression of these three genes has been documented in adult zebrafish tissues as well as during embryonic and larval development indicating equimolar stoichiometry in forming the protomers (Gistelincx et al., 2016). However, different glycosylated forms of COL1A1a have been found in zebrafish embryos indicating altered PTM levels during development (Gistelincx et al., 2016). Thus, tissue-specific mapping of collagen I site-specific PTMs are of the highest importance in order to further delineate their function. Here, we have comprehensively mapped the site-specific prolyl-hydroxylation, lysyl-hydroxylation, and lysine O-glycosylation in three different chains (COL1A1a, COL1A1b, and COL1A2) of collagen I in zebrafish heart ECM.

3.4.1 Comprehensive Site-specific Map of PTMs in COL1A1a From the Zebrafish Heart ECM–

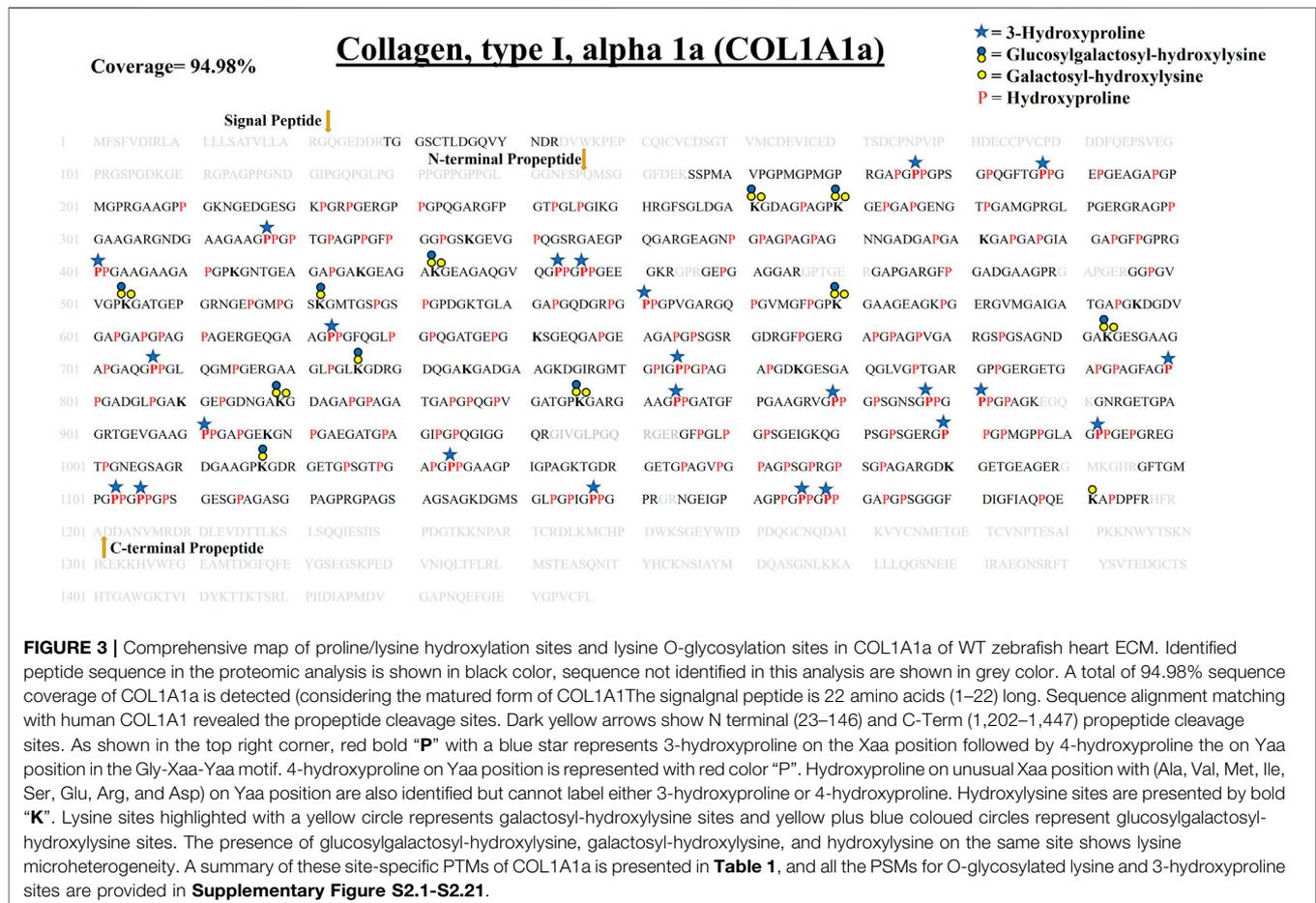
COL1A1a chain was found to be the highest abundant out of all the three chains of collagen 1 triple helix in ECM of zebrafish heart. This chain has an important role in the stability and proper functioning of zebrafish cardiac tissues (Gistelincx et al., 2016). It is a large 1,447 amino acid long collagen chain. For determining the N (23–146, FSP/QMS), and C terminal propeptides cleavage site (1,202–1,447, RA/DDAN), we performed cleavage site sequence alignment matching with human COL1A1, which has also been reported previously (Gistelincx et al., 2016). However, we also detected a peptide sequence from N terminal propeptide hinting toward the dynamics of collagen assembly in the ECM. PTM module enabled dual search

TABLE 1 | Mass-Spectrometry-based identification of site-specific hydroxylation and glycosylation of different collagens present in the zebrafish cardiac ECM.

Collagen chains	3-Hydroxyproline Sites (95)	Hydroxylysine Sites (108)	Galactosyl-Hydroxylysine Sites (29)	Glycosylgalactosyl-Hydroxylysine Sites (128)
COL1A1a	P ¹⁷⁶ , P ¹⁸⁸ , P ³¹⁷ , P ⁴⁰¹ , P ⁴⁴³ , P ⁴⁴⁶ , P ⁵⁵¹ , P ⁶²³ , P ⁷⁰⁷ , P ⁷⁵⁵ , P ⁸⁰⁰ , P ⁸⁵⁴ , P ⁸⁶⁹ , P ⁸⁷⁸ , P ⁸⁸¹ , P ⁹¹¹ , P ⁹⁸⁰ , P ⁹⁹² , P ¹⁰³⁴ , P ¹¹⁰³ , P ¹¹⁰⁶ , P ¹¹⁴⁸ , P ¹¹⁶⁶ , P ¹¹⁶⁹	K ²⁶¹ , K ²⁷⁰ , K ³³⁶ , K ³⁸¹ , K ⁴¹⁴ , K ⁴²⁶ , K ⁴³² , K ⁵⁰⁴ , K ⁵²² , K ⁵⁷⁰ , K ⁵⁹⁶ , K ⁶⁴¹ , K ⁶⁹³ , K ⁷²⁶ , K ⁷³⁵ , K ⁷⁶⁵ , K ⁸¹⁰ , K ⁸¹⁹ , K ⁸⁴⁶ , K ⁹¹⁸ , K ¹⁰¹⁷ , K ¹⁰⁸⁰ , K ¹¹⁹¹	K ²⁶¹ , K ²⁷⁰ , K ⁴³² , K ⁵⁰⁴ , K ⁵⁷⁰ , K ⁶⁹³ , K ⁸¹⁹ , K ⁸⁴⁶ , K ¹¹⁹¹	K ²⁶¹ , K ²⁷⁰ , K ⁴³² , K ⁵⁰⁴ , K ⁵²² , K ⁵⁷⁰ , K ⁶⁹³ , K ⁷²⁶ , K ⁸¹⁹ , K ⁸⁴⁶ , K ¹⁰¹⁷
COL1A1b	P ¹⁷⁶ , P ¹⁹¹ , P ³⁰² , P ³²⁰ , P ⁴⁰⁴ , P ⁴⁷⁰ , P ⁵⁵⁴ , P ⁷⁵⁸ , P ⁸⁷² , P ⁸⁸¹ , P ⁹¹⁴ , P ⁹³⁸ , P ⁹⁸³ , P ¹⁰³¹ , P ¹¹⁰⁹ , P ¹¹⁵¹ , P ¹¹⁷⁵	K ²¹⁵ , K ²⁶⁴ , K ²⁷³ , K ³³⁹ , K ³⁸⁴ , K ⁴¹⁷ , K ⁴²⁹ , K ⁵²⁵ , K ⁵⁷³ , K ⁶¹⁸ , K ⁶⁹⁶ , K ⁷³⁸ , K ⁷⁴⁶ , K ⁷⁶⁸ , K ⁸¹³ , K ⁸⁴⁹ , K ⁹⁰⁰ , K ¹⁰²⁰ , K ¹¹⁹³	K ²⁶⁴ , K ²⁷³ , K ⁸⁴⁹ , K ¹⁰²⁰	K ²⁶⁴ , K ²⁷³ , K ³³⁹ , K ⁸⁴⁹
COL1A2	P ⁵⁶ , P ⁵⁹ , P ²⁴⁴ , P ²⁸⁶ , P ³⁶¹ , P ⁴⁵¹ , P ⁴⁶⁹ , P ⁶⁷³ , P ⁷¹⁸ , P ⁷⁸⁷ , P ⁸²⁰ , P ⁸³⁸ , P ⁹²⁵ , P ¹⁰⁶⁶ , P ¹⁰⁸¹	K ⁷⁴ , K ¹⁷⁹ , K ¹⁸⁸ , K ²⁵⁴ , K ²⁹⁹ , K ³⁴⁴ , K ³⁵⁰ , K ⁴⁵⁴ , K ⁴⁸⁸ , K ⁵⁰⁰ , K ⁵³³ , K ⁵⁷⁸ , K ⁶⁴⁴ , K ⁶⁴⁷ , K ⁶⁸³ , K ⁷³¹ , K ⁷³⁷ , K ⁸³⁶ , K ⁹³⁵ , K ⁹⁹⁸ , K ¹⁰⁰⁴	K ⁷⁴ , K ¹⁸⁸ , K ²⁵⁴ , K ³⁴⁴ , K ³⁵⁰ , K ⁶⁴⁴	K ⁷⁴ , K ¹⁶⁷ , K ¹⁸⁸ , K ²⁵⁴ , K ²⁹⁹ , K ⁵⁷⁸ , K ⁶⁴⁴ , K ⁶⁴⁷ , K ⁹³⁵
COL4A1	P ²⁰¹ , P ²⁰⁴ , P ²⁹⁴ , P ²⁹⁷ , P ¹³³⁸			K ⁴⁶³ , K ⁴⁶⁶ , K ⁸⁸² , K ⁹⁰⁹ , K ¹¹⁴⁹ , K ¹¹⁷⁹ , K ¹¹⁸²
COL5A1	P ⁶⁸⁶ , P ⁷⁴⁶ , P ⁸²⁴ , P ¹¹¹² , P ¹¹¹⁵ , P ¹¹⁶³ , P ¹¹⁶⁶ , P ¹¹⁹⁰ , P ¹¹⁹³ , P ¹²⁵³ , P ¹⁴²¹ , P ¹⁴²⁴	K ¹¹³⁴ , K ¹²⁹⁶ , K ¹³²⁹ , K ¹⁴⁹⁷ , K ¹⁵⁷⁸ , K ¹⁵⁸¹ , K ¹⁵⁹² , K ¹⁶⁰⁴ , K ¹⁶⁴⁴	K ¹³⁴⁴	K ⁶²³ , K ⁶²⁶ , K ⁷⁵³ , K ⁸¹³ , K ⁹⁴⁵ , K ¹¹⁰⁴ , K ¹²⁰⁹ , K ¹²⁴⁵ , K ¹²⁹⁰ , K ¹³³⁸ , K ¹³⁸³ , K ¹⁴⁵⁵ , K ¹⁴⁸² , K ¹⁵⁷⁸ , K ¹⁵⁸¹ , K ³²⁹ , K ³³⁸ , K ⁵⁰⁰ , K ⁵⁵⁷ , K ⁵⁷² , K ⁷⁶¹ , K ⁷⁶¹ , K ⁷⁹⁴ , K ⁸³⁸ , K ⁸⁰³ , K ⁸⁸⁷
COL5A2a	P ²⁴¹ , P ²⁷⁷ , P ⁵³² , P ⁶⁶¹ , P ⁸⁶⁸ , P ⁹⁰¹ , P ⁹³⁷ , P ⁹⁶⁷ , P ¹¹⁰⁵	K ²⁸⁰ , K ³²⁹ , K ⁴⁴⁹ , K ⁴⁸² , K ⁴⁹⁴ , K ⁷⁰⁹ , K ⁷⁴⁹ , K ⁸⁰³ , K ⁸⁷⁸ , K ⁹⁸⁶ , K ¹⁰³⁶ , K ¹⁰⁷⁸ , K ¹¹²⁴ , K ¹¹³⁵		K ³⁹⁸ , K ⁴⁸⁸ , K ⁵³⁸ , K ⁵⁶⁰ , K ⁵⁶³ , K ¹⁶⁶⁰ , K ¹⁷⁷⁸ , K ¹⁷⁹⁶ , K ¹⁷¹¹ , K ¹⁹⁶³ , K ¹⁸⁸⁰ , K ¹⁸⁸⁹
COL6A1	P ⁵⁷⁴ , P ⁵⁷⁷	K ⁴⁸² , K ⁵³⁸ , K ¹⁷⁹⁶ , K ¹⁹⁷⁵		K ³³⁹ , K ³⁸⁷ , K ⁴²⁶ , K ⁴³⁵ , K ⁴³⁸ , K ⁴⁵⁶ , K ⁴⁵⁹ , K ⁵³⁹ , K ⁵⁴⁵
COL6A3				K ⁹¹¹ , K ¹¹⁴⁸ , K ¹¹⁵¹ , K ⁵⁴ , K ⁷² , K ³¹⁸ , K ⁵⁶⁷ , K ⁵⁸² , K ⁵⁹³ , K ⁶⁶² , K ⁶⁷⁴ , K ⁶⁸⁹ , K ⁷⁰⁴ , K ⁷⁰⁷ , K ⁷²⁵ , K ⁹⁴⁸ , K ⁹⁵¹ , K ⁹⁶⁹ , K ¹⁰⁴⁵ , K ¹⁰⁴⁸ , K ¹⁰⁷⁶ , K ¹²³⁵ , K ¹²⁴⁷ , K ¹²⁸³ , K ¹²⁹² , K ¹³³¹ , K ¹³⁴⁶
COL6A2	P ⁴⁷⁰	K ³⁸⁷	K ⁵³⁹	K ²²⁵ , K ¹⁴⁹⁰ , K ¹⁷⁸⁷ , K ¹⁷⁹⁶ , K ²¹⁴⁸ , K ⁴²³ , K ¹¹⁷⁸
COL11A1a	P ⁹¹⁹	K ¹³¹³	K ¹²¹⁷ , K ¹³⁰⁴	
COL4A2	P ³³⁹ , P ⁵⁵⁵ , P ⁶¹³	K ⁵⁹³ , K ⁷⁰⁴ , K ⁷⁰⁷ , K ¹⁶²⁰ , K ¹⁶⁵¹	K ⁶⁸⁹	
COL4A5	P ⁹³	K ¹²⁷ , K ²¹⁴	K ²²⁵	
COL6A4a	-	-	K ¹⁵⁰¹	
COL2A1a	-	K ⁶¹²	K ⁴²³	
COL2A1b	-	-	-	
COL16A1	-	K ¹⁰⁰⁸ , K ¹³⁴³ , K ¹⁴⁹⁵ , K ¹⁴⁹⁸	-	K ⁶⁴⁴ , K ⁶⁵⁴ , K ¹⁰⁰⁸ , K ¹²³³ , K ¹¹⁹¹ , K ¹⁴⁰⁰
COL11A1b	-	-	-	
COL5A2b	P ⁸³⁸	-	-	K ⁷³¹
COL11A2	P ¹¹³³	K ⁹⁵⁹ , K ⁹⁶³	-	K ¹¹⁸⁸ , K ¹¹⁹⁷ , K ¹²⁰⁰ , K ¹²⁰³
COL7A1	-	-	-	K ¹⁵⁰⁴ , K ²¹³⁶
COL5A3b	P ⁴³⁷ , P ⁴⁶⁷	-	-	-
COL22A1	-	K ⁴⁹⁵ , K ¹¹²⁸	K ⁷⁸⁸ , K ¹⁵³⁸	K ⁴⁹⁸ , K ⁵¹³ , K ⁷⁸⁵
COL17A1a	P ¹¹⁵⁶	-	-	-

engine-based database search yielded 94.98% sequence coverage (considering the processed chain as the full length) (**Figure 3; Table 1, Supplementary Figure S2**). A total of 92 4-HyP sites on the “Yaa” position of “Gly-Xaa-Yaa” motif were detected. This estimate is suggestive of about 95% 4-HyP occupancy in the Yaa positioned (in Gly-Xaa-Yaa motif) prolines corroborating the previous Edman sequencing analysis in COL1A1 from mammals (Fietzek and Kühn, 1975). Furthermore, we have also detected 41 hydroxyproline sites in the “Xaa” position of the “Gly-Xaa-Yaa” motif. This finding is similar (n = 37) to human COL1A1 (Merl-Pham et al., 2019). In addition, we analyzed the occurrence of prolyl-3-hydroxylation sites providing crucial structural equilibrium in the collagen molecules deposited in

ECM and affecting interactions with other ECM proteins (Pokidysheva et al., 2014; Grafe et al., 2014; Montgomery et al., 2018). We assigned these HyP as 3-hydroxyproline (3-HyP) as determined by previous experimental inferences (Fietzek and Kühn, 1975). A total of twenty-four 3-hydroxyproline sites (see **Table 1**) at the “Xaa” position of the G-Xaa-HyP motif in COL1A1a were identified. Out of these, 10 sites (P¹⁷⁶, P⁵⁵¹, P⁷⁵⁵, P⁸⁰⁰, P⁸⁶⁹, P⁸⁸¹, P⁹¹¹, P¹¹⁰³, P¹¹⁰⁶, and P¹¹⁴⁸) are evolutionarily conserved in mice (Padmanabhan Iyer et al., 2016) and human heart tissue (Barallobre-Barreiro et al., 2021) (**Figure 4; Table 1, Supplementary Figure S2 and Supplementary Figure S3**). This highlights probable conserved functions of these sites in collagen molecule providing structural support and



maintaining homeostatic ECM assembly. Interestingly, we report the identification of 14 novels (P^{188} , P^{317} , P^{401} , P^{443} , P^{446} , P^{623} , P^{707} , P^{854} , P^{878} , P^{980} , P^{992} , P^{1034} , P^{1166} and P^{1169}) 3-HyP sites in the COL1A1a from zebrafish heart ECM. In zebrafish heart ECM, COL1A1a harbors numerous unique 3-HyP sites close to the N-terminal propeptide cleavage sites (3-HyP^{188,317}), spanning across the triple-helical region as well. Two clusters of “Gly-Pro-Pro” motif present close to C-terminal propeptide were found to be 3-hydroxylated (3-HyP^{1103,1106} and 3-HyP^{1166,1169}). The presence of many new 3-HyP sites compared to human and mouse hearts is itself a unique finding in the zebrafish COL1A1a (**Figure 4**) protein structure. These unique prolyl-3-hydroxylation sites may contribute to ECM remodeling during heart regeneration and warrants further investigations.

3.4.2 Comprehensive Site-specific Map of PTMs in COL1A1b From the Zebrafish Heart ECM–

In zebrafish cardiac ECM, the second chain contributing to the collagen 1 triple helix is COL1A1b. It is a 1,449 amino acid long chain. Our analysis revealed COL1A1b to be the next highest abundant collagen chain present in the zebrafish cardiac ECM. We identified and comprehensively mapped site-specific PTMs of

COL1A1b. N terminal (150, FLS/QMA) and C terminal propeptide cleavage sites (1,204, YRA/DDA) were determined by sequence alignment, corroborating the previous report (Gistelinck et al., 2016). We detected 96.48% sequence coverage (considering the matured form) for COL1A1b. A total of 99 4-HyP sites on the “Yaa” position of “Gly-Xaa-Yaa” motif were detected out of the possible 103 sites. Similar to COL1A1a, we detected seventeen 3-HyP sites in COL1A1b on the ‘Xaa’ position of the Gly-Xaa-HyP motif out of the possible 27 sites. It seems that COL1A1b is less 3-hydroxylated at proline residues compared to COL1A1a. Furthermore, we have also detected 51 hydroxyproline sites in the “Xaa” position of “Gly-Xaa-Yaa” motif. As described previously, mass-spectrometry-based methods will not be able to resolve whether these prolyl-hydroxylations are occurring at the 3’ or 4’ position. However, recent studies have shown that these are most probably 4-prolyl-hydroxylations (Van Huizen et al., 2019). Surprisingly, we have detected a smaller number of lysine O-glycosylation sites in COL1A1b. A total of 19 hydroxylysines, 4 galactosyl-hydroxylysine, and 4 glucosylgalactosyl-hydroxylysine sites have been identified (**Table 1**). These comprehensive PTM maps of COL1A1b of zebrafish heart ECM are summarized in **Figure 5** (the

3-hydroxyproline Sites in Collagen 1 Alpha 1 Protein From Heart ECM

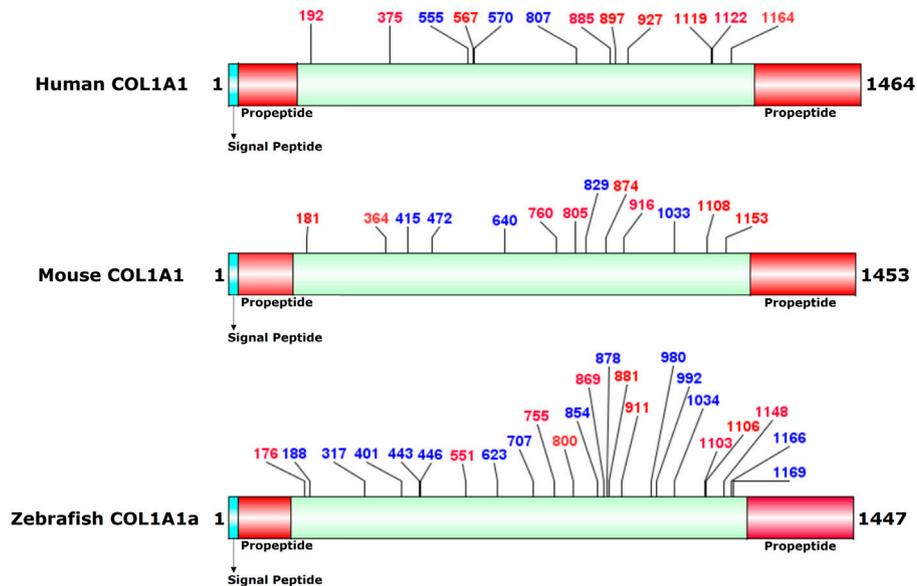


FIGURE 4 | Comparison of 3-HyP sites identified in COL1A1a from zebrafish heart ECM to COL1A1 of human and mice heart ECM. The horizontal box represents the full-length COL1A1 sequence and the vertical black lines indicate the corresponding 3-HyP sites. The information for 3-HyP sites of human and mice heart ECM were re-analyzed in this manuscript from the available raw MS data from Barallobre-Barreiro et al. and Padmanabhan et al. The 3-HyP sites marked with red represent the conserved sites among human, mouse, and zebrafish.

Coverage= 96.48%

Collagen, type I, alpha 1b (COL1A1b)

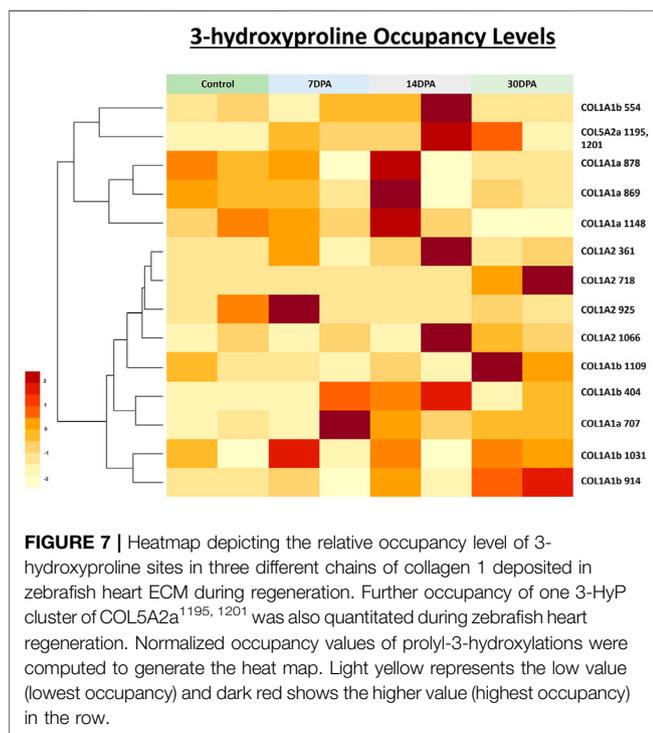
★ = 3-Hydroxyproline
 ⚙ = Glucosylgalactosyl-hydroxylysine
 ⚙ = Galactosyl-hydroxylysine
 P = Hydroxyproline

	Signal Peptide ↓									
1	MFSFVDIRLG	LLLAAVVLAV	RGQGEDDISP	GSCTI EGQNY	NDRKDVWRKPEP	CQICVCSGDT	VMCDEVICED	TSDCANPEIP	DGECCPICPD	EEETAGPEPP
	N-terminal Propeptide ↓									
101	YVDTVGPA GP	RGDPGPVGP	GRDGTPEGNG	LPGPPGPGT	SGLGGGSFLS	QMAYGSEKSS	GPPQPPAGP	MGPRGPPGA	GSPPGQFTG	PPGEPGPGA
201	PGPMGSRGPP	GPFGKNGEDG	EAGKAGRPE	RGAAGPQGAR	GFPGTGLPG	IKGHRGFPL	DGAKGDAGIA	GPKGEAGSPG	EAGTPGAMGA	RGLPGERGRP
301	GPFGPSGARG	NDGNTGAA GP	PGTGPAGPP	GFPGAGAKG	ETGAAAGRGS	EGPQGRGEP	GNP GPAGAAG	PAGGPGSDGT	PGGKGPSGAA	GIAGAPGFPG
401	ARGPPGAPG	NGAPGPKNN	GDPGLQPKG	DAGPKGEPG	VGPQGLAGNA	GEEGKRGARG	EPGAAGPVGP	PGAR GAPGNR	GLPGSDGAPG	PLGAPGERGA
501	NGPMGAQGAS	GESGRPEEPG	LPGSKGMTGA	PGSPGPDGKA	GPSGAPGQDG	RPGPPGAPS	RGAPGVMGFP	GPKGTDGEGS	KPGERGLAGP	TGPLGAPGKD
601	GDIGSPGAPG	PAGPAGEKGE	QGSAGPSGFQ	GLPGAQATG	ETGKPEGQGL	PGEAGPSGSPS	GPRGDRGFPG	ERGSPPGPGT	TGPRGAPGSP	GNDGAKGEPG
701	AAGANGVGP	AGMQMPGER	GAGGMPGARG	DRGDGGPKGP	DGAPGKDGVR	GMTGPIGPPG	PSGAPGDKGE	PGAVGPPGLT	GPRGGPGERG	EVGAPGPAF
801	AGPPGGNGQP	GAKGETGDSG	PKG DAGAPG	AGPVGASGPQ	GPAGATGPKG	ARGGAGPPGA	TGFPGAAGR	GPFGPSGASG	PPGTPGPAK	EGQRGARGEK
901	GPAGRPGEAG	AAGPPPSGSE	RGSPPGDPGL	GLAGAPGPPG	VIGSRGIVGL	PGQRGERGFG	GLAGPSGEPG	KQGPSGPVGE	RGPPGPMGPP	GLSGAPGEAG
1001	REGNPGHDGV	SGRDGPPGPK	GDRGESGNSG	PPGAPGAPGA	PGPIGPSGKT	GDRGEAGPAG	PAGPSGPAGV	RGALGPAGPR	GDRGEAGEAG	ERGMKGHRGL
1101	SGMPGVVPP	GTSGEQGPAG	ATGPAGPRGP	AGSNGSPGKD	GMNGVGPVVG	PPGPRRSGE	MGPAGAPGLP	GPFGPPGSAG	PGEFPLMPQ	QEKGPDLR
1201	GYRAPPDASVR	DRDSFVDTSL	KSLSQKIENI	RSPEGTQANP	ARMCRDLRMC	HPWEKSGSYW	VDPNQGSPLD	AIKVFCNMET	GETCVSPYHS	DIPRKNWYTS
	↓ C-terminal Propeptide									
1301	KNIREKKHYV	FGESMPNGFQ	FQYGS DSDP	EDVNIQLTFM	RLMSNQGSQN	ITYHCKNSIA	YMDETSGNLK	KALLLQGSND	IEIRAEGNSR	FTYRVSEDGC
1401	TSHTGTWGKT	VIDYKTKTKS	RLPIIDIAPM	DIGAPDQDFG	VEYGPVCFI					

FIGURE 5 | Comprehensive PTM map of COL1A1b of WT zebrafish heart ECM, presenting proline/lysine hydroxylation sites and lysine O-glycosylation sites. The representation of PTM sites is similar to the COL1A1a PTM map as shown in top right corner. Peptides identified in proteomics analysis are shown in black color and unidentified peptides are shown in grey color. A total of 96.48% sequence coverage of COL1A1a is detected (considering the matured form of COL1A1b). The signal peptide is 22 amino acids (1–22) long. Sequence alignment matching and previous analysis by Gistelink et al. with human COL1A1 revealed the propeptide cleavage sites. Dark yellow arrows show N terminal (23–150) and C-Term (1,204–1,447) propeptide cleavage sites. Red bold “P” with a blue star represents 3-HyP and red “P” represents 4-HyP. Hydroxylysine is represented with bold “K” and yellow and blue circle represents the lysine O-glycosylation. A summary of these site-specific PTMs of COL1A1b is presented in **Table 1**, and all the PSMs for O-glycosylated lysine and 3-hydroxyproline sites are provided in Supplementary Figures. S2.22–S2.52.

TABLE 2 | Quantitative occupancy (%) of 3-hydroxyproline (3-HyP) sites identified in three different chains of collagen I (COL1A1a, COL1A1b, and COL1A2) from zebrafish cardiac ECM. Results are expressed as mean \pm S.D. Significant statistical differences were estimated by ANOVA (* $p < 0.05$, ^{ns} $p > 0.05$).

3-HyP site	3-Hydroxyproline average occupancy %			
	Sham average	7 DPA Average	14 DPA Average	30 DPA Average
COL1A1a P ⁷⁰⁷	7.55 \pm 4.3	31.49 \pm 38.1	21.21 \pm 7.8	20.50 \pm 1.5 ^{ns}
COL1A1a P ⁸⁶⁹	35.37 \pm 6.9	22.82 \pm 13.3	48.10 \pm 63.8	20.12 \pm 2.9 ^{ns}
COL1A1a P ⁸⁷⁸	28.87 \pm 7.2	17.57 \pm 18.4	27.63 \pm 35.8	12.76 \pm 1.8 ^{ns}
COL1A1a P ¹¹⁴⁸	48.73 \pm 14.7	44.03 \pm 17.0	68.96 \pm 41.1	7.23 \pm 4.6 ^{ns}
COL5A2a P ^{1195, 1201}	1.77 \pm 1.3	23.97 \pm 4.5	55.24 \pm 47.4	30.04 \pm 41.1 ^{ns}
COL1A2 P ⁷¹⁸	0.19 \pm 0.1	0.20 \pm 0.3	0.10 \pm 0.1	10.54 \pm 7.4 ^{ns}
COL1A2 P ⁹²⁵	5.14 \pm 5.7	10.00 \pm 13.2	0.53 \pm 0.7	1.86 \pm 1.8 ^{ns}
COL1A2 P ¹⁰⁶⁶	3.92 \pm 3.4	3.87 \pm 3.1	12.95 \pm 15.7	6.91 \pm 2.0 ^{ns}
COL1A2 P ³⁶¹	0.96 \pm 0.3	1.52 \pm 1.6	4.66 \pm 4.5	1.18 \pm 0.4 ^{ns}
COL1A1b P ⁴⁰⁴	7.88 \pm 1.9	18.28 \pm 12.8	30.08 \pm 6.7	11.95 \pm 7.1 ^{ns}
COL1A1b P ⁵⁵⁴	47.73 \pm 4.0	46.52 \pm 14.8	76.14 \pm 29.7	42.91 \pm 1.3 ^{ns}
COL1A1b P ⁹¹⁴	17.09 \pm 1.4	16.74 \pm 10.6	21.16 \pm 14.0	47.35 \pm 6.2 ^{ns}
COL1A1b P ¹⁰³¹	12.44 \pm 16.1	32.42 \pm 34.0	18.31 \pm 24.2	33.45 \pm 3.6 ^{ns}
COL1A1b P ¹¹⁰⁹	10.36 \pm 7.6	3.93 \pm 1.7	4.76 \pm 6.0	32.64 \pm 18.6 ^{ns}



Surprisingly, this conserved site present in zebrafish (3-HyP¹¹⁴⁸) control cardiac ECM is only about half (48%) 3-hydroxylated (Table 2; Figure 7). Furthermore, the occupancy of this site decreased to ~7% in the 30 DPA regenerated zebrafish heart. The occupancy level of two other 3-HyP sites (P⁸⁶⁹, P⁸⁷⁸) of COL1A1a decreased at 30 DPA compared to control and 7 DPA occupancy (Table 2; Figure 7). Only 3-HyP⁷⁰⁷ occupancy level was increased (7.5–20%) at 30 DPA regenerated zebrafish heart compared to sham. This increased occupancy may favor the assembly of ECM by enhancing collagen fibril formation by forming a water bridge with the available free carbonyl group of the subsequent collagen

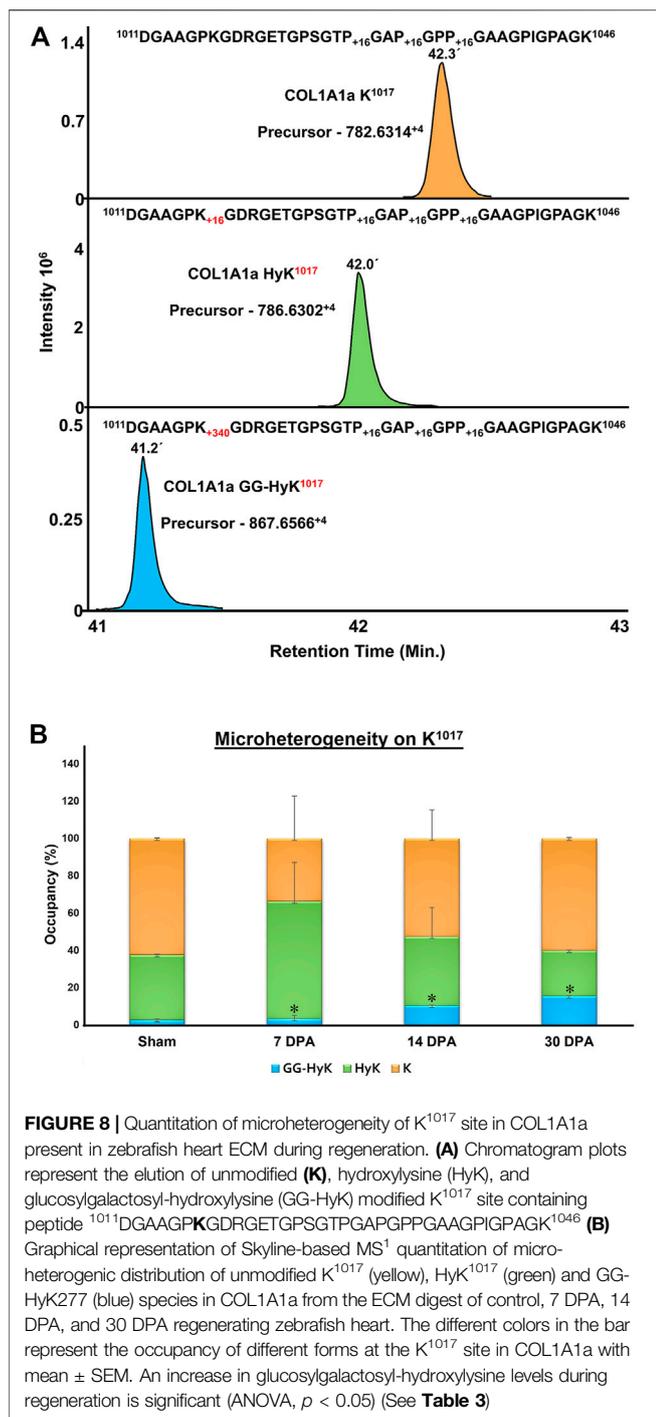
chain. However, a decrease in occupancy of other 3-HyP sites may favor the reduced interaction of deposited collagen molecules in ECM thereby favoring the transition to regeneration from the fibrotic milieu. Although, the exact functional role of these site-specific 3-HyPs needs to be experimentally assessed. A total of five 3-hydroxylation sites (P^{404,554,914,1031,1109}) occupancy levels were quantitated in COL1A1b. Out of these, 4 sites (P^{404,914,1031,1109}) showed the trend of increased occupancy level at 30 DPA regenerated zebrafish heart. We quantitated the occupancy level of 4 sites in COL1A2. Most of these sites were having <10% prolyl-3-hydroxylation. Interestingly, we identified a cluster of 3-HyP sites (P^{1195,1201}) in COL5A2a to be highly 3-hydroxylated (24%) after 7 DPA samples compared to only 1.7% occupancy in control. The occupancy of this 3-HyP cluster showed dynamic changes at 14 DPA (55%) and further reverting to 30% at 30 DPA, in the regenerated zebrafish heart. The site-specific variation in the occupancy level of 3-hydroxyproline could be due to varied gene expression of specific enzymes expressed in the zebrafish heart during post-amputation regeneration. A recent report from Sanchez-Iranzo et al. showed the expression of prolyl-3-hydroxylase-3 (P3H3) and prolyl-3-hydroxylase-4 (P3H4) decreased at 7 days post cryoinjury (log fold change -0.4, -0.2, respectively), in the zebrafish heart. However, in the same RNA seq analysis, the expression level of prolyl-3-hydroxylase-1 (P3H1) and prolyl-3-hydroxylase-2 (P3H2) were found to be increased (log fold change 1.41, 1.75 respectively) in the zebrafish heart at 7 days post-cryoinjury model. Following the regeneration in the same model till 60 days post-cryoinjury, the expression level of P3H2 and P3H4 increased (log fold change 0.94, 1.73 respectively) and the expression of P3H1 and P3H3 decreased (log fold change -0.56, -1.13) respectively highlighting the dynamic changes in the expression of these isoforms of prolyl-3-hydroxylases. However, the enzyme-substrate specificity for these different isoforms of prolyl-3-hydroxylases is not known yet. Thus, ascertaining a direct correlation between the expression level of one enzyme isoform with the substrate (occupancy level of site-specific modifications)

TABLE 3 | Quantitative micro-heterogenic occupancy (%) of O-glycosylated lysine sites identified three different chains of collagen I (COL1A1a, COL1A1b, and COL1A2) from zebrafish cardiac ECM. Results are expressed as mean \pm SEM. Statistically significant differences were estimated by ANOVA ($p < 0.05$, $^{ns} > 0.05$). N/D, not detected; HyK, hydroxylysine; G-HyK, galactosyl-hydroxylysine; GG-HyK, glucosylgalactosylhydroxylysine.

Lysine sites	Modification	Microheterogeneity %			
		Sham	7 DPA	14 DPA	30 DPA
COL1A1a K ²⁶¹	K	36.62 \pm 9.7	43.75 \pm 4.38	46.09 \pm 4.89	32.30 \pm 5.08 ^{ns}
	HyK	38.38 \pm 10.38	34.81 \pm 12.3	21.13 \pm 2.85	45.72 \pm 8.65 ^{ns}
	G-HyK	16.71 \pm 4.9	17.85 \pm 12.0	31.26 \pm 3.01	20.22 \pm 3.01 ^{ns}
	GG-HyK	8.28 \pm 6.5	3.59 \pm 4.3	1.32 \pm 0.97	2.35 \pm 0.57 ^{ns}
COL1A1a K ²⁷⁰	K	50.46 \pm 5.3	32.92 \pm 4.4	26.58 \pm 1.7	46.48 \pm 4.1 ^{ns}
	HyK	34.97 \pm 0.7	37.13 \pm 1.3	46.12 \pm 1.1	40.69 \pm 4.4 ^{ns}
	G-HyK	12.26 \pm 5.0	29.03 \pm 4.4	24.56 \pm 2.7	11.52 \pm 0.4 ^{ns}
	GG-HyK	2.32 \pm 1.1	0.93 \pm 1.3	2.74 \pm 0.2	1.30 \pm 0.7 ^{ns}
COL1A1a K ⁵⁰⁴	K	35.12 \pm 0.2	6.54 \pm 0.3	34.96 \pm 46.9	51.61 \pm 20.2 ^{ns}
	HyK	7.20 \pm 4.4	3.49 \pm 0.3	16.03 \pm 22.4	16.85 \pm 5.2 ^{ns}
	G-HyK	57.68 \pm 4.2	89.97 \pm 0.0	49.01 \pm 69.3	31.54 \pm 15.0 ^{ns}
	GG-HyK	ND	ND	ND	ND
COL1A1a K ⁵⁷⁰	K	59.52 \pm 12.5	24.79 \pm 15.4	17.67 \pm 1.6	42.84 \pm 51.5 ^{ns}
	HyK	14.06 \pm 4.6	6.83 \pm 7.6	4.50 \pm 0.1	0.21 \pm 0.3 ^{ns}
	G-HyK	ND	ND	ND	ND
	GG-HyK	26.42 \pm 17.1	68.39 \pm 7.8	77.83 \pm 1.6	56.95 \pm 51.8 ^{ns}
COL1A1a K ⁶⁹³	K	16.05 \pm 1.8	11.35 \pm 1.4	21.14 \pm 15.4	17.46 \pm 2.3 ^{ns}
	HyK	73.86 \pm 2.2	83.40 \pm 0.1	70.60 \pm 23.8	74.25 \pm 1.7 ^{ns}
	GK	1.34 \pm 0.3	0.25 \pm 0.4	0.56 \pm 0.8	1.47 \pm 0.2 ^{ns}
	GGK	8.75 \pm 0.4	5.00 \pm 1.0	7.69 \pm 9.2	6.82 \pm 3.8 ^{ns}
COL1A1a K ⁸⁴⁶	K	ND	ND	ND	ND
	HyK	ND	ND	ND	ND
	G-HyK	48.69 \pm 29.7	85.04 \pm 19.4	51.02 \pm 9.0	68.87 \pm 5.5 ^{ns}
	GG-HyK	51.31 \pm 29.7	14.96 \pm 19.4	48.98 \pm 9.0	31.13 \pm 5.5 ^{ns}
COL1A1a K ¹⁰¹⁷	K	62.44 \pm 0.3	33.69 \pm 22.8	52.55 \pm 21.8	60.09 \pm 0.5 ^{ns}
	HyK	34.82 \pm 0.4	62.96 \pm 20.9	36.95 \pm 22.1	24.44 \pm 0.2 ^{ns}
	G-HyK	ND	ND	ND	ND
	GG-HyK	2.74 \pm 0.7	3.35 \pm 1.9	10.50 \pm 0.3	15.47 \pm 0.3 *
COL1A1b K ²⁶⁴	K	50.52 \pm 15.5	40.63 \pm 50.7	52.94 \pm 15.8	20.39 \pm 20.2 ^{ns}
	HyK	41.23 \pm 8.2	53.16 \pm 56.0	30.76 \pm 34.6	46.65 \pm 44.7 ^{ns}
	G-HyK	5.19 \pm 4.4	4.64 \pm 3.6	9.75 \pm 10.8	27.51 \pm 20.7 ^{ns}
	GG-HyK	3.05 \pm 2.9	1.57 \pm 1.6	6.54 \pm 7.9	5.45 \pm 3.7 ^{ns}
COL1A1b K ²⁷³	K	ND	ND	ND	ND
	HyK	41.65 \pm 47.5	50.93 \pm 39.7	24.43 \pm 25.6	33.90 \pm 36.0 ^{ns}
	G-HyK	5.58 \pm 4.6	33.38 \pm 20.8	27.55 \pm 13.3	19.99 \pm 10.5 ^{ns}
	GG-HyK	52.77 \pm 42.9	15.69 \pm 19.0	48.02 \pm 38.9	46.11 \pm 25.5 ^{ns}
COL1A1b K ⁸⁴⁹	K	27.06 \pm 1.1	10.35 \pm 13.7	11.17 \pm 2.2	22.95 \pm 3.4 ^{ns}
	HyK	67.46 \pm 1.1	68.53 \pm 16.8	70.77 \pm 9.7	62.89 \pm 0.3 ^{ns}
	G-HyK	3.35 \pm 0.4	6.76 \pm 9.4	12.97 \pm 4.5	9.48 \pm 1.5 ^{ns}
	GG-HyK	2.13 \pm 0.4	14.36 \pm 12.4	5.09 \pm 2.9	4.68 \pm 2.2 ^{ns}
COL1A2 K ²⁵⁴	K	6.48 \pm 3.2	20.75 \pm 14.6	21.61 \pm 14.3	55.69 \pm 18.4 ^{ns}
	HyK	89.24 \pm 1.6	78.24 \pm 13.4	69.76 \pm 2.2	43.90 \pm 18.1 ^{ns}
	G-HyK	0.11 \pm 0.0	0.10 \pm 0.1	0.22 \pm 0.3	0.05 \pm 0.0 ^{ns}
	GG-HyK	4.17 \pm 4.8	0.91 \pm 1.1	8.40 \pm 11.7	0.36 \pm 0.3 ^{ns}
COL1A2 K ⁶⁴⁴	K	ND	ND	ND	ND
	HyK	84.59 \pm 1.8	61.15 \pm 40.2	95.15 \pm 6.5	90.77 \pm 11.1 ^{ns}
	G-HyK	ND	ND	ND	ND
	GGK	15.41 \pm 1.8	38.85 \pm 40.2	4.85 \pm 6.5	9.23 \pm 11.1 ^{ns}

specificity in this dataset is difficult. But, the varied expression of these enzymes during the post-cryoinjury regeneration model in zebrafish heart further supports the specific regulation of site-specific dynamic prolyl-3-hydroxylation changes in collagens

deposited in the zebrafish heart ECM during regeneration. The dynamics of prolyl-3-hydroxylation occupancy in collagen molecules deposited in ECM during zebrafish heart regeneration are reported here for the first time.



3.6 Comprehensive Mapping and Microheterogeneity of Collagen 1 Lysine O-Glycosylation Sites During Regeneration of Zebrafish Heart

In collagens, lysine hydroxylation and glycosylation play important role in collagen assembly and fibril formation (Terajima et al., 2014; Terajima et al., 2016). Musculoskeletal

defects and cerebral small vessel disease are caused due to lack of collagen glycosylation (Geister et al., 2019; Miyatake et al., 2018; Kivirikko et al., 1973). A total of 23 HyK, 9 G-HyK, and 11 GG-HyK sites in zebrafish COL1A1a were identified from wild-type (Sham) zebrafish heart ECM (Figure 3). Classical amino acid analysis has revealed that fibrillar collagens are less glycosylated compared to basement membrane collagen IV (Terajima et al., 2014; Perdivara et al., 2013; Yamauchi and Sricholpech, 2012). However, collagen glycosylation in fibrillar collagen is important for crosslinking mediated fibrillar assembly. Microheterogeneity of lysine sites has been determined in collagens chains (Yamauchi et al., 1982). A single lysine site of a collagen chain could be present in the form of unmodified, hydroxylysine (HyK), galactosyl-hydroxylysine (G-HyK), and glucosyl-galactosyl-hydroxylysine (GG-HyK). This microheterogeneity was found to be highly dynamic in the COL1A1a chain during zebrafish heart regeneration. We detected microheterogeneity on a total of eight lysine sites in the COL1A1a chain. Figure 5 and Figure 6 respectively show the site-specific mapping of lysine O-glycosylation sites in COL1A1b and COL1A2 chains from wild-type zebrafish heart ECM. In COL1A1b a total of 24 HyK, 4 G-HyK (K²⁶⁴, K²⁷³, K⁸⁴⁹, K¹⁰²⁰), and 4 GG-HyK (K²⁶⁴, K²⁷³, K³⁹⁹, K⁸⁴⁹) sites were identified. Three of these sites showed microheterogeneity in glycosylation patterns. Similarly, a total of 21 HyK, 6 G-HyK (K⁷⁴, K¹⁸⁸, K²⁵⁴, K³⁴⁴, K³⁵⁰, K⁶⁴⁴), and 9 GG-HyK (K⁷⁴, K¹⁶⁷, K¹⁸⁸, K²⁵⁴, K²⁹⁹, K⁵⁷⁸, K⁶⁴⁴, K⁶⁴⁷, K⁹³⁵) sites were identified in COL1A2 chain for the first time from zebrafish heart ECM. Four of these sites showed microheterogeneity in different glycosylation moieties (galactosyl or glucosyl-galactosyl). Overall, our analysis mapped the different lysine O-glycosylation sites in three different chains of collagen 1 chains. Collagen 1A1a was found to be less deposited during the regeneration of the zebrafish heart by our as well as Garcia-Puig et al. analysis (Garcia-Puig et al., 2019). Furthermore, we wanted to assess the changes in O-glycosylation of collagen I chains during zebrafish heart regeneration.

In order to assess the changes in the microheterogeneity of lysine O-glycosylation sites on collagen I during regeneration of zebrafish heart, the Skyline-based targeted MS¹ quantitative method was integrated into our pipeline as described previously (Merl-Pham et al., 2019). The relative microheterogeneity analysis of site-specific lysine in three different chains of collagen I was performed. A total of 12 sites of lysine microheterogeneity from three different chains of collagen I was assessed (Table 3). Figure 8 represents an example of a COL1A1a peptide $^{1011}\text{DGAAGPKGDRGETGPSGTPGAPGPPGAAGPIGPAGK}^{1046}$ eluting at 42.3 min during the chromatographic separation as non-glycosylated form with three 4-HyP (P^{1029,1032,1035}) sites. Hydroxylation of K¹⁰¹⁷ resulted in an early elution of this peptide at ~42 min in the C18 column-based separation. Further, the glucosyl-galactosyl-hydroxylysine form of K¹⁰¹⁷ was further eluted earlier (at 41.2 min) than the hydroxylation form. We determined the microheterogeneity of this site (K¹⁰¹⁷) across control, 7-day post-amputation (DPA), 14 DPA, and 30 DPA samples across zebrafish heart regeneration. Interestingly, the glucosyl-galactosyl-hydroxylation of K¹⁰¹⁷ increased significantly (*p* < 0.01) to 15.4% in the 30 DPA regenerated heart as compared to

control (2.74%), (3.35%) 7 DPA. and 14 DPA (10.5%) samples (**Figure 8; Table 3**). The increase in the micro heterogenic distribution of O-glycosylation the in K¹⁰¹⁷ site was evident by a decrease in the HyK level during regeneration (**Figure 8; Table 3**). Using a similar strategy, we found K⁸⁴⁶ (COL1A1a) to be specifically present in a full glycosylated form (either G-HyK or GG-HyK). The GG-HyK⁸⁴⁶ was decreased at 30 DPA regenerated zebrafish heart. We also quantitated the microheterogeneity of three lysine sites (K^{264,273,849}) in COL1A1b and 2 sites in COL1A2 (K^{254,644}) (see **Table 3**). The micro-heterogenic distribution of G-HyK²⁶⁴ of COL1A1b was found to have an increasing trend at 30 DPA. The changes in the level of hydroxylation and O-glycosylation of lysine sites present in collagen I could be potentially due to the gene expression changes of lysyl-hydroxylases and glycosyltransferases during zebrafish heart regeneration. In a similar model of zebrafish heart regeneration, proteomic analysis performed by Ma *et al.* (2018) revealed a significant elevation of LH2 levels at 2 DPA (**Supplementary Figure S4**). However, at 14 DPA the LH2 protein level stabilizes to sham levels. Similar oscillatory gene expressional observations of the genes (PLOD1a, PLOD2, and PLOD3) coding for different isoforms of lysyl-hydroxylases has been documented in the zebrafish heart cryoinjury model (Sánchez-Iranzo *et al.*, 2018). Global gene expression analysis revealed increased expression of these three genes in the zebrafish heart at 7 days post-injury. However, the level of gene expression of PLOD1a, PLOD2, and PLOD3 decreases at 60 days post injury in the regenerated heart (Sánchez-Iranzo *et al.*, 2018). These micro-heterogenic site-specific glycosylation profiles point towards the differences present in three different chains of collagen I that may play an important role in the ECM remodeling mediated zebrafish heart regeneration.

4 CONCLUSION

In conclusion, our analysis presents the first comprehensive mapping and dynamics of site-specific collagen PTMs from the crude extracellular matrix proteins in wild-type and regenerating zebrafish hearts. We optimized an in-house dual database search engine-based mass-spectrometry data analysis strategy to efficiently identify long post-translationally modified collagen peptides from crude ECM analysis. We identified the highest number of collagen chains (36) present in the zebrafish heart ECM. Our analysis revealed that the expression of collagens gets altered during regeneration. Global characterization of zebrafish heart ECM revealed the identification of site-specific PTMs (proline hydroxylation, lysine hydroxylation, and lysine glycosylation sites) in 23 different collagen chains. These are the first comprehensive site-specific collagen PTM maps of zebrafish cardiac ECM highlighting some novel sites. Further, we present the first

comprehensive PTM maps of three chains (COL1A1a, COL1A1b, and COL1A2) of collagen 1. Our analysis also identified the conserved 3-hydroxyproline sites occurring in the COL1A1 chain in humans, mice, and zebrafish heart ECM. This finding highlights the importance of these conserved 3-HyP sites involved in the basic function of collagens in humans, mice, and zebrafish myocardium. Moreover, we established that site-specific PTM (Prolyl-3-hydroxylations and lysyl O-glycosylation) changes in collagen chains are correlated with the regeneration process. Glycosylation on one site (GG-HyK¹⁰¹⁷) of COL1A1a is found to be significantly increased during the regeneration process. Taken together, our analysis will widen new avenues to dig deeper in understanding the role of collagen PTMs during zebrafish heart regeneration.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. The datasets PXD011627 and PXD010092 can be accessed at: <https://www.proteomexchange.org/>.

AUTHOR CONTRIBUTIONS

VS, SS, and TB worked on analyzing, interpreting, and summarizing the study. VS and SS prepared the initial drafts of the manuscript. VS prepared all the figures. TB conceptualize the overall structure of the study and finalized the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge Prof. Angel Raya for sharing the raw mass-spectrometry data (*.d files) for the regenerating samples. Authors also acknowledge the HPC facility of IIT-Mandi for performing intensive database searches. The Seed grant (IITM/SG/TB/75) from IIT-Mandi and the Indian Council of Medical Research (ICMR) funded “High-Resolution Plasma Proteomic and Lipidomic Analyses for Fibrosis-Related Metabolic Assessment In Dilated Cardiomyopathy (DCM) Patients in India: A Multi-center Based Study (NCD/Adhoc/40/2020–2021)” to TB is acknowledged for this work. VS and SS also acknowledge the HTRA fellowship (MHRD, Govt. of India) for the doctoral and post-graduate program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2022.892763/full#supplementary-material>

REFERENCES

- Barallobre-Barreiro, J., Radovits, T., Fava, M., Mayr, U., Lin, W.-Y., Ermolaeva, E., et al. (2021). Extracellular Matrix in Heart Failure: Role of ADAMTS5 in Proteoglycan Remodeling. *Circulation* 144 (25), 2021–2034. doi:10.1161/circulationaha.121.055732
- Basak, T., Vega-Montoto, L., Zimmerman, L. J., Tabb, D. L., Hudson, B. G., and Vanacore, R. M. (2016). Comprehensive Characterization of Glycosylation and Hydroxylation of Basement Membrane Collagen IV by High-Resolution Mass

- Spectrometry. *J. Proteome Res.* 15 (1), 245–258. doi:10.1021/acs.jproteome.5b00767
- Bootorabi, F., Manouchehri, H., Changizi, R., Barker, H., Palazzo, E., Saltari, A., et al. (2017). Zebrafish as a Model Organism for the Development of Drugs for Skin Cancer. *Int. J. Mol. Sci.* 18 (7), 1550. doi:10.3390/ijms18071550
- Cabral, W. A., Chang, W., Barnes, A. M., Weis, M., Scott, M. A., Leikin, S., et al. (2007). Prolyl 3-hydroxylase 1 Deficiency Causes a Recessive Metabolic Bone Disorder Resembling Lethal/severe Osteogenesis Imperfecta. *Nat. Genet.* 39 (3), 359–365. doi:10.1038/ng1968
- Chen, W. C., Wang, Z., Missinato, M. A., Park, D. W., Long, D. W., Liu, H. J., et al. (2016). Decellularized Zebrafish Cardiac Extracellular Matrix Induces Mammalian Heart Regeneration. *Sci. Adv.* 2 (11), e1600844. doi:10.1126/sciadv.1600844
- Cowling, R. T., Kupsky, D., Kahn, A. M., Daniels, L. B., and Greenberg, B. H. (2019). Mechanisms of Cardiac Collagen Deposition in Experimental Models and Human Disease. *Transl. Res.* 209, 138–155. doi:10.1016/j.trsl.2019.03.004
- Cutie, S., and Huang, G. N. (2021). Vertebrate Cardiac Regeneration: Evolutionary and Developmental Perspectives. *Cell Regen.* 10 (1), 6–10. doi:10.1186/s13619-020-00068-y
- Di Lullo, G. A., Sweeney, S. M., Körkkö, J., Ala-Kokko, L., and San Antonio, J. D. (2002). Mapping the Ligand-Binding Sites and Disease-Associated Mutations on the Most Abundant Protein in the Human, Type I Collagen. *J. Biol. Chem.* 277 (6), 4223–4231. doi:10.1074/jbc.m110709200
- Duran, I., Csukasi, F., Taylor, S. P., Krakow, D., Becerra, J., Bombarely, A., et al. (2015). Collagen Duplicate Genes of Bone and Cartilage Participate during Regeneration of Zebrafish Fin Skeleton. *Gene Expr. Patterns* 19 (1–2), 60–69. doi:10.1016/j.gep.2015.07.004
- Fietzek, P. P., and Kühn, K. (1975). Information Contained in the Amino Acid Sequence of the alpha(I)-chain of Collagen and its Consequences upon the Formation of the Triple Helix, of Fibrils and Crosslinks. *Mol. Cell Biochem.* 8 (3), 141–157. doi:10.1007/bf01792765
- Frangogiannis, N. G. (2019). The Extracellular Matrix in Ischemic and Nonischemic Heart Failure. *Circ. Res.* 125 (1), 117–146. doi:10.1161/circresaha.119.311148
- García-Puig, A., Mosquera, J. L., Jiménez-Delgado, S., García-Pastor, C., Jorba, I., Navajas, D., et al. (2019). Proteomics Analysis of Extracellular Matrix Remodeling during Zebrafish Heart Regeneration. *Mol. Cell. Proteomics* 18 (9), 1745–1755. doi:10.1074/mcp.ra118.001193
- Geister, K. A., Lopez-Jimenez, A. J., Houghtaling, S., Ho, T. H., Vanacore, R., and Beier, D. R. (2019). Loss of Function of Colgalt1 Disrupts Collagen Post-translational Modification and Causes Musculoskeletal Defects. *Dis. Model Mech.* 12 (6), dmm037176. doi:10.1242/dmm.037176
- Gistelink, C., Gioia, R., Gagliardi, A., Tonelli, F., Marchese, L., Bianchi, L., et al. (2016). Zebrafish Collagen Type I: Molecular and Biochemical Characterization of the Major Structural Protein in Bone and Skin. *Sci. Rep.* 6, 21540. doi:10.1038/srep21540
- González, A., Schelbert, E. B., Diez, J., and Butler, J. (2018). Myocardial Interstitial Fibrosis in Heart Failure: Biological and Translational Perspectives. *J. Am. Coll. Cardiol.* 71 (15), 1696–1706. doi:10.1016/j.jacc.2018.02.021
- Grafe, I., Yang, T., Alexander, S., Homan, E. P., Lietman, C., Jiang, M. M., et al. (2014). Excessive Transforming Growth Factor- β Signaling Is a Common Mechanism in Osteogenesis Imperfecta. *Nat. Med.* 20 (6), 670–675. doi:10.1038/nm.3544
- Hennet, T. (2019). Collagen Glycosylation. *Curr. Opin. Struct. Biol.* 56, 131–138. doi:10.1016/j.sbi.2019.01.015
- Hudson, D. M., and Eyre, D. R. (2013). Collagen Prolyl 3-hydroxylation: a Major Role for a Minor Post-translational Modification? *Connect. Tissue Res.* 54 (4–5), 245–251. doi:10.3109/03008207.2013.800867
- Hudson, D. M., Joeng, K. S., Werther, R., Rajagopal, A., Weis, M., Lee, B. H., et al. (2015). Post-translationally Abnormal Collagens of Prolyl 3-hydroxylase-2 Null Mice Offer a Pathobiological Mechanism for the High Myopia Linked to Human LEPREL1 Mutations. *J. Biol. Chem.* 290 (13), 8613–8622. doi:10.1074/jbc.m114.634915
- Jürgensen, H. J., Madsen, D. H., Ingvarsen, S., Melander, M. C., Gårdsvoll, H., Pathy, L., et al. (2011). A Novel Functional Role of Collagen Glycosylation: Interaction with the Endocytic Collagen Receptor uparap/ENDO180. *J. Biol. Chem.* 286 (37), 32736–32748. doi:10.1074/jbc.M111.266692
- Kivirikko, K. I., Ryhanen, L., Anttinen, H., Bornstein, P., and Prockop, D. J. (1973). Hydroxylation of Lysyl Residues in Collagen by Protocollagen Lysyl Hydroxylase *In Vitro*. *Biochemistry* 12 (24), 4966–4971. doi:10.1021/bi00748a023
- Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017). MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry-Based Proteomics. *Nat. Methods* 14 (5), 513–520. doi:10.1038/nmeth.4256
- Li, L., Zhao, Q., and Kong, W. (2018). Extracellular Matrix Remodeling and Cardiac Fibrosis. *Matrix Biol.* 68–69, 490–506. doi:10.1016/j.matbio.2018.01.013
- Ma, D., Tu, C., Sheng, Q., Yang, Y., Kan, Z., Guo, Y., et al. (2018). Dynamics of Zebrafish Heart Regeneration Using an HPLC-ESI-MS/MS Approach. *J. Proteome Res.* 17 (3), 1300–1308. doi:10.1021/acs.jproteome.7b00915
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. *Bioinformatics* 26 (7), 966–968. doi:10.1093/bioinformatics/btq054
- Merl-Pham, J., Basak, T., Knüppel, L., Ramanujam, D., Athanason, M., Behr, J., et al. (2019). Quantitative Proteomic Profiling of Extracellular Matrix and Site-specific Collagen Post-translational Modifications in an *In Vitro* Model of Lung Fibrosis. *Matrix Biol. Plus* 1, 100005. doi:10.1016/j.mbplus.2019.04.002
- Missinato, M. A., Tobita, K., Romano, N., Carroll, J. A., and Tsang, M. (2015). Extracellular Component Hyaluronic Acid and its Receptor Hmhr Are Required for Epicardial EMT during Heart Regeneration. *Cardiovasc Res.* 107 (4), 487–498. doi:10.1093/cvr/cvv190
- Miyatake, S., Schneeberger, S., Koyama, N., Yokochi, K., Ohmura, K., Shiina, M., et al. (2018). Biallelic COLGALT1 Variants Are Associated with Cerebral Small Vessel Disease. *Ann. Neurol.* 84 (6), 843–853. doi:10.1002/ana.25367
- Montgomery, N. T., Zientek, K. D., Pokidysheva, E. N., and Bächinger, H. P. (2018). Post-translational Modification of Type IV Collagen with 3-hydroxyproline Affects its Interactions with Glycoprotein VI and Nidogen 1 and 2. *J. Biol. Chem.* 293 (16), 5987–5999. doi:10.1074/jbc.ra117.000406
- Morello, R., Bertin, T. K., Chen, Y., Hicks, J., Tonachini, L., Monticone, M., et al. (2006). CRTAP Is Required for Prolyl 3-Hydroxylation and Mutations Cause Recessive Osteogenesis Imperfecta. *Cell* 127 (2), 291–304. doi:10.1016/j.cell.2006.08.039
- Myllyharju, J. (2005). Intracellular Post-Translational Modifications of Collagens. *Top. Curr. Chem.* 247, 115–147. doi:10.1007/b103821
- Nauroy, P., Hughes, S., Naba, A., and Ruggiero, F. (2018). The In-Silico Zebrafish Matrixome: A New Tool to Study Extracellular Matrix Gene and Protein Functions. *Matrix Biol.* 65, 5–13. doi:10.1016/j.matbio.2017.07.001
- Padmanabhan Iyer, R., Chiao, Y. A., Flynn, E. R., Hakala, K., Cates, C. A., Weintraub, S. T., et al. (2016). Matrix Metalloproteinase-9-dependent Mechanisms of Reduced Contractility and Increased Stiffness in the Aging Heart. *Prot. Clin. Appl.* 10 (1), 92–107. doi:10.1002/prca.201500038
- Pedrioli, P. G. A. (2010). Trans-proteomic Pipeline: a Pipeline for Proteomic Analysis. *Methods Mol. Biol. Clift. NJ* 604, 213–238. doi:10.1007/978-1-60761-444-9_15
- Perdivara, I., Yamauchi, M., and Tomer, K. B. (2013). Molecular Characterization of Collagen Hydroxylysine O-Glycosylation by Mass Spectrometry: Current Status. *Aust. J. Chem.* 66 (7), 760–769. doi:10.1071/ch13174
- Pokidysheva, E., Boudko, S., Vranka, J., Zientek, K., Maddox, K., Moser, M., et al. (2014). Biological Role of Prolyl 3-hydroxylation in Type IV Collagen. *Proc. Natl. Acad. Sci. U.S.A.* 111 (1), 161–166. doi:10.1073/pnas.1307597111
- Risteli, J., and Kivirikko, K. I. (1974). Activities of Prolyl Hydroxylase, Lysyl Hydroxylase, Collagen Galactosyltransferase and Collagen Glucosyltransferase in the Liver of Rats with Hepatic Injury. *Biochem. J.* 144 (1), 115–122. doi:10.1042/bj1440115
- Salo, A. M., Cox, H., Farndon, P., Moss, C., Grindulis, H., Risteli, M., et al. (2008). A Connective Tissue Disorder Caused by Mutations of the Lysyl Hydroxylase 3 Gene. *Am. J. Hum. Genet.* 83 (4), 495–503. doi:10.1016/j.ajhg.2008.09.004
- Salo, A. M., and Myllyharju, J. (2021). Prolyl and Lysyl Hydroxylases in Collagen Synthesis. *Exp. Dermatol.* 30 (1), 38–49. doi:10.1111/exd.14197
- Sánchez-Iranzo, H., Galardi-Castilla, M., Sanz-Morejón, A., González-Rosa, J. M., Costa, R., Ernst, A., et al. (2018). Transient Fibrosis Resolves via Fibroblast Inactivation in the Regenerating Zebrafish Heart. *Proc. Natl. Acad. Sci. U. S. A.* 115 (16), 4188–4193. doi:10.1073/pnas.1716713115

- Schegg, B., Hülsmeier, A. J., Rutschmann, C., Maag, C., and Hennet, T. (2009). Core Glycosylation of Collagen Is Initiated by Two $\beta(1 \rightarrow O)$ Galactosyltransferases. *Mol. Cell Biol.* 29 (4), 943–952. doi:10.1128/mcb.02085-07
- Sciatti, L., Chiapparino, A., De Giorgi, F., Fumagalli, M., Khoraiuli, L., Nergadze, S., et al. (2018). Molecular Architecture of the Multifunctional Collagen Lysyl Hydroxylase and Glycosyltransferase LH3. *Nat. Commun.* 99 (11), 3163. doi:10.1038/s41467-018-05631-5
- Shao, X., Taha, I. N., Clauser, K. R., Gao, Y., and Naba, A. (2020). MatrisomeDB: the ECM-Protein Knowledge Database. *Nucleic acids Res.* 48 (D1), D1136–D1144. doi:10.1093/nar/gkz849
- Sipilä, L., Ruotsalainen, H., Sormunen, R., Baker, N. L., Lamandé, S. R., Vapola, M., et al. (2007). Secretion and Assembly of Type IV and VI Collagens Depend on Glycosylation of Hydroxylysines. *J. Biol. Chem.* 282 (46), 33381–33388. doi:10.1074/jbc.M704198200
- Stawikowski, M. J., Aukszi, B., Stawikowska, R., Cudic, M., and Fields, G. B. (2014). Glycosylation Modulates Melanoma Cell $\alpha 2\beta 1$ and $\alpha 3\beta 1$ Integrin Interactions with Type IV Collagen. *J. Biol. Chem.* 289 (31), 21591–21604. doi:10.1074/jbc.m114.572073
- Tabb, D. L., Fernando, C. G., and Chambers, M. C. (2007). MyriMatch: Highly Accurate Tandem Mass Spectral Peptide Identification by Multivariate Hypergeometric Analysis. *J. Proteome Res.* 6 (2), 654–661. doi:10.1021/pr0604054
- Terajima, M., Perdivara, I., Sricholpech, M., Deguchi, Y., Pleshko, N., Tomer, K. B., et al. (2014). Glycosylation and Cross-Linking in Bone Type I Collagen. *J. Biol. Chem.* 289 (33), 22636–22647. doi:10.1074/jbc.m113.528513
- Terajima, M., Taga, Y., Chen, Y., Cabral, W. A., Hou-Fu, G., Srisawasdi, S., et al. (2016). Cyclophilin-B Modulates Collagen Cross-Linking by Differentially Affecting Lysine Hydroxylation in the Helical and Telopeptidyl Domains of Tendon Type I Collagen. *J. Biol. Chem.* 291 (18), 9501–9512. doi:10.1074/jbc.m115.699470
- Travers, J. G., Kamal, F. A., Robbins, J., Yutzey, K. E., and Blaxall, B. C. (2016). Cardiac Fibrosis. *Circ. Res.* 118 (6), 1021–1040. doi:10.1161/circresaha.115.306565
- Van Huizen, N. A., Burgers, P. C., Saintmont, F., Brocogens, P., Gerbaux, P., Stingl, C., et al. (2019). Identification of 4-Hydroxyproline at the Xaa Position in Collagen by Mass Spectrometry. *J. Proteome Res.* 18 (5), 2045–2051. doi:10.1021/acs.jproteome.8b00930
- Weis, M. A., Hudson, D. M., Kim, L., Scott, M., Wu, J.-J., and Eyre, D. R. (2010). Location of 3-Hydroxyproline Residues in Collagen Types I, II, III, and V/XI Implies a Role in Fibril Supramolecular Assembly. *J. Biol. Chem.* 285 (4), 2580–2590. doi:10.1074/jbc.m109.068726
- Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I., and Birk, D. E. (2004). Type V Collagen Controls the Initiation of Collagen Fibril Assembly. *J. Biol. Chem.* 279 (51), 53331–53337. doi:10.1074/jbc.m409622200
- Willyard, C. (2018). Unlocking the Secrets of Scar-free Skin Healing. *Nature* 563 (7732), S86–S88. doi:10.1038/d41586-018-07430-w
- Yamauchi, M., Noyes, C., Kuboki, Y., and Mechanic, G. L. (1982). Collagen Structural Microheterogeneity and a Possible Role for Glycosylated Hydroxylysine in Type I Collagen. *Proc. Natl. Acad. Sci. U.S.A.* 79, 7684–7688. doi:10.1073/pnas.79.24.7684
- Yamauchi, M., and Sricholpech, M. (2012). Lysine Post-translational Modifications of Collagen. *Essays Biochem.* 52 (1), 113–133. doi:10.1042/bse0520113

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Saroji, Srivastava and Basak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.