Novel bioinformatics tools for analysis of gene expression in the skate, *Leucoraja erinacea*

Taylor Hersh[1], Ben King[2] and Bram V. Lutton[3]
[1]Carnegie Mellon University, Pittsburgh, PA 15213
[2]Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672
[3]Endicott College, Beverly, MA 01915

The skate, *Leucoraja erinacea*, is emerging as a potentially useful model organism for studies of stem cell production and mobilization, as well as tissue regeneration. The benefit of using *L. erinacea* in these studies is that these organisms are cartilaginous, which allows for easier access to stem cell niches and more accurate findings. This study further annotates genes within the hematopoietic niches of *L. erinacea* using newly available sequence data. Within the epigonal organ of the skate, we have found evidence of both CXCR4 and CXCL12 expression, which in mammalian bone marrow are known to play significant roles in the regulation of hematopoietic stem cell activity.

Mammalian bone marrow is composed of both vascular (blood vessel) and endosteal (bone) niches.[5] Within these two niches, stem cell activation and the accompanying leukocyte differentiation begins and leukocytes are mobilized into the blood. The importance of the vascular niche in comparison with the endosteal niche has been a controversial topic in recent years, with scientists sometimes placing more value on the endosteal niche.[2] However, both niches are now considered to have significant autonomous importance.[5] Elasmobranchs are cartilaginous fishes, and therefore have hematopoietic tissues possessing only the vascular niche. These animals and their ancestors have continued to survive without the endosteum for more than 400 million years. The epigonal organ, one of these vascular hematopoietic niches, is functionally similar to mammalian bone marrow in that it is responsible for production and differentiation of leukocytes from hematopoietic stem cells.[1]

The epigonal organ has a direct association with the gonads as it is physically bound to the germinal bed via cellular and vascular connections in *L.erinacea*.[4] In females, this structure is referred to as the epigonal-ovary complex. As *L. erinacea* mature from non-reproductively active to reproductively active states, cellular proliferation and angiogenesis within the epigonal-ovary complex increase significantly.[4] These observations indicate that factors secreted during reproductive activity directly or indirectly stimulate activation of hematopoietic stem cells, which leads to leukocyte mobilization and angiogenesis. According to studies published over the past decade using mammalian models, these two processes appear to be linked by cellular and molecular interactions within the hematopoietic environment.[6]

Hematopoietic stem cells in mammals are encased within the solid bone matrix, which makes accessing them without damaging the extremely fragile and complex marrow a daunting task. The lack of endosteum characteristic of the epigonal organ permits greater ease of analysis since decalcification, which may limit accuracy in histochemical preparations, is unnecessary. Thus, while mice and zebrafish are the primary models used in studies of hematopoiesis and angiogenesis, we believe the epigonal organ of elasmobranchs will provide novel information for studies of cell and tissue regeneration. However, significant limitations also exist in the elasmobranch model because no cell differentiation labels exist. Our aim is to identify candidate genes to begin developing micro RNAs, recombinant proteins, and antibodies for functional analyses.

The goal of this study was to look for the expression of genes with known hematopoietic and angiogenic functions in mammalian bone marrow in *L. erinacea*. One receptor-ligand pair that plays crucial roles in this tissue is chemokine receptor (CXCR) 4 and its cognate ligand, CXCL12.[7] CXCR4 is expressed on hematopoietic stem cells, and the binding between CXCR4 and CXCL12 maintains hematopoietic stem cells in the niche.[7] In 2012, our lab demonstrated CXCR4 expression in the epigonal organs of both reproductively active and non-reproductively active *L. erinacea*.[3] Utilizing the bioinformatics expertise of collaborators at the MDIBL, who have assembled large genome and transcriptome data sets for *L. erinacea*, the purpose of the current study was to demonstrate CXCL12 expression for the first time in the epigonal organ of this species.

Primers designed to validate the expression of the CXCR4 and CXCL12 genes in the epigonal organ of *L. erinacea* were generated using bioinformatics analysis of existing (http://skatebase.org) and novel (provided by collaborators at the MDIBL) sequence data. PCR verified CXCR4 expression in the epigonal organ of both non-

reproductively active and reproductively active *L. erinacea*. In addition, CXCL12 expression was demonstrated in the epigonal organ of a reproductively active *L. erinacea*. CXCL12 expression has not yet been observed in the epigonal organ of non-reproductively active *L. erinacea*. This may signify that the CXCR4-CXCL12 interaction is restricted to times when *L. erinacea* are reproductively active, the physiological condition in which we observe significantly increased hematopoietic stem cell mobilization and angiogenesis.[4] Using quantitative PCR to assess the epigonal organs of a greater number of reproductively active and non-reproductively active *L. erinacea* in future studies will likely elucidate the significance of these findings.

Sequencing was carried out for both CXCR4 and CXCL12. The data for each gene was translated into amino acids and then entered into the NCBI protein BLAST (blastp) program. The default BLOSOM62 scoring matrix was used for these alignments. Tables 1 and 2 show some of the highest scoring alignments in different vertebrate species that were returned for CXCL12 and CXCR4, respectively. The values are listed in order of descending score; a higher score signifies decreased divergence from the query sequence (the *L. erinacea* sequence for each gene). The E value reflects the statistical significance of each alignment.

Table 1. NCBI Protein BLAST Alignment Results for CXCL12

Species Name	Max Score	Sequence Accession Number	Query Cover	E value	Identity to *L. erinacea*
C. milii	87.4	AFM88233.1	98%	2e-20	47%
L. chalumnae	79.3	XP_006009497.1	100%	3e-17	43%
X. laevis	70.9	NP_001083632.1	100%	5e-14	38%
C. lupus familiaris	68.2	NP_001121569.1	100%	5e-13	41%
D. rerio	63.9	NP_932334.1	100%	2e-11	40%
T. rubripes	60.8	XP_003971985.1	100%	5e-10	40%
H. sapiens	60.5	NP_954637.1	100%	4e-10	41%
M. musculus	59.7	NP_068350.1	75%	1e-09	44%
F. catus	59.3	BAA28601.1	75%	1e-09	44%
G. gallus	56.6	AAR91696.1	79%	2e-08	42%

Table 2. NCBI Protein BLAST Alignment Results for CXCR4

Species Name	Max Score	Sequence Accession Number	Query Cover	E value	Identity to *L. erinacea*
C. milii	473	AFM88361.1	99%	2e-163	69%
X. laevis	402	NP_001080681.1	99%	9e-136	59%
G. gallus	402	AAG09054.1	95%	6e-136	61%
L. chalumnae	402	XP_006006253.1	95%	5e-136	61%
D. rerio	402	NP_571957.2	100%	3e-136	60%
H. sapiens	397	CAA12166.1	100%	4e-134	59%
C. lupus familiaris	395	NP_001041491.1	96%	2e-133	59%
M. musculus	383	CAB02202.1	99%	2e-128	58%
T. rubripes	365	XP_003962124.1	99%	3e-121	54%
P. marinus	251	AAO21209.1	86%	5e-77	42%

Using alignments obtained from BLAST (including additional alignments not represented in Tables 1 and 2) and Jalview alignment software, multiple sequence alignments were created that compared known CXCR4 and CXCL12 gene sequences in other species to sequence data for *L. erinacea* (manual annotation assistance from Dr. Dannie Durand, Carnegie Mellon University). These alignments provided strong confirmation the genes identified using semi-quantitative PCR were indeed our target chemokine genes. For example, the multiple sequence alignment for CXCL12 (Figure 1). Due to the size of the alignment for CXCR4, it is not shown.

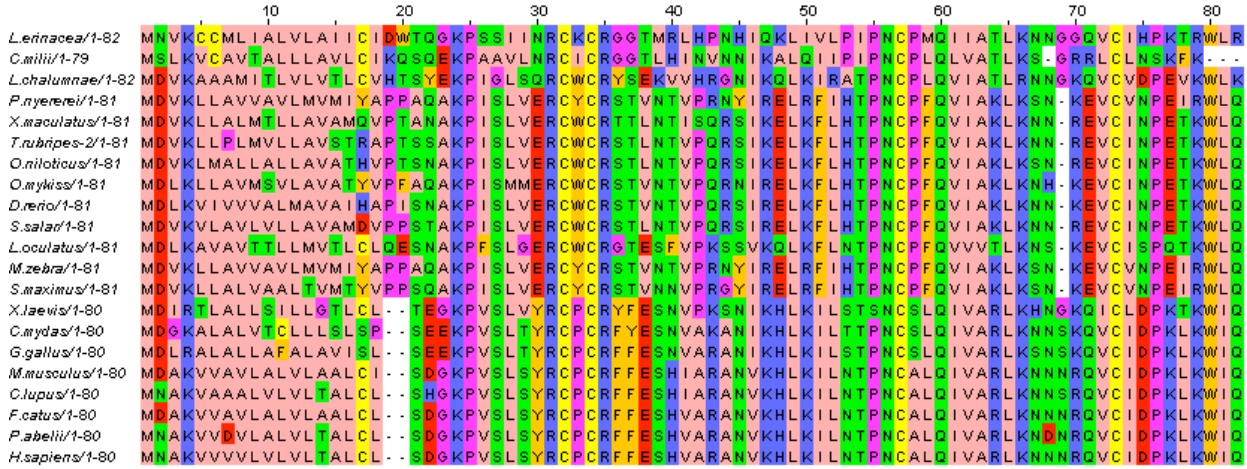

Figure 1. Multiple sequence alignment comparing the *L. erinacea* CXCL12 sequence to the CXCL12 sequence in other species.

Using the multiple sequence alignments, coupled with UniProt data on the human version of the genes, some preliminary analyses regarding conservation of structure and function were conducted. For CXCL12, all the species examined had a conserved KP motif from position 24-25, which suggests this motif has been conserved throughout evolution. Within the human CXCL12 gene, this region acts as a receptor activation motif. Additionally, the amino acid sequence PNC is conserved in all of the species (with the exception of *X. laevis*) at positions 55-57. This region creates a turn in the secondary structure of the human CXCL12 sequence, and its conservation across many species suggests that some aspects of CXCL12 secondary structure are conserved even as far back as elasmobranchs. Conservation of cysteines, important for disulfide bonding, is also apparent when assessing the alignments for both genes. The multiple sequence alignment for CXCR4 also showed conservation of certain regions across species (data not shown). For example, an arginine cage DRY motif at positions 133-135 was uniformly conserved. Within human CXCR4, this region plays a role in activation of the receptor. Additionally, a stretch of 17 amino acids known to create an alpha helix in the human CXCR4 gene was found in all species examined.

In conclusion, CXCR4 expression was verified in both reproductively active and non-reproductively active *L. erinacea*. In addition, CXCL12 expression was demonstrated for the first time in reproductively active *L. erinacea*. Furthermore, primers have been designed for seventeen other genes known to be involved in hematopoietic and angiogenic signaling in mammals. Future studies will continue to characterize the unique vascular hematopoietic niches of *L. erinacea* and the suitability of this model organism for cellular regeneration studies will continue to be improved.

This research was supported by an NSF REU grant (NSF DBI-0453391). We also acknowledge the assistance of Chris Smith (MDIBL), Dr. Dannie Durand (Carnegie Mellon University), Pieter Spealman (Carnegie Mellon University) and Dr. James Sulikowski (University of New England).

1. **Anderson MK, Pant R, Miracle AL, Sun X, Luer CA, Walsh CJ, Telfer JC, Litman GW, Rothenberg EV**. Evolutionary origins of lymphocytes: ensembles of T cell and B cell transcriptional regulators in a cartilaginous fish. *J Immunol*. 172:5851-5860, 2004.
2. **Kiel MJ, Iwashita T, Yilmaz OH, Morrison SJ**. Spatial differences in hematopoiesis but not in stem cells indicated a lack of regional patterning in definitive hematopoietic stem cells. *Dev. Biol*. 283:29-39, 2005.
3. **Lupica NV, Cacciatore C, Lutton BV**. Chemokine Receptor 4 Expression in the Little Skate. *The Bulletin*. 51:47-48, 2012.
4. **Lutton BV, Callard IP**. Effects of reproductive activity and sex hormones on apoptosis in the epigonal organ of the skate (*Leucoraja erinacea*). *Gen Comp Endocrinol*. 154:75-84, 2007.
5. **Morrison SJ, Scadden DT**. The bone marrow niche for haematopoietic stem cells. *Nature*, 505:327-34, 2014.
6. **Ding L, Morrison SJ**. Hematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature*. 495:231-236, 2013.
7. **Tanegashima K, Suzuki K, Nakayama Y, Tsuji K, Shigenaga A, Otaka A, Hara T**. CXCL14 is a natural inhibitor of the CXCL12-CXCR4 signaling axis. *FEBS Lett*. 587:1731-5, 2013.