

고래의 게놈에서 hypoxia-inducible factor binding site의 예측과 target gene에 대한 분석

Prediction of Hypoxia-inducible Factor Binding Site in Whale Genome and Analysis of Target Genes Regulated by Predicted Sites

임형순*, 이재학

Hyung-Soon Yim*, Jae-Hak Lee

한국해양과학기술원, 안산, 15627, 대한민국

Korea Institute of Ocean Science & Technology, Ansan 15627, Republic of Korea

(Received 26 October 2015, Revised 2 November 2015, Accepted 3 November 2015)

Abstract Whales are marine mammals that are fully adapted to aquatic environment. Whales breathe by lungs so they require adaptive system to low oxygen concentration (hypoxia) while deep and prolonged diving. However, the study for the molecular mechanism underlying cetacean adaptation to hypoxia has been limited. Hypoxia-inducible factor (HIF) is the central transcription factor that regulates hypoxia-related gene expression. Here we identified HIF-binding sites in whale genome by phylogenetic footprinting and analyzed HIF-target genes to understand how whales cope with hypoxia. By comparison with the HIF-target genes of terrestrial mammals, it was suggested that whales may retain unique adaptation mechanisms to hypoxia.

Keywords : whale, hypoxia, HIF (hypoxia-inducible factor), HIF-target gene, adaptation mechanism

고래목은 수염고래 (baleen whale), 돌고래 (dolphin), 상괘이 (porpoise)가 포함된 분류군으로 진화단계에 있어서 매우 특이한 위치에 놓여 있다. 생물의 진원은 바다이며 진화를 거쳐 육상에서 적응하여 생활하는 것이 일반적이나 고래는 육상에서 서식하던 포유류가 해양 환경에 적응하여 다시 해양으로 서식지를 옮겨간 것으로 알려져 있다. 카슈미르 (Kashmir) 지방에서 발견된 *Indohyus* 화석의 해부적 특성과 치아의 동위원소 연구를 통해 이 동물이 고래의 조상으로 간주된다 [11]. *Indohyus*는 사슴과 같이 생긴 우제류 (Artiodactyla,

even-toed mammal)이며 수륙을 오가며 생활했을 것으로 보고 있다. 고래는 해양에서 생활하지만 포유류이기 때문에 허파 호흡, 항온, 모유 수유 등의 포유류의 기본적인 특성을 지니고 있다. 따라서 생존을 위해서는 폐에 산소가 지속적으로 공급이 되어야 하나 고래는 장시간 잠수하는 능력을 가지고 있어서 저산소 (hypoxia)에 적응하기 위한 기전을 가지고 있어야 한다. sperm whale 같은 경우 1시간가량 잠수하기도 한다 [15]. myoglobin은 주로 근육에 많이 존재하면서 산소를 공급해 주는 역할을 하는 단백질이다 [17]. 최근 연구에서 해양포유

* Corresponding author
Phone: 82-31-400-6447 Fax: 82-31-406-2495
E-mail: yimh@kiost.ac.kr

This is an open-access journal distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/bync/3.0>)

류가 더 높은 농도로 myoglobin을 가지고 있으며 이는 해양포유류 myoglobin net surface charge가 육상 포유류의 myoglobin net surface charge보다 더 커서 서로 척력(repulsion)이 작용하기 때문에 가능하다고 밝혀졌다 [7]. 또한 밍크고래 유전체 연구를 통해 glutathione 대사 단백질, haptoglobin 같은 항산화 관련 유전자들이 고래 특이적 변이를 보여 hypoxia에 대응할 수 있는 기전들을 가지고 있음을 확인하였다 [18].

생존을 위해 산소를 필요로 하는 생물들은 hypoxia (저산소)에 노출되면 이에 적응하기 위해 유전자 발현을 조절하여 에너지 대사와 산소 소모를 유지하려고 한다. 전사인자인 HIF와 HIF에 의해 조절 받는 target gene (HIF-target gene)들은 hypoxia 상태에서 세포들의 대표적인 적응 기전으로 알려져 있다. 따라서 hypoxia 상태에서 고래 HIF-target gene들을 확인하고 이를 분석함으로써 고래가 해수 환경에 적응하기 위한 기전을 분자 수준에서 설명할 수 있게 된다. HIF는 target gene들의 promoter region에 결합하여 target gene들의 발현을 조절하고 이를 통해 hypoxia에 대응하게 된다 [16]. 따라서 HIF가 결합하는 위치 (HIF-binding site, HBS)를 확인하면 HIF-target gene들을 찾게 된다. Transcription-factor-binding sites (TFBS)의 생물정보학적 (*in silico*) 예측은 전사인자의 target gene을 실험적으로 찾는 데 매우 강력한 보조 도구이며 [2], 이와 같이 찾아진 target gene들의 정보들을 chromatin immunoprecipitation같은 실험을 통해 78%의 sensitivity와 97.8%의 specificity를 보여 매우 유용한 방법임을 확인할 수 있었다 [8].

본 연구에서는 해양포유류인 밍크고래 (minke whale, *Balaenoptera acutorostrata*), 병코돌고래 (bottlenose dolphin, *Tursiops truncatus*)와 육상포유류 중에 고래류와 계통적으로 가까운 소 (cow, *Bos taurus*), 돼지 (pig, *Sus scrofa*) 유전체 정보에서 HBS를 찾아내고 HIF-target gene을 비교함으로써 HIF에 의한 고래의 hypoxia 적응 기전을 설명하고자 한다. 이와 같이 서로 다른 생물종의 orthologous sequences을 비교하여 TFBS를 찾는 기술을 phylogenetic footprinting이라 한다.

이전 연구에서 사용되었던 4종 (minke whale, bottlenose dolphin, cow, pig)의 scaffold data와 orthologous cluster 파일을 사용하여 분석을 진행하였다 [18]. 먼저 orthologous cluster 파일에서 4종이 모두 가지는

gene family를 추출하였다 (Set 1). 그리고 set 1에서 minke whale과 bottlenose dolphin 2종의 gene만 추출하여 따로 세트를 구성하였다 (Set 2). orthologous cluster의 protein ID와 gtf 파일의 transcription ID와 링크하여 각각의 start position / end position / strand type을 parsing하여 각 종의 scaffold로부터 upstream region 5K bp 구간을 추출하였다. filtering은 quality control을 통해서 false positive를 후보에서 제거하였다. 찾아진 promoter region의 질은 base-calling error rate를 포함하여 많은 요인에 영향을 받는데, 생성된 scaffold는 모든 read를 사용하는 samtools를 이용하였다. 이 때 low mapping quality의 read를 사용하지 않고 GATK의 Base quality score recalibration (BQSR)을 적용하여 quality를 높였다. 따라서 추출한 promoter region에 raw read를 alignment하여 BQSR을 적용하고, 이 영역은 FDR (False Discovery Rate)에 의해서 5% 이하의 FDR을 선택하여 사용하였다. HBS를 찾기 위해 Gibbs motif sampler을 이용하여 promoter 영역에 존재하는 motif를 추정하였다 [12]. Gibbs program은 전체 탐색이 아닌 확률을 사용한 부분 탐색을 이용함으로써 빠르고 쉽게 다중 DNA의 반복서열을 찾아낸다. 대표적인 HBS의 core motif는 hypoxia response element (HRE)내에 있는 [A/G]CGTG로 알려져 있다 [16]. Gibbs motif sampler에 1, 2 sets를 input으로 하여 추출한 upstream 구간으로부터 common motif를 예측하였다. 예측된 motif들을 대상으로 하여 HRE와 부합하는 motif의 gene 정보를 수집하고 고래류 (Set 2)에서 예측된 HRE 결과와 4종 전체에서 예측된 결과 (Set 1)에서 중복되는 결과물을 제외하여 결과를 도출하였다. 이 같은 과정을 통해 소, 돼지에게는 없지만 고래류에서만 확인되는 HBS의 정보를 얻게 된다. 위의 단계로 얻어진 motif 추정 영역은 annotation 정보를 이용하여 validation을 한다 (Fig 1).

Orthologous cluster 파일에서 4종이 모두 가지는 gene family를 추출하였는데 8191개의 orthologous cluster가 있으며 이 cluster에 포함되는 gene의 숫자는 minke whale이 12114개, bottlenose dolphin이 11515개, cow가 13061개, pig는 11806개이다 (Set 1). 이들로부터 4종 공통으로 HRE의 motif sequence [A/G]CGTG를 갖는 gene이 56개 (Set 1) 이며 고래류에서 예측된 gene이 86개이다 (Set 2). Set 2에서 예측된 gene 결과에서 Set 1과 중복되는 결과를 제외한 30개

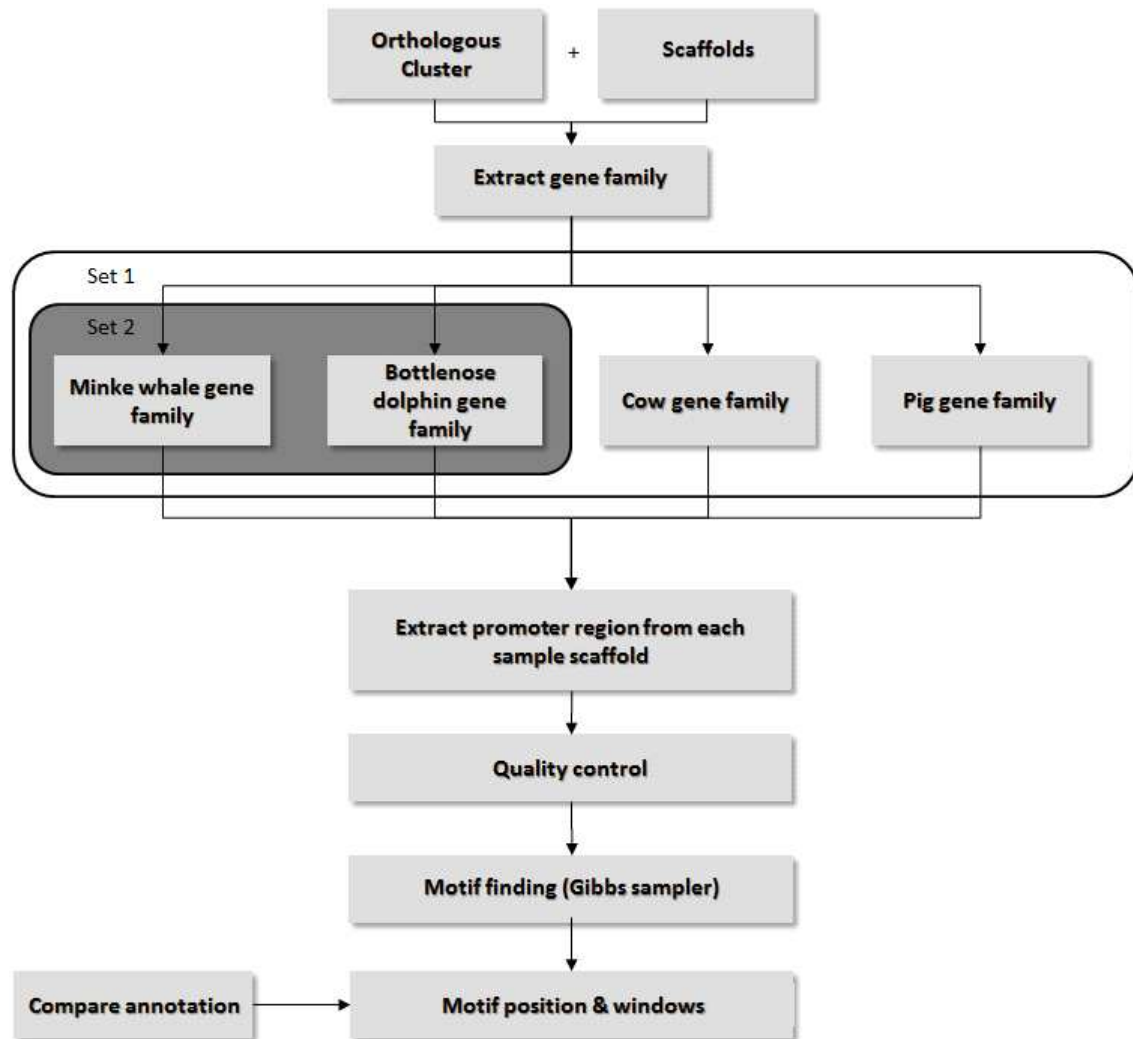


Figure 1. Flowchart of data collection and HIF-target gene analysis in 2 terrestrial mammals (cow and pig) and 2 whales (minke whale and bottlenose dolphin).

의 gene는 hypoxia에서 고래에서만 특이적으로 조절되는 gene으로 예측되는 것들이다 (Table 1). 고래의 gene 중에 promoter region에 HRE의 core motif를 가지고 있어서 hypoxia에 의해 발현이 조절될 것으로 예상되는 30개 중에는 CCDC152 (coiled-coil domain containing 152) 같이 그 기능이 전혀 밝혀지지 않은 것들도 있다. 그러나 그 중에는 알려진 유전자의 기능을 통해 hypoxia와 고래의 생리적 특성에 대한 설명이 가능하다 (Table 2). SLC5A2는 신장에서 sodium의 농도에 따라 glucose를 재흡수하는 역할을 하는 단백질이다 [14]. 고래는 염도가 높은 해수에서 생활하고 있기 때문에 sodium 농도와 깊은 관련이 있으며 또한 저산소 상태에서는 anaerobic glycolysis를 하게 되어 glucose의 소모량이 달라진다 [5]. 따라서 SL

C5A2가 고래에서 HIF-target gene으로 기능할 것으로 추정된다. SLC5A2는 2형 당뇨 치료의 타겟 가능성도 높아서 hypoxia와 관련하여 흥미 있는 연구를 기대할 수 있다 [13]. DRAM1은 autophagy과정에 관여하는 lysosomal membrane protein인데 autophagy는 세포내 기능을 못하거나 불필요한 단백질 등을 체계적으로 분해하고 이를 재활용할 수 있도록 하는 과정을 말한다 [1]. autophagy는 hypoxia를 포함한 다양한 스트레스에 의해 유도되는데 hypoxia와 DRAM1과 관련성에 대한 연구는 없다 [6]. 따라서 HIF-target gene으로 DRAM1이 확인되었기 때문에 고래가 hypoxia상태에서 다른 과정을 통해 autophagy가 유도될 수 있음을 보여 주었다. SLC7A9는 cystine, neutral amino acid, dibasic amino acid를 흡수하는 역할을 한

Table 1. Summary of whale-specific genes with HBS in their promoter regions. BACU and ENSTTRP indicates protein ID of minke whale and bottlenose dolphin respectively. All genome sequences in this analysis were downloaded from www.marinegenome.kr.

Group	Protein ID	Scaffold	Start	End	Gene name	Motif
1	ENSTTRP00000010955	GeneScaffold_2089	118560	118707	"WIF1"	GCTGGGCGTGT T GTTTCA
1	BACUMR048475	scaffold23	17299746	17390573	"WIF1"	GCTGGGCGTGT C GTTTCA
2	ENSTTRP00000006209	GeneScaffold_674	124212	125414	"MET"	CTGGCA C GTGGTAAACAC
2	BACU10008115	scaffold80	8484700	8570626	"MET"	CTGGCA C GTGGTAAACAC
3	ENSTTRP00000014818	GeneScaffold_2372	211455	211668	"CYP4V2"	TAGGGA C GTGCTGCTAGG
3	BACU10014401	scaffold17	12728983	12793063	"CYP4V2"	TAGGGA C GTGCTGCTAGG
4	ENSTTRP00000003269	GeneScaffold_1696	251788	251961	"PTPN14"	GCGGCA C GTGGGATCTTC
4	BACU10002716	scaffold186	1354039	1462335	"PTPN14"	GCGGCA T GTGGGATCTTT
5	ENSTTRP00000011861	GeneScaffold_2614	168787	168912	"SLC5A2"	GCGCCA C GTGTGGCCTCC
5	BACU10014694	scaffold14	46021	104332	"SLC5A1"	GCAGCA C GTGGGATCTAG
5	BACU10015436	scaffold10	18792771	18812016	"SLC5A9"	GCAGCA T GTGGGATCTTC
5	ENSTTRP00000000237	GeneScaffold_487	8260	8394	"SLC5A1"	GTGGCA T GTGGGATCTTT
5	BACU10016882	scaffold4	5413203	5420027	"SLC5A2"	GCGCCA C GTGTGGCCTCC
5	ENSTTRP00000011722	GeneScaffold_3586	79001	79141	"SLC5A9"	GCCCCA C TGGGGCCTCT
5	ENSTTRP00000003728	GeneScaffold_1627	99878	99985	"SLC5A10"	GCAGCA T GTGGGATCTTC
6	ENSTTRP00000015523	scaffold_106287	76438	77136		TTTCCA C GTGCGCGTAGG
6	BACU10008095	scaffold80	1238224	1243440	"THAP5"	TTTCCA C GTGCGCGTAGG
7	BACUMR005627	scaffold11	9865531	9912154	"DRAM1"	AGCCCG C GTGCCCTAGAG
7	ENSTTRP00000016321	GeneScaffold_1961	1041455	1041585	"DRAM1"	AGCCCG C GTGCCCTAGAG
8	ENSTTRP00000011585	GeneScaffold_3616	115479	115565	"SLC7A9"	GAGCCA C GTGGCATGTGG
8	BACU10005787	scaffold104	3470844	3501181	"SLC7A9"	GAGCCA C GTGGCATGTGG
9	ENSTTRP00000007999	GeneScaffold_3199	250161	250244	"C16orf80"	AGCCTG C GTGCTGCAACT
9	BACU10015269	scaffold12	16074247	16087808	"GTL3"	AGCCTG C GTGCTGCAACT
10	ENSTTRP00000008134	scaffold_103859	79206	79377	"CTSC"	GCGGCA C GTGGGCCCTAG
10	BACU10005031	scaffold125	1281702	1322317	"CTSC"	GCGGCA C GTGGGCCCTGA
11	BACU10015951	scaffold7	2778296	2781041	"FAM107A"	CATGGA C GTGTGTGTGTG
11	ENSTTRP00000013044	GeneScaffold_2541	68678	68847	"FAM107A"	CATGGA C GTGTGTGTGTG
12	ENSTTRP00000000637	GeneScaffold_3416	30841	30859	"CCDC152"	TGTGTA C GTGTGAATTTT
12	BACU10009489	scaffold82	3678121	3708783	"CCDC152"	TGTGTA C GTGTGAATTTT
13	BACU10003798	scaffold163	2530788	2558555	"SYAP1"	CAGAGA C GTGGATGGACC
13	ENSTTRP00000012647	GeneScaffold_129	164993	165046	"SYAP1"	CAGAGA C GTGGATGGACC
14	BACU10002242	scaffold200	2250646	2257941	"COX11"	GACCTG C GTGGGGCGGAC
14	ENSTTRP00000011327	GeneScaffold_1630	258845	259000	"COX11"	GACCTG C GTGGGGCGGAC
15	ENSTTRP00000013940	GeneScaffold_1582	7792	8229	"NAA30"	CCCAGG C GTGTTTGCAGC
15	BACU10012306	scaffold71	11112630	11128399	"NAA30"	CCCAGG C GTGCTTGCAGC
16	BACUMR091311	scaffold72	3434963	3527557	"MEF2C"	TCTGTG C GTGTTTCTGC

16	ENSTTRP0000003719	scaffold_115684	507412	507465	"MEF2C"	TCTGTGCGTGTTCCTGC
17	BACU10001902	scaffold210	1523619	1529812	"PNN"	CAAGGGCGTGTGGCAGCA
17	ENSTTRP00000016236	GeneScaffold_519	238400	238512	"PNN"	CAAGGGCGTGTGGCAGCA
18	BACU10005638	scaffold108	2778806	2819533	"GEMIN5"	CGGCCGCGTGTTCACGGT
18	ENSTTRP00000015120	GeneScaffold_2073	164047	164212	"GEMIN5"	CGGCCGCGTGTTCACGGT
19	ENSTTRP00000011170	GeneScaffold_715	147359	147544	"RG9MTD3"	CTCATACGTGCATTTCAA
19	BACU10006633	scaffold90	2896200	2906162	"RG9MTD3"	CTCATACGTGCATTTCAA
20	BACU10014918	scaffold14	19863077	19868695	"RAD9A"	CTAACACGTGTTAAATGA
20	ENSTTRP00000012537	GeneScaffold_2747	15644	15677	"RAD9A"	CTAACACGTGTTAAATGA
21	ENSTTRP00000003071	GeneScaffold_1767	371752	371995	"NFXL1"	TCCCGGCGTGTCTTGGAG
21	BACU10015904	scaffold24	27944386	27997036	"NFXL1"	TCCCGGCGTGTCTTGGGG
22	ENSTTRP00000001251	GeneScaffold_2252	601523	602193	"ZNHIT6"	GATCCGCGTGAGGCTACG
22	BACU10004942	scaffold159	3079286	3171944	"ZNHIT6"	GATCCGCGTGAGGCCACG
23	ENSTTRP00000004954	GeneScaffold_609	719732	719862	"SLTM"	AAGGGGCGTGTGCGCGCC
23	BACU10016691	scaffold11	28245151	28289307	"SLTM"	AAGGGGCGTGTGCGCGCC
24	BACU10005783	scaffold104	3280357	3283303	"PDCD5"	GGACAGCGTGTGTGTGTG
24	ENSTTRP00000003761	GeneScaffold_647	41626	41670	"PDCD5"	GGACAGCGTGTGTGTGTG
25	ENSTTRP00000002534	GeneScaffold_3518	287562	287579	"ANLN"	GAGAGGCGTGCCCTGCC
25	BACU10008092	scaffold80	84635	157226	"ANLN"	GAGAGGCGTGCCCTGCC
26	BACU10009514	scaffold82	8430288	8431735	"PELO"	AAAAAGCGTGTGGCGTCT
26	ENSTTRP00000001747	GeneScaffold_1998	106406	107131	"PELO"	AAAAAGCGTGTGGCGTCT
27	ENSTTRP00000008472	GeneScaffold_3250	356546	356569	"ORC3"	TAGGTGCGTGGCTTCAG
27	BACU10008463	scaffold85	2722491	2776606	"ORC3"	TAGGCGCGTGGCTTCAG
28	BACU10009367	scaffold50	5310009	5388882	"MCPH1"	AGCCCACTGCCACAAC
28	ENSTTRP00000016188	GeneScaffold_1851	124581	124670	"MCPH1"	AGCCCACTGCCACAAC
29	ENSTTRP00000005403	GeneScaffold_3173	70756	70816	"C14orf166"	CCCCGACGTGGCCCCGCC
29	BACU10012272	scaffold71	6202788	6222995	"CN166_PONAB"	CCCCGACGTGGCCCCGCC
30	ENSTTRP00000005084	GeneScaffold_417	93544	93756	"NDUFAF5"	CCTCAGCGTGACGCGCAG
30	BACU10005754	scaffold105	5599240	5667557	"NDUF5_RAT"	CCTCAGCGTGACGCGCAG

Table 2. The list of whale-specific HIF-target genes that have the characteristics related to whale physiology.

Group	Protein ID	Gene symbol	Activity	Reference
5	ENSTTRP00000011861 BACU10016882	SLC5A2 (solute carrier family 5, member 2)	sodium/glucose cotransporter (SGLT2)	[13]
7	BACUMR005627 ENSTTRP00000016321	DRAM1	DNA-damage regulated autophagy modulator 1	[1]
10	ENSTTRP00000011585 BACU10005787	SLC7A9 (Solute carrier family 7, member 9)	cationic amino acid transporter, y+ system	[3]
16	BACUMR091311 ENSTTRP00000003719	MEF2C	Myocyte-specific enhancer factor 2C or MADS box transcription enhancer factor 2, polypeptide C	[9]

다 [3,4]. SLC5A2, DRAM1 그리고 SLC7A9 등 3가지 유전자는 모두 영양분 흡수와 관련이 있어서 고래가 hypoxia에 적응하기 위해 대사와 관련된 유전자가 HIF를 통해 조절을 받는다는 사실을 추정할 수 있다. MEF2C는 전사인자로서 다양한 기관과 조직의 발생, 분화 등에 관여하는 것으로 알려졌다 [9]. 특히 근육발생, 혈관 발달에 영향을 미치는데 이들 조직은 hypoxia 상태에서도 지속적으로 일정한 산소를 가지고 있어야 하기 때문에 MEF2C는 HIF-target gene으로 역할을 해야 할 것이다. 그 외에 apoptosis에서 기능을 하는 PDCD5 (programmed cell death 5)같은 유전자도 HIF-target gene으로 확인되었지만 고래의 생리적 특성과의 연관성을 이해하기는 어렵다.

본 연구에서는 Gibbs motif sampler를 활용한 phylogenetic footprinting을 통해 고래 특이적인 HIF-target gene들을 예측할 수 있었다. chromatin immunoprecipitation을 통해 실험적으로 증명할 수 있지만 세계적으로 밍크고래 혹은 병코돌고래 세포주가 없기 때문에 현재로서는 불가능하다. 그러나 이런 분석을 통해 해양포유류의 hypoxia 적응 기전에 대한 새로운 이해를 하는데 단초를 제공할 수 있다. 또한 현재까지 유전체 분석에서 고래의 해양환경 적응기전을 설명할 경우 gene-coding sequences에 대한 고래 특이적 아미노산 변이 혹은 positively selected gene (PSG) 같은 분석을 통해서 이루어 졌던 것에 반해 이번 연구에서는 regulatory sequences 분석을 활용함으로써 고래 유전체 연구에 새로운 접근이 가능해졌다. hypoxia와 암, 심혈관질환, 당뇨 등 인간의 질환과 깊은 관련성은 이미 많은 연구를 통해 알려져 있다 [10]. 본 연구를 통해 육상 포유류와 다른 고래 특이적 hypoxia 대응 기전을 이해함으로써 인간 질환에 대한 이해와 새로운 치료 방법에 대한 제안을 할 것으로 기대한다.

감사의 글

본 연구는 한국해양과학기술원의 주요과제 연구비 (PE99314)의 지원을 받아 수행하였다.

References

1. Crighton, D., Wilkinson, S., O'Prey, J., Syed, N., Smith, P., Harrison, P., Gasco, M., Garrone, O., Crook, T. and Ryan, K. M. 2006. DRAM, a p53-induced modulator of autophagy, is critical for apoptosis. *Cell*, **126**, 121-134.
2. Elnitski, L., Jin, V. X., Farnham, P. J. and Jones, S. J. 2006. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Res.* **16**, 1455-1464.
3. Feliubadaló, L., Arbonés, M. L., Mañas, S., Chillarón, J., Visa, J., Rodés, M., Rousaud, F., Zorzano, A., Palacín, M. and Nunes, V. 2003. Slc7a9-deficient mice develop cystinuria non-I and cystine urolithiasis. *Hum. Mol. Genet.* **12**, 2097-2108.
4. Goodyer, P., Boutros, M. and Rozen, R. 2000. The molecular basis of cystinuria: an update. *Exp. Nephrol.* **8**, 123-127.
5. Hochachka, P. W., Owen, T. G., Allen, J. F. and Whittow, G. C. 1975. Multiple end products of anaerobiosis in diving vertebrates. *Comp. Biochem. Physiol. B.* **50**, 17-22.
6. Lee, J., Giordano, S. and Zhang, J. 2012. Autophagy, mitochondria and oxidative stress: cross-talk and redox signalling. *Biochem. J.* **441** 523-540.
7. Mirceta S., Signore A.V., Burns J.M., Cossins A.R., Campbell K.L. and Berenbrink M. 2013. Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science.* **340**, 1234192. doi: 10.1126/science.1234192.
8. Ortiz-Barahona, A., Villar, D., Pescador, N., Amigo, J. and del Peso, L. 2010. Genome-wide identification of hypoxia-inducible factor binding sites and target genes by a probabilistic model integrating transcription-profiling data and *in silico* binding site prediction. *Nucleic Acids Res.* **38**, 2332-2345.
9. Potthoff, M. J. and Olson, E. N. 2007. MEF2: a central regulator of diverse developmental programs. *Development.* **134**, 4131-4140.
10. Semenza, G. L. 2012. Hypoxia-inducible factors in physiology and medicine. *Cell.* **148**, 399-408.
11. Thewissen, J.G.M., Cooper, L.N., Clementz, M.T., Bajpai, S. and Tiwari, B.N. 2007. Whales originated

- from aquatic artiodactyls in the Eocene epoch of India. *Nature*. **450**, 1190 - 1194.
12. Thompson, W., Conlan, S., McCue, L. A. and Lawrence, C. E. 2007. Using the Gibbs Motif Sampler for phylogenetic footprinting. *Methods Mol. Biol.* **395**, 403-424.
13. Tönjes, A. and Kovacs, P. 2013. SGLT2: a potential target for the pharmacogenetics of Type 2 diabetes? *Pharmacogenomics*. **14**, 825-833.
14. Vallon, V., Platt, K. A., Cunard, R., Schroth, J., Whaley, J., Thomson, S. C., Koepsell, H. and Rieg, T. 2011. SGLT2 mediates glucose reabsorption in the early proximal tubule. *J. Am. Soc. Nephrol.* **22**, 104-112.
15. Watwood, S. L., Miller, P. J. O., Johnson, M., Madsen, P. T. and Tyack, P. L. 2006. Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). *J. Anim. Ecol.* **75**, 814-825.
16. Wenger, R. H., Stiehl, D. P., and Camenisch, G. 2005. Integration of oxygen signaling at the consensus HRE. *Sci STKE*. **2005**, re12.
17. Wittenberg J. B. 1970. Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. *Physiol Rev.* **50**, 559-636.
18. Yim H. S., Cho, Y.S., Guang, X., Kang, S. G., Jeong, J. Y., Cha, S. S., Oh, H. M., Lee, J. H., Yang, E. C., Kwon, K. K., Kim, Y. J., Kim, T. W., Kim, W., Jeon J. H., Kim, S. J., Choi, D. H., Jho, S., Kim, H. M., Ko, J., Kim, H., Shin, Y. A., Jung, H. J., Zheng, Y., Wang, Z., Chen, Y., Chen, M., Jiang, A., Li, E., Zhang, S., Hou, H., Kim, T. H., Yu, L., Liu, S., Ahn, K., Cooper, J., Park, S. G., Hong, C. P., Jin, W., Kim, H. S., Park, C., Lee, K., Chun, S., Morin, P. A., O'Brien, S. J., Lee, H., Kimura, J., Moon, D. Y., Manica, A., Edwards, J., Kim, B. C., Kim, S., Wang, J., Bhak, J., Lee, H. S. and Lee, J. H. 2014. Minke whale genome and aquatic adaptation in cetaceans. *Nat. Genet.* **46**, 88-92