HighAltitudeOmicsDB: An integrated resource for High-Altitude associated genes and proteins, interacting networks and semantic-similarities.

Pankaj Khurana (✉ pkhurana08@gmail.com)

  DIPAS, DRDO

Rajeev Varshney

  DIPAS,DRDO

Abstract

Millions of people worldwide visit, live, or work in the hypoxic environment encountered at high altitudes and it is important to understand the biomolecular responses to this stress. This would help design mitigation strategies for High Altitude Illnesses (HAIs). Inspite of a number of studies spanned over 100 years, complex mechanisms controlling acclimatization to hypoxia remain largely unknown. Some biomolecules, though, have been proposed as potential diagnostic, therapeutic and predictive markers for HA stress.

HighAltitudeOmicsDB is a unique resource that provides a comprehensive, curated, user- friendly and detailed compilation of various genes/proteins which have been experimentally validated to be associated with various HA conditions; their Protein Protein Interactions (PPI) and Gene Ontology (GO) semantic similarities. For each database entry, HighAltitudeOmicsDB stores the level of regulation (up/down-regulation), fold change, control (low landers or high landers), duration and altitude of exposure, tissue of expression, source organism, level of hypoxia, experimental validation, place/country of study, ethnicity, geographical location, link to respective publication etc.Tthe database also collates information on disease and drug association, Gene Ontology and KEGG pathway associations. The web resource is a unique server platform which constructs PPI networks and extracts GO semantic similarities among the interactors. These unique features help to offer mechanistic insights in disease pathology. Hence, HighAltitudeOmicsDB is a unique platform for researchers working in this area to explore, fetch, compare and analyze HA associated gens/proteins, their PPI networks and GO semantic similarities. The database is available at http://www.altitudeomicsdb.in

Introduction

A large percentage of world's population live at high altitude areas and many also visit the mountains above 2500m for outdoor activities such as trekking, climbing, and other adventure sports. Rapid ascent to high altitude leads to an instantaneous decrease in barometric pressure, the oxygen concentration remains the same but, the number of oxygen molecules per breath is reduced e.g. at altitude 3,600 meters the barometric pressure decreases to 483 mmHg, and <40% of oxygen molecules are available to breathe. Since the amount of oxygen required for activity is the same, the body must adjust to having less oxygen or hypobaric hypoxia [1]. Some lowland residents adjust to the reduced oxygen availability at high altitude through a process known as acclimatization but some suffer from various disorders like Acute Mountain Sickness (AMS), High-Altitude Cerebral Edema (HACE), and High-Altitude Pulmonary Edema (HAPE), etc. ([2], [3]) . Therefore research for the identification of the early signs of these physiological alterations is gaining momentum. A recent comparison in protein profiles of low landers with their induction at high altitude has identified differentially expressed proteins like serum proteins Irisin, myostatin, acute precursor proteins (APPs), apolipoprotein A1, etc in high altitude acclimatization [4],[5], [6]. These proteins are associated with energy-related processes, skeletal muscle regeneration, inflammatory responses, and other hallmark molecular responses at high altitude. Henceforth these proteins are being proposed as biomarkers to predict early acclimatization of individuals at high altitude.

Hunting for novel protein biomarkers in low landers and native samples using peptide profiling has become a key method ([3], [6]). Identification of the differentially expressed proteins that play a key role in the acclimatization process has helped to uncover the mechanisms responsible for the acclimatization at high altitudes. Eg a genome-wide study has uncovered plasma proteins that have the potential to predict vascular homeostasis during HAPE ([7]). Similarly, a transcriptomic study indicated the modulation of multiple pathways and proteins involved in the early phase of hypobaric hypoxia exposure like VIM, CORO1A, CD37, STMN1, etc. ([8]). Though there is enormous literature available that have reported '-omics' profiles of the human and animals exposed to high altitude; the real challenge remains to integrate all these studies to produce a holistic understanding of continuously evolving mechanisms involved in functional adaptations of cells, tissues and organs, as well as the whole organism in the high-altitude hypoxic environment. Hence we developed a HighAltitudeOmicsDB where all this scattered data is collected, curated, analyzed and visualized. The database currently contains ~1300 proteins that are manually curated from peer-reviewed publications. The database stores the association of each protein with HA-stress in terms of level of regulation (up/down-regulation), fold change, control (low landers or high landers), duration and altitude of exposure, tissue of expression, source organism, level of hypoxia, experimental validation, place/country of study, ethnicity, geographical location etc. The database also stored whether the protein has been associated as a high altitude associated biomarker with a corresponding link to the respective publication. The database also collates protein official symbol, protein aliases, chromosomal location, length, Uniprot ID, Enzyme Commission (EC) Number, Protein family Information (Pfam) Id, Protein Databank (PDB) Id, The Integrative Protein Signature Database (InterPro) Id, Single Nucleotide Polymorphism Database (dbSNP) Id. The database has collates protein's functional information like Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways association; their association with other diseases and drugs. The database also provides protein-protein network interactions of each protein with its top-50 interacting partners. The network can be visualized interactively on the webserver. Additionally, HighAltitudeOmicsDB additionally calculates the GO semantic similarity with these 50 interactors to identify functionally related proteins. The database additionally stores the Transcription factors interacting with the gene and their regulation type (repression, activation, distal, proximal etc). Additionally, the miRNAs interacting with the gene is also listed. Thus, HighAltitudeOmicsDB is a unique integrated platform to explore, retrieve, compare and evaluate genes/proteins associated with HA-stress, their PPI networks and semantic similarity and regulation by Transcription factors and miRNAs. This will help uncover the underlying crosstalk between proteins that exists to acclimatize to HA and provide mechanistic insights in these complex molecular responses. It will thus be useful in identifying novel and robust molecular biomarker candidates that can further help in the development of new diagnostic, prognostics, and therapeutic strategies for high altitude disorders.

Methodology

Data collection

A combination of various keywords such as "high altitude", "protein", "gene", "omics", "hypobaric-hypoxia", "anoxia" were used for extensive literature mining from PubMed and google search engines as in 2020 ([9]). The publications were manually scrutinised to identify differentially expressed genes/proteins. After removing redundancy and duplicity, a comprehensive list of proteins that have been Differentially Expressed (DE) at HA were curated from these publications. For each DE protein, its associated information was also fetched that includes 'Name of the protein', 'Protein Official Symbol', 'Aliases', homologous 'Human Entrez ID', 'Source Organism', 'Tissue of expression', 'level of hypoxia', 'altitude', 'duration of experiment', 'Level of regulation', 'Fold change', 'Experiment details', 'geographical location', 'ethnicity', 'Control group', 'Associated as Biomarker'. For the studies in which the source organism was other than human, the homologous human gene/protein was identified and stored. This way, even for experiments conducted on different experimental organisms (mice/rats/yak/bird/toad/sheep), human equivalence/translation would be easier. The collection was stored in JavaScript Object Notation (JSON) file format and stored in MongoDB.

Data Processing and Enrichment

For each protein additional details like protein official symbol, protein aliases, chromosomal location, length, Uniprot ID, Enzyme Commission (EC) Number, Protein family Information (Pfam) Id, Protein Databank (PDB) Id, The Integrative Protein Signature Database (InterPro) Id, Single Nucleotide Polymorphism Database (dbSNP) Id was collected to help in cross-linking with other databases. For each protein, its GO-functional enrichment and KEGG-pathway annotation was identified by Database for Annotation, Visualization and Integrated Discovery (DAVID), and KEGG mapper respectively ([10],

For each protein in the database, its top-50 protein interactors were identified by Search tool for retrieval of interacting proteins (STRING) webserver ([11]). The stringency for the search was kept at the highest level (0.9) and the filter was placed to allow a maximum of 50 associated proteins as the direct interactor of queried protein. STRING database constructs the protein-protein interactions network based on seven sources of information i.e. neighbourhood on the chromosome, gene fusion, phylogenetic co-occurrence, homology, co-expression, experimentally determined interaction, database annotated, automated textmining. The interaction file was downloaded from the STRING database was stored in JSON format.

To make the database more informative several other attributes were also added; protein-disease associations were mined from Diseagenet [12] ; protein–drug relationship from DGIdb 3.0 database ([13]). All these attributes were also stored in JSON files.

Gene Ontology (GO) annotation based semantic comparisons between genes is an innovative approach to quantitatively assess the functional similarities between them. They have been extensively used across varied bioinformatics analysis ([14], [15]). The higher the semantic similarity score more is the probability that two genes/proteins are likely to have a similar molecular function or involved in a common biological process [14]. Whereas the low semantic similarity score shows two genes imparting different molecular functions. To identify semantic similarity, each protein in the HighAltitudeOmicsDB and its top-50 direct interacting proteins was submitted to GOSemSim R algorithm. GOSemSim is an R

package for semantic similarity computation among GO terms, sets of GO terms, gene products and gene clusters ([15]). The results were represented in a 51 X 51 matrix. All these matrix files were also stored in the JSON file format.

Database Development

All constructed JSON files were transferred to the MongoDB database collection and uploaded on the server localhost using pymongo. Server query commands were made in the MongoDB compass. Vis.js library specifically was used to display the Protein-protein interactions network. The IDs such as Human Entrez ID, Uniprot ID, Protein Official Symbol, EC Number, PDB_ID, InterPro ID, Pfam ID, dbSNP ID, and reference PMIDs present in all tables are hyperlinked to the corresponding databases to provide additional details. The web interface also has a 'Contact us' page which includes option to send email to the developers; for the submission of any new data by the user. It would be reviewed and appended to database.

Results

Web Interface

HighAltitudeOmicsDB (Figure1) is a user-friendly, free-to-access resource which requires no prior registration. It is a comprehensive, non-redundant, manually curated resource of genes/proteins whose expression level are experimentally validated to be associated with high-altitude stress. The database surveyed using "browse" and "Search" options.

The "Browse" option allows the user to choose easily single or multiple genes/proteins of the database from a pull-down menu. Alternatively, the user may upload a file containing the protein official symbols or alternatively type the protein-official symbols. Clicking the adjacent 'Browse' button connects to a tabular format which hyperlinks the individual protein page. If the user-list contains protein symbols which are not in the database, a separate table highlighting the same is also provided (Figure 1)

The Search Option of the database offers multiple options to explore the database based on user research interests. Search by chromosome allows to click on the any human chromosome number and identify the proteins of HighAltitudeOmicsDB which lie on the respective chromosome (Figure 2). Search by 'duration of experiment' allows to identify the list of genes/proteins whose expression changes in hours/days/weeks/months/years. Searching by 'Tissue of expression' opens a pull-down menu from which the user can choose the tissue of interest. Searching by 'Ethnicity', 'source organism', 'level of regulation', 'geographical location' similarly opens a pull-down menu from which the user may choose the ethnicity, source organism, up/down regulation and location respectively and get a tabular list of genes/proteins which are hyperlinked to the respective detailed information page of the protein (as discussed in following sections).

Additionally 'Associated as Biomarker' option leads to a tabular list of proteins which have been proposed/validated as molecular biomarkers for HA-stress (Figure 2). The protein symbols are hyperlinked to the respective protein page which provides a link to the Pubmed which validates the protein as a biomarker. Additionally to fetch proteins which are DE in an altitude-dependent manner, a user-interactive slider (ranging from 2200mt to 9800 mt) is provided. The user may set the slider values and fetch genes/proteins which are associated with a defined altitude-range. This has been combined with (AND/OR) options with time of exposure to HA and level of regulation (Up/Down). The user may thus be able to make combination queries like up/down-regulated proteins expressed in days at an altitude range of 2200mt to 4500 mt. The list of these proteins can be downloaded in Excel /CSV format for further analysis.

The webserver also allows to explore the proteins of HighAltitudeOmicsDB associated with a particular Transcription Factor (TF), microRNA (miRNA), disease, drug, GO or Kegg pathway (Figure 3).

The details of the protein and its association with HA is provided in the detailed information page which may be divided into six sections.

- Knowledge Base

This is the first section of the database that gives general information about the protein like Protein Official Symbol, Aliases, Chromosomal location, Length, Uniprot Id, EC number, Pfam Id, PDB Id, InterProID, dbSNP Id which allows cross-linking to additional databases easy and quick. The Uniprot Id is hyperlinked to the Uniprot database (Figure 4(i)).

- Interactions and Semantics

The top-50 direct protein interactors of each protein is identified from STRING database using cut-offs described in the methodology section. The network is displayed in a user-interactive format with translation, zoom-in and zoom-out features. The nodes are color coded (yellow: the protein being studied; blue : the top-50 interactors) (Figure 4). The edges are also color coded (yellow: interactions between the protein being studied and its 50 direct interactors; blue: interactions among the top-50 interactors). The network may easily be downloaded in .sif format which can be easily visualised in network visualization software like cystoscape, bina etc. The list of interactions between them and their combined score is realidy provided in a tabular format which can be downloaded in excel/CSV format. The table is also provided with a 'search' option to easily search the protein of interest.

The pai-wise GO semantic similarity score is calculated between the protein being studied and its top-50 interacting proteins as described in methodology section. The results are visualised as 51 X 51 matrix. The GO semantic similarity score > 0.8 is highlighted in red-color in the matrix. If any protein among the top-50 interactors is also a part of HighAltitudeOmicsDB, the protein symbol in the matrix is hyperlinked to the respective detailed protein information page within the database. This helps to identify any

functional hubs of proteins that would be associated HA stress and hence could shed light om molecular basis for acclimatization/adaptation (Figure 4).

- **Association with High Altitude**

For each protein, its association with HA stress is compiled in a tabular format. The details are presented as the human protein symbol, source organism (organism in which the study was performed), tissue of expression, Level of Hypoxia, Altitude, Duration of experiment, Level of expression, Fold Change, Experiment Details, Geographical location, Ethnicity, Control group expression, control group details and reference paper (Figure 4). The association of the protein as a biomarker is also compiled i.e. if the protein is ever experimentally validated to be a biomarker, the entry in the column will be "Yes" otherwise "No". The papers are hyperlinked to PubMed which allows ready access to the original publication. In this format, the expression changes of a protein in different durations, tissues and altitude-conditions can be easily and quickly explored, compared and analysed.

- **Association with TFs and miRNAs**

Transcription Factors and miRNAs are two most important transcriptional and post-transcriptional regulatory molecules fine-tuning the expression of genes. Thus the list of TF and miRNAs that are known the regulate the protein being studied is presented in a tabular format. The TF association table lists the TF symbol (hyperlinked to Genecards Database), its entrez id, symbol and entrez id of the protein being studied, type of association, link to publication which ascertained this association and the database from which the association is extracted. The tables are downloadable in Excel/csv format and provided with 'search' option to explore the table with a user-defined keyword (Figure 5).

Similary the miRNA-gene association table lists the miRNA miRTarBase Id, miRNA, symbol and entrez id of the protein being studied, experiment (luciferase reporter assay/western blot/ PCR/ Immunohostochemistry etc), support type and link to respective publication (hyperlinked to PubMed) which ascertained this association. The tables may be downloaded in Excel/csv format. The table is also provided with 'search' option to explore the table with a user-defined keyword.

- **Gene Ontology and KEGG Pathway annotations**

The Gene Ontology annotations are presented in a tabular format. The GO ID, GO Term and GO type are listed. The GO ID is also hyperlinked to QuickGO which provides detailed GO annotations. The KEGG pathway annotations are also compiled and presented as KEGG ID and KEGG Term. The KEGG ID is hyperlinked to KEGG database that provides additional details about the respective pathways (Figure 5).

Both these tables can eb downloaded in Excel/CSV format and have an in-built 'search' option for keyword search.

- **Association of proteins with other diseases and drugs**

This section provides details of drug, tissue, and disease-association of HA-associated genes/proteins. The information is represented in two tables belonging to each category respectively (Figure 5). The first table shows information about protein and its associated drug. This type of information can help the users to guide/design any protein-based drug-targeting experiment. These two tables are equipped with the "search" option which help in easy search of user-defined terms across lengthy tables. The tables can also be downloaded in Excel/CSV format.

Web Statistics

HighAltitudeOmicsDB contains ~1300 associations of 820 proteins that have been found differentially expressed at High altitude. A detailed review of the database shows that all proteins were sourced from experimental studies in 25 tissues (Figure 6a). These tissues are source from 7 animal species i.e. Human, Sheep, Rat, Mice, Yak, Bird, Toad (Figure 6b). Humans as source organisms can be further characterized in terms of their ethnicity i.e. Americans, Tibetans, Han-Chinese, Italians, Nepali, Ladakhi, and Germans. The time of exposure is dependent on the source organism and it ranges from 0.5 hours to 110 Days for the native population.

The database contains two types of functional annotations- GO and KEGG pathway enrichment. The GO enrichment shows 'Metabolic Process' (GO: 0042572), 'Outer Dynein Arm Assembly' (GO: 0036158), 'Response To Reactive Oxygen Species' (GO: 0000302) as the top biological processes (Figure 7a). 'Metabolic process' is highly associated with weight loss due to the adaptation mechanism at high altitude ([16]). At high altitude, induction of hypobaric hypoxia activates HIF protein that further regulates genes responsible for mediating changes in cellular metabolism/energetics leading to weight loss due to increases in energy expenditure ([17]). The second biological process 'Outer Dynein Arm Assembly' is the process for axonemal assemblies. The increase in the length and density of axoneme-like cilia due to hypoxia has been associated with cell death ([18]). Lastly, 'Response To Reactive Oxygen Species' is the reflection of the redox status of the cell, and disturbances in redox status due to hypobaric hypoxia can lead to oxidative stress and DNA damage ([3]). Similarly, terms like 'Fructose-Bisphosphate Aldolase Activity', 'Oxidoreductase Activity', 'Acting On Paired Donors', 'Incorporation Or Reduction Of Molecular Oxygen', 'Oxidoreductase Activity', 'Acting On Peroxide As Acceptor', 'Electron Transfer Activity' and 'ATP Binding', etc. are found to be top molecular functions of proteins present in the database (Figure 7b). All the molecular functions are direct steps or feedback mechanisms associated with oxidative phosphorylation (aerobic respiration). Mitochondria plays important role in oxidative phosphorylation and recent clinical studies have revealed a high percentage of mitochondria are present in gastrocnemius muscle tissue of high lander that helps to adapt to high energy expenditure environment ([19]). 'COP9 signalosome' and 'Actomyosin' are the two cellular components terms that are found most enriched in differentially expressed protein sets present in the database (Figure 7c). COP9 Signalosome is the part of the ubiquitin proteasomal degradation complex that controls the expression of pVHL, HIF-1α, and other oxygen responsive transcription factors regulated during hypobaric hypoxia ([20]). Whereas Actomyosin is a cytoskeleton of actin-myosin fiber complex present in different muscle tissues like skeletal muscle.

The muscle fiber-type composition of both adult animals and humans is markedly altered during chronic exposure to high altitude.

The KEGG pathway enrichment shows 'hsa00910: Nitrogen metabolism' as the most enriched pathways in the differentially expressed HA protein set (Figure 7d). Nitrogen metabolism is a process of nitrogen oxides production and these oxides such as nitrous, nitrite, nitrate have been found to play important role in high altitude acclimatization responses ([21]). Thus the proteins in the database are associated with hallmark responses to hypobaric hypoxic stress responses, which supports the comprehensiveness of the database.

Conclusions

HighAltitudeOmicsDB is an interactive resource and a server platform that captures and organizes knowledge for genes/proteins associated with HA stress. It provides the comprehensive view of different HA related studies; offers the annotations and visualization of PPI networks and semantic similarities associated with gene/protein in the database. HighAltitudeOmicsDB is the first of its kind database that has a collection of manually curated differentially expressed genes/proteins that were fetched using text mining and manual curation. The information enables the user to browse biomolecules based on different query filters in the database i.e. level of expression; duration of experiment; altitude and source organism. HighAltitudeOmicsDB also encompasses protein-associated information such as protein-disease association, protein-drug association. Hence the information-base of HighAltitudeOmicsDB is much larger than any other database. HighAltitudeOmicsDB also identifies PPIs for each protein in the database and calculates GO semantic similarity between them. The analysis of PPI networks and similarities would enable the user to infer mechanistic insights during HA stress. The webserver also offers functional correlation of proteins. The functional correlation includes both GO enrichment and KEGG pathways enrichment.

Declarations

Acknowledgements

Conflict of Interest

The authors declare no competing interests.

References

1. Simonson, T.S., *Altitude Adaptation: A Glimpse Through Various Lenses.* High Alt Med Biol, 2015. **16**(2): p. 125-37.

2. Basnyat, B. and D.R. Murdoch, *High-altitude illness.* Lancet, 2003. **361**(9373): p. 1967-74.

3. Ahmad, Y., et al., *An insight into the changes in human plasma proteome on adaptation to hypobaric hypoxia.* PLoS One, 2013. **8**(7): p. e67548.

4. Wang, C., et al., *Exploration of Acute Phase Proteins and Inflammatory Cytokines in Early Stage Diagnosis of Acute Mountain Sickness.* High Alt Med Biol, 2018. **19**(2): p. 170-177.

5. Sliwicka, E., et al., *Serum irisin and myostatin levels after 2 weeks of high-altitude climbing.* PLoS One, 2017. **12**(7): p. e0181259.

6. Ahmad, Y., et al., *Identification of haptoglobin and apolipoprotein A-I as biomarkers for high altitude pulmonary edema.* Funct Integr Genomics, 2011. **11**(3): p. 407-17.

7. Sharma, M., S.B. Singh, and S. Sarkar, *Genome wide expression analysis suggests perturbation of vascular homeostasis during high altitude pulmonary edema.* PLoS One, 2014. **9**(1): p. e85902.

8. Gaur, P., et al., *Temporal transcriptome analysis suggest modulation of multiple pathways and gene network involved in cell-cell interaction during early phase of high altitude exposure.* PLoS One, 2020. **15**(9): p. e0238117.

9. Fiorini, N., D.J. Lipman, and Z. Lu, *Towards PubMed 2.0.* Elife, 2017. **6**.

10. Dennis, G., Jr., et al., *DAVID: Database for Annotation, Visualization, and Integrated Discovery.* Genome Biol, 2003. **4**(5): p. P3.

11. von Mering C, H.M., Jaeggi D, Schmidt S, Bork P, Snel B. , *STRING: a database of predicted functional associations between proteins.* Nucleic Acids Res., 2003(31(1)): p. 258-61.

12. Pinero, J., et al., *DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants.* Nucleic Acids Res, 2017. **45**(D1): p. D833-D839.

13. Cotto, K.C., et al., *DGIdb 3.0: a redesign and expansion of the drug-gene interaction database.* Nucleic Acids Res, 2018. **46**(D1): p. D1068-D1073.

14. Pesquita, C., *Semantic Similarity in the Gene Ontology.* Methods Mol Biol, 2017. **1446**: p. 161-173.

15. Yu, G., et al., *GOSemSim: an R package for measuring semantic similarity among GO terms and gene products.* Bioinformatics, 2010. **26**(7): p. 976-8.

16. Moore, L.G., *Measuring high-altitude adaptation.* J Appl Physiol (1985), 2017. **123**(5): p. 1371-1385.

17. Palmer, B.F. and D.J. Clegg, *Ascent to altitude as a weight loss method: the good and bad of hypoxia inducible factor activation.* Obesity (Silver Spring), 2014. **22**(2): p. 311-7.

18. Brown, J.M., et al., *Hypoxia regulates assembly of cilia in suppressors of Tetrahymena lacking an intraflagellar transport subunit gene.* Mol Biol Cell, 2003. **14**(8): p. 3192-207.

19. Scott, G.R., K.H. Guo, and N.J. Dawson, *The Mitochondrial Basis for Adaptive Variation in Aerobic Performance in High-Altitude Deer Mice.* Integr Comp Biol, 2018. **58**(3): p. 506-518.

20. Mikus, P. and W. Zundel, *COPing with hypoxia.* Semin Cell Dev Biol, 2005. **16**(4-5): p. 462-73.

21. Levett, D.Z., et al., *The role of nitrogen oxides in human adaptation to hypoxia.* Sci Rep, 2011. **1**: p. 109.
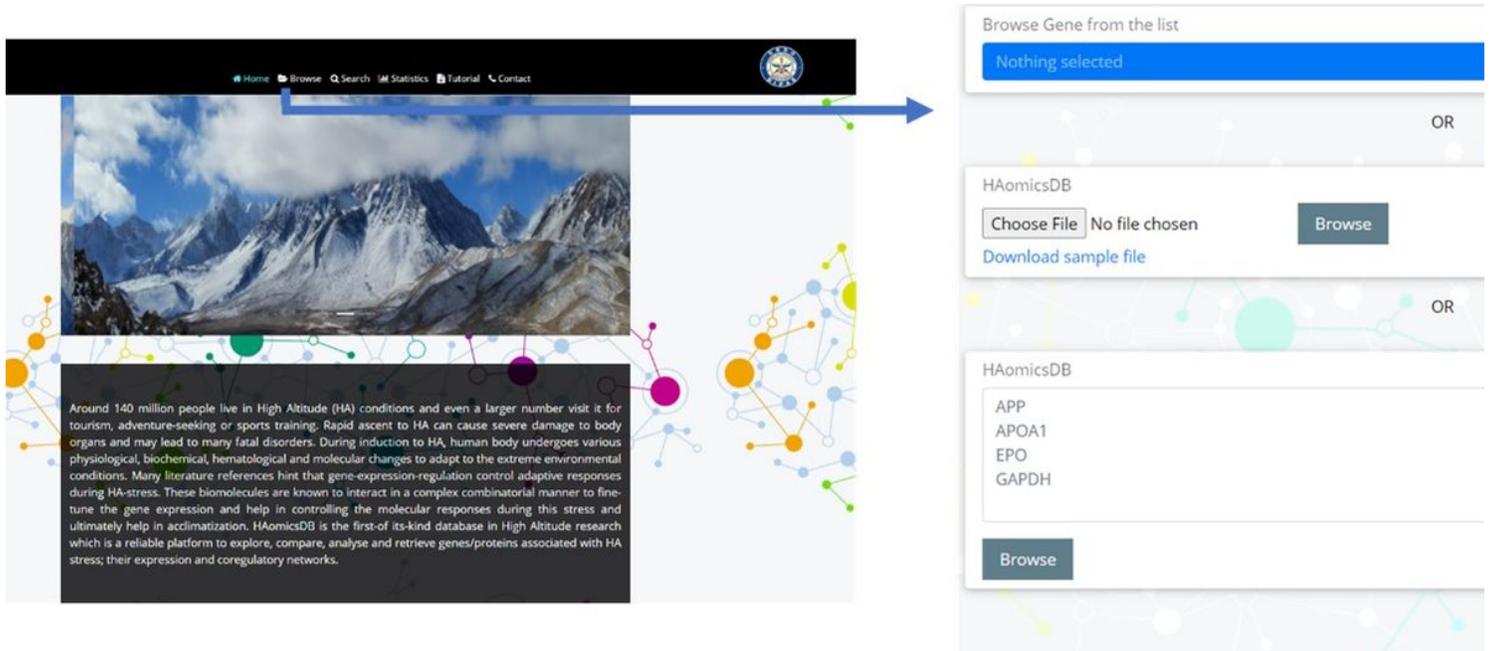
Figures

Figure 1
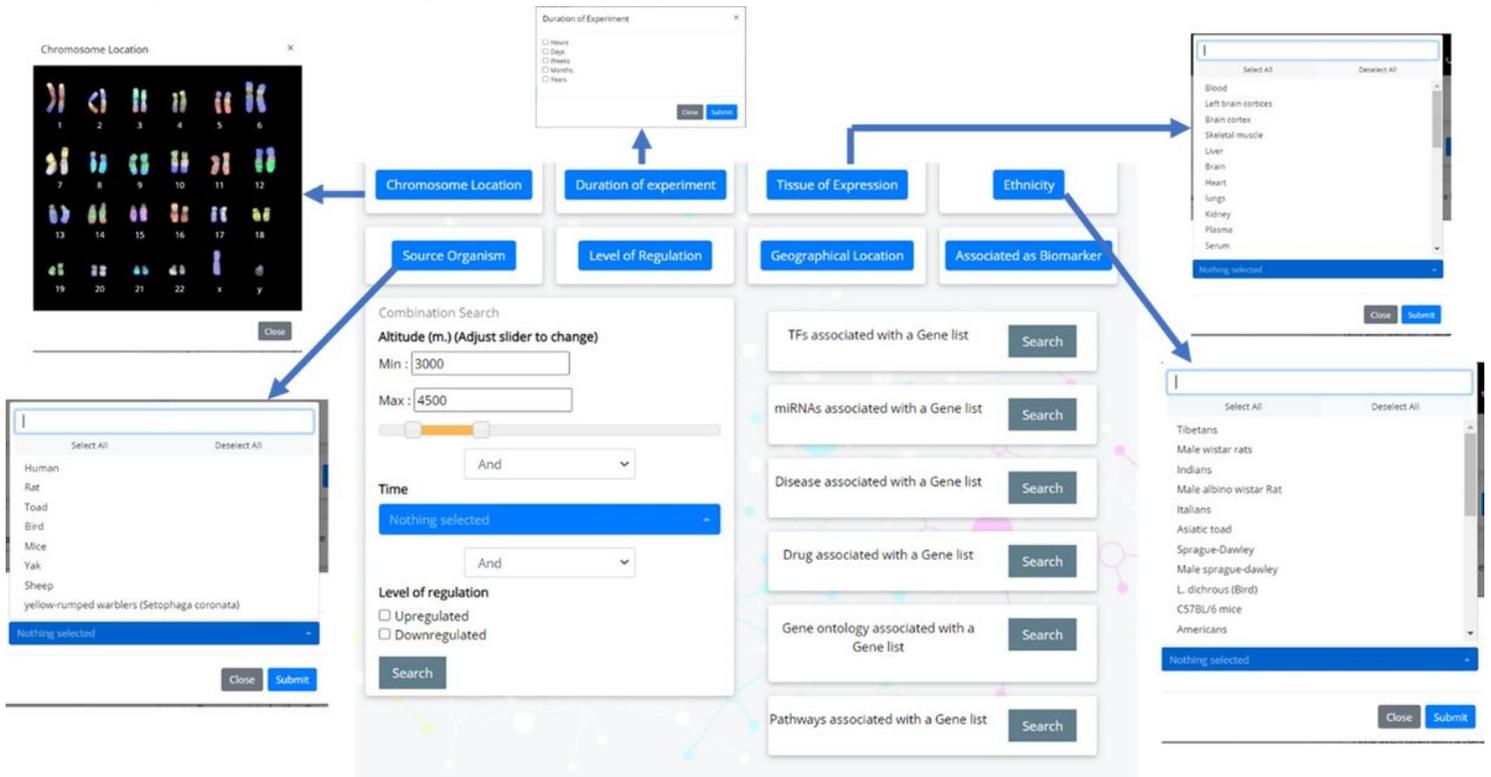
The home page and browse option of the webserver

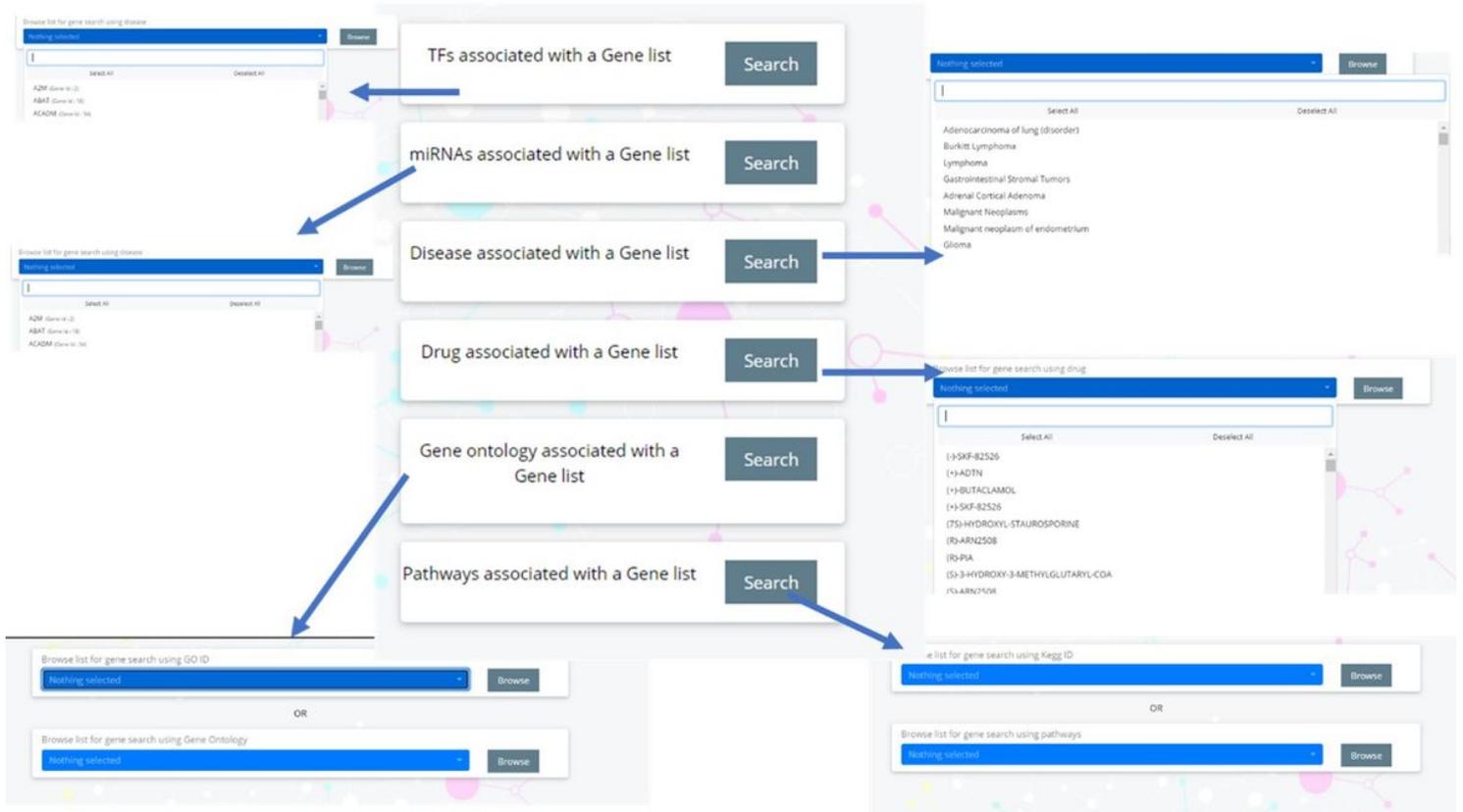## Figure 2

Screenshot of the "Search" module

Figure 3

Screenshot of the "search module"

Figure 4

Protein information page details

(iv) Association with TFs and miRNAs

Association with TF

Excel CSV

Search:

TF	TF Entrez	Gene	Gene Entrez	Type	PMID	Database
EP300	2033	A2M	2	distal	22955619	TRANSFAC
FGF9	2254	A2M	2	Unknown	20110263	TRUSST
FOXA1	3169	A2M	2	proximal_filtered	22955619	TRANSFAC
FOXA2	3170	A2M	2	proximal_filtered	22955619	TRANSFAC
NFKB1	4790	A2M	2	Unknown	11466367	TRUSST
STAT3	6774	A2M	2	Unknown	11466367	TRUSST
TCF4	6925	A2M	2	proximal_filtered	22955619	TRANSFAC
TFCP2	7024	A2M	2	Unknown	10455131	TRUSST

(v) Gene Ontology and KEGG Pathway annotations

Gene Ontology

Excel CSV

Search:

ID	GO ID	GO Term	GO Type
2	GO:0001869	negative regulation of complement activation	GOTERM_BP_DIRECT
2	GO:0002020	protease binding	GOTERM_MF_DIRECT
2	GO:0002576	platelet degranulation	GOTERM_BP_DIRECT
2	GO:0004866	endopeptidase inhibitor activity	GOTERM_MF_DIRECT
2	GO:0004867	serine-type endopeptidase inhibitor activity	GOTERM_MF_DIRECT
2	GO:0005096	GTPase activator activity	GOTERM_MF_DIRECT

Pathways

Excel CSV

Search:

Human Entrez ID	KEGG ID	KEGG Term
2	hsa04610	Complement and coagulation cascades

Showing 1 to 1 of 1 entries

(vi) Association of proteins with other diseases and drugs

Association with Disease

Excel CSV

Search:

Protein Official Symbol	Human Entrez ID	Disease Name	Disease Id	Disease Semantic Type	Semantic score	DSI	DPI	Disease Type
A2M	2	Alzheimer's Disease	C0002395	Disease or Syndrome	0.4	0.564	0.724	disease
A2M	2	Malignant tumor of colon	C0007102	Neoplastic Process	0.3	0.564	0.724	disease
A2M	2	Colonic Neoplasms	C0009375	Neoplastic Process	0.3	0.564	0.724	group
A2M	2	Presenile dementia	C0011265	Mental or Behavioral Dysfunction	0.3	0.564	0.724	disease
A2M	2	Hepatolenticular Degeneration	C0019202	Disease or Syndrome	0.3	0.564	0.724	disease
A2M	2	Kidney Failure, Acute	C0022660	Disease or Syndrome	0.3	0.564	0.724	disease

Association with Drug

Excel CSV

Search:

Protein Official Symbol	Human Entrez ID	drug_claim_primary_name	drug_name	drug_chembl_id	interaction_types
A2M	2	COBALT	None	None	None
A2M	2	PROSTAGLANDIN E1	ALPROSTADIL	CHEMBL495	None
A2M	2	THROMBIN	THROMBIN	CHEMBL2108110	None

Figure 5
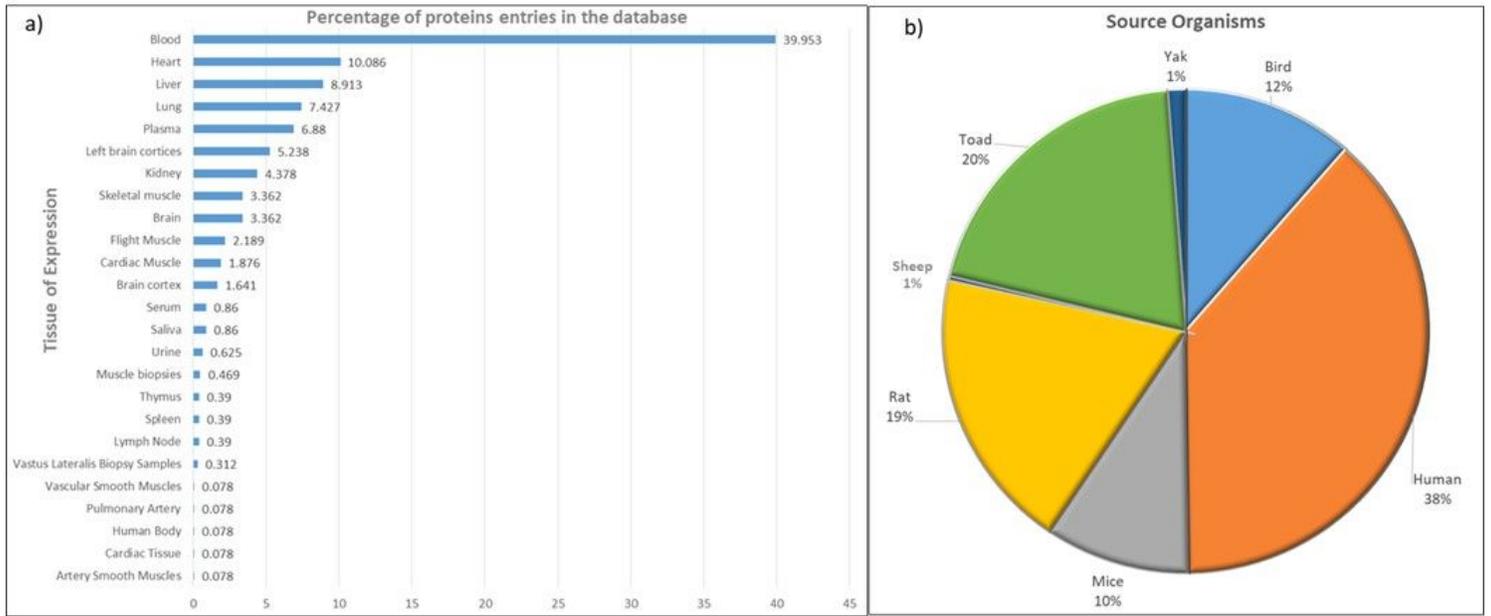
Screenshot of the protein information page

Figure 6

HighAltitudeOmicsDB Statistics (a) Distribution of high altitude proteins as per their tissue of expression. (b) Distribution of high altitude proteins studied in different source organism.
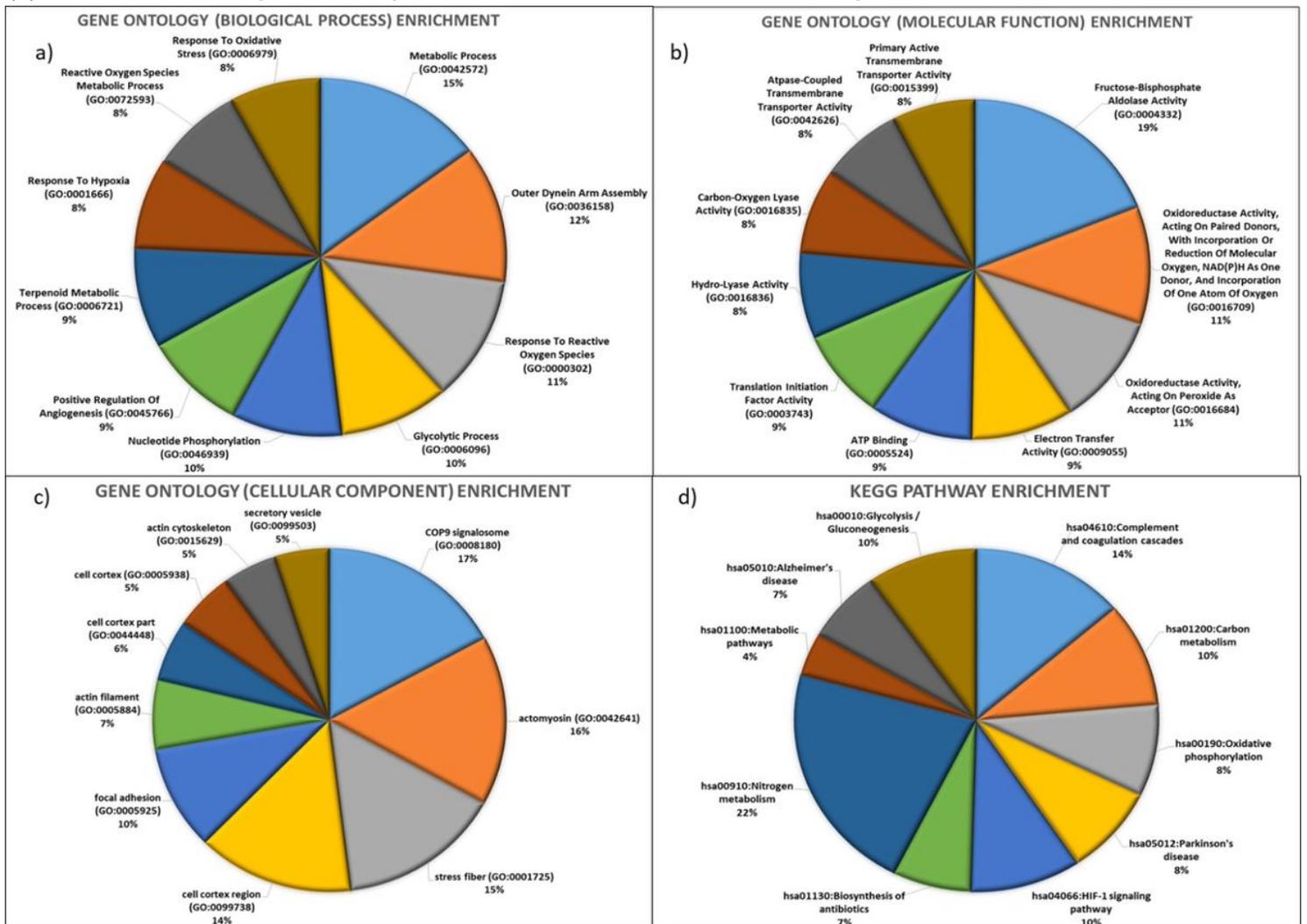
Figure 7

Functional characterization of differentially expressed High altitude proteins a) GO: Biological processes. b) GO: Molecular functions c) GO: Cellular Compartment and d) KEGG Pathway Enrichment.