

BGvar - a comprehensive resource for blood group immunogenetics

Mercy Rophina^{1,2}, Kavita Pandhare^{1,2}, Sudhir Jadhao^{3,4}, Shivashankar H. Nagaraj^{3,4} and Vinod Scaria^{1,2,\$}.

¹CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, Delhi 110025, INDIA

²Academy of Scientific and Innovative Research (AcSIR), CSIR-HRDC Campus, Sector 19, Kamla Nehru Nagar, Ghaziabad, Uttar Pradesh 201002, INDIA

³ Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia

⁴Translational Research Institute, Brisbane, Australia

^{\$}Address for correspondence

Corresponding author Email: vinods@igib.in (Scaria V)

Address of the author for correspondence

Vinod Scaria, MBBS, PhD

Senior Scientist

G N Ramachandran Knowledge Centre for Genome Informatics,
CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB),
South Campus, Mathura Road,
New Delhi 110020.

Phone: +91 9650466002

Fax: +91 011-27667471

Email: vinods@igib.in

Email of authors:

Mercy Rophina: mercy.rophina@igib.res.in

Kavita Pandhare: pandharekavita@gmail.com

Sudhir Jadhao: sudhirjadhao2009@gmail.com

S Hiriya-Nagaraj: shivashank@gmail.com

Running title: BGvar - A Compendium of Human blood group alleles

Abstract

Background

Blood groups form the basis of effective and safe blood transfusion. There are about 41 well recognized human blood group systems presently known. Blood groups are molecularly determined by the presence of specific antigens on the red blood cells and are genetically determined and inherited following Mendelian principles. The lack of a comprehensive, relevant, manually compiled and genome-ready dataset of red cell antigens limited the widespread application of genomic technologies to characterise and interpret the blood group complement of an individual from genomic datasets.

Materials and Methods

A range of public datasets were used to systematically annotate the variation compendium for its functionality and allele frequencies across global populations. Details on phenotype or relevant clinical importance were collated from reported literature evidence.

Results

We have compiled the Blood Group Associated Genomic Variant Resource (BGvar), a manually curated online resource comprising all known human blood group related allelic variants including a total of 1672 ISBT approved alleles and 1552 alleles predicted and curated from literature reports. This repository includes 1606 Single Nucleotide Variations (SNVs), 270 Insertions, Deletions (InDels) and Duplications and about 1310 combination mutations corresponding to 41 human blood group systems and 2 transcription factors. This compendium also encompasses gene fusion and rearrangement events occurring in human blood group genes.

Conclusion

To the best of our knowledge, BGvar is a comprehensive and a user friendly resource with most relevant collation of blood group alleles in humans. BGvar is accessible online at URL: <http://clingen.igib.res.in/bgvar/>

Introduction

Accurate and expedient characterization of blood group antigen profile is central to the safe and effective practice of transfusion medicine. These blood groups are defined according to the presence of specific antigens on the surface of Red Blood Cells (RBCs) with an individual's blood type corresponding to their specific complement of RBC surface antigens¹. Serological tests used to identify these antigens thus form the backbone of blood group characterization efforts in clinical settings.

Blood group antigens encompass a diverse range of specific genetically-determined protein sequences and oligosaccharide chains. With the exception of de-novo events, blood groups thus exhibit a Mendelian pattern of inheritance. Human blood group systems are often regulated by a single gene locus or two or by more closely linked homologous genes. There exists about 360 human blood group antigens, 322 of which belong to 36 well recognized blood group systems².

Virtually all genes involved in encoding these human blood group antigens have been identified and mapped³. One of the earliest reviews elaborating the genetic basis of the human ABO blood group system extensively explained the types of variations and their functional impacts^{4,5}. RH is yet another major human blood group system whose molecular characterization has been pervasively studied^{6,7,8}. The accurate characterization of these blood groups is important both in the context of preventing transfusion-related hemolytic reactions, and in research efforts aimed at understanding the pathogenesis of a number of hematological disorders^{1,9,10,11}.

The advent of high throughput genomic modalities including array based genotyping and Next Generation Sequencing (NGS) technologies has revolutionized the application of genomics in clinical settings. In recent years, these technologies have enabled the characterization of potentially clinically-relevant genetic variations, among individuals. Such characterization is potentially relevant for the diagnosis of genetic conditions for pharmacogenomics, for optimizing therapeutic dosing strategies, and for minimizing adverse events while simultaneously identifying genetic correlates of a number of clinical traits and conditions. With its efficacy in overcoming most of the limitations borne by serological and molecular diagnostics, it has now become the future choice in blood group genotyping strategies. One of the major and significant advantage of NGS based blood typing is the ability to determine the sequence of the entire gene coding for the red blood cell/platelet antigen, thereby uncovering the role of novel variants in human blood group systems.^{12,13,14,15} Extensive blood typing procedures are also commonly employed to minimize transfusion reactions in clinical settings¹⁶. The need for such extensive characterization is also highlighted by a recent annual report of the Food and Drug Administration (FDA) regarding transfusion associated fatalities, which emphasizes that ~7% of Hemolytic Transfusion Reactions (HTRs) observed in the United States from 2013 - 2017 were due to ABO incompatibilities whereas, 11% of these reactions were due to non-ABO incompatibilities¹⁷.

Despite the promise of these personalized blood typing strategies, the widespread use of genomic technologies in transfusion medicine is presently hampered by the lack of a systematic and relevant compendium of alleles associated with blood group antigens that can aid the clinical interpretation of genetic variants based upon the results of high throughput experiments. One of the initiatives by the National Center for Blood Group Genomics (NCBGG) has maintained a compilation of about 148 RHCE alleles along with their relevant information like ancestry, prevalence and variation IDs¹⁸. To that end, we created BGvar, a comprehensive catalog of genetic variants associated with human blood group systems. To the best of our knowledge, BGvar is the most comprehensive and up-to-date compendium of human blood group alleles. The resource is accessible online at URL: <http://clingen.igib.res.in/bgvar/>

Materials and Methods

Datasets

Human blood group variant information was compiled from literature sources and public documents. The compiled variants were additionally checked in public resources and appropriate identifiers were mapped. Blood group antigens and allele data was manually collated from resources including the International Society of Blood Transfusion (ISBT)², The Blood Group Antigen Gene Mutation Database (BGMUT)¹⁹, ErythroGene²⁰, The Human RhesusBase²¹, Blood Antigens^{22,23} and literature references including *The Blood group antigen factsbook*²⁴ and *Blood groups and Red cell antigens*²⁵. ISBT has systematically developed and maintained a repository of blood antigen information corresponding to 41 human blood groups and 2 transcription factors. BGMUT, a now inactive online resource, comprises a manually curated compilation of allelic variations in various human blood group related genes accounting for about 1251 alleles from 41 genic loci. ErythroGene is an online search engine which comprises multiethnic human blood group related alleles predicted from 1000 Genomes Project in addition to known and recognized alleles. RhesusBase is a recent resource that exclusively encompasses Rhesus and Colton blood group antigen information. Blood Antigens resource, in addition to compilation of molecular changes occurring in Red Blood Cells and Platelets, also provides an automated blood antigen - typing algorithm from genome sequencing data (BloodTyper). A schematic representation of the data compilation process is summarized in **Figure 1**.

Compilation and annotation of genetic variants

All variants were first individually compiled to a master sheet and identifiers were mapped. Blood group IDs and gene names were obtained for each variant. Phenotype and allele names (as per ISBT nomenclature) were also collected. Genomic loci corresponding to these variants were obtained from the RefSeq database for both GRCh37/hg19 and GRCh38/hg38 human genome assemblies. Sequence variant descriptions for DNA, RNA and protein levels (according to HGVS nomenclature)²⁶ were added and verified using Mutalyzer^{27,28}. Variant annotation was performed using Annotate Variation (ANNOVAR)²⁹ integrating a number of public databases and computational tools. The allelic frequency of each variant was retrieved from global population datasets including the 1000 Genomes project³⁰, Exome Sequencing Project (ESP6500), Exome Aggregation Consortium (ExAC v.0.3)³¹, Genome Aggregation Database (gnomAD v2.1.1 Exomes) and The Greater Middle East (GME) Variome project³². Comprehensive list of datasets used for variant annotation is summarized in **Supplementary Table 1**. The clinical significance of the variants in the compendium was evaluated individually through manual literature searches and curation. Comparison of BGvar features with other existing databases is summarized in **Supplementary table 2**.

Database architecture

The variant data and corresponding annotations were transformed to JavaScript Object Notation format and ported onto MongoDB 3.4.1. The user-friendly web interface for querying the database was coded in PHP 7.0, AngularJS, HTML, Bootstrap 4 and CSS. The web server was configured in Apache HTTP server. The Highcharts javascript library was used for the graphical representation of variant data.

Results

Compendium of blood group variants

BGvar comprises 3224 human blood group related variants. This compilation precisely includes 1672 ISBT approved alleles and 1552 alleles predicted from 1000 Genomes Project and compiled from literature. The repository holds 1606 Single Nucleotide Variations (SNVs), 270 Insertions, Deletions (InDels) and duplications and about 1310 combination mutations corresponding to 41 human blood group systems and 2 transcription factors. Genetic variations due to events like gene fusion and rearrangements were also included. Of the total SNVs, 1016 were found to be non synonymous, 66 were synonymous, 126 and 2 were stop gain and stop loss variants respectively. Detailed summary of blood group related variants compiled in the resource is provided in **Table 1a and 1b**. **Figure 2** illustrates the category and function classification of the compiled variants.

Database features and navigation

BGvar has been designed to have a user friendly interface. Blood group systems in this database are organized on the homepage for rapid access to relevant information. In addition, the search interface enables the user to query the database based on gene name, genetic variant ID (rsID), nucleotide change, protein change and ISBT allele name. This database navigates through 3 major sections: (i) Basic information regarding the blood group system, providing a layman's summary of the blood group and its history; (ii) A blood group variant list with summary information and (iii) A detailed description of the variants associated with the blood group. Users are also provided with an option to search for ISBT approved alleles and other predicted and curated alleles separately. **Figure 3a and 3b** illustrates various sections of result display of the resource.

Blood group information

BGvar provides a brief summary pertaining to each blood group system. The section briefly details the history of the blood group system including its discovery, antigen expression patterns, molecular details and information regarding gene functionality. This is followed by the amino acid sequence of the protein, and information regarding reported disease associations, if any. Brief summary on the list of ISBT approved antigens and alleles pertaining to each blood group system along with the counts of alleles predicted from the 1000 Genomes project data and curated from literature sources are also provided. Additionally, this section details the clinical correlates or transfusion complications, if any, derived from literature.

Genetic Variant Information

This section details the genetic variants associated with each blood group. These include basic details pertaining to these variants, such as allele names, allele phenotypes, genomic loci, cDNA sequences and protein changes conforming to the HGVS nomenclature. This section also contains information regarding the corresponding global population frequencies of a given variant in major global datasets including the gnomAD (version 2.1.1 exomes), ExAC, 1000 Genomes Project, Greater Middle East and ESP6500. This page also details any known information regarding the clinical significance of the variant and phenotypic or genotypic descriptions thereof.

Community participation in variant curation/annotation

As blood group systems and genetic variants associated with blood groups are being continuously discovered, it is imperative that reference databases be up-to-date. BGvar therefore invites the community to contribute and enrich the utility this resource. BGvar provides an option for users to contribute any blood group related variant discoveries to the repository using a standard template. A detailed user manual to guide users in accessing the database is also provided.

Discussions

Genome-scale technologies have significantly impacted many clinical fields, and have promised to pave the way for precision medicine. Such genomic technologies are anticipated to significantly impact transfusion medicine, by facilitating a more precise understanding of the genetic repertoire of variants associated with blood group systems^{33,34}. Such an approach would have implications for the identification of rare blood groups, banking of blood and blood products and safer transfusion. As the numbers of new blood antigens being identified and reported continues to rise, a comprehensive compendium has the potential to serve as a reliable reference for transfusion research. BGvar is designed to fit this niche by compiling data pertaining to 3224 antigens belonging to 41 human blood group systems. We publish this resource in the hopes that it will be widely used by the community and will form a central resource for blood group immunogenetics.

Funding

This work was supported by The Council of Scientific and Industrial Research, India [MLP1806 RareGen]

Acknowledgements

MR curated the data for the database. KP designed the database. SJ and SHN helped in fact checking. MR and VS contributed in writing the manuscript. VS conceived and designed the project. All authors approved the final manuscript. Authors acknowledge help, support and constructive comments from Mukta Poojary, Bani Jolly and Ambily Sivadas. Authors acknowledge funding from CSIR India. The funders had no role in the preparation of the manuscript or decision to publish.

References

1. Mitra R, Mishra N, Rath GP. Blood groups systems. *Indian J Anaesth*. 2014;58(5):524-528.
2. Storry JR, Clausen FB, Castilho L, et al. International Society of Blood Transfusion Working Party on Red Cell Immunogenetics and Blood Group Terminology: Report of the Dubai, Copenhagen and Toronto meetings. *Vox Sang*. 2019;114(1):95-102.
3. Storry JR, Olsson ML. Genetic basis of blood group diversity. *British Journal of Haematology*. 2004;126(6):759-771. doi:10.1111/j.1365-2141.2004.05065.x
4. Yamamoto F-I. Molecular Genetics of the ABO Histo-Blood Group System. *Vox Sanguinis*.

1995;69(1):1-7. doi:10.1159/000462783

5. Yamamoto F-I, Clausen H, White T, Marken J, Hakomori S-I. Molecular genetic basis of the histo-blood group ABO system. *Nature*. 1990;345(6272):229-233. doi:10.1038/345229a0
6. Flegel WA, Wagner FF. Molecular genetics of RH. *Vox Sang*. 2000;78 Suppl 2:109-115.
7. Huang CH, Liu PZ, Cheng JG. Molecular biology and genetics of the Rh blood group system. *Semin Hematol*. 2000;37(2):150-165.
8. Flegel WA. Molecular genetics of RH and its clinical application. *Transfus Clin Biol*. 2006;13(1-2):4-12.
9. Ma CM, Wu QS, Yu TT, et al. [ABO gene subtypes and gene expression analysis in three cases of hematological malignancies patients]. *Zhonghua Yi Xue Za Zhi*. 2020;100(43):3443-3447.
10. Legge SE, Christensen RH, Petersen L, et al. The Duffy-null genotype and risk of infection. *Hum Mol Genet*. Published online September 22, 2020. doi:10.1093/hmg/ddaa208
11. Tuson M, Hue-Roye K, Koval K, et al. Possible suppression of fetal erythropoiesis by the Kell blood group antibody anti-Kp(a). *Immunohematology*. 2011;27(2):58-60.
12. Tilley L, Grimsley S. Is Next Generation Sequencing the future of blood group testing? *Transfus Apher Sci*. 2014;50(2):183-188.
13. Mori A, Deola S, Xumerle L, Mijatovic V, Malerba G, Monsurrò V. Next generation sequencing: new tools in immunology and hematology. *Blood Res*. 2013;48(4):242-249.
14. Lane WJ, Westhoff CM, Uy JM, et al. Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle. *Transfusion*. 2016;56(3):743-754.
15. Fürst D, Tsamadou C, Neuchel C, Schrezenmeier H, Mytilineos J, Weinstock C. Next-Generation Sequencing Technologies in Blood Group Typing. *Transfus Med Hemother*. 2020;47(1):4-13.
16. Hendrickson JE, Tormey CA, Shaz BH. Red blood cell alloimmunization mitigation strategies. *Transfus Med Rev*. 2014;28(3):137-144.
17. FY2017 Annual Summary.pdf.
18. RHCE Table. Accessed December 3, 2020. <https://www.bloodgroupgenomics.org/rhce/rhce-table/>
19. Patnaik SK, Helmsberg W, Blumenfeld OO. BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. *Nucleic Acids Res*. 2012;40(Database issue):D1023-D1029.
20. Möller M, Jöud M, Storry JR, Olsson ML. ErythroGene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. *Blood Adv*. 2016;1(3):240-249.
21. Wagner FF, Flegel WA. The Rhesus Site. *Transfus Med Hemother*. 2014;41(5):357-363.

22. Lane WJ, Westhoff CM, Gleadall NS, et al. Automated typing of red blood cell and platelet antigens: a whole-genome sequencing study. *Lancet Haematol.* 2018;5(6):e241-e251.
23. Lane WJ, Vege S, Mah HH, et al. Automated typing of red blood cell and platelet antigens from whole exome sequences. *Transfusion* . 2019;59(10):3253-3263.
24. Reid ME, Lomas-Francis C, Olsson ML. *The Blood Group Antigen FactsBook*. Academic Press; 2012.
25. Website. Accessed November 28, 2020. Dean L. Blood Groups and Red Cell Antigens [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK2261/>
26. den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat.* 2016;37(6):564-569.
27. Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PEM. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat.* 2008;29(1):6-13.
28. Vis JK, Vermaat M, Taschner PEM, Kok JN, Laros JFJ. An efficient algorithm for the extraction of HGVS variant descriptions from sequences. *Bioinformatics.* 2015;31(23):3751-3757.
29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
30. Consortium T 1000 GP, The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74. doi:10.1038/nature15393
31. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.
32. Scott EM, Halees A, Itan Y, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet.* 2016;48(9):1071-1076.
33. Liu Z, Liu M, Mercado T, Illoh O, Davey R. Extended Blood Group Molecular Typing and Next-Generation Sequencing. *Transfusion Medicine Reviews.* 2014;28(4):177-186. doi:10.1016/j.tmr.2014.08.003
34. Fichou Y, Mariez M, Le Maréchal C, Férec C. The experience of extended blood group genotyping by next-generation sequencing (NGS): investigation of patients with sickle-cell disease. *Vox Sang.* 2016;111(4):418-424.

BGvar - a comprehensive resource for blood group immunogenetics

Figure and Table legends

Figure 1 - Schematic summarising the data curation, database mapping, variant annotation and database/interface development of BGvar

Figure 2- Illustration of categories and functional classifications of variants compiled in the resource

Figure 3. Illustration of query search and sections of result display in BGvar. Homepage of the web resource with query search examples and user contribution options (*top*). Result display sections which provide brief information pertaining to the blood group system, list of alleles known and reported and the complete genomic summary of a selected variant (*bottom*).

Table 1a - Summary of ISBT approved human blood group alleles and their classifications

Table 1b - Summary of human blood group alleles predicted from the 1000 Genomes Project and curated from literature evidence

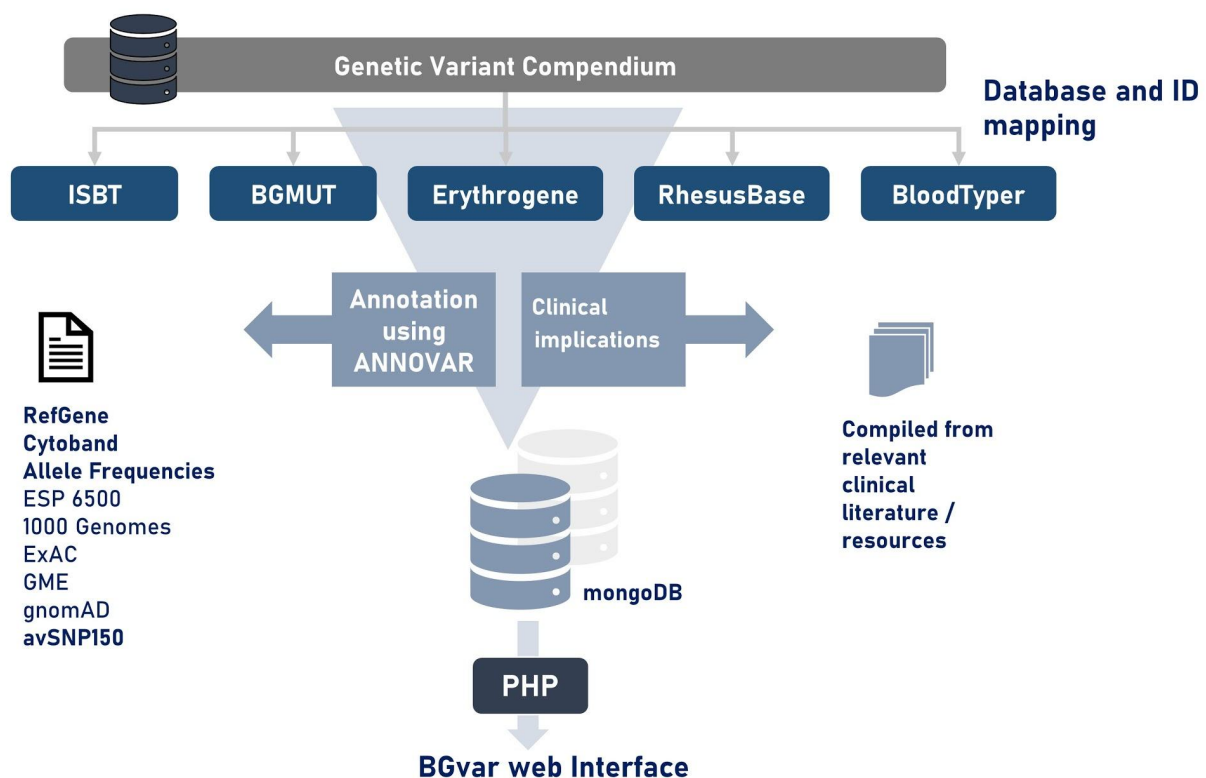


Figure 1. Schematic summary of data curation, database mapping, variant annotation and database/interface development for BGvar

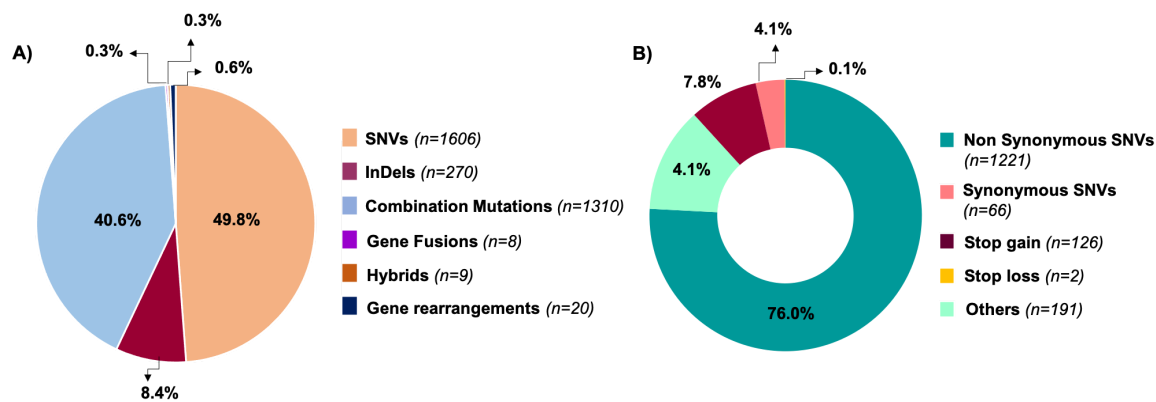
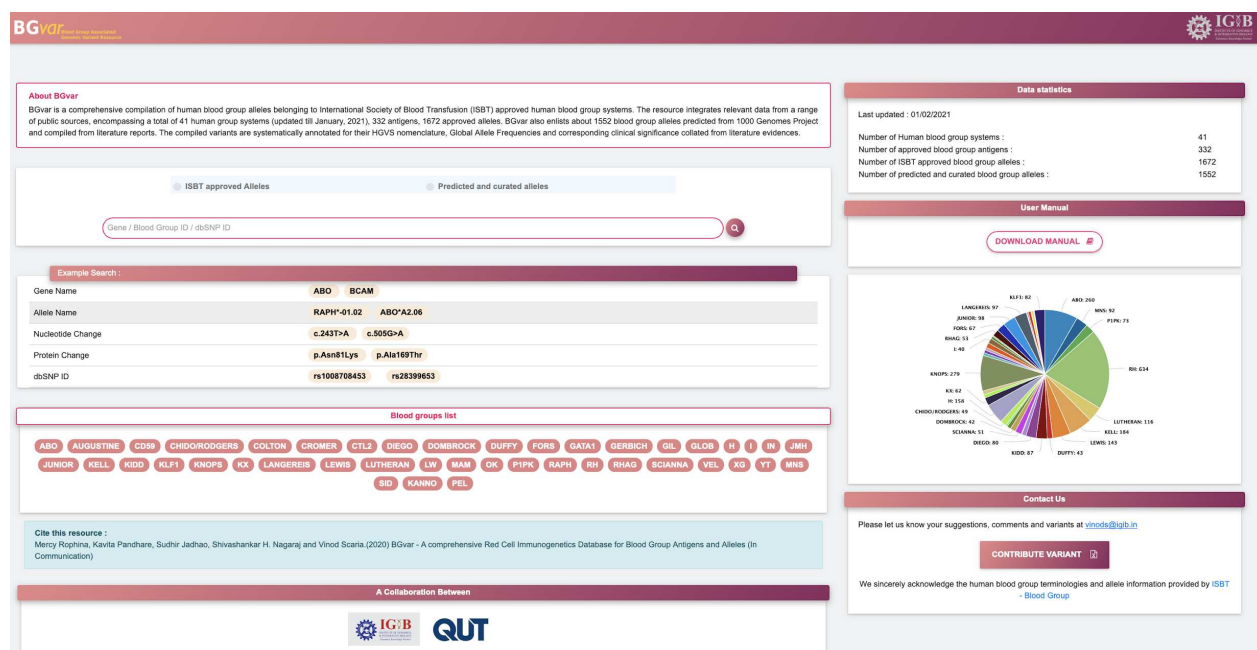


Figure 2. A) Distribution of various categories of human blood group related variations. **B)** Functional classification of Single Nucleotide Variants compiled in the resource.



COLTON	CO	015	9	5	4	0
LW	LW	016	2	1	1	0
CHIDO/RODGERS	CH/RG	017	5	3	0	2
H	H	018	99	74	14	9
KX	XK	019	60	22	38	0
GERBICH	GE	020	12	7	3	1
CROMER	CROM	021	17	16	1	0
KNOPS	KN	022	6	6	0	0
IN	IN	023	5	5	0	0
OK	OK	024	3	3	0	0
RAPH	RAPH	025	4	2	1	1
JMH	JMH	026	6	4	0	2
I	I	027	16	12	3	1
GLOB	GLOB	028	12	8	3	1
GIL	GIL	029	1	1	0	0
RHAG	RHAG	030	32	27	4	1
FORS	FORS	031	4	3	0	1
JUNIOR	JR	032	38	18	13	7
LANGEREIS	LAN	033	49	29	17	3
VEL	VEL	034	6	5	1	0
CD59	CD59	035	5	3	1	1
AUGUSTINE	AUG	036	4	4	0	0
KANNO	KANNO	037	1	1	0	0
SID	SID	038	1	1	0	0
CTL2	CTL2	039	2	1	1	0
PEL	PEL	040	1	0	1	0
MAM	MAM	041	5	0	4	1
GATA1	GATA1	Transcription Factor	1	1	0	0
KLF1	KLF1	Transcription Factor	57	38	15	4

Table 1a. Summary of ISBT approved blood group alleles in each blood group system.

Blood Group System	ISBT Blood Group ID	ISBT Blood Group System	Predicted and curated allele	SNVs Count	InDels Count	Combination Mutation Count
--------------------	---------------------	-------------------------	------------------------------	------------	--------------	----------------------------

		Number	count			
ABO	ABO	001	72	38	6	28
MNS	MNS	002	41	12	0	29
P1PK	P1PK	003	33	4	0	29
RH	RH	004	104	13	0	91
LUTHERAN	LU	005	80	32	0	48
KELL	KEL	006	71	32	0	39
LEWIS	LE	007	72	22	0	50
DUFFY	FY	008	13	2	0	11
KIDD	JK	009	27	0	0	27
DIEGO	DI	010	63	44	0	19
YT	YT	011	29	16	0	13
XG	XG	012	5	5	0	0
SCIANNA	SC	013	44	25	0	19
DOMBROCK	DO	014	28	8	0	20
COLTON	CO	015	22	20	0	2
LW	LW	016	11	10	0	1
CHIDO/RODGERS	CH/RG	017	44	17	0	27
H	H	018	79	34	8	37
KX	XK	019	2	2	0	0
GERBICH	GE	020	15	4	0	11
CROMER	CROM	021	16	16	0	0
KNOPS	KN	022	273	4	0	269
IN	IN	023	17	14	0	3
OK	OK	024	19	7	0	12
RAPH	RAPH	025	17	8	0	9
JMH	JMH	026	34	13	0	21
I	I	027	32	30	0	2
GLOB	GLOB	028	4	4	0	0
GIL	GIL	029	9	1	0	8
RHAG	RHAG	030	27	24	0	3
FORS	FORS	031	63	20	0	43
JUNIOR	JR	032	60	35	0	25
LANGEREIS	LAN	033	48	5	0	43
VEL	VEL	034	14	12	2	0
CD59	CD59	035	5	4	0	1
AUGUSTINE	AUG	036	18	18	0	0

KANNO	KANNO	037	0	0	0	0
SID	SID	038	0	0	0	0
CTL2	CTL2	039	0	0	0	0
PEL	PEL	040	0	0	0	0
MAM	MAM	041	0	0	0	0
GATA1	GATA1	Transcription Factor	16	14	1	1
KLF1	KLF1	Transcription Factor	25	11	1	13

Table 1b. Summary of blood group alleles in each blood group system predicted from the 1000 Genomes Project