

Structural bioinformatics

3DBIONOTES v3.0: crossing molecular and structural biology data with genomic variations

Joan Segura*, Ruben Sanchez-Garcia , C. O. S. Sorzano and J. M. Carazo

Spanish National Institute for Bioinformatics (INB ELIXIR-ES) and Biocomputing Unit, National Centre of Biotechnology (CSIC)/Instruct Image Processing Centre, 28049 Madrid, Spain

*To whom correspondence should be addressed.
Associate Editor: Alfonso Valencia

Received and revised on December 17, 2018; editorial decision on February 9, 2019; accepted on February 13, 2019

Abstract

Motivation: Many diseases are associated to single nucleotide polymorphisms that affect critical regions of proteins as binding sites or post translational modifications. Therefore, analysing genomic variants with structural and molecular biology data is a powerful framework in order to elucidate the potential causes of such diseases.

Results: A new version of our web framework 3DBIONOTES is presented. This version offers new tools to analyse and visualize protein annotations and genomic variants, including a contingency analysis of variants and amino acid features by means of a Fisher exact test, the integration of a gene annotation viewer to highlight protein features on gene sequences and a protein–protein interaction viewer to display protein annotations at network level.

Availability and implementation: The web server is available at <https://3dbionotes.cnb.csic.es>

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

Contact: Spanish National Institute for Bioinformatics (INB ELIXIR-ES) and Biocomputing Unit, National Centre of Biotechnology (CSIC)/Instruct Image Processing Centre, C/ Darwin n° 3, Campus of Cantoblanco, 28049 Madrid, Spain. jsegura@cnb.csic.es, jsegurmor@gmail.com

1 Introduction

Next-generation sequencing has flooded many databases with biomedical data where single-nucleotide variations are associated with phenotypes or diseases (Zerbino *et al.*, 2018). This information comprises collections of variant–disease pairs that can be used to infer which genomic variations might be involved in a particular disease. However, changes on the biochemical or structural features of the affected amino acids (if applicable) can be more informative in order to understand the causes of diseases. For that reason, some of the existent resources compiling variant–disease knowledge also annotates protein residues with biochemical features displaying what properties could be affected (Dingerdissen *et al.*, 2018).

In this work, we present a new version of 3DBIONOTES (Segura *et al.*, 2017; Tabas-Madrid *et al.*, 2016) where different analysis tools and viewers have been integrated to find how genomic variants may affect the different protein residues. 3DBIONOTES is

a web framework that integrates biological annotations and structural information of proteins from multiple sources (see [Supplementary Section S1](#)). In this version the application computes Fisher's exact test in order to find what biochemical or structural features are statistically affected by the variants associated to a particular disease. Moreover, a new panel displays those annotated regions where the co-occurrence between protein features and variants are statistically enriched. In addition, a gene annotation viewer has been fully integrated to display protein features at gene level. Also, a protein–protein interaction (PPI) viewer has been included in such a way that the different annotations can be displayed at network level. Finally, and as an additional tool, a new type of query, request by set of proteins, has been implemented to explore and analyse PPI networks. Moreover, custom annotations, including variants, can be submitted and analysed with the biological features integrated in the application.

2 New features

2.1 Gene annotation viewer

In this version of 3DBIONOTES, a gene annotation viewer has been fully integrated (see [Supplementary Section S2](#)). This panel displays gene information from ENSEMBL database ([Zerbino *et al.*, 2018](#)); the collected information includes: introns, exons, codifying regions and genomic variants. Moreover, ENSEMBL gene sequences are aligned with UniProt ([UniProt Consortium, 2018](#)) and PDB ([Burley *et al.*, 2018](#)) amino acids in such a way that protein annotated regions can be highlighted on gene sequences and vice versa (see [Supplementary Section S2](#)).

2.2 Genomic variants contingency analysis

In order to find which protein regions are statistically affected by the genomic variants associated to a particular disease, Fisher's exact test between the different annotations and the variants associated to the different diseases is computed. The main objective is to find co-occurrence of protein residues between the different structural or biochemical annotations and the variants associated to diseases. For example, most cancer related variants of the KRAS human protein map on its nucleotide binding region (see Section 3.1 and [Supplementary Section S3](#)).

2.3 Exploring PPI networks

A new type of query to request information for a set of proteins is now available. Moreover, a panel to visualize PPI networks using a graph-based representation has been integrated. This panel displays the physical binding between proteins when the information for a multimeric entry is requested or the PPIs that have been experimentally observed when a given set of proteins is submitted. In the first case, the contacts are computed using a distance threshold of 6 Å between heavy atoms. For the second case, PPI data is collected from Interactome3D ([Mosca *et al.*, 2013](#)). Moreover, the network panel can display annotations at network level using a similar approach as dSysMap ([Mosca *et al.*, 2015](#)) (see [Supplementary Section S4](#)).

2.4 Submitting custom annotations

The application supports the submission of custom data in such a way that users can analyse their own genomic variants or other annotations and compare them with 3DBIONOTES integrated data. The submitted information is fully integrated and the different visualization and analysis tools can be used to display and process the external data.

3 Use cases

3.1 Analysis on the human KRAS genomic variants

In this example we have analysed the genomic variants associated to KRAS human protein (UniProt accession P01116). KRAS protein is a GTPase that acts as a signalling switch in many transduction pathways including cell proliferation. The active state of KRAS occurs when the protein is bound to GTP. In this state, the protein recruits and activates other growth factors and cell signalling receptors. Upon GTP hydrolysis and conversion to GDP, KRAS is inactivated. KRAS mutations are known to be involved in different diseases such as multiple cancer types or neurofibromatosis ([Simanshu, *et al.*, 2017](#)). We used 3DBIONOTES to analyse the co-occurrence of KRAS variants associated to diseases with the different biochemical annotations. The main reason was to check whether those variants occur in particular regions of KRAS or randomly distributed. [Supplementary Figure S4](#) and [Supplementary Tables S1 and S2](#) display the analysis panel of

3DBIONOTES and clearly show that many of those variants occur in the 'Nucleotide Binding Site' annotated regions. KRAS acts as on/off switch in many processes and its active or inactive form depends on the interaction with GTP or GDP, respectively. Then, mutations affecting KRAS GTP/GDP-binding sites may affect its activation and therefore, many cell regulatory processes.

3.2 GNB1 neurodevelopmental disability

This example illustrates how 3DBIONOTES can be used to analyse external variants. We have collected the variants of the G protein subunit beta (GNB1) associated to neurodevelopmental delay, hypotonia and seizures available in the work of [Petrovski *et al.* \(2016\)](#) (see [Supplementary Table S3](#)). GNB1 protein modulates transmembrane signalling pathways controlled by G protein-coupled receptors. We have requested the PPI network information for the GNB1 protein (UniProt accession P62873) and attached the collected variants to 3DBIONOTES. When the variants are mapped to the PPI network, many of them appear affecting the binding sites between GNB1 and other G proteins. Moreover, the contingency analysis identified that the co-occurrence between many of the GNB1-binding sites and the submitted variants was statistically significant (see [Supplementary Fig. S8](#)). Consequently, mutations of the GNB1-binding sites may affect the interaction with other G proteins and, thus, some of the cell signalling pathways involving G proteins.

Funding

This work was supported by Ministerio de Economía, Industria y Competitividad, Gobierno de España [grant No. BIO2016-76400-R(AEI/FEDER, UE)]; Comunidad de Madrid [grant No. S2017/BMD-3817]; Instituto de Salud Carlos III [grant No. PT13/0001/0009; INB Grant PT17/0009/0010 - ISCIII-SGFI/ERDF]; Horizon 2020 [grant No. Elixir - EXCELERATE INFRADEV-3-2015, Proposal 676559] and iNEXT [INFRAIA-1-2014-2015, Proposal 653706]; Ministerio de Ciencia, Innovación y Universidades, Gobierno de España [Juan de la Cierva-E-28-2018-0015407 to J.S.]; and Ministerio de Educación, Cultura y Deporte [FPU-2015/264 to R.S.-G.].

Conflict of Interest: none declared.

References

- Burley,S.K. *et al.* (2018) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.*, **47**, D464–D474.
- Dingerdisen,H.M. *et al.* (2018) BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. *Nucleic Acids Res.*, **46**, D1128–D1136.
- Mosca,R. *et al.* (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.
- Mosca,R. *et al.* (2015) dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods*, **12**, 167–168.
- Petrovski,S. *et al.* (2016) Germline De Novo Mutations in GNB1 Cause Severe Neurodevelopmental Disability, Hypotonia, and Seizures. *Am. J. Hum. Genet.*, **98**, 1001–1010.
- Segura,J. *et al.* (2017) 3DBIONOTES v2.0: a web server for the automatic annotation of macromolecular structures. *Bioinformatics*, **33**, 3655–3657.
- Simanshu,D.K. *et al.* (2017) RAS proteins and their regulators in human disease. *Cell*, **170**, 17–33.
- Tabas-Madrid,D. *et al.* (2016) 3DBIONOTES: a unified, enriched and interactive view of macromolecular information. *J. Struct. Biol.*, **194**, 231–234.
- UniProt Consortium. (2018) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Zerbino,D.R. *et al.* (2018) Ensembl 2018. *Nucleic Acids Res.*, **46**, D754–D761.

3DBIONOTES v3.0: Crossing molecular and structural biology data with genomic variations (Supplementary Material)

J. Segura, R. Sanchez-Garcia, C.O.S. Sorzano, J.M. Carazo

Spanish National Institute for Bioinformatics (INB ELIXIR-ES) and Biocomputing Unit, National Centre of Biotechnology (CSIC), C/ Darwin nº 3, Campus of Cantoblanco, 28049 Madrid, Spain.

S1 3DBIONOTES v3.0 sources of information

3DBIONOTES integrates information from multiple sources of different nature:

1. Posttranslational modifications from PhosphoSitePlus (Hornbeck, et al., 2015) and PTMdb (Huang, et al., 2016)
2. Genetic variants and diseases from dSysMap (Mosca, et al., 2015) and BioMuta (Wu, et al., 2014)
3. Short linear motifs from the Eukaryotic Linear Motifs database (Dinkel, et al., 2016)
4. Epitopes from the Immune Epitope database (Vita, et al., 2015)
5. Protein disorder information from MobiDB database (Necci, et al., 2017)
6. Protein domain families from Pfam (Finn, et al., 2005), SMART (Letunic, et al., 2015) and InterPro (Finn, et al., 2017)
7. Annotations from UniProt database (UniProt, 2015)
8. Genomic information from ENSEMBL database (Zerbino, et al., 2018)

S2 The gene annotations panel

In this version of 3DBIONOTES, a fully integrated panel to visualize gene annotations has been included (see Figure S1). The main function of this component is to display the gene information of the selected proteins and to highlight the integrated annotations at gene sequence. The gene panel collects and displays the information from ENSEMBL database (Zerbino, et al., 2018) including exons, introns, codifying regions and genomic variants. Gene and protein sequences are aligned in such a way that 3DBIONOTES can highlight the selected annotations at gene sequence level or vice versa (see Figure S1 and Figure S2). To that end, gene and protein

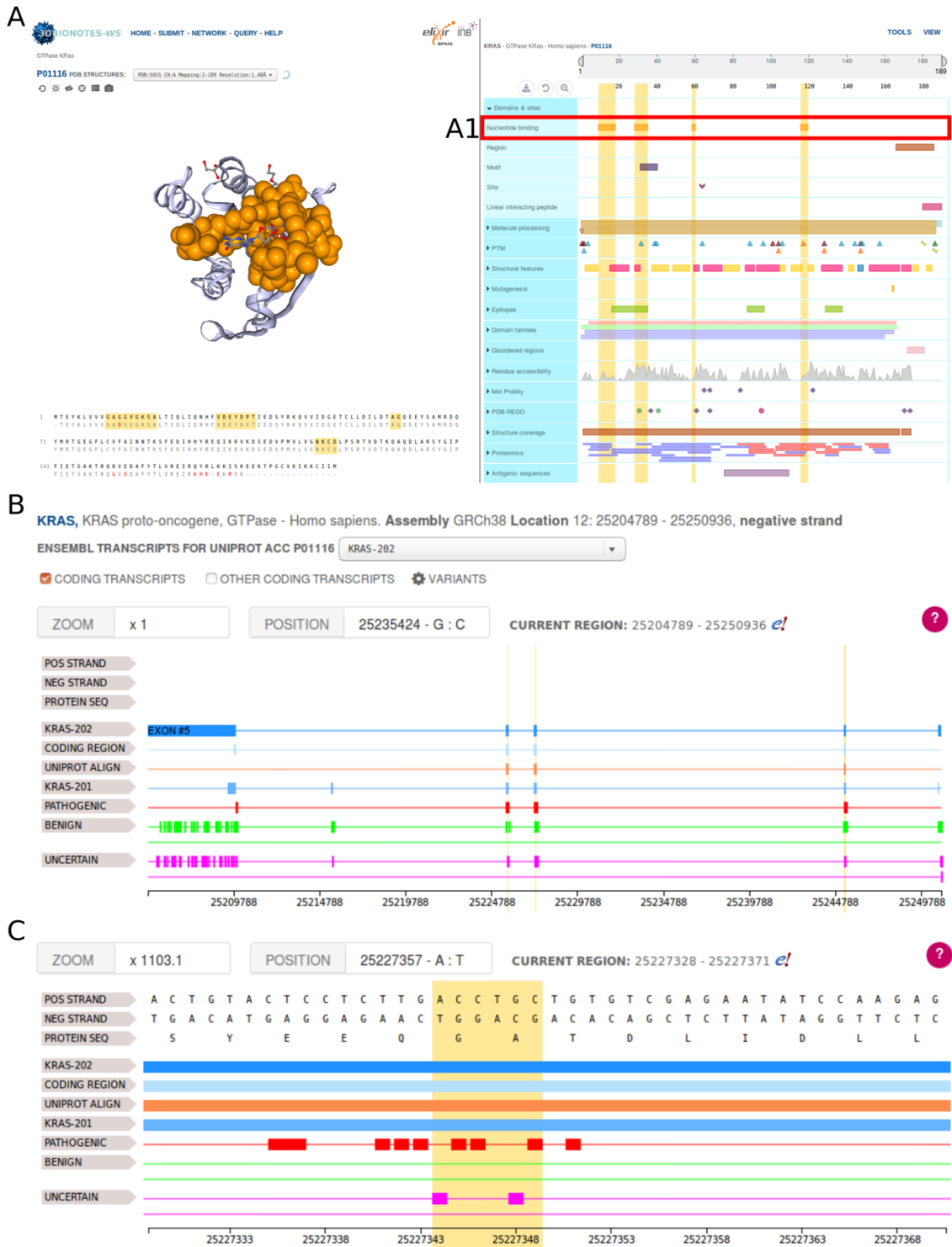


Figure S1 Integration of the gene annotation panel in 3DBIONOTES. (A) Selection and visualization of 'Nucleotide Binding Site' annotations (red box marked as A1) of the KRAS protein (UniProt accession P01116) at structure and sequence level. (B) Gene annotation panel for the KRAS protein in 3DBIONOTES and the visualization of the selected 'Nucleotide Binding Site' annotation at gene level. (C) Zoom within the third exon of KRAS gene where some of the selected 'Nucleotide Binding Site' annotations can be appreciated at gene sequence scale.

alignments are computed in three different steps. First, ENSEMBL gene and protein sequences are aligned. This is a simple step since ENSEMBL protein sequences are translated from ENSEMBL genes. Second, ENSEMBL protein sequences are aligned with their corresponding UniProt sequences using the Smith-Waterman algorithm (Smith and Waterman, 1981). Finally, UniProt and PDB sequences are aligned using the SIFT service (Velankar, et al., 2013), when suitable, or the Smith-Waterman algorithm otherwise. The gene annotation viewer is a modified version of the neXtprot viewer (Schaeffer, et al., 2017) where some additional features, such as multiple selection or multiple sequence visualizations, have been included.

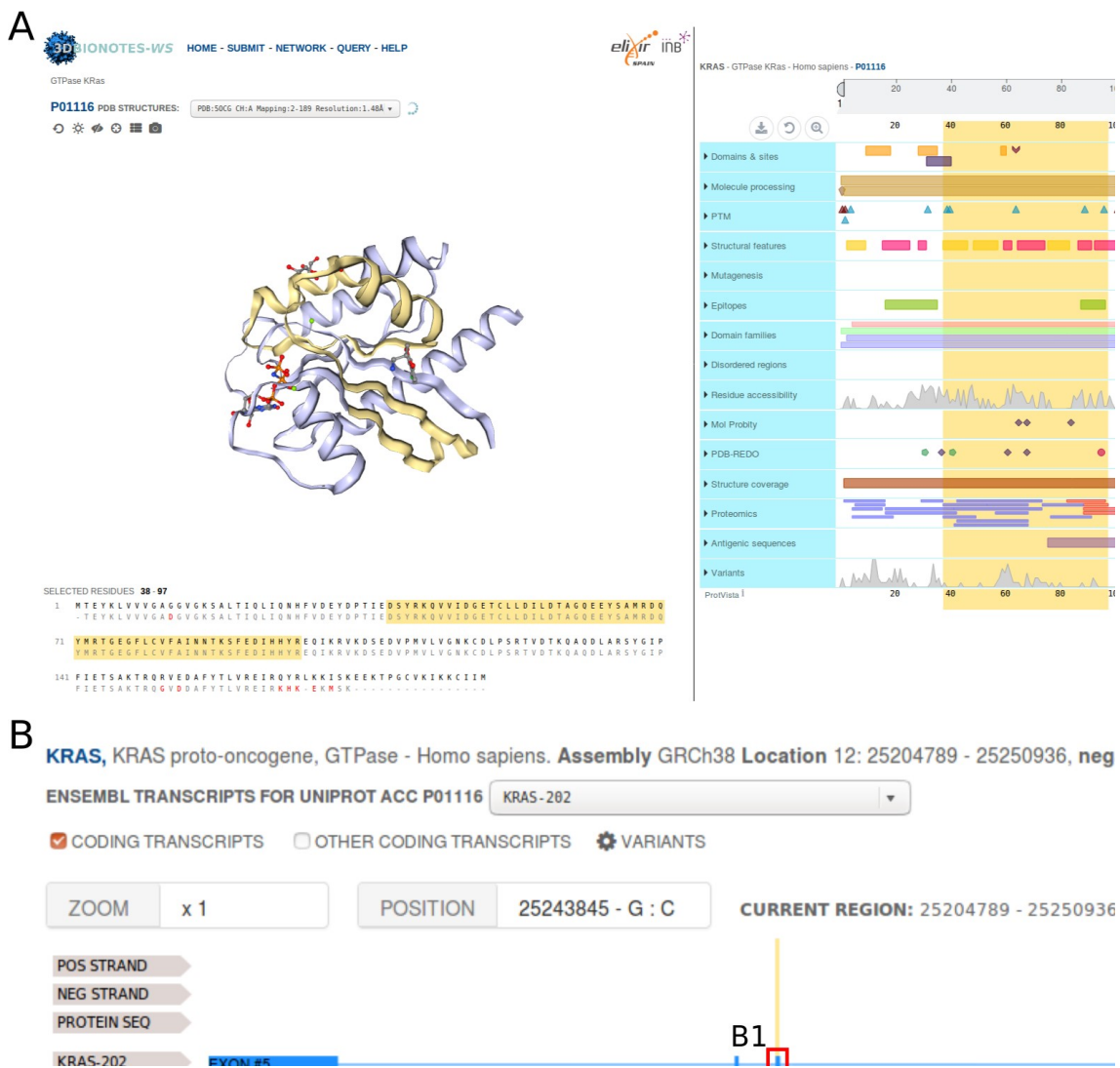


Figure S2 Display of gene annotations at protein level. (A) Highlighted region at protein sequence and structure level of the KRAS gene third exon. (B) Gene annotation viewer displaying KRAS annotation where the third exon has been selected (red box marked as B1).

S3 Contingency between genomic variants and biological annotations.

One of the main tools in this version is the contingency analysis between genomic variants and integrated features. For that purpose, the application computes Fisher's exact test between the variants associated to a particular disease and the protein regions corresponding to the different biochemical annotations in a similar way as was done in dSysMap (Mosca, et al., 2015) with variants and protein-protein binding sites. For example, Figure S3 displays the cooccurrence between the genomic variants associated to colorectal cancer for KRAS protein and its nucleotide binding site (NBS) that is comprised of 22 residues. The analysis viewer (Figure S3) represents the NBS regions with orange boxes while the genomic variants are displayed in red or grey dots depending whether they map on NBS amino acids or not. A total of 22 variants associated with colorectal cancer can be observed. From those, 12 variants were located at NBS residues while the remaining variants affected other amino acids. In this case, Fisher's exact test computed a p-value of 2.53e-6.



Figure S3 3DBIONOTES analysis viewer. Visualization of cooccurrence between KRAS colorectal cancer variants and KRAS nucleotide binding site (NBS). NBS regions are displayed in orange boxes, variants affecting NBS residues are shown in red, while variants outside NBS regions in grey. Additionally, the contingency table for colorectal cancer variants and NBS residues is displayed on the right bottom side of the figure.

The analysis panel displays those diseases for which the cooccurrence of their associated variants with a specific biochemical feature is statistically significant when Fisher's exact test is computed. The statistical significance is defined in terms of Fisher's exact test p-value and the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) with a false discovery rate of 1% to control the amount of false positives or Type I error. This information is displayed in two different tracks (see Figure S4): the first track displays the biochemical feature regions (Figure S4A) and the second track the list of diseases with the results of Fisher's exact test (Figure S4B) and the genomic variants (Figure S4C) using a colour code of red, if they are within the selected biochemical feature region (NBS's, in this case), or grey, if they are outside. Table S1 shows the full list of statistically significant diseases affecting KRAS nucleotide binding sites that can be found in the analysis panel (Figure S4B).

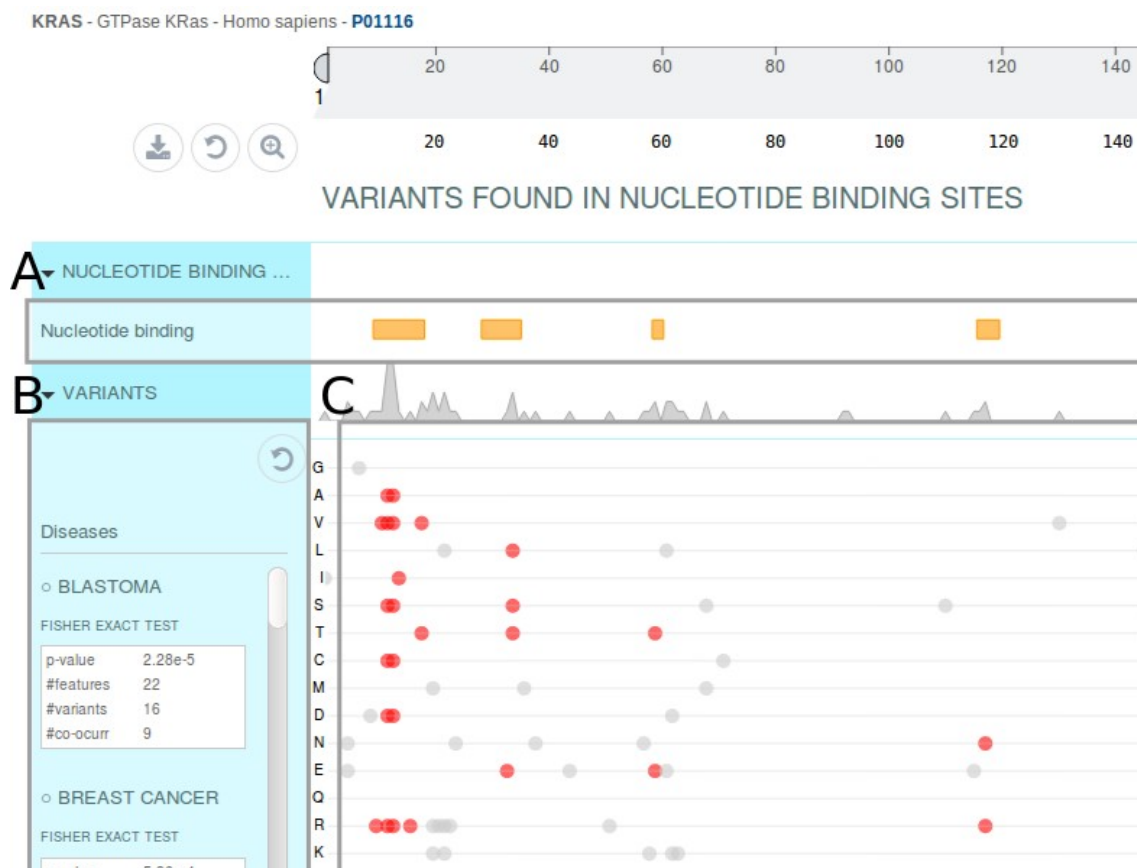


Figure S4 Contingency analysis viewer for the KRAS protein. (A) Biochemical feature track. In this example the nucleotide binding site of KRAS protein is displayed in orange boxes. (B) List of diseases and results of Fisher's exact test where the cooccurrence of genomic variants and the nucleotide binding region was statistically significant. (C) Specific location of the genomic variants, in red colour those that mapped on the nucleotide binding site and in grey if they map outside.

Table S1 KRAS genomic variants associated to diseases cooccurring in the nucleotide binding region.

First column, disease name. Second column, number of variants associated to the corresponding disease. Third column, number of variants cooccurring in the nucleotide binding site region of KRAS. Four column, p-value computed by Fisher's exact test.

DISEASE NAME	#VARIANTS	#COOCURRENCES	P-VALUE
PANCREATIC CANCER	11	10	1,09E-08
STOMACH CANCER	12	9	6,25E-07
LUNG CANCER	21	12	1,19E-06
COLORECTAL CANCER	22	12	2,53E-06
HEMATOLOGIC CANCER	13	9	2,61E-06
BLASTOMA	16	9	2,28E-05
GASTRIC CANCER	6	5	1,16E-04
URINARY BLADDER CANCER	6	5	1,41E-04
UTERINE CANCER	24	10	3,58E-04
CERVICAL CANCER	7	5	3,72E-04
BREAST CANCER	7	5	5,36E-04
LEUKEMIA	3	3	1,57E-03
KIDNEY CANCER	3	3	1,77E-03
MELANOMA	30	10	2,28E-03
LIVER CANCER	10	5	2,83E-03

S4 The protein-protein interaction network viewer.

A new tool of this version is the protein-protein interaction (PPI) network viewer that can be used to display biological annotations at PPI network level. When the user requests the information for a multimeric entry, thus, a protein complex with multiple subunits, 3DBIONOTES can display the entry at network level using a graph representation. In this graph, nodes are proteins and edges interactions, i.e. physical contacts, between the proteins. The application can also map the different annotations at network level. As examples, Figure S5 shows phosphorylation (blue colour) and ubiquitination (orange colour) sites mapped on the Human APC/C-CDH1-EMI1 complex (EMDB code EMD-2924). The annotations can be located in four different positions. (1) within a graph node, meaning that the annotation occurs in some buried residues of the protein. (2) on a graph node border, representing that the annotation occurs on some surface residues. (3) on a graph edge, when the annotation is located on the residue binding site. Finally, (4) under the graph node name, meaning in this case that the annotation occurs in a region of the protein sequence that has not been solved at structural level.

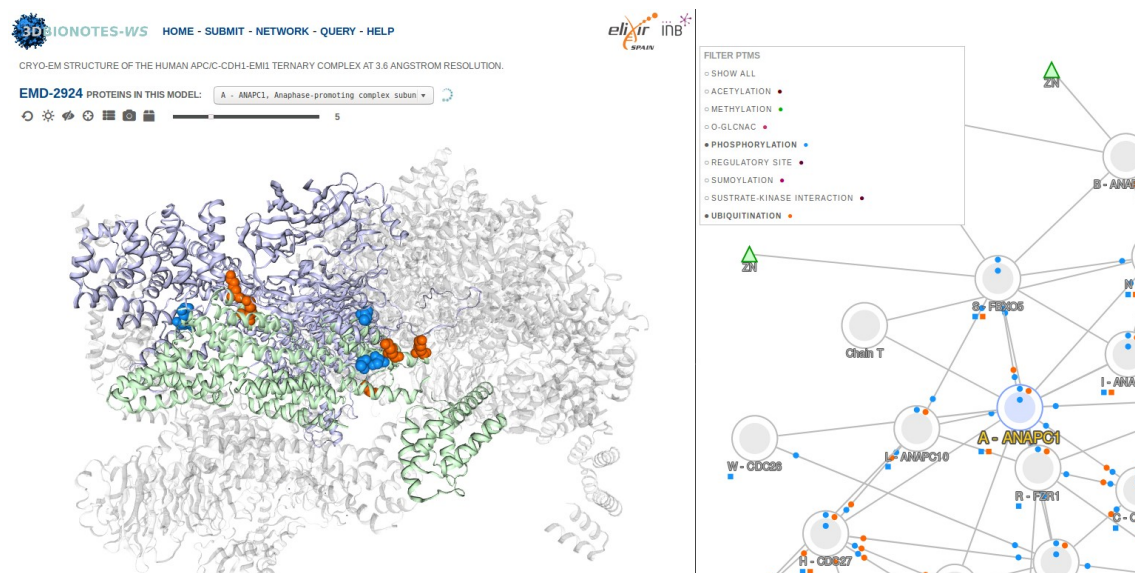


Figure S5 Protein-protein network viewer of 3DBIONOTES. On the right hand side of the image, graph representation of the Human APC/C-CDH1-EMI1 complex (EMDB code EMD-2924). Graph nodes represent proteins and edges interactions (physical bindings) between proteins. The phosphorylation (blue) and ubiquitination (orange) sites are displayed at network level. On the left hand side of the image, interaction between anaphase-promoting subunit 1

(blue) and subunit 5 (green), where the interface phosphorylation and ubiquitination residues are highlighted in blue and orange colour, respectively.

The PPI network viewer also interacts with the contingency analysis of the genomic variants. In this way, when 3DBIONOTES visualizes the genomic variants at network level, those variants associated to a particular disease whose cooccurrence with a specific biochemical feature is statistically significant (see Section S3) are displayed with two concentric circles (see Figure S6).

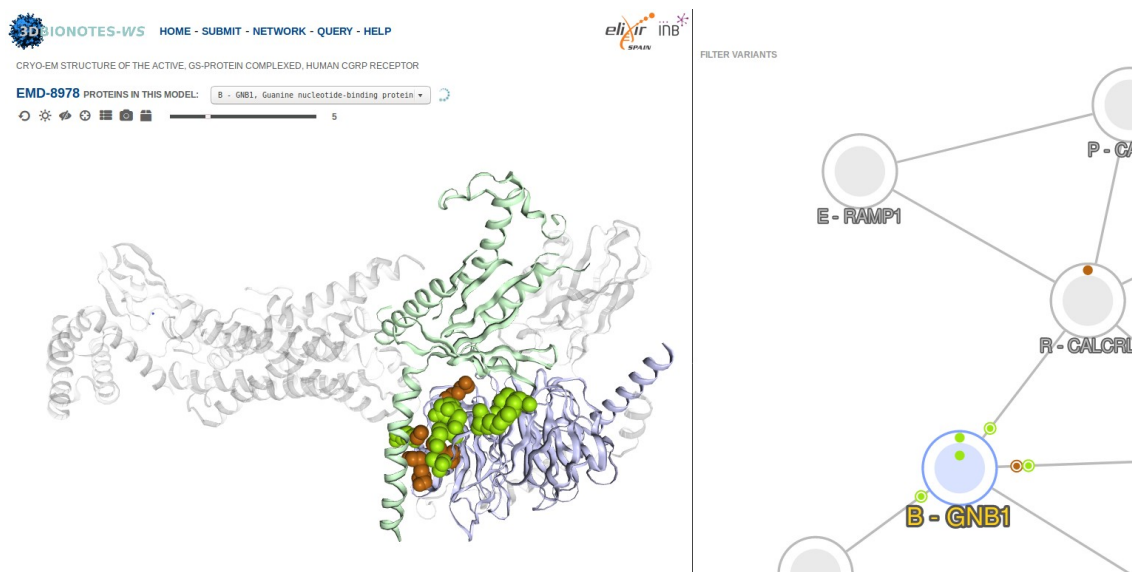


Figure S6 Protein-protein network viewer displaying genomic variants. In this example 3DBIONOTES displays on the right hand side the genomic variants associated to mental disability (brown) and hematologic cancer (green) on the Active GS complex in human. The concentric circles on the graph edges indicate that the cooccurrence between those variants and the protein binding site is statistically significant in terms of Fisher's exact test (see Section S3). The structural panel on the left hand side highlights the interaction between the proteins GNB1 and GNAS and the variants that mapped on the interface.

Additionally, this version includes a new type of query: query by a set of UniProt accessions. In this type of query the application collects the PPI structural information that is available from Interactome3d database (Mosca, et al., 2013) and builds a PPI network. Then, the application allows the user to explore the biological annotations for the different proteins and interactions of the network. All analysis tools and viewers of 3DBIONOTES are available for this type of query. Figure S7 shows the resulting network when the requested information included simultaneously KRAS (UniProt accession P01116), NRAS (UniProt accession P01111) and HRAS (UniProt

accession P01112) Human proteins. The network includes all interacting proteins involving the set of requested proteins and their neighbours for which structural data is available. Finally, in this example, the application is displaying the phosphorylation (blue colour) and methylation (green colour) sites at PPI network level. On the left side of Figure S7, the structural panel displays the interaction between NRAS and RASA1 proteins, highlighting the phosphorylated and methylated residues of the interface in blue and green colour, respectively.

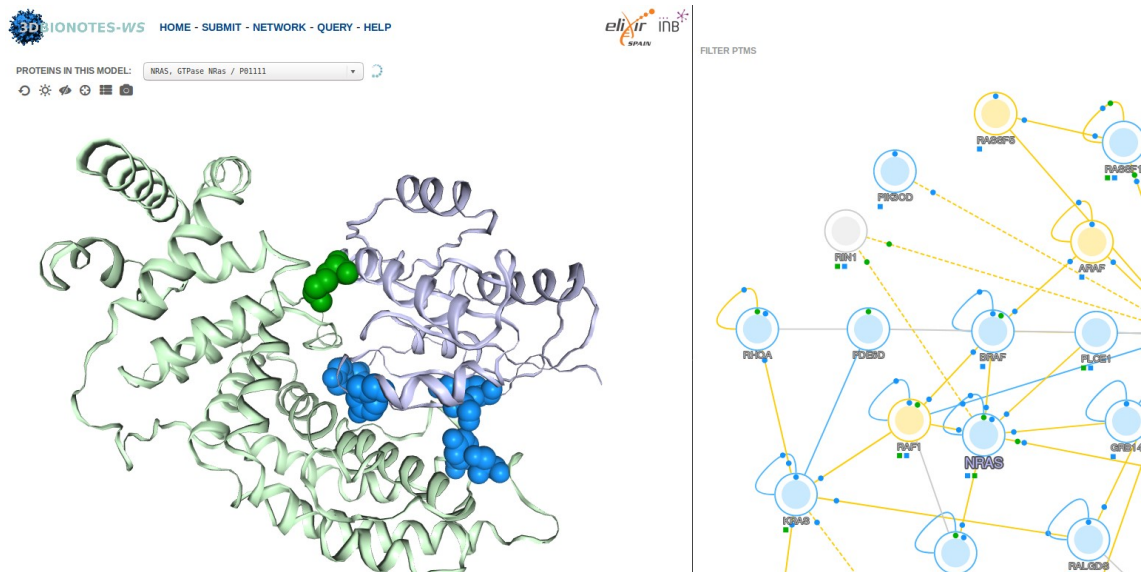


Figure S7 Protein-protein interaction network of HRAS, KRAS and NRAS. On the right site of the image it is shown the protein-protein interaction network involving HRAS, KRAS and NRAS and all proteins which interacts at least with one of them and for which structural information of the interaction is available. Blue nodes of the network indicate that experimental structure data is available for the proteins, while for the yellow nodes only homology models have been obtained. Colour edges also indicate whether structure information is available from experimental data, in blue, or from homology models, in yellow. In addition, discontinuous lines indicate that the structural models are obtained from domain-domain interaction templates.

S5 Use cases.

S5.1 Analysis on human KRAS genomic variants

In this example we used 3DBIONOTES to analyse the genomic variants associated to diseases for KRAS Human protein (UniProt accession P01116). The main purpose was to check whether these variants were affecting specific regions with particular biochemical features or were randomly distributed. From the analysis panel of 3DBIONOTES (Figure S4) we could observe that most disease variants were located in the nucleotide binding site of KRAS. In particular,

from 22 different diseases associated with variants that affect KRAS amino acids (see Table S2), the cooccurrence of the genomic variants associated to 16 of these diseases (see Table S1) with KRAS nucleotide binding site was statistically significant when the contingency analysis was computed (see Section S3).

Table S2 3DBIONOTES diseases associated to variants for KRAS protein

Blastoma	Malignant glioma
Brain cancer	Malignant mesothelioma
Breast cancer	Melanoma
Cervical cancer	Ovarian cancer
Colorectal cancer	Pancreatic cancer
Esophageal cancer	Sarcoma
Germ cell cancer	Stomach cancer
Hematologic cancer	Thymus cancer
Kidney cancer	Thyroid carcinoma
Liver cancer	Urinary bladder cancer
Lung cancer	Uterine cancer

A possible explanation is that KRAS protein acts as a switch (on/off) in the regulation of many signalling pathways and its state depends on whether KRAS interacts with GTP or GDP. Then, mutations that affect the nucleotide binding region of KRAS may be affecting the regulation of many signalling pathway and finally giving rise to different cancer diseases. In this example, the contingency analysis of 3DBIONOTES has found a clear relation between KRAS nucleotide binding site and the variants associated to different cancer diseases.

S5.2 GNB1 neurodevelopmental disability

In this example we have analysed the GNB1 mutations that were studied by Petrovski et al. (Petrovski, et al., 2016) using 3DBIONOTES to analyse external data (see Section 2.4 in Main Text). We used the option 'query by set of proteins' to study how the mutations were affecting GNB1 and its interacting partners for which structural information is known. Table S3 contains the list of mutated amino acids of GNB1 that were submitted and analysed.

Table S3 Observed mutations on the GNB1 protein residues

53	76	77	78	80	89
92	95	101	114	118	326

Figure S8 displays the PPI network of GNB1 and its interacting partners for which structural information is available. The external variations that were submitted to the application were mapped at network level in cyan colour. We can observe that most of the GNB1 interaction binding sites were affected by the submitted variants (Figure S8B and S8E). Moreover, the contingency analysis shows that the cooccurrence between GNB1 binding sites and submitted variants is statistically significant for many of the GNB1 interactions. This example illustrates how the combination of contingency analysis and network viewer can be helpful in order to study the effects of genomic variations.

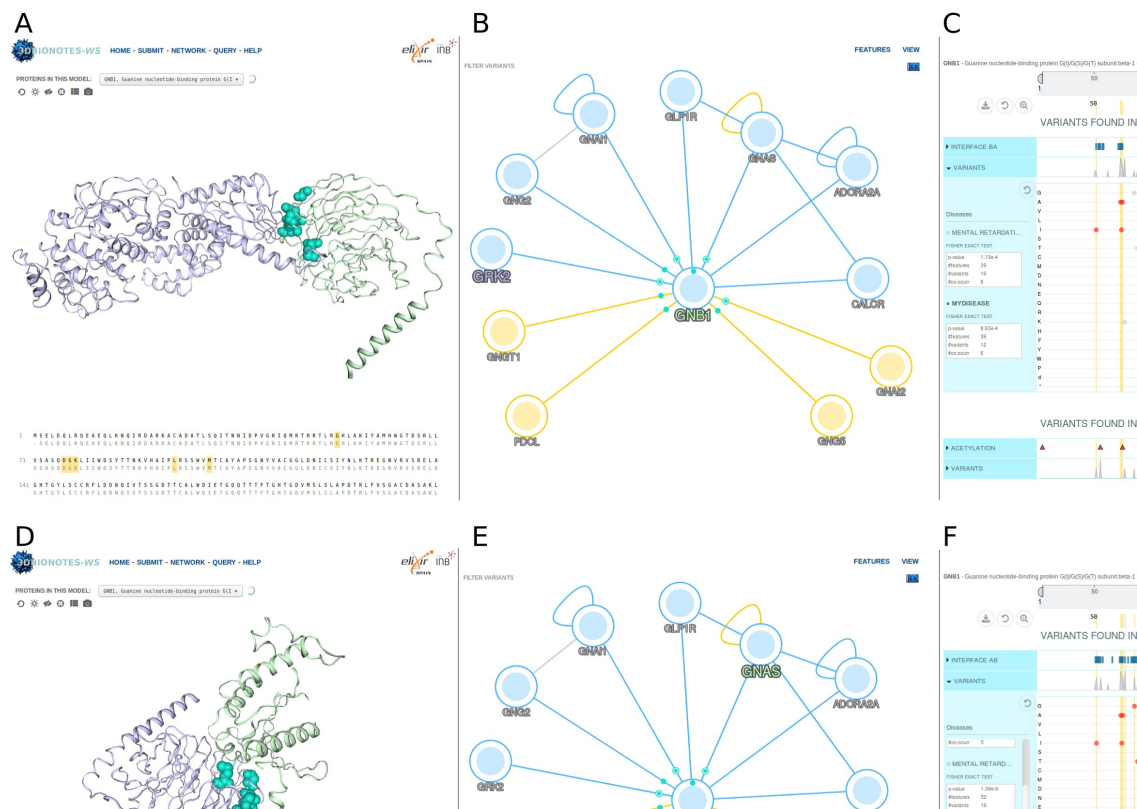


Figure S8 Analysis of external variants for the GNB1 protein. (A) Interaction between GNB1 (green colour) and GRK2 (blue colour), the variants affecting residues on the interface are highlighted in cyan colour. (B) PPI network of GNB1 where the interaction with GRK2 has been selected and the external variants are mapped and displayed in cyan colour. (C) Contingency analysis panel displaying the cooccurrence of the external variants and GNB1-GRK2 binding site. (D) Interaction between GNB1 (blue colour) and GNAS (green colour), the variants

affecting residues on the interface are highlighted in cyan colour. (E) PPI network of GNB1 where the interaction with GNAS has been selected and the external variants are mapped and displayed in cyan colour. (C) Contingency analysis panel displaying the cooccurrence of the external variants and GNB1-GNAS binding site.

References

- Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995;57(1):289-300.
- Dinkel, H., *et al.* ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic acids research* 2016;44(D1):D294-300.
- Finn, R.D., *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic acids research* 2017;45(D1):D190-D199.
- Finn, R.D., Marshall, M. and Bateman, A. iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 2005;21(3):410-412.
- Hornbeck, P.V., *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research* 2015;43(Database issue):D512-520.
- Huang, K.Y., *et al.* dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. *Nucleic acids research* 2016;44(D1):D435-446.
- Letunic, I., Doerks, T. and Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic acids research* 2015;43(Database issue):D257-260.
- Mosca, R., Ceol, A. and Aloy, P. Interactome3D: adding structural details to protein networks. *Nat Methods* 2013;10(1):47-53.
- Mosca, R., *et al.* dSysMap: exploring the edgetic role of disease mutations. *Nat Methods* 2015;12(3):167-168.
- Necci, M., *et al.* MobiDB-lite: Fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics* 2017.
- Petrovski, S., *et al.* Germline De Novo Mutations in GNB1 Cause Severe Neurodevelopmental Disability, Hypotonia, and Seizures. *Am J Hum Genet* 2016;98(5):1001-1010.
- Schaeffer, M., *et al.* The neXtProt peptide uniqueness checker: a tool for the proteomics community. *Bioinformatics* 2017;33(21):3471-3472.
- Smith, T.F. and Waterman, M.S. Identification of common molecular subsequences. *Journal of molecular biology* 1981;147(1):195-197.
- UniProt, C. UniProt: a hub for protein information. *Nucleic acids research* 2015;43(Database issue):D204-212.

- Velankar, S., *et al.* SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic acids research* 2013;41(Database issue):D483-489.
- Vita, R., *et al.* The immune epitope database (IEDB) 3.0. *Nucleic acids research* 2015;43(Database issue):D405-412.
- Wu, T.J., *et al.* A framework for organizing cancer-related variations from existing databases, publications and NGS data using a High-performance Integrated Virtual Environment (HIVE). *Database (Oxford)* 2014;2014:bau022.
- Zerbino, D.R., *et al.* Ensembl 2018. *Nucleic acids research* 2018;46(D1):D754-D761.