

Prioritization of variants and genes: *BiERapp*

Álex Alemán

March 1st, 2016



GDA

International Course on
Genomic **D**ata **A**nalysis



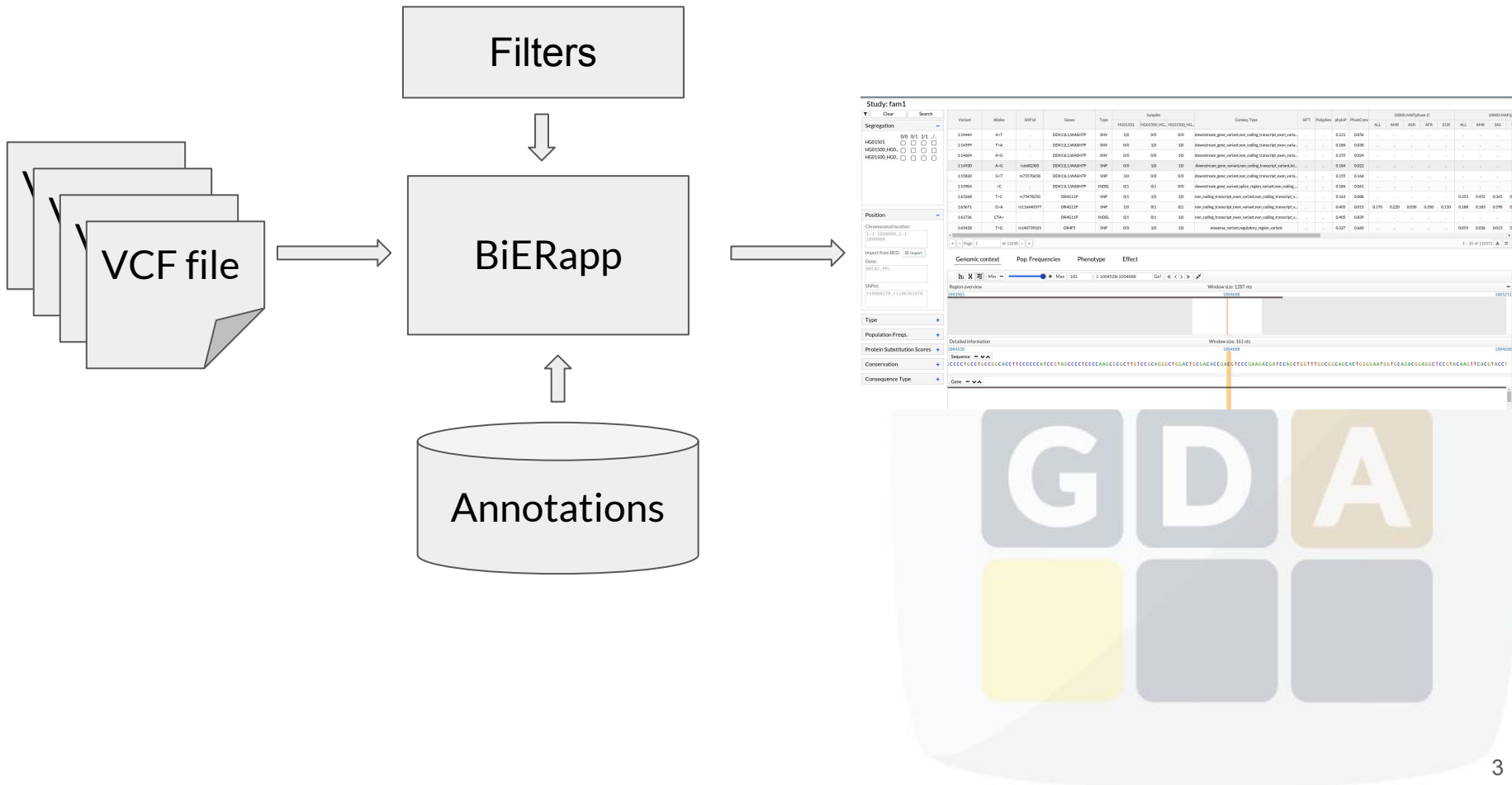
PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Introduction

- ❑ Whole-exome sequencing has become a fundamental tool for the discovery of disease-related genes of familial diseases but there are difficulties to **find the causal mutation among the enormous background.**
- ❑ There are different scenarios, so we need **different and immediate strategies of prioritization.**
- ❑ Vast amount of **biological knowledge available** in many databases.
- ❑ We need a tool to **integrate this information and filter immediately** to select candidate variants related to the disease



How does BiERapp work?



Input: VCF file

Primary Analysis

1. Sequence preprocessing

2. Mapping

3. Variant calling

4. Variant prioritization

→ VCF files

G D A

BiERapp

Secondary

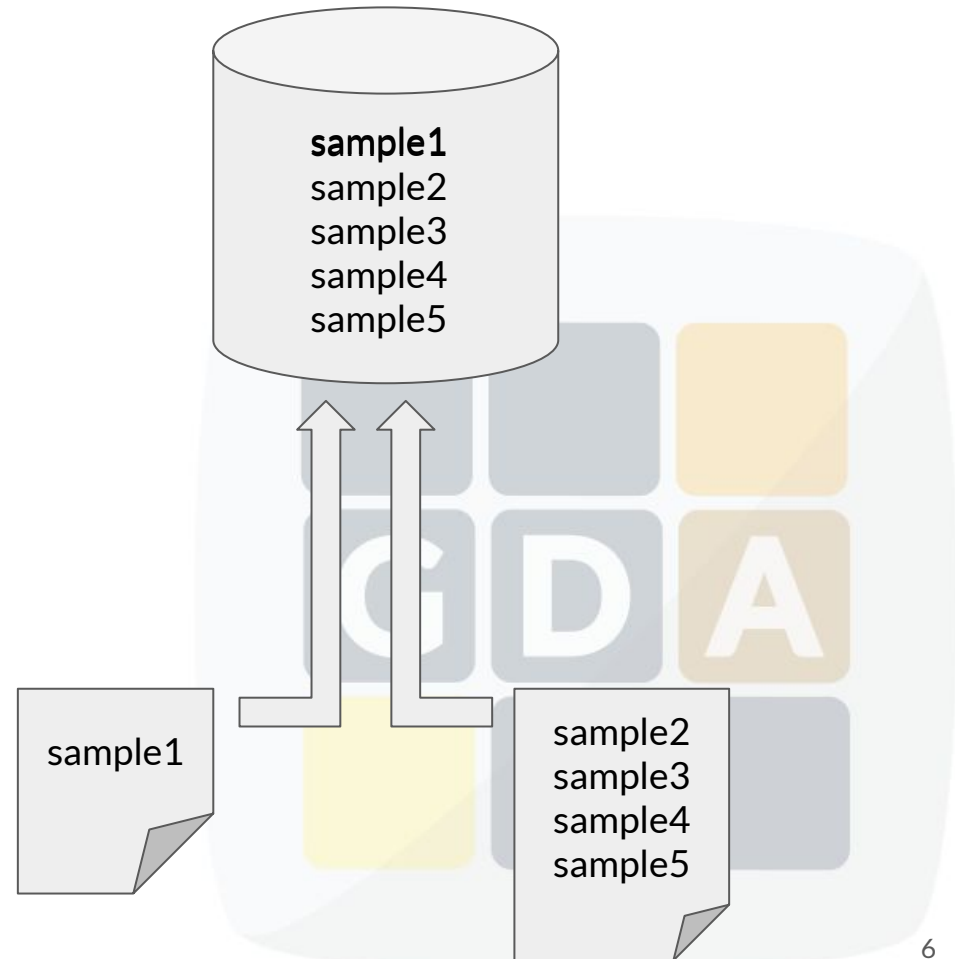
Input: VCF file

```
##fileformat=VCFv4.1
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,md5=f126cdf8a6e0c7f379d618ff66beb2da,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 PASS NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:.,.
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3
20 1110696 rs6040355 A G,T 67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:35:4
20 1230237 . T . 47 PASS NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G GT:GQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

- ❑ We can upload multiple VCF single/multi sample.
- ❑ You do not need to create a multi-sample file with all the samples.
- ❑ BiERapp will merge all the samples from those files in the database.

Input: VCF multisample

- Create a Study
- Upload a new single-sample file
- BiERapp stores the sample in the created study
- Now we upload a new multisample file with 4 more samples
- BiERapp merges these samples in the study



Getting information: SIFT & PolyPhen

- **SIFT**

- **SIFT** predicts whether an amino acid substitution affects protein function
- **Interpretation:** 1 (tolerated) to 0 (deleterious)

<http://sift.jcvi.org/>

J. Craig Venter™
INSTITUTE

SIFT

- **PolyPhen**

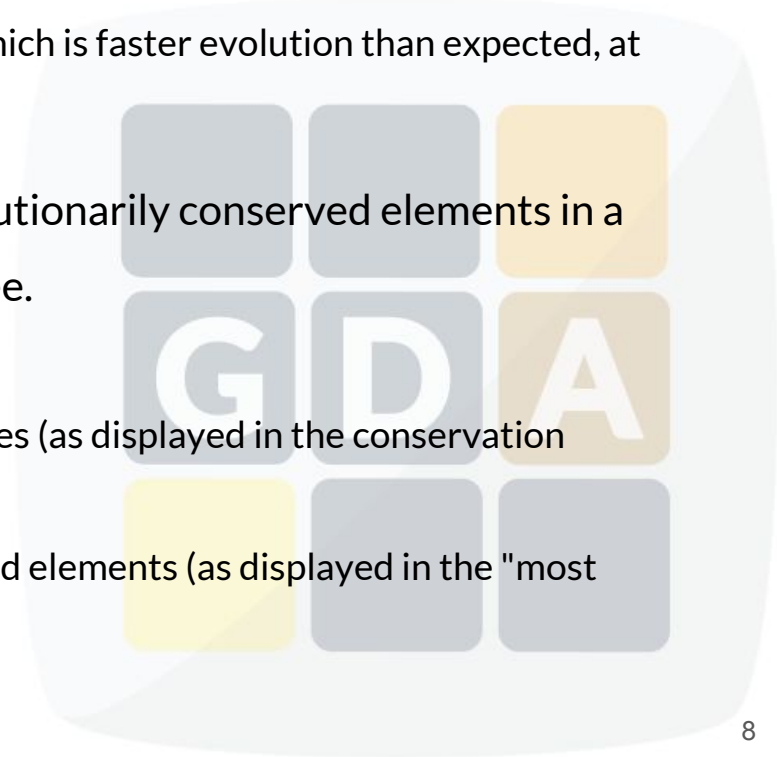
- Polymorphism Phenotyping is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein.
- **Interpretation:** 1 (probably damage) to 0 (benign)

<http://genetics.bwh.harvard.edu/pph2/index.shtml>

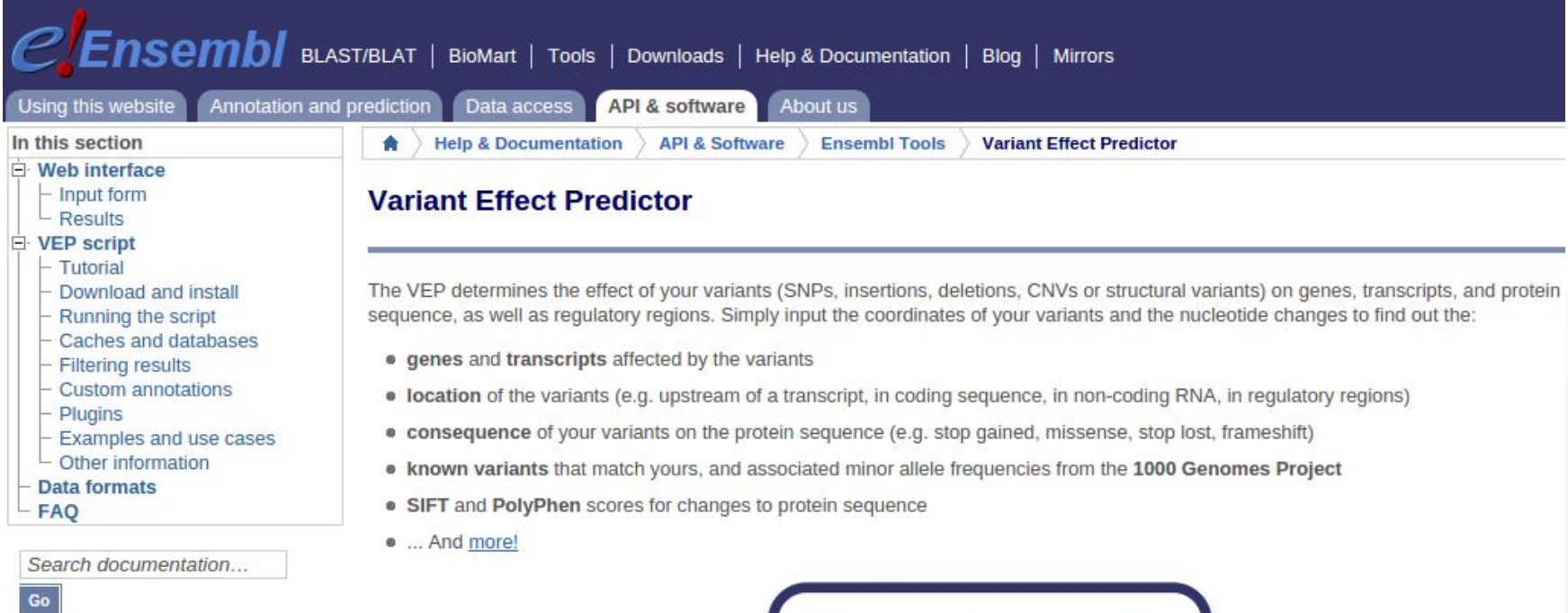


Getting information: Conservation

- **PhyloP**
 - PhyloP scores measure evolutionary conservation at individual alignment sites. The scores are interpreted as follows compared to the evolution expected under neutral drift:
 - Positive scores -- Measure conservation, which is slower evolution than expected, at sites that are predicted to be conserved.
 - Negative scores -- Measure acceleration, which is faster evolution than expected, at sites that are predicted to be fast-evolving.
- **PhastCons**
 - PhastCons is a program for identifying evolutionarily conserved elements in a multiple alignment, given a phylogenetic tree.
 - PhastCons essentially does three things:
 - It produces base-by-base conservation scores (as displayed in the conservation tracks in the UCSC browser)
 - It produces predictions of discrete conserved elements (as displayed in the "most conserved" tracks in the browser)
 - It estimates free parameters.



Getting information: Effect



The screenshot shows the Ensembl website navigation and the VEP documentation page. The top navigation bar includes links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. Below this is a secondary navigation bar with tabs for 'Using this website', 'Annotation and prediction', 'Data access', 'API & software', and 'About us'. The 'API & software' tab is active, and a breadcrumb trail shows the path: Home > Help & Documentation > API & Software > Ensembl Tools > Variant Effect Predictor.

In this section


- Web interface
 - Input form
 - Results
- VEP script
 - Tutorial
 - Download and install
 - Running the script
 - Caches and databases
 - Filtering results
 - Custom annotations
 - Plugins
 - Examples and use cases
 - Other information
- Data formats
- FAQ

Search documentation...
Go

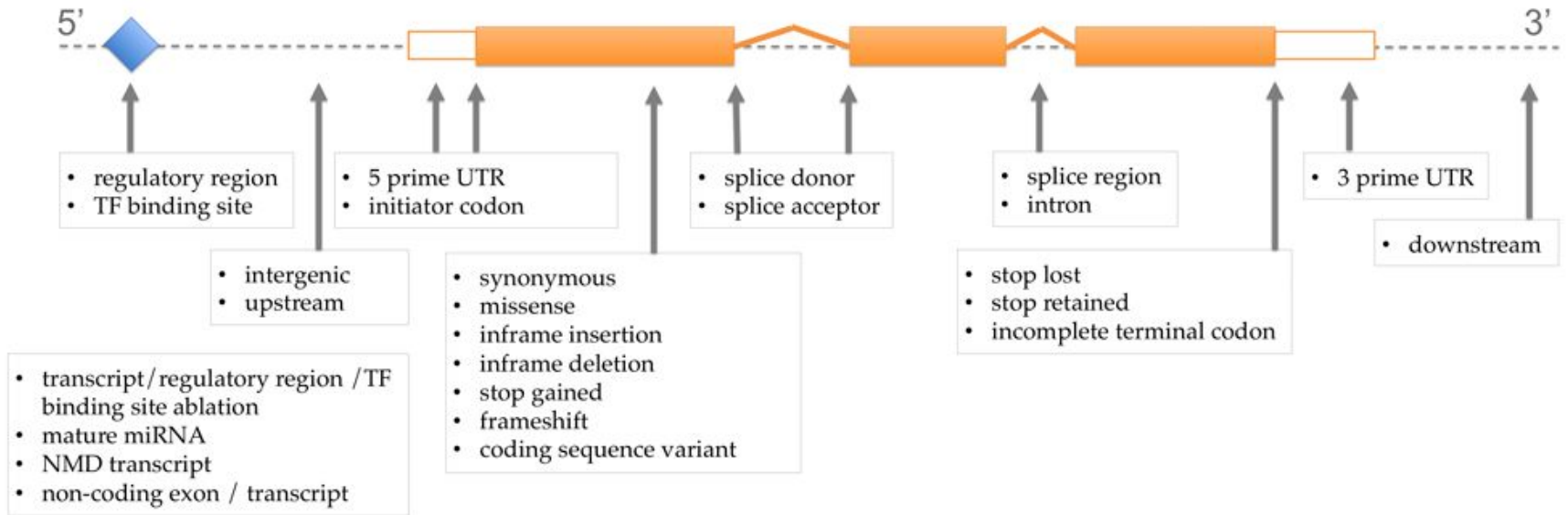
Variant Effect Predictor

The VEP determines the effect of your variants (SNPs, insertions, deletions, CNVs or structural variants) on genes, transcripts, and protein sequence, as well as regulatory regions. Simply input the coordinates of your variants and the nucleotide changes to find out the:

- **genes** and **transcripts** affected by the variants
- **location** of the variants (e.g. upstream of a transcript, in coding sequence, in non-coding RNA, in regulatory regions)
- **consequence** of your variants on the protein sequence (e.g. stop gained, missense, stop lost, frameshift)
- **known variants** that match yours, and associated minor allele frequencies from the **1000 Genomes Project**
- **SIFT** and **PolyPhen** scores for changes to protein sequence
- ... And [more!](#)

 Launch Ve!P

Getting information: Effect



http://www.ensembl.org/info/genome/variation/predicted_data.html

Getting information: Phenotype

ClinVar

ClinVar aggregates information about genomic variation and its relationship to human health.



GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Getting information: Pop. Frequencies

1000 Genomes

A Deep Catalog of Human Genetic Variation



NHLBI Exome Sequencing Project (ESP)
Exome Variant Server

Tool interface: Official release

<http://bierapp.babelomics.org/>

> Menu

BierApp 

sign in 

Overview

Welcome to the gene/variant prioritization tool of the BIER (the Team of Bioinformatics for Rare Diseases). This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees, case-controls or sporadic samples.

Try an Example

Here you can try all the filtering options and discover the gene affected in a test family.

Analyze your own families or case-control data

Here you can upload your VCF file containing the exomes to be analyzed. Define the thresholds of allele frequencies, pathogenicity, conservation; the type of variants sought; and define the type of inheritance and the segregation schema along the family.

Supported by

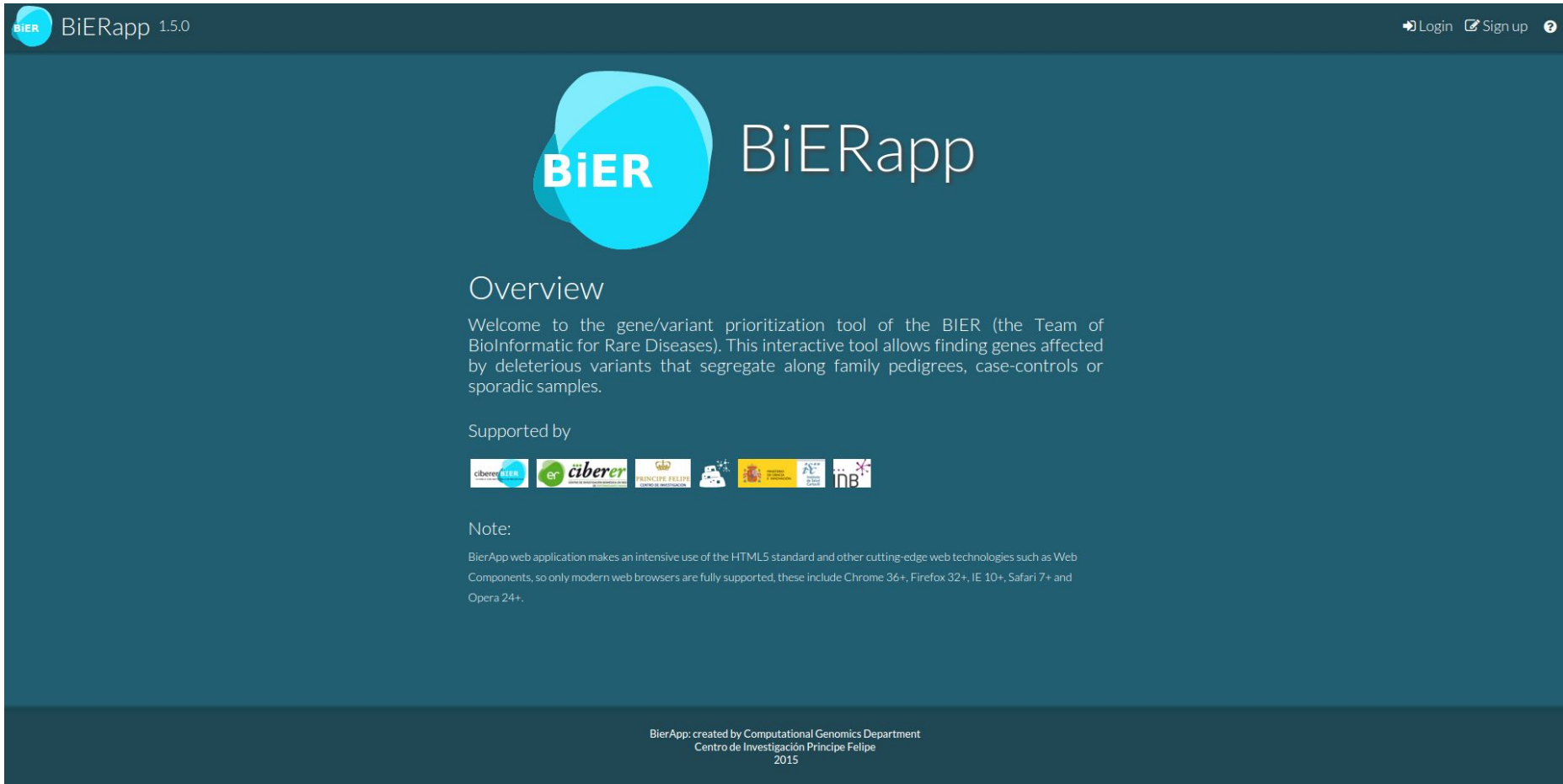


Note

This web application makes an intensive use of new web technologies and standards like HTML5, so browsers that are fully supported for this site are: Chrome 14+, Firefox 7+, Safari 5+ and Opera 11+. Older browser like Chrome13-, Firefox 5- or Internet Explorer 9 may rise some errors. Internet Explorer 6 and 7 are no supported at all.

Tool interface: Beta

<http://bierapp.babelomics.org/beta>



BiERapp 1.5.0


Login Sign up

BiERapp

Overview

Welcome to the gene/variant prioritization tool of the BIER (the Team of BioInformatic for Rare Diseases). This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees, case-controls or sporadic samples.

Supported by



Note:

BierApp web application makes an intensive use of the HTML5 standard and other cutting-edge web technologies such as Web Components, so only modern web browsers are fully supported, these include Chrome 36+, Firefox 32+, IE 10+, Safari 7+ and Opera 24+.

BierApp: created by Computational Genomics Department
Centro de Investigación Príncipe Felipe
2015

Tool interface: Sign up /Log in

BiERapp 1.5.0

BiER BiERapp

Overview

Welcome to the gene/variant prioritization tool of the BIER (the Team of BioInformatic for Rare Diseases). This interactive tool allows finding genes affected by deleterious variants that segregate along family pedigrees, case-controls or sporadic samples.

Supported by

Note:

BierApp web application makes an intensive use of the HTML5 standard and other cutting-edge web technologies such as Web Components, so only modern web browsers are fully supported, these include Chrome 36+, Firefox 32+, IE 10+, Safari 7+ and Opera 24+.

BiERapp: created by Computational Genomics Department
Centro de Investigación Príncipe Felipe
2015

Tool interface: Sign up /Log in

BiERapp 1.5.0

Create a new user



User name

E-mail

Create a password

Confirm your password

Sign up

[Login](#) [Sign up](#) [?](#)



Tool interface: Sign up /Log in

The screenshot displays the BiERapp 1.5.0 interface. The top left corner shows the BiER logo and the text "BiERapp 1.5.0". The main content area is a white box with a dark blue border, containing the following elements:

- Login** header with a right-pointing arrow icon.
- User** label above a text input field.
- Create new user...** link below the user input field.
- Password** label above a text input field.
- Remember password...** link below the password input field.
- Login** button below the password input field.
- Login as anonymous** link below the login button.
- All data will be deleted on logout** text below the anonymous login link.

In the top right corner of the interface, there is a navigation bar with three items: a right-pointing arrow icon followed by the text "Login", a checkmark icon followed by the text "Sign up", and a question mark icon. A red box highlights these three items, and a red arrow points upwards from below the box towards the "Sign up" text.

Tool interface: Project/Study

BiERapp 1.5.0 [My data](#) Study Browser

File Browser

/family > fam1

Upload

Name	Date	Status
HG01500_HG01501_m1_f2.vcf	Feb 26, 2016	✓ READY
analysis	Feb 26, 2016	--
data	Feb 26, 2016	--

view
to t
natic f
erious
sampl

Supported by

Tool interface: Project/Study

New Project

The screenshot shows the BiERapp 1.5.0 interface. At the top, there is a navigation bar with the BiER logo, the text 'BiERapp 1.5.0', and two menu items: 'My data' (with a cloud icon) and 'Study Browser' (with a group of people icon). Below this is a 'File Browser' window. The window title is 'File Browser' and it has standard window controls (pin, up, close). The current path is '/family > fam1'. Below the path, there are two buttons: a folder icon with a plus sign and a document icon with a plus sign. A red arrow points from the text 'New Project' to the folder icon with a plus sign. To the right of these buttons are 'Upload' (with a cloud icon) and a folder icon with a plus sign. Below the buttons is a table with columns 'Name', 'Date', and 'Status'. The table contains three rows of data:

Name	Date	Status
HG01500_HG01501_m1_f2.vcf	Feb 26, 2016	✓ READY
analysis	Feb 26, 2016	--
data	Feb 26, 2016	--

On the left side of the File Browser window, there is a tree view showing the following structure:

- Default project
 - Default study
- family
 - fam1
- family2
 - fam2
- family3
 - fam3

Tool interface: Project/Study

New Project

The screenshot displays the BiERapp 1.5.0 interface. At the top, there is a navigation bar with the BiER logo, the text 'BiERapp 1.5.0', and two menu items: 'My data' (with a cloud icon) and 'Study Browser' (with a group of people icon). Below this is a 'File Browser' window showing a directory structure. The current path is '/family > fam1'. The toolbar includes icons for adding files and folders, an 'Upload' button, and a grid view icon. A 'Create Project' dialog box is open in the foreground, containing a 'Project name:' label, an input field with the text 'New project', and a 'Create' button. A red arrow points from the text 'New Project' to the '+' icon in the file browser toolbar.

Name	Date	Status
Default project	26, 2016	✓ READY
Default study	26, 2016	--
family	26, 2016	--
fam1		
family2		
fam2		
family3		
fam3		

Tool interface: Project/Study

BiERapp 1.5.0 My data Study Browser

File Browser

/family ▶ fam1

Upload +

Default project
Default study
family
fam1
family2
fam2
family3
fam3

Name	Date	Status
HG01500_HG01501_m1_f2.vcf	Feb 26, 2016	✓ READY
analysis	Feb 26, 2016	--
data	Feb 26, 2016	--

New Study

Supported by

Tool interface: Project/Study

New Study

The screenshot displays the BiERap tool interface. A central 'Create Study' dialog box is open, showing a list of projects: Default project, .meta, family (selected), family2, and family3. Below the project list, the 'Study type' section includes radio buttons for Family (selected), Case / Control, Trio, Paired, Case set, and Control set. The 'Study name' field contains 'New study'. A 'Create' button is at the bottom of the dialog. In the background, a 'File Browser' window shows a tree view with folders for 'family', 'family2', and 'family3'. A red arrow points from the 'New Study' text to the '+ New Study' button in the File Browser. To the right, a 'rowser' window shows a table with columns 'Date' and 'Status'.

Date	Status
Feb 26, 2016	✓ READY
Feb 26, 2016	--
Feb 26, 2016	--

Tool interface: Project/Study

Select the parent Project

Select the Study type

BiERap

File Browser

/family ▶ fam1

- Default project
 - Default s
- family
 - fam1
- family2
 - fam2
- family3
 - fam3

Projects:

- Default project
- .meta
- family
- family2
- family3

Study type

- Family
- Case / Control
- Trio
- Paired
- Case set
- Control set

Study name:

New study

Create

Date	Status
Feb 26, 2016	✓ READY
Feb 26, 2016	--
Feb 26, 2016	--

view

to t
natic f
erious
sampl

Supported by

Tool interface: Study Browser

Select Study Browser

BiERapp 1.5.0 My data Study Browser

Study Browser

/ family ▶ fam1

Upload Search by name...

Name	Status	Source
HG01500	✓ READY	HG01500_HG0...
HG01501	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...

Study Type: FAMILY Variant Browser

Tool interface: Study Browser

Choose your study

Select Study Browser

BiERapp 1.5.0 My data Study Browser

Study Browser

/ family ▶ fam1

Upload Search by name...

Name	Status	Source
HG01500	✓ READY	HG01500_HG0...
HG01501	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...

Study Type: FAMILY Variant Browser

Tool interface: Study Browser

Choose your study

Select Study Browser

BiERapp 1.5.0 My data Study Browser

Study Browser

/ family ▶ fam1

Upload Search by name...

Name	Status	Source
HG01500	✓ READY	HG01500_HG0...
HG01501	✓ READY	HG01501_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...

Study Type: FAMILY Variant Browser

Select your samples

Tool interface: Study Browser

Choose your study

Select Study Browser

BiERapp 1.5.0 My data Study Browser

Study Browser

/ family ▶ fam1

Upload Search by name...

Name	Status	Source
HG01500	✓ READY	HG01500_HG0...
HG01501	✓ READY	HG01501_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...
HG01500_HG0...	✓ READY	HG01500_HG0...

Study Type: FAMILY Variant Browser

Select your samples

Run the Variant Browser

Tool interface: Variant Browser

Study: fam1

Clear Search

Segregation

HG01501 0/0 0/1 1/1 ./.
 HG01500_HG0...
 HG01500_HG0...

Position

Chromosomal location:
 1:1-1000000,2:1-1000000

Import from BED:

Gene:
 BRCA2, PPL

SNPId:
 rs9988179, rs140361978

Type +

Population Freqs. +

Protein Substitution Scores +

Conservation +

Consequence Type +

Variant	Alleles	SNP Id	Genes	Type	Samples			Conseq. Type	SIFT	Polyphen	phyloP	PhastCons	1000G MAF(phase 1)					1000G MAF(G)		
					HG01501	HG01500_HG...	HG01500_HG...						ALL	AMR	ASN	AFR	EUR	ALL	AMR	SAS
1:14464	A>T	.	DDX11L1.WASH7P	SNV	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.121	0.056
1:14599	T>A	.	DDX11L1.WASH7P	SNV	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.184	0.030
1:14604	A>G	.	DDX11L1.WASH7P	SNV	0 0	0 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.024
1:14930	A>G	rs6682385	DDX11L1.WASH7P	SNP	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_variant,int...	.	.	0.184	0.022
1:15820	G>T	rs75570658	DDX11L1.WASH7P	SNP	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.166
1:15904	T>C	.	DDX11L1.WASH7P	INDEL	0 1	0 1	0 0	downstream_gene_variant,splice_region_variant,non_coding...	.	.	0.184	0.061
1:63268	T>C	rs75478250	OR4G11P	SNP	0 1	1 0	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.163	0.088	0.353	0.452	0.365
1:63671	G>A	rs116440577	OR4G11P	SNP	1 0	0 1	0 1	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.015	0.170	0.220	0.050	0.350	0.110	0.188	0.183	0.190
1:63736	CTA>	.	OR4G11P	INDEL	0 1	0 1	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.839
1:69428	T>G	rs140739101	OR4F5	SNP	0 0	1 0	1 0	missense_variant,regulatory_region_variant	.	.	0.327	0.660	0.019	0.036	0.015

Page 1 of 11058

1 - 10 of 110572

Genomic context Pop. Frequencies Phenotype Effect

Region overview Window size: 1287 nts

1003965 1004608 1005251

Detailed information Window size: 161 nts

1004528 1004608 1004688

Sequence

GCCTGCTGCGGACCTTCCCCCATCCGTAGCCCCCTCCCCAAGCGCGCTTTCGGCAGGGCTGGACTGGCAGCCGACGTCGCCGAAGACGATCCAGCTGGTTTGGCGGACGACATGGGGAATGGTGCAGACGGAGGCTCCGTACAAGTTCAGTACC1

Gene

Tool interface: Variant Browser

Study: fam1

Clear Search

Segregation

0/0 0/1 1/1 ./.

HG01501

HG01500_HGO...

HG01500_HGO...

Position

Chromosomal location:
1:1-1000000,2:1-1000000

Import from BED:

Gene:
BRCA2, PPL

SNPId:
rs9988179, rs140361978

Type +

Population Freqs. +

Protein Substitution Scores +

Conservation +

Consequence Type +

Variant	Alleles	SNP id	Genes	Type	Samples			Conseq. Type	SIFT	Polyphen	phyloP	PhastCons	1000G MAF(phase 1)					1000G MAF(G)		
					HG01501	HG01500_HG...	HG01500_HG...						ALL	AMR	ASN	AFR	EUR	ALL	AMR	SAS
1:14464	A>T	.	DDX11L1.WASH7P	SNV	1/0	0/0	0/0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.121	0.056
1:14599	T>A	.	DDX11L1.WASH7P	SNV	0/0	1/0	1/0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.184	0.030
1:14604	A>G	.	DDX11L1.WASH7P	SNV	0/0	0/0	1/0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.024
1:14930	A>G	rs6682385	DDX11L1.WASH7P	SNP	0/0	1/0	1/0	downstream_gene_variant,non_coding_transcript_variant,int...	.	.	0.184	0.022
1:15820	G>T	rs75570658	DDX11L1.WASH7P	SNP	1/0	0/0	0/0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.166
1:15904	T>C	.	DDX11L1.WASH7P	INDEL	0/1	0/1	0/0	downstream_gene_variant,splice_region_variant,non_coding...	.	.	0.184	0.061
1:63268	T>C	rs75478250	OR4G11P	SNP	0/1	1/0	1/0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.163	0.088	0.353	0.452	0.365
1:63671	G>A	rs116440577	OR4G11P	SNP	1/0	0/1	0/1	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.015	0.170	0.220	0.050	0.350	0.110	0.188	0.183	0.190
1:63736	CTA>	.	OR4G11P	INDEL	0/1	0/1	1/0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.839
1:69428	T>G	rs140739101	OR4F5	SNP	0/0	1/0	1/0	missense_variant,regulatory_region_variant	.	.	0.327	0.660	0.019	0.036	0.015

Page 1 of 11058

1 - 10 of 110572

Genomic context Pop. Frequencies Phenotype Effect

Region overview Window size: 1287 nts

Detailed information Window size: 161 nts

Sequence
CCCTGCCCTGCCGGCACCTTCCCCCATCCGTAGCCCCCTCCCCAAGCGCGCTTTCGGCAGGGCTGGACTGGCAGACCGACGTCGCCAAGACGATCCAGCTGGTTTGGCGGCAGCAC TGGGAATGGTGCAGACGGAGGCTCCGTACAAGTTCAGTACC1

Gene

Filters

Tool interface: Variant Browser

Variants

Study: fam1

Clear Search

Segregation

HG01501 0/0 0/1 1/1 ./.
 HG01500_HG0...
 HG01500_HG0...

Position

Chromosomal location:
 1:1-1000000,2:1-1000000

Import from BED:

Gene:
 BRCA2, PPL

SNPId:
 rs9988179, rs140361978

Type +

Population Freqs. +

Protein Substitution Scores +

Conservation +

Consequence Type +

Variant	Alleles	SNP id	Genes	Type	Samples			Conseq. Type	SIFT	Polyphen	phyloP	PhastCons	1000G MAF(phase 1)					1000G MAF(
					HG01501	HG01500_HG...	HG01500_HG...						ALL	AMR	ASN	AFR	EUR	ALL	AMR	SAS
1:14464	A>T	.	DDX11L1.WASH7P	SNV	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.121	0.056
1:14599	T>A	.	DDX11L1.WASH7P	SNV	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.184	0.030
1:14604	A>G	.	DDX11L1.WASH7P	SNV	0 0	0 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.024
1:14930	A>G	rs6682385	DDX11L1.WASH7P	SNP	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_variant,int...	.	.	0.184	0.022
1:15820	G>T	rs75570658	DDX11L1.WASH7P	SNP	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.166
1:15904	>C	.	DDX11L1.WASH7P	INDEL	0 1	0 1	0 0	downstream_gene_variant,splice_region_variant,non_coding...	.	.	0.184	0.061
1:63268	T>C	rs75478250	OR4G11P	SNP	0 1	1 0	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.163	0.088	0.353	0.452	0.365
1:63671	G>A	rs116440577	OR4G11P	SNP	1 0	0 1	0 1	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.015	0.170	0.220	0.050	0.350	0.110	0.188	0.183	0.190
1:63736	CTA>	.	OR4G11P	INDEL	0 1	0 1	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.839
1:69428	T>G	rs140739101	OR4F5	SNP	0 0	1 0	1 0	missense_variant,regulatory_region_variant	.	.	0.327	0.660	0.019	0.036	0.015

Genomic context Pop. Frequencies Phenotype Effect

Region overview Window size: 1287 nts

Detailed information Window size: 161 nts

Sequence
 G C C C C T G C C T G C C G G C A C C T T C C C C C A T C C G T A G C C C C T C C C C A A G C G C G C T T G C C G C A G G G C T G G A C T G C G A C A C C G A C G T C C C G A A G A C G A T C C A G C T G G T T T G G C G G C A G C A C T G G G G A A T G G T G C A G A C G G A G G C T C C G T A C A A G T T C A C G T A C C 1

Gene

Tool interface: Variant Browser

Study: fam1

Clear Search

Segregation

HG01501 0/0 0/1 1/1 ./.
 HG01500_HG0...
 HG01500_HG0...

Position

Chromosomal location:
 1:1-1000000,2:1-1000000

Import from BED:

Gene:
 BRCA2, PPL

SNPId:
 rs9988179, rs140361978

Type +

Population Freqs. +

Protein Substitution Scores +

Conservation +

Consequence Type +

Variant	Alleles	SNP id	Genes	Type	Samples			Conseq. Type	SIFT	Polyphen	phyloP	PhastCons	100G MAF(phase 1)					100G MAF(G)		
					HG01501	HG01500_HG...	HG01500_HG...						ALL	AMR	ASN	AFR	EUR	ALL	AMR	SAS
1:14464	A>T	.	DDX11L1.WASH7P	SNV	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.121	0.056
1:14599	T>A	.	DDX11L1.WASH7P	SNV	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.184	0.030
1:14604	A>G	.	DDX11L1.WASH7P	SNV	0 0	0 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.024
1:14930	A>G	rs6682385	DDX11L1.WASH7P	SNP	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_variant,int...	.	.	0.184	0.022
1:15820	G>T	rs75570658	DDX11L1.WASH7P	SNP	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.166
1:15904	T>C	.	DDX11L1.WASH7P	INDEL	0 1	0 1	0 0	downstream_gene_variant,splice_region_variant,non_coding...	.	.	0.184	0.061
1:63268	T>C	rs75478250	OR4G11P	SNP	0 1	1 0	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.163	0.088	0.353	0.452	0.365
1:63671	G>A	rs116440577	OR4G11P	SNP	1 0	0 1	0 1	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.015	0.170	0.220	0.050	0.350	0.110	0.188	0.183	0.190
1:63736	CTA>	.	OR4G11P	INDEL	0 1	0 1	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.839
1:69428	T>G	rs140739101	OR4F5	SNP	0 0	1 0	1 0	missense_variant,regulatory_region_variant	.	.	0.327	0.660	0.019	0.036	0.015

Page 1 of 11058

1 - 10 of 110572

Genomic context Pop. Frequencies Phenotype Effect

Region overview Window size: 1287 nts

1004528 1004688 100525

Detailed information Window size: 161 nts

1004528 1004688 1004688

Sequence - v ^

CCCTGCCCTGCCGGCACCTTCCCCCATCCGTAGCCCCCTCCCCAAGCGCGCTTTCGGCAGGGCTGGACTGCGACACCGACGTCGCCAAGACGATCCAGCTGGTTTGGCGGCAGCAC TGGGGAATGGTGCAGACGGAGGCTCCGTACAAGTTCACGTACC1

Gene - v ^

More info

Tool interface: Filters

Segregation

	0/0	0/1	1/1	./.
HG01501	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HG01500_HG0...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HG01500_HG0...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Choose your genotypes.

Position

Chromosomal location:
1:1-1000000, 2:1-1000000

Import from BED:

Gene:
BRCA2, PPL

SNPId:
rs9988179, rs140361978

You can filter by region, Chromosome, Gene and SNPId. You can also import regions from a BED file

Type

SNV MNV
 INDEL SV
 CNV



Tool interface: Filters

Population Freqs.

▼ 1000 Genomes population phase 1

All populations MAF [ALL]

< ▾

American MAF [AMR]

< ▾

Asian MAF [ASN]

< ▾

African MAF [AFR]

< ▾

European MAF [EUR]

< ▾

▼ 1000 Genomes population phase 3

All populations MAF [ALL]

< ▾

American MAF [AMR]

< ▾

South Asian MAF [SAS]

< ▾

East Asian MAF [EAS]

< ▾

African MAF [AFR]

< ▾

European MAF [EUR]

< ▾

▼ ESP 6500

European american MAF

< ▾

African american MAF

< ▾

Protein Substitution Scores

SIFT

< ▾

Polyphen

< ▾

Conservation

PhyloP

< ▾

PhastCons

< ▾

Consequence Type

transcript ablation

splice donor variant

splice acceptor variant

stop gained

frameshift variant

stop lost

initiator codon variant

inframe insertion

inframe deletion

missense variant

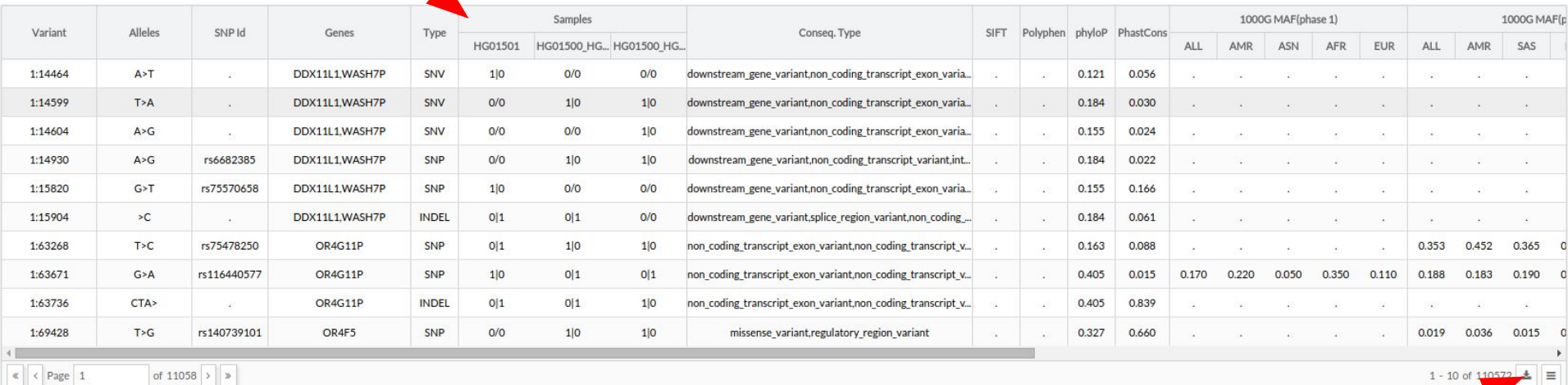
transcript amplification

Filter by MAF

- 1000G phase1
- 1000G phase3
- ESP 6500

Tool interface: Variant grid

Resizable columns



Variant	Alleles	SNP Id	Genes	Type	Samples			Conseq. Type	SIFT	Polyphen	phyloP	PhastCons	1000G MAF(phase 1)					1000G MAF(p...				
					HG01501	HG01500_HG...	HG01500_HG...						ALL	AMR	ASN	AFR	EUR	ALL	AMR	SAS		
1:14464	A>T	.	DDX11L1.WASH7P	SNV	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.121	0.056	
1:14599	T>A	.	DDX11L1.WASH7P	SNV	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.184	0.030	
1:14604	A>G	.	DDX11L1.WASH7P	SNV	0 0	0 0	1 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.024	
1:14930	A>G	rs6682385	DDX11L1.WASH7P	SNP	0 0	1 0	1 0	downstream_gene_variant,non_coding_transcript_variant,int...	.	.	0.184	0.022	
1:15820	G>T	rs75570658	DDX11L1.WASH7P	SNP	1 0	0 0	0 0	downstream_gene_variant,non_coding_transcript_exon_varia...	.	.	0.155	0.166	
1:15904	>C	.	DDX11L1.WASH7P	INDEL	0 1	0 1	0 0	downstream_gene_variant,splice_region_variant,non_coding...	.	.	0.184	0.061	
1:63268	T>C	rs75478250	OR4G11P	SNP	0 1	1 0	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.163	0.088	0.353	0.452	0.365	0
1:63671	G>A	rs116440577	OR4G11P	SNP	1 0	0 1	0 1	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.015	0.170	0.220	0.050	0.350	0.110	0.188	0.183	0.190	0	
1:63736	CTA>	.	OR4G11P	INDEL	0 1	0 1	1 0	non_coding_transcript_exon_variant,non_coding_transcript_v...	.	.	0.405	0.839	0
1:69428	T>G	rs140739101	OR4F5	SNP	0 0	1 0	1 0	missense_variant,regulatory_region_variant	.	.	0.327	0.660	0.019	0.036	0.015	0

Pagination system

Export to CSV

Column Selector

Tool interface: Genomic Context

Genomic context Pop. Frequencies Phenotype Effect

Region overview Window size: 1287 nts

63028 63671 64314

gene

Detailed information Window size: 161 nts

63591 63671 63751

Sequence - v v ^

AATTGATCATACATTTTTATTCGGTCACTGTCCAACGACATTCCTCAAATGATTTTATCCAAAGCATTCTTCACCTCGCGCTCACATCACCGTAGTGGTTTTGTTTTTGTCCATGCATGTTTCTACGTGTGGCCTTCCCTACTAAGTCATTGG#

Gene - v v ^

SNP - v v ^

rs202004563 rs80011619 rs116440577 rs111440589 rs201582574 rs112317269 rs201888535 rs77426996 rs61158452

Powered by Genome Maps 63671

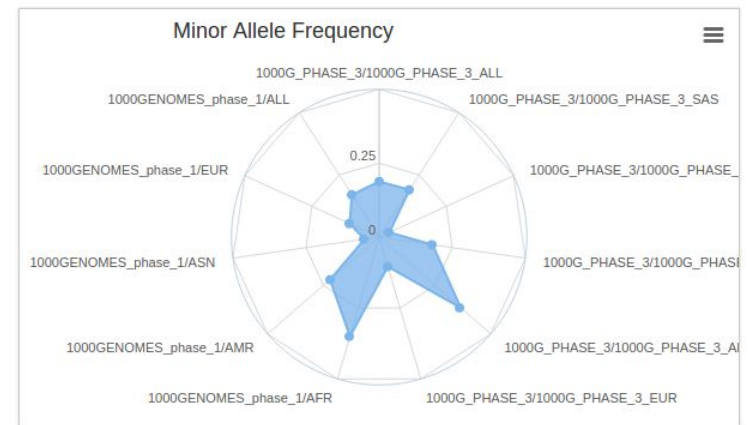
The screenshot displays the Genomic Context tool interface. At the top, there are tabs for 'Genomic context', 'Pop. Frequencies', 'Phenotype', and 'Effect'. Below the tabs is a control bar with a zoom slider, a window size input field set to '161', a 'Go!' button, and navigation arrows. The main interface is divided into three sections: 'Region overview', 'Detailed information', and 'SNP'. The 'Region overview' section shows a horizontal bar representing a genomic region from 63028 to 64314, with a window size of 1287 nts. A vertical orange line marks the position 63671. The 'Detailed information' section shows a window size of 161 nts, with a sequence viewer displaying the DNA sequence: AATTGATCATACATTTTTATTCGGTCACTGTCCAACGACATTCCTCAAATGATTTTATCCAAAGCATTCTTCACCTCGCGCTCACATCACCGTAGTGGTTTTGTTTTTGTCCATGCATGTTTCTACGTGTGGCCTTCCCTACTAAGTCATTGG#. Below the sequence is a 'Gene' section with a dropdown menu. The 'SNP' section shows a list of SNPs with their IDs: rs202004563, rs80011619, rs116440577, rs111440589, rs201582574, rs112317269, rs201888535, rs77426996, and rs61158452. The interface is powered by Genome Maps, as indicated at the bottom left, and the current position 63671 is shown at the bottom right.

Tool interface: Pop. Frequencies

Genomic context Pop. Frequencies Phenotype Effect

Study	Population	SuperPopulation	Ref. Allele	Alt. Allele	Ref. Allele Fr...	Alt. Allele Fr...	MAF	0/0	0/1	
1000G_PHASE_3	1000G_PHASE_3_ALL	1000G_PHASE_3_ALL	G	A	0.813	0.188	0.188	0	0	0
1000G_PHASE_3	1000G_PHASE_3_SAS	1000G_PHASE_3_SAS	G	A	0.810	0.190	0.190	0	0	0
1000G_PHASE_3	1000G_PHASE_3_EAS	1000G_PHASE_3_EAS	G	A	0.962	0.038	0.038	0	0	0
1000G_PHASE_3	1000G_PHASE_3_AMR	1000G_PHASE_3_AMR	G	A	0.817	0.183	0.183	0	0	0
1000G_PHASE_3	1000G_PHASE_3_AFR	1000G_PHASE_3_AFR	G	A	0.635	0.365	0.365	0	0	0
1000G_PHASE_3	1000G_PHASE_3_EUR	1000G_PHASE_3_EUR	G	A	0.896	0.104	0.104	0	0	0
1000GENOMES_phase_1	AFR	AFR	G	A	0.650	0.350	0.350	0	0	0
1000GENOMES_phase_1	AMR	AMR	G	A	0.780	0.220	0.220	0	0	0
1000GENOMES_phase_1	ASN	ASN	G	A	0.950	0.050	0.050	0	0	0
1000GENOMES_phase_1	EUR	EUR	G	A	0.890	0.110	0.110	0	0	0
1000GENOMES_phase_1	ALL	ALL	G	A	0.830	0.170	0.170	0	0	0

Total: 11



Tool interface: Phenotype & Effect

Genomic context Pop. Frequencies Phenotype Effect

Cosmic:

Gene name	Histology subtype	Mutation ID	Mutation somatic status	Primary histology	Primary site	Sample source	Site subtype	Tumour orig
AGRN	adenocarcinoma	1126908	Confirmed somatic variant	carcinoma	prostate	fresh/frozen - NOS	NS	primary
AGRN	adenocarcinoma	1126908	Confirmed somatic variant	carcinoma	large_intestine	NS	colon	NS
AGRN	neoplasm	1126908	Confirmed somatic variant	other	thyroid	NS	NS	NS
AGRN	neoplasm	1126908	Confirmed somatic variant	other	thyroid	NS	NS	NS
AGRN	adenocarcinoma	1126908	Confirmed somatic variant	carcinoma	large_intestine	NS	colon	NS

Total: 5

GWAS:

No results found.

Clinvar:

Accession	Clinical significance	Gene name	Review status	Traits
RCV000116259	Benign	AGRN	CLASSIFIED_BY_SINGLE_SUBMITTER	not specified.AllHighlyPenetrant.Not Specified

Total: 1

Genomic context Pop. Frequencies Phenotype Effect

Gene Name	Ensembl Gene Id	Ensembl Transcript Id	Conseq. type	Relative Position	Codon	Strand	Biotype	cDna Position	cds Position	AA Position	AA Change	Sift	Polyphen
AGRN	ENSG00000188157	ENST00000379370	synonymous_variant		tCA/tcG	+	protein_coding	3116	3066	1022	SER/SER		
AGRN	ENSG00000188157	ENST00000479707	2KB_downstream_gene_variant			+	retained_intron						
AGRN	ENSG00000188157	ENST00000466223	2KB_upstream_gene_variant			+	retained_intron						
AGRN	ENSG00000188157	ENST00000478677	2KB_upstream_gene_variant			+	retained_intron						
AGRN	ENSG00000188157	ENST00000492947	2KB_upstream_gene_variant			+	retained_intron						
AGRN	ENSG00000188157	ENST00000419249	upstream_gene_variant			+	protein_coding						
			regulatory_region_variant										

Total: 7

Results

Segregation

	0/0	0/1	1/1	./.
HG01501	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HG01500_HG0...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HG01500_HG0...	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Position

Chromosomal location:

1:1-1000000,2:1-1000000

Import from BED:

Gene:

NBPF1

SNPId:

rs9988179,rs140361978

Population Freqs.

1000 Genomes population phase 1

All populations MAF [ALL]

< 0.001

American MAF [AMR]

<

Asian MAF [ASN]

<

African MAF [AFR]

<

European MAF [EUR]

<

1000 Genomes population phase 3

ESP 6500

Type

- SNV MNV
 INDEL SV
 CNV

Click on "Search" and view the results

Results

Study: fam1

Clear Search

Segregation

HG01501 0/0 0/1 1/1 ./.

HG01500_HG0...

HG01500_HG0...

Position

Type

SNV MNV

INDEL SV

CNV

Population Freqs.

1000 Genomes population phase 1

All populations MAF [ALL]

American MAF [AMR]

Asian MAF [ASN]

African MAF [AFR]

European MAF [EUR]

1000 Genomes population phase 3

ESP 6500

Variant	Alleles	SNP Id	Genes	Type	Samples			Conseq. Type	SIFT	Polyphen	phyloP	PhastCons	1000G MAF(phase 1)					1000G MAF(p...		
					HG01501	HG01500_HG...	HG01500_HG...						ALL	AMR	ASN	AFR	EUR	ALL	AMR	SAS
1:16902894	A>G	rs28453011	NBPF1	SNP	1 0	0 1	1 1	missense_variant,3_prime_UTR_variant,NMD_transcript_vari...	.	.	0.000	0.000	0.318	0.354	0.415	0
1:16916603	C>A	rs11583443	NBPF1	SNP	1 0	0 1	1 1	intron_variant,NMD_transcript_variant,downstream_gene_v...	.	.	0.164	0.018	0.228	0.291	0.324	0

Page 1 of 1

1 - 2 of 2

Genomic context Pop. Frequencies Phenotype Effect

Region overview Window size: 1287 nts

16902251 16902894 16903537

Detailed information Window size: 161 nts

16902814 16902894 16902974

Sequence

Gene

Who is using BierApp?

ciberer *isciii*

Centro de Investigación Biomédica en Red
Enfermedades Raras

European Variation Archive

Home Submit Data Study Browser Variant Browser Clinical Browser GA4GH API FAQ Contact

Filter

Reset Submit

Genome Assembly

Organism / Assembly: Human / GRCH37

Position

Filter By: Chromosomal Location

1:3000000-3100000

Consequence Type

Minor Allele Frequency

Protein Substitution Score

Studies Mapped To Assembly

search

Name ↑

1000 Genomes Phase 1 Analysis

Variant Browser

Page 1 of 547 Variants 1 - 10 of 5462

Chr	Position	Variant ID	Alleles	Class	Most Severe Consequence Type	Most Severe Protein Substitution Score		View
						PolyPhen2	Sift	
1	3000017	rs5578667	dbSNP ID(Human), TransPlant ID(Plant) and Submitted ID(others)			-	-	dbSNP
1	3000087	rs189829684	G/A	SNV	intron_variant	-	-	dbSNP
1	3000090	rs570767133	G/A	SNV	intron_variant	-	-	dbSNP
1	3000100	rs539360517	A/G	SNV	intron_variant	-	-	dbSNP
1	3000124	rs7605569	A/G	SNV	intron_variant	-	-	dbSNP
1	3000126	rs575845718	G/A	SNV	intron_variant	-	-	dbSNP
1	3000146	rs115632128	C/T	SNV	intron_variant	-	-	dbSNP
1	3000150	rs555464840	G/A	SNV	intron_variant	-	-	dbSNP
1	3000156	rs34722621	TG/-	INDEL	intron_variant	-	-	dbSNP
1	3000157	rs569678638	-/CC	INDEL	intron_variant	-	-	dbSNP

Results per Page: 10 Export as CSV

Variant Data

Annotation Files Genotypes Population Statistics

IT4Innovations
national!\$11€0
supercomputing
center1001\$1\$0



The EVA has its own customized version of BiERApp.

Conclusions

- ❑ The proposed web-based interactive framework has **great potential to detect disease-related variants** in familial diseases as demonstrated by its successful use in several studies.
- ❑ **The use of the filters is interactive** and the results are almost instantaneously displayed in a panel that includes the genes affected, the variants and specific information for them.
- ❑ Candidate variants are **new knowledge useful for future diagnostic.**



More info: publication

Nucleic Acids Research Advance Access published May 6, 2014

Nucleic Acids Research, 2014 **1**
doi: 10.1093/nar/gku407

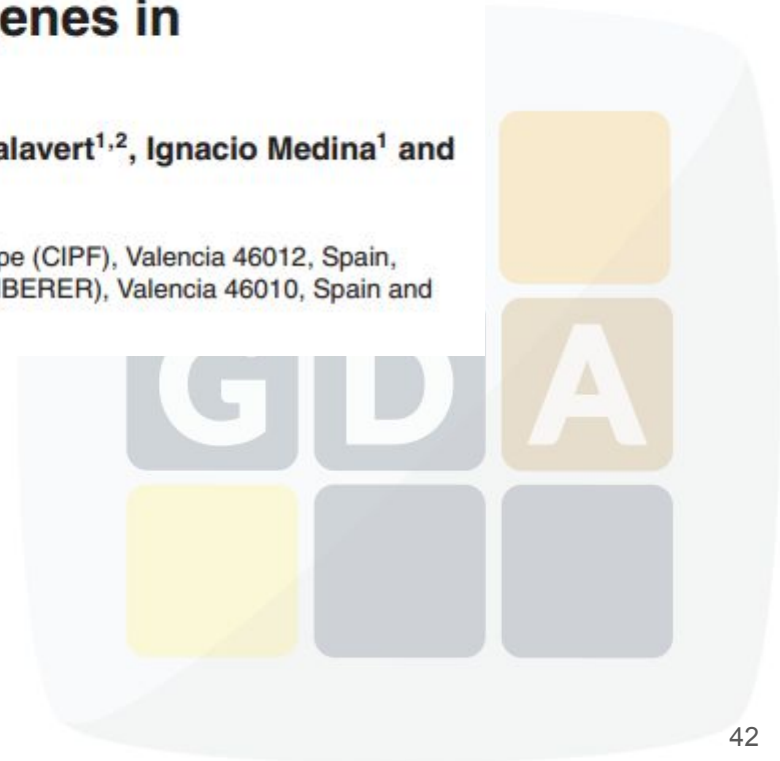
A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies

Alejandro Alemán^{1,2}, Francisco Garcia-Garcia¹, Francisco Salavert^{1,2}, Ignacio Medina¹ and Joaquín Dopazo^{1,2,3,*}

¹Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia 46012, Spain,

²Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia 46010, Spain and

³Functional Genomics Node, (INB) at CIPF, Valencia 46012, Spain



More info: new features

- ❑ Case-Control studies.
- ❑ Export results to VCF.
- ❑ Improve the database to allow larger datasets.
- ❑ Add new filters: HPO, GO, tissues, pathways, ...



More info: BiERApp behind the scenes

