

Booklet for the PiER

# Contents

<b>1</b>	<b>Background</b>	<b>2</b>
<b>2</b>	<b>Facilities</b>	<b>4</b>
<b>3</b>	<b>Compatibility</b>	<b>6</b>
<b>4</b>	<b>Runtime</b>	<b>7</b>
<b>5</b>	<b>Frontpage</b>	<b>8</b>
<b>6</b>	<b>Development</b>	<b>10</b>
<b>7</b>	<b>Help buttons</b>	<b>11</b>
<b>8</b>	<b>Error messages</b>	<b>12</b>
<b>9</b>	<b>eV2CG</b>	<b>13</b>
9.1	Interface . . . . .	13
9.2	Linking results . . . . .	15
<b>10</b>	<b>eCG2PG</b>	<b>17</b>
10.1	Interface . . . . .	17
10.2	Networking results . . . . .	17
<b>11</b>	<b>eCrosstalk</b>	<b>20</b>
11.1	Interface . . . . .	20
11.2	Crosstalk results . . . . .	20
<b>12</b>	<b>cTGene</b>	<b>22</b>
12.1	Interface . . . . .	22
12.2	Prioritisation results . . . . .	23
<b>13</b>	<b>cTCrosstalk</b>	<b>26</b>
13.1	Interface . . . . .	26
13.2	Prioritisation results . . . . .	28

# Chapter 1

## Background



Figure 1.1: The logo for the PiER. The above-water pillar structure in red (symbolising the infrastructure) and water waves in blue (by analogy the piano stave) collectively illustrate the web-based PiER facilities enabling ab initio and real-time genetic target prioritisation.

### Motivation

The field of target discovery has been advanced by genetics-led target prioritisation approaches. Integrative prioritisation for early-stage genetic target discovery has proven cost-effective in promoting the translational use of disease genetic associations, which is increasingly recognised in reducing drug attrition rate in late-stage clinical trials.

### Design

Building on the verified Pi approach (see Nature Genetics 2019), here I introduce web-based servers/facilities called PiER. The PiER is free and open to all users and there is no login requirement, allowing the users to perform ab initio and real-time target prioritisation harnessing human disease genetics, functional genomics and protein interactions.

By analogy to the piano stave, the PiER consists of five horizontal lines, with three lines representing the elementary facility (**eV2CG**, **eCG2PG** and **eCrosstalk**), each doing specific tasks on their own, and the rest two lines signifying the combinatory facility (**cTGene** and **cTCrosstalk**).

- **eV2CG**, linking variants to core genes; see Example Output
- **eCG2PG**, networking core genes to peripheral genes; see Example Output
- **eCrosstalk**, identifying the crosstalk between pathways; see Example Output
- **cTGene**, prioritising targets at the gene level; see Example Output
- **cTCrosstalk**, prioritising targets at the crosstalk level; see Example Output

# Chapter 2

## Facilities

The elementary facility supports three specific tasks, including three online tools: (i) **eV2CG**, utilising functional genomics to link disease-associated variants (including those located at the non-coding genome) to core genes likely responsible for genetic associations; (ii) **eCG2PG**, using knowledge of protein interactions to ‘network’ core genes with each other and with additional peripheral genes as well, producing a ranked list of core and peripheral genes; and (iii) **eCrosstalk**, exploiting the information of pathway-derived interactions to identify highly ranked genes that mediate the crosstalk between molecular pathways. By chaining together elementary tasks supported in the elementary facility, the combinatory facility enables the automation of genetics-led and network-based integrative prioritisation for genetic targets, both at the gene level (**cTGene**) and at the crosstalk level (**cTCrosstalk**). Notably, in addition to target crosstalk, the **cTCrosstalk** further supports target pathway prioritisation and crosstalk-based drug repurposing analysis (that is, repositioning approved drugs from original disease indications into new ones).

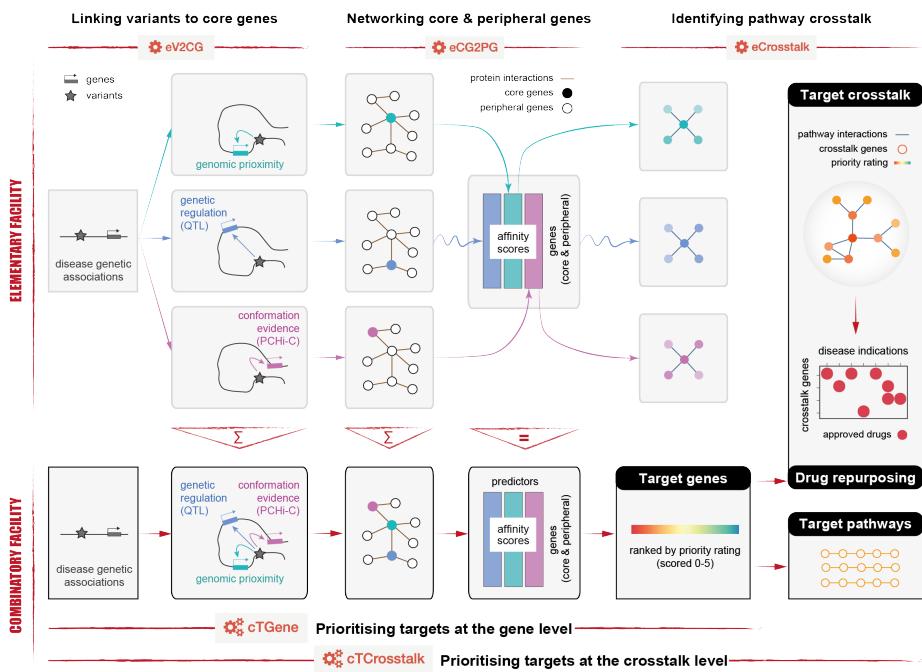


Figure 2.1: Schematic illustration of two facilities supported in the PiER.

# Chapter 3

# Compatibility

Table 3.1: A summary of the PiER browser compatibility.

	MacOS (Big Sur)	Windows (10)	Linux (Ubuntu)
Safari	14.1.2	N/A	N/A
Microsoft Edge	N/A	85.0.564.67	N/A
Google Chrome	96.0.4664.110	90.0.4430.93	96.0.4664.110
Firefox	95.0.2	95.0.2	95.0.2

## Chapter 4

# Runtime

Table 4.1: A summary of the runtime (on the server and client sides) per tool estimated using Google Chrome.

Facilities	Tools	Runtime (Server + Client)
Elementary	eV2CG	(67 + 82) seconds
Elementary	eCG2PG	(15 + 70) seconds
Elementary	eCrosstalk	(53 + 71) seconds
Combinatory	cTGene	(90 + 91) seconds
Combinatory	cTCrosstalk	(143 + 97) seconds

## Chapter 5

# Frontpage

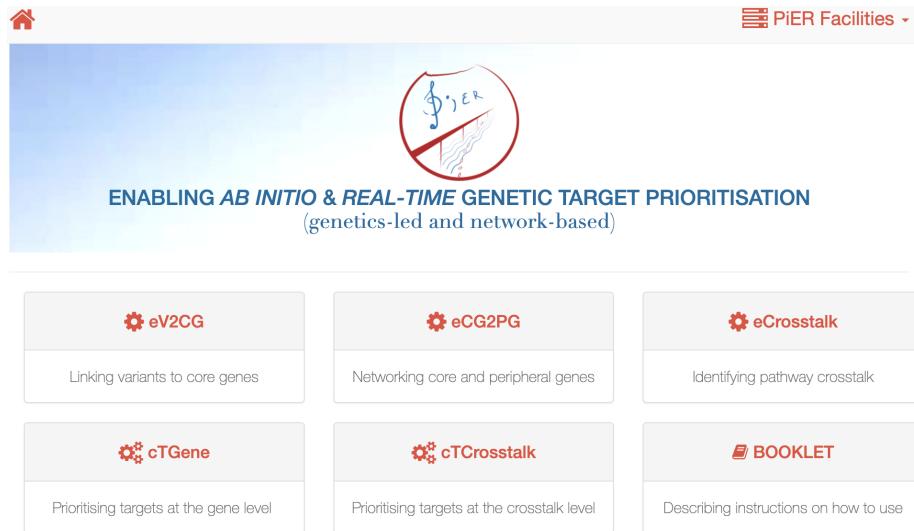


Figure 5.1: The landing frontpage (visited using Google Chrome in MacBook Pro) of the PiER, featuring two facilities ('elementary' and 'combinatory'). The elementary facility includes: (i) 'eV2CG', linking disease associated variants (particularly located at the non-coding genomic region) to core genes likely responsible for genetic associations, based on either promoter capture Hi-C (PCHi-C, that is, conformation evidence), quantitative trait locus (QTL) mapping (that is, genetic regulation of gene expression or protein abundance), or simply genomic proximity; (ii) 'eCG2PG', using knowledge of protein interactions to 'network' core genes with each other and with additional peripheral genes as well, producing a ranked list of core and peripheral genes; and (iii) 'eCrosstalk', exploiting the information of pathway-derived interactions to identify highly-ranked genes that mediate the crosstalk between molecular pathways. By chaining together elementary tasks supported in the elementary facility, the combinatory facility enables automation of genetics-led and network-based integrative prioritisation for genetic targets: (iv) at the gene level ('cTGene'); and (v) at the crosstalk level ('cTCrosstalk'). Also included is the tutorial-like booklet (in an HTML format) describing step-by-step instructions on how to use.

# Chapter 6

# Development

The PiER was developed using a next-generation Perl web framework Mojolicious that requires nearly zero-effort maintenance for interface updates. The PiER was also built using Bootstrap that supports the mobile-first and responsive webserver. The source codes are made available at GitHub.

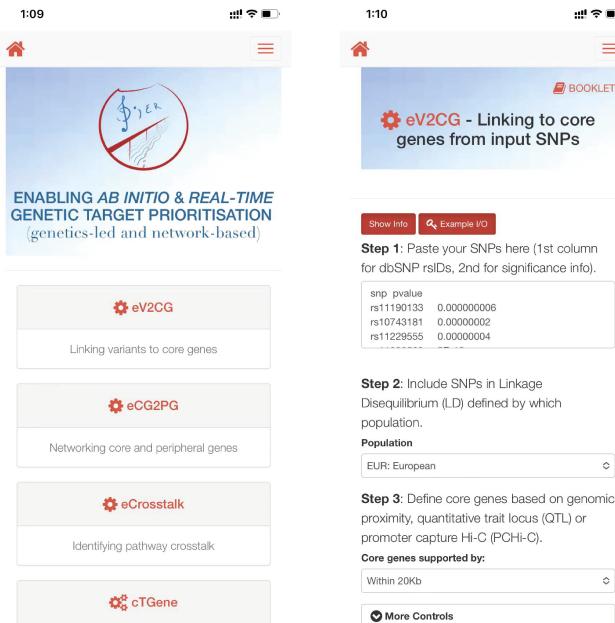


Figure 6.1: The screenshots for the PiER visited using Google Chrome in iPhone. Left: the frontpage; Right: the ‘eV2CG’ interface.

# Chapter 7

## Help buttons

Each user-request interface has the **Show/Hide Info** toggle button that contains the help information on use, including the details on input, output, mechanism and other useful information, while the **Example I/O** button showcases the example input/output. For example, shown below is the screenshot in the **cTCrosstalk** interface.

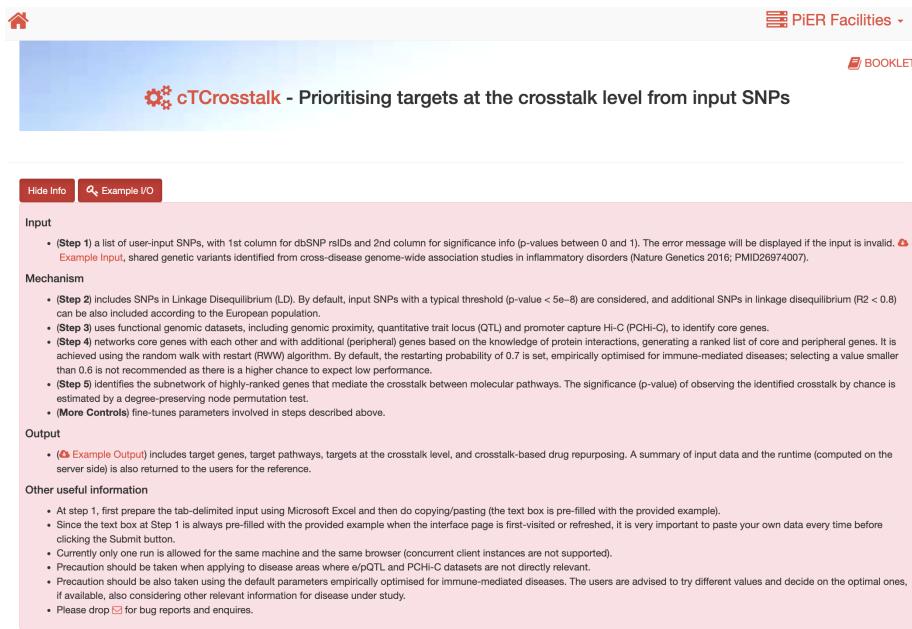


Figure 7.1: The screenshots for the ‘Show/Hide Info’ toggle button in the ‘cTCrosstalk’ interface.

# Chapter 8

## Error messages

The error messages will be displayed, for example, if the input into the cTCrosstalk is invalid (see the screenshot below). Notably, in the results page, a summary of input data is also returned to the users for the reference.



Figure 8.1: The screenshot for the error messages shown when the input is invalid, for example, in the ‘cTCrosstalk’ interface.

# Chapter 9

## eV2CG

### 9.1 Interface

#### Input

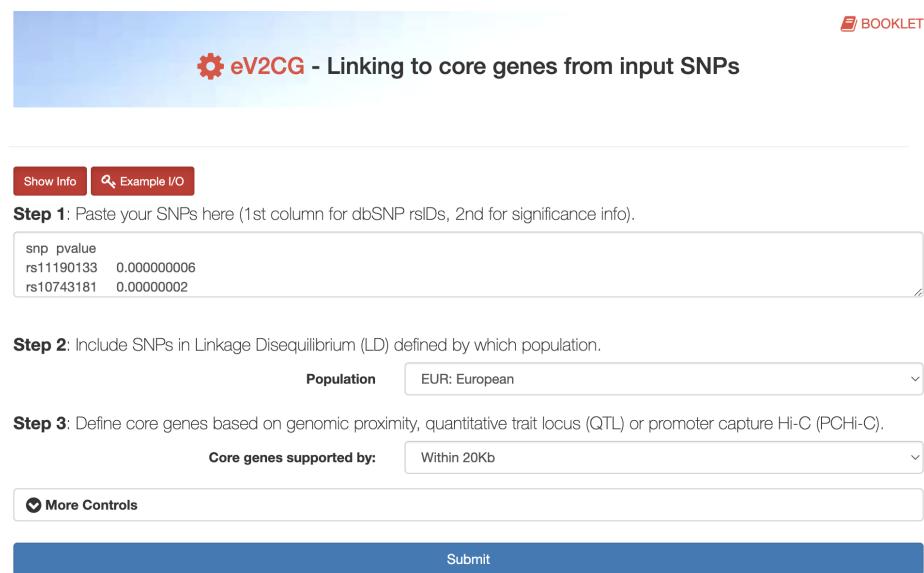
- Step 1: a list of user-input SNPs, with 1st column for dbSNP rsIDs and 2nd column for significance info (p-values between 0 and 1). The error message will be displayed if the input is invalid. Example input data are shared genetic variants identified from cross-disease genome-wide association studies in inflammatory disorders; see Nature Genetics 2016.

#### Mechanism

- Step 2: includes SNPs in Linkage Disequilibrium (LD). By default, input SNPs with a typical threshold ( $p\text{-value} < 5e-8$ ) are considered, and additional SNPs in linkage disequilibrium ( $R^2 < 0.8$ ) can be also included according to the European population.
- Step 3: uses genomic proximity, quantitative trait locus (QTL), or promoter capture Hi-C (PCHi-C) to identify core genes.
- More Controls: fine-tunes parameters involved in steps described above.

#### Output

- Example Output includes two interactive tables for core genes and evidence used, and a manhattan plot (illustrating scored core genes color-coded by chromosomes). A summary of input data and the runtime (computed on the server side) is also returned to the users for the reference.



The screenshot shows the eV2CG web application interface. At the top right is a 'BOOKLET' icon. Below it, the title 'eV2CG - Linking to core genes from input SNPs' is displayed with a gear icon. There are two buttons at the top left: 'Show Info' and 'Example I/O'. A text input area contains the following data:

snp	pvalue
rs11190133	0.000000006
rs10743181	0.00000002

**Step 1:** Paste your SNPs here (1st column for dbSNP rsIDs, 2nd for significance info).

**Step 2:** Include SNPs in Linkage Disequilibrium (LD) defined by which population. A dropdown menu shows 'Population: EUR: European'.

**Step 3:** Define core genes based on genomic proximity, quantitative trait locus (QTL) or promoter capture Hi-C (PCHi-C). A dropdown menu shows 'Core genes supported by: Within 20Kb'.

A 'More Controls' button is located below the dropdowns. At the bottom is a large blue 'Submit' button.

Figure 9.1: The interface of eV2CG, linking disease associated variants (particularly located at the non-coding genomic region) to (core) genes likely responsible for associations, based on either promoter capture Hi-C (PCHi-C; conformation evidence), quantitative trait locus (QTL) mapping (that is, genetic regulation of gene expression or protein abundance), or simply genomic proximity. The ‘Show/Hide Info’ toggle button contains the help information on how to use the ‘eV2CG’, including input, output, mechanism, etc.

## 9.2 Linking results

- Under the tab **Output: core genes**, **Manhattan plot** illustrates scored core genes that are color-coded by chromosomes. Also provided is the downloadable PDF file.
- Under the tab **Output: core genes**, **An interactive table** lists core genes linked from the input SNPs, with scores quantifying the level of genes responsible for genetic associations (capped at 100). Genes are cross-referenced and hyperlinked to GeneCards. Also provided is the column **Evidence** used to define core genes.
- Under the tab **Output: core genes**, **Evidence table** for core genes, showing which SNPs (see the column **SNPs**) are used to define core genes (the column **Core genes**) based on which evidence (see the column **Evidence**). The column **SNP type** tells the SNP type (either **Input** for use-input SNPs or **LD** for LD SNPs). Notably, the column **Evidence** details datasets used: the prefix **Proximity\_** indicative of SNPs in the proximity, the prefix **PCHiC\_** for PCHi-C datasets, and the prefix **QTL\_** for e/pQTL datasets.

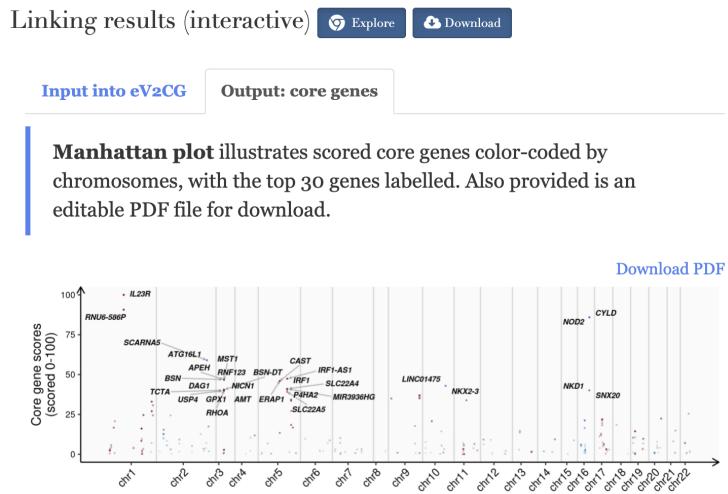


Figure 9.2: Interactive results for the ‘eV2CG’. Under the tab ‘Output: core genes’ is a manhattan plot illustrating scores for core genes. The user-input data under the tab ‘Input into eV2CG’ are also returned for the exploration.

**An interactive table** of core genes linked from the input SNPs, with the column `Scores` quantifying the degree of genes responsible for genetic associations (capped at 100). `Core genes` are cross-referenced and hyperlinked to [GeneCards](#). Also provided is the column `Evidence` used to define core genes; please refer to [Evidence table](#) for details.

	CSV	Copy	Search:
Core genes	Scores	Evidence	Description
<a href="#">IL23R</a>	100	Proximity_2000obp	interleukin 23 receptor
<a href="#">RNU6-586P</a>	90.7	Proximity_2000obp	RNA, U6 small nuclear 586, pseudogene
<a href="#">NOD2</a>	85.92	Proximity_2000obp	nucleotide binding oligomerization domain containing 2
<a href="#">CYLD</a>	85.92	Proximity_2000obp	CYLD lysine 63 deubiquitinase
<a href="#">ATG16L1</a>	59	Proximity_2000obp	autophagy related 16 like 1

Showing 1 to 5 of 410 entries

Previous 1 2 3 4 5 ... 82 Next

**Evidence table** for core genes, showing which SNPs (see the column `SNPs`) are used to define core genes (the column `core genes`) based on which evidence (see the column `Evidence`). The column `SNP type` tells the SNP type (either `Input` for use-input SNPs or `LD` for LD SNPs). Notably, the column `Evidence` details datasets used: the prefix `Proximity_` indicative of SNPs in the proximity, the prefix `PCHiC_` for PCHi-C datasets, and the prefix `QTL_` for e/QTL datasets.

	CSV	Copy	Search:
Core genes	SNPs	SNP type	Evidence
IL23R	rs10889676	Input	Proximity_2000obp
IL23R	rs183686347	Input	Proximity_2000obp
IL23R	rs7517847	Input	Proximity_2000obp
IL23R	rs80174646	Input	Proximity_2000obp
IL23R	rs1004820	LD	Proximity_2000obp

Showing 1 to 5 of 5,978 entries

Previous 1 2 3 4 5 ... 1,196 Next

Figure 9.3: Two tabular displays about core genes (top) and evidence (bottom) under the tab ‘Output: core genes’.

# Chapter 10

## eCG2PG

### 10.1 Interface

#### Input

- Step 1: a list of user-defined core genes, with 1st column for gene symbols, 2nd columns for weights (positive values), such as results from eV2CG above. The error message will be displayed if the input is invalid.

#### Mechanism

- Step 2: networks core genes with each other and with additional (peripheral) genes based on the knowledge of protein interactions, generating a ranked list of core and peripheral genes. It is achieved using the random walk with restart (RW) algorithm. By default, the restarting probability of 0.7 is set, empirically optimised for immune-mediated diseases; selecting a value smaller than 0.6 is not recommended as there is a higher chance to expect low performance.
- More Controls: fine-tunes parameters involved in steps described above.

#### Output

- Example Output includes an interactive table for core and peripheral genes, and a manhattan plot (illustrating scores for genes color-coded by chromosomes). A summary of input data and the runtime (computed on the server side) is also returned to the users for the reference.

### 10.2 Networking results

- Under the tab Output: core and peripheral genes, Manhattan plot illustrates affinity scores for genes that are color-coded by chromosomes.

The screenshot shows the eCG2PG interface. At the top, there's a header bar with a logo, the title 'eCG2PG - Networking core and peripheral genes based on input core genes', and a 'BOOKLET' button. Below the header, there are two buttons: 'Show Info' and 'Example I/O'. A text area labeled 'Step 1' contains the following input:

gene	weight
IL23R	100
RNU6-586P	90.7

Below this, 'Step 2' asks for a network type, with 'Protein interactions with high confidence (0.7)' selected. There's also a 'More Controls' link and a large blue 'Submit' button at the bottom.

Figure 10.1: The interface of the ‘eCG2PG‘, using the knowledge of protein interactions to ‘network’ core genes with each other and with additional (peripheral) genes as well, generating a ranked list of core and peripheral genes. The ‘Show/Hide Info‘ toggle button contains the help information on how to use the ‘eCG2PG‘, including input, output, mechanism, etc.

Also provided is the downloadable PDF file.

- Under the tab **Output: core and peripheral genes**, An interactive table lists core and peripheral genes, with scores quantifying the affinity to core genes (sum up to 1). Genes are cross-referenced and hyperlinked to GeneCards.

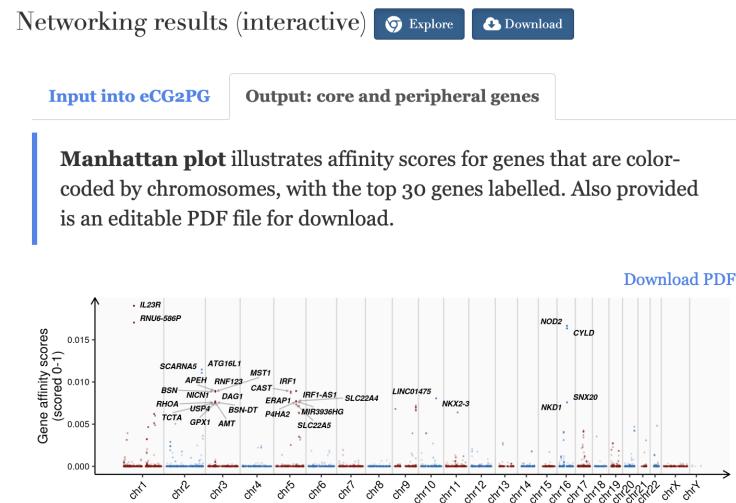


Figure 10.2: Interactive results for the ‘eCG2PG‘ under the tab ‘Output: core and peripheral genes‘. The user-input data the tab ‘Input into eCG2PG‘ are also returned for the exploration.

# Chapter 11

## eCrosstalk

### 11.1 Interface

#### Input

- Step 1: a ranked list of genes, with 1st column for gene symbols, 2nd columns for scores (positive values), such as results from eCG2PG above. The error message will be displayed if the input is invalid.

#### Mechanism

- Step 2: identifies the subnetwork of highly-ranked genes that mediate the crosstalk between molecular pathways. The significance (p-value) of observing the identified crosstalk by chance is estimated by a degree-preserving node permutation test.

#### Output

- Example Output includes an interactive table for pathway crosstalk genes, and a network visualisation (illustrating the crosstalk between pathways).

### 11.2 Crosstalk results

- Under the tab **Output: pathway crosstalk**, A network visualisation illustrates crosstalk genes color-coded by input scores. The significance (p-value) of observing the identified crosstalk by chance is estimated by a degree-preserving node permutation test. Also provided is the downloadable PDF file.
- Under the tab **Output: pathway crosstalk**, An interactive table lists crosstalk genes together with input scores. Genes are cross-referenced and hyperlinked to GeneCards.

BOOKLET

**eCrosstalk - Identifying pathway crosstalk based on input genes**

[Show Info](#) [Example I/O](#)

**Step 1:** Paste your genes here (1st column for gene symbols, 2nd for scoring info).

gene	score
IL23R	0.01903
RNU6-586P	0.01705

**Step 2:** Identify the crosstalk mediating molecular pathways with the desired number of genes.

Number of crosstalk genes:	30
Significance of the crosstalk:	Degree-preserving node permutation test

**Submit**

Figure 11.1: The interface of the ‘eCrosstalk’, exploiting the information of well-curated pathway-derived interactions to identify the subnetwork of highly ranked genes that mediate pathway crosstalk. The ‘Show/Hide Info’ toggle button introducing how to use the ‘eCrosstalk’, including input, output, mechanism, etc.

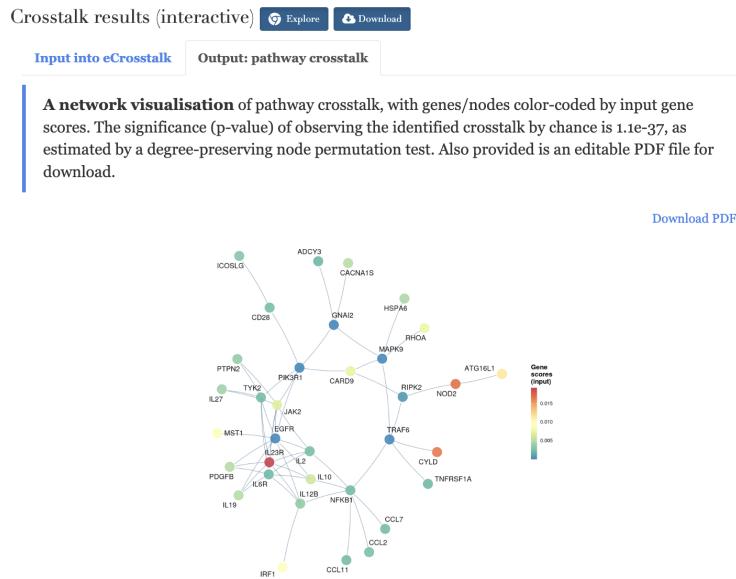


Figure 11.2: Interactive results for the ‘eCrosstalk’ under the tab ‘Output: pathway crosstalk’. The user-input data under the tab ‘Input into eCrosstalk’ are also returned for the exploration.

# Chapter 12

## cTGene

### 12.1 Interface

#### Input

- **Step 1:** a list of user-input SNPs, with 1st column for dbSNP rsIDs and 2nd column for significance info (p-values between 0 and 1). The error message will be displayed if the input is invalid. Example input data are shared genetic variants identified from cross-disease genome-wide association studies in inflammatory disorders; see Nature Genetics 2016.

#### Mechanism

- **Step 2:** includes SNPs in Linkage Disequilibrium (LD). By default, input SNPs with a typical threshold ( $p\text{-value} < 5e-8$ ) are considered, and additional SNPs in linkage disequilibrium ( $R^2 < 0.8$ ) can be also included according to the European population.
- **Step 3:** uses functional genomic datasets, including genomic proximity, quantitative trait locus (QTL) and promoter capture Hi-C (PCHi-C), to identify core genes.
- **Step 4:** networks core genes with each other and with additional (peripheral) genes based on the knowledge of protein interactions, generating a ranked list of core and peripheral genes. It is achieved using the random walk with restart (RWWR) algorithm. By default, the restarting probability of 0.7 is set, empirically optimised for immune-mediated diseases; selecting a value smaller than 0.6 is not recommended as there is a higher chance to expect low performance.
- **More Controls:** fine-tunes parameters involved in steps described above.

#### Output

- Example Output includes a manhattan plot (illustrating priority rating for target genes color-coded by chromosomes), and two tabular displays about prioritisation and evidence. A summary of input data and the runtime (computed on the server side) is also returned to the users for the reference.

The screenshot shows the 'cTGene - Prioritising targets at the gene level from input SNPs' interface. At the top right is a 'BOOKLET' icon. Below it is a logo with three red dots and the text 'cTGene - Prioritising targets at the gene level from input SNPs'. There are two buttons: 'Show Info' (highlighted in red) and 'Example I/O'. A text area labeled 'Step 1: Paste your SNPs here (1st column for dbSNP rsIDs, 2nd for significance info).' contains the following data:

snp	pvalue
rs11190133	0.00000006
rs10743181	0.00000002

**Step 2:** Include SNPs in Linkage Disequilibrium (LD) defined by which population. A dropdown menu labeled 'Population' is set to 'EUR: European'.

**Step 3:** Define core genes based on genomic proximity, quantitative trait locus (QTL) and promoter capture Hi-C (PCHi-C). Three dropdown menus are shown:

- Distance-to-SNP window: Within 20Kb
- QTL datasets: pQTL (plasma)
- PCHi-C datasets: Monocytes

**Step 4:** Network core and peripheral genes using knowledge of protein interactions. A dropdown menu labeled 'Network' is set to 'Protein interactions with high confidence (0.7)'.

At the bottom left is a 'More Controls' button, and at the bottom center is a large blue 'Submit' button.

Figure 12.1: The interface of the ‘cTGene’, enabling/automating genetics-led and network-based identification and prioritisation of drug targets at the gene level. The ‘Show/Hide Info’ toggle button contains the help information on how to use the ‘cTGene’, including input, output, mechanism, etc.

## 12.2 Prioritisation results

- Under the tab **Output: target genes, Manhattan plot** illustrates priority rating for target genes that are color-coded by chromosomes. Also provided is the downloadable PDF file.
- Under the tab **Output: target genes, Prioritisation table** lists all prioritised genes, each receiving 5-star priority rating (scored 0-5). Genes are cross-referenced and hyperlinked to GeneCards. The column Type tells the target gene type (either **Core** for core genes or **Peripheral** for peripheral genes). Also provided is a summary of evidence used to define

core genes, including columns **Proximity** (evidence of genomic proximity), **QTL** (e/pQTL evidence) and **PCHiC** (conformation evidence).

- Under the tab **Output: target genes**, **Evidence table** for core genes, showing which SNPs (see the column **SNPs**) are used to define core genes (the column **Core genes**) based on which evidence (see the column **Evidence**). The column **SNP type** tells the SNP type (either **Input** for user-input SNPs or **LD** for LD SNPs). Notably, the column **Evidence** details datasets used: the prefix **Proximity\_** indicative of SNPs in the proximity, the prefix **PCHiC\_** for PCHi-C datasets, and the prefix **QTL\_** for e/pQTL datasets.

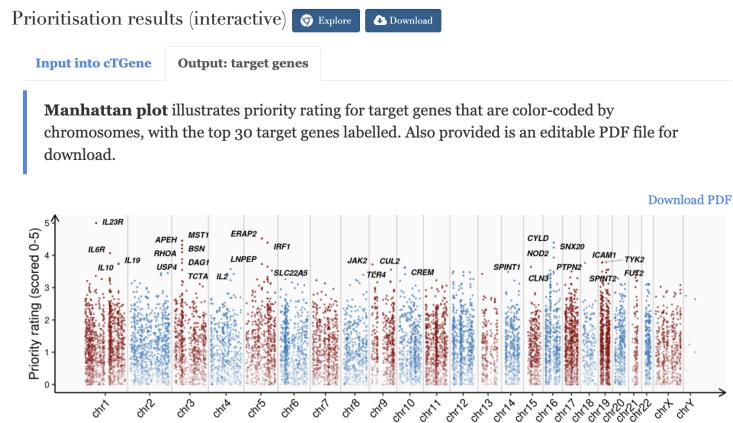


Figure 12.2: Prioritisation results for the ‘cTGene’. Under the tab ‘Output: target genes’ is a manhattan plot illustrating priority rating for target genes. The user-input data under the tab ‘Input into cTGene’ are also returned for the exploration.

**Prioritisation table** for targets at the gene level, each receiving 5-star priority rating (scored 0–5; see the column `Rating`) and its rank (see the column `Rank`). The prioritised target genes are cross-referenced and hyperlinked to [GeneCards](#); see the column `Genes`. The column `Type` tells the target gene type (either `core` for core genes or `Peripheral` for peripheral genes). Also provided is a summary of evidence used to define core genes, including columns `Proximity` (evidence of genomic proximity), `QTL` (e/pQTL evidence) and `PCHiC` (conformation evidence). For details on evidence, please refer to [Evidence table](#).

	CSV	Copy	Search:				
<b>Genes</b> <b>Rank</b> <b>Rating</b> <b>Type</b> <b>Proximity</b> <b>QTL</b> <b>PCHiC</b> <b>Description</b>							
IL23R	1	5	Core	1	1	1	interleukin 23 receptor
ERAP2	2	4.519	Core	1	1	1	endoplasmic reticulum aminopeptidase 2
MST1	3	4.454	Core	1	1	0	macrophage stimulating 1
CYLD	4	4.396	Core	1	0	1	CYLD lysine 63 deubiquitinase
IRF1	5	4.393	Core	1	0	1	interferon regulatory factor 1

Showing 1 to 5 of 13,973 entries

Previous 1 2 3 4 5 ... 2,795 Next

**Evidence table** for core genes, showing which SNPs (see the column `SNPs`) are used to define core genes (the column `Core genes`) based on which evidence (see the column `Evidence`). The column `SNP type` tells the SNP type (either `Input` for use-input SNPs or `LD` for LD SNPs). Notably, the column `Evidence` details datasets used: the prefix `Proximity_` indicative of SNPs in the proximity, the prefix `PCHiC_` for PCHi-C datasets, and the prefix `QTL_` for e/pQTL datasets.

	CSV	Copy	Search:
<b>Core genes</b> <b>SNPs</b> <b>SNP type</b> <b>Evidence</b>			
IL23R	rs10889676	Input	Proximity_2000obp
IL23R	rs183686347	Input	Proximity_2000obp
IL23R	rs7517847	Input	Proximity_2000obp
IL23R	rs80174646	Input	Proximity_2000obp
IL23R	rs1004820	LD	Proximity_2000obp

Showing 1 to 5 of 9,359 entries

Previous 1 2 3 4 5 ... 1,872 Next

Figure 12.3: Two tabular displays about target genes (top) and evidence (bottom) under the tab ‘Output: target genes’.

# Chapter 13

## cTCrosstalk

### 13.1 Interface

#### Input

- **Step 1:** a list of user-input SNPs, with 1st column for dbSNP rsIDs and 2nd column for significance info (p-values between 0 and 1). The error message will be displayed if the input is invalid. Example input data are shared genetic variants identified from cross-disease genome-wide association studies in inflammatory disorders; see Nature Genetics 2016.

#### Mechanism

- **Step 2:** includes SNPs in Linkage Disequilibrium (LD). By default, input SNPs with a typical threshold ( $p\text{-value} < 5e-8$ ) are considered, and additional SNPs in linkage disequilibrium ( $R^2 < 0.8$ ) can be also included according to the European population.
- **Step 3:** uses functional genomic datasets, including genomic proximity, quantitative trait locus (QTL) and promoter capture Hi-C (PCHi-C), to identify core genes.
- **Step 4:** networks core genes with each other and with additional (peripheral) genes based on the knowledge of protein interactions, generating a ranked list of core and peripheral genes. It is achieved using the random walk with restart (RWWR) algorithm. By default, the restarting probability of 0.7 is set, empirically optimised for immune-mediated diseases; selecting a value smaller than 0.6 is not recommended as there is a higher chance to expect low performance.
- **Step 5:** identifies the subnetwork of highly-ranked genes that mediate the crosstalk between molecular pathways. The significance (p-value)

of observing the identified crosstalk by chance is estimated by a degree-preserving node permutation test.

- **More Controls:** fine-tunes parameters involved in steps described above.

### Output

- Example Output includes target genes, target pathways, targets at the crosstalk level, and crosstalk-based drug repurposing. A summary of input data and the runtime (computed on the server side) is also returned to the users for the reference.

The screenshot shows the 'cTCrosstalk' web application interface. At the top, there's a blue header bar with the title 'cTCrosstalk - Prioritising targets at the crosstalk level from input SNPs'. Below the header, there are two buttons: 'Show Info' and 'Example I/O'. The main area is divided into five numbered steps:

- Step 1:** Paste your SNPs here (1st column for dbSNP rsIDs, 2nd for significance info). A text input field contains the following data:
 

snp	pvalue
rs11190133	0.000000006
rs10743181	0.00000002
- Step 2:** Include SNPs in Linkage Disequilibrium (LD) defined by which population. A dropdown menu is set to 'EUR: European'.
- Step 3:** Define core genes based on genomic proximity, quantitative trait locus (QTL) and promoter capture Hi-C (PCHi-C). Three dropdown menus are shown:
  - Distance-to-SNP window:** Within 20Kb
  - QTL datasets:** pQTL (plasma)
  - PCHi-C datasets:** Monocytes
- Step 4:** Network core and peripheral genes using knowledge of protein interactions. A dropdown menu is set to 'Protein interactions with high confidence (0.7)'.
- Step 5:** Identify the crosstalk mediating molecular pathways with the desired number of genes. Two dropdown menus are shown:
  - Number of crosstalk genes:** 30
  - Significance of the crosstalk:** Degree-preserving node permutation test

At the bottom left is a 'More Controls' button, and at the bottom right is a large blue 'Submit' button.

Figure 13.1: The interface of the ‘cTCrosstalk’, enabling/automating genetics-led and network-based identification and prioritisation of drug targets at the crosstalk level. The ‘Show/Hide Info’ toggle button contains the help information on how to use the ‘cTCrosstalk’, including input, output, mechanism, etc.

## 13.2 Prioritisation results

- **Output: target genes:** includes **Manhattan plot** illustrating priority rating for target genes that are color-coded by chromosomes. Also provided is the downloadable PDF file. It also includes **Prioritisation table** listing all prioritised genes, each receiving 5-star priority rating (scored 0-5), and **Evidence table** for core genes showing which SNPs are used to define core genes based on which evidence. Genes are cross-referenced and hyperlinked to GeneCards.
- **Output: target pathways:** includes a dot plot and a prioritisation table for target pathways. Also provided is the downloadable PDF file.
- **Output: targets at the crosstalk level:** includes **A network visualisation** illustrating the crosstalk between pathways, with genes colored by priority rating and labelled in the form of **rating@rank**, **Prioritisation table** listing crosstalk genes, each receiving 5-star priority rating (scored 0-5), and **Evidence table** for pathway crosstalk genes, showing which SNPs are used to crosstalk genes based on which evidence. Genes are cross-referenced and hyperlinked to GeneCards.
- **Output: crosstalk-based drug repurposing:** includes **A heatmap-like illustration** showing drug repurposing analysis of approved drugs (licensed medications) based on pathway crosstalk genes, with crosstalk genes on y-axis, disease indications on x-axis, red dots indexed in number and referenced beneath in the table where the information on approved drugs and mechanisms of action is detailed. It also includes **An interactive table** of crosstalk genes (the column **Crosstalk genes**), disease indications (the column **Disease indications**), approved drugs and mechanisms (the column **Approved drugs [mechanisms of action]**), and drug index (the column **Index**) shown above within the dot plot.

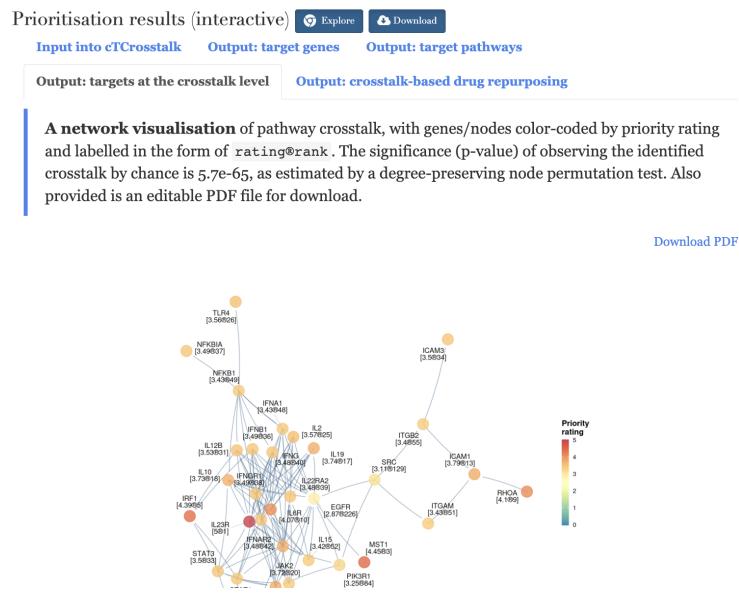


Figure 13.2: Prioritisation results for the ‘cTCrosstalk’. In addition to a summary of input data and the runtime (computed on the server side) under the tab ‘Input into cTCrosstalk’, the prioritisation results page provides the output, including target genes under the tab ‘Output: target genes’ (the same as shown in the ‘cTGene’), target pathways under the tab ‘Output: target pathways’, and targets at the crosstalk level under the tab ‘Output: targets at the crosstalk level’, and crosstalk-based drug repurposing under the tab ‘Output: crosstalk-based drug repurposing’. Under the tab ‘Output: target genes’ include network visualisation of the crosstalk, with genes/nodes colour-coded by priority rating and labelled in the form of ‘rating@rank’, and two tabular displays about prioritisation and evidence for crosstalk genes.

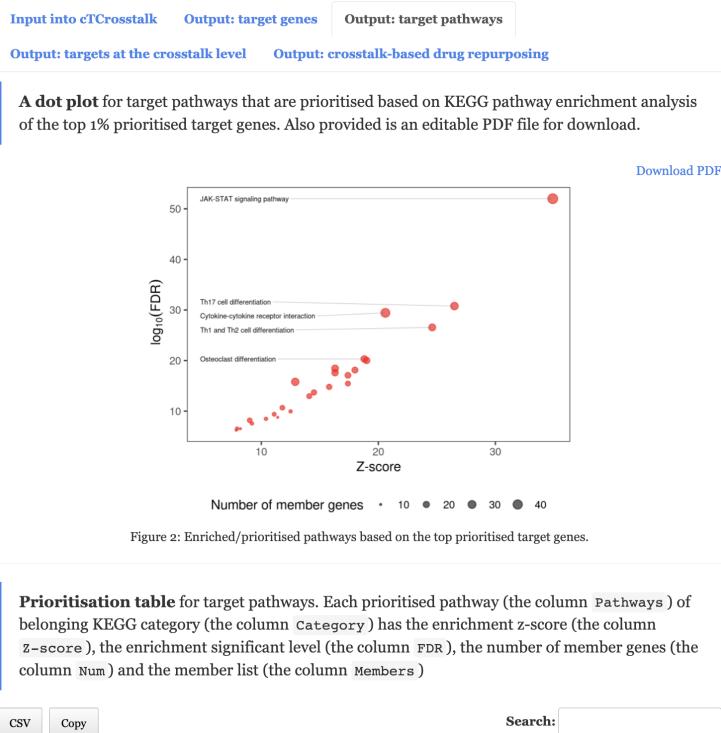


Figure 13.3: A dot plot for prioritised target pathways, with the top five labelled, available under the tab ‘Output: target pathways’. Also available is ‘Prioritisation table’ for target pathways.

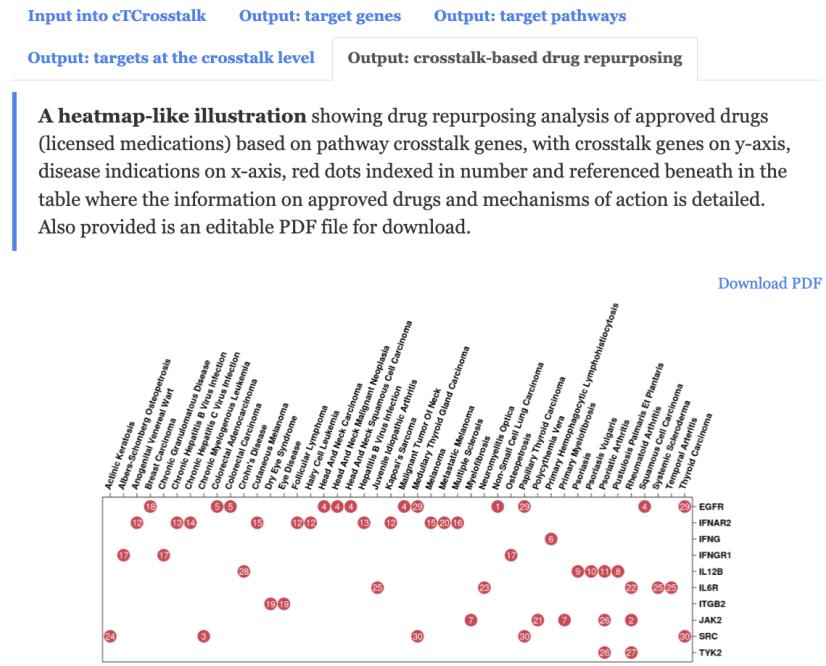


Figure 4: Drug repurposing analysis of approved drugs (licensed medications) based on pathway crosstalk genes. The heatmap shows crosstalk genes (y-axis) and disease indications (x-axis), with red dots indexed in number and referenced beneath the table where the information on approved drugs and mechanisms of action is also listed.

An **interactive table** of crosstalk genes (the column `crosstalk genes`), disease indications (the column `Disease indications`), approved drugs and mechanisms (the column `Approved drugs (mechanisms of action)`), and drug index (the column `Index`) shown above within the dot plot. The crosstalk genes are cross-referenced and hyperlinked to [GeneCards](#).



Figure 13.4: A heatmap-like illustration, with crosstalk genes on the y-axis, disease indications on the x-axis, and red dots indexed in numbers under the tab ‘Output: crosstalk-based drug repurposing’. The index numbers are referenced in a table where the information on approved drugs and mechanisms of action is detailed.