



Tutorial 1.1: Import Data Matrix From IntOGen

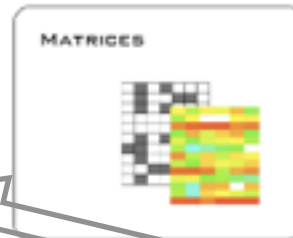
Import a data matrix from IntOGen that contains p-values for significantly up-regulated genes in each tumor morphology type



Import data

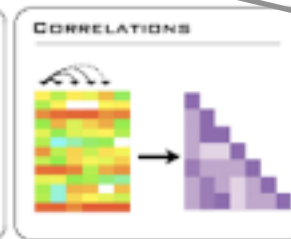
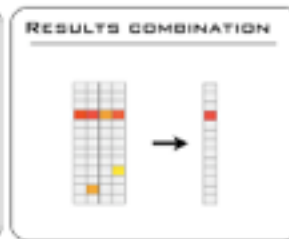
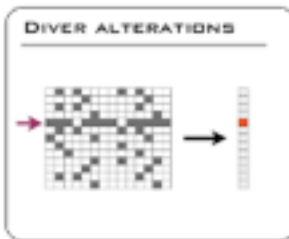


Data types



These buttons will open a box that explains the different types of data used in Gitools

Start an analysis



STEP 1

Click Import
Data from
IntOGen



Welcome x Demo x



Import data

bio::mart



Start an analysis

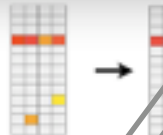
ENRICHMENT ANALYSIS



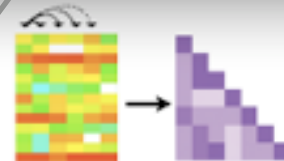
DIVER ALTERATIONS



RESULTS COMBINATION



CORRELATIONS



From IntOGen you can
download Matrices or
Modules.

Import IntOGen data



What kind of data do you want to import ?

☒ Matrix

Select this option to download a matrix with
one gene per row and one experiment or
tumor type per column with p-values per cell.

☐ Oncomodules

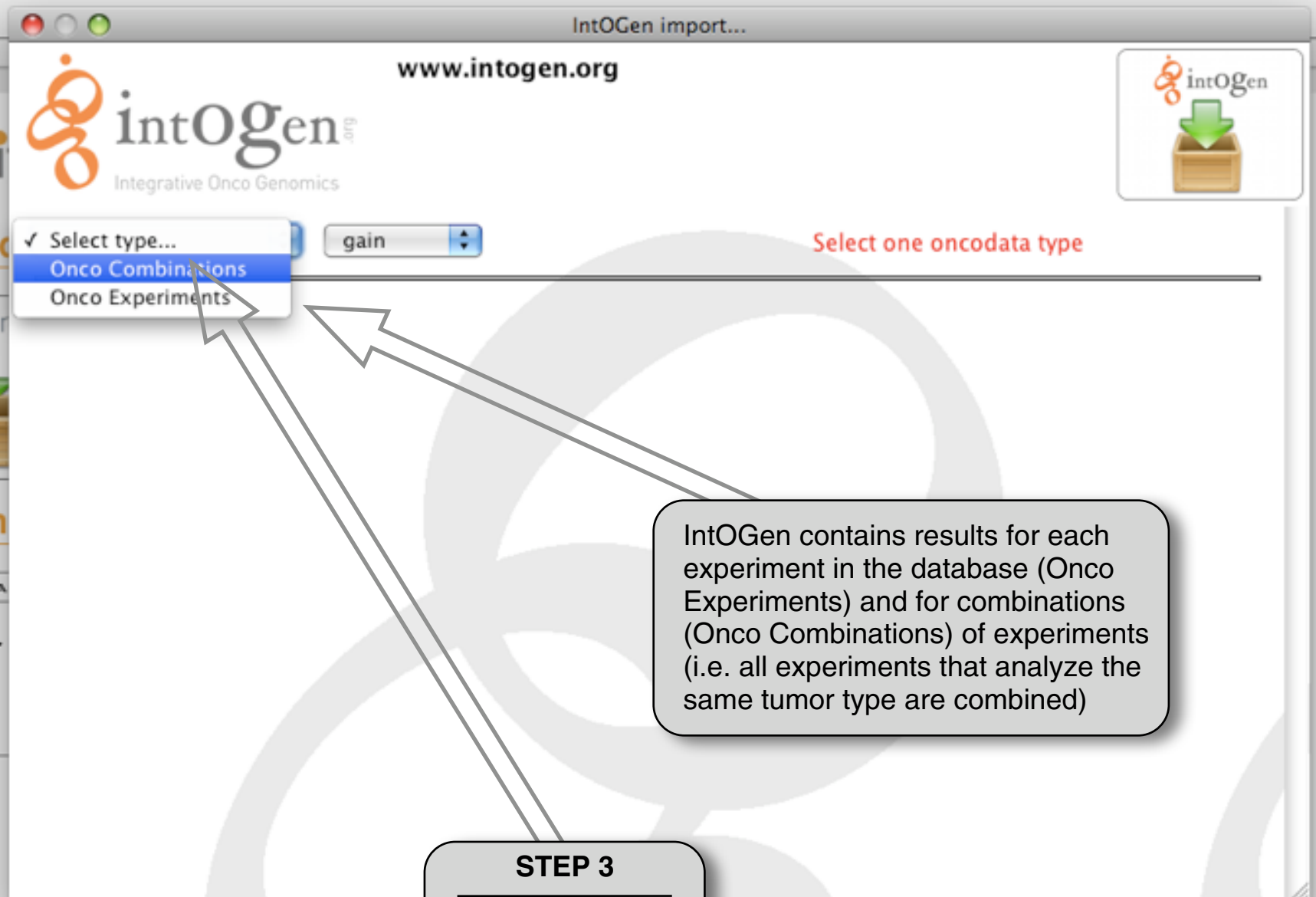
Select this option to download cancer
related modules (gene sets).

OK

Cancel

STEP 2

Select Matrix
and click OK.



IntOGen contains results for each experiment in the database (Onco Experiments) and for combinations (Onco Combinations) of experiments (i.e. all experiments that analyze the same tumor type are combined)

STEP 3

Select Onco
Combinations



www.intogen.org

Integrative Onco Genomics

Onco Combinations

- ✓ gain
- loss
- upreg**
- downreg

Add some filters below (use CTRL or ⌘ key to select more than one)

Topography

- adrenal cortex
- adrenal glands
- anal canal and anus
- bones and joints
- brain, nos
- breast
- cerebellum

Pathology

- teratoid malignant melanoma, mali
- acute lymphoblastic leukemia, no
- acute myeloid leukemia, nos
- adenocarcinoma, intestinal type
- adenocarcinoma, nos
- adenoid cystic carcinoma
- adenoma, nos

IntOGen contains results for different types of alterations in cancer samples. Gain and Loss of DNA and over (upreg) or under-expression (downreg) of genes/transcripts.

STEP 4

Select upreg to get results for upregulation



Import



Onco Combinations

upreg

You have selected 20 columns.

Download

Topography

adrenal cortex
bones and joints
brain, nos
breast
cervix uteri
colon, nos
hematopoietic and reticuloendothelial

Morphology

ALL
acute lymphoblastic leukemia, no
acute myeloid leukemia, nos
adenocarcinoma, nos
adenoma, nos
adrenal cortical carcinoma
astrocytoma, nos

STEP 5

Select all
Topographies
using SHIFT key

STEP 6

Select ALL in
Morphology and
click Download

ALL in Morphology means that samples are grouped independently of their morphology type.

In IntOGen samples are grouped by Topography and Morphology according to International Classification of Disease Oncology (ICD). All samples with same ICD terms and analyzed together in the same platform form an experiment unit. Experiment units with the same ICD terms are combined to obtain results for Onco Combinations.

For more information see
<http://www.intogen.org>



▼ datamatrices

intogen_combinations_all_tumortypes.cdm.gz
 intogen_combinations_all_tumortypes_row_annotations.tsv
 intogen_combinations_all_tumortypes_column_annotations.tsv

Once the download has been completed you will have three new files in the directory that you indicated.

	A	B	C	D	E	F	G	H	I	J
1		110-4	118-4	120-4	131-4	140-4	145-4	160-4	168-4	172-4
2	ENSG000000000003	0.01269054	1	0.40276613	0.94055216	1	1	0.999999	1	5.64E-11
3	ENSG000000000005	-	0.999999	1	1	1	1	0.99998161	1	1
4	ENSG000000000419	1	0.94613856	0.98268852	1	1	1	0.999999	1	1
5	ENSG000000000457	1	0.999999	1	1	1	1	0.99998469	1	1
6	ENSG000000000460	0.999999	1	0.00012587	0.99999986	0.9998687	0.9999997	0.61551355	0.99776226	0.96119548
7	ENSG000000000938	0.96289456	1	0.9999926	0.99998991	1	0.99864867	1	1	0.9817801
8	ENSG000000000971	1	1	3.33E-16	0.91307589	1	0	0.99642733	1	0.00012065
9	ENSG000000001036	-	-	6.16E-05	5.91E-05	1	1	1	1	0.999999
10	ENSG000000001084	0.99957262	0.00011951	1	0.99999992	1	0.999999	1	1	1
11	ENSG000000001167	1	1	1	1	0.99920007	0.9999997	1	1	0.96119548
12	ENSG000000001460	-	-	0.999999	0.999999	1	1	-	1	0.999999
13	ENSG000000001461	1	1	1	1	1	1	0.95000265	0.99999993	1
14	ENSG000000001497	1	1	1	1	1	1	1	0.99999698	1
15	ENSG000000001561	0.999999	1	0.4386625	0.99999742	1	1	0.999999	1	1
16	ENSG000000001617	0.999999	1	1	1	0.95858806	0.47434021	0.047148	1	0.99981133
17	ENSG000000001626	1.00E-06	1	0.73280025	1	1	0.46039017	0.90212403	0.93020477	1
18	ENSG000000001629	0.999999	-	0.00072387	0.999999	1	0.99924487	0.999999	1	0.999999
19	ENSG000000001630	1	1	1	0.99999567	1	1	0.99997859	0.43544254	0.9999854
20	ENSG000000001631	1	0.98756234	1	1	1	1	0.999999	0.88148552	1
21	ENSG000000002016	0.59272721	1	1	1	1	0.98876601	0.5461895	1	0.96119548
22	ENSG000000002079	-	-	0.999999	0.999999	1	1	-	-	0.999999
23	ENSG000000002330	1	0.999999	1	1	1	1	1	1	1
24	ENSG000000002549	0.71507907	0.999999	0.04193981	0.00037975	1	1	1	1	0.999999
25	ENSG000000002586	0.999999	1	0.99999998	1	0.94841068	1	1	1	1
26	ENSG000000002587	1	1	0.92965581	0.99950351	1	1	1	1	1
27	ENSG000000002726	1	1	0.99999266	1	1	1	1	1	1
28	ENSG000000002745	-	0.999999	1	1	1	1	1	1	1
29	ENSG000000002746	1	0.999999	1	1	1	1	1	1	1
30	ENSG000000002822	1	1	1	1	0.99992542	0	1	1	1

The file with extension cdm contains a Continuous Data Matrix with one row per gene and one column per tumor type. The values in each cell are corrected p-values that indicate if the gene is significantly up-regulated in this tumor type.

▼ datamatrices

intogen_combinations_all_tumortypes.cdm.gz
 intogen_combinations_all_tumortypes_row_annotations.tsv
 intogen_combinations_all_tumortypes_column_annotations.tsv

intogen.combinations.all.tumortypes.row.annotations.tsv

id	symbol	chromosome	band
ENSG000000000003	TSPAN6	X	q22.1
ENSG000000000005	TNMD	X	q22.1
ENSG000000000419	DPM1	20	q13.13
ENSG000000000457	SCYL3	1	q24.2
ENSG000000000460	C1orf112	1	q24.2
ENSG000000000938	FGR	1	p36.11
ENSG000000000971	CFH	1	q31.3
ENSG000000001036	FUCA2	6	q24.2
ENSG000000001084	GCLC	6	p12.1
ENSG000000001167	NFYA	6	p21.1
ENSG000000001460	C1orf201	1	p36.11
ENSG000000001461	NIPAL3	1	p36.11
ENSG000000001497	LAS1L	X	q12
ENSG000000001561	ENPP4	6	p21.1
ENSG000000001617	SEMA3F	3	p21.31
ENSG000000001626	CFTR	7	q31.2
ENSG000000001629	ANKIB1	7	q21.2
ENSG000000001630	CYP51A1	7	q21.2
ENSG000000001631	KRIT1	7	q21.2
ENSG000000002016	RAD52	12	p13.33
ENSG000000002079	AC004834.1	7	q22.1
ENSG000000002330	BAD	11	q13.1
ENSG000000002549	LAP3	4	p15.32
ENSG000000002586	CD99	X	p22.33
ENSG000000002587	HS3ST1	4	p15.33
ENSG000000002726	ABP1	7	q36.1
ENSG000000002745	WNT16	7	q31.31
ENSG000000002746	HECW1	7	p14.1
ENSG000000002822	USP44	7	p22.3

This file is a table with the annotations of rows. For each gene ensembl id it indicates the corresponding gene symbol, chromosome and cytogenetic band.

intogen.combinations.all.tumortypes.column.annotations.tsv

id	topography	morphology
110-4	testis	ALL
118-4	mouth, nos	ALL
120-4	cervix uteri	ALL
131-4	hematopoietic and reticuloendothelial systems	ALL
140-4	breast	ALL
145-4	kidney	ALL
160-4	thyroid gland	ALL
168-4	prostate gland	ALL
172-4	brain, nos	ALL
18-4	adrenal cortex	ALL
188-4	bones and joints	ALL
189-4	pancreas	ALL
201-4	liver	ALL
202-4	ovary	ALL
42-4	lung	ALL
44-4	skin	ALL
49-4	lymph nodes	ALL
85-4	colon, nos	ALL
9-4	urinary bladder	ALL
94-4	stomach	ALL

This file is a table with the annotations of columns. For each tumor type id it indicates the corresponding topography and morphology.



THANKS FOR USING GITTOOLS

<http://www.gittools.org>