Nozzle: a framework to generate reports for high-throughput analysis pipelines

---

Nils Gehlenborg

TCGA GDAC Research Presentation / 4 August 2011

Acknowledgements

Michael S Noble
Douglas Voet

Lihua Zou
Gordon Saksena
Pei Lin

Dan DiCara
Hailei Zhang
John Zhang
Richard Park
Robert Zupko
Sachet Shukla
Spring Liu
Terrence Wu

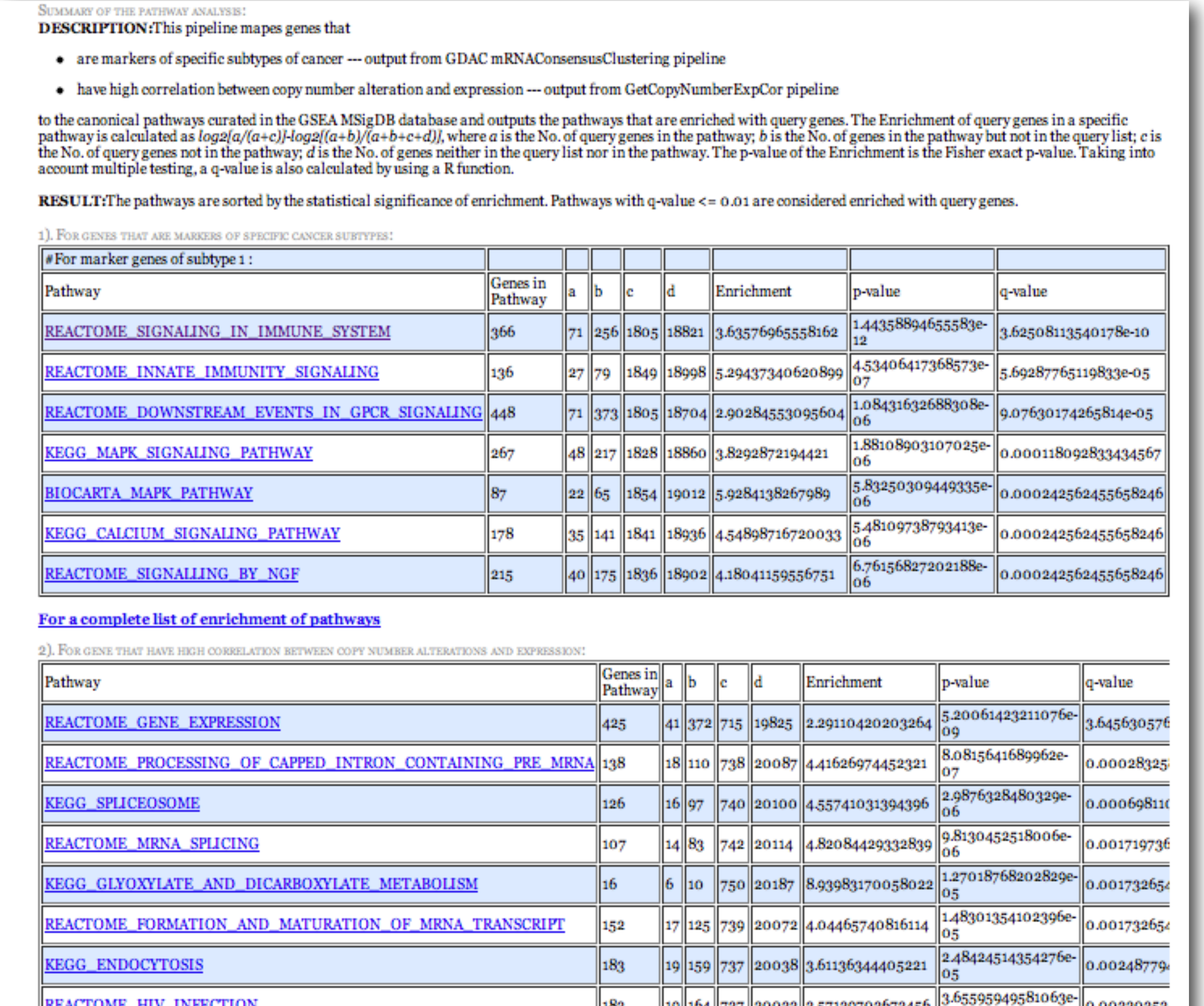
Peter J Park
Gad Getz
Lynda Chin

What is “Nozzle”?

- "A device attached to the end of a fire hose that **directs**, **shapes** and **regulates** the flow of the water or fire fighting agent pumped into the hose. (Wikipedia)

- Our R Package to generate Firehose reports based on HTML, CSS (cascading style sheets), JavaScript and semantic markup.

Pre-Nozzle Reports



SUMMARY OF THE PATHWAY ANALYSIS:

DESCRIPTION: This pipeline mapes genes that

- are markers of specific subtypes of cancer --- output from GDAC mRNAConsensusClustering pipeline
- have high correlation between copy number alteration and expression --- output from GetCopyNumberExpCor pipeline

to the canonical pathways curated in the GSEA MSigDB database and outputs the pathways that are enriched with query genes. The Enrichment of query genes in a specific pathway is calculated as *log2[a/(a+c)]-log2[(a+b)/(a+b+c+d)]*, where *a* is the No. of query genes in the pathway; *b* is the No. of genes in the pathway but not in the query list; *c* is the No. of query genes not in the pathway; *d* is the No. of genes neither in the query list nor in the pathway. The p-value of the Enrichment is the Fisher exact p-value. Taking into account multiple testing, a q-value is also calculated by using a R function.

RESULT: The pathways are sorted by the statistical significance of enrichment. Pathways with q-value <= 0.01 are considered enriched with query genes.

1). FOR GENES THAT ARE MARKERS OF SPECIFIC CANCER SUBTYPES:

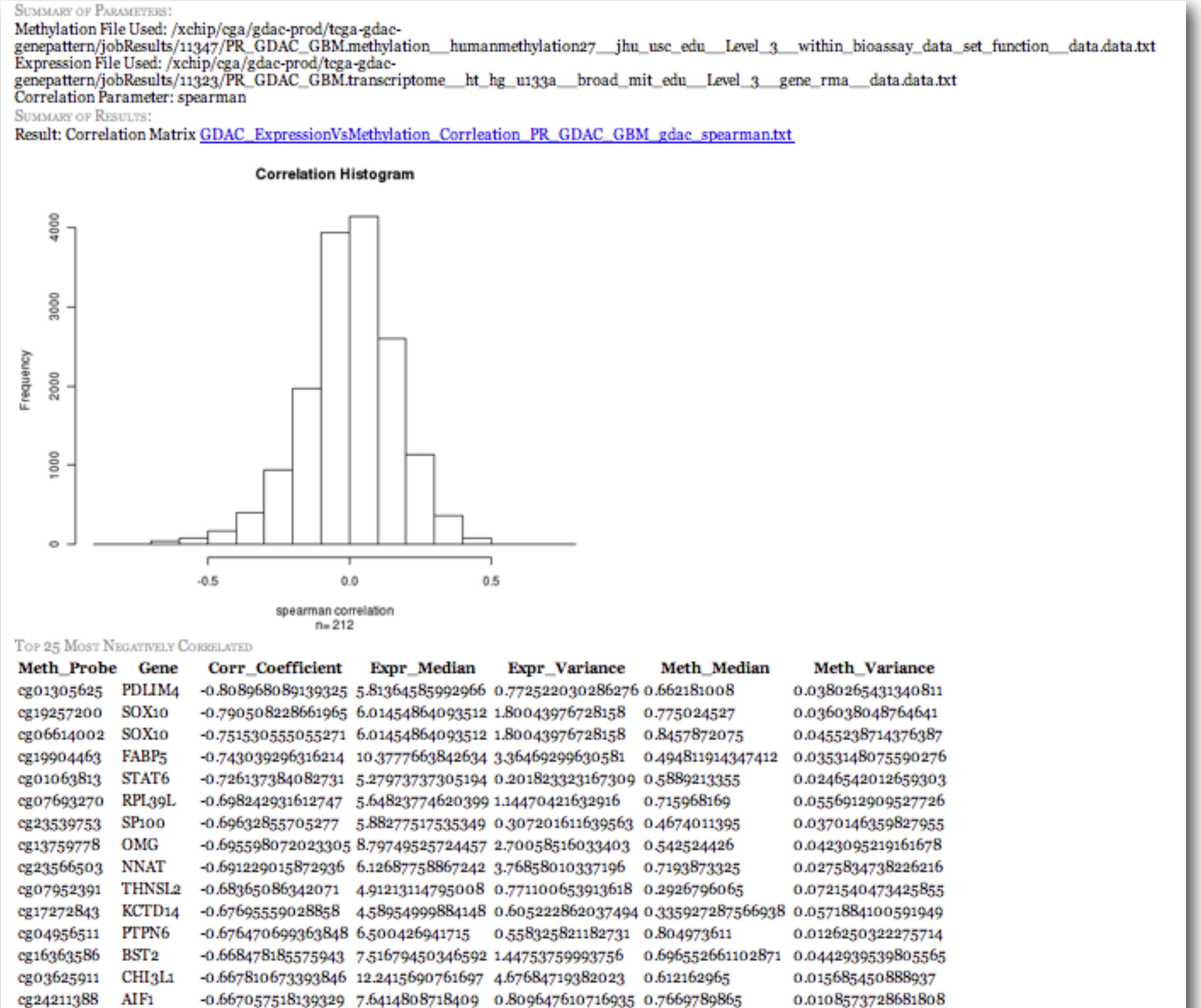
#For marker genes of subtype 1 :								
Pathway	Genes in Pathway	a	b	c	d	Enrichment	p-value	q-value
REACTOME_SIGNALING_IN_IMMUNE_SYSTEM	366	71	256	1805	18821	3.63576965558162	1.44358894655583e-12	3.62508113540178e-10
REACTOME_INNATE_IMMUNITY_SIGNALING	136	27	79	1849	18998	5.29437340620899	4.53406417368573e-07	5.69287765119833e-05
REACTOME_DOWNSTREAM_EVENTS_IN_GPCR_SIGNALING	448	71	373	1805	18704	2.90284553095604	1.08431632688308e-06	9.07630174265814e-05
KEGG_MAPK_SIGNALING_PATHWAY	267	48	217	1828	18860	3.8292872194421	1.88108903107025e-06	0.000118092833434567
BIOCARTA_MAPK_PATHWAY	87	22	65	1854	19012	5.9284138267989	5.83250309449335e-06	0.000242562455658246
KEGG_CALCIUM_SIGNALING_PATHWAY	178	35	141	1841	18936	4.54898716720033	5.48109738793413e-06	0.000242562455658246
REACTOME_SIGNALLING_BY_NGF	215	40	175	1836	18902	4.18041159556751	6.76156827202188e-06	0.000242562455658246

For a complete list of enrichment of pathways

2). FOR GENE THAT HAVE HIGH CORRELATION BETWEEN COPY NUMBER ALTERATIONS AND EXPRESSION:

Pathway	Genes in Pathway	a	b	c	d	Enrichment	p-value	q-value
REACTOME_GENE_EXPRESSION	425	41	372	715	19825	2.29110420203264	5.20061423211076e-09	3.64563057
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	138	18	110	738	20087	4.41626974452321	8.0815641689962e-07	0.00028325
KEGG_SPLICEOSOME	126	16	97	740	20100	4.55741031394396	2.9876328480329e-06	0.00069811
REACTOME_MRNA_SPLICING	107	14	83	742	20114	4.82084429332839	9.81304525180060e-06	0.00171973
KEGG_GLYOXYLATE_AND_DICARBOXYLATE_METABOLISM	16	6	10	750	20187	8.93983170058022	1.27018768202829e-05	0.00173265
REACTOME_FORMATION_AND_MATURATION_OF_MRNA_TRANSCRIPT	152	17	125	739	20072	4.04465740816114	1.48301354102396e-05	0.00173265
KEGG_ENDOCYTOSIS	183	19	159	737	20038	3.61136344405221	2.48424514354276e-05	0.00248779
REACTOME_HIV_INFECTION	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.65595949581063e-	[illegible]

Pre-Nozzle Reports



SUMMARY OF PARAMETERS:

Methylation File Used: /xchip/cga/gdac-prod/tcga-gdac-genepattern/jobResults/11347/PR_GDAC_GBM.methylation__humanmethylation27__jhu_usc_edu__Level_3__within_bioassay_data_set_function__data.data.txt
Expression File Used: /xchip/cga/gdac-prod/tcga-gdac-genepattern/jobResults/11323/PR_GDAC_GBM.transcriptome__ht_hg_u133a__broad_mit_edu__Level_3__gene_rma__data.data.txt
Correlation Parameter: spearman

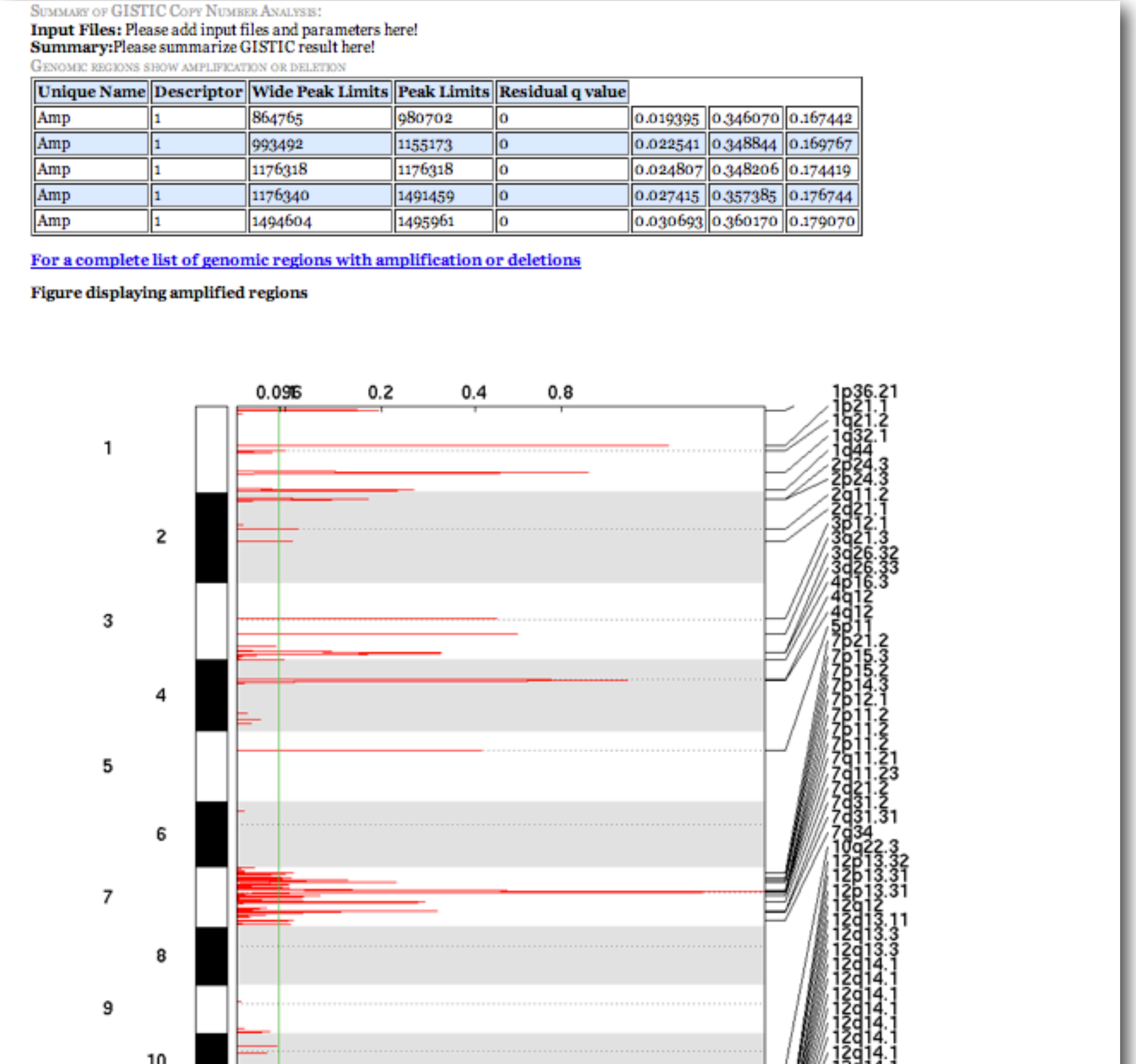
SUMMARY OF RESULTS:

Result: Correlation Matrix GDAC_ExpressionVsMethylation_Corrleation_PR_GDAC_GBM_gdac_spearman.txt

TOP 25 MOST NEGATIVELY CORRELATED

Meth_Probe	Gene	Corr_Coefficient	Expr_Median	Expr_Variance	Meth_Median	Meth_Variance
cg01305625	PDLIM4	-0.808968089139325	5.81364585992966	0.772522030286276	0.662181008	0.0380265431340811
cg19257200	SOX10	-0.790508228661965	6.01454864093512	1.80043976728158	0.775024527	0.036038048764641
cg06614002	SOX10	-0.751530555055271	6.01454864093512	1.80043976728158	0.8457872075	0.0455238714376387
cg19904463	FABP5	-0.743039296316214	10.3777663842634	3.3646929963058	0.494811914347412	0.0353148075590276
cg01063813	STAT6	-0.726137384082731	5.27973737305194	0.201823323167309	0.5889213355	0.0246542012659303
cg07693270	RPL39L	-0.698242931612747	5.64823774620399	1.14470421632916	0.715968169	0.0556912909527726
cg23539753	SP100	-0.69632855705277	5.88277517535349	0.307201611639563	0.4674011395	0.0370146359827955
cg13759778	OMG	-0.695598072023305	8.79749525724457	2.70058516033403	0.542524426	0.0423095219161678
cg23566503	NNAT	-0.691229015872936	6.12687758867242	3.76858010337196	0.7193873325	0.0275834738226216
cg07952391	THNSL2	-0.68365086342071	4.91213114795008	0.771100653913618	0.2926796065	0.0721540473425855
cg17272843	KCTD14	-0.67695559028858	4.58954999884148	0.605222862037494	0.335927287566938	0.0571884100591949
cg04956511	PTPN6	-0.676470699363848	6.500426941715	0.558325821182731	0.804973611	0.0126250322275714
cg16363586	BST2	-0.668478185575943	7.51679450346592	1.44753759993756	0.696552661102871	0.0442939539805565
cg03625911	CHI3L1	-0.667810673393846	12.2415690761697	4.67684719382023	0.612162965	0.015685450888937
cg24211388	AIF1	-0.667057518139329	7.6414808718409	0.809647610716935	0.7669789865	0.0108573728681808

Pre-Nozzle Reports



SUMMARY OF GISTIC COPY NUMBER ANALYSIS:

Input Files: Please add input files and parameters here!

Summary: Please summarize GISTIC result here!

GENOMIC REGIONS SHOW AMPLIFICATION OR DELETION

Unique Name	Descriptor	Wide Peak Limits	Peak Limits	Residual q value			
Amp	1	864765	980702	0	0.019395	0.346070	0.167442
Amp	1	993492	1155173	0	0.022541	0.348844	0.169767
Amp	1	1176318	1176318	0	0.024807	0.348206	0.174419
Amp	1	1176340	1491459	0	0.027415	0.357385	0.176744
Amp	1	1494604	1495961	0	0.030693	0.360170	0.179070

For a complete list of genomic regions with amplification or deletions

Figure displaying amplified regions

Why are we doing this?

- **Organization is difficult to navigate.**
    - Because every report has a different structure.
    - Because every report has a different layout.
    - Because some reports contain a lot of detail.
- **Content is difficult to interpret.**
    - Because important explanations are missing.

What are the goals for Nozzle?

- All reports should have the same **structure**.
- All reports should have the same **layout**.
- All reports should have **advanced features** such as folding sections and subsections, zoomable figures, etc.
- All reports should be created with a **simple** set of instructions and without any knowledge of the technologies used to render the reports.
- **Developers focus on the content!**

Example: Text Formatting

without Nozzle (from an existing report):

```
write(paste("<p><b>Figure 1:</b> Consensus NMF clustering of ", genenumber, " variably expressed genes and ",
samplenum, " samples.</p>", sep=""), file="report.html", append=TRUE);
```

Example: Text Formatting

without Nozzle (from an existing report):

```
write(paste("<p><b>Figure 1:</b> Consensus NMF clustering of ", genenumber, " variably expressed genes and ",
samplenum, " samples.</p>", sep=""), file="report.html", append=TRUE);
```

with Nozzle:

```
p <- newParagraph( asStrong( "Figure 1: "), "Consensus NMF clustering of ", genenumber, " variably expressed genes
and ", samplenum, " samples." );

# insert p into a section, ..., write report
```

Example: Figure

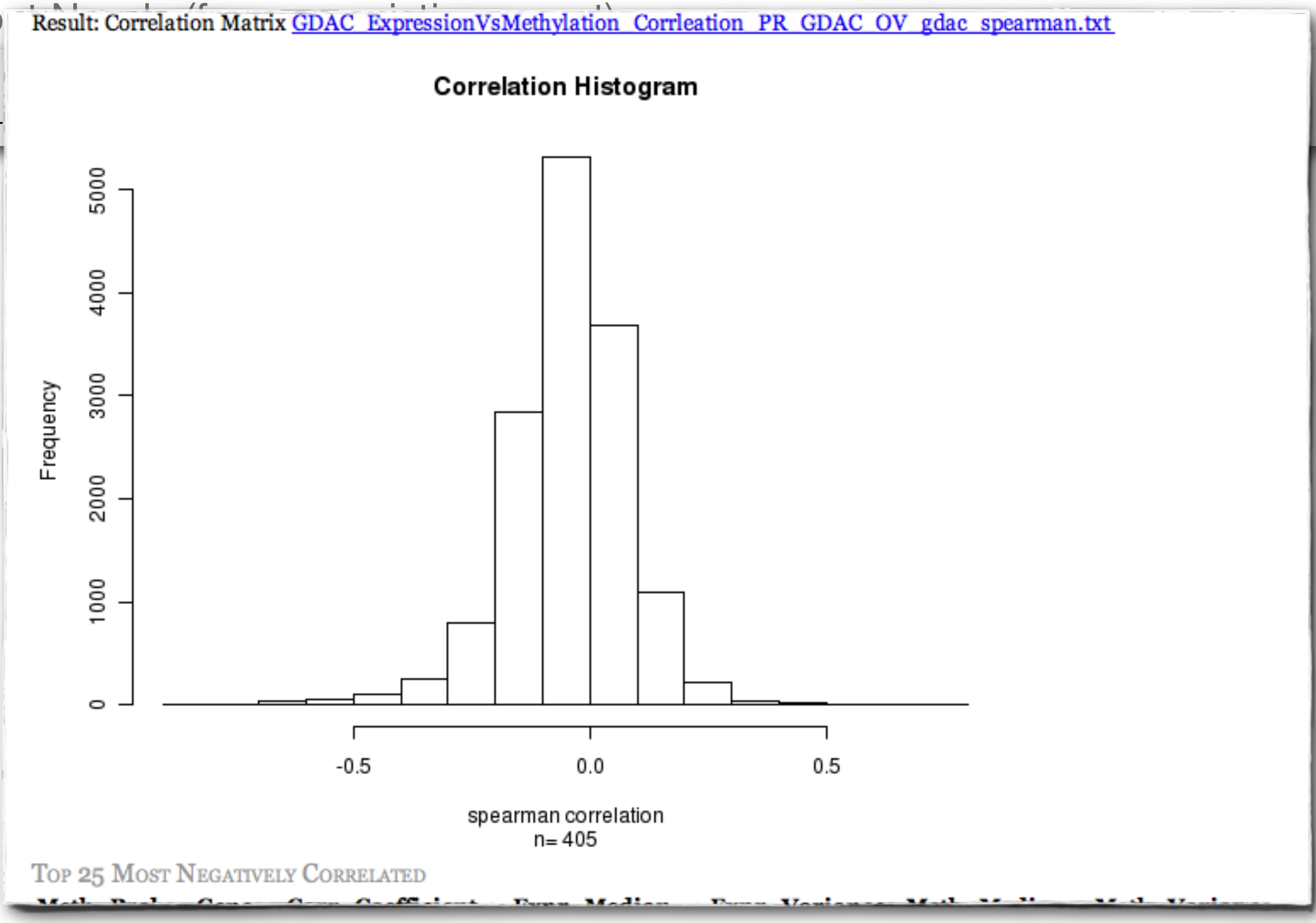
without Nozzle (from an existing report):

```r
write( paste( "<img src=", plot_file, " alt=", "Histogram of Correlation Values", "/><br>", sep=" "),
    file="report.html", append=TRUE);
```

Example: Figure

witho

```
write(                                                                                    "),
    fi
```



Result: Correlation Matrix GDAC_ExpressionVsMethylation_Corrleation_PR_GDAC_OV_gdac_spearman.txt

TOP 25 MOST NEGATIVELY CORRELATED

Example: Figure

without Nozzle (from an existing report):

```
write( paste( "<img src=", plot_file, " alt=", "Histogram of Correlation Values", "/><br>", sep=" "),
     file="report.html", append=TRUE);
```

with Nozzle:

```
imageFile <- "files/histogram.png";
imageFileHighRes <- "files/histogram.pdf"; # could also be a high-res pixel-based image file

f <- newFigure( imageFile, imgeFileHighRes, "Histogram of Correlation Values. ..." );

# insert figure into a section, ..., write report
```

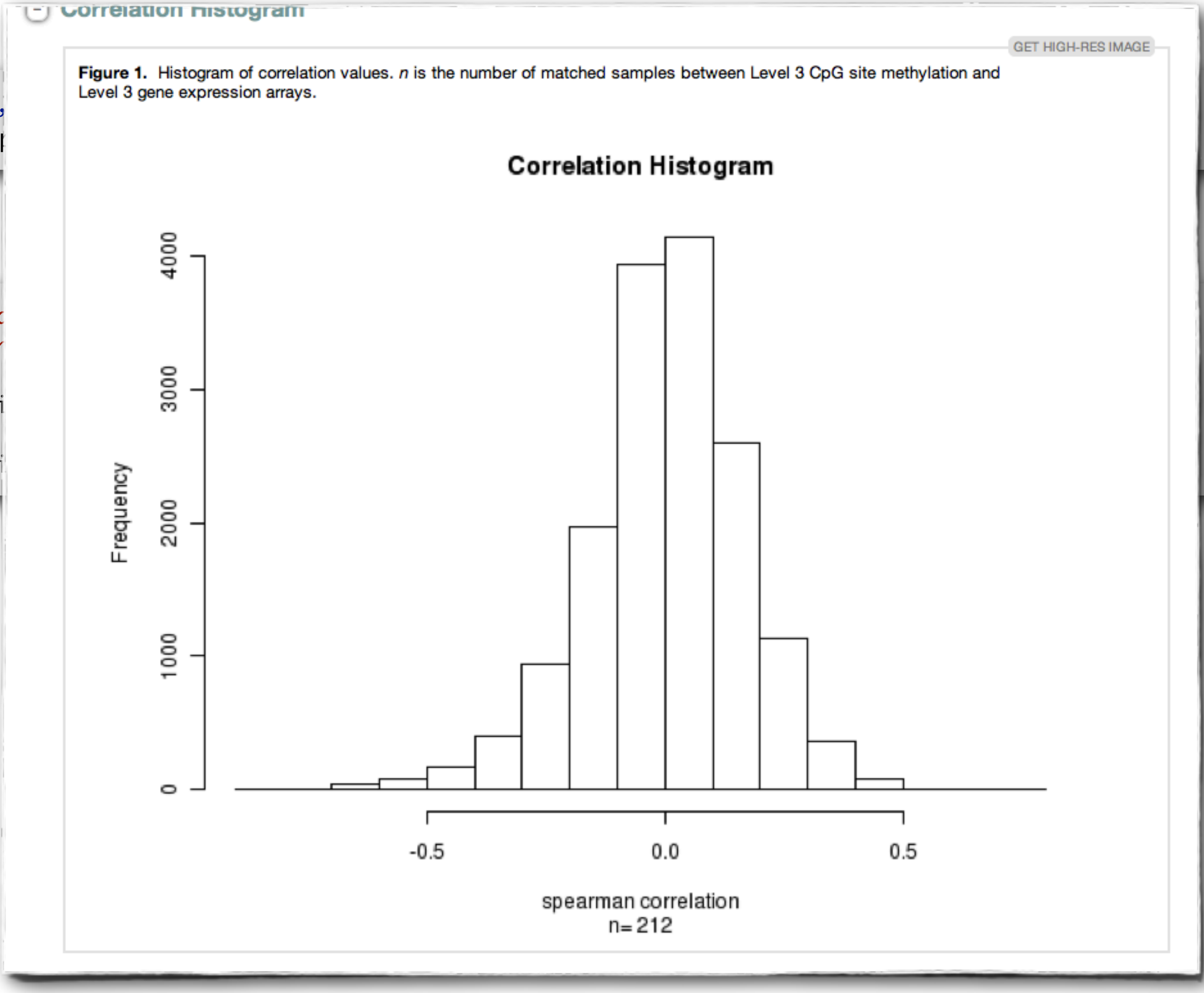
Example: Figure

without Nozzle (from

```
write( paste( "<img src=",
     file="report.html", ap
```

with Nozzle:

```
imageFile <- "files/histogra
imageFileHighRes <- "files/h

f <- newFigure( imageFile, i

# insert figure into a secti
```



Correlation Histogram

Figure 1. Histogram of correlation values. *n* is the number of matched samples between Level 3 CpG site methylation and Level 3 gene expression arrays.

Example: Figure

without Nozzle (from

```
write( paste( "<img src=",
     file="report.html", ap
```

with Nozzle:

```
imageFile <- "files/histogra
imageFileHighRes <- "files/h

f <- newFigure( imageFile, i

# insert figure into a secti
```

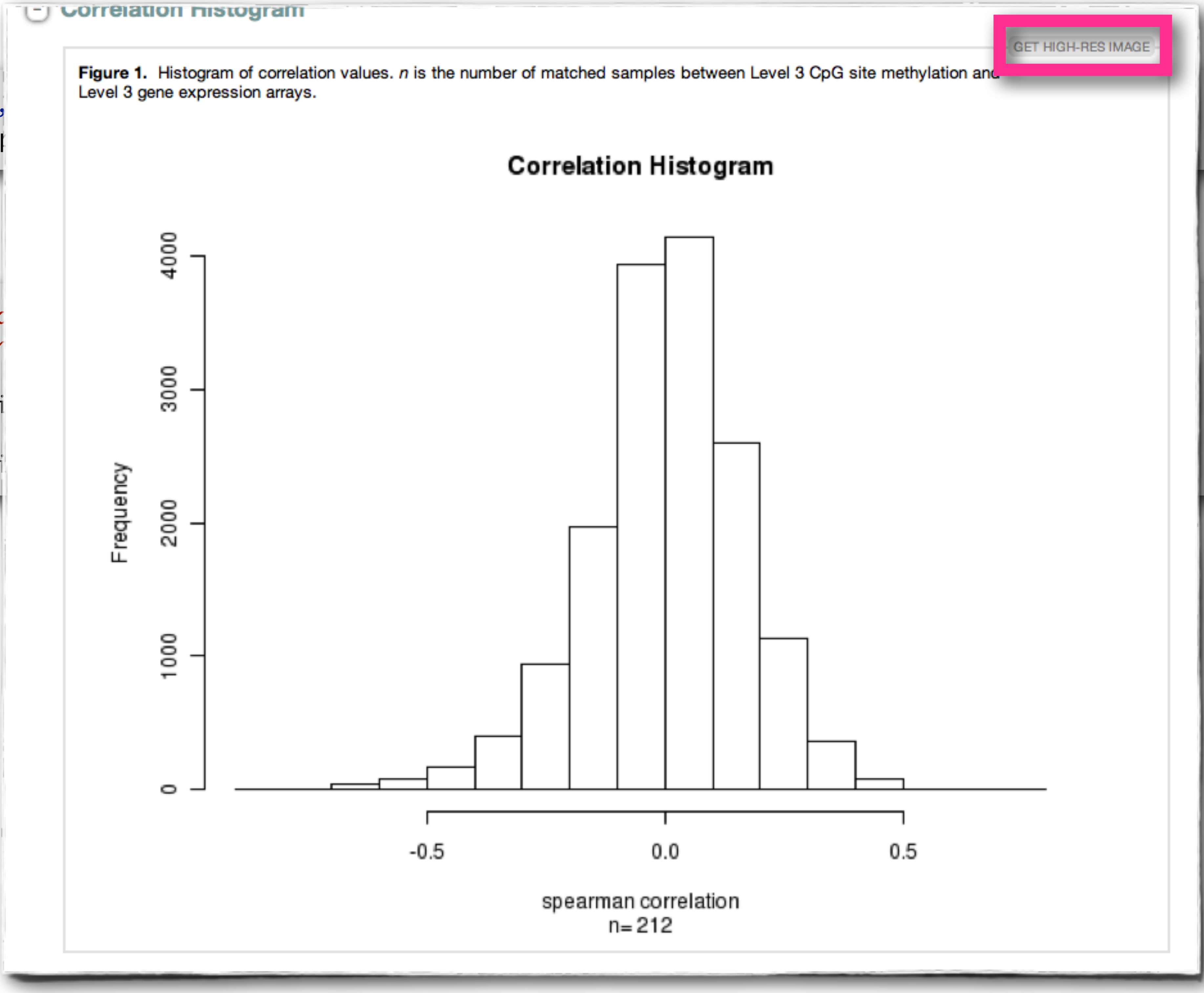


Figure 1. Histogram of correlation values. *n* is the number of matched samples between Level 3 CpG site methylation and Level 3 gene expression arrays.

Example: Table

without Nozzle (from an existing report):

```
top = "<h3> Top 25 Most Negatively Correlated</h3><table><tr>";

title = dimnames(ov_final)[[2]];
for (t1 in 1:length(title)){
    top = paste(top, "<th>", title[t1], "</th>", sep="");
}
top = paste(top,"</tr>", sep="");

### TOP 25 ###
for (a in 1:25){
    top = paste(top, "<tr>", sep="");
    for (b in 1:ncol(ov_final)){
        top = paste(top, "<td>", ov_final[a,b], "</td>", sep="");
    }
    top = paste(top, "</tr>", sep="");
}
top = paste(top,"</table><br>", sep="");
write(top, file="report.html", append=TRUE);
```

Example: Table

without Nozzle (from an existing report):

```
top = "<h3> Top 25 Most Negatively Correlated</h3><table><tr>";

title = dimnames(ov_final)[[2]];
for (t1 in 1:length(title)){
    top = paste(top, "<th>", title[t1], "</th>", sep="");
}
top = paste(top,"</tr>", sep="");

### TOP 25 ###
for (a in 1:25){
    top = paste(top, "<tr>", sep="");
    for (b in 1:ncol(ov_final)){
        top = paste(top, "<td>", ov_final[a,b], "</td>", sep="");
    }
    top = paste(top, "</tr>", sep="");
}
top = paste(top,"</table><br>", sep="");
write(top, file="report.html", append=TRUE);
```

with Nozzle:

```
corFile <- "files/correlations_gdac_spearman.txt";
corTable <- read.table( corFile, sep="\t", nrow=25, stringsAsFactors=FALSE, header=TRUE );

t <- newTable( corTable, corFile,"Top 25 most negatively correlated methylation probe/gene pairs. ..." );

# insert table into a section, ..., write report
```

Example: Tables (cont'd)

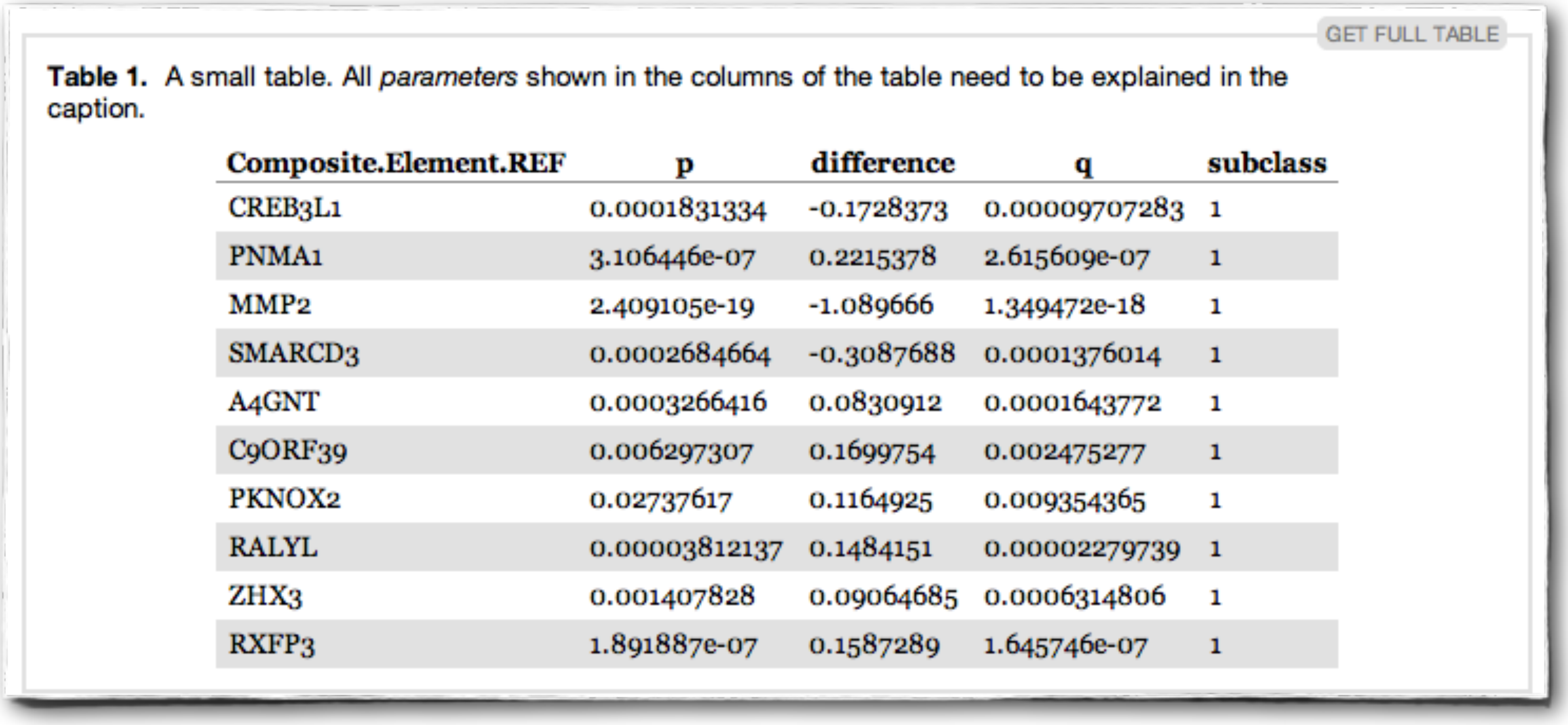
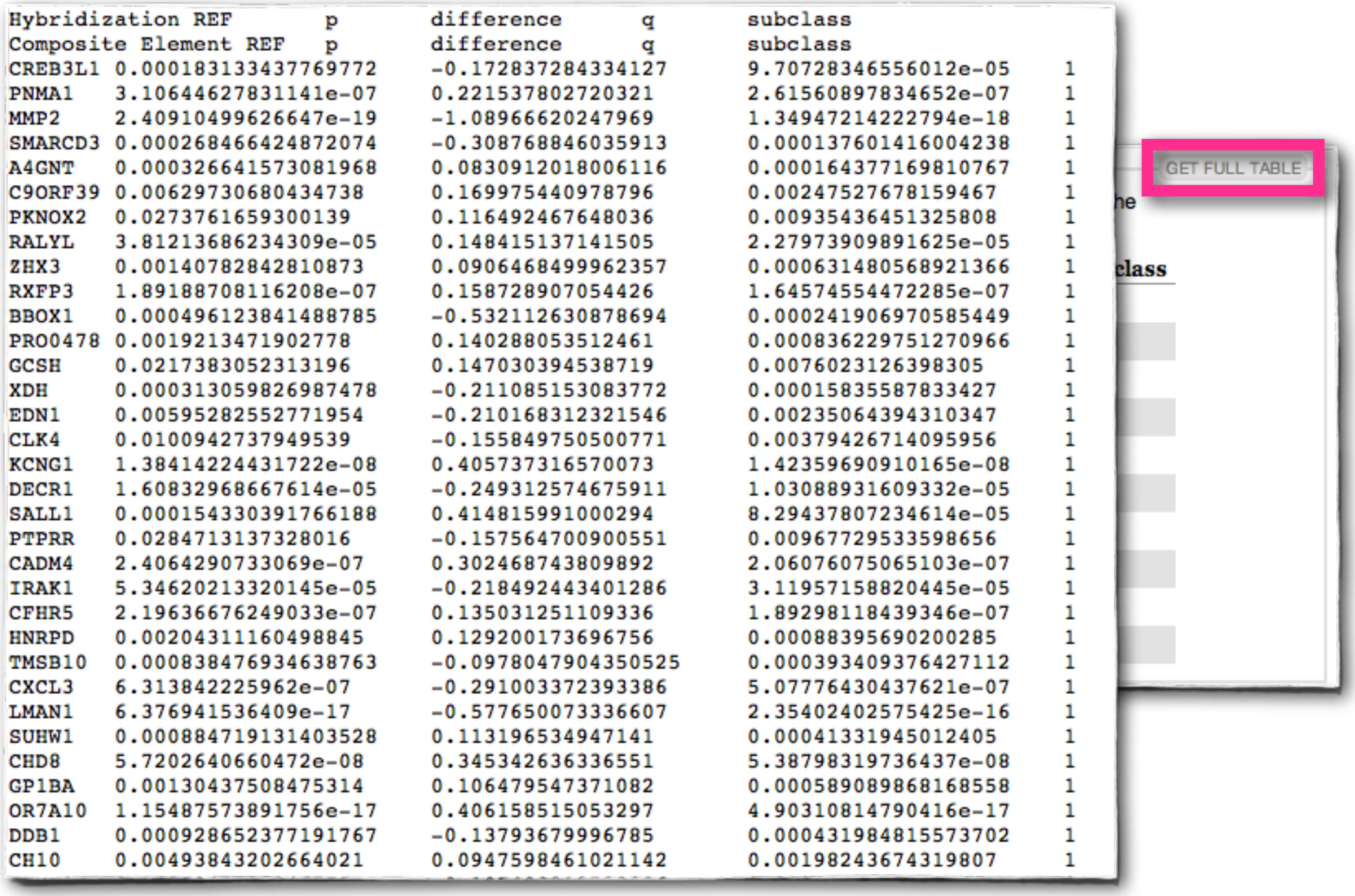


Table 1. A small table. All *parameters* shown in the columns of the table need to be explained in the caption.

Composite.Element.REF	p	difference	q	subclass
CREB3L1	0.0001831334	-0.1728373	0.00009707283	1
PNMA1	3.106446e-07	0.2215378	2.615609e-07	1
MMP2	2.409105e-19	-1.089666	1.349472e-18	1
SMARCD3	0.0002684664	-0.3087688	0.0001376014	1
A4GNT	0.0003266416	0.0830912	0.0001643772	1
C9ORF39	0.006297307	0.1699754	0.002475277	1
PKNOX2	0.02737617	0.1164925	0.009354365	1
RALYL	0.00003812137	0.1484151	0.00002279739	1
ZHX3	0.001407828	0.09064685	0.0006314806	1
RXFP3	1.891887e-07	0.1587289	1.645746e-07	1

Example: Tables (cont’d)



```
Hybridization REF       p       difference      q       subclass
Composite Element REF   p       difference      q       subclass
CREB3L1 0.000183133437769772    -0.172837284334127      9.70728346556012e-05    1
PNMA1   3.10644627831141e-07    0.221537802720321       2.61560897834652e-07    1
MMP2    2.40910499626647e-19    -1.08966620247969       1.34947214222794e-18    1
SMARCD3 0.000268466424872074    -0.308768846035913      0.000137601416004238    1
A4GNT   0.000326641573081968    0.0830912018006116      0.000164377169810767    1
C9ORF39 0.00629730680434738     0.169975440978796       0.00247527678159467     1
PKNOX2  0.0273761659300139      0.116492467648036       0.00935436451325808     1
RALYL   3.81213686234309e-05    0.148415137141505       2.27973909891625e-05    1
ZHX3    0.00140782842810873     0.0906468499962357      0.000631480568921366    1
RXFP3   1.89188708116208e-07    0.158728907054426       1.64574554472285e-07    1
BBOX1   0.000496123841488785    -0.532112630878694      0.000241906970585449    1
PRO0478 0.0019213471902778      0.140288053512461       0.000836229751270966    1
GCSH    0.0217383052313196      0.147030394538719       0.0076023126398305      1
XDH     0.000313059826987478    -0.211085153083772      0.00015835587833427     1
EDN1    0.00595282552771954     -0.210168312321546      0.00235064394310347     1
CLK4    0.0100942737949539      -0.155849750500771      0.00379426714095956     1
KCNG1   1.38414224431722e-08    0.405737316570073       1.42359690910165e-08    1
DECR1   1.60832968667614e-05    -0.249312574675911      1.03088931609332e-05    1
SALL1   0.000154330391766188    0.414815991000294       8.29437807234614e-05    1
PTPRR   0.0284713137328016      -0.157564700900551      0.00967729533598656     1
CADM4   2.4064290733069e-07     0.302468743809892       2.06076075065103e-07    1
IRAK1   5.34620213320145e-05    -0.218492443401286      3.11957158820445e-05    1
CFHR5   2.19636676249033e-07    0.135031251109336       1.89298118439346e-07    1
HNRPD   0.00204311160498845     0.129200173696756       0.00088395690200285     1
TMSB10  0.000838476934638763    -0.0978047904350525     0.000393409376427112    1
CXCL3   6.313842225962e-07      -0.291003372393386      5.07776430437621e-07    1
LMAN1   6.376941536409e-17      -0.577650073336607      2.35402402575425e-16    1
SUHW1   0.000884719131403528    0.113196534947141       0.00041331945012405     1
CHD8    5.7202640660472e-08     0.345342636336551       5.38798319736437e-08    1
GP1BA   0.00130437508475314     0.106479547371082       0.000589089868168558    1
OR7A10  1.15487573891756e-17    0.406158515053297       4.90310814790416e-17    1
DDB1    0.000928652377191767    -0.13793679996785       0.000431984815573702    1
CH10    0.00493843202664021     0.0947598461021142      0.00198243674319807     1
```

Example: Tables (cont'd)

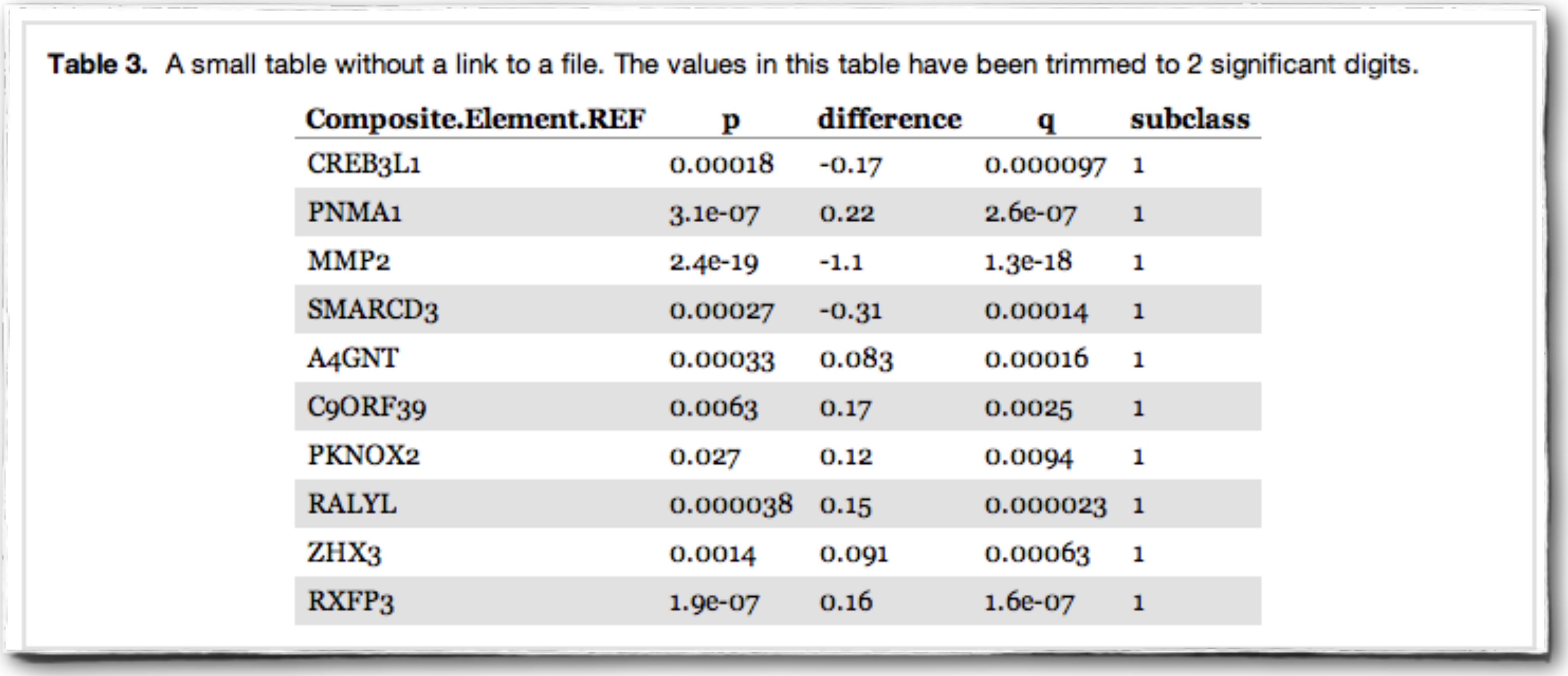


Table 3. A small table without a link to a file. The values in this table have been trimmed to 2 significant digits.

Composite.Element.REF	p	difference	q	subclass
CREB3L1	0.00018	-0.17	0.000097	1
PNMA1	3.1e-07	0.22	2.6e-07	1
MMP2	2.4e-19	-1.1	1.3e-18	1
SMARCD3	0.00027	-0.31	0.00014	1
A4GNT	0.00033	0.083	0.00016	1
C9ORF39	0.0063	0.17	0.0025	1
PKNOX2	0.027	0.12	0.0094	1
RALYL	0.000038	0.15	0.000023	1
ZHX3	0.0014	0.091	0.00063	1
RXFP3	1.9e-07	0.16	1.6e-07	1

values trimmed to 2 significant digits (default)

Example: Tables (cont'd)

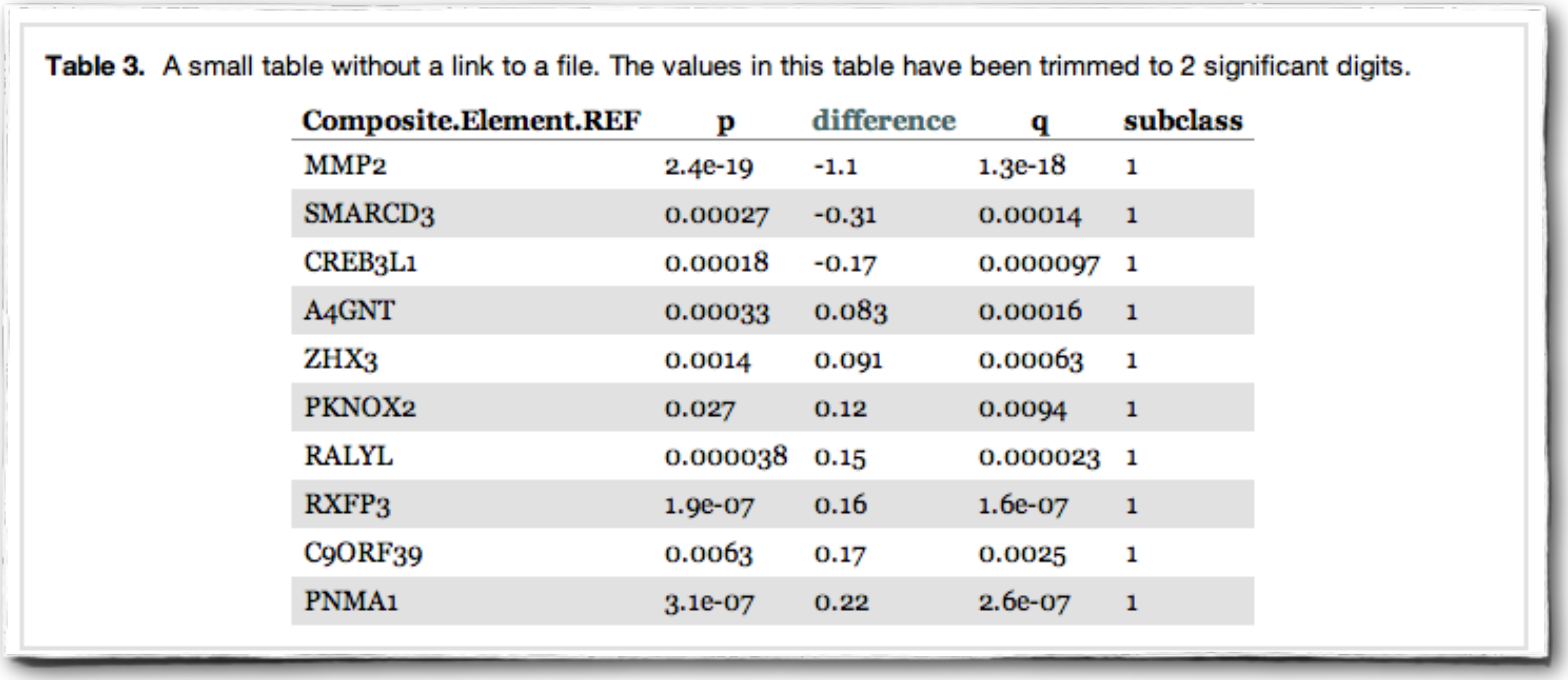


Table 3. A small table without a link to a file. The values in this table have been trimmed to 2 significant digits.

Composite.Element.REF	p	difference	q	subclass
MMP2	2.4e-19	-1.1	1.3e-18	1
SMARCD3	0.00027	-0.31	0.00014	1
CREB3L1	0.00018	-0.17	0.000097	1
A4GNT	0.00033	0.083	0.00016	1
ZHX3	0.0014	0.091	0.00063	1
PKNOX2	0.027	0.12	0.0094	1
RALYL	0.000038	0.15	0.000023	1
RXFP3	1.9e-07	0.16	1.6e-07	1
C9ORF39	0.0063	0.17	0.0025	1
PNMA1	3.1e-07	0.22	2.6e-07	1

rows sorted by "difference" column

Example: Tables (cont'd)

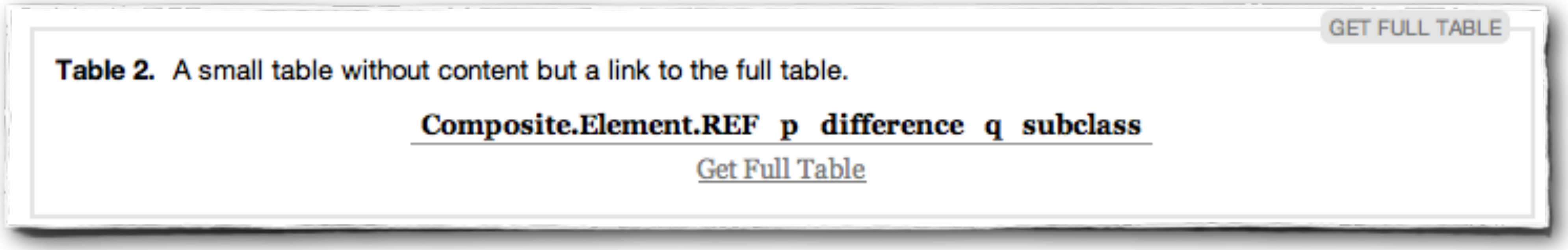


Table 2. A small table without content but a link to the full table.

Composite.Element.REF	p	difference	q	subclass

Get Full Table

link to full table only for long tables
caption can be used to describe columns

Example: Table (cont’d)

newTable(**table**, ..., *file=NA*, *significantDigits=2*, *protection*=PROTECTION.PUBLIC)

This function creates a Table element.

Arguments:

- **table** - A matrix or data.frame R data structure containing the table that should be included in the report. The table header will be created from the column names of the table data structure. The number of lines should be limited to no more than 30 unless there are good reasons to include more.
- ... - Zero or more strings and scalar variables that will be concatenated and used as caption for the table. This may also contain references and Result elements.
- *file* - Path and filename to a file containing the full table. Ideally, this should be a tab-delimited text file with a single line header.
- significantDigits - Number of significant digits for trimming of numeric columns in the table.
- *protection* - Protection level for the new element.

Value:

A Table element.

Example: References

```
simpleCitation <- newCitation( authors="Nils Gehlenborg", title="The Nozzle Report Package", year="2011",
url="http://www.gehlenborg.com" );

webCitation <- newCitation( title="The Cancer Genome Atlas Website", url="http://tcga.cancer.gov" );

fullCitation <- newCitation( authors="Nils Gehlenborg", title="Nozzle: An R Package for Report Generation",
publication="Yet Another Bioinformatics Journal", issue="14", number="44", pages="783-789", year="2011",
url="http://wwww.wikipedia.org" );

r <- addToReferences( r, simpleCitation, webCitation, fullCitation );
```

References

[1] Nils Gehlenborg, The Nozzle Report Package (2011)

[2] The Cancer Genome Atlas Website

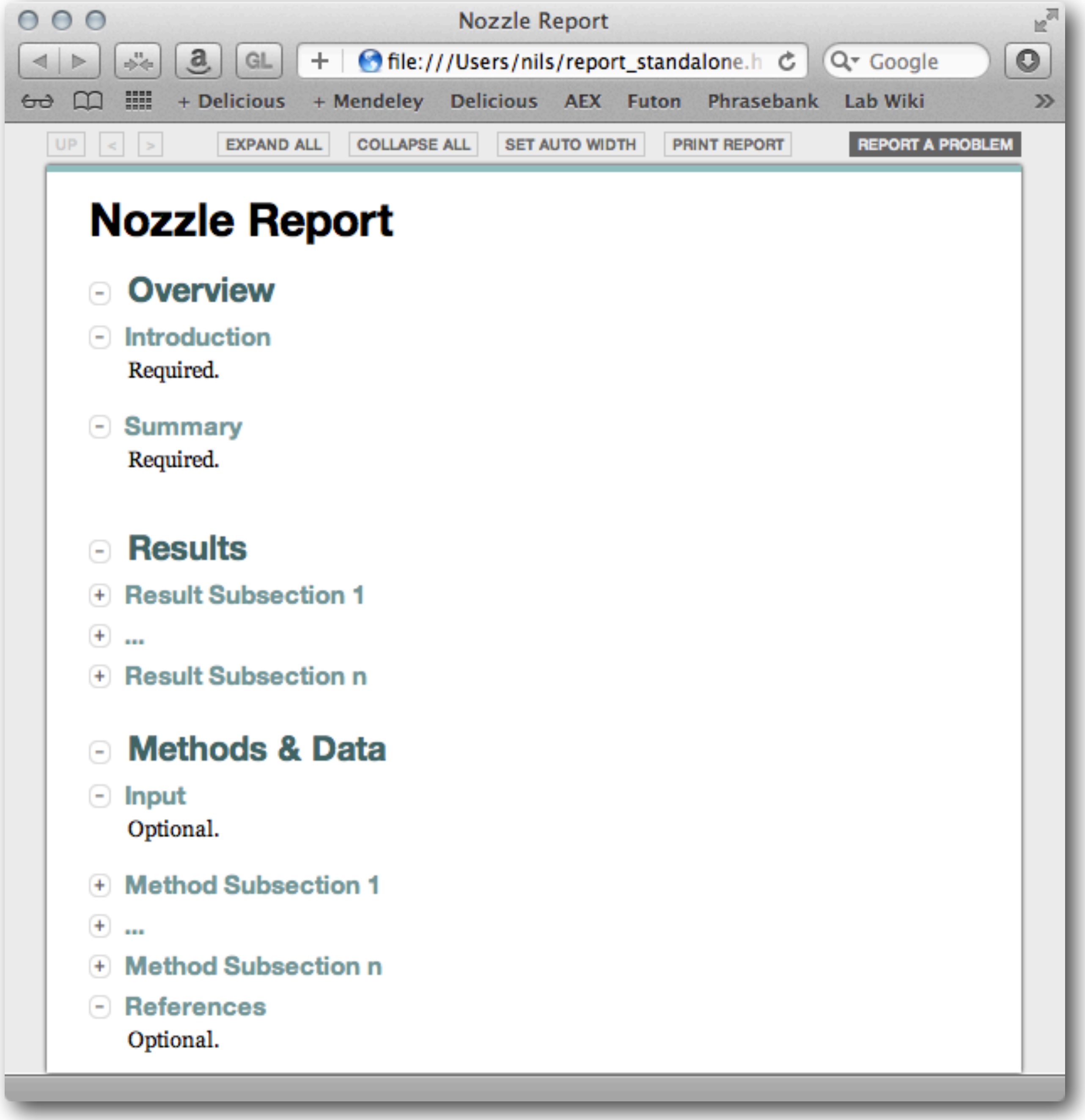
[3] Nils Gehlenborg, Nozzle: An R Package for Report Generation, *Yet Another Bioinformatics Journal* **14**(44):783-789 (2011)

Structure of a Nozzle Report

```
r <- newReport( "Nozzle Report" );

r <- addToIntroduction( r, newParagraph( "Required." ) );
r <- addToSummary( r, newParagraph( "Required." ) );

r <- addToResults( r, newSubSection( "Result Subsection 1" ) );
r <- addToResults( r, newSubSection( "..." ) );
r <- addToResults( r, newSubSection( "Result Subsection n" ) );

r <- addToMethods( r, newSubSection( "Method Subsection 1" ) );
r <- addToMethods( r, newSubSection( "..." ) );
r <- addToMethods( r, newSubSection( "Method Subsection n" ) );

r <- addToInput( r, newParagraph( "Optional." ) );
r <- addToReferences( r, newParagraph( "Optional." ) );

writeReport( r );
```

Structure of a Nozzle Report

```
r <- newReport( "Nozzle Report" );

r <- addToIntroduction( r, newParagraph( "Required." ) );
r <- addToSummary( r, newParagraph( "Required." ) );

r <- addToResults( r, newSubSection( "Result Subsection 1" ) );
r <- addToResults( r, newSubSection( "..." ) );
r <- addToResults( r, newSubSection( "Result Subsection n" ) );

r <- addToMethods( r, newSubSection( "Method Subsection 1" ) );
r <- addToMethods( r, newSubSection( "..." ) );
r <- addToMethods( r, newSubSection( "Method Subsection n" ) );

r <- addToInput( r, newParagraph( "Optional." ) );
r <- addToReferences( r, newParagraph( "Optional." ) );

writeReport( r );
```



Nozzle Report

⊖ Overview

⊖ Introduction

Required.

⊖ Summary

Required.

⊖ Results

⊕ Result Subsection 1

⊕ ...

⊕ Result Subsection n

⊖ Methods & Data

⊖ Input

Optional.

⊕ Method Subsection 1

⊕ ...

⊕ Method Subsection n

⊖ References

Optional.

Report Generation with Nozzle

Generate **Assemble** **Write**

- Conceptually a three stage process:
  1. **Generate** the elements that make up the report.
     a. *Content elements* (paragraphs, figures, tables, etc.)
     b. *Structural elements* (sections, lists, etc.)
  2. **Assemble** the elements into the structure of the report.
  3. **Write** the report to file.
- In practice it will be easier to combine parts of stage 1 and 2.

API Overview: Generation Stage

- **Structural elements**:
  - newReport, newSection, newSubSection, newSubSubSection, newList, newResult
- **Content elements**:
  - newParagraph, newTable, newFigure, newCitation, newParameterList
- **Content formatting**:
  - asLink, asStrong, asEmphasis, asParameter, asValue, asFilename, asCode
  - asReference, asSummary

Example

```
paragraph <- newParagraph( "Nozzle is an R package for report generation in ", asStrong( “Firehose” ) );
subsection <- newSubSection( "About the Package" );
report <- newReport( "The Nozzle Report" );
```

API Overview: Assembly Stage

- Reports are assembled in a bottom-up approach: generate content elements first and then attach them to structural elements and so on.

- General function to attach an element to a parent element:

```
addTo( parent, ..., row=NA, column=NA )
```

- Specialized functions to attach to predefined report elements:

```
addToResults( report, ... )
addToMethods( report, ... )
addToIntroduction( report, ... )
addToSummary( report, ... )
addToInput( report, ... )
addToReferences( report, ... )
```

Example (cont’d)

```
paragraph <- newParagraph( "Nozzle is an R package for report generation in ", asStrong( “Firehose” ) );
subsection <- newSubSection( "About the Package" );
report <- newReport( "The Nozzle Report" );
```

```
subsection <- addTo( subsection, paragraph );
report <- addToMethods( report, subsection );
```

API Overview: Writing Stage

writeReport(**report**, *filename*=DEFAULT.REPORT.FILENAME, *debug*=FALSE, *level*=PROTECTION.PUBLIC)

This function writes the report to a file as an HTML fragment for further processing by Firehose. Before the report is written, the optional, predefined “Methods & Data” subsections “Input” and “References” are removed from the report if no elements have been attached to them.

Arguments:

- **report** - The report element.
- *filename* - The filename for the report. The default is “report.html”.
- *debug* - If set to TRUE, a “stand-alone” version of the report will be generated that will work outside Firehose. This is for testing purposes only and this argument must be set to FALSE in production code.
- *level* - Determines which elements should be included in the report. If set to PROTECTION.PUBLIC, only elements marked with PROTECTION.PUBLIC will be included. If set to PROTECTION.TCGA, elements marked with PROTECTION.PUBLIC and PROTECTION.TCGA will be included. If set to PROTECTION.PRIVATE all elements will be included.

Value:

None.

Example (cont'd)

```
paragraph <- newParagraph( "Nozzle is an R package for report generation in ", asStrong( “Firehose” ) );
subsection <- newSubSection( "About the Package" );
report <- newReport( "The Nozzle Report" );
```

```
subsection <- addTo( subsection, paragraph );
report <- addToMethods( report, subsection );
```

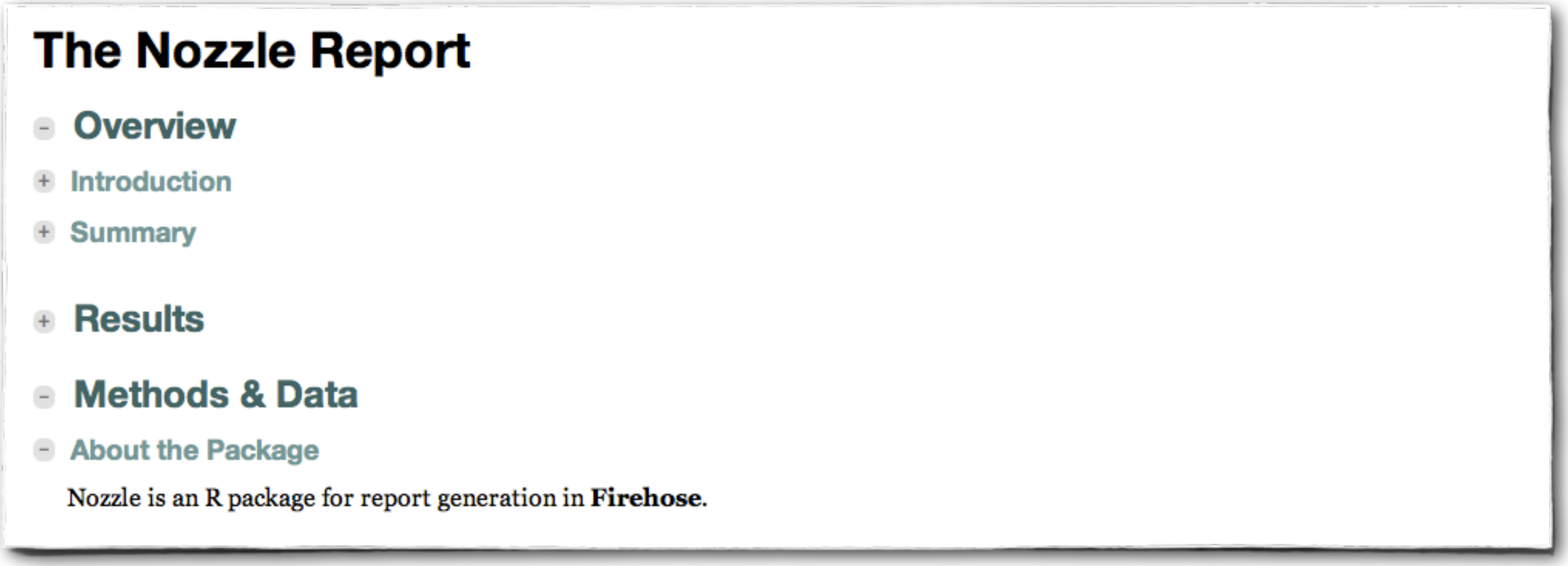
```
writeReport( report );
```

Example (cont’d)

```
paragraph <- newParagraph( "Nozzle is an R package for report generation in ", asStrong( “Firehose” ) );
subsection <- newSubSection( "About the Package" );
report <- newReport( "The Nozzle Report" );
```

```
subsection <- addTo( subsection, paragraph );
report <- addToMethods( report, subsection );
```

```
writeReport( report );
```



The Nozzle Report

Overview

Introduction

Summary

Results

Methods & Data

About the Package

Nozzle is an R package for report generation in **Firehose**.

Supplementary Results

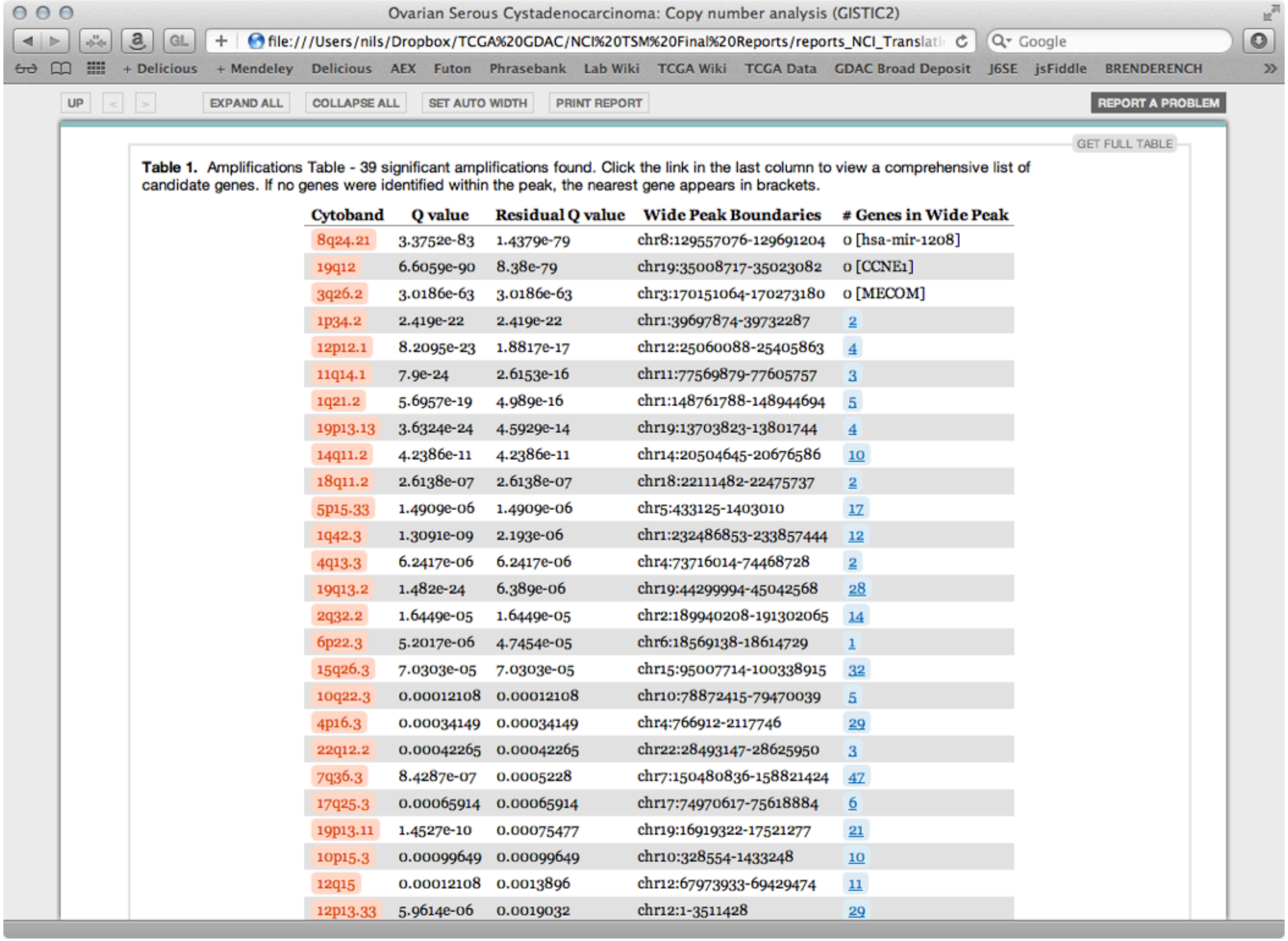
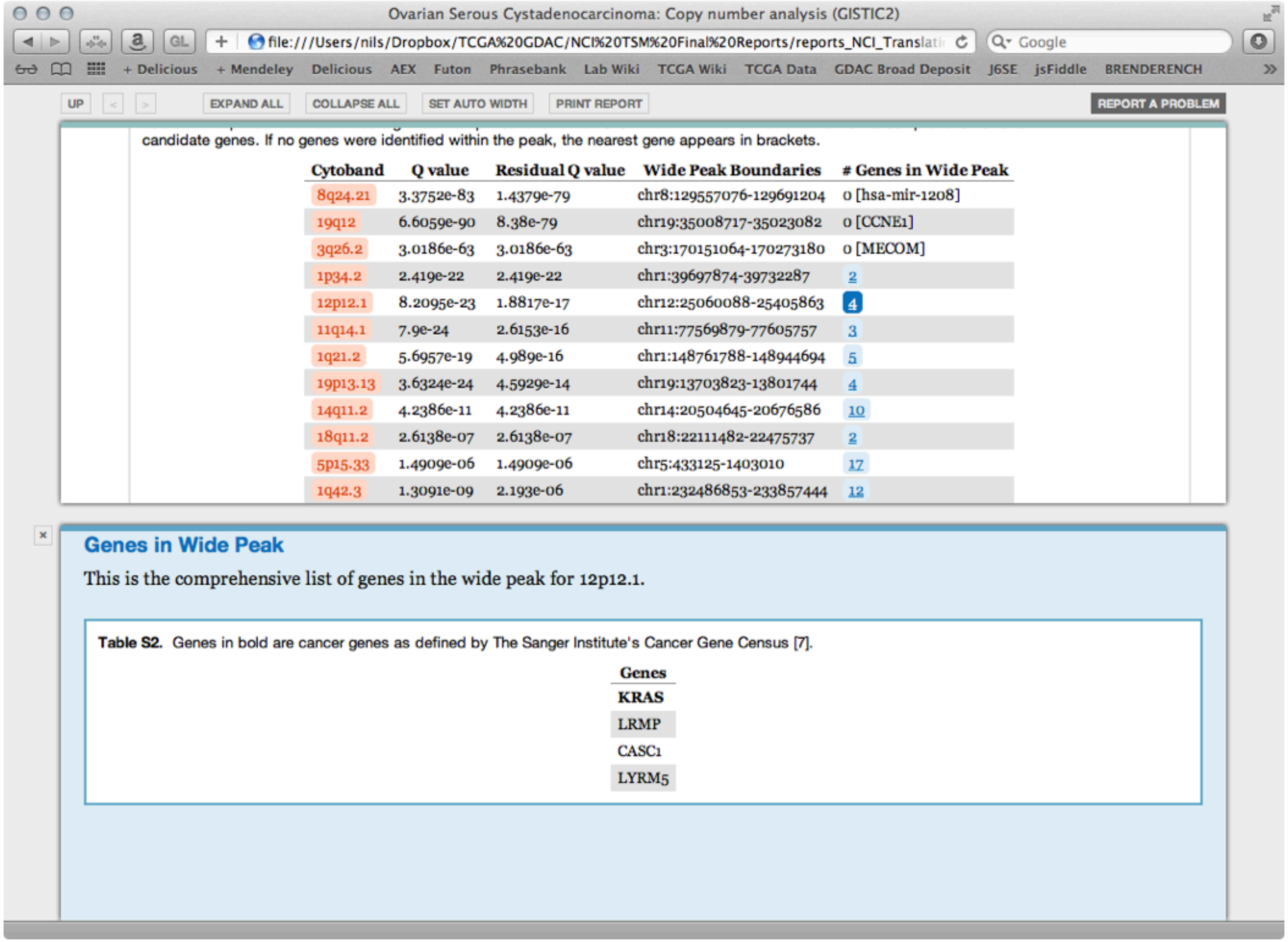


Table 1. Amplifications Table - 39 significant amplifications found. Click the link in the last column to view a comprehensive list of candidate genes. If no genes were identified within the peak, the nearest gene appears in brackets.

Cytoband	Q value	Residual Q value	Wide Peak Boundaries	# Genes in Wide Peak
8q24.21	3.3752e-83	1.4379e-79	chr8:129557076-129691204	0 [hsa-mir-1208]
19q12	6.6059e-90	8.38e-79	chr19:35008717-35023082	0 [CCNE1]
3q26.2	3.0186e-63	3.0186e-63	chr3:170151064-170273180	0 [MECOM]
1p34.2	2.419e-22	2.419e-22	chr1:39697874-39732287	2
12p12.1	8.2095e-23	1.8817e-17	chr12:25060088-25405863	4
11q14.1	7.9e-24	2.6153e-16	chr11:77569879-77605757	3
1q21.2	5.6957e-19	4.989e-16	chr1:148761788-148944694	5
19p13.13	3.6324e-24	4.5929e-14	chr19:13703823-13801744	4
14q11.2	4.2386e-11	4.2386e-11	chr14:20504645-20676586	10
18q11.2	2.6138e-07	2.6138e-07	chr18:22111482-22475737	2
5p15.33	1.4909e-06	1.4909e-06	chr5:433125-1403010	17
1q42.3	1.3091e-09	2.193e-06	chr1:232486853-233857444	12
4q13.3	6.2417e-06	6.2417e-06	chr4:73716014-74468728	2
19q13.2	1.482e-24	6.389e-06	chr19:44299994-45042568	28
2q32.2	1.6449e-05	1.6449e-05	chr2:189940208-191302065	14
6p22.3	5.2017e-06	4.7454e-05	chr6:18569138-18614729	1
15q26.3	7.0303e-05	7.0303e-05	chr15:95007714-100338915	32
10q22.3	0.00012108	0.00012108	chr10:78872415-79470039	5
4p16.3	0.00034149	0.00034149	chr4:766912-2117746	29
22q12.2	0.00042265	0.00042265	chr22:28493147-28625950	3
7q36.3	8.4287e-07	0.0005228	chr7:150480836-158821424	47
17q25.3	0.00065914	0.00065914	chr17:74970617-75618884	6
19p13.11	1.4527e-10	0.00075477	chr19:16919322-17521277	21
10p15.3	0.00099649	0.00099649	chr10:328554-1433248	10
12q15	0.00012108	0.0013896	chr12:67973933-69429474	11
12p13.33	5.9614e-06	0.0019032	chr12:1-3511428	29

Supplementary Results



candidate genes. If no genes were identified within the peak, the nearest gene appears in brackets.

Cytoband	Q value	Residual Q value	Wide Peak Boundaries	# Genes in Wide Peak
8q24.21	3.3752e-83	1.4379e-79	chr8:129557076-129691204	0 [hsa-mir-1208]
19q12	6.6059e-90	8.38e-79	chr19:35008717-35023082	0 [CCNE1]
3q26.2	3.0186e-63	3.0186e-63	chr3:170151064-170273180	0 [MECOM]
1p34.2	2.419e-22	2.419e-22	chr1:39697874-39732287	2
12p12.1	8.2095e-23	1.8817e-17	chr12:25060088-25405863	4
11q14.1	7.9e-24	2.6153e-16	chr11:77569879-77605757	3
1q21.2	5.6957e-19	4.989e-16	chr1:148761788-148944694	5
19p13.13	3.6324e-24	4.5929e-14	chr19:13703823-13801744	4
14q11.2	4.2386e-11	4.2386e-11	chr14:20504645-20676586	10
18q11.2	2.6138e-07	2.6138e-07	chr18:22111482-22475737	2
5p15.33	1.4909e-06	1.4909e-06	chr5:433125-1403010	17
1q42.3	1.3091e-09	2.193e-06	chr1:232486853-233857444	12

Genes in Wide Peak

This is the comprehensive list of genes in the wide peak for 12p12.1.

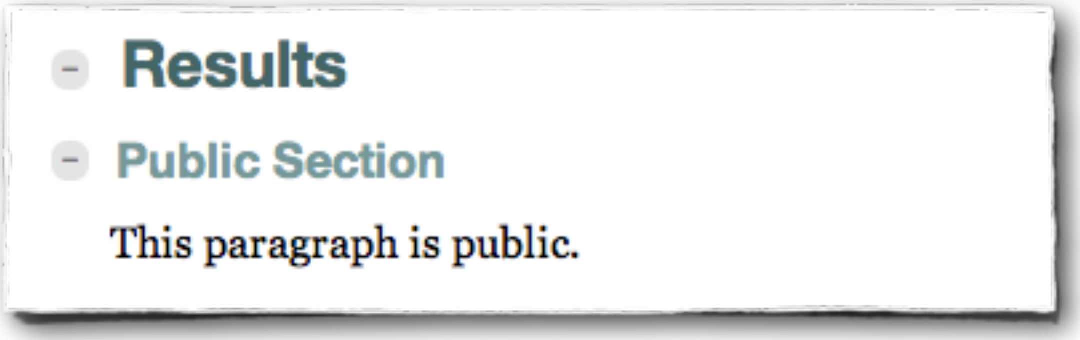
Table S2. Genes in bold are cancer genes as defined by The Sanger Institute's Cancer Gene Census [7].

Genes
KRAS
LRMP
CASC1
LYRM5

Protecting Report Elements for Privacy

```
p1 <- newParagraph( "This paragraph is public.", protection=PROTECTION.PUBLIC );
p2 <- newParagraph( "This paragraph is TCGA-only.", protection=PROTECTION.GROUP );
subsection <- newSubSection( "Public Section" );
subsection <- addTo( subsection, p1, p2 );

# write a report (fragment) for public consumption: p2 will be stripped from the report before it is written to file
writeReport( addToResults( newReport( "Report: Public" ), subsection ), level=PROTECTION.PUBLIC );

# write a report (fragment) for consumption by TCGA members: both p1 and p2 will be included in the report
writeReport( addToResults( newReport( "Report: TCGA" ), subsection ), level=PROTECTION.GROUP );
```

public report:

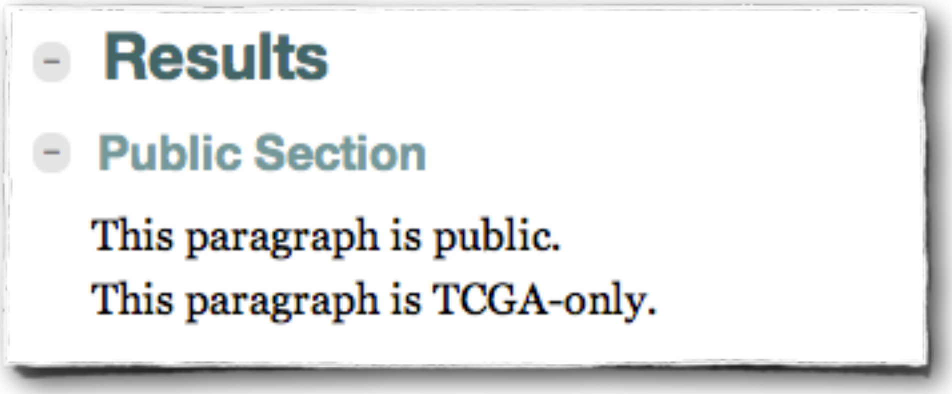


Results

Public Section

This paragraph is public.

group report:

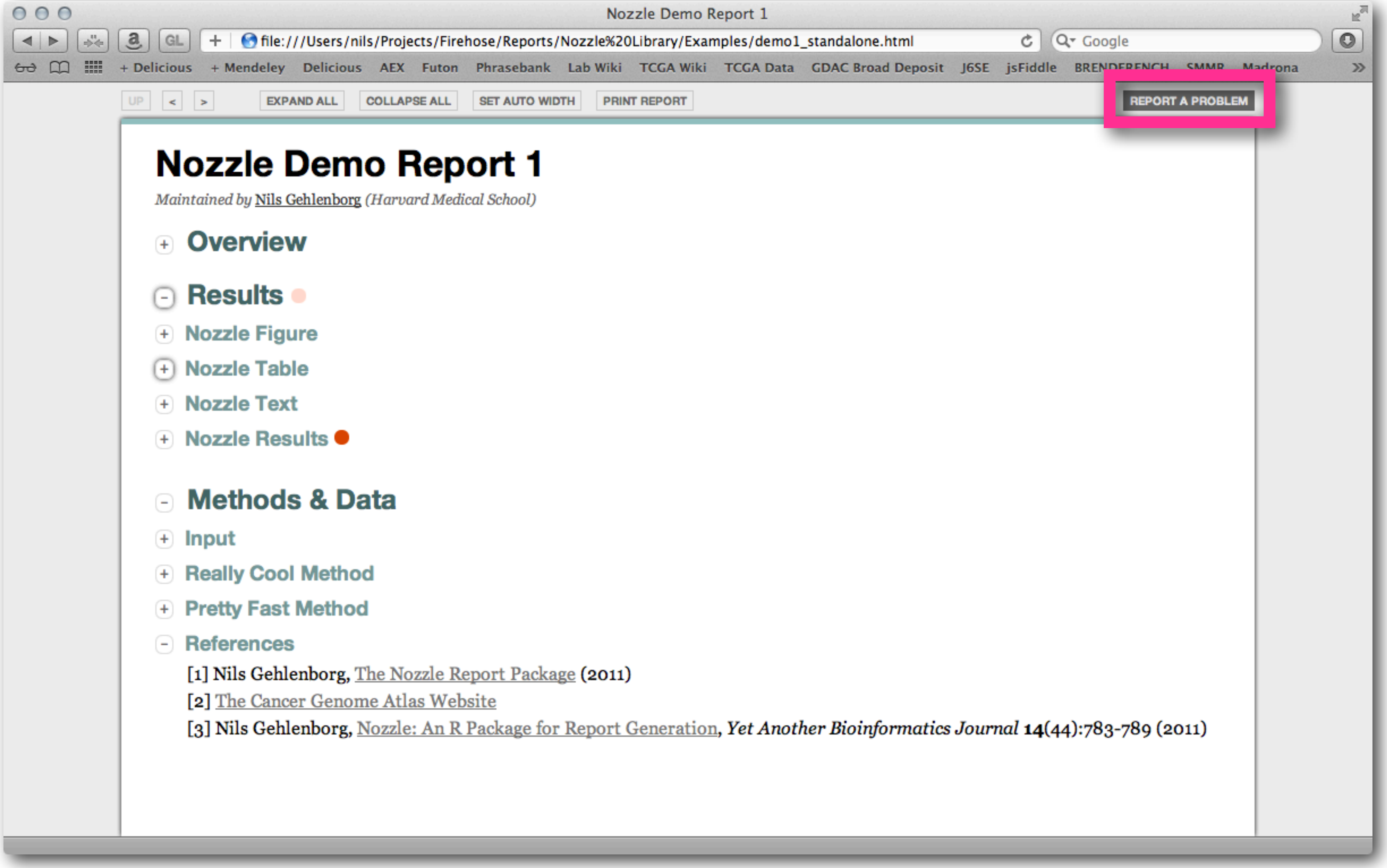


Results

Public Section

This paragraph is public.
This paragraph is TCGA-only.

Getting in Touch



Nozzle Demo Report 1

Maintained by Nils Gehlenborg (Harvard Medical School)

⊞ Overview

⊟ Results ●

⊞ Nozzle Figure

⊞ Nozzle Table

⊞ Nozzle Text

⊞ Nozzle Results ●

⊟ Methods & Data

⊞ Input

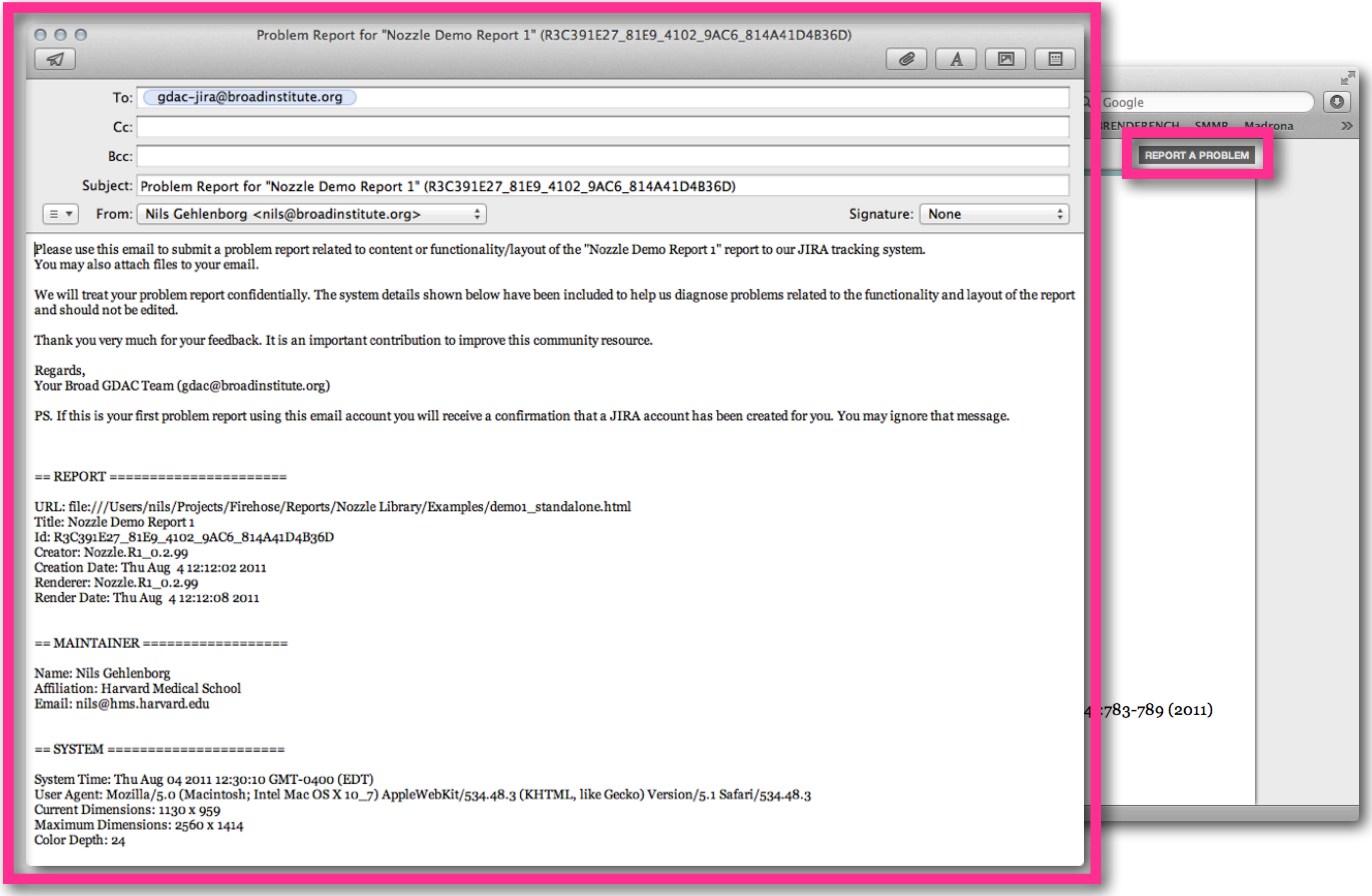
⊞ Really Cool Method

⊞ Pretty Fast Method

⊟ References

[1] Nils Gehlenborg, The Nozzle Report Package (2011)
[2] The Cancer Genome Atlas Website
[3] Nils Gehlenborg, Nozzle: An R Package for Report Generation, *Yet Another Bioinformatics Journal* **14**(44):783-789 (2011)

Getting in Touch



Problem Report for "Nozzle Demo Report 1" (R3C391E27_81E9_4102_9AC6_814A41D4B36D)

To: gdac-jira@broadinstitute.org
Cc:
Bcc:
Subject: Problem Report for "Nozzle Demo Report 1" (R3C391E27_81E9_4102_9AC6_814A41D4B36D)
From: Nils Gehlenborg <nils@broadinstitute.org>
Signature: None

Please use this email to submit a problem report related to content or functionality/layout of the "Nozzle Demo Report 1" report to our JIRA tracking system.
You may also attach files to your email.

We will treat your problem report confidentially. The system details shown below have been included to help us diagnose problems related to the functionality and layout of the report and should not be edited.

Thank you very much for your feedback. It is an important contribution to improve this community resource.

Regards,
Your Broad GDAC Team (gdac@broadinstitute.org)

PS. If this is your first problem report using this email account you will receive a confirmation that a JIRA account has been created for you. You may ignore that message.

```
== REPORT ======================

URL: file:///Users/nils/Projects/Firehose/Reports/Nozzle Library/Examples/demo1_standalone.html
Title: Nozzle Demo Report 1
Id: R3C391E27_81E9_4102_9AC6_814A41D4B36D
Creator: Nozzle.R1_0.2.99
Creation Date: Thu Aug  4 12:12:02 2011
Renderer: Nozzle.R1_0.2.99
Render Date: Thu Aug  4 12:12:08 2011


== MAINTAINER ==================

Name: Nils Gehlenborg
Affiliation: Harvard Medical School
Email: nils@hms.harvard.edu


== SYSTEM ======================

System Time: Thu Aug 04 2011 12:30:10 GMT-0400 (EDT)
User Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7) AppleWebKit/534.48.3 (KHTML, like Gecko) Version/5.1 Safari/534.48.3
Current Dimensions: 1130 x 959
Maximum Dimensions: 2560 x 1414
Color Depth: 24
```

Application of Nozzle in Firehose

- Only at the very end of a pipeline in a separate GenePattern module.
- Best practice is to write all tables and figures to file before the report module is run.

Developer-friendly deployment in Firehose

- Single installation of Nozzle for all our modules.
- Allows updates to Nozzle without module developers having to change their code.
- Only changes that result in backwards incompatibilities will require developers to act
    - This a rare event (has not happened so far despite many new features).

Behind the Scenes

- Every report element is a nested R “list” data structure and the report is assembled into a tree of nested lists (try `str( report )`).

- The writeReport(...) function preprocesses the report tree (e.g. it numbers figures, tables, etc. based on their position in the report and inserts the correct reference strings where requested) and then traverses the tree to produce the HTML code.

- The CSS and JavaScript code required to display a report are embedded in the HTML page, making it a permanent, self-contained record once it is written to file.

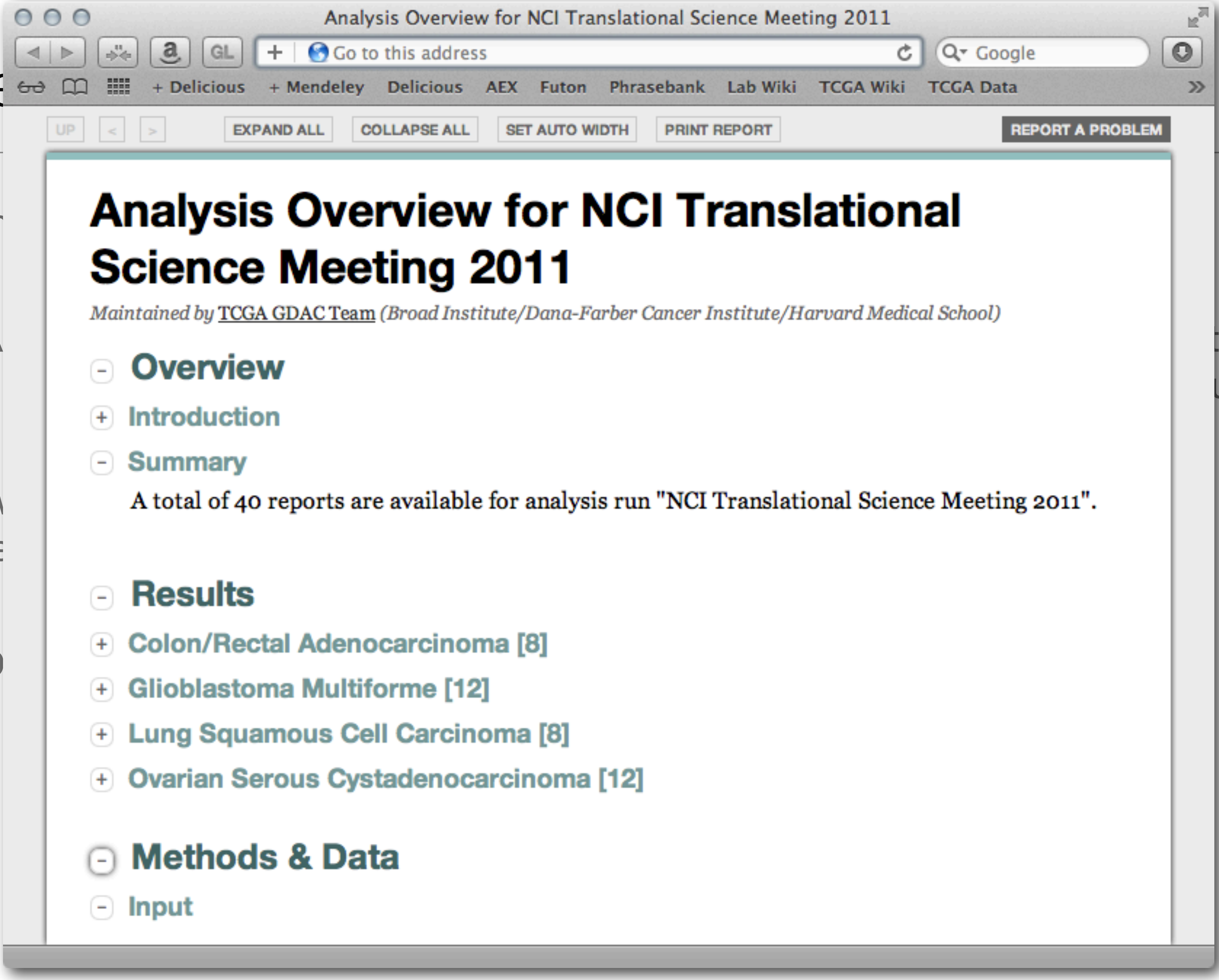
  - **Caveat:** Figures and other external files can’t be included in the HTML.

Behind the Scenes (cont’d)

- In Firehose we additionally save the report R data structure in an “Rdata” file.
- At the end of a Firehose run we copy the “Rdata” files from the job output directories into a new directory structure along with the associated images and files.
- We create an additional Nozzle report that contains links to the pipeline reports for all tumor types.
- During this stage we can apply updated formatting and introduce new features.

Be

- Ir
- A t
  d ure.
- W
  re
- D



Analysis Overview for NCI Translational Science Meeting 2011

Maintained by *TCGA GDAC Team* *(Broad Institute/Dana-Farber Cancer Institute/Harvard Medical School)*

⊖ Overview

⊕ Introduction

⊖ Summary

A total of 40 reports are available for analysis run "NCI Translational Science Meeting 2011".

⊖ Results

⊕ Colon/Rectal Adenocarcinoma [8]

⊕ Glioblastoma Multiforme [12]

⊕ Lung Squamous Cell Carcinoma [8]

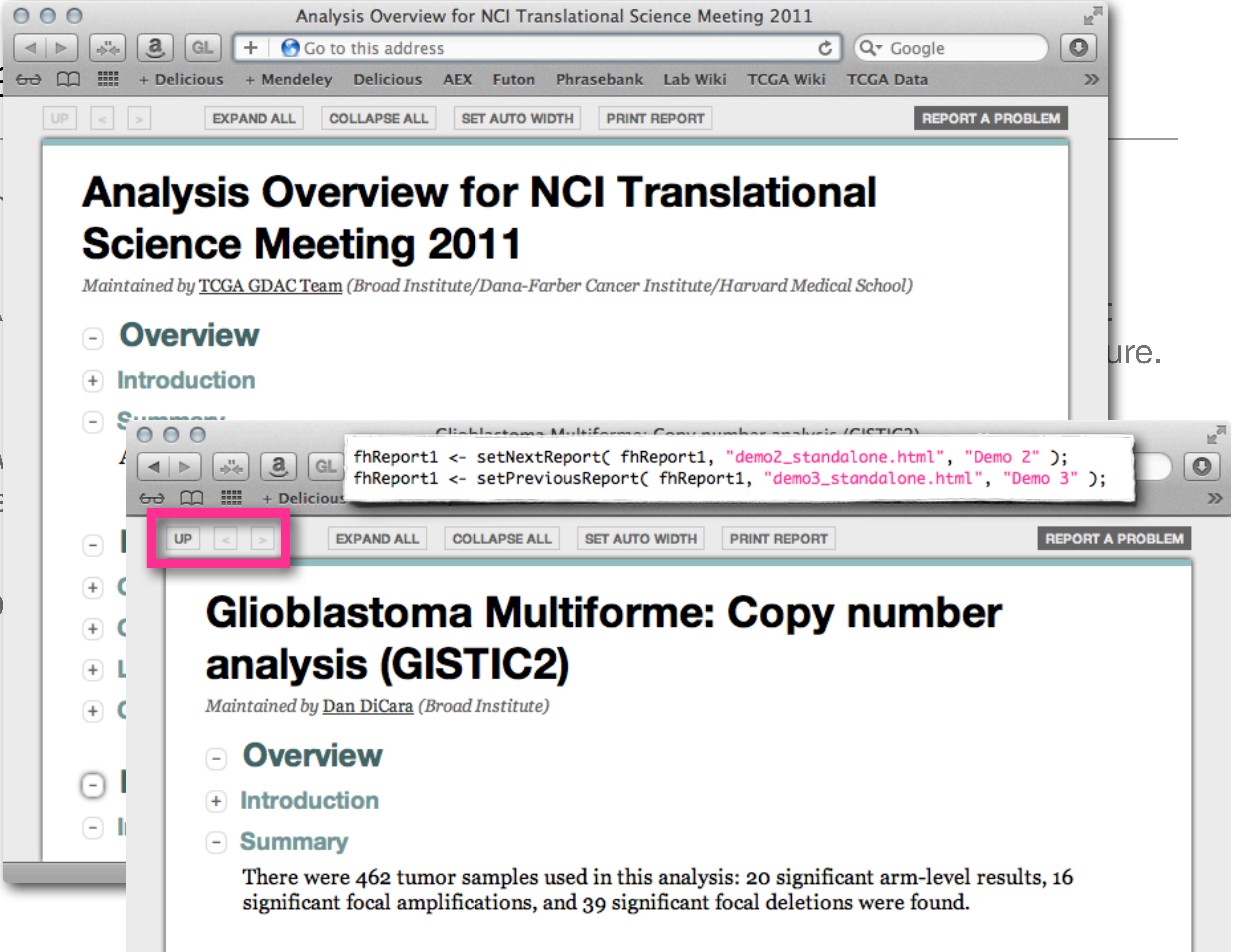
⊕ Ovarian Serous Cystadenocarcinoma [12]

⊖ Methods & Data

⊖ Input

Be

- In
- A
  d ure.
- W
  re
- D



Analysis Overview for NCI Translational Science Meeting 2011

Maintained by TCGA GDAC Team (Broad Institute/Dana-Farber Cancer Institute/Harvard Medical School)

Overview

Introduction

```
fhReport1 <- setNextReport( fhReport1, "demo2_standalone.html", "Demo 2" );
fhReport1 <- setPreviousReport( fhReport1, "demo3_standalone.html", "Demo 3" );
```

Glioblastoma Multiforme: Copy number analysis (GISTIC2)

Maintained by Dan DiCara (Broad Institute)

Overview

Introduction

Summary

There were 462 tumor samples used in this analysis: 20 significant arm-level results, 16 significant focal amplifications, and 39 significant focal deletions were found.

Finally ...

- Nozzle is not tied to Firehose, can be used for any pipeline or problem.
- We are very close to releasing a completely generalized version.

- Questions and comments to nils@hms.harvard.edu
- Download and documentation (slightly outdated):