

Supplementary File 7. Search Comparison

We compare here approaches using Transcriptomine, GEO, PubMed and Google to answering biological questions related to nuclear receptor signaling transcriptomics. Reasonable attempts have been made to obtain results with each method. We used the following approach.

1. Use several different alternative text search strategies for GEO, PubMed and Google when comparing results to Transcriptomine.
2. Evaluate search results (clicking on hits in the first two pages of search results).

Positive results are defined as data points in peer-reviewed studies.

1. Use Case 1

Transcriptomine

Query:

- Gene: Single Gene: Any
- Fold Change: 2-fold
- Direction: Up
- Significance: $p < 0.05$
- Regulatory Molecule: 1,25-Dihydroxyvitamin D3 (1,25-Dihydroxyvitamin D3); Any
- RNA Source: Cardiovascular, all cell lines, all tissues

Results: This query returns over 4,000 results, which is a large amount of data for a user unexperienced in bioinformatics to consume. To reduce the number of results, the user clicks on the “Modify Query” button, and adjusting the fold change cutoff to a higher value, or the significance cut-off to a more stringent confidence level. Modifying the query as follows:

- Gene: Single Gene: Any
- Fold Change: 2-fold
- Direction: Up
- Significance: $P < 0.0001$
- Regulatory Molecule: 1,25-Dihydroxyvitamin D3 (1,25-Dihydroxyvitamin D3); Any
- RNA Source: Cardiovascular, all cell lines, all tissues

This returns 1036 results (See Supplemental File S4) and Discussion in text.

Google

Query: Genes AND rapidly induced AND (estrogen OR estradiol) AND MCF-7

Results: Google returns 4,600,00 search results.

One positive hit was returned (1), but were all individual of already very well characterized 17β -estradiol genes. Other positives included older papers describe dated characterizations of individual target genes using older method (21835891). False positives included hits that describe actions of estradiol on MCF-7 cells unrelated to regulation of gene expression (17363451), regulation of genes in MCF-7 cells by other signaling molecules (8417822), and regulation of ER α by other molecules (15012603). Moreover, investigation of some hits requires journal subscription credentials, even though some are nearly 10 years old.

Query: Genes AND rapidly induced AND (estrogen OR estradiol) AND MCF-7 AND microarray

Results: Google returns 381,000 search results.

While the addition of the term “microarray” refined the search, none of the results on the first two results page corresponded to short-term treatment of MCF-7 cells

Query: Genes AND minutes AND (estrogen OR estradiol) AND MCF-7 AND microarray

Results: Google returns 381,000 search results.

The number of results was further refined and the first two pages included two true positive hits 17013392 AND 19103744. A third true positive (11888893) described induction of a single gene (STC2) but did not contain a gene list either in the main paper nor in supplementary file and did not reference a GEO deposition.

Substantial noise was introduced by the term “minutes”, due primarily to its appearance in experimental descriptions of protocols such as PCR, Northern, Western etc.

PubMed

Query: estradiol AND MCF-7 AND <1h

Results: 18 search results returned. None describe short term estradiol treatment of MCF-7 cells

Query: estradiol AND MCF-7 AND rapid

Results: 135 search results returned. Display Settings was adjusted to display full abstract text and pages searched for “rapid” to quickly identify positive results. True positives Results included rapid E2 dependent post-translational processing of cyclin E (22449977), rapid degradation of an estradiol metabolite (22152028), rapid ER-dependent phosphorylation of MAPK 20974640, rapid E2-dependent tubulin deacetylation (19318565), a rapid assay to study estrogenic endocrine disruptors (18980187) and other rapid events in MCF-7 cells related to estradiol or ERalpha unrelated to regulation of gene expression.

GEO Profiles

In order to obtain the same metadata instantaneously provided by Transcriptome, all GEO-based search strategies require a text-based query, drill-down on individual search results, scanning of descriptions to establish experimental conditions. Investigating each individual result further requires downloading the .cel files corresponding to that query, importing them into a spreadsheet, pairing sample and control values correctly, obtaining the ratio of sample vs control expression values to obtain fold changes, carrying out a t-test to obtain p-values and manually sorting cutting and pasting to generate custom lists with specific cutoffs.

Query: estradiol AND MCF-7 AND <1h

Results: 28 search results returned. None are from short term estradiol treatment of MCF-7 cells

Query: estradiol AND MCF-7 AND rapid

Results: No search results returned

Query: estradiol AND MCF-7 AND short-term

Results: No search results returned

Query: estradiol AND MCF-7 AND minutes

Results: No search results returned

2. Use Case 2

Transcriptome

Query:

- Gene: Single Gene: ACACA
- Fold Change: 2-fold
- Direction: Up
- Significance: Any
- Regulatory Molecule: Any
- RNA Source: Any

Results: 46 fold changes (**see Supplemental File S5**)

Google

Note: returning of results containing the word “acacia” contributed substantially to search result noise.

Query: ACACA AND PPAR

Results: No true positives on pages 1-3

Query: ACACA AND rosiglitazone

Results: 1 true positive (20154153).

Query: ACACA AND RORA

Results: No true positives on pages 1-3

Query: ACACA AND retinoid orphan receptor related

Results: No true positives on pages 1-3

GEO (Note: neither fold change nor significance of results was evaluated)

Query: ACACA AND rosiglitazone AND liver

Results: No results

Query: ACACA AND rosiglitazone AND adipose

Results: No results

Query: ACACA AND rosiglitazone AND fat

Results: No results

Query: ACACA AND rosiglitazone

Results: 6 results, no clear effect of rosiglitazone on ACACA expression was discernible.

Query: ACACA AND RORA

Results: No results

Query: ACACA AND retinoid orphan receptor-related

Results: No results

PUBMED

Query: ACACA AND PPAR

Results: No true positives

While the search returned no results in metabolic tissues, these papers describe induction by rosiglitazone of ACACA in bovine mammary epithelial cells.

Query: ACACA AND rosiglitazone

Results: 1 result, 1 true positive (19700688).

Query: ACACA AND RORA

Results: No result

Query: ACACA AND retinoid orphan receptor related

Results: 3 results, no true positives

Query: ACACA AND estradiol

Results: 2 results, no true positives

Query: ACACA AND dihydrotestosterone

Results: No results

Query: ACACA AND estrogen receptor

Results: 1 result, no true positives

Query: ACACA AND androgen receptor

Results: No results

Query: ACACA AND retinoic acid

Results: No results

Query: ACACA AND RAR

Results: No results

Query: ACACA AND vitamin D3

Results: No results

Query: ACACA AND vitamin D receptor

Results: No results

Use Case 3

Transcriptomine

- Gene: GO term: purine base biosynthetic process

- Fold Change: 2-fold
- Direction: Up or Down
- Significance: $p < 0.05$
- Regulatory Molecule: Any
- RNA Source: Any

This search returns 93 results (**See Supplementary Files S6**).

Google

Query: purine base biosynthetic process “nuclear receptor”

Results: No true positives on pages 1-3

Query: PRPS1 “nuclear receptor”

Results: No true positives on pages 1-3. One false positive result matched PRPS1 and Nuclear Receptor Binding Factor 2.

Query: PPAT “nuclear receptor”

Results: No true positives on pages 1-3. One false positive result wrongly gave PPAR γ the name PPAT.

Query: GART “nuclear receptor”

Results: No true positives on pages 1-3. One false positive result matched PRPS1 and Nuclear Receptor Binding Factor 2.

Query: PAICS “nuclear receptor”

Results: No true positives on pages 1-3

Query: GMPS “nuclear receptor”

Results: No true positives on pages 1-3. One false positive showed that GMPS was required for nuclear receptor target silencing in *Drosophila*.

Query: SHMT1 “nuclear receptor”

Results: No true positives on pages 1-3 One study cited repression of an SHMT gene by retinoic acid but was unable to determine if the gene was SHMT1 or SHMT2.

GEO (Note: neither fold change nor significance of results was evaluated)

Query: purine base biosynthetic process “nuclear receptor”

Results: No results

Query: PRPS1 “nuclear receptor”

Results: 19 results. Many showed a high degree of variance between samples, but not significance value or fold change is available. One shows upregulation by PCN, a PXR agonist and another shows downregulation in retinal cells overexpressing NR2E3.

Query: PPAT “nuclear receptor”

Results: 30 results. Many contained results for Coasy. One shows downregulation in retinal cells overexpressing NR2E3. Another shows downregulation in a HNF4 α KO, consistent with data retrieved from Transcriptomine.

Query: GART “nuclear receptor”

Results: 19 results. One shows downregulation in retinal cells overexpressing NR2E3.

Query: PAICS “nuclear receptor”

Results: 19 results. One shows downregulation in retinal cells overexpressing NR2E3.

Query: GMPS “nuclear receptor”

Results: 15 results. One shows downregulation in retinal cells overexpressing NR2E3; another shows upregulation by PCN in the liver.

Query: SHMT1 “nuclear receptor”

Results: 26 results. One shows downregulation in retinal cells overexpressing NR2E3.

PubMed

Query: "purine base biosynthesis" nuclear receptor

Results: No results

Query: PRPS1 “nuclear receptor”

Results: No results

Query: PPAT “nuclear receptor”

Results: 1 result – false positive (typo – PPATgamma for PPARgamma)

Query: GART “nuclear receptor”

Results: No results

Query: PAICS “nuclear receptor”

Results: No results

Query: GMPS “nuclear receptor”

Results: 1 result. One false positive showed that GMPS was required for nuclear receptor target silencing in *Drosophila*.

Query: SHMT1 “nuclear receptor”

Results: No results

Query: SHMT1 “estradiol”

Results: No results

Use Case 4

Google

Query: FOXG1 retinoic acid

Results: No results

Query: GAS1 retinoic acid

Results: One result: identified paper by Su & Gudas from which one of the Transcriptome microarrays was retrieved and which referred to *Gas1* induction in the abstract.

Query: GATA3

Results: One result identified regulation of GATA3 by retinoic acid.

Query: GREB1L

Results: No results

Query: NCOA3

Results: No results. Substantial noise from NCOA3 coactivation of RARs.

GEO (Note: neither fold change nor significance of results was evaluated)

Query: FOXG1 all trans retinoic acid

Results: 5 results.

Query: GAS1 retinoic acid

Results: 6 results, one shows induction of GAS1 in muscle cells.

Query: GATA3 retinoic acid

Results: 6 results, one shows induction of GATA3 in muscle cells.

Query: GREB1L retinoic acid

Results: 2 results showing induction of GREB1L by retinoic acid in leukemia cell lines.

Query: NCOA3 retinoic acid

Results: Multiple results showing induction by retinoic acid.

PubMed

Query: FOXG1 retinoic acid

Results: No results

Query: GAS1 retinoic acid

Results: One result: identified paper by Su & Gudas from which one of the Transcriptome microarrays was retrieved and which referred to *Gas1* induction in the abstract.

Query: GATA3 retinoic acid

Results: One result: retinoic acid down-regulates GATA3.

Query: GREB1L retinoic acid

Results: No results

Query: NCOA3 retinoic acid

Results: No results