# Part 1[¶](#Part-1)

Description:

Some more reports to compile:

On the "To Keep" list that Dushi sent, could I get:

1a) How many UniProt IDs on the list are currently found on ZFIN gene records?

1b) How many of those ZFIN gene records are unique?

## 1a[¶](#1a)

How many UniProt IDs on the list are currently found on ZFIN gene records? **Answer: 24,252**.

## 1b[¶](#1b)

How many of those ZFIN gene records are unique? **Answer: 17,991**.

# Part 2[¶](#Part-2)

Of the remaining not associated with the above,

2a) How many UniProt IDs have RefSeq IDs found on ZFIN gene records?

2b) How many of those ZFIN gene records are unique?

## 2a[¶](#2a)

How many UniProt IDs have RefSeq IDs found on ZFIN gene records? Answer: 21,898

## 2b[¶](#2b)

How many of those ZFIN gene records are unique from 2a? Answer: 10,493

## Part 2.5[¶](#Part-2.5)

As far as the “third pass” (#3) in this request, looking at Ensembl IDs, I was thinking more along the lines of what you did over in ticket ZFIN-8376: UniProt Review of Next Release to\_delete and to\_load filesIN DEVELOPMENT where you used the data in NCBI to try to match UniProt IDs to ZFIN through NCBI. This wasn’t clear from my request.

UniProt IDs are associated with RefSeqs in NCBI Gene records. Those NCBI gene records have links back to ZFIN but we don’t have reciprocal links because we don’t match on GFIL sequence IDs. I’m not going to ask you to do that again here because I think we’ve got that covered. It will definitely allow us to load more UniProt IDs.

So it looks like if we load the “To Keep” IDs based on 1) UniProt IDs currently in ZFIN, 2) associations we can make through RefSeqs and 3) associations we can make through NCBI linking to us where we don’t have links to NCBI (eliminate the GFIL problem), we will not lose much relevant coverage.

### Data for Part 2.5:[¶](#Data-for-Part-2.5:)

# Part 3[¶](#Part-3)

Of the remaining not associated with the above,

3a) How many UniProt IDs have Ensembl IDs found on ZFIN gene records? 3b) How many of those ZFIN gene records are unique?

### Strategy[¶](#Strategy)

For step 3, we need some mapping between uniprot IDs and ensembl IDs. We get that here: <https://ftp.expasy.org/databases/uniprot/current_release/knowledgebase/idmapping/by_organism/DANRE_7955_idmapping.dat.gz>

We've imported that file to a table called uniprot2ensembl with 2 columns (accession, ensdarg).

## 3a[¶](#3a)

Joining uniprot to\_keep through ensembl and zfin gives 39 results:

ZDB-GENE-070705-18

ZDB-GENE-060526-280

ZDB-GENE-030411-2

...

### Note[¶](#Note)

I noticed that mapping accessions through ensembl to our db\_link table gives the following 3 entries:

Q5BJA5 ENSDARG00000070297 ZDB-GENE-121214-199

Q5BJA5 ENSDARG00000075482 ZDB-GENE-071004-56

Q5BJA5 ENSDARG00000091728 ZDB-GENE-050320-24

Those don't show up in the above table because we have already processed Q5BJA5 as part of 2a (though that step only matched ZDB-GENE-121214-199). How should we handle a single accession that maps to multiple genes?

## 3b[¶](#3b)

How many of those ZFIN gene records are unique? 40 genes

# Part 4[¶](#Part-4)

Of the remaining not associated with the above,

4a) How many UniProt IDs have Ensembl IDs found in the notes field of "Genomic Feature" records in ZFIN? 4b) How many unique ZFIN gene records are these allele records associated with?

Confused on this one? Look at the notes field on this page: ZFIN Feature: sa35159 We were unable to assign these alleles to a specific gene in ZFIN but we needed to keep the alleles and assign a gene. We created a gene record (in this case unm\_sa35159), attached the allele and made a standard note as to which Ensembl ID the gene corresponds. We could attach a UniProt ID to the gene based on this relationship (maybe shakey and would have to discuss).

## 4a[¶](#4a)

Looks like there are 838 ENSDARGs in the extnote\_note table (606 uniques). Of those, I can match 328 to the uniprot accessions through uniprot2ensembl:

A0A2R8Q3S9 ENSDARG00000100186

Q5BJA5 ENSDARG00000104501

F1QQC3 ENSDARG00000044010

Q08CH3 ENSDARG00000061989

F1RCR6 ENSDARG00000061397

Q1L673 ENSDARG00000070721

## 4b[¶](#4b)

How many unique ZFIN gene records are these allele records associated with?

Using the same uniprot2ensembl table to cross reference from these ENSDARGs to uniprot accessions and cross referencing the feature IDs with markers using the feature\_marker\_relationship table, I found 232 markers.

# Summary[¶](#Summary)

5a) How many unique ZFIN gene records are listed on the above lists, total? Answer: 20,829

5b) How many ZFIN gene records currently have UniProt IDs? Answer: 23,396

5c) How many UniProt IDs on the "To Keep" list are not matched using matching criteria 1-4? Answer: 10,377

(10,299 RefSeq entries, 78 by ZFIN ID)

# Queries supporting these results[¶](#Queries-supporting-these-results)

## Initialize Database[¶](#Initialize-Database)

In [1]:

!rm -f zfin-db-slice.db

!curl -L -o zfin-db-slice.db.gz https://github.com/rtaylorzfin/8395-uniprot-notebook/raw/main/zfin-db-slice.db.gz

!gunzip -f zfin-db-slice.db.gz

% Total % Received % Xferd Average Speed Time Time Time Current

Dload Upload Total Spent Left Speed

0 0 0 0 0 0 0 0 --:--:-- --:--:-- --:--:-- 0

100 24.3M 100 24.3M 0 0 11.2M 0 0:00:02 0:00:02 --:--:-- 22.4M

In [2]:

%reload\_ext sql

In [3]:

%%sql

sqlite:///zfin-db-slice.db

## Database Queries[¶](#Database-Queries)

Create a copy of to\_keep for modifying

In [4]:

%%sql

-- create copy of to\_keep from uniprot

create table to\_keep\_m as select \* from to\_keep;

-- Remove the version from the refseq numbers:

update to\_keep\_m set primary\_id = substr(primary\_id, 0, LENGTH(rtrim(primary\_id, '0123456789'))) where database = 'RefSeq' and primary\_id like '%.%';

update to\_keep\_m set secondary\_id = substr(secondary\_id, 0, LENGTH(rtrim(secondary\_id, '0123456789'))) where database = 'RefSeq' and secondary\_id like '%.%';

create index primary\_id\_key on to\_keep\_m (primary\_id);

create index secondary\_id\_key on to\_keep\_m (secondary\_id);

-- Remove the version from ensdarg in uniprot2ensembl

update uniprot2ensembl set ensdarg = substr(ensdarg, 0, LENGTH(rtrim(ensdarg, '0123456789'))) where ensdarg like '%.%';

-- Remove the version from refseq in refseq2ncbi

update refseq2ncbi set refseq = substr(refseq, 0, LENGTH(rtrim(refseq, '0123456789'))) where refseq like '%.%';

\* sqlite:///zfin-db-slice.db

Done.

62344 rows affected.

62344 rows affected.

Done.

Done.

45704 rows affected.

55485 rows affected.

Out[4]:

[]

In [5]:

%%sql

-- Move the comma delimited dbXrefs from the ncbi\_map table to its own table for ease of querying

-- See: https://stackoverflow.com/questions/51571854/how-to-split-comma-delimited-values-into-multiple-rows-using-sqlite

create table ncbi\_xrefs as

WITH RECURSIVE split(id, value, rest) AS (

SELECT GeneID, '', dbXrefs || '|' FROM ncbi\_map

UNION ALL SELECT

id,

substr(rest, 0, instr(rest, '|')),

substr(rest, instr(rest, '|')+1)

FROM split WHERE rest!=''

)

SELECT id, value

FROM split

WHERE value!='';

\* sqlite:///zfin-db-slice.db

Done.

Out[5]:

[]

In [6]:

%%sql

-- Move the comma delimited dbXrefs from the ncbi\_map table to its own table for ease of querying

-- See: https://stackoverflow.com/questions/51571854/how-to-split-comma-delimited-values-into-multiple-rows-using-sqlite

drop table if exists xrefs;

create table xrefs as

WITH RECURSIVE split(GeneID, dbXrefs, rest) AS (

SELECT GeneID, '', dbXrefs || '|' FROM ncbi\_map

UNION ALL SELECT

GeneID,

substr(rest, 0, instr(rest, '|')),

substr(rest, instr(rest, '|')+1)

FROM split WHERE rest!=''

)

SELECT GeneID, dbXrefs, '' as org, '' as acc

FROM split

WHERE dbXrefs!='';

-- separate the organization and the accession into columns

update xrefs set acc = replace(dbXrefs, rtrim(dbXrefs, replace(dbXrefs, ':', '')), '');

update xrefs set org = substr(dbXrefs, 0, INSTR(dbXrefs, acc) - 1);

\* sqlite:///zfin-db-slice.db

Done.

Done.

90831 rows affected.

90831 rows affected.

Out[6]:

[]

## Section 1 Queries[¶](#Section-1-Queries)

### 1a[¶](#1a)

1a) How many UniProt IDs on the list are currently found on ZFIN gene records? Answer: 54,715.

In [7]:

%%sql

drop table if exists "to\_keep\_accessions\_only";

create table to\_keep\_accessions\_only as select distinct accession from to\_keep;

drop table if exists "1a";

create table "1a" as select \* from to\_keep\_accessions\_only inner join db\_link on accession = dblink\_acc\_num;

select count(\*) from "1a";

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Done.

Done.

Out[7]:

|  |
| --- |
| **count(\*)** |
| 24252 |

### 1b[¶](#1b)

How many of those ZFIN gene records are unique? Answer: 17,991.

In [8]:

%%sql

drop table if exists "1b";

create table "1b" as select mrkr\_abbrev, mrkr\_zdb\_id from marker where mrkr\_zdb\_id in (select dblink\_linked\_recid from "1a");

\* sqlite:///zfin-db-slice.db

Done.

Done.

Out[8]:

[]

In [9]:

%%sql

select count(\*) from "1b";

\* sqlite:///zfin-db-slice.db

Done.

Out[9]:

|  |
| --- |
| **count(\*)** |
| 17991 |

## Section 2 Queries[¶](#Section-2-Queries)

### Prepare DB[¶](#Prepare-DB)

Delete entries from to\_keep\_m that have already been matched in part 1a (48,739 deletes). Small note, instead of doing deletions, I prefer to create a view of to\_keep\_m that mimics what that table would look like if we did a deletion. That way we can run each step independently.

In [10]:

%%sql

drop view if exists to\_keep\_after\_1a;

create view to\_keep\_after\_1a as select \* from to\_keep\_m where accession not in (select accession from "1a");

\* sqlite:///zfin-db-slice.db

Done.

Done.

Out[10]:

[]

### 2a[¶](#2a)

In [11]:

%%sql

create index dblink\_acc\_num\_key on db\_link (dblink\_acc\_num);

create table "2a" as

select \* from to\_keep\_after\_1a

join db\_link on

(primary\_id = dblink\_acc\_num or secondary\_id = dblink\_acc\_num)

and database = 'RefSeq';

\* sqlite:///zfin-db-slice.db

Done.

Done.

Out[11]:

[]

In [12]:

%%sql

select count(distinct accession) from "2a";

\* sqlite:///zfin-db-slice.db

Done.

Out[12]:

|  |
| --- |
| **count(distinct accession)** |
| 21899 |

### 2b[¶](#2b)

How many of those ZFIN gene records are unique from 2a? Answer: 10,493

In [13]:

%%sql

select count(distinct dblink\_linked\_recid) from "2a";

\* sqlite:///zfin-db-slice.db

Done.

Out[13]:

|  |
| --- |
| **count(distinct dblink\_linked\_recid)** |
| 10494 |

In [14]:

%%sql

create table "2b" as select mrkr\_abbrev, mrkr\_zdb\_id from marker where mrkr\_zdb\_id in (select dblink\_linked\_recid from "2a");

\* sqlite:///zfin-db-slice.db

Done.

Out[14]:

[]

In [15]:

%%sql

-- mimic this: delete from to\_keep where accession in (select accession from "2a");

drop view if exists to\_keep\_after\_2a;

create view to\_keep\_after\_2a as select \* from to\_keep\_after\_1a where accession not in (select accession from "2a");

\* sqlite:///zfin-db-slice.db

Done.

Done.

Out[15]:

[]

## Section 2.5 Queries[¶](#Section-2.5-Queries)

In [16]:

%%sql

drop table if exists "2.5a";

create table "2.5a" as

select accession, refseq, ncbi, dblink\_linked\_recid from to\_keep\_after\_2a k join refseq2ncbi r on k.primary\_id = r.refseq join db\_link d on d.dblink\_acc\_num = r.ncbi;

select count(\*) from "2.5a";

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[16]:

|  |
| --- |
| **count(\*)** |
| 62 |

In [17]:

%%sql

-- mimic this: delete from to\_keep where accession in (select accession from "2a");

drop view if exists "to\_keep\_after\_2.5a";

create view "to\_keep\_after\_2.5a" as select \* from to\_keep\_after\_2a where accession not in (select accession from "2.5a");

select count(\*) from "to\_keep\_after\_2.5a";

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[17]:

|  |
| --- |
| **count(\*)** |
| 10318 |

In [18]:

%%sql

select \* from "2.5a";

\* sqlite:///zfin-db-slice.db

Done.

Out[18]:

|  |  |  |  |
| --- | --- | --- | --- |
| **accession** | **refseq** | **ncbi** | **dblink\_linked\_recid** |
| A0A8M6Z2V1 | XP\_017213386 | 101886280 | ZDB-GENE-201021-1 |
| A0A8M6Z2V1 | XP\_021334760 | 101886280 | ZDB-GENE-201021-1 |
| A0A8M9PI15 | XP\_021324394 | 100537633 | ZDB-GENE-201110-1 |
| A0A8M9PTJ7 | XP\_021324395 | 100537633 | ZDB-GENE-201110-1 |
| A0A2R8QN98 | XP\_005157000 | 101883994 | ZDB-GENE-201111-2 |
| A0A8M2B308 | XP\_005157000 | 101883994 | ZDB-GENE-201111-2 |
| A0A8M1RHE9 | XP\_002663314 | 100333146 | ZDB-GENE-201105-1 |
| E7FBD9 | XP\_001922909 | 566378 | ZDB-GENE-201110-2 |
| A0A8M6YZB7 | XP\_017210347 | 566378 | ZDB-GENE-201110-2 |
| A0A8M6Z7P2 | XP\_017212877 | 100334628 | ZDB-GENE-140106-268 |
| A0A8M1RPE8 | XP\_003201788 | 100537771 | ZDB-GENE-210112-1 |
| A0A8M2BGY4 | XP\_005169738 | 100333863 | ZDB-GENE-201217-1 |
| A0A8N7TBG4 | XP\_693964 | 565603 | ZDB-GENE-211020-1 |
| A0A8M9PNK8 | XP\_021333077 | 100005536 | ZDB-GENE-211110-1 |
| A0A8M9Q4R9 | XP\_021334473 | 110437809 | ZDB-GENE-201221-1 |
| A0A8M2B8H5 | XP\_005161336 | 796314 | ZDB-GENE-201223-1 |
| A0A8M9PI63 | XP\_021326653 | 100535428 | ZDB-GENE-221121-1 |
| A0A8M9PP63 | XP\_021326654 | 100535428 | ZDB-GENE-221121-1 |
| A0A8M9PZX1 | XP\_021326655 | 100535428 | ZDB-GENE-221121-1 |
| A0A8M2BGX7 | XP\_005169737 | 100333734 | ZDB-GENE-201217-2 |
| A0A8M1RDA8 | XP\_002660634 | 100334365 | ZDB-GENE-201223-2 |
| A0A8M1PSL8 | XP\_001342135 | 100002327 | ZDB-GENE-201223-3 |
| A0A0G2L879 | XP\_005174278 | 101882451 | ZDB-GENE-200318-2 |
| A0A8M9PYK9 | XP\_021326245 | 101882451 | ZDB-GENE-200318-2 |
| A0A8M2B898 | XP\_005161335 | 100147871 | ZDB-GENE-201223-4 |
| A0A8M2B706 | XP\_005159907 | 567498 | ZDB-GENE-130522-3 |
| A0A8M6Z0C4 | XP\_017213239 | 100331465 | ZDB-GENE-201012-1 |
| A0A0G2L0N1 | XP\_017207581 | 100007225 | ZDB-GENE-200723-1 |
| A0A8M6YUB7 | XP\_017207581 | 100007225 | ZDB-GENE-200723-1 |
| A0A8M6YT62 | XP\_017206601 | 100330653 | ZDB-GENE-200330-2 |
| A0A2R8PY68 | NP\_001313426 | 100537904 | ZDB-GENE-161128-1 |
| A0A8M1P2R9 | NP\_001313426 | 100537904 | ZDB-GENE-161128-1 |
| A0A8M3AWI8 | XP\_009303410 | 100537904 | ZDB-GENE-161128-1 |
| A0A8M1REC6 | XP\_002666419 | 100333891 | ZDB-GENE-141223-1 |
| A0A8M2BGI8 | XP\_005168683 | 100333891 | ZDB-GENE-141223-1 |
| A0A8M9P965 | XP\_021322953 | 796202 | ZDB-GENE-210812-1 |
| A0A8M9P965 | XP\_021328735 | 796202 | ZDB-GENE-210812-1 |
| A0A8M1RLH6 | XP\_002667589 | 100332723 | ZDB-GENE-200729-1 |
| A0A8M1RLH6 | XP\_021326579 | 100332723 | ZDB-GENE-200729-1 |
| A0A8M9PZP2 | XP\_021326580 | 100332723 | ZDB-GENE-200729-1 |
| A0A8M9PW51 | XP\_021326581 | 100332723 | ZDB-GENE-200729-1 |
| A0A8M9QL68 | XP\_021336971 | 101882648 | ZDB-GENE-090609-4 |
| A0A8M2BCL8 | XP\_005165254 | 559162 | ZDB-GENE-060526-336 |
| A0A8M1PV41 | XP\_001342688 | 100003043 | ZDB-GENE-050208-530 |
| A0A8M1PV41 | XP\_021329231 | 100003043 | ZDB-GENE-050208-530 |
| A0A8M9Q954 | XP\_021332046 | 567114 | ZDB-GENE-220119-1 |
| A0A8M9PKP9 | XP\_021332047 | 567114 | ZDB-GENE-220119-1 |
| A0A8M9PY86 | XP\_021332048 | 567114 | ZDB-GENE-220119-1 |
| A0A8M9PXR7 | XP\_021325935 | 795763 | ZDB-GENE-050616-13 |
| A0A8M9P965 | XP\_021322953 | 796202 | ZDB-GENE-201118-2 |
| A0A8M9P965 | XP\_021328735 | 796202 | ZDB-GENE-201118-2 |
| A0A8M9PIW5 | XP\_021322181 | 792889 | ZDB-GENE-110804-1 |
| A0A8M6YWJ3 | XP\_017207875 | 101886447 | ZDB-GENE-220328-1 |
| A0A8M6Z8I4 | XP\_017213238 | 101886421 | ZDB-GENE-220321-1 |
| G8XQM4 | NP\_001243568 | 100537567 | ZDB-GENE-220329-1 |
| A0A8M3AR39 | XP\_009291047 | 100330960 | ZDB-GENE-060515-1 |
| A0A8M3B7J6 | XP\_009305031 | 559133 | ZDB-GENE-190528-1 |
| E7F7E1 | XP\_009305031 | 559133 | ZDB-GENE-190528-1 |
| A0A8M9QE05 | XP\_021328060 | 557816 | ZDB-GENE-200107-1 |
| A0A8M9QE05 | XP\_021335659 | 557816 | ZDB-GENE-200107-1 |
| A0A8M9QMZ8 | XP\_021336025 | 110440145 | ZDB-GENE-220914-1 |
| A0A8M9PF49 | XP\_021323204 | 108179693 | ZDB-GENE-210219-1 |

## Section 3 Queries[¶](#Section-3-Queries)

In [19]:

%%sql

drop table if exists "3a";

create table "3a" as

select \* from to\_keep\_after\_2a tkm

left join uniprot2ensembl e on tkm.accession = e.accession

left join db\_link on ensdarg = dblink\_acc\_num

where dblink\_linked\_recid is not null;

select \* from "3a" limit 3;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[19]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **accession** | **entry\_type** | **database** | **primary\_id** | **secondary\_id** | **isoform** | **accession:1** | **db** | **ensdarg** | **dblink\_linked\_recid** | **dblink\_acc\_num** | **dblink\_info** | **dblink\_zdb\_id** | **dblink\_acc\_num\_display** | **dblink\_length** | **dblink\_fdbcont\_zdb\_id** |
| A0A0G2KC95 | TrEMBL | RefSeq | XP\_021330976 | XM\_021475301 |  | A0A0G2KC95 | Ensembl | ENSDARG00000113971 | ZDB-GENE-161017-91 | ENSDARG00000113971 |  | ZDB-DBLINK-201118-20 | ENSDARG00000113971 |  | ZDB-FDBCONT-200123-1 |
| A0A0G2KC95 | TrEMBL | RefSeq | XP\_021330976 | XM\_021475301 |  | A0A0G2KC95 | Ensembl | ENSDARG00000113971 | ZDB-GENE-161017-91 | ENSDARG00000113971 |  | ZDB-DBLINK-200727-3 | ENSDARG00000113971 |  | ZDB-FDBCONT-061018-1 |
| A0A0G2KM47 | TrEMBL | RefSeq | XP\_696046 | XM\_690954 |  | A0A0G2KM47 | Ensembl | ENSDARG00000101138 | ZDB-GENE-090303-2 | ENSDARG00000101138 |  | ZDB-DBLINK-220412-5 | ENSDARG00000101138 |  | ZDB-FDBCONT-061018-1 |

In [20]:

%%sql

select distinct accession from "3a";

\* sqlite:///zfin-db-slice.db

Done.

Out[20]:

|  |
| --- |
| **accession** |
| A0A0G2KC95 |
| A0A0G2KM47 |
| A0A0G2KM96 |
| A0A0G2KV16 |
| A0A0G2KZI5 |
| A0A0G2L0N1 |
| A0A0G2L8W6 |
| A0A1L1QZG2 |
| A0A1L1QZI0 |
| A0A1L1QZN4 |
| A0A286Y931 |
| A0A2R8QK73 |
| A0A5H1ZRH9 |
| A0A8E7UDD7 |
| A0A8M9PQ61 |
| A0A8N7XJA6 |
| A2RV10 |
| A3KQ92 |
| A5WV69 |
| A8KC24 |
| B3DJS3 |
| C4XVB5 |
| D2Y178 |
| E7EXK5 |
| E7F7E1 |
| E7FCU2 |
| E7FD79 |
| E7FFT2 |
| E7FGS2 |
| E9QCW0 |
| F1Q5R8 |
| G3G7U2 |
| Q1LUU9 |
| Q2PR87 |
| Q2PRL6 |
| Q5TKT2 |
| Q803H4 |
| Q8AW20 |
| U3N6I5 |

In [21]:

%%sql

drop table if exists "3b";

create table "3b" as

select distinct accession, mrkr\_zdb\_id, mrkr\_abbrev from marker join "3a" on mrkr\_zdb\_id = dblink\_linked\_recid ;

select \* from "3b" order by accession;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[21]:

|  |  |  |
| --- | --- | --- |
| **accession** | **mrkr\_zdb\_id** | **mrkr\_abbrev** |
| A0A0G2KC95 | ZDB-GENE-161017-91 | si:cabz01069016.2 |
| A0A0G2KM47 | ZDB-GENE-090303-2 | cxcl11.4 |
| A0A0G2KM96 | ZDB-GENE-130726-2 | kcnj10b |
| A0A0G2KV16 | ZDB-GENE-200114-3 | ifit17 |
| A0A0G2KZI5 | ZDB-GENE-161017-26 | si:dkey-240n22.9 |
| A0A0G2L0N1 | ZDB-GENE-200318-4 | olfcg13 |
| A0A0G2L0N1 | ZDB-GENE-200723-1 | olfcg14 |
| A0A0G2L8W6 | ZDB-GENE-161017-35 | si:dkey-172k15.10 |
| A0A1L1QZG2 | ZDB-GENE-161017-3 | si:dkey-172k15.18 |
| A0A1L1QZI0 | ZDB-GENE-161017-132 | si:cabz01021435.4 |
| A0A1L1QZN4 | ZDB-GENE-161017-2 | si:dkey-172k15.17 |
| A0A286Y931 | ZDB-GENE-120709-85 | insyn2ab |
| A0A2R8QK73 | ZDB-GENE-131121-319 | tcerg1l |
| A0A5H1ZRH9 | ZDB-GENE-060526-261 | tmc1 |
| A0A8E7UDD7 | ZDB-GENE-170413-1 | gpr6 |
| A0A8M9PQ61 | ZDB-GENE-030131-1662 | dspb |
| A0A8N7XJA6 | ZDB-GENE-070705-19 | clcn2a |
| A2RV10 | ZDB-GENE-030707-2 | her15.1 |
| A3KQ92 | ZDB-GENE-060503-166 | syt1b |
| A5WV69 | ZDB-GENE-070705-18 | acbd5b |
| A8KC24 | ZDB-GENE-040426-2149 | mhc1zba |
| B3DJS3 | ZDB-GENE-031118-50 | snx9b |
| C4XVB5 | ZDB-GENE-100922-169 | ikzf2 |
| D2Y178 | ZDB-GENE-060526-110 | npffr1l2 |
| E7EXK5 | ZDB-GENE-181119-4 | zmp:0000001331 |
| E7F7E1 | ZDB-GENE-190528-1 | uts2r1 |
| E7FCU2 | ZDB-GENE-091208-1 | acot12 |
| E7FD79 | ZDB-GENE-080723-21 | zgc:193593 |
| E7FFT2 | ZDB-GENE-060526-280 | tmc2a |
| E7FGS2 | ZDB-GENE-200102-1 | ncoa7a |
| E9QCW0 | ZDB-GENE-060526-185 | tacr1a |
| F1Q5R8 | ZDB-GENE-050208-742 | aox6 |
| G3G7U2 | ZDB-GENE-990415-25 | rrm2 |
| Q1LUU9 | ZDB-GENE-020430-1 | flot1a |
| Q2PR87 | ZDB-GENE-081031-102 | si:dkey-92k1.9 |
| Q2PRL6 | ZDB-GENE-070804-2 | or80a13 |
| Q5TKT2 | ZDB-GENE-061031-5 | il17c |
| Q803H4 | ZDB-GENE-030411-2 | selenoh |
| Q8AW20 | ZDB-GENE-030616-264 | stat4 |
| U3N6I5 | ZDB-GENE-110421-2 | grm8a |

In [22]:

%%sql

select count(\*) from "3b";

\* sqlite:///zfin-db-slice.db

Done.

Out[22]:

|  |
| --- |
| **count(\*)** |
| 40 |

In [23]:

%%sql

--mimic: delete from to\_keep where accession in (select accession from "3a");

drop view if exists to\_keep\_after\_3a;

create view to\_keep\_after\_3a as select \* from "to\_keep\_after\_2.5a" where accession not in (select accession from "3a");

select count(\*) from to\_keep\_after\_3a;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[23]:

|  |
| --- |
| **count(\*)** |
| 10255 |

## Section 4 Queries[¶](#Section-4-Queries)

### 4a[¶](#4a)

In [24]:

%%sql

select \* from extnote\_note where extnote\_note like '%ensdarg%' limit 1;

\* sqlite:///zfin-db-slice.db

Done.

Out[24]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **extnote\_zdb\_id** | **extnote\_data\_zdb\_id** | **extnote\_note** | **extnote\_note\_type** | **extnote\_source\_zdb\_id** | **extnote\_tag** |
| ZDB-EXTNOTE-160601-517 | ZDB-ALT-130502-1 | carries a cysteine to threonine mutation in exon 4 of syne3 (ENSDARG00000023237) leading to the introduction of a premature stop codon at position 195 | feature | ZDB-PUB-160331-1 | variant with ID ZDB-FGMD-190401-1961 |

In [25]:

%%sql

--create tempnotes table with the ENSDARG isolated to its own field

create table tempnotes as select SUBSTR(extnote\_note, INSTR(extnote\_note, 'ENSDARG'), 18) AS ensdarg, \* from extnote\_note;

\* sqlite:///zfin-db-slice.db

Done.

Out[25]:

[]

In [26]:

%%sql

select count(ensdarg) from tempnotes;

\* sqlite:///zfin-db-slice.db

Done.

Out[26]:

|  |
| --- |
| **count(ensdarg)** |
| 838 |

In [27]:

%%sql

drop table if exists "4a";

create table "4a" as

select n.ensdarg, u2e.accession from tempnotes n left join uniprot2ensembl u2e on n.ensdarg = u2e.ensdarg;

select count(distinct accession) from "4a";

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[27]:

|  |
| --- |
| **count(distinct accession)** |
| 328 |

In [28]:

%%sql

select \* from "4a" limit 2;

\* sqlite:///zfin-db-slice.db

Done.

Out[28]:

|  |  |
| --- | --- |
| **ensdarg** | **accession** |
| ENSDARG00000023237 | X1WGA9 |
| ENSDARG00000070074 | E7EZD4 |

### 4b[¶](#4b)

In [29]:

%%sql

-- create supporting table (feature\_marker\_ensdarg) of joins from feature\_marker\_relationship, notes(with ensdarg), uniprot2ensembl, marker

drop table if exists "feature\_marker\_ensdarg";

create table "feature\_marker\_ensdarg" as

select fmrel\_ftr\_zdb\_id, fmrel\_type, fmrel\_mrkr\_zdb\_id as marker\_id, m.mrkr\_abbrev as gene\_abbrev, tempnotes.ensdarg, extnote\_note

from feature\_marker\_relationship

join tempnotes on extnote\_data\_zdb\_id = fmrel\_ftr\_zdb\_id

left join uniprot2ensembl u2e on tempnotes.ensdarg = u2e.ensdarg

left join marker m on fmrel\_mrkr\_zdb\_id = mrkr\_zdb\_id;

select \* from feature\_marker\_ensdarg limit 5;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[29]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **fmrel\_ftr\_zdb\_id** | **fmrel\_type** | **marker\_id** | **gene\_abbrev** | **ensdarg** | **extnote\_note** |
| ZDB-ALT-070426-3 | is allele of | ZDB-GENE-050114-2 | lmx1bb | ENSDARG00000068365 | Mutation in lmx1b (ENSDARG00000068365) that introduces a premature termination codon at 34,129,111 (T>A) on chromosome 8. Overall ear size reduced, abnormally formed semicircular canals, and slightly reduced otoliths. |
| ZDB-ALT-070426-3 | is allele of | ZDB-GENE-050114-2 | lmx1bb | ENSDARG00000068365 | Mutation in lmx1b (ENSDARG00000068365) that introduces a premature termination codon at 34,129,111 (T>A) on chromosome 8. Overall ear size reduced, abnormally formed semicircular canals, and slightly reduced otoliths. |
| ZDB-ALT-070427-1 | is allele of | ZDB-GENE-011128-4 | jag1b | ENSDARG00000013168 | G>A mutation at position 35,593,144 on chromosome 13 in jag1b (ENSDARG00000013168) disrupting a splice donor site resulting in a frameshift. Antero-posterior axis is shortened in the ear. |
| ZDB-ALT-070427-1 | is allele of | ZDB-GENE-011128-4 | jag1b | ENSDARG00000013168 | G>A mutation at position 35,593,144 on chromosome 13 in jag1b (ENSDARG00000013168) disrupting a splice donor site resulting in a frameshift. Antero-posterior axis is shortened in the ear. |
| ZDB-ALT-090714-8 | contains innocuous sequence feature | ZDB-TGCONSTRCT-090713-1 | Tg(ClzpAU14I) | ENSDARG00000071738 | Family pa008 22; 1,567,618 Plus strand, Target gene- ENSDARG00000071738 180 bp upstream XM\_001922442 |

In [30]:

%%sql

--match ensdarg to uniprot accession

drop table if exists "4b";

create table "4b" as

select accession, fme.\* from feature\_marker\_ensdarg fme join uniprot2ensembl u2e on fme.ensdarg = u2e.ensdarg;

select \* from "4b" limit 2;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[30]:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **accession** | **fmrel\_ftr\_zdb\_id** | **fmrel\_type** | **marker\_id** | **gene\_abbrev** | **ensdarg** | **extnote\_note** |
| Q4L1M6 | ZDB-ALT-070426-3 | is allele of | ZDB-GENE-050114-2 | lmx1bb | ENSDARG00000068365 | Mutation in lmx1b (ENSDARG00000068365) that introduces a premature termination codon at 34,129,111 (T>A) on chromosome 8. Overall ear size reduced, abnormally formed semicircular canals, and slightly reduced otoliths. |
| Q4L1M6 | ZDB-ALT-070426-3 | is allele of | ZDB-GENE-050114-2 | lmx1bb | ENSDARG00000068365 | Mutation in lmx1b (ENSDARG00000068365) that introduces a premature termination codon at 34,129,111 (T>A) on chromosome 8. Overall ear size reduced, abnormally formed semicircular canals, and slightly reduced otoliths. |

In [31]:

%%sql

select count(\*) from (select distinct accession, marker\_id from "4b");

\* sqlite:///zfin-db-slice.db

Done.

Out[31]:

|  |
| --- |
| **count(\*)** |
| 408 |

In [32]:

%%sql

select count(\*) from (select distinct marker\_id from "4b");

\* sqlite:///zfin-db-slice.db

Done.

Out[32]:

|  |
| --- |
| **count(\*)** |
| 232 |

In [33]:

%%sql

select accession, ensdarg, marker\_id, gene\_abbrev from "4b" limit 2;

\* sqlite:///zfin-db-slice.db

Done.

Out[33]:

|  |  |  |  |
| --- | --- | --- | --- |
| **accession** | **ensdarg** | **marker\_id** | **gene\_abbrev** |
| Q4L1M6 | ENSDARG00000068365 | ZDB-GENE-050114-2 | lmx1bb |
| Q4L1M6 | ENSDARG00000068365 | ZDB-GENE-050114-2 | lmx1bb |

In [34]:

%%sql

select count( distinct accession ) from "4a" limit 5;

\* sqlite:///zfin-db-slice.db

Done.

Out[34]:

|  |
| --- |
| **count( distinct accession )** |
| 328 |

In [35]:

%%sql

--mimic: delete from to\_keep where accession in (select accession from "4a");

drop view if exists to\_keep\_after\_4a;

create view to\_keep\_after\_4a as select \* from to\_keep\_after\_3a k where k.accession not in (select distinct accession from "4b");

select count(\*) from to\_keep\_after\_4a;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[35]:

|  |
| --- |
| **count(\*)** |
| 10223 |

## Section 5 Queries[¶](#Section-5-Queries)

## 5a[¶](#5a)

In [36]:

%%sql

select count (distinct dblink\_linked\_recid) as gene\_id from (

select dblink\_linked\_recid from "1a"

union

select dblink\_linked\_recid from "2a"

union

select dblink\_linked\_recid from "3a"

union

select marker\_id as dblink\_linked\_recid from "4b"

)

-- select dblink\_linked\_recid from "2a" limit 5; --get gene IDs from 2a

-- select dblink\_linked\_recid from "3a" ; --get gene IDs from 3a

-- select marker\_id from "4b"; --get gene IDs from 4b

\* sqlite:///zfin-db-slice.db

Done.

Out[36]:

|  |
| --- |
| **gene\_id** |
| 21075 |

## 5b[¶](#5b)

In [37]:

%%sql

create table "5b" as

select distinct

dblink\_linked\_recid as gene\_id,

mrkr\_abbrev as gene\_abbrev

FROM

db\_link dl

left join marker m on dl.dblink\_linked\_recid = m.mrkr\_zdb\_id

WHERE

dblink\_fdbcont\_zdb\_id in ( 'ZDB-FDBCONT-040412-47' ) -- uniprot is ZDB-FDBCONT-040412-47

order by dblink\_linked\_recid;

select count(\*) from "5b";

\* sqlite:///zfin-db-slice.db

Done.

Done.

Out[37]:

|  |
| --- |
| **count(\*)** |
| 23396 |

In [38]:

%%sql

select \* from to\_keep\_after\_4a where database <> 'RefSeq';

\* sqlite:///zfin-db-slice.db

Done.

Out[38]:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **accession** | **entry\_type** | **database** | **primary\_id** | **secondary\_id** | **isoform** |
| A0A0G2L877 | TrEMBL | ZFIN | ZDB-GENE-070410-35 | taco1 |  |
| A0A2R8QTM9 | TrEMBL | ZFIN | ZDB-GENE-030131-437 | ybey |  |
| A0A8M1NZ01 | TrEMBL | ZFIN | ZDB-GENE-130530-791 | prr5a |  |
| A0A8M1P317 | TrEMBL | ZFIN | ZDB-GENE-081107-2 | si:ch1073-291c23.2 |  |
| A0A8M1P3Z2 | TrEMBL | ZFIN | ZDB-GENE-131121-624 | si:ch211-189e2.2 |  |
| A0A8M1P9C3 | TrEMBL | ZFIN | ZDB-GENE-081107-2 | si:ch1073-291c23.2 |  |
| A0A8M1PZ20 | TrEMBL | ZFIN | ZDB-GENE-091204-17 | fam234a |  |
| A0A8M1Q7M7 | TrEMBL | ZFIN | ZDB-GENE-060503-247 | emilin2a |  |
| A0A8M1QK92 | TrEMBL | ZFIN | ZDB-GENE-090312-22 | mxra5a |  |
| A0A8M1RI32 | TrEMBL | ZFIN | ZDB-GENE-170421-1 | spata22 |  |
| A0A8M1RJ47 | TrEMBL | ZFIN | ZDB-GENE-120709-60 | si:ch211-201o1.1 |  |
| A0A8M1RJV8 | TrEMBL | ZFIN | ZDB-GENE-090311-19 | lrba |  |
| A0A8M1RSW5 | TrEMBL | ZFIN | ZDB-GENE-130530-871 | rhobtb1 |  |
| A0A8M2B8K6 | TrEMBL | ZFIN | ZDB-GENE-030616-129 | patj |  |
| A0A8M2BCP5 | TrEMBL | ZFIN | ZDB-GENE-070705-337 | cspg4ba |  |
| A0A8M2BIG9 | TrEMBL | ZFIN | ZDB-GENE-131127-71 | si:ch73-40a17.4 |  |
| A0A8M2BKC8 | TrEMBL | ZFIN | ZDB-GENE-140106-261 | zmp:0000001301 |  |
| A0A8M2BKM7 | TrEMBL | ZFIN | ZDB-GENE-081113-1 | cplx2 |  |
| A0A8M3AJX9 | TrEMBL | ZFIN | ZDB-GENE-030616-152 | heatr5a |  |
| A0A8M3AP30 | TrEMBL | ZFIN | ZDB-GENE-131120-167 | si:dkey-106n21.1 |  |
| A0A8M3AQU2 | TrEMBL | ZFIN | ZDB-GENE-100922-42 | hivep3b |  |
| A0A8M3ASW3 | TrEMBL | ZFIN | ZDB-GENE-110411-120 | tecta |  |
| A0A8M3AZ57 | TrEMBL | ZFIN | ZDB-GENE-131120-131 | gabrr3b |  |
| A0A8M3B0F5 | TrEMBL | ZFIN | ZDB-GENE-170330-1 | pik3r6a |  |
| A0A8M3B1J1 | TrEMBL | ZFIN | ZDB-GENE-130530-634 | kcng4b |  |
| A0A8M3B6E4 | TrEMBL | ZFIN | ZDB-GENE-110411-276 | fyco1a |  |
| A0A8M3B6K5 | TrEMBL | ZFIN | ZDB-GENE-091118-25 | aak1b |  |
| A0A8M6YX16 | TrEMBL | ZFIN | ZDB-GENE-081103-3 | si:dkey-265e15.2 |  |
| A0A8M6Z0Y4 | TrEMBL | ZFIN | ZDB-GENE-141219-15 | ttc16 |  |
| A0A8M6Z479 | TrEMBL | ZFIN | ZDB-GENE-120709-73 | cspp1a |  |
| A0A8M6Z6L2 | TrEMBL | ZFIN | ZDB-GENE-110913-180 | si:ch73-266f23.1 |  |
| A0A8M6Z959 | TrEMBL | ZFIN | ZDB-GENE-131121-398 | si:dkeyp-97e7.9 |  |
| A0A8M9P2G6 | TrEMBL | ZFIN | ZDB-GENE-070912-667 | si:dkeyp-68b7.10 |  |
| A0A8M9PDG6 | TrEMBL | ZFIN | ZDB-GENE-050208-336 | cts12 |  |
| A0A8M9PFP2 | TrEMBL | ZFIN | ZDB-GENE-100921-73 | clstn3 |  |
| A0A8M9PFT5 | TrEMBL | ZFIN | ZDB-GENE-050419-171 | pgpep1l |  |
| A0A8M9PGQ5 | TrEMBL | ZFIN | ZDB-GENE-050320-122 | zgc:113295 |  |
| A0A8M9PP83 | TrEMBL | ZFIN | ZDB-GENE-131127-246 | si:ch211-15d5.12 |  |
| A0A8M9PQI7 | TrEMBL | ZFIN | ZDB-GENE-090313-92 | snphb |  |
| A0A8M9PRG4 | TrEMBL | ZFIN | ZDB-GENE-081104-348 | lmod1a |  |
| A0A8M9PST5 | TrEMBL | ZFIN | ZDB-GENE-050419-202 | pias1b |  |
| A0A8M9PVW5 | TrEMBL | ZFIN | ZDB-GENE-050419-126 | ttc23 |  |
| A0A8M9PWM8 | TrEMBL | ZFIN | ZDB-GENE-131127-210 | si:dkey-172k15.4 |  |
| A0A8M9PWY5 | TrEMBL | ZFIN | ZDB-GENE-121214-37 | si:dkey-269i1.4 |  |
| A0A8M9PZN0 | TrEMBL | ZFIN | ZDB-GENE-131127-267 | adgrl1a |  |
| A0A8M9Q1N6 | TrEMBL | ZFIN | ZDB-GENE-131127-582 | si:ch211-284e13.14 |  |
| A0A8M9Q1X1 | TrEMBL | ZFIN | ZDB-GENE-100921-78 | cgna |  |
| A0A8M9Q293 | TrEMBL | ZFIN | ZDB-GENE-050517-2 | abca3b |  |
| A0A8M9Q3K9 | TrEMBL | ZFIN | ZDB-GENE-090313-92 | snphb |  |
| A0A8M9Q5L8 | TrEMBL | ZFIN | ZDB-GENE-131119-82 | si:dkey-111k8.3 |  |
| A0A8M9Q7C5 | TrEMBL | ZFIN | ZDB-GENE-050208-475 | lama3 |  |
| A0A8M9Q7D4 | TrEMBL | ZFIN | ZDB-GENE-050420-335 | si:dkeyp-35e5.5 |  |
| A0A8M9Q9Z6 | TrEMBL | ZFIN | ZDB-GENE-081107-2 | si:ch1073-291c23.2 |  |
| A0A8M9QBG1 | TrEMBL | ZFIN | ZDB-GENE-081104-416 | mcf2lb |  |
| A0A8M9QFD9 | TrEMBL | ZFIN | ZDB-GENE-040724-111 | si:rp71-1g18.13 |  |
| A0A8M9QFD9 | TrEMBL | ZFIN | ZDB-GENE-040724-111 | SO:0001217 |  |
| A0A8M9QGN1 | TrEMBL | ZFIN | ZDB-GENE-141212-301 | sh3rf2 |  |
| A0A8M9QHY5 | TrEMBL | ZFIN | ZDB-GENE-160728-123 | si:ch73-174h16.5 |  |
| A0A8N7TFD8 | TrEMBL | ZFIN | ZDB-GENE-081104-431 | lzts1 |  |
| A0A8N7UVK1 | TrEMBL | ZFIN | ZDB-GENE-040724-39 | ankrd33ba |  |
| A6P6V5 | TrEMBL | ZFIN | ZDB-GENE-050510-3 | tas2r203 |  |
| Q5SYA4 | TrEMBL | ZFIN | ZDB-GENE-020225-16 | nitr1b |  |

In [39]:

%%sql

select count(\*) from to\_keep\_after\_4a where database = 'RefSeq';

\* sqlite:///zfin-db-slice.db

Done.

Out[39]:

|  |
| --- |
| **count(\*)** |
| 10161 |

In [40]:

%%sql

create view extnotes as select \* from tempnotes;

\* sqlite:///zfin-db-slice.db

Done.

Out[40]:

[]

# Appendix[¶](#Appendix)

## Appendix 1. Error Check[¶](#Appendix-1.-Error-Check)

Implement the logic described in comment:

To find potential problems, I was thinking it would be a good idea to 1) find all of the UniProt IDs on the “To Keep” list in ZFIN and their corresponding ZFIN gene. Then, from the original ”To Keep” list of the UniProt IDs, 2) run all of the RefSeqs to find the corresponding ZFIN gene. Then make sure that the ZFIN gene identified by the UniProt ID was the same ZFIN gene identified by the RefSeq.

For example, the lines in the “to Keep” file:

Q75ZI3,TrEMBL,RefSeq,NP\_999853.1,NM\_214688.1, Q75ZI3,TrEMBL,ZFIN,ZDB-GENE-030131-2185,vcanb,

In the first search, Q75ZI3 is associated with vcanb and in the second search, RefSeq,NP\_999853 is also associated with vcanb → OK

In the first search, Q75ZI3 is associated with vcanb and in the second search, RefSeq,NP\_999853 is associated with vcana → Not OK

In [41]:

%%sql

drop table if exists to\_keep\_wide;

create table to\_keep\_wide as

select a.accession, a.entry\_type, a.primary\_id as zfin\_id, a.secondary\_id as abbrev, b.primary\_id as refseq1, b.secondary\_id as refseq2 from (select \* from to\_keep\_m where database = 'ZFIN') as a

left join (select \* from to\_keep\_m where database = 'RefSeq') b on a.accession = b.accession;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Out[41]:

[]

In [42]:

%%sql

drop table if exists appendix1;

create table appendix1 as

SELECT

k.\*,

d.dblink\_linked\_recid,

m.mrkr\_abbrev

FROM

to\_keep\_wide k

JOIN db\_link d ON d.dblink\_acc\_num = k.refseq1

JOIN marker m ON d.dblink\_linked\_recid = m.mrkr\_zdb\_id

WHERE zfin\_id not in (select dblink\_linked\_recid from db\_link where dblink\_acc\_num = k.refseq1);

-- select count(\*) from appendix1;

select \* from appendix1;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[42]:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **accession** | **entry\_type** | **zfin\_id** | **abbrev** | **refseq1** | **refseq2** | **dblink\_linked\_recid** | **mrkr\_abbrev** |
| A0A0G2KES1 | TrEMBL | ZDB-GENE-050522-331 | zgc:110249 | XP\_005163751 | XM\_005163694 | ZDB-GENE-030131-4447 | wu:fd06e11 |
| A0A0H2UKA4 | TrEMBL | ZDB-GENE-141215-61 | pimr118 | XP\_001339551 | XM\_001339515 | ZDB-GENE-141215-37 | pimr210 |
| A0A0N4STV9 | TrEMBL | ZDB-GENE-080215-17 | ftr50 | XP\_017210397 | XM\_017354908 | ZDB-GENE-080218-19 | ftr46 |
| A0A0R4IA89 | TrEMBL | ZDB-GENE-041014-40 | si:dkey-27n6.1 | XP\_021331404 | XM\_021475729 | ZDB-GENE-131118-8 | si:ch211-203c5.3 |
| A0A0R4IBX9 | TrEMBL | ZDB-GENE-141216-84 | si:ch73-1a9.3 | XP\_005169261 | XM\_005169204 | ZDB-GENE-050809-70 | im:7159173 |
| A0A0R4ICR4 | TrEMBL | ZDB-GENE-081106-1 | fgfbp1b | XP\_005157409 | XM\_005157352 | ZDB-GENE-140430-1 | fgfbp1a |
| A0A0R4IFL8 | TrEMBL | ZDB-GENE-070424-98 | zgc:162952 | XP\_001343419 | XM\_001343383 | ZDB-GENE-131127-583 | b3galt8 |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_002662464 | XM\_002662418 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_005168803 | XM\_005168746 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_021332779 | XM\_021477104 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_021332780 | XM\_021477105 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHS5 | TrEMBL | ZDB-GENE-110914-181 | si:dkey-262g12.7 | XP\_009299301 | XM\_009301026 | ZDB-GENE-071004-60 | zgc:173624 |
| A0A0R4IIX8 | TrEMBL | ZDB-GENE-131121-97 | si:dkey-172k15.3 | XP\_005165023 | XM\_005164966 | ZDB-GENE-161017-82 | si:dkey-172k15.13 |
| A0A0R4IJA2 | TrEMBL | ZDB-GENE-051120-168 | ppp2r2ca | XP\_009289579 | XM\_009291304 | ZDB-GENE-101026-5 | ppp2r2ba |
| A0A0R4IJR7 | TrEMBL | ZDB-GENE-091204-431 | si:ch1073-392o20.2 | XP\_003199201 | XM\_003199153 | ZDB-GENE-200323-2 | tgoln2 |
| A0A0R4IK53 | TrEMBL | ZDB-GENE-060526-127 | cmklrl2 | XP\_001335526 | XM\_001335490 | ZDB-GENE-060526-128 | cmklr1l1 |
| A0A0R4IKB2 | TrEMBL | ZDB-GENE-071004-121 | zgc:174353 | XP\_021328071 | XM\_021472396 | ZDB-GENE-080214-3 | zgc:174355 |
| A0A0R4IKB2 | TrEMBL | ZDB-GENE-071004-121 | zgc:174353 | XP\_021330305 | XM\_021474630 | ZDB-GENE-080214-3 | zgc:174355 |
| A0A0R4IKB2 | TrEMBL | ZDB-GENE-071004-121 | zgc:174353 | XP\_021330343 | XM\_021474668 | ZDB-GENE-080214-3 | zgc:174355 |
| A0A0R4IKB2 | TrEMBL | ZDB-GENE-071004-121 | zgc:174353 | XP\_021335632 | XM\_021479957 | ZDB-GENE-080214-3 | zgc:174355 |
| A0A0R4ILY2 | TrEMBL | ZDB-GENE-030131-5958 | dtx3lb.2 | XP\_009304026 | XM\_009305751 | ZDB-GENE-131126-78 | dtx3lb.3 |
| A0A0R4IMZ3 | TrEMBL | ZDB-GENE-110913-88 | si:dkey-26i24.1 | XP\_021330575 | XM\_021474900 | ZDB-GENE-080208-8 | zgc:174690 |
| A0A0R4INQ1 | TrEMBL | ZDB-GENE-141215-36 | si:ch211-160o17.4 | XP\_009304049 | XM\_009305774 | ZDB-GENE-140324-1 | hipk1b |
| A0A0R4IQM2 | TrEMBL | ZDB-GENE-060929-84 | fbxo44.3 | XP\_021325462 | XM\_021469787 | ZDB-GENE-040801-53 | fbxo44.5 |
| A0A0R4IRM1 | TrEMBL | ZDB-GENE-070629-2 | kif5ba | XP\_002664065 | XM\_002664019 | ZDB-GENE-070629-4 | kif5bb |
| A0A0R4IUE1 | TrEMBL | ZDB-GENE-070806-44 | or94a1 | XP\_009289844 | XM\_009291569 | ZDB-GENE-141211-45 | si:ch211-24o8.2 |
| A0A0R4IUN5 | TrEMBL | ZDB-GENE-070806-84 | or63a2 | XP\_001920470 | XM\_001920435 | ZDB-GENE-131121-386 | or63h1 |
| A0A0R4IV27 | TrEMBL | ZDB-GENE-080204-2 | oscp1a | XP\_009292251 | XM\_009293976 | ZDB-GENE-090508-7 | lsm10 |
| A0A0R4IV86 | TrEMBL | ZDB-GENE-140106-124 | si:ch211-221f10.2 | XP\_009289330 | XM\_009291055 | ZDB-GENE-030131-8507 | wu:fb13g09 |
| A0A0R4IV86 | TrEMBL | ZDB-GENE-140106-124 | si:ch211-221f10.2 | XP\_017214608 | XM\_017359119 | ZDB-GENE-030131-8507 | wu:fb13g09 |
| A0A0R4IV86 | TrEMBL | ZDB-GENE-140106-124 | si:ch211-221f10.2 | XP\_021328455 | XM\_021472780 | ZDB-GENE-030131-8507 | wu:fb13g09 |
| A0A0R4IV86 | TrEMBL | ZDB-GENE-140106-124 | si:ch211-221f10.2 | XP\_021328456 | XM\_021472781 | ZDB-GENE-030131-8507 | wu:fb13g09 |
| A0A0R4IV86 | TrEMBL | ZDB-GENE-140106-124 | si:ch211-221f10.2 | XP\_021329979 | XM\_021474304 | ZDB-GENE-030131-8507 | wu:fb13g09 |
| A0A0R4IVJ0 | TrEMBL | ZDB-GENE-030131-9631 | ogt.1 | XP\_001921543 | XM\_001921508 | ZDB-GENE-051128-1 | ogt.2 |
| A0A0R4IVW1 | TrEMBL | ZDB-GENE-050208-210 | si:ch73-217b7.1 | XP\_021325713 | XM\_021470038 | ZDB-GENE-090313-100 | map7b |
| A0A0R4IWI1 | SwissProt | ZDB-GENE-070112-292 | ebf3a | XP\_009289484 | XM\_009291209 | ZDB-GENE-081028-51 | ebf1a |
| A0A0R4IWI1 | SwissProt | ZDB-GENE-070112-292 | ebf3a | XP\_009289485 | XM\_009291210 | ZDB-GENE-081028-51 | ebf1a |
| A0A0R4IZK4 | TrEMBL | ZDB-GENE-141216-18 | si:dkey-225f5.5 | XP\_001340413 | XM\_001340377 | ZDB-GENE-140820-4 | mhc1lca |
| A0A0R4IZN4 | TrEMBL | ZDB-GENE-050809-146 | zgc:109744 | XP\_017207674 | XM\_017352185 | ZDB-GENE-030131-2552 | slc52a2 |
| A0A140LG23 | TrEMBL | ZDB-GENE-160113-53 | si:dkey-23a13.21 | XP\_009301168 | XM\_009302893 | ZDB-GENE-030722-8 | hist2h3ca1 |
| A0A140LGW1 | TrEMBL | ZDB-GENE-160113-120 | si:rp71-46j2.7 | XP\_003199962 | XM\_003199914 | ZDB-LINCRNAG-030616-71 | si:rp71-46j2.3 |
| A0A140LH17 | TrEMBL | ZDB-GENE-040728-3 | pdcd7 | XP\_002667041 | XM\_002666995 | ZDB-GENE-120406-5 | ubap1lb |
| A0A1D5NS55 | TrEMBL | ZDB-GENE-160728-28 | si:ch211-235p24.2 | XP\_017212528 | XM\_017357039 | ZDB-GENE-160215-1 | ric3a |
| A0A1D5NSB4 | TrEMBL | ZDB-GENE-030131-3671 | adgrg1 | XP\_005166455 | XM\_005166398 | ZDB-GENE-160728-126 | si:dkey-181f22.4 |
| A0A1L1QZT6 | TrEMBL | ZDB-GENE-071004-26 | zgc:171727 | XP\_009299309 | XM\_009301034 | ZDB-GENE-161017-48 | si:cabz01021428.1 |
| A0A286YAZ5 | TrEMBL | ZDB-GENE-040426-2772 | hdac8 | XP\_005166576 | XM\_005166519 | ZDB-GENE-031118-56 | phka1a |
| A0A286YB74 | TrEMBL | ZDB-GENE-040912-7 | lyn | XP\_003197913 | XM\_003197865 | ZDB-GENE-070802-2 | tgs1 |
| A0A286YB74 | TrEMBL | ZDB-GENE-040912-7 | lyn | XP\_005163470 | XM\_005163413 | ZDB-GENE-070802-2 | tgs1 |
| A0A2R8PYF9 | TrEMBL | ZDB-GENE-060503-121 | si:ch211-216l23.1 | XP\_005159187 | XM\_005159130 | ZDB-GENE-030131-4368 | znf423 |
| A0A2R8PYX0 | TrEMBL | ZDB-GENE-091204-463 | si:ch73-63e15.2 | XP\_009294614 | XM\_009296339 | ZDB-GENE-060130-85 | sbno2b |
| A0A2R8Q097 | TrEMBL | ZDB-GENE-030131-7310 | eef1b2 | NP\_001313402 | NM\_001326473 | ZDB-GENE-101101-1 | zdbf2 |
| A0A2R8QAK7 | TrEMBL | ZDB-GENE-030131-6520 | mrpl39 | NP\_001314937 | NM\_001328008 | ZDB-GENE-200617-1 | cenpq |
| A0A2R8QK04 | TrEMBL | ZDB-GENE-030131-657 | cyth1a | XP\_021330428 | XM\_021474753 | ZDB-GENE-130530-529 | cyth3b |
| A0A2R8QLP7 | TrEMBL | ZDB-GENE-050411-48 | im:7142702 | XP\_002667906 | XM\_002667860 | ZDB-GENE-110201-2 | bcl11bb |
| A0A2R8QLQ1 | TrEMBL | ZDB-GENE-040718-54 | ppfibp2b | XP\_001342693 | XM\_001342657 | ZDB-GENE-090422-2 | olfml1 |
| A0A2R8QMY3 | TrEMBL | ZDB-GENE-030616-83 | si:rp71-79p20.2 | XP\_002666496 | XM\_002666450 | ZDB-GENE-090313-332 | slc2a4rg |
| A0A2U3TVM2 | TrEMBL | ZDB-GENE-060503-121 | si:ch211-216l23.1 | XP\_005159186 | XM\_005159129 | ZDB-GENE-030131-4368 | znf423 |
| A0A8M1Q7N9 | TrEMBL | ZDB-GENE-091204-463 | si:ch73-63e15.2 | XP\_001342116 | XM\_001342080 | ZDB-GENE-060130-85 | sbno2b |
| A0A8M1Q7N9 | TrEMBL | ZDB-GENE-091204-463 | si:ch73-63e15.2 | XP\_001342116 | XM\_001342080 | ZDB-GENE-060130-85 | sbno2b |
| A0A8M1Q7N9 | TrEMBL | ZDB-GENE-091204-463 | si:ch73-63e15.2 | XP\_009294613 | XM\_009296338 | ZDB-GENE-060130-85 | sbno2b |
| A0A8M1Q7N9 | TrEMBL | ZDB-GENE-091204-463 | si:ch73-63e15.2 | XP\_009294613 | XM\_009296338 | ZDB-GENE-060130-85 | sbno2b |
| A0A8M1QR37 | TrEMBL | ZDB-GENE-131121-97 | si:dkey-172k15.3 | XP\_001922501 | XM\_001922466 | ZDB-GENE-161017-82 | si:dkey-172k15.13 |
| A0A8M1QR37 | TrEMBL | ZDB-GENE-131121-97 | si:dkey-172k15.3 | XP\_001922501 | XM\_001922466 | ZDB-GENE-161017-82 | si:dkey-172k15.13 |
| A0A8M1QR37 | TrEMBL | ZDB-GENE-131121-97 | si:dkey-172k15.3 | XP\_021331505 | XM\_021475830 | ZDB-GENE-161017-82 | si:dkey-172k15.13 |
| A0A8M1RIN8 | TrEMBL | ZDB-GENE-131127-375 | si:ch211-194e18.2 | XP\_003198798 | XM\_003198750 | ZDB-GENE-130530-709 | pld5 |
| A0A8M1RIN8 | TrEMBL | ZDB-GENE-131127-375 | si:ch211-194e18.2 | XP\_003198798 | XM\_003198750 | ZDB-GENE-130530-709 | pld5 |
| A0A8M1RIN8 | TrEMBL | ZDB-GENE-131127-375 | si:ch211-194e18.2 | XP\_021332663 | XM\_021476988 | ZDB-GENE-130530-709 | pld5 |
| A0A8M2BEI4 | TrEMBL | ZDB-GENE-141211-7 | si:ch211-269m15.3 | XP\_005167155 | XM\_005167098 | ZDB-GENE-130530-814 | znf335 |
| A0A8M2BEI4 | TrEMBL | ZDB-GENE-141211-7 | si:ch211-269m15.3 | XP\_005167155 | XM\_005167098 | ZDB-GENE-130530-814 | znf335 |
| A0A8M2BEI4 | TrEMBL | ZDB-GENE-141211-7 | si:ch211-269m15.3 | XP\_005167156 | XM\_005167099 | ZDB-GENE-130530-814 | znf335 |
| A0A8M3APS4 | TrEMBL | ZDB-GENE-110913-104 | si:ch211-103e16.5 | XP\_009296335 | XM\_009298060 | ZDB-GENE-081223-1 | lrrc56 |
| A0A8M3APS4 | TrEMBL | ZDB-GENE-110913-104 | si:ch211-103e16.5 | XP\_009296335 | XM\_009298060 | ZDB-GENE-081223-1 | lrrc56 |
| A0A8M3APS4 | TrEMBL | ZDB-GENE-110913-104 | si:ch211-103e16.5 | XP\_017209805 | XM\_017354316 | ZDB-GENE-081223-1 | lrrc56 |
| A0A8M3APS4 | TrEMBL | ZDB-GENE-110913-104 | si:ch211-103e16.5 | XP\_017209805 | XM\_017354316 | ZDB-GENE-081223-1 | lrrc56 |
| A0A8M3AYM0 | TrEMBL | ZDB-GENE-131119-12 | si:dkey-239b22.1 | XP\_009305155 | XM\_009306880 | ZDB-GENE-131126-52 | si:ch211-39f2.3 |
| A0A8M3AYM0 | TrEMBL | ZDB-GENE-131119-12 | si:dkey-239b22.1 | XP\_009305155 | XM\_009306880 | ZDB-GENE-131126-52 | si:ch211-39f2.3 |
| A0A8M3AYM0 | TrEMBL | ZDB-GENE-131119-12 | si:dkey-239b22.1 | XP\_017214053 | XM\_017358564 | ZDB-GENE-131126-52 | si:ch211-39f2.3 |
| A0A8M6Z983 | TrEMBL | ZDB-GENE-061009-48 | si:dkeyp-2e4.3 | XP\_017214342 | XM\_017358853 | ZDB-GENE-061009-51 | si:dkeyp-2e4.7 |
| A0A8M6Z983 | TrEMBL | ZDB-GENE-061009-48 | si:dkeyp-2e4.3 | XP\_017214342 | XM\_017358853 | ZDB-GENE-061009-51 | si:dkeyp-2e4.7 |
| A0A8M6Z983 | TrEMBL | ZDB-GENE-061009-48 | si:dkeyp-2e4.3 | XP\_021336570 | XM\_021480895 | ZDB-GENE-061009-51 | si:dkeyp-2e4.7 |
| A0A8M9QA37 | TrEMBL | ZDB-GENE-081028-2 | pimr155 | XP\_009305397 | XM\_009307122 | ZDB-GENE-081028-10 | si:dkey-265c15.9 |
| A0A8M9QD30 | TrEMBL | ZDB-GENE-110913-180 | si:ch73-266f23.1 | XP\_009299290 | XM\_009301015 | ZDB-GENE-080219-37 | zgc:171604 |
| A0A8M9QD30 | TrEMBL | ZDB-GENE-110913-180 | si:ch73-266f23.1 | XP\_021331480 | XM\_021475805 | ZDB-GENE-080219-37 | zgc:171604 |
| A2BHM3 | TrEMBL | ZDB-GENE-030131-1250 | wu:fb64b08 | NP\_001315076 | NM\_001328147 | ZDB-GENE-060526-147 | si:ch211-271e10.3 |
| A2BHM3 | TrEMBL | ZDB-GENE-030131-1250 | wu:fb64b08 | XP\_021331949 | XM\_021476274 | ZDB-GENE-060526-147 | si:ch211-271e10.3 |
| A2CEY2 | TrEMBL | ZDB-GENE-061009-48 | si:dkeyp-2e4.3 | NP\_001121755 | NM\_001128283 | ZDB-GENE-061009-50 | si:dkeyp-2e4.6 |
| A3KP86 | TrEMBL | ZDB-GENE-070410-131 | zgc:163057 | NP\_001076303 | NM\_001082834 | ZDB-GENE-980526-286 | hbae4 |
| A3KPG5 | TrEMBL | ZDB-GENE-030131-5826 | crp2 | XP\_017209283 | XM\_017353794 | ZDB-GENE-060503-220 | crp3 |
| A3KPG5 | TrEMBL | ZDB-GENE-030131-5826 | SO:0001217 | XP\_017209283 | XM\_017353794 | ZDB-GENE-060503-220 | crp3 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | NP\_001083038 | NM\_001089569 | ZDB-GENE-070424-79 | zgc:163040 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | NP\_001099176 | NM\_001105706 | ZDB-GENE-070927-10 | hist1h4l |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_002666964 | XM\_002666918 | ZDB-GENE-131121-158 | si:dkey-108k21.28 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_005163249 | XM\_005163192 | ZDB-GENE-121214-170 | si:ch211-113a14.6 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_009296456 | XM\_009298181 | ZDB-GENE-131121-139 | si:dkey-108k21.18 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_009301164 | XM\_009302889 | ZDB-GENE-160113-88 | si:dkey-23a13.7 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_017209641 | XM\_017354152 | ZDB-GENE-070927-10 | hist1h4l |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_017209642 | XM\_017354153 | ZDB-GENE-131127-177 | si:ch73-36p18.4 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_017209670 | XM\_017354181 | ZDB-GENE-121214-197 | si:ch211-113a14.21 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_021326437 | XM\_021470762 | ZDB-GENE-121214-196 | si:ch211-113a14.26 |
| A3KPR4 | TrEMBL | ZDB-GENE-070620-17 | zgc:165555 | XP\_688877 | XM\_683785 | ZDB-GENE-121214-187 | si:ch211-113a14.17 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | NP\_001083038 | NM\_001089569 | ZDB-GENE-070424-79 | zgc:163040 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | NP\_001092231 | NM\_001098761 | ZDB-GENE-070620-17 | zgc:165555 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_002666964 | XM\_002666918 | ZDB-GENE-131121-158 | si:dkey-108k21.28 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_005163249 | XM\_005163192 | ZDB-GENE-121214-170 | si:ch211-113a14.6 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_009296456 | XM\_009298181 | ZDB-GENE-131121-139 | si:dkey-108k21.18 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_009301164 | XM\_009302889 | ZDB-GENE-160113-88 | si:dkey-23a13.7 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_017209642 | XM\_017354153 | ZDB-GENE-131127-177 | si:ch73-36p18.4 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_017209670 | XM\_017354181 | ZDB-GENE-121214-197 | si:ch211-113a14.21 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_021326437 | XM\_021470762 | ZDB-GENE-121214-196 | si:ch211-113a14.26 |
| A3KPR4 | TrEMBL | ZDB-GENE-070927-10 | hist1h4l | XP\_688877 | XM\_683785 | ZDB-GENE-121214-187 | si:ch211-113a14.17 |
| A4FVM2 | TrEMBL | ZDB-GENE-030131-6292 | hist1h2a6 | XP\_001334871 | XM\_001334835 | ZDB-GENE-160113-30 | si:ch73-368j24.13 |
| A4IGB0 | TrEMBL | ZDB-GENE-141215-12 | si:ch73-42k18.1 | NP\_001077286 | NM\_001083817 | ZDB-GENE-040724-246 | si:ch211-155i14.1 |
| A4IGE9 | TrEMBL | ZDB-GENE-120214-9 | pimr161 | NP\_001077043 | NM\_001083574 | ZDB-GENE-070424-73 | pimr166 |
| A5WUN3 | TrEMBL | ZDB-GENE-120214-42 | si:dkey-13n23.3 | NP\_001093542 | NM\_001100072 | ZDB-GENE-070705-527 | si:dkeyp-106c3.2 |
| A5WV15 | TrEMBL | ZDB-GENE-030131-6493 | acsl4b | NP\_001002050 | NM\_001002050 | ZDB-GENE-040625-10 | cwc15 |
| A6H8Q4 | TrEMBL | ZDB-GENE-070720-9 | crygm2d17 | XP\_002663401 | XM\_002663355 | ZDB-GENE-081105-90 | crygm2d19 |
| A6H8Q4 | TrEMBL | ZDB-GENE-070720-9 | crygm2d17 | XP\_021334574 | XM\_021478899 | ZDB-GENE-081105-90 | crygm2d19 |
| A7MBT5 | TrEMBL | ZDB-GENE-040801-183 | rpl15 | NP\_001096586 | NM\_001103116 | ZDB-GENE-071004-16 | ube2e1 |
| A7MBT5 | TrEMBL | ZDB-GENE-040801-183 | rpl15 | XP\_005159563 | XM\_005159506 | ZDB-GENE-071004-16 | ube2e1 |
| A7MC74 | TrEMBL | ZDB-GENE-071004-51 | zgc:173425 | NP\_001098612 | NM\_001105142 | ZDB-GENE-081028-17 | si:dkey-33c12.13 |
| A7MCM3 | TrEMBL | ZDB-GENE-980526-285 | ctslb | XP\_003199634 | XM\_003199586 | ZDB-GENE-121214-36 | si:dkey-26g8.4 |
| A7MCR8 | TrEMBL | ZDB-GENE-980526-285 | ctslb | NP\_001096585 | NM\_001103115 | ZDB-GENE-121214-19 | si:dkey-26g8.5 |
| A8DZ97 | TrEMBL | ZDB-GENE-041111-101 | im:7140055 | XP\_017207918 | XM\_017352429 | ZDB-GENE-060531-2 | pals1b |
| A8E587 | TrEMBL | ZDB-GENE-980526-285 | ctslb | NP\_001099150 | NM\_001105680 | ZDB-GENE-121214-37 | si:dkey-269i1.4 |
| A8KB68 | TrEMBL | ZDB-GENE-071004-25 | zgc:171717 | XP\_021325279 | XM\_021469604 | ZDB-GENE-070705-50 | si:ch211-154a22.8 |
| A8KB68 | TrEMBL | ZDB-GENE-071004-25 | zgc:171717 | XP\_021329182 | XM\_021473507 | ZDB-GENE-070705-50 | si:ch211-154a22.8 |
| A8KBJ5 | TrEMBL | ZDB-GENE-080220-24 | zgc:173552 | XP\_009296488 | XM\_009298213 | ZDB-GENE-121214-143 | si:ch211-113a14.27 |
| A8KBJ5 | TrEMBL | ZDB-GENE-080220-24 | zgc:173552 | XP\_009301184 | XM\_009302909 | ZDB-GENE-160113-117 | si:ch1073-153i20.2 |
| A8KBM3 | TrEMBL | ZDB-GENE-110913-19 | si:dkey-156k2.3 | XP\_017211162 | XM\_017355673 | ZDB-GENE-080208-5 | zgc:173709 |
| A8WG67 | TrEMBL | ZDB-GENE-050208-695 | si:ch211-226h8.8 | NP\_001103856 | NM\_001110386 | ZDB-GENE-050208-629 | si:ch211-226h8.11 |
| A8WG67 | TrEMBL | ZDB-GENE-050208-695 | si:ch211-226h8.8 | XP\_003201119 | XM\_003201071 | ZDB-GENE-080220-1 | zgc:171597 |
| A8WG67 | TrEMBL | ZDB-GENE-050208-695 | si:ch211-226h8.8 | XP\_021329181 | XM\_021473506 | ZDB-GENE-080220-1 | zgc:171597 |
| A8WG99 | TrEMBL | ZDB-GENE-041210-203 | acot15 | XP\_005164881 | XM\_005164824 | ZDB-GENE-041210-254 | acot18 |
| A8WHX5 | TrEMBL | ZDB-GENE-070912-707 | si:rp71-45g20.10 | XP\_001920294 | XM\_001920259 | ZDB-GENE-070912-708 | si:rp71-45g20.11 |
| A8WHX5 | TrEMBL | ZDB-GENE-070912-707 | si:rp71-45g20.10 | XP\_021334633 | XM\_021478958 | ZDB-GENE-070912-708 | si:rp71-45g20.11 |
| A8WHX5 | TrEMBL | ZDB-GENE-070912-708 | si:rp71-45g20.11 | XP\_003199271 | XM\_003199223 | ZDB-GENE-070912-707 | si:rp71-45g20.10 |
| A8WHX5 | TrEMBL | ZDB-GENE-070912-708 | si:rp71-45g20.11 | XP\_021334634 | XM\_021478959 | ZDB-GENE-070912-707 | si:rp71-45g20.10 |
| A9JTE3 | TrEMBL | ZDB-GENE-090313-304 | nckap5l | XP\_005162220 | XM\_005162163 | ZDB-GENE-041010-67 | kansl2 |
| B0R190 | TrEMBL | ZDB-GENE-060526-78 | cxcl11.6 | XP\_005165469 | XM\_005165412 | ZDB-GENE-060526-79 | cxcl11.7 |
| B0R190 | TrEMBL | ZDB-GENE-060526-78 | cxcl11.6 | XP\_021327340 | XM\_021471665 | ZDB-GENE-060526-79 | cxcl11.7 |
| B0S5E6 | TrEMBL | ZDB-GENE-070912-459 | pimr49 | XP\_021322781 | XM\_021467106 | ZDB-GENE-070912-426 | pimr53 |
| B0S638 | TrEMBL | ZDB-GENE-070912-666 | si:dkeyp-67a8.4 | XP\_694045 | XM\_688953 | ZDB-GENE-070912-138 | si:ch211-165f21.6 |
| B0S673 | TrEMBL | ZDB-GENE-040625-39 | rpl37 | XP\_017207846 | XM\_017352357 | ZDB-GENE-070912-694 | coa1 |
| B0S7I1 | TrEMBL | ZDB-GENE-081104-23 | abo | NP\_001116741 | NM\_001123269 | ZDB-GENE-080402-11 | gbgt1l5 |
| B0UXR9 | TrEMBL | ZDB-GENE-110405-1 | hsp70.2 | XP\_003198158 | XM\_003198110 | ZDB-GENE-990415-91 | hsp70.1 |
| B2GPI5 | TrEMBL | ZDB-GENE-010507-2 | clic3 | NP\_001074124 | NM\_001080655 | ZDB-GENE-070112-1812 | pnpla7a |
| B2GQW3 | TrEMBL | ZDB-GENE-030131-5590 | calm3a | NP\_955864 | NM\_199570 | ZDB-GENE-030131-527 | calm3b |
| B2GQW3 | TrEMBL | ZDB-GENE-030131-5590 | calm3a | NP\_956290 | NM\_199996 | ZDB-GENE-030804-3 | calm2a |
| B2GQW3 | TrEMBL | ZDB-GENE-030131-5590 | calm3a | NP\_956376 | NM\_200082 | ZDB-GENE-030804-2 | calm1b |
| B2GQW3 | TrEMBL | ZDB-GENE-030131-5590 | calm3a | NP\_998516 | NM\_213351 | ZDB-GENE-030131-8308 | calm1a |
| B2GQW3 | TrEMBL | ZDB-GENE-030131-5590 | calm3a | NP\_999901 | NM\_214736 | ZDB-GENE-020415-2 | calm2b |
| B2GSF5 | TrEMBL | ZDB-GENE-050417-65 | h3f3b.1 | NP\_956297 | NM\_200003 | ZDB-GENE-030131-8175 | h3f3c |
| B2GSF5 | TrEMBL | ZDB-GENE-050417-65 | h3f3b.1 | NP\_957395 | NM\_201101 | ZDB-GENE-040426-1023 | h3f3d |
| B2GSF5 | TrEMBL | ZDB-GENE-050417-65 | h3f3b.1 | XP\_002664801 | XM\_002664755 | ZDB-GENE-120214-5 | si:ch1073-429i10.3 |
| B2GSF5 | TrEMBL | ZDB-GENE-050417-65 | h3f3b.1 | XP\_009290278 | XM\_009292003 | ZDB-GENE-120214-5 | si:ch1073-429i10.3 |
| B2GSF5 | TrEMBL | ZDB-GENE-050417-65 | h3f3b.1 | XP\_017209248 | XM\_017353759 | ZDB-GENE-040426-1928 | h3f3a |
| B3DGB1 | TrEMBL | ZDB-GENE-050610-7 | pcdh2g17 | NP\_001019372 | NM\_001024201 | ZDB-GENE-050610-20 | pcdh2g29 |
| B3DH93 | TrEMBL | ZDB-GENE-050522-428 | fthl31 | NP\_001124139 | NM\_001130667 | ZDB-GENE-080722-16 | fthl29 |
| B3DH93 | TrEMBL | ZDB-GENE-050522-428 | fthl31 | XP\_021324849 | XM\_021469174 | ZDB-GENE-080722-16 | fthl29 |
| B3DHP3 | TrEMBL | ZDB-GENE-080723-25 | zgc:194215 | NP\_001124121 | NM\_001130649 | ZDB-GENE-080723-51 | zgc:194186 |
| B3DHP3 | TrEMBL | ZDB-GENE-080723-51 | zgc:194186 | NP\_001124102 | NM\_001130630 | ZDB-GENE-080723-25 | zgc:194215 |
| B5DDQ4 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001170815 | NM\_001177344 | ZDB-GENE-080227-3 | ugt1a4 |
| B8A5A9 | TrEMBL | ZDB-GENE-081028-8 | pimr152 | XP\_017214435 | XM\_017358946 | ZDB-GENE-081028-10 | si:dkey-265c15.9 |
| B8JI22 | TrEMBL | ZDB-GENE-090313-377 | si:dkeyp-13d11.1 | XP\_017208182 | XM\_017352693 | ZDB-GENE-130531-32 | si:dkey-6f10.5 |
| B8JM89 | TrEMBL | ZDB-GENE-081105-8 | si:dkey-56i24.1 | XP\_002663252 | XM\_002663206 | ZDB-GENE-081105-10 | si:ch73-211l2.1 |
| E7EXD7 | TrEMBL | ZDB-GENE-121214-198 | si:ch211-113a14.19 | XP\_009296467 | XM\_009298192 | ZDB-GENE-131127-102 | si:dkey-261m9.11 |
| E7EXD7 | TrEMBL | ZDB-GENE-121214-198 | si:ch211-113a14.19 | XP\_009296490 | XM\_009298215 | ZDB-GENE-121214-152 | si:ch211-113a14.25 |
| E7EXM4 | TrEMBL | ZDB-GENE-030131-738 | si:ch211-236h17.3 | XP\_694030 | XM\_688938 | ZDB-GENE-200318-3 | slc41a3 |
| E7EY81 | TrEMBL | ZDB-GENE-030131-4623 | olfm4.2 | XP\_697903 | XM\_692811 | ZDB-GENE-131120-11 | olfm4.1 |
| E7EZQ8 | TrEMBL | ZDB-GENE-100406-5 | ugt5b3 | NP\_001170970 | NM\_001177499 | ZDB-GENE-100406-4 | ugt5b2 |
| E7F016 | TrEMBL | ZDB-GENE-130530-933 | zmp:0000000930 | XP\_021322425 | XM\_021466750 | ZDB-GENE-060825-61 | ccdc127a |
| E7F1C0 | TrEMBL | ZDB-GENE-131127-163 | si:dkey-178o16.4 | XP\_003198871 | XM\_003198823 | ZDB-GENE-121023-1 | avpr2b.1 |
| E7F1C0 | TrEMBL | ZDB-GENE-131127-163 | si:dkey-178o16.4 | XP\_021327514 | XM\_021471839 | ZDB-GENE-121023-1 | avpr2b.1 |
| E7F4B5 | TrEMBL | ZDB-GENE-131121-60 | si:dkey-27o4.1 | XP\_689861 | XM\_684769 | ZDB-GENE-110922-6 | fam83ga |
| E7F4R5 | TrEMBL | ZDB-GENE-040426-2772 | hdac8 | XP\_009301729 | XM\_009303454 | ZDB-GENE-031118-56 | phka1a |
| E7F551 | TrEMBL | ZDB-GENE-141215-36 | si:ch211-160o17.4 | XP\_701123 | XM\_696031 | ZDB-GENE-140324-1 | hipk1b |
| E7F5V3 | TrEMBL | ZDB-GENE-090603-1 | ccdc57 | XP\_009305081 | XM\_009306806 | ZDB-GENE-030131-7802 | fasn |
| E7F5V3 | TrEMBL | ZDB-GENE-090603-1 | ccdc57 | XP\_009305082 | XM\_009306807 | ZDB-GENE-030131-7802 | fasn |
| E7F5W3 | TrEMBL | ZDB-GENE-131127-92 | hist1h2a10 | XP\_693579 | XM\_688487 | ZDB-GENE-160113-115 | si:ch1073-159d7.10 |
| E7F6B3 | TrEMBL | ZDB-GENE-110913-144 | si:dkeyp-85d8.5 | XP\_017211059 | XM\_017355570 | ZDB-GENE-141212-227 | si:ch211-142b24.6 |
| E7F6Y5 | TrEMBL | ZDB-GENE-061027-239 | mettl7a.1 | XP\_697500 | XM\_692408 | ZDB-GENE-120215-64 | mettl7a.3 |
| E7F704 | TrEMBL | ZDB-GENE-121214-325 | si:ch1073-272o11.3 | XP\_005169488 | XM\_005169431 | ZDB-GENE-101203-4 | cdr2b |
| E7F787 | TrEMBL | ZDB-GENE-041014-165 | ccl38.6 | XP\_017208203 | XM\_017352714 | ZDB-GENE-110411-218 | ccl38a.3 |
| E7F8K3 | TrEMBL | ZDB-GENE-050222-2 | kcnj11l | XP\_009296055 | XM\_009297780 | ZDB-GENE-131127-232 | si:ch211-286c4.6 |
| E7F8K3 | TrEMBL | ZDB-GENE-050222-2 | kcnj11l | XP\_009296056 | XM\_009297781 | ZDB-GENE-131127-232 | si:ch211-286c4.6 |
| E7F8K3 | TrEMBL | ZDB-GENE-050222-2 | kcnj11l | XP\_017209637 | XM\_017354148 | ZDB-GENE-131127-232 | si:ch211-286c4.6 |
| E7F8Q1 | TrEMBL | ZDB-GENE-120703-49 | si:dkeyp-44b5.5 | XP\_005168082 | XM\_005168025 | ZDB-GENE-080208-6 | zgc:174702 |
| E7F8Q1 | TrEMBL | ZDB-GENE-120703-49 | si:dkeyp-44b5.5 | XP\_021331305 | XM\_021475630 | ZDB-GENE-080208-6 | zgc:174702 |
| E7F965 | TrEMBL | ZDB-GENE-980526-174 | ascl1b | XP\_688932 | XM\_683840 | ZDB-GENE-161017-126 | si:ch211-210p4.6 |
| E7F9B4 | TrEMBL | ZDB-GENE-141216-123 | si:dkey-264d12.4 | XP\_021335574 | XM\_021479899 | ZDB-GENE-150211-1 | emp3a |
| E7F9Y7 | TrEMBL | ZDB-GENE-030131-4199 | tagln3b | XP\_002666688 | XM\_002666642 | ZDB-GENE-041111-196 | abhd10b |
| E7FC94 | TrEMBL | ZDB-GENE-081104-61 | si:ch211-200p22.4 | XP\_001344303 | XM\_001344267 | ZDB-GENE-030131-7935 | wu:fj42b12 |
| E7FD02 | TrEMBL | ZDB-GENE-131125-81 | si:dkey-83h2.5 | XP\_009289708 | XM\_009291433 | ZDB-GENE-130530-576 | ano7 |
| E7FF76 | TrEMBL | ZDB-GENE-050315-1 | adcyap1r1a | XP\_021327469 | XM\_021471794 | ZDB-GENE-110127-1 | adcyap1r1b |
| E7FF76 | TrEMBL | ZDB-GENE-050315-1 | adcyap1r1a | XP\_682980 | XM\_677888 | ZDB-GENE-110127-1 | adcyap1r1b |
| E7FFZ5 | TrEMBL | ZDB-GENE-060503-248 | smad10a | XP\_001922725 | XM\_001922690 | ZDB-GENE-110408-38 | smad10b |
| E7FFZ5 | TrEMBL | ZDB-GENE-060503-248 | smad10a | XP\_005158136 | XM\_005158079 | ZDB-GENE-110408-38 | smad10b |
| E7FGJ8 | TrEMBL | ZDB-GENE-091204-253 | gdpd5a | XP\_009303864 | XM\_009305589 | ZDB-GENE-030131-8810 | wu:fk20b05 |
| E7FGT6 | TrEMBL | ZDB-GENE-030131-7307 | wu:fj05g07 | XP\_017209505 | XM\_017354016 | ZDB-GENE-070424-101 | zgc:163022 |
| E9QDC9 | TrEMBL | ZDB-GENE-091118-98 | SO:0001217 | XP\_001343958 | XM\_001343922 | ZDB-GENE-030131-2924 | wu:fc21g01 |
| E9QDC9 | TrEMBL | ZDB-GENE-091118-98 | ypel2a | XP\_001343958 | XM\_001343922 | ZDB-GENE-030131-2924 | wu:fc21g01 |
| E9QE43 | TrEMBL | ZDB-GENE-060503-165 | si:dkey-28d5.5 | XP\_001919193 | XM\_001919158 | ZDB-GENE-160113-18 | si:ch211-63p21.3 |
| E9QEE5 | TrEMBL | ZDB-GENE-041114-102 | cyp3c3 | XP\_021330384 | XM\_021474709 | ZDB-GENE-081103-62 | cyp3c2 |
| E9QEM8 | TrEMBL | ZDB-GENE-091204-417 | pgbd4 | XP\_005161688 | XM\_005161631 | ZDB-GENE-030131-2555 | wu:fc11c11 |
| E9QFV1 | TrEMBL | ZDB-GENE-030131-1927 | ddx41 | NP\_001315125 | NM\_001328196 | ZDB-GENE-100921-22 | si:ch1073-75o15.3 |
| E9QG09 | TrEMBL | ZDB-GENE-040801-94 | pibf1 | XP\_001344916 | XM\_001344880 | ZDB-GENE-090312-167 | klf5a |
| E9QG10 | TrEMBL | ZDB-GENE-091113-52 | rps6kb1a | XP\_690726 | XM\_685634 | ZDB-GENE-050411-103 | im:7155962 |
| E9QGT6 | TrEMBL | ZDB-GENE-110914-204 | znf1047 | XP\_003198571 | XM\_003198523 | ZDB-GENE-080215-19 | zgc:173713 |
| E9QHK9 | TrEMBL | ZDB-GENE-130530-732 | zmp:0000000729 | XP\_001341915 | XM\_001341879 | ZDB-GENE-070912-285 | hbl1 |
| E9QHK9 | TrEMBL | ZDB-GENE-130530-732 | zmp:0000000729 | XP\_009296948 | XM\_009298673 | ZDB-GENE-070912-285 | hbl1 |
| E9QHZ3 | TrEMBL | ZDB-GENE-091113-21 | si:dkey-84o3.3 | XP\_017208436 | XM\_017352947 | ZDB-GENE-091113-30 | si:dkey-84o3.7 |
| F1Q7N5 | TrEMBL | ZDB-GENE-121214-88 | si:dkeyp-9d4.5 | XP\_017211446 | XM\_017355957 | ZDB-GENE-110922-5 | slc5a10 |
| F1Q918 | TrEMBL | ZDB-GENE-080215-8 | zgc:174193 | XP\_017213555 | XM\_017358066 | ZDB-GENE-141212-238 | si:dkey-10b15.8 |
| F1QAF6 | TrEMBL | ZDB-GENE-080227-6 | SO:0001217 | NP\_001032505 | NM\_001037428 | ZDB-GENE-071004-4 | ugt1a1 |
| F1QAF6 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001032505 | NM\_001037428 | ZDB-GENE-071004-4 | ugt1a1 |
| F1QAT3 | TrEMBL | ZDB-GENE-080303-7 | si:ch211-165d12.4 | XP\_017209594 | XM\_017354105 | ZDB-GENE-050320-8 | nmrk1 |
| F1QAU0 | TrEMBL | ZDB-GENE-071004-32 | zgc:171759 | XP\_001922605 | XM\_001922570 | ZDB-GENE-160113-108 | si:dkey-23a13.1 |
| F1QDR1 | TrEMBL | ZDB-GENE-030131-657 | cyth1a | XP\_017210323 | XM\_017354834 | ZDB-GENE-130530-529 | cyth3b |
| F1QGJ3 | TrEMBL | ZDB-GENE-040426-2056 | zgc:77486 | XP\_005166988 | XM\_005166931 | ZDB-GENE-030131-7218 | plin3 |
| F1QGJ3 | TrEMBL | ZDB-GENE-040426-2056 | zgc:77486 | XP\_021334076 | XM\_021478401 | ZDB-GENE-030131-7218 | plin3 |
| F1QIH9 | TrEMBL | ZDB-GENE-050626-121 | zp3.2 | XP\_005169843 | XM\_005169786 | ZDB-GENE-071004-37 | zgc:171778 |
| F1QIH9 | TrEMBL | ZDB-GENE-071004-54 | zgc:173556 | XP\_005169843 | XM\_005169786 | ZDB-GENE-071004-37 | zgc:171778 |
| F1QJD1 | TrEMBL | ZDB-GENE-030131-3143 | apoa4b.2 | XP\_001338037 | XM\_001338001 | ZDB-GENE-100921-47 | apoa4b.3 |
| F1QM88 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001166241 | NM\_001172770 | ZDB-GENE-071004-5 | ugt1a2 |
| F1QNF8 | TrEMBL | ZDB-GENE-100406-4 | ugt5b2 | XP\_005170903 | XM\_005170846 | ZDB-GENE-100406-3 | ugt5b1 |
| F1QNI1 | TrEMBL | ZDB-GENE-070912-577 | cbln5 | XP\_017207460 | XM\_017351971 | ZDB-GENE-040724-89 | c1ql4l |
| F1QNI1 | TrEMBL | ZDB-GENE-070912-577 | cbln5 | XP\_021326933 | XM\_021471258 | ZDB-GENE-040724-89 | c1ql4l |
| F1QRI1 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001170810 | NM\_001177339 | ZDB-GENE-080227-5 | ugt1a6 |
| F1QRJ7 | TrEMBL | ZDB-GENE-110913-41 | si:dkey-199m13.4 | XP\_003198383 | XM\_003198335 | ZDB-GENE-071004-91 | zgc:173716 |
| F1QRX6 | TrEMBL | ZDB-GENE-070424-265 | si:dkey-88n24.10 | NP\_001314984 | NM\_001328055 | ZDB-GENE-041014-86 | si:dkey-88n24.9 |
| F1QTL1 | TrEMBL | ZDB-GENE-081205-1 | hist1h2a11 | XP\_001335390 | XM\_001335354 | ZDB-GENE-160113-112 | h2ax2 |
| F1QTL1 | TrEMBL | ZDB-GENE-081205-1 | hist1h2a11 | XP\_001337024 | XM\_001336988 | ZDB-GENE-160113-109 | si:dkey-23a13.20 |
| F1QTL1 | TrEMBL | ZDB-GENE-081205-1 | hist1h2a11 | XP\_001920286 | XM\_001920251 | ZDB-GENE-160113-102 | si:ch1073-159d7.11 |
| F1QTL1 | TrEMBL | ZDB-GENE-081205-1 | hist1h2a11 | XP\_009301187 | XM\_009302912 | ZDB-GENE-160113-105 | hist1h2a8 |
| F1QW80 | TrEMBL | ZDB-GENE-161017-27 | si:ch73-120g24.4 | XP\_009299293 | XM\_009301018 | ZDB-GENE-161017-9 | si:ch73-120g24.5 |
| F1QWI8 | TrEMBL | ZDB-GENE-050522-333 | trim35-7 | XP\_005163760 | XM\_005163703 | ZDB-GENE-131119-18 | si:ch73-106l15.3 |
| F1QYD4 | TrEMBL | ZDB-GENE-151216-1 | ccr2 | XP\_001344220 | XM\_001344184 | ZDB-GENE-160113-21 | si:ch211-207g17.3 |
| F1QZA2 | TrEMBL | ZDB-GENE-030616-83 | si:rp71-79p20.2 | XP\_005162071 | XM\_005162014 | ZDB-GENE-090313-332 | slc2a4rg |
| F1R096 | TrEMBL | ZDB-GENE-030131-6779 | gtf2h4 | XP\_687006 | XM\_681914 | ZDB-GENE-060503-310 | sh3bp5lb |
| F1R0T3 | TrEMBL | ZDB-GENE-100922-162 | si:dkey-40m6.8 | XP\_005158431 | XM\_005158374 | ZDB-GENE-100812-13 | lmtk3 |
| F1R1P2 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_998587 | NM\_213422 | ZDB-GENE-040426-2762 | ugt1ab |
| F1R4A4 | TrEMBL | ZDB-GENE-071004-32 | zgc:171759 | XP\_005163255 | XM\_005163198 | ZDB-GENE-121214-207 | si:ch211-113a14.15 |
| F1R4A4 | TrEMBL | ZDB-GENE-071004-32 | zgc:171759 | XP\_017209640 | XM\_017354151 | ZDB-GENE-131127-87 | si:dkey-108k21.27 |
| F1R707 | TrEMBL | ZDB-GENE-140106-283 | mfap4.6 | XP\_009294456 | XM\_009296181 | ZDB-GENE-110411-23 | si:zfos-2330d3.6 |
| F1R891 | TrEMBL | ZDB-GENE-040912-88 | msrb1b | XP\_021335967 | XM\_021480292 | ZDB-GENE-120104-8 | meiob |
| F1R9N6 | TrEMBL | ZDB-GENE-121214-99 | si:ch211-113a14.12 | NP\_001013495 | NM\_001013477 | ZDB-GENE-050320-42 | zgc:110216 |
| F1RB61 | TrEMBL | ZDB-GENE-060414-5 | taar20i | XP\_009303900 | XM\_009305625 | ZDB-GENE-060414-19 | taar20l |
| F1RCB6 | TrEMBL | ZDB-GENE-040520-4 | actc1a | XP\_009296202 | XM\_009297927 | ZDB-GENE-040625-46 | acte1 |
| F1RCB6 | TrEMBL | ZDB-GENE-040520-4 | actc1a | XP\_021329509 | XM\_021473834 | ZDB-GENE-040625-46 | acte1 |
| F1REH3 | TrEMBL | ZDB-GENE-080218-7 | znf1085 | XP\_017211205 | XM\_017355716 | ZDB-GENE-141212-233 | znf1086 |
| F6NGD6 | TrEMBL | ZDB-GENE-041212-67 | zgc:101540 | XP\_001336957 | XM\_001336921 | ZDB-GENE-081104-49 | slc27a2b |
| F6NN09 | TrEMBL | ZDB-GENE-070806-48 | or92a4 | XP\_017208527 | XM\_017353038 | ZDB-GENE-081031-102 | si:dkey-92k1.9 |
| F6NZA7 | TrEMBL | ZDB-GENE-141212-252 | si:ch73-160i9.3 | XP\_001332899 | XM\_001332863 | ZDB-GENE-081031-42 | si:ch211-171h4.6 |
| F6PCD2 | TrEMBL | ZDB-GENE-050208-457 | ifi44f5 | XP\_003201083 | XM\_003201035 | ZDB-GENE-050208-616 | si:ch211-197g15.9 |
| F6PCI7 | TrEMBL | ZDB-GENE-081104-12 | si:ch211-74m13.1 | XP\_017210320 | XM\_017354831 | ZDB-GENE-060503-219 | icam3 |
| F8W4J0 | TrEMBL | ZDB-GENE-041014-165 | ccl38.6 | XP\_003200783 | XM\_003200735 | ZDB-GENE-110411-218 | ccl38a.3 |
| F8W539 | TrEMBL | ZDB-GENE-030131-7370 | wu:fj11g02 | XP\_002660664 | XM\_002660618 | ZDB-GENE-110408-28 | mfap4.13 |
| H0WE71 | TrEMBL | ZDB-GENE-110913-49 | si:dkey-72l17.6 | XP\_017210990 | XM\_017355501 | ZDB-GENE-050306-7 | zgc:113135 |
| H0WF55 | TrEMBL | ZDB-GENE-080218-3 | znf1078 | XP\_009299032 | XM\_009300757 | ZDB-GENE-110913-64 | znf1076 |
| I3IT29 | TrEMBL | ZDB-GENE-120215-63 | si:dkey-151m15.5 | XP\_003198160 | XM\_003198112 | ZDB-GENE-100311-1 | rundc1 |
| I6LC33 | TrEMBL | ZDB-GENE-051113-68 | zgc:123181 | NP\_001019272 | NM\_001024101 | ZDB-GENE-050203-2 | pcdh1a6 |
| K7DY30 | TrEMBL | ZDB-GENE-120709-3 | si:dkey-16p19.1 | XP\_005168225 | XM\_005168168 | ZDB-GENE-071004-63 | zgc:173721 |
| K7DYJ1 | TrEMBL | ZDB-GENE-080215-13 | znf995 | XP\_017211118 | XM\_017355629 | ZDB-GENE-120703-28 | si:dkey-22h13.1 |
| K7DYJ1 | TrEMBL | ZDB-GENE-080215-13 | znf995 | XP\_021331235 | XM\_021475560 | ZDB-GENE-120703-28 | si:dkey-22h13.1 |
| O73712 | TrEMBL | ZDB-GENE-990415-138 | ldb1a | NP\_571388 | NM\_131313 | ZDB-GENE-990415-135 | ldb1b |
| O93548 | TrEMBL | ZDB-GENE-090501-3 | hbbe1.2 | XP\_005164391 | XM\_005164334 | ZDB-GENE-131120-128 | hbbe1.3 |
| O93548 | TrEMBL | ZDB-GENE-131120-128 | hbbe1.3 | NP\_001096600 | NM\_001103130 | ZDB-GENE-090501-3 | hbbe1.2 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_001335214 | XM\_001335178 | ZDB-GENE-160113-100 | si:ch73-368j24.10 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_002667787 | XM\_002667741 | ZDB-GENE-160113-113 | si:ch1073-159d7.5 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_003198935 | XM\_003198887 | ZDB-GENE-160113-114 | si:ch1073-159d7.6 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_005168860 | XM\_005168803 | ZDB-GENE-030131-4921 | si:ch1073-159d7.13 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_005170787 | XM\_005170730 | ZDB-GENE-131121-71 | si:dkey-108k21.13 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_009301169 | XM\_009302894 | ZDB-GENE-160113-83 | si:dkey-23a13.16 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_009301182 | XM\_009302907 | ZDB-GENE-160113-81 | si:ch73-368j24.9 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_017212444 | XM\_017356955 | ZDB-GENE-160113-111 | si:ch1073-153i20.3 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_021333351 | XM\_021477676 | ZDB-GENE-160113-44 | si:dkey-23a13.5 |
| Q05AK8 | TrEMBL | ZDB-GENE-071004-46 | ano9a | XP\_692972 | XM\_687880 | ZDB-GENE-160113-28 | si:ch73-368j24.8 |
| Q15I85 | TrEMBL | ZDB-GENE-060918-2 | crygm2d3 | NP\_001082900 | NM\_001089431 | ZDB-GENE-060918-3 | crygm2d4 |
| Q15I85 | TrEMBL | ZDB-GENE-060918-3 | crygm2d4 | NP\_001073530 | NM\_001080061 | ZDB-GENE-060918-2 | crygm2d3 |
| Q1LVW4 | TrEMBL | ZDB-GENE-050208-760 | si:dkeyp-98a7.4 | NP\_001103581 | NM\_001110111 | ZDB-GENE-050208-360 | si:dkeyp-98a7.5 |
| Q1LVW4 | TrEMBL | ZDB-GENE-050208-760 | si:dkeyp-98a7.4 | NP\_001153834 | NM\_001160362 | ZDB-GENE-050208-679 | si:dkeyp-98a7.3 |
| Q1LVW4 | TrEMBL | ZDB-GENE-050208-760 | si:dkeyp-98a7.4 | XP\_003201138 | XM\_003201090 | ZDB-GENE-050208-520 | si:dkeyp-98a7.7 |
| Q1LW00 | TrEMBL | ZDB-GENE-030131-2100 | he1.2 | NP\_001038639 | NM\_001045174 | ZDB-GENE-021211-3 | he1.1 |
| Q1LWV8 | TrEMBL | ZDB-GENE-050419-140 | meak7 | NP\_001038433 | NM\_001044968 | ZDB-GENE-030131-9532 | plcg2 |
| Q1MTC7 | TrEMBL | ZDB-GENE-990415-150 | mhc2a | NP\_001004521 | NM\_001004521 | ZDB-GENE-980526-201 | mhc2daa |
| Q2PR90 | TrEMBL | ZDB-GENE-070806-42 | or93a3 | XP\_017208427 | XM\_017352938 | ZDB-GENE-070806-4 | or113-4 |
| Q2PRH6 | TrEMBL | ZDB-GENE-010309-2 | or70a12 | XP\_001923166 | XM\_001923131 | ZDB-GENE-070806-6 | or70a1 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050522-93 | hist1h2a2 | NP\_001026996 | NM\_001031826 | ZDB-GENE-050731-6 | hist1h2a4 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050522-93 | hist1h2a2 | XP\_009301165 | XM\_009302890 | ZDB-GENE-160113-107 | si:dkey-23a13.6 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050522-93 | hist1h2a2 | XP\_009301166 | XM\_009302891 | ZDB-GENE-160113-98 | hist1h2a7 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050522-93 | hist1h2a2 | XP\_009301180 | XM\_009302905 | ZDB-GENE-160113-99 | si:ch73-368j24.14 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050522-93 | hist1h2a2 | XP\_017209638 | XM\_017354149 | ZDB-GENE-131127-88 | hist1h2a9 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050731-6 | hist1h2a4 | NP\_001019567 | NM\_001024396 | ZDB-GENE-050522-93 | hist1h2a2 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050731-6 | hist1h2a4 | XP\_005168861 | XM\_005168804 | ZDB-GENE-050522-93 | hist1h2a2 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050731-6 | hist1h2a4 | XP\_009301165 | XM\_009302890 | ZDB-GENE-160113-107 | si:dkey-23a13.6 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050731-6 | hist1h2a4 | XP\_009301166 | XM\_009302891 | ZDB-GENE-160113-98 | hist1h2a7 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050731-6 | hist1h2a4 | XP\_009301180 | XM\_009302905 | ZDB-GENE-160113-99 | si:ch73-368j24.14 |
| Q4FZZ7 | TrEMBL | ZDB-GENE-050731-6 | hist1h2a4 | XP\_017209638 | XM\_017354149 | ZDB-GENE-131127-88 | hist1h2a9 |
| Q4KMD2 | TrEMBL | ZDB-GENE-050706-140 | exosc3 | XP\_009296660 | XM\_009298385 | ZDB-GENE-030131-3295 | wu:fc33b09 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | NP\_001019567 | NM\_001024396 | ZDB-GENE-050522-93 | hist1h2a2 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_001338956 | XM\_001338920 | ZDB-GENE-131127-50 | si:dkey-261m9.7 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_001338956 | XM\_001338920 | ZDB-GENE-131127-50 | si:dkey-261m9.7 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_005163251 | XM\_005163194 | ZDB-GENE-121214-148 | si:ch211-113a14.7 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_005163251 | XM\_005163194 | ZDB-GENE-121214-148 | si:ch211-113a14.7 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_009296458 | XM\_009298183 | ZDB-GENE-131121-80 | si:dkey-108k21.15 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_009296465 | XM\_009298190 | ZDB-GENE-131127-83 | si:dkey-261m9.18 |
| Q561S9 | TrEMBL | ZDB-GENE-070822-28 | hist1h2a3 | XP\_009296465 | XM\_009298190 | ZDB-GENE-131127-83 | si:dkey-261m9.18 |
| Q568F0 | TrEMBL | ZDB-GENE-050706-188 | eef1a1l2 | NP\_001017795 | NM\_001017795 | ZDB-GENE-050417-327 | eef1a1b |
| Q568F0 | TrEMBL | ZDB-GENE-050706-188 | SO:0001217 | NP\_001017795 | NM\_001017795 | ZDB-GENE-050417-327 | eef1a1b |
| Q5BL99 | TrEMBL | ZDB-GENE-030828-11 | dynll2a | NP\_001025171 | NM\_001030000 | ZDB-GENE-050320-132 | dynll2b |
| Q5BL99 | TrEMBL | ZDB-GENE-050320-132 | dynll2b | NP\_956393 | NM\_200099 | ZDB-GENE-030828-11 | dynll2a |
| Q5RG00 | TrEMBL | ZDB-GENE-141215-12 | si:ch73-42k18.1 | XP\_003198008 | XM\_003197960 | ZDB-GENE-041014-76 | si:ch211-182e10.4 |
| Q5WQV6 | TrEMBL | ZDB-GENE-030219-125 | efemp2b | XP\_021324469 | XM\_021468794 | ZDB-GENE-040630-2 | fibpb |
| Q5WQV6 | TrEMBL | ZDB-GENE-030219-125 | efemp2b | XP\_021324470 | XM\_021468795 | ZDB-GENE-040630-2 | fibpb |
| Q5WQV6 | TrEMBL | ZDB-GENE-030219-125 | efemp2b | XP\_021324471 | XM\_021468796 | ZDB-GENE-040630-2 | fibpb |
| Q6DC41 | TrEMBL | ZDB-GENE-040801-35 | zgc:100864 | NP\_001003528 | NM\_001003528 | ZDB-GENE-041014-242 | gfus.2 |
| Q6DRK9 | TrEMBL | ZDB-GENE-040426-2676 | kyat1 | XP\_005171861 | XM\_005171804 | ZDB-GENE-040426-1333 | surf6 |
| Q6DRM3 | TrEMBL | ZDB-GENE-070410-99 | rfc1 | NP\_001003861 | NM\_001003861 | ZDB-GENE-030131-8646 | rpl9 |
| Q6DRM3 | TrEMBL | ZDB-GENE-070410-99 | rfc1 | XP\_005170963 | XM\_005170906 | ZDB-GENE-030131-8646 | rpl9 |
| Q6IQR3 | TrEMBL | ZDB-GENE-040520-4 | actc1a | XP\_002665778 | XM\_002665732 | ZDB-GENE-030131-172 | actc1c |
| Q6NZZ2 | TrEMBL | ZDB-GENE-010724-9 | arpc4 | NP\_001003762 | NM\_001003762 | ZDB-GENE-040808-18 | arpc4l |
| Q6NZZ2 | TrEMBL | ZDB-GENE-040808-18 | arpc4l | NP\_991100 | NM\_205537 | ZDB-GENE-010724-9 | arpc4 |
| Q6P2U5 | TrEMBL | ZDB-GENE-030131-8007 | arf6a | NP\_001154847 | NM\_001161375 | ZDB-GENE-030131-9056 | arf6b |
| Q6P2U5 | TrEMBL | ZDB-GENE-030131-9056 | arf6b | NP\_956287 | NM\_199993 | ZDB-GENE-030131-8007 | arf6a |
| Q7SXA1 | TrEMBL | ZDB-GENE-041010-133 | atp6v0e1 | NP\_998278 | NM\_213113 | ZDB-GENE-040426-2117 | rpl26 |
| Q7ZT21 | TrEMBL | ZDB-GENE-980526-80 | hbae1.1 | XP\_001333555 | XM\_001333519 | ZDB-GENE-061207-39 | hbae1.3 |
| Q801W9 | TrEMBL | ZDB-GENE-001106-8 | nitr3a | NP\_001007212 | NM\_001007211 | ZDB-GENE-001106-16 | nitr3b |
| Q8WM94 | TrEMBL | ZDB-GENE-030616-318 | si:busm1-194e12.11 | NP\_001007206 | NM\_001007205 | ZDB-GENE-030616-505 | mhc2dga |
| Q9DDE0 | TrEMBL | ZDB-GENE-020122-2 | meis2a | XP\_009292852 | XM\_009294577 | ZDB-GENE-000210-23 | meis2b |
| Q9DDE0 | TrEMBL | ZDB-GENE-020122-2 | meis2a | XP\_009292852 | XM\_009294577 | ZDB-GENE-000210-23 | meis2b |
| R4GDN4 | TrEMBL | ZDB-GENE-121214-244 | si:ch73-105m5.1 | XP\_002667479 | XM\_002667433 | ZDB-GENE-101021-2 | med9 |
| R4GE15 | TrEMBL | ZDB-GENE-121214-203 | si:ch211-113a14.22 | XP\_017209765 | XM\_017354276 | ZDB-GENE-121214-146 | si:ch211-113a14.29 |
| R4GE15 | TrEMBL | ZDB-GENE-121214-203 | si:ch211-113a14.22 | XP\_691943 | XM\_686851 | ZDB-GENE-160113-110 | si:dkey-23a13.8 |
| R4GEN1 | TrEMBL | ZDB-GENE-121214-92 | si:dkey-31m14.7 | XP\_699946 | XM\_694854 | ZDB-GENE-100330-2 | mark4b |
| R4GEQ7 | TrEMBL | ZDB-GENE-121214-112 | mfap4.9 | XP\_003197840 | XM\_003197792 | ZDB-GENE-121214-168 | mfap4.8 |
| R4GEZ5 | TrEMBL | ZDB-GENE-120703-46 | si:zfos-364h11.2 | XP\_002665163 | XM\_002665117 | ZDB-GENE-121214-26 | si:zfos-364h11.1 |
| U3JA39 | TrEMBL | ZDB-GENE-061103-70 | fbxo11a | XP\_021328223 | XM\_021472548 | ZDB-GENE-110810-2 | fbxo11b |
| U3JA39 | TrEMBL | ZDB-GENE-061103-70 | fbxo11a | XP\_696134 | XM\_691042 | ZDB-GENE-110810-2 | fbxo11b |
| X1WB90 | TrEMBL | ZDB-GENE-040801-253 | dhrs13b.1 | XP\_001341662 | XM\_001341626 | ZDB-GENE-131121-649 | si:ch211-108p6.4 |
| X1WCJ9 | TrEMBL | ZDB-GENE-060414-4 | taar20o | XP\_001920823 | XM\_001920788 | ZDB-GENE-060414-17 | taar20u |
| X1WD42 | TrEMBL | ZDB-GENE-131122-80 | ftr90 | XP\_017206748 | XM\_017351259 | ZDB-GENE-131125-2 | ftr88 |
| X1WEP8 | TrEMBL | ZDB-GENE-050222-2 | kcnj11l | XP\_001336789 | XM\_001336753 | ZDB-GENE-131127-232 | si:ch211-286c4.6 |
| X1WEP8 | TrEMBL | ZDB-GENE-050222-2 | kcnj11l | XP\_009296057 | XM\_009297782 | ZDB-GENE-131127-232 | si:ch211-286c4.6 |
| X1WEV4 | TrEMBL | ZDB-GENE-070705-267 | flrt2 | XP\_005158737 | XM\_005158680 | ZDB-GENE-131120-42 | si:dkey-87k14.1 |
| X1WEV4 | TrEMBL | ZDB-GENE-070705-267 | flrt2 | XP\_021328761 | XM\_021473086 | ZDB-GENE-131120-42 | si:dkey-87k14.1 |
| X1WEZ5 | TrEMBL | ZDB-GENE-131120-28 | si:dkey-37f18.2 | XP\_003201514 | XM\_003201466 | ZDB-GENE-100212-1 | otud7a |
| X1WG38 | TrEMBL | ZDB-GENE-030131-3481 | si:ch73-95l15.5 | NP\_956460 | NM\_200166 | ZDB-GENE-040426-762 | ppt2a.2 |
| X1WHH5 | TrEMBL | ZDB-GENE-040912-55 | ms4a17a.5 | XP\_001339255 | XM\_001339219 | ZDB-GENE-131126-22 | si:ch73-56d11.3 |
| X1WHH8 | TrEMBL | ZDB-GENE-030131-657 | cyth1a | XP\_017210324 | XM\_017354835 | ZDB-GENE-130530-529 | cyth3b |

## Appendix 2.[¶](#Appendix-2.)

Same as above, but instead of cross-checking on the refseq, just check on the uniprot accession.

In [43]:

%%sql

drop table if exists appendix2;

create table appendix2 as

SELECT

k.\*,

d.dblink\_linked\_recid,

m.mrkr\_abbrev

FROM

to\_keep\_wide k

JOIN db\_link d ON d.dblink\_acc\_num = k.accession

JOIN marker m ON d.dblink\_linked\_recid = m.mrkr\_zdb\_id

WHERE zfin\_id not in (select dblink\_linked\_recid from db\_link where dblink\_acc\_num = k.accession);

-- select count(\*) from appendix1;

select \* from appendix2;

\* sqlite:///zfin-db-slice.db

Done.

Done.

Done.

Out[43]:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **accession** | **entry\_type** | **zfin\_id** | **abbrev** | **refseq1** | **refseq2** | **dblink\_linked\_recid** | **mrkr\_abbrev** |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_002662464 | XM\_002662418 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_005168803 | XM\_005168746 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_017212106 | XM\_017356617 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_021332779 | XM\_021477104 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4IHM6 | TrEMBL | ZDB-GENE-120215-200 | si:dkey-44g17.6 | XP\_021332780 | XM\_021477105 | ZDB-GENE-091020-2 | gng12b |
| A0A0R4INQ1 | TrEMBL | ZDB-GENE-141215-36 | si:ch211-160o17.4 | XP\_009304049 | XM\_009305774 | ZDB-GENE-140324-1 | hipk1b |
| A0A0R4IVW1 | TrEMBL | ZDB-GENE-050208-210 | si:ch73-217b7.1 | XP\_021325713 | XM\_021470038 | ZDB-GENE-090313-100 | map7b |
| A0A0R4IZK4 | TrEMBL | ZDB-GENE-141216-18 | si:dkey-225f5.5 | XP\_001340413 | XM\_001340377 | ZDB-GENE-140820-4 | mhc1lca |
| A0A140LG23 | TrEMBL | ZDB-GENE-160113-53 | si:dkey-23a13.21 | XP\_009301168 | XM\_009302893 | ZDB-GENE-030722-8 | hist2h3ca1 |
| A0A2R8PYX0 | TrEMBL | ZDB-GENE-091204-463 | si:ch73-63e15.2 | XP\_009294614 | XM\_009296339 | ZDB-GENE-060130-85 | sbno2b |
| A0A2R8Q097 | TrEMBL | ZDB-GENE-030131-7310 | eef1b2 | NP\_001313402 | NM\_001326473 | ZDB-GENE-101101-1 | zdbf2 |
| A0A2R8QAK7 | TrEMBL | ZDB-GENE-030131-6520 | mrpl39 | NP\_001314937 | NM\_001328008 | ZDB-GENE-200617-1 | cenpq |
| A0A2R8QLP7 | TrEMBL | ZDB-GENE-050411-48 | im:7142702 | XP\_002667906 | XM\_002667860 | ZDB-GENE-110201-2 | bcl11bb |
| A0A2R8QMY3 | TrEMBL | ZDB-GENE-030616-83 | si:rp71-79p20.2 | XP\_002666496 | XM\_002666450 | ZDB-GENE-090313-332 | slc2a4rg |
| A3KP86 | TrEMBL | ZDB-GENE-070410-131 | zgc:163057 | NP\_001076303 | NM\_001082834 | ZDB-GENE-980526-286 | hbae4 |
| A5WV15 | TrEMBL | ZDB-GENE-030131-6493 | acsl4b | NP\_001002050 | NM\_001002050 | ZDB-GENE-040625-10 | cwc15 |
| B3DH93 | TrEMBL | ZDB-GENE-080722-16 | fthl29 | NP\_001124139 | NM\_001130667 | ZDB-GENE-050522-428 | fthl31 |
| B3DH93 | TrEMBL | ZDB-GENE-080722-16 | fthl29 | XP\_021324849 | XM\_021469174 | ZDB-GENE-050522-428 | fthl31 |
| B3DHP3 | TrEMBL | ZDB-GENE-080723-51 | zgc:194186 | NP\_001124102 | NM\_001130630 | ZDB-GENE-080723-25 | zgc:194215 |
| B3DHP3 | TrEMBL | ZDB-GENE-080723-51 | zgc:194186 | NP\_001124121 | NM\_001130649 | ZDB-GENE-080723-25 | zgc:194215 |
| E7F016 | TrEMBL | ZDB-GENE-130530-933 | zmp:0000000930 | XP\_021322425 | XM\_021466750 | ZDB-GENE-060825-61 | ccdc127a |
| E7F1C0 | TrEMBL | ZDB-GENE-131127-163 | si:dkey-178o16.4 | XP\_003198871 | XM\_003198823 | ZDB-GENE-121023-1 | avpr2b.1 |
| E7F1C0 | TrEMBL | ZDB-GENE-131127-163 | si:dkey-178o16.4 | XP\_021327514 | XM\_021471839 | ZDB-GENE-121023-1 | avpr2b.1 |
| E7F4B5 | TrEMBL | ZDB-GENE-131121-60 | si:dkey-27o4.1 | XP\_689861 | XM\_684769 | ZDB-GENE-110922-6 | fam83ga |
| E7F551 | TrEMBL | ZDB-GENE-141215-36 | si:ch211-160o17.4 | XP\_701123 | XM\_696031 | ZDB-GENE-140324-1 | hipk1b |
| E7F704 | TrEMBL | ZDB-GENE-121214-325 | si:ch1073-272o11.3 | XP\_005169488 | XM\_005169431 | ZDB-GENE-101203-4 | cdr2b |
| E7F9B4 | TrEMBL | ZDB-GENE-141216-123 | si:dkey-264d12.4 | XP\_009304549 | XM\_009306274 | ZDB-GENE-150211-1 | emp3a |
| E7F9B4 | TrEMBL | ZDB-GENE-141216-123 | si:dkey-264d12.4 | XP\_021335574 | XM\_021479899 | ZDB-GENE-150211-1 | emp3a |
| E7F9Y7 | TrEMBL | ZDB-GENE-030131-4199 | tagln3b | XP\_002666688 | XM\_002666642 | ZDB-GENE-041111-196 | abhd10b |
| E7FD02 | TrEMBL | ZDB-GENE-131125-81 | si:dkey-83h2.5 | XP\_009289708 | XM\_009291433 | ZDB-GENE-130530-576 | ano7 |
| E7FG24 | TrEMBL | ZDB-GENE-121214-254 | si:dkeyp-9d4.4 | XP\_009304437 | XM\_009306162 | ZDB-GENE-041008-52 | fam83gb |
| E7FGT6 | TrEMBL | ZDB-GENE-070424-101 | zgc:163022 | XP\_017209505 | XM\_017354016 | ZDB-GENE-030131-7307 | wu:fj05g07 |
| E9QG09 | TrEMBL | ZDB-GENE-040801-94 | pibf1 | XP\_001344916 | XM\_001344880 | ZDB-GENE-090312-167 | klf5a |
| F1Q7N5 | TrEMBL | ZDB-GENE-121214-88 | si:dkeyp-9d4.5 | XP\_017211446 | XM\_017355957 | ZDB-GENE-110922-5 | slc5a10 |
| F1QAF6 | TrEMBL | ZDB-GENE-080227-6 | SO:0001217 | NP\_001032505 | NM\_001037428 | ZDB-GENE-071004-4 | ugt1a1 |
| F1QAF6 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001032505 | NM\_001037428 | ZDB-GENE-071004-4 | ugt1a1 |
| F1QIH9 | TrEMBL | ZDB-GENE-071004-54 | zgc:173556 | XP\_005169843 | XM\_005169786 | ZDB-GENE-050626-121 | zp3.2 |
| F1QM88 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001166241 | NM\_001172770 | ZDB-GENE-071004-5 | ugt1a2 |
| F1QNF8 | TrEMBL | ZDB-GENE-100406-4 | ugt5b2 | XP\_005170903 | XM\_005170846 | ZDB-GENE-100406-3 | ugt5b1 |
| F1QRI1 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_001170810 | NM\_001177339 | ZDB-GENE-080227-5 | ugt1a6 |
| F1QRX6 | TrEMBL | ZDB-GENE-041014-86 | si:dkey-88n24.9 | NP\_001314984 | NM\_001328055 | ZDB-GENE-070424-265 | si:dkey-88n24.10 |
| F1QRX6 | TrEMBL | ZDB-GENE-041014-86 | si:dkey-88n24.9 | XP\_017212352 | XM\_017356863 | ZDB-GENE-070424-265 | si:dkey-88n24.10 |
| F1QVS3 | TrEMBL | ZDB-GENE-041001-97 | si:ch211-285c6.4 | XP\_005166693 | XM\_005166636 | ZDB-GENE-080215-3 | zgc:172075 |
| F1QZA2 | TrEMBL | ZDB-GENE-030616-83 | si:rp71-79p20.2 | XP\_005162071 | XM\_005162014 | ZDB-GENE-090313-332 | slc2a4rg |
| F1R0T3 | TrEMBL | ZDB-GENE-100922-162 | si:dkey-40m6.8 | XP\_005158431 | XM\_005158374 | ZDB-GENE-100812-13 | lmtk3 |
| F1R1P2 | TrEMBL | ZDB-GENE-080227-6 | ugt1a7 | NP\_998587 | NM\_213422 | ZDB-GENE-040426-2762 | ugt1ab |
| F8W539 | TrEMBL | ZDB-GENE-030131-7370 | wu:fj11g02 | XP\_002660664 | XM\_002660618 | ZDB-GENE-110408-28 | mfap4.13 |
| I3IT29 | TrEMBL | ZDB-GENE-120215-63 | si:dkey-151m15.5 | XP\_003198160 | XM\_003198112 | ZDB-GENE-100311-1 | rundc1 |
| Q1LW00 | TrEMBL | ZDB-GENE-030131-2100 | he1.2 | NP\_001038639 | NM\_001045174 | ZDB-GENE-021211-3 | he1.1 |
| Q1MTC7 | TrEMBL | ZDB-GENE-990415-150 | mhc2a | NP\_001004521 | NM\_001004521 | ZDB-GENE-980526-201 | mhc2daa |
| Q1MTC7 | TrEMBL | ZDB-GENE-990415-150 | mhc2a | NP\_001092896 | NM\_001099426 | ZDB-GENE-980526-201 | mhc2daa |
| Q6DC41 | TrEMBL | ZDB-GENE-040801-35 | zgc:100864 | NP\_001003528 | NM\_001003528 | ZDB-GENE-041014-242 | gfus.2 |
| Q8WM94 | TrEMBL | ZDB-GENE-030616-318 | si:busm1-194e12.11 | NP\_001007206 | NM\_001007205 | ZDB-GENE-030616-505 | mhc2dga |
| Q9DDE0 | TrEMBL | ZDB-GENE-020122-2 | meis2a | XP\_009292852 | XM\_009294577 | ZDB-GENE-000210-23 | meis2b |
| Q9DDE0 | TrEMBL | ZDB-GENE-020122-2 | meis2a | XP\_009292852 | XM\_009294577 | ZDB-GENE-000210-23 | meis2b |
| R4GDN4 | TrEMBL | ZDB-GENE-121214-244 | si:ch73-105m5.1 | XP\_002667479 | XM\_002667433 | ZDB-GENE-101021-2 | med9 |
| R4GEN1 | TrEMBL | ZDB-GENE-121214-92 | si:dkey-31m14.7 | XP\_699946 | XM\_694854 | ZDB-GENE-100330-2 | mark4b |
| X1WEZ5 | TrEMBL | ZDB-GENE-131120-28 | si:dkey-37f18.2 | XP\_003201514 | XM\_003201466 | ZDB-GENE-100212-1 | otud7a |

## Playground[¶](#Playground)

Here is where to put some temporary sql for playing around with the data

In [44]:

%%sql

-- select accession, primary\_id from to\_keep\_after\_3a limit 5;

select refseq from refseq2ncbi limit 5;

\* sqlite:///zfin-db-slice.db

Done.

Out[44]:

|  |
| --- |
| **refseq** |
| NP\_001001398 |
| NP\_001001399 |
| NP\_001001400 |
| NP\_001001401 |
| NP\_001001402 |

In [45]:

%%sql

-- select count(\*) from to\_keep\_after\_3a k join refseq2ncbi r on k.primary\_id = r.refseq join db\_link on ncbi = dblink\_acc\_num

select accession, primary\_id, ncbi, db\_link.\* from to\_keep\_after\_3a k join refseq2ncbi r on k.primary\_id = r.refseq join db\_link on ncbi = dblink\_acc\_num

\* sqlite:///zfin-db-slice.db

Done.

Out[45]:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **accession** | **primary\_id** | **ncbi** | **dblink\_linked\_recid** | **dblink\_acc\_num** | **dblink\_info** | **dblink\_zdb\_id** | **dblink\_acc\_num\_display** | **dblink\_length** | **dblink\_fdbcont\_zdb\_id** |

# Export Excel Spreadsheet[¶](#Export-Excel-Spreadsheet)

In [46]:

import sqlite3

import pandas as pd

def main():

tables = ['1a', '1b', '2a', '2b', '3a', '3b', '4a', '4b', '5b', 'extnotes', 'appendix1', 'appendix2', 'to\_keep\_after\_4a']

# Create a Pandas Excel writer using the openpyxl engine

writer = pd.ExcelWriter('8395 - UniProt Data.xlsx', engine='openpyxl')

# Loop over the CSV files

for i, table in enumerate(tables):

# Read the CSV file

df = get\_table\_rows\_as\_data\_frame(table)

# Write the dataframe to a sheet in the Excel file

df.to\_excel(writer, table, index=False)

writer.close()

def get\_table\_rows\_as\_data\_frame(tablename):

# Connect to the database

conn = sqlite3.connect('zfin-db-slice.db')

# Create a cursor

cursor = conn.cursor()

cursor.execute('SELECT \* FROM "' + tablename + '"')

results = cursor.fetchall()

column\_names = [description[0] for description in cursor.description]

# Convert the results to a Pandas DataFrame

df = pd.DataFrame(results, columns=column\_names)

cursor.close()

conn.close()

return df

main()

In [ ]: