An Introduction to BigQuery

(in less than 10 minutes)

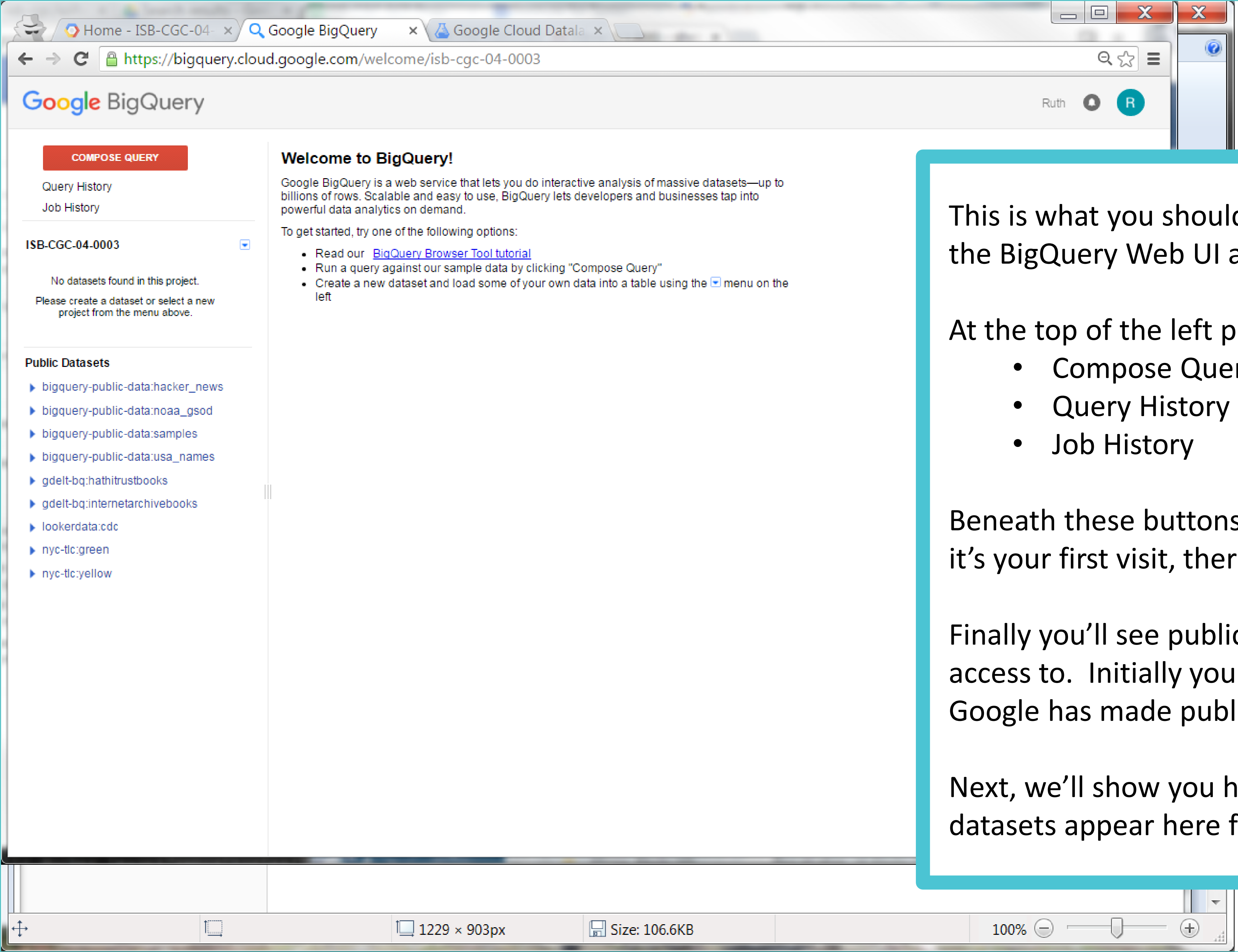
brought to you by

The ISB Cancer Genomics Cloud

Institute for Systems Biology
Revolutionizing Science. Enhancing Life.

Google Cloud Platform

CSRA



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

- Read our BigQuery Browser Tool tutorial
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the ▾ menu on the left

This is what you should see the first time you go to the BigQuery Web UI at bigquery.cloud.google.com

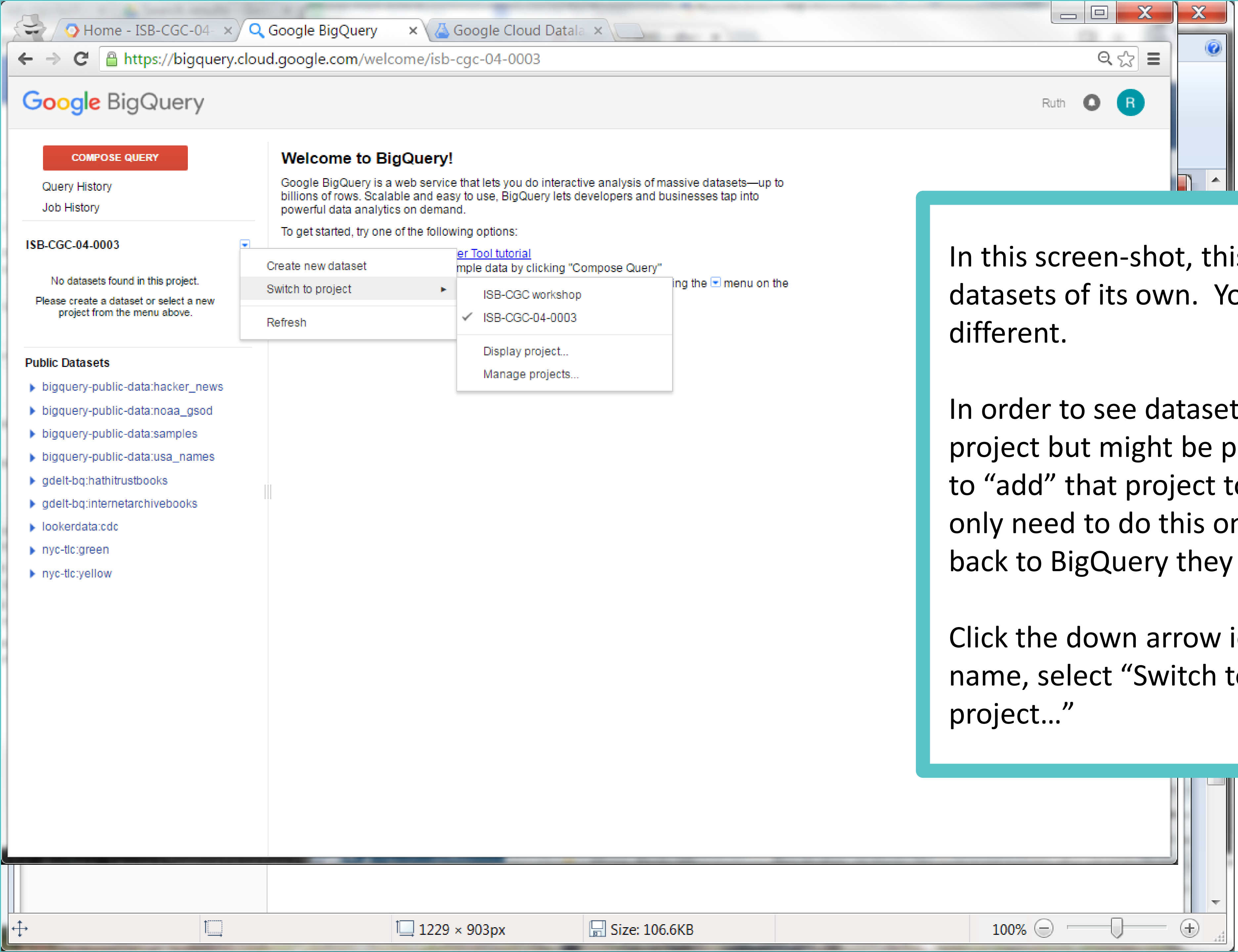
At the top of the left panel are three buttons:

- Compose Query
- Query History
- Job History

Beneath these buttons is your project space. Since it's your first visit, there are no datasets.

Finally you'll see public datasets that you may have access to. Initially you will see a few datasets that Google has made public.

Next, we'll show you how to make the ISB-CGC datasets appear here for easy access.



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

er Tool tutorial

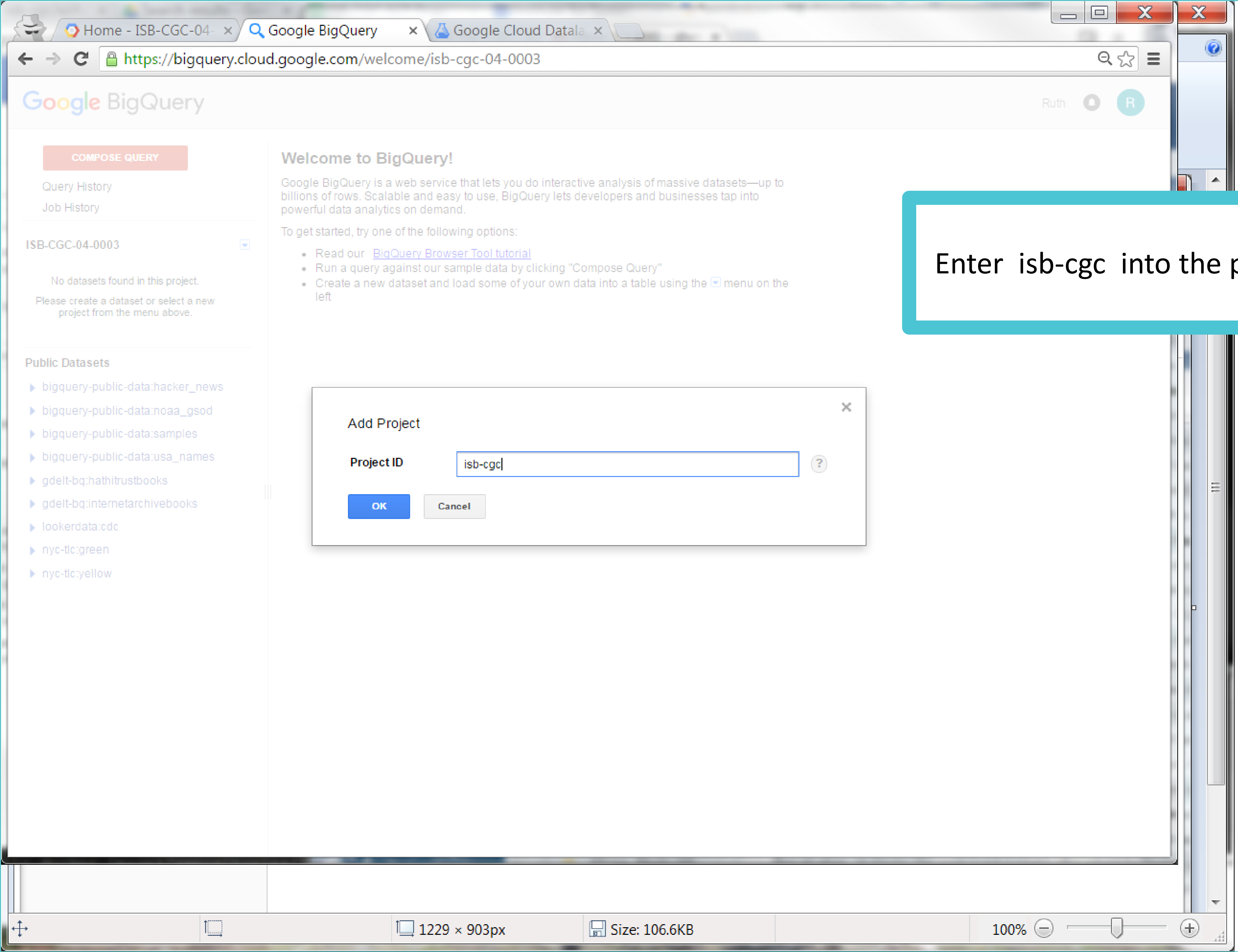
mple data by clicking "Compose Query"

ing the menu on the

In this screen-shot, this particular project has no datasets of its own. Your project might look different.

In order to see datasets that are owned by another project but might be publicly-accessible, you need to “add” that project to your BigQuery view. (You’ll only need to do this once – next time you come back to BigQuery they will already be there.)

Click the down arrow icon next to your project name, select “Switch to project”, and then “Display project…”



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

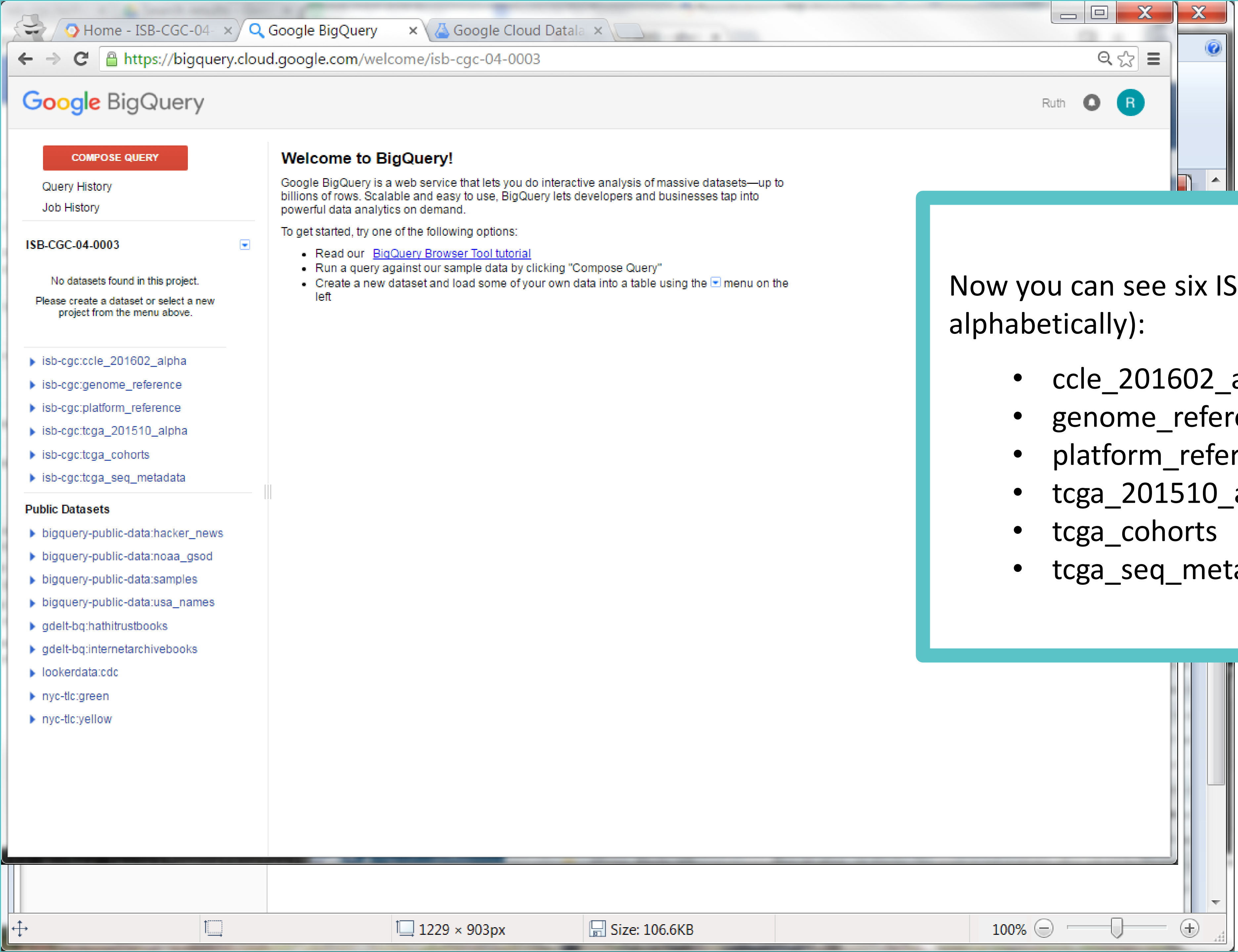
To get started, try one of the following options:

- Read our BigQuery Browser Tool tutorial
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the menu on the left

Add Project

Project ID: isb-cgc

Enter isb-cgc into the pop-up window, and click **OK**



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

- isb-cgc:ccle_201602_alpha
- isb-cgc:genome_reference
- isb-cgc:platform_reference
- isb-cgc:tcga_201510_alpha
- isb-cgc:tcga_cohorts
- isb-cgc:tcga_seq_metadata

Public Datasets

- bigquery-public-data:hacker_news
- bigquery-public-data:noaa_gsod
- bigquery-public-data:samples
- bigquery-public-data:usa_names
- gdelt-bq:hathitrustbooks
- gdelt-bq:internetarchivebooks
- lookerdata:cdc
- nyc-tlc:green
- nyc-tlc:yellow

Welcome to BigQuery!

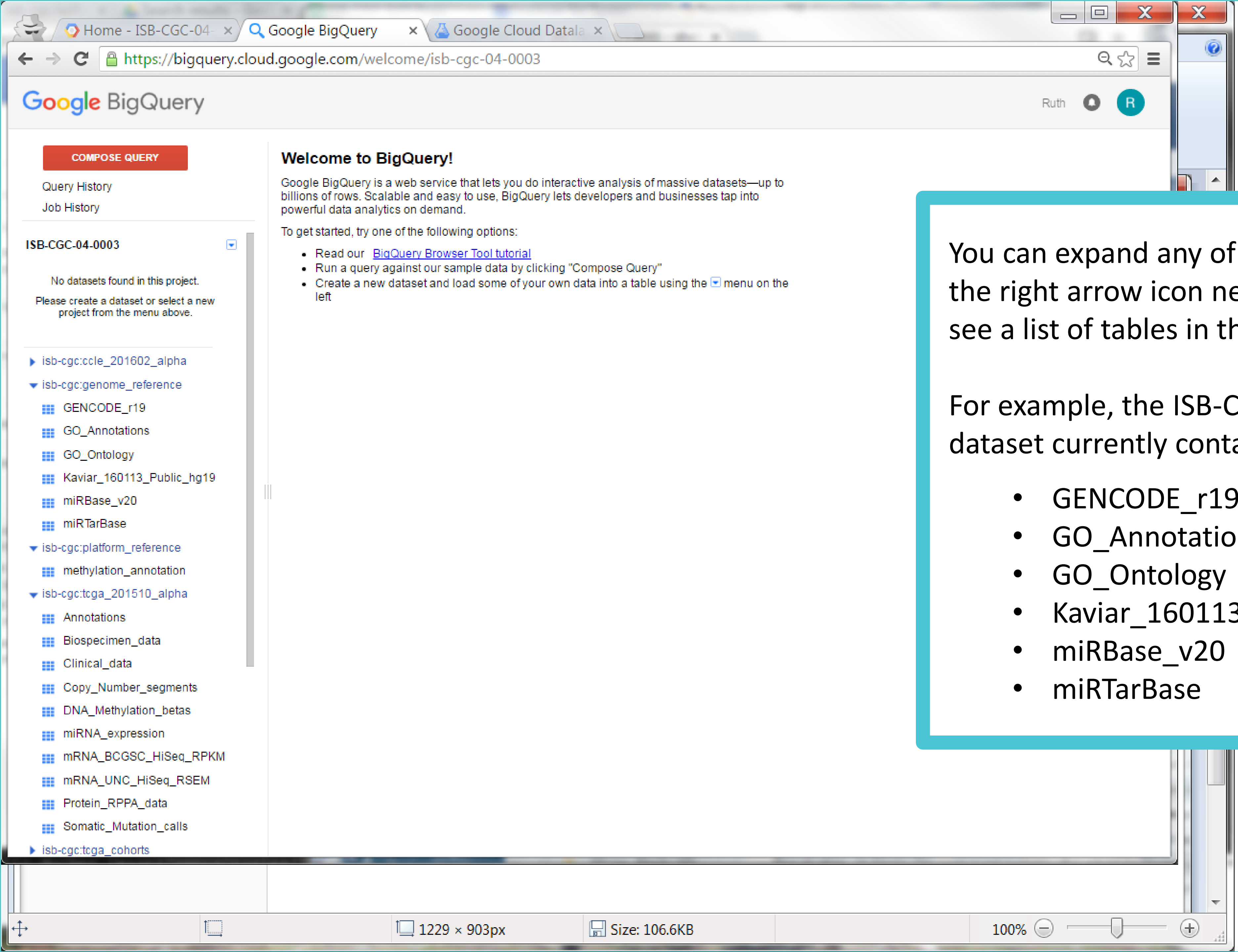
Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

- Read our BigQuery Browser Tool tutorial
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the ▾ menu on the left

Now you can see six ISB-CGC datasets (arranged alphabetically):

- ccle_201602_alpha
- genome_reference
- platform_reference
- tcga_201510_alpha
- tcga_cohorts
- tcga_seq_metadata



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

- isb-cgc:ccle_201602_alpha
- isb-cgc:genome_reference
  - GENCODE_r19
  - GO_Annotations
  - GO_Ontology
  - Kaviar_160113_Public_hg19
  - miRBase_v20
  - miRTarBase
- isb-cgc:platform_reference
  - methylation_annotation
- isb-cgc:tcga_201510_alpha
  - Annotations
  - Biospecimen_data
  - Clinical_data
  - Copy_Number_segments
  - DNA_Methylation_betas
  - miRNA_expression
  - mRNA_BCGSC_HiSeq_RPKM
  - mRNA_UNC_HiSeq_RSEM
  - Protein_RPPA_data
  - Somatic_Mutation_calls
- isb-cgc:tcga_cohorts

Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

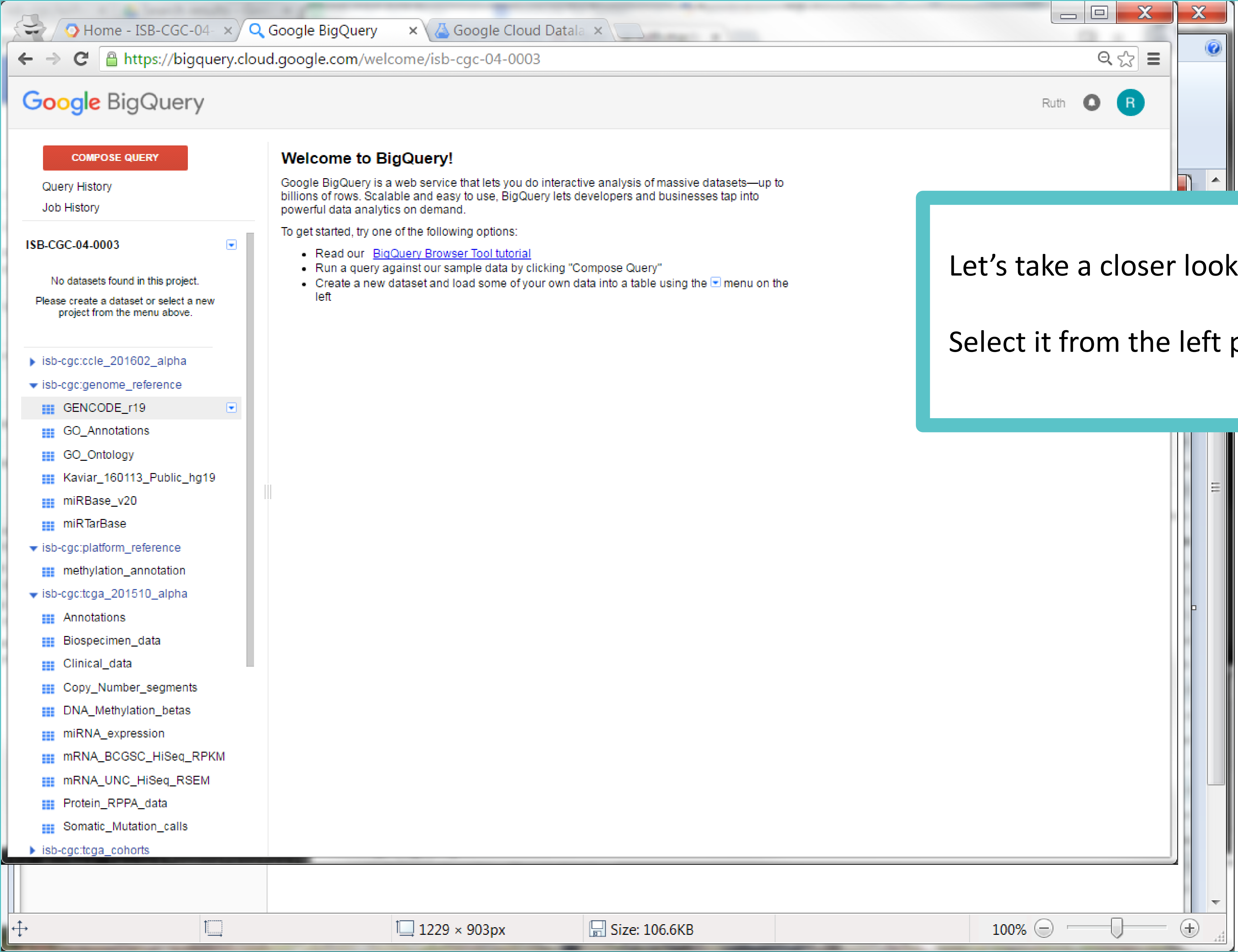
To get started, try one of the following options:

- Read our BigQuery Browser Tool tutorial
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the ▾ menu on the left

You can expand any of the datasets by clicking on the right arrow icon next to the dataset name, to see a list of tables in that dataset.

For example, the ISB-CGC “genome_reference” dataset currently contains the following tables:

- GENCODE_r19
- GO_Annotations
- GO_Ontology
- Kaviar_160113_Public_hg19
- miRBase_v20
- miRTarBase



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

Welcome to BigQuery!

Google BigQuery is a web service that lets you do interactive analysis of massive datasets—up to billions of rows. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

To get started, try one of the following options:

- Read our BigQuery Browser Tool tutorial
- Run a query against our sample data by clicking "Compose Query"
- Create a new dataset and load some of your own data into a table using the menu on the left

Let’s take a closer look at the GENCODE_r19 table.

Select it from the left panel and click.

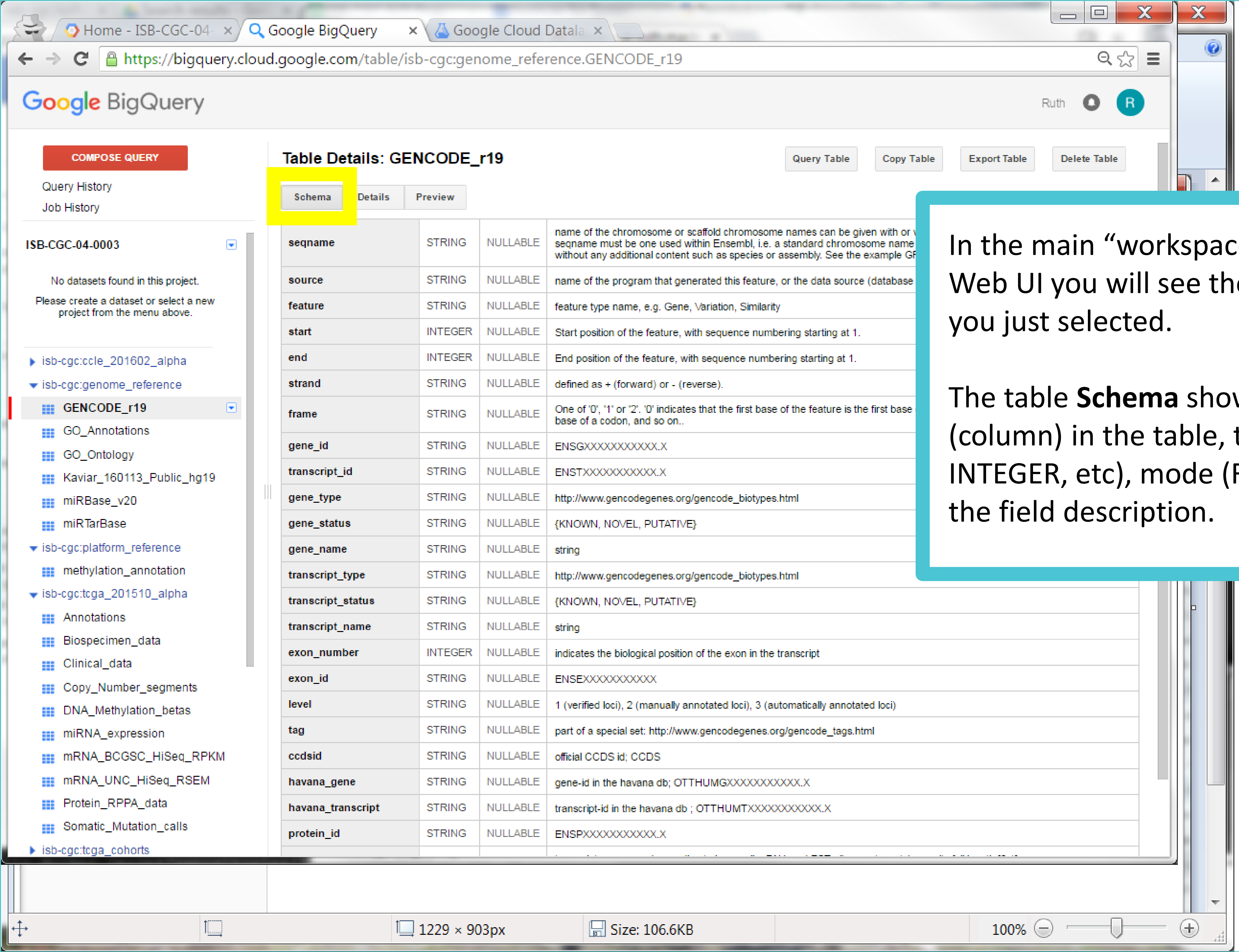


Table Details: GENCODE_r19

seqname	STRING	NULLABLE	name of the chromosome or scaffold chromosome names can be given with or seqname must be one used within Ensembl, i.e. a standard chromosome name without any additional content such as species or assembly. See the example GF
source	STRING	NULLABLE	name of the program that generated this feature, or the data source (database
feature	STRING	NULLABLE	feature type name, e.g. Gene, Variation, Similarity
start	INTEGER	NULLABLE	Start position of the feature, with sequence numbering starting at 1.
end	INTEGER	NULLABLE	End position of the feature, with sequence numbering starting at 1.
strand	STRING	NULLABLE	defined as + (forward) or - (reverse).
frame	STRING	NULLABLE	One of '0', '1' or '2'. '0' indicates that the first base of the feature is the first base base of a codon, and so on..
gene_id	STRING	NULLABLE	ENSGXXXXXXXXXXX.X
transcript_id	STRING	NULLABLE	ENSTXXXXXXXXXXX.X
gene_type	STRING	NULLABLE	http://www.gencodegenes.org/gencode_biotypes.html
gene_status	STRING	NULLABLE	{KNOWN, NOVEL, PUTATIVE}
gene_name	STRING	NULLABLE	string
transcript_type	STRING	NULLABLE	http://www.gencodegenes.org/gencode_biotypes.html
transcript_status	STRING	NULLABLE	{KNOWN, NOVEL, PUTATIVE}
transcript_name	STRING	NULLABLE	string
exon_number	INTEGER	NULLABLE	indicates the biological position of the exon in the transcript
exon_id	STRING	NULLABLE	ENSEXXXXXXXXXXX
level	STRING	NULLABLE	1 (verified loci), 2 (manually annotated loci), 3 (automatically annotated loci)
tag	STRING	NULLABLE	part of a special set: http://www.gencodegenes.org/gencode_tags.html
ccdsid	STRING	NULLABLE	official CCDS id; CCDS
havana_gene	STRING	NULLABLE	gene-id in the havana db; OTTHUMGXXXXXXXXXXX.X
havana_transcript	STRING	NULLABLE	transcript-id in the havana db ; OTTHUMTXXXXXXXXXXX.X
protein_id	STRING	NULLABLE	ENSPXXXXXXXXXXX.X

In the main "workspace" portion of the BigQuery Web UI you will see the "Table Details" for the table you just selected.

The table **Schema** shows the name of each field (column) in the table, the data type (STRING, INTEGER, etc), mode (REQUIRED or NULLABLE), and the field description.

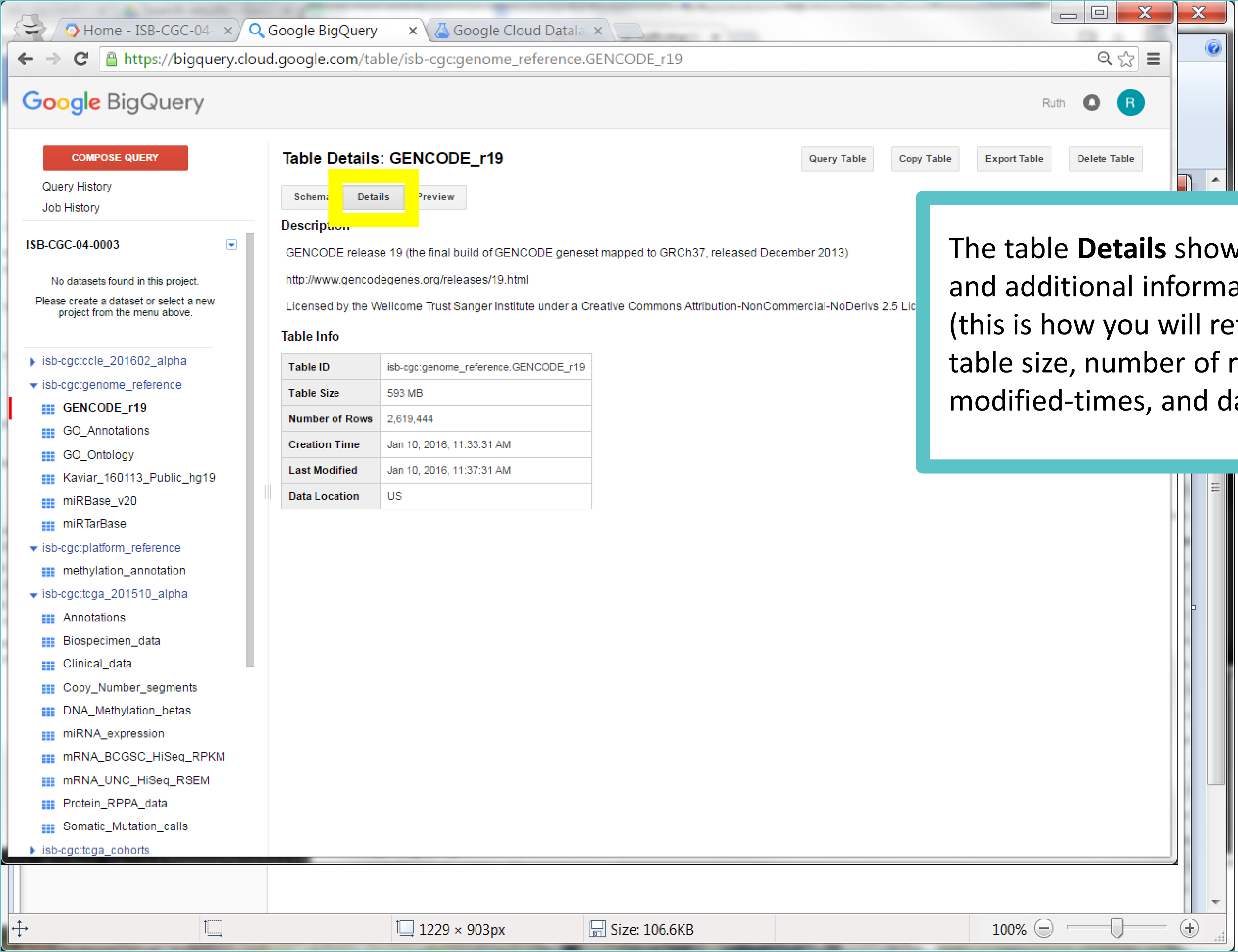


Table Details: GENCODE_r19

Description

GENCODE release 19 (the final build of GENCODE geneset mapped to GRCh37, released December 2013)

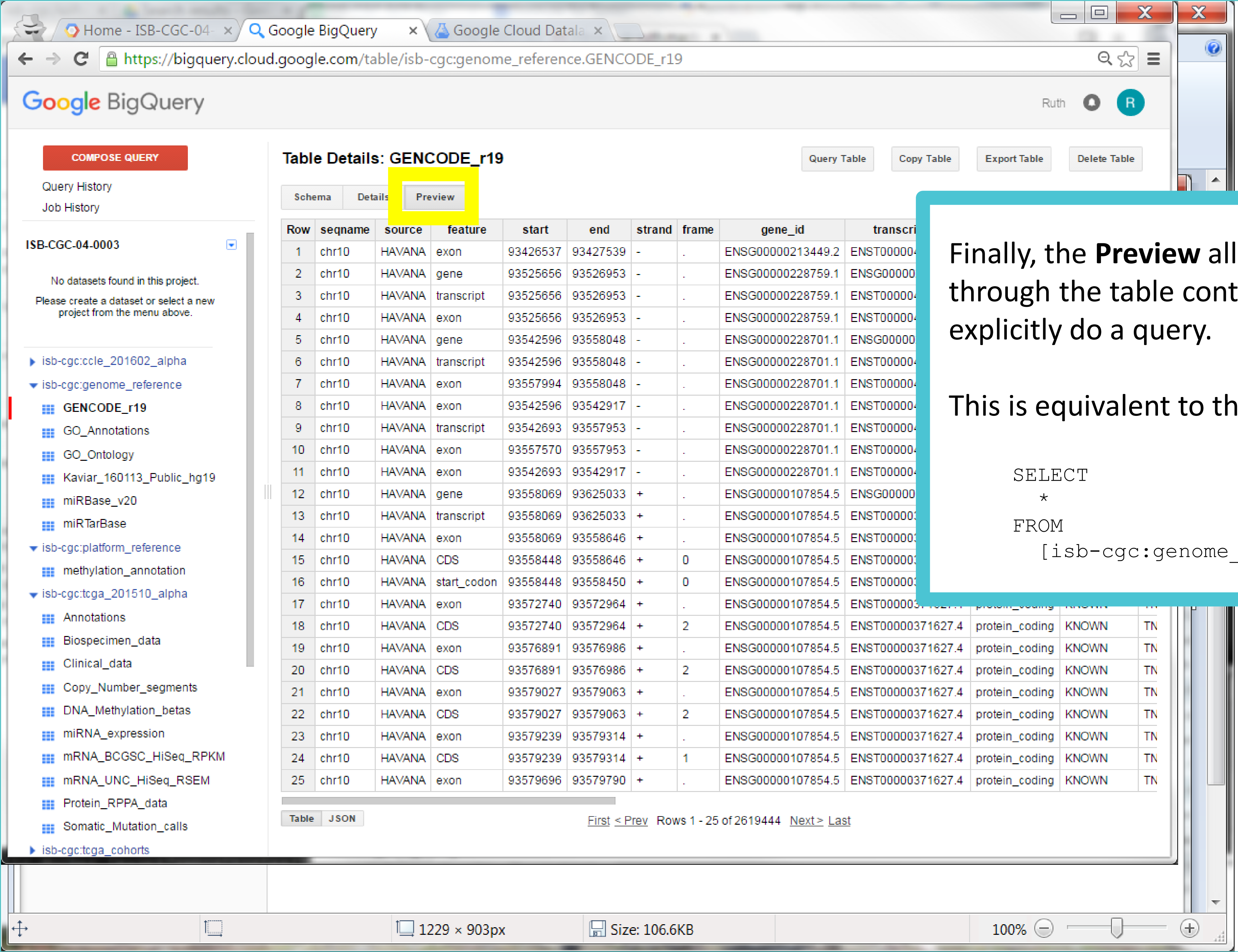
http://www.gencodegenes.org/releases/19.html

Licensed by the Wellcome Trust Sanger Institute under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 Lic

Table Info

Table ID	isb-cgc:genome_reference.GENCODE_r19
Table Size	593 MB
Number of Rows	2,619,444
Creation Time	Jan 10, 2016, 11:33:31 AM
Last Modified	Jan 10, 2016, 11:37:31 AM
Data Location	US

The table **Details** shows you the table Description and additional information including the table ID (this is how you will refer to it in a SQL query), the table size, number of rows, creation- and last-modified-times, and data location.



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

Table Details: GENCODE_r19

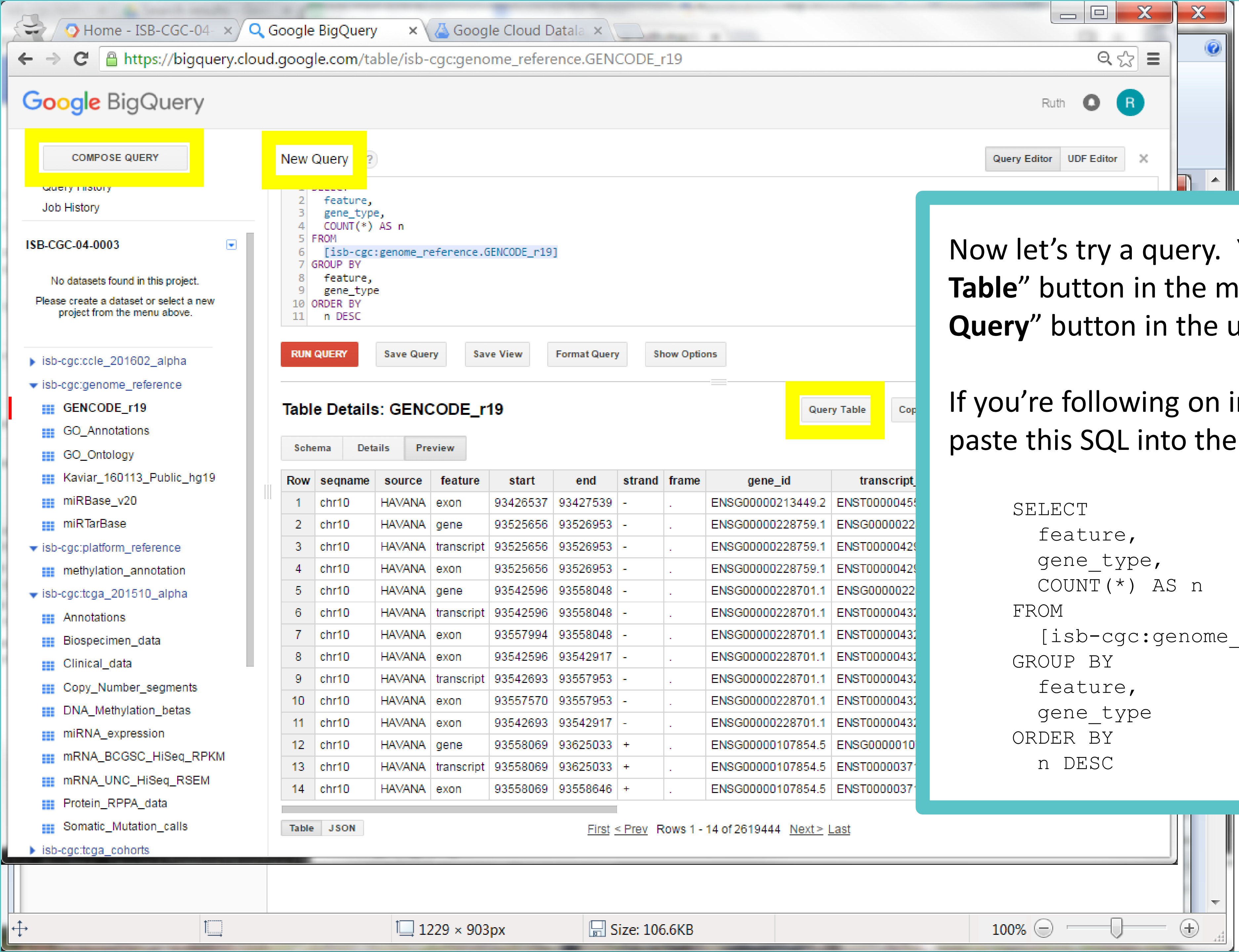
Row	seqname	source	feature	start	end	strand	frame	gene_id	transcri			
1	chr10	HAVANA	exon	93426537	93427539	-	.	ENSG00000213449.2	ENST00000			
2	chr10	HAVANA	gene	93525656	93526953	-	.	ENSG00000228759.1	ENSG00000			
3	chr10	HAVANA	transcript	93525656	93526953	-	.	ENSG00000228759.1	ENST00000			
4	chr10	HAVANA	exon	93525656	93526953	-	.	ENSG00000228759.1	ENST00000			
5	chr10	HAVANA	gene	93542596	93558048	-	.	ENSG00000228701.1	ENSG00000			
6	chr10	HAVANA	transcript	93542596	93558048	-	.	ENSG00000228701.1	ENST00000			
7	chr10	HAVANA	exon	93557994	93558048	-	.	ENSG00000228701.1	ENST00000			
8	chr10	HAVANA	exon	93542596	93542917	-	.	ENSG00000228701.1	ENST00000			
9	chr10	HAVANA	transcript	93542693	93557953	-	.	ENSG00000228701.1	ENST00000			
10	chr10	HAVANA	exon	93557570	93557953	-	.	ENSG00000228701.1	ENST00000			
11	chr10	HAVANA	exon	93542693	93542917	-	.	ENSG00000228701.1	ENST00000			
12	chr10	HAVANA	gene	93558069	93625033	+	.	ENSG00000107854.5	ENSG00000			
13	chr10	HAVANA	transcript	93558069	93625033	+	.	ENSG00000107854.5	ENST00000			
14	chr10	HAVANA	exon	93558069	93558646	+	.	ENSG00000107854.5	ENST00000			
15	chr10	HAVANA	CDS	93558448	93558646	+	0	ENSG00000107854.5	ENST00000			
16	chr10	HAVANA	start_codon	93558448	93558450	+	0	ENSG00000107854.5	ENST00000			
17	chr10	HAVANA	exon	93572740	93572964	+	.	ENSG00000107854.5	ENST00000			
18	chr10	HAVANA	CDS	93572740	93572964	+	2	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
19	chr10	HAVANA	exon	93576891	93576986	+	.	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
20	chr10	HAVANA	CDS	93576891	93576986	+	2	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
21	chr10	HAVANA	exon	93579027	93579063	+	.	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
22	chr10	HAVANA	CDS	93579027	93579063	+	2	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
23	chr10	HAVANA	exon	93579239	93579314	+	.	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
24	chr10	HAVANA	CDS	93579239	93579314	+	1	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN
25	chr10	HAVANA	exon	93579696	93579790	+	.	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TN

First < Prev Rows 1 - 25 of 2619444 Next > Last

Finally, the **Preview** allows you see to see and scroll through the table contents without having to explicitly do a query.

This is equivalent to the following SQL query:

```
SELECT
  *
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
```



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

- isb-cgc:ccle_201602_alpha
- isb-cgc:genome_reference
  - GENCODE_r19
  - GO_Annotations
  - GO_Ontology
  - Kaviar_160113_Public_hg19
  - miRBase_v20
  - miRTarBase
- isb-cgc:platform_reference
  - methylation_annotation
- isb-cgc:tcga_201510_alpha
  - Annotations
  - Biospecimen_data
  - Clinical_data
  - Copy_Number_segments
  - DNA_Methylation_betas
  - miRNA_expression
  - mRNA_BCGSC_HiSeq_RPKM
  - mRNA_UNC_HiSeq_RSEM
  - Protein_RPPA_data
  - Somatic_Mutation_calls
- isb-cgc:tcga_cohorts

```
SELECT
  feature,
  gene_type,
  COUNT(*) AS n
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
GROUP BY
  feature,
  gene_type
ORDER BY
  n DESC
```

Table Details: GENCODE_r19

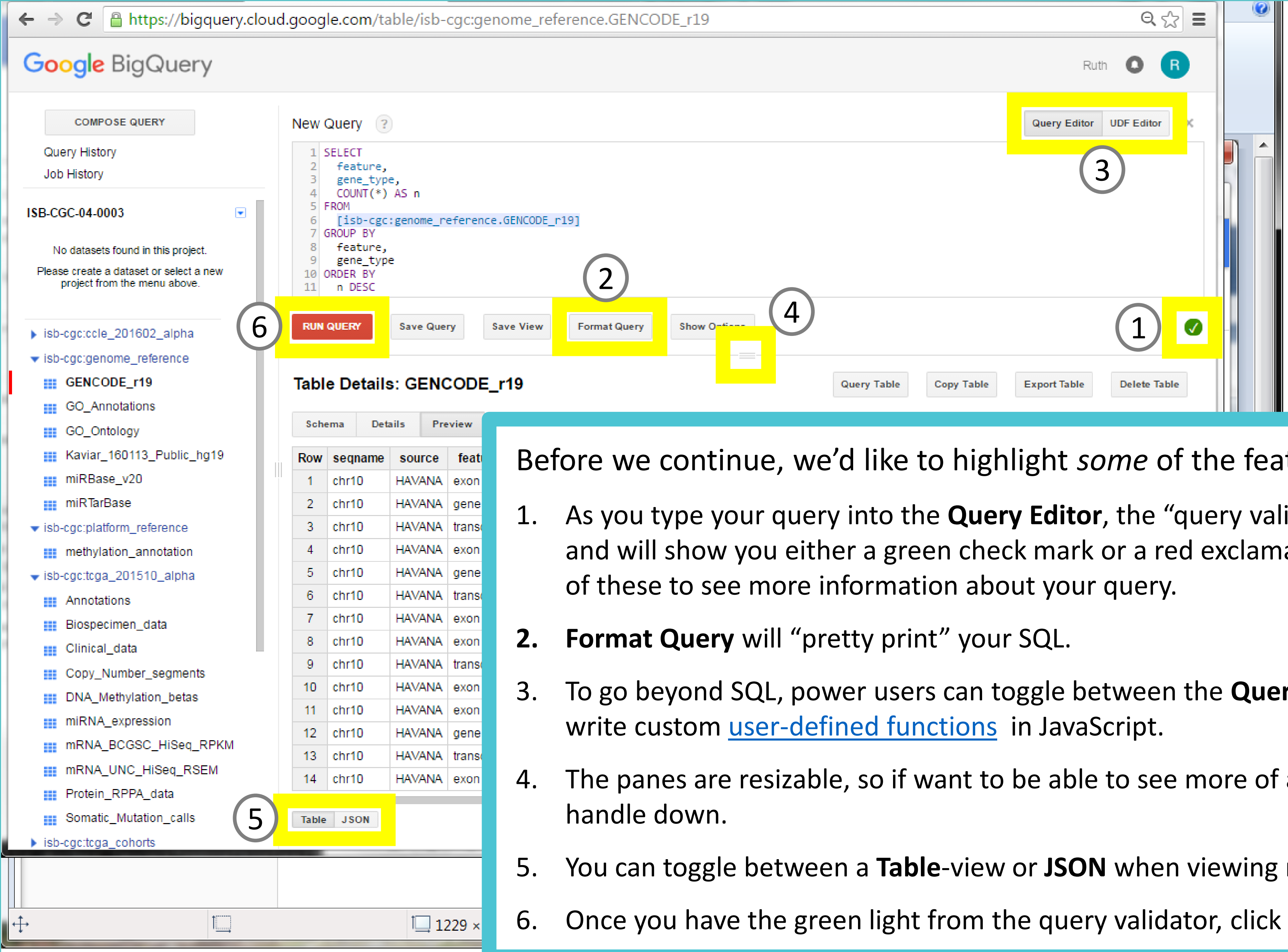
Row	seqname	source	feature	start	end	strand	frame	gene_id	transcript_
1	chr10	HAVANA	exon	93426537	93427539	-	.	ENSG00000213449.2	ENST0000045
2	chr10	HAVANA	gene	93525656	93526953	-	.	ENSG00000228759.1	ENSG0000022
3	chr10	HAVANA	transcript	93525656	93526953	-	.	ENSG00000228759.1	ENST0000042
4	chr10	HAVANA	exon	93525656	93526953	-	.	ENSG00000228759.1	ENST0000042
5	chr10	HAVANA	gene	93542596	93558048	-	.	ENSG00000228701.1	ENSG0000022
6	chr10	HAVANA	transcript	93542596	93558048	-	.	ENSG00000228701.1	ENST0000043
7	chr10	HAVANA	exon	93557994	93558048	-	.	ENSG00000228701.1	ENST0000043
8	chr10	HAVANA	exon	93542596	93542917	-	.	ENSG00000228701.1	ENST0000043
9	chr10	HAVANA	transcript	93542693	93557953	-	.	ENSG00000228701.1	ENST0000043
10	chr10	HAVANA	exon	93557570	93557953	-	.	ENSG00000228701.1	ENST0000043
11	chr10	HAVANA	exon	93542693	93542917	-	.	ENSG00000228701.1	ENST0000043
12	chr10	HAVANA	gene	93558069	93625033	+	.	ENSG00000107854.5	ENSG0000010
13	chr10	HAVANA	transcript	93558069	93625033	+	.	ENSG00000107854.5	ENST0000037
14	chr10	HAVANA	exon	93558069	93558646	+	.	ENSG00000107854.5	ENST0000037

First < Prev Rows 1 - 14 of 2619444 Next > Last

Now let's try a query. You can click on the "**Query Table**" button in the main panel or in the "**Compose Query**" button in the upper left corner.

If you're following on in your own browser, cut and paste this SQL into the **New Query** text area:

```
SELECT
  feature,
  gene_type,
  COUNT(*) AS n
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
GROUP BY
  feature,
  gene_type
ORDER BY
  n DESC
```



```
SELECT
  feature,
  gene_type,
  COUNT(*) AS n
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
GROUP BY
  feature,
  gene_type
ORDER BY
  n DESC
```

Row	seqname	source	feat
1	chr10	HAVANA	exon
2	chr10	HAVANA	gene
3	chr10	HAVANA	trans
4	chr10	HAVANA	exon
5	chr10	HAVANA	gene
6	chr10	HAVANA	trans
7	chr10	HAVANA	exon
8	chr10	HAVANA	exon
9	chr10	HAVANA	trans
10	chr10	HAVANA	exon
11	chr10	HAVANA	exon
12	chr10	HAVANA	gene
13	chr10	HAVANA	trans
14	chr10	HAVANA	exon

Before we continue, we'd like to highlight *some* of the features in the BigQuery Web UI:

1. As you type your query into the **Query Editor**, the "query validator" is automatically running, and will show you either a green check mark or a red exclamation point. You can click on either of these to see more information about your query.

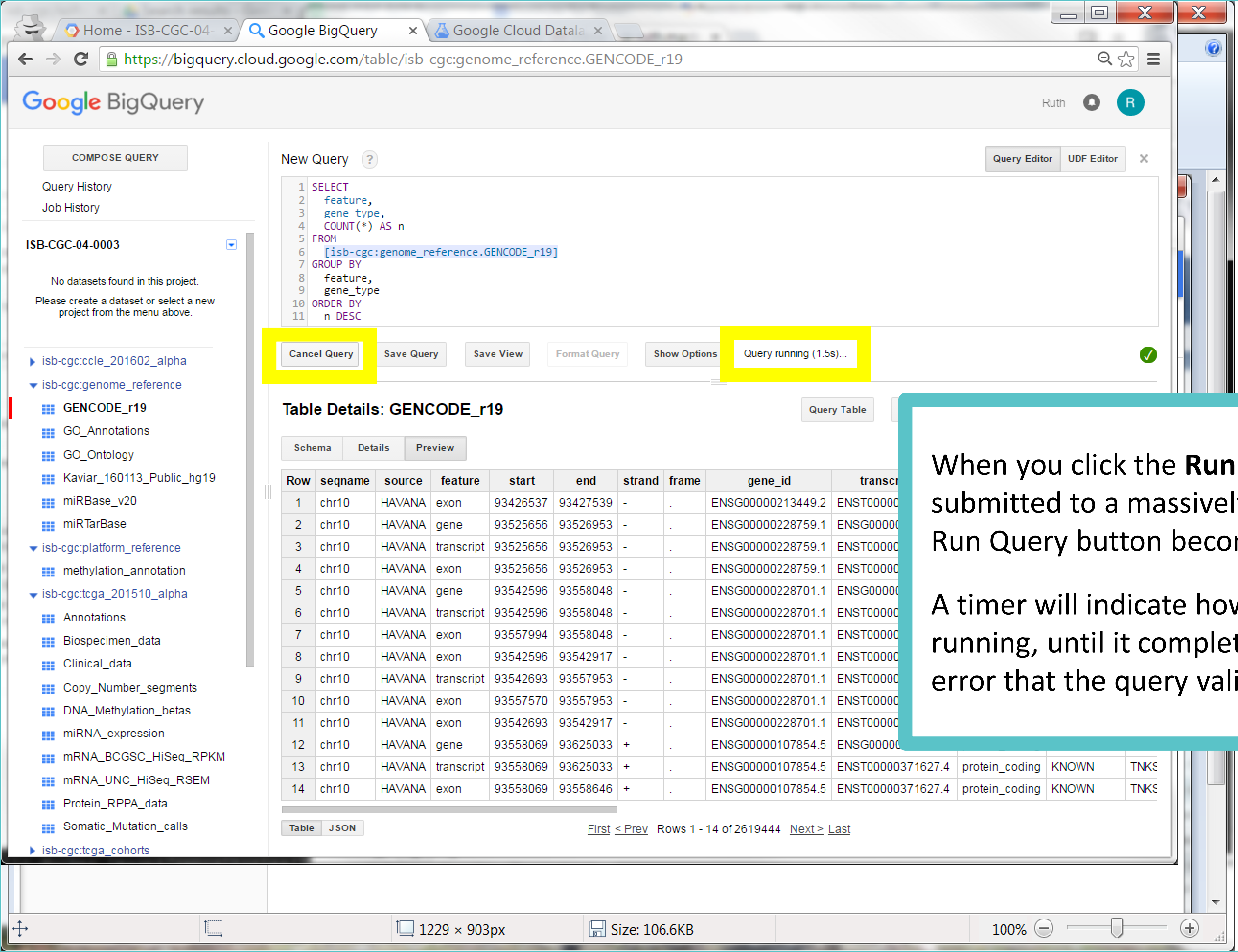
2. **Format Query** will "pretty print" your SQL.

3. To go beyond SQL, power users can toggle between the **Query Editor** and the **UDF Editor** and write custom user-defined functions in JavaScript.

4. The panes are resizable, so if want to be able to see more of a long query you can drag the sash handle down.

5. You can toggle between a **Table**-view or **JSON** when viewing results.

6. Once you have the green light from the query validator, click the red **Run Query** button.



No datasets found in this project.

Please create a dataset or select a new project from the menu above.

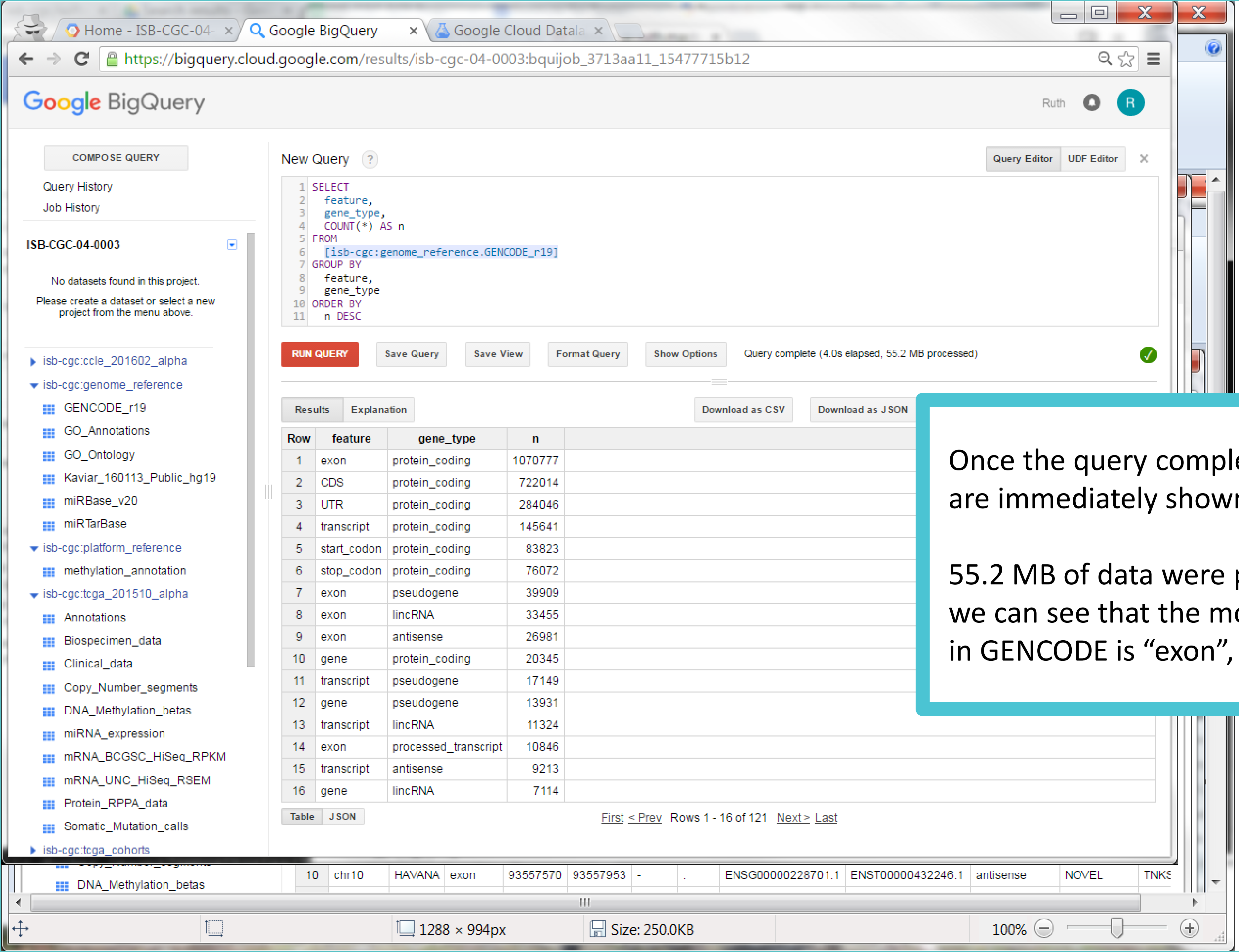
```
SELECT
  feature,
  gene_type,
  COUNT(*) AS n
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
GROUP BY
  feature,
  gene_type
ORDER BY
  n DESC
```

Table Details: GENCODE_r19

Row	seqname	source	feature	start	end	strand	frame	gene_id	transcr			
1	chr10	HAVANA	exon	93426537	93427539	-	.	ENSG00000213449.2	ENST0000			
2	chr10	HAVANA	gene	93525656	93526953	-	.	ENSG00000228759.1	ENSG0000			
3	chr10	HAVANA	transcript	93525656	93526953	-	.	ENSG00000228759.1	ENST0000			
4	chr10	HAVANA	exon	93525656	93526953	-	.	ENSG00000228759.1	ENST0000			
5	chr10	HAVANA	gene	93542596	93558048	-	.	ENSG00000228701.1	ENSG0000			
6	chr10	HAVANA	transcript	93542596	93558048	-	.	ENSG00000228701.1	ENST0000			
7	chr10	HAVANA	exon	93557994	93558048	-	.	ENSG00000228701.1	ENST0000			
8	chr10	HAVANA	exon	93542596	93542917	-	.	ENSG00000228701.1	ENST0000			
9	chr10	HAVANA	transcript	93542693	93557953	-	.	ENSG00000228701.1	ENST0000			
10	chr10	HAVANA	exon	93557570	93557953	-	.	ENSG00000228701.1	ENST0000			
11	chr10	HAVANA	exon	93542693	93542917	-	.	ENSG00000228701.1	ENST0000			
12	chr10	HAVANA	gene	93558069	93625033	+	.	ENSG00000107854.5	ENSG0000			
13	chr10	HAVANA	transcript	93558069	93625033	+	.	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TNKS
14	chr10	HAVANA	exon	93558069	93558646	+	.	ENSG00000107854.5	ENST00000371627.4	protein_coding	KNOWN	TNKS

When you click the **Run Query** button, your query is submitted to a massively parallel engine (and the Run Query button becomes a **Cancel Query** button.)

A timer will indicate how long the query has been running, until it completes (or until it encounters an error that the query validator was not able to catch).



```
SELECT
  feature,
  gene_type,
  COUNT(*) AS n
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
GROUP BY
  feature,
  gene_type
ORDER BY
  n DESC
```

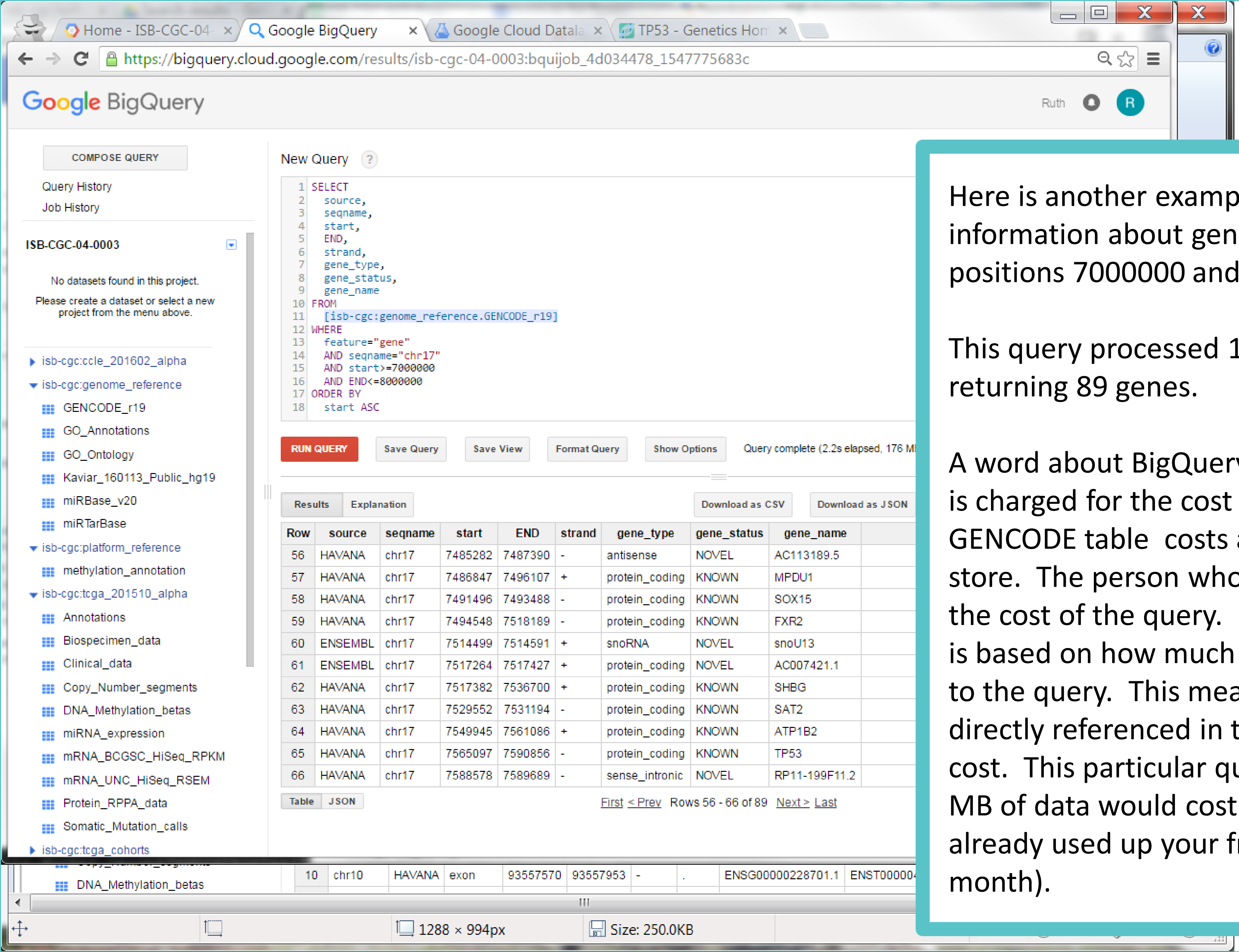
Query complete (4.0s elapsed, 55.2 MB processed)

Row	feature	gene_type	n
1	exon	protein_coding	1070777
2	CDS	protein_coding	722014
3	UTR	protein_coding	284046
4	transcript	protein_coding	145641
5	start_codon	protein_coding	83823
6	stop_codon	protein_coding	76072
7	exon	pseudogene	39909
8	exon	lincRNA	33455
9	exon	antisense	26981
10	gene	protein_coding	20345
11	transcript	pseudogene	17149
12	gene	pseudogene	13931
13	transcript	lincRNA	11324
14	exon	processed_transcript	10846
15	transcript	antisense	9213
16	gene	lincRNA	7114

10	chr10	HAVANA	exon	93557570	93557953	-	.	ENSG00000228701.1	ENST00000432246.1	antisense	NOVEL	TNKS

Once the query completes successfully, the results are immediately shown in the lower pane.

55.2 MB of data were processed in 4 seconds, and we can see that the most common type of feature in GENCODE is “exon”, followed by “CDS” *etc*



```
SELECT
  source,
  seqname,
  start,
  END,
  strand,
  gene_type,
  gene_status,
  gene_name
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
WHERE
  feature="gene"
  AND seqname="chr17"
  AND start>=7000000
  AND END<=8000000
ORDER BY
  start ASC
```

Query complete (2.2s elapsed, 176 M

Row	source	seqname	start	END	strand	gene_type	gene_status	gene_name
56	HAVANA	chr17	7485282	7487390	-	antisense	NOVEL	AC113189.5
57	HAVANA	chr17	7486847	7496107	+	protein_coding	KNOWN	MPDU1
58	HAVANA	chr17	7491496	7493488	-	protein_coding	KNOWN	SOX15
59	HAVANA	chr17	7494548	7518189	-	protein_coding	KNOWN	FXR2
60	ENSEMBL	chr17	7514499	7514591	+	snoRNA	NOVEL	snoU13
61	ENSEMBL	chr17	7517264	7517427	+	protein_coding	NOVEL	AC007421.1
62	HAVANA	chr17	7517382	7536700	+	protein_coding	KNOWN	SHBG
63	HAVANA	chr17	7529552	7531194	-	protein_coding	KNOWN	SAT2
64	HAVANA	chr17	7549945	7561086	+	protein_coding	KNOWN	ATP1B2
65	HAVANA	chr17	7565097	7590856	-	protein_coding	KNOWN	TP53
66	HAVANA	chr17	7588578	7589689	-	sense_intronic	NOVEL	RP11-199F11.2

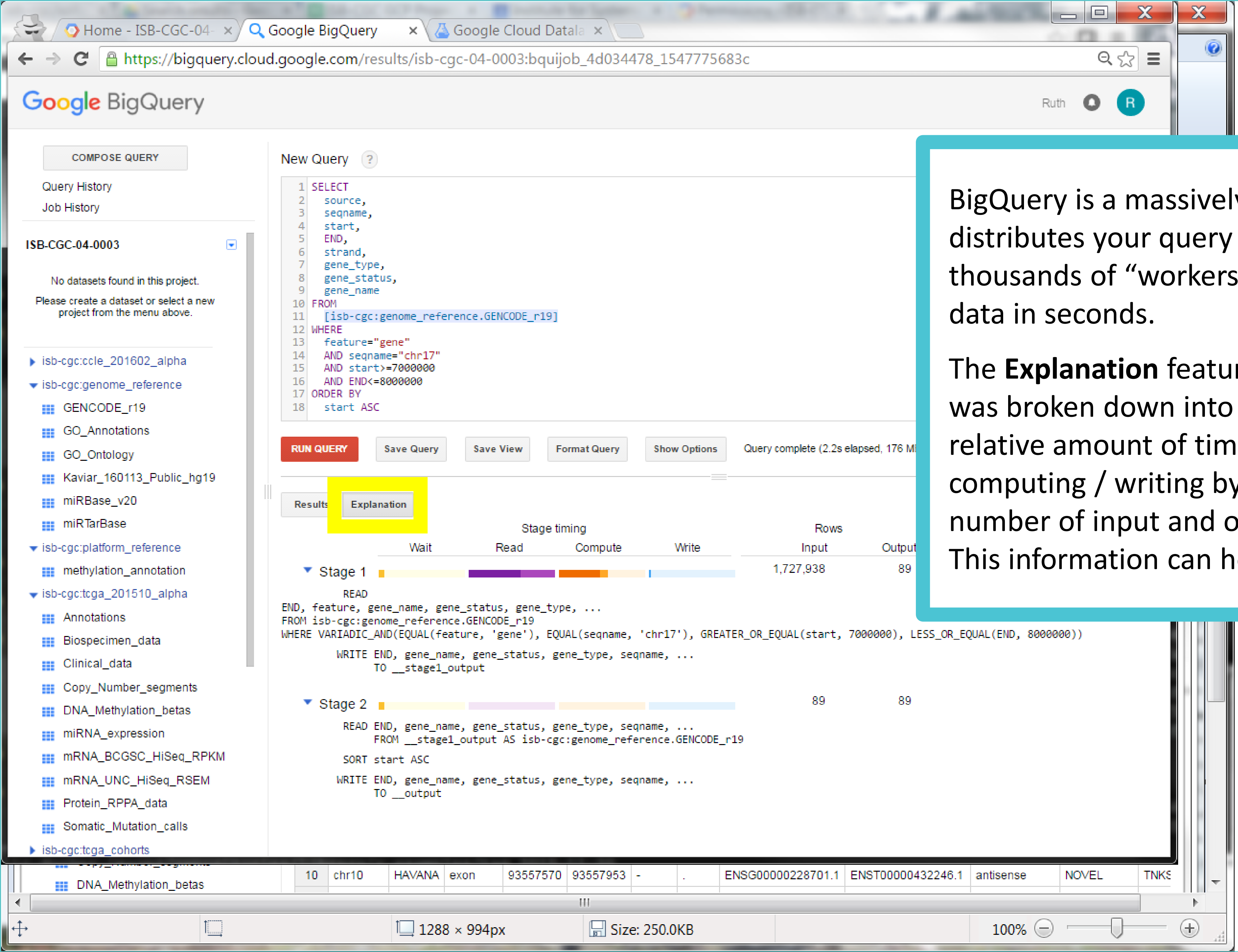
Rows 56 - 66 of 89

10	chr10	HAVANA	exon	93557570	93557953	-	.	ENSG00000228701.1	ENST00000

Here is another example query, which asks for information about genes on chr17 between positions 7000000 and 8000000.

This query processed 176 MB in just 2.2 seconds, returning 89 genes.

A word about BigQuery costs. The owner of a table is charged for the cost of the storage, and this GENCODE table costs about 7 cents per year to store. The person who runs a query gets charged the cost of the query. For most queries, this charge is based on how much data is "scanned" to respond to the query. This means only columns that are directly referenced in the query count towards the cost. This particular query, which processed 176 MB of data would cost less than one cent (if you've already used up your free $5 worth of queries this month).



```
SELECT
  source,
  seqname,
  start,
  END,
  strand,
  gene_type,
  gene_status,
  gene_name
FROM
  [isb-cgc:genome_reference.GENCODE_r19]
WHERE
  feature="gene"
  AND seqname="chr17"
  AND start>=7000000
  AND END<=8000000
ORDER BY
  start ASC
```

Query complete (2.2s elapsed, 176 M

	Stage timing				Rows	
	Wait	Read	Compute	Write	Input	Output
Stage 1					1,727,938	89
Stage 2					89	89

Stage 1:

```
        READ
END, feature, gene_name, gene_status, gene_type, ...
FROM isb-cgc:genome_reference.GENCODE_r19
WHERE VARIADIC_AND(EQUAL(feature, 'gene'), EQUAL(seqname, 'chr17'), GREATER_OR_EQUAL(start, 7000000), LESS_OR_EQUAL(END, 8000000))

       WRITE END, gene_name, gene_status, gene_type, seqname, ...
             TO __stage1_output
```

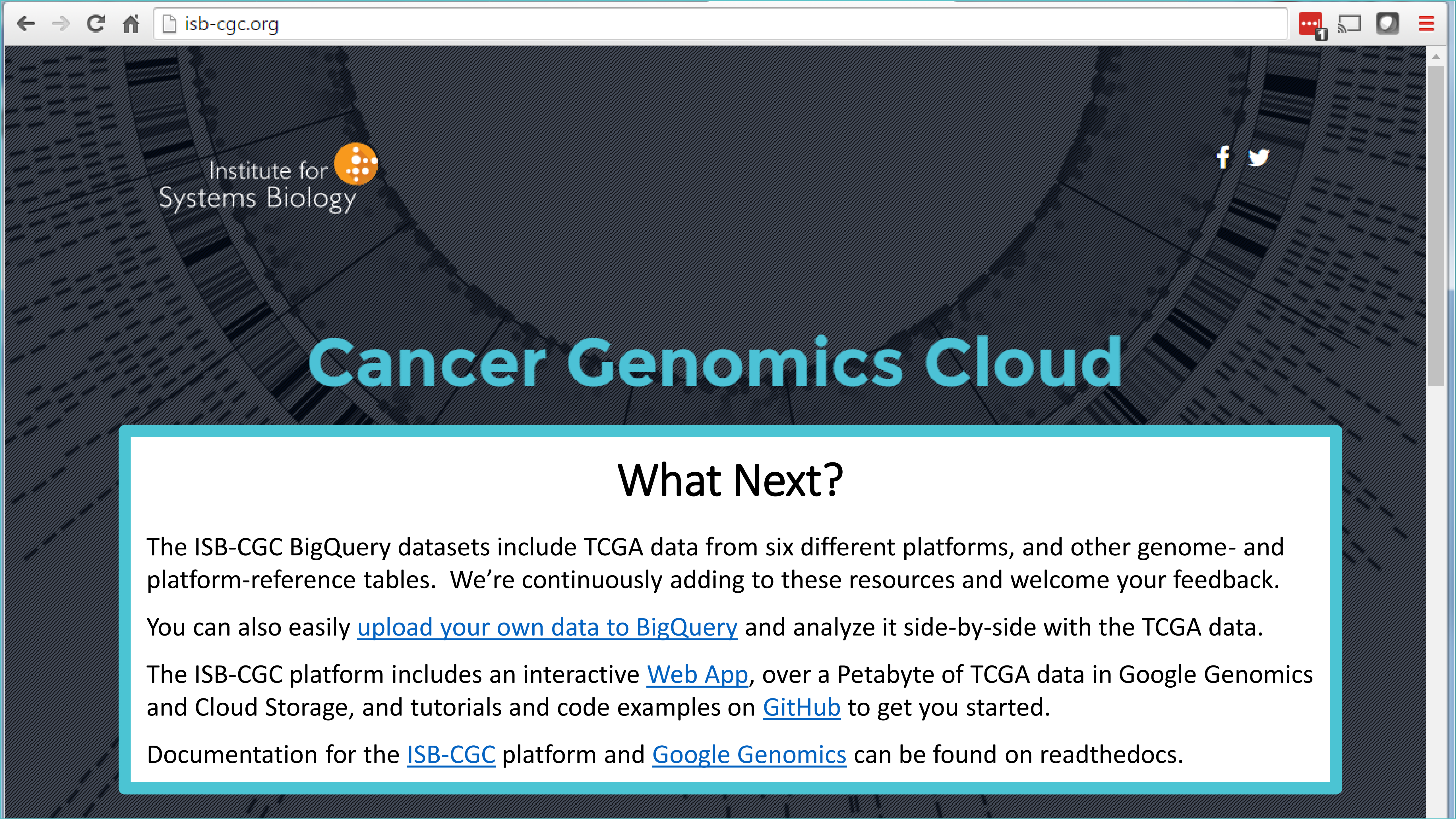
Stage 2:

```
        READ END, gene_name, gene_status, gene_type, seqname, ...
             FROM __stage1_output AS isb-cgc:genome_reference.GENCODE_r19

        SORT start ASC

       WRITE END, gene_name, gene_status, gene_type, seqname, ...
             TO __output
```

10	chr10	HAVANA	exon	93557570	93557953	-	.	ENSG00000228701.1	ENST00000432246.1	antisense	NOVEL	TNKS

BigQuery is a massively parallel engine which distributes your query across hundreds or thousands of “workers” and can scan terabytes of data in seconds.

The **Explanation** feature shows you how your query was broken down into a series of stages, the relative amount of time spent waiting / reading / computing / writing by the “workers”, and the number of input and output rows at each stage. This information can help you optimize your query.



Cancer Genomics Cloud

What Next?

The ISB-CGC BigQuery datasets include TCGA data from six different platforms, and other genome- and platform-reference tables. We're continuously adding to these resources and welcome your feedback.

You can also easily upload your own data to BigQuery and analyze it side-by-side with the TCGA data.

The ISB-CGC platform includes an interactive Web App, over a Petabyte of TCGA data in Google Genomics and Cloud Storage, and tutorials and code examples on GitHub to get you started.

Documentation for the ISB-CGC platform and Google Genomics can be found on readthedocs.