

```
In [1]: import hail as hl

# hl.init(backend="spark")

# Say, "the default sans-serif font is COMIC SANS"
# matplotlib.rcParams['font.sans-serif'] = "Arial"
# Then, "ALWAYS use sans-serif fonts"
# matplotlib.rcParams['font.family'] = "sans-serif"
```

```
In [2]: # import holoviews as hv
import numpy as np
import pandas as pd

from bokeh.layouts import column, gridplot, Plot, row
from bokeh.models import *
from bokeh.plotting import *
from bokeh.palettes import Category10, Spectral6
from gnomad import *
from gnomad.sample_qc.ancestry import POP_COLORS
from gnomad.utils.filtering import add_filters_expr
from gnomad.utils.plotting import *
# from gnomad_qc.v3.resources import *
from IPython.core.display import display, HTML
from statsmodels.robust.scale import mad

# hv.extension("bokeh")

TOOLS = "hover,save,pan,box_zoom,reset,wheel_zoom"
output_notebook()
display(HTML("<style>.container { width:100% !important; }</style>"))

if 'old_show' not in dir():
    old_show = hl.Table.show
    def new_show(t, n=10, width=90, truncate=None, types=True):
        old_show(t, 10, 170, 40)
    hl.Table.show = new_show
```

```
/tmp/ipykernel_37220/3819452698.py:14: DeprecationWarning: Importing display
from IPython.core.display is deprecated since IPython 7.14, please import fro
m IPython display
from IPython.core.display import display, HTML
```



BokehJS 3.1.1 successfully loaded.

```
In [3]: rgp_path = 'gs://marten-seqr-sandbox-storage/rgp_entire_mt_from_vcf_20240404
rgp_mt = hl.read_matrix_table(rgp_path)
```

```

Initializing Hail with default parameters...
/opt/conda/miniconda3/lib/python3.10/site-packages/hailtop/aiocloud/aiogoogle/user_config.py:43: UserWarning: Reading spark-defaults.conf to determine GCS requester pays configuration. This is deprecated. Please use `hailctl config set gcs_requester_pays/project` and `hailctl config set gcs_requester_pays/buckets`.
  warnings.warn(
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
SPARKMONITOR_LISTENER: Started SparkListener for Jupyter Notebook
SPARKMONITOR_LISTENER: Port obtained from environment: 51611
SPARKMONITOR_LISTENER: Application Started: application_1712844629510_0002
...Start Time: 1712858699407

Running on Apache Spark version 3.3.0
SparkUI available at http://dmdragen-m.us-central1-b.c.marten-seqr-sandbox-9123.internal:43107
Welcome to

      <>
  //  //  //  //  //
 /  _  /  _  \  //  //
/_/  /_/\_/_/  //  //
LOGGING: writing to /home/hail/hail-20240411-1804-0.2.120-f00f916faf78.log

```

```
In [4]: dragen_path = 'gs://marten-seqr-sandbox-storage/dragen_entire_mt_from_vcf_20240411-1804-0.2.120-f00f916faf78'
dragen_mt = hl.read_matrix_table(dragen_path)
```

```
In [5]: # Hana queries:
# 1: Remove the X chromosome - Stephanie mentioned that in GATK males
# -> on the X chromosome have a ton of GQs between 30 and 40m so that might
# 2: Look at the GQ distribution for non-ref calls only, to see if there are
```

```
In [8]: sex_set = hl.set({"chrX", "chrY"})
```

```
In [9]: dragen_mt_autosomes = dragen_mt.filter_rows(~sex_set.contains(dragen_mt.locus))
```

```
In [10]: dragen_mt_nonref = dragen_mt.filter_entries(dragen_mt.GT.is_non_ref())
```

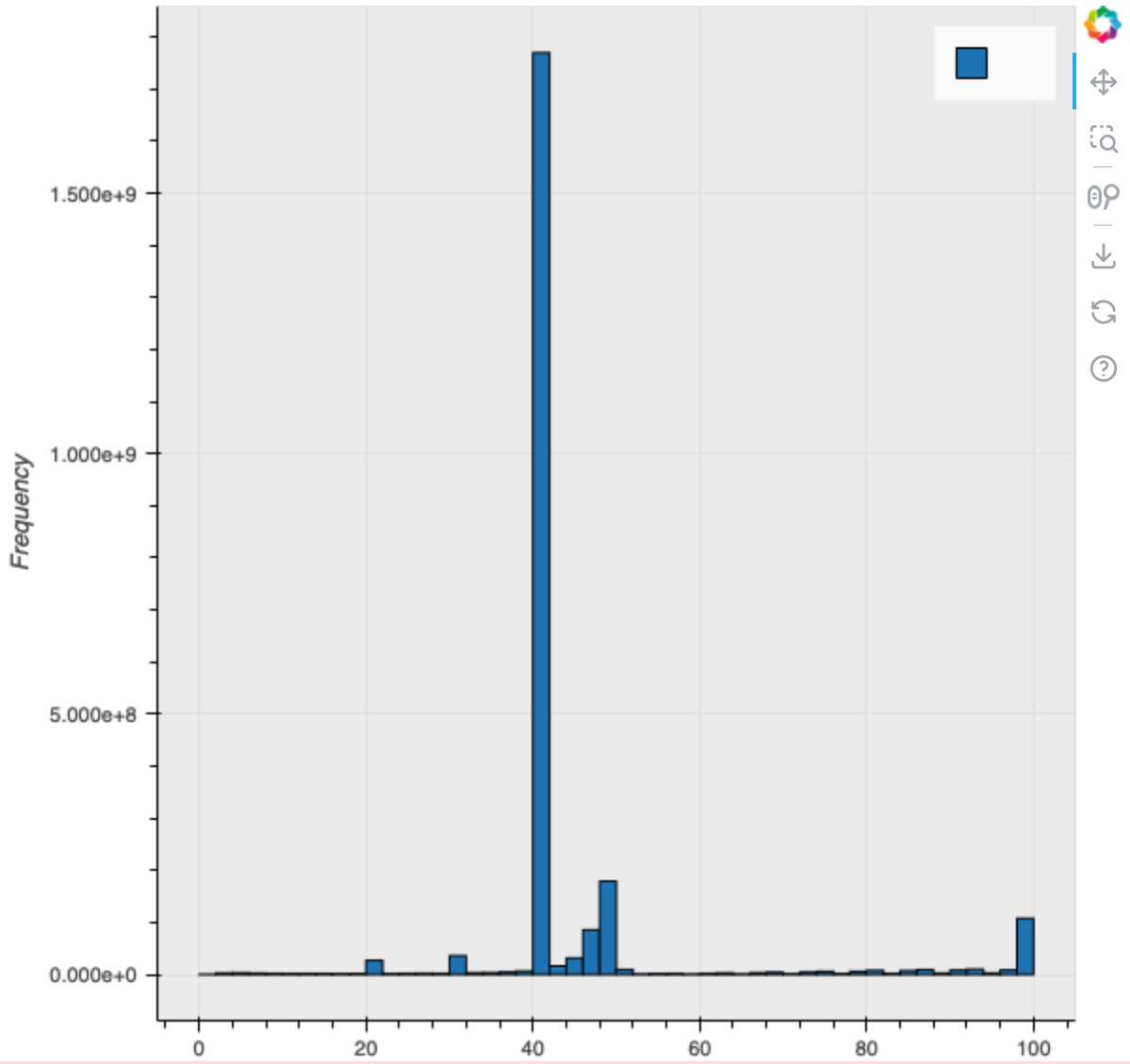
```
In [16]: # GQ distribution: overall vs autosomes vs nonref
```

```
In [17]: hail_expr_set = {'All':dragen_mt.GQ , 'Autosomes':dragen_mt_autosomes.GQ , 'NonRef':dragen_mt_nonref.GQ}
```

```
In [19]: for hail_title, hail_expr in hail_expr_set.items():
  plotting_obj = hl.plot.histogram(hail_expr, range=(0, 100))
  plotting_obj.title = f"GQ of DRAGEN {hail_title}"
  plotting_obj.title.text_font_size = '18pt'
  show(plotting_obj)
```

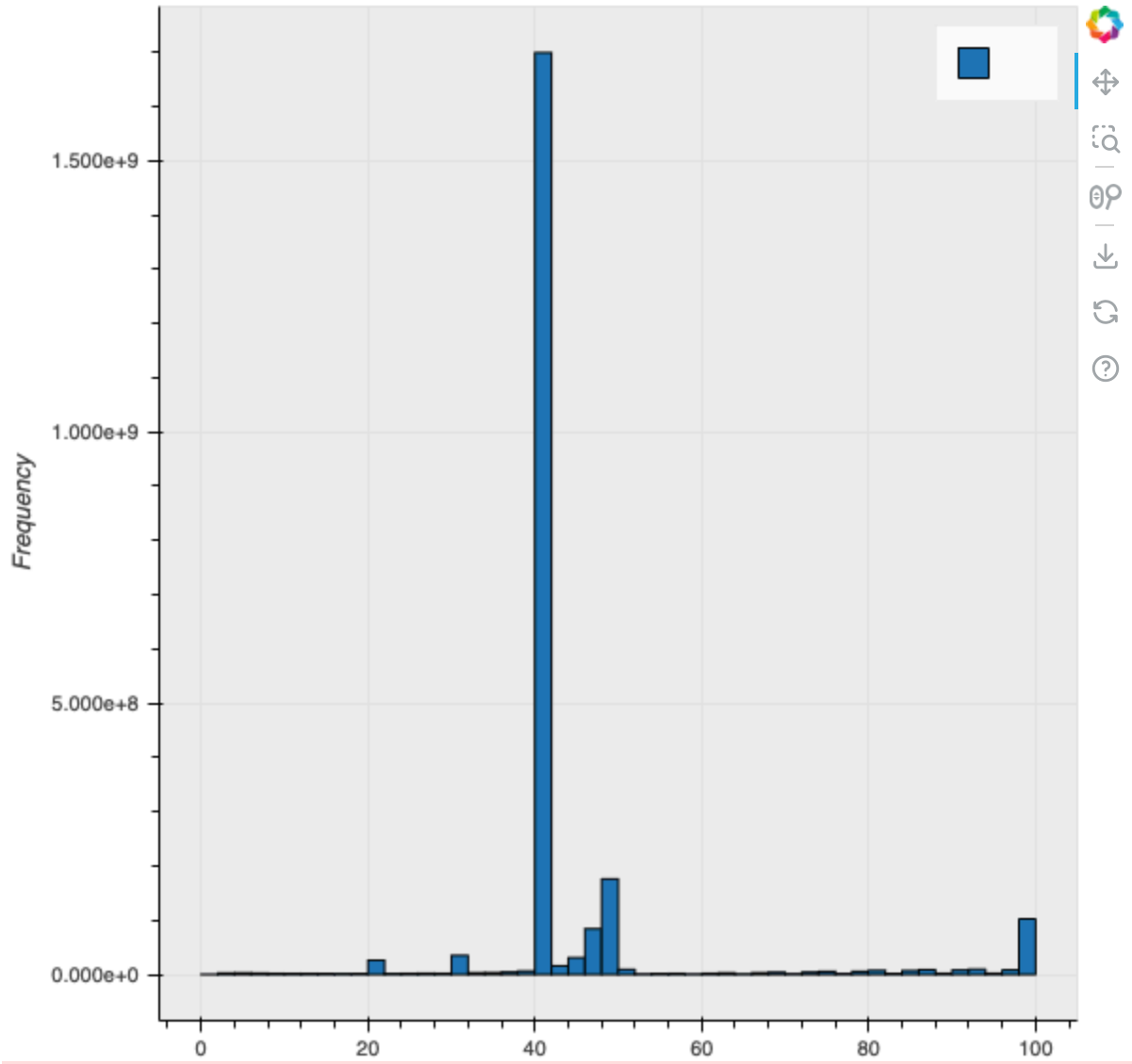
```
[Stage 4:===== (510 + 1) / 510] [Stage 5:===== > (10 + 2) / 11]
```

GQ of DRAGEN AII



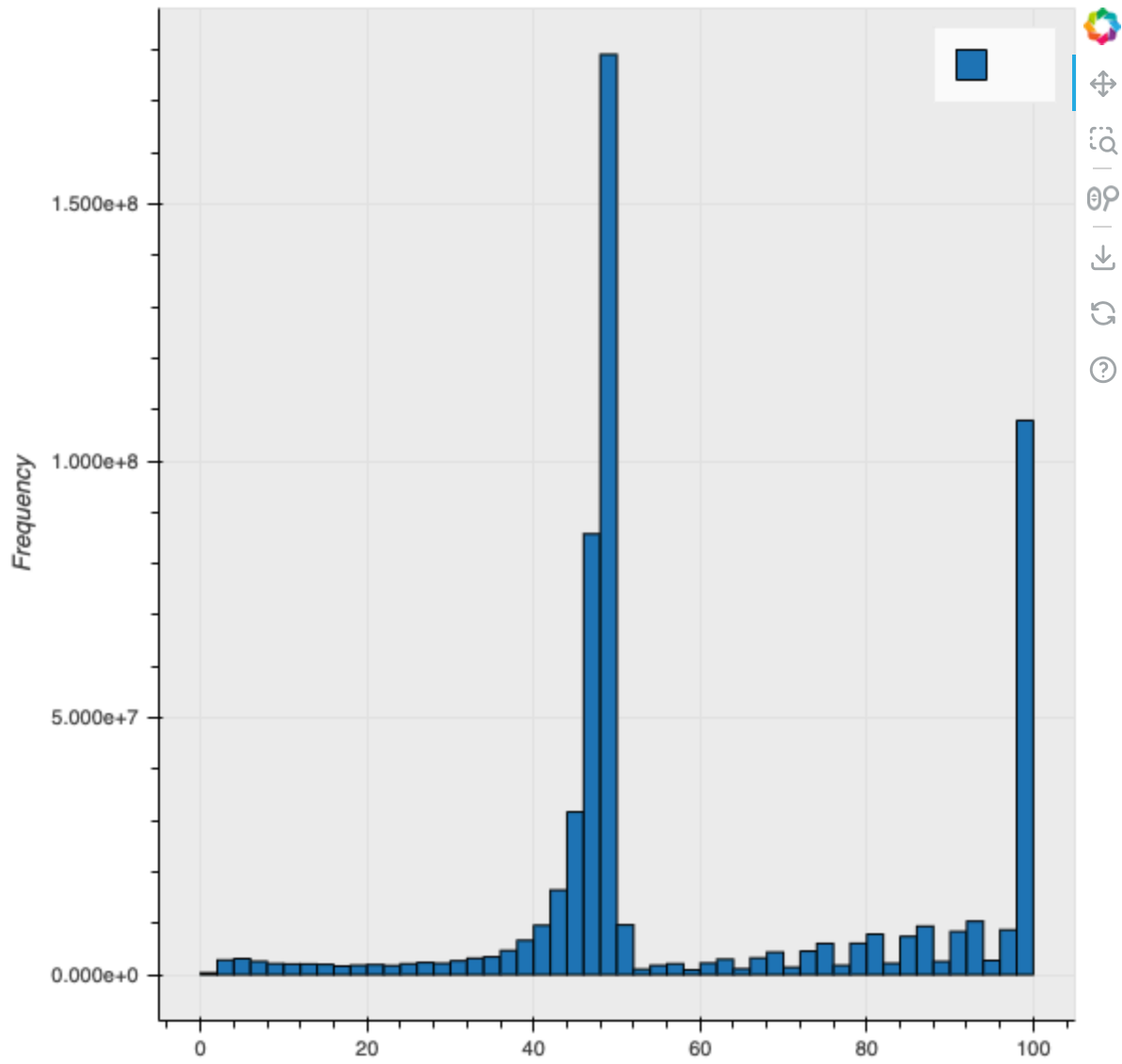
[Stage 6: =====(510 + 1) / 5
10]

GQ of DRAGEN Autosomes



[Stage 8: =====>(509 + 1) / 5
10]

GQ of DRAGEN Nonref



In []: