VMH graphs

combined

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
# seaborn style
sns.set_context("talk")
sns.set_style("whitegrid")
plt.tight_layout() # for saving
```

```
<Figure size 432x288 with 0 Axes>
```

In [3]:
```
# categories
con_set = {1,2,3,4,9,11,14,15,17,18,21,22}
csd_set = {5,6,7,8,10,12,13,16,19,20,23,24}
```

In [4]:
```
# data file
f = "Data/02/processed/vmh_01.csv"
df1 = pd.read_csv(f)
```

In [5]:
```
# drop NaN from Sample Name
df1 = df1.dropna(subset =['Sample Name'])
df1['Sample Name'] = df1['Sample Name'].str.split(expand = Tru
e)[1]
df1['Sample Name'] = pd.to_numeric(df1['Sample Name'], errors = 'c
oerce')
df1 = df1.dropna(subset =['Sample Name'])
df1['Sample Name'] = df1['Sample Name'].astype(int)

# drop expfail = Y
df1 = df1[df1['EXPFAIL'] == 'N']

# convert target name to lower case to stay consistent with previo
us data files
df1['Target Name']  = df1['Target Name'].str.lower()

# fix typo acbt
df1.loc[ df1['Target Name'] == 'acbt', 'Target Name'] = 'actb'
```

In [6]:
```
# data file #2
f = "Data/02/processed/vmh_02.csv"
df2 = pd.read_csv(f)
```

In [7]:
```
# drop NaN from Sample Name
df2 = df2.dropna(subset =['Sample Name'])
df2['Sample Name'] = df2['Sample Name'].str.split(expand = Tru
e)[1]
df2['Sample Name'] = pd.to_numeric(df2['Sample Name'], errors = 'c
oerce')
df2 = df2.dropna(subset =['Sample Name'])
df2['Sample Name'] = df2['Sample Name'].astype(int)

# drop expfail = Y
#df2 = df2[df2['EXPFAIL'] == 'N']


# convert target name to lower case to stay consistent with previo
us data files
df2['Target Name']  = df2['Target Name'].str.lower()

# fix typo acbt
df2.loc[ df2['Target Name'] == 'acbt', 'Target Name'] = 'actb'
```

In [8]:
```
# data file #3
f = "Data/02/processed/vmh_03.csv"
df3 = pd.read_csv(f)
```

In [9]:
```
# drop NaN from Sample Name
df3 = df3.dropna(subset =['Sample Name'])
df3['Sample Name'] = df3['Sample Name'].str.split(expand = Tru
e)[1]
df3['Sample Name'] = pd.to_numeric(df2['Sample Name'], errors = 'c
oerce')
df3 = df3.dropna(subset =['Sample Name'])
df3['Sample Name'] = df3['Sample Name'].astype(int)

# drop expfail = Y
#df2 = df2[df2['EXPFAIL'] == 'N']


# convert target name to lower case to stay consistent with previo
us data files
df3['Target Name']  = df3['Target Name'].str.lower()

# fix typo acbt
df3.loc[ df3['Target Name'] == 'acbt', 'Target Name'] = 'actb'
```

In [10]:
```
# append dataframes
df1 = df1.append(df2)
df1 = df1.append(df3)
```

In [11]:
```
df1["CON"] = df1["Sample Name"].isin(con_set)
df1['CSD'] = df1["Sample Name"].isin(csd_set)

df1["Check Category"] = df1["CON"] ^ df1["CSD"]
```

In [12]:
```
# check if all the samples are labeled (OK if 0)
df1[df1['Check Category'] == False].size
```

Out[12]:
```
0
```

In [13]:
```
# create string column with category labels
df1['Category'] = 'CSD'
df1.loc[ df1['CON'] == True, 'Category'] = 'CON'
```

In [14]:
```
# fix data types
df1['Ct'] = pd.to_numeric( df1['Ct'],errors ='coerce')
df1 = df1.dropna(subset =['Ct'])
```

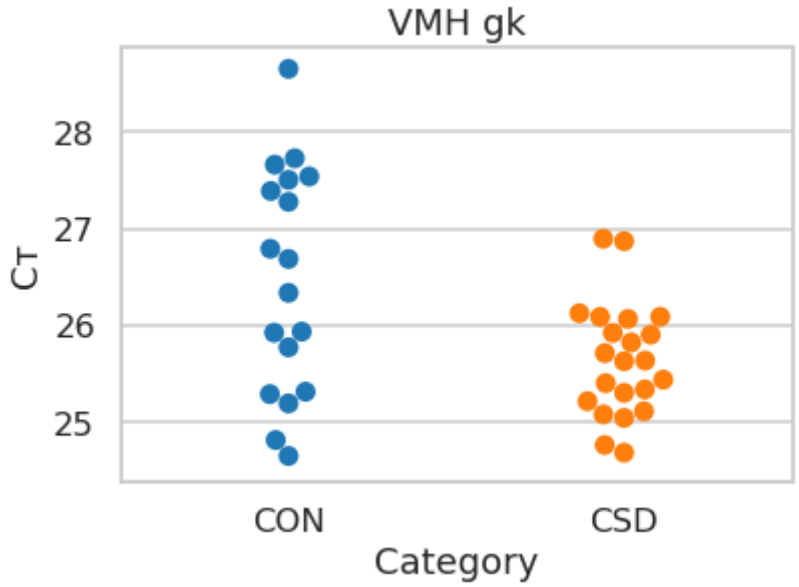
In [15]:
```
df1 = df1[df1['Sample Name'] != 18]  # drop sample 18
```

In [16]:
```
# drop samples with Ct mean > 29

for sn in df1[df1['Ct Mean']>29]['Sample Name'].unique():
    df1 = df1[df1['Sample Name'] != sn]
```

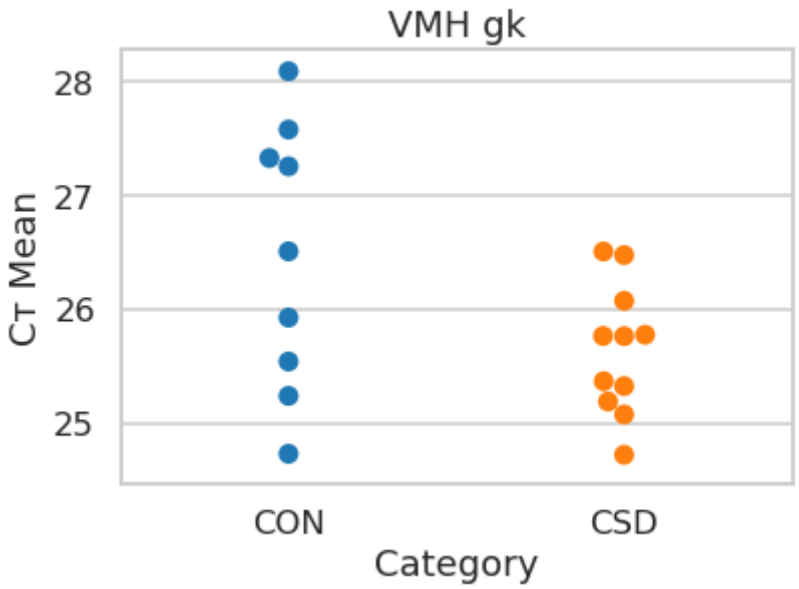
Generating plots

gk

In [17]:
```
plt.title("VMH gk")
sns.swarmplot(x='Category', y='Ct', data=df1[df1['Target Name'] ==
'gk'], size = 10)
plt.savefig("vmh_gk_ct.png",  bbox_inches = 'tight', dpi = 300)
```

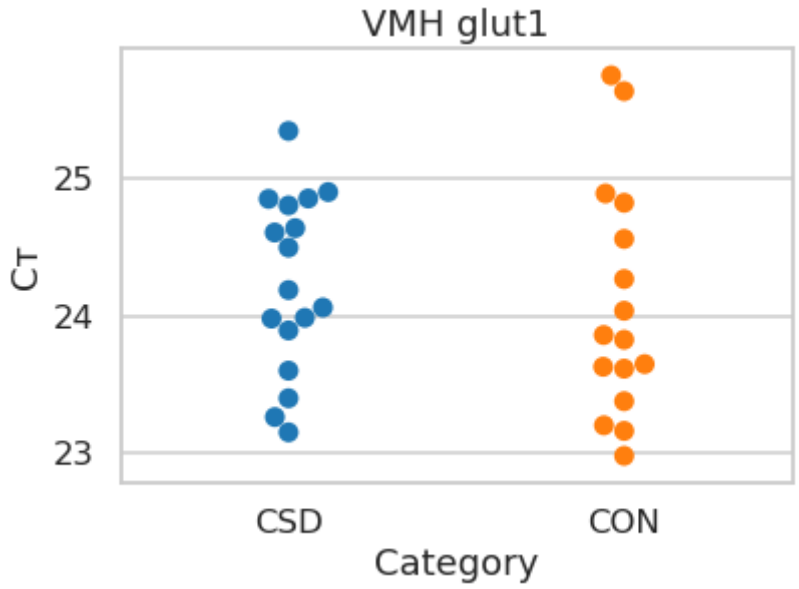


In [18]:
```
plt.title("VMH gk")
sns.swarmplot(x='Category', y='Cт Mean', data=df1[df1['Target Name
'] == 'gk'].drop_duplicates(subset=['Sample Name']), size = 10)
plt.savefig("vmh_gk_ctmean.png",  bbox_inches = 'tight', dpi = 30
0)
```

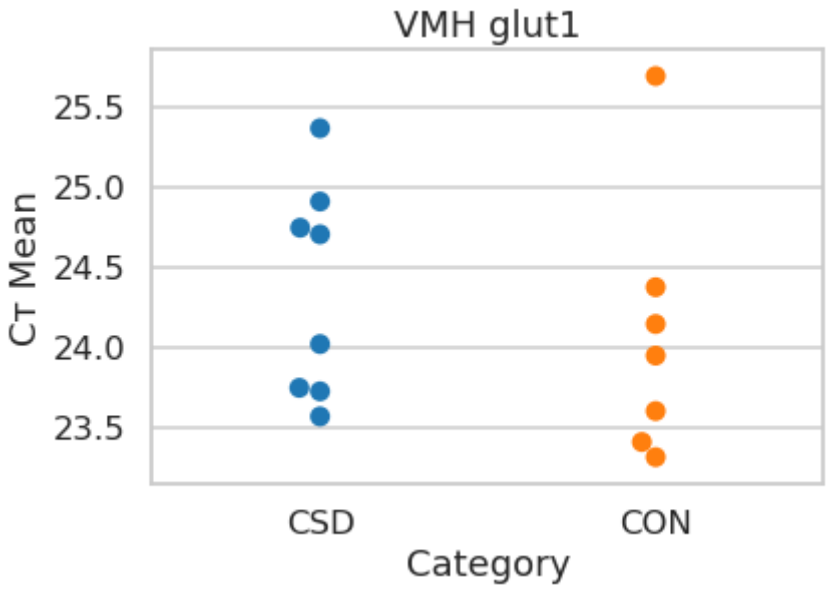


glut1

In [19]:
```
plt.title("VMH glut1")
sns.swarmplot(x='Category', y='Cт', data=df1[df1['Target Name'] ==
'glut1'], size = 10)
plt.savefig("vmh_glut1_ct.png",  bbox_inches = 'tight', dpi = 300)
```

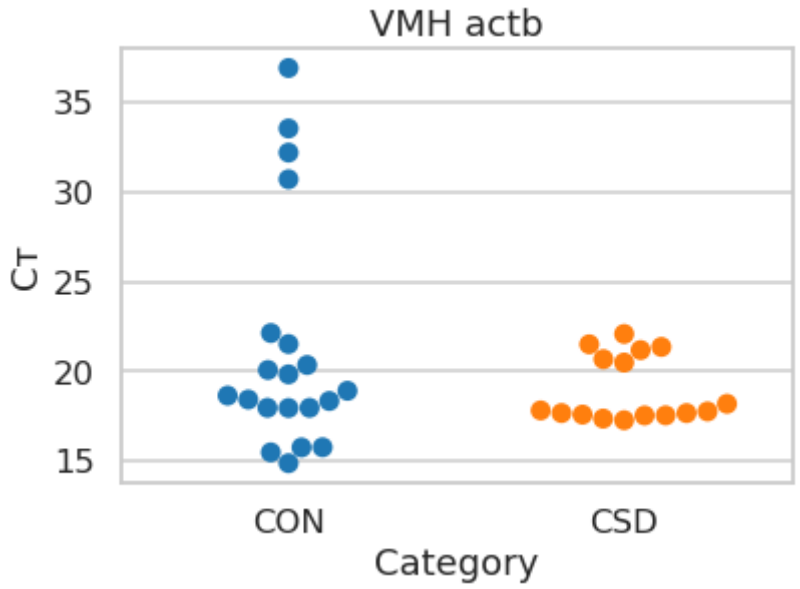


In [20]:
```
plt.title("VMH glut1")
sns.swarmplot(x='Category', y='Cт Mean', data=df1[df1['Target Name
'] == 'glut1'].drop_duplicates(subset=['Sample Name']), size = 10)
plt.savefig("vmh_glut1_ctmean.png",  bbox_inches = 'tight', dpi =
300)
```

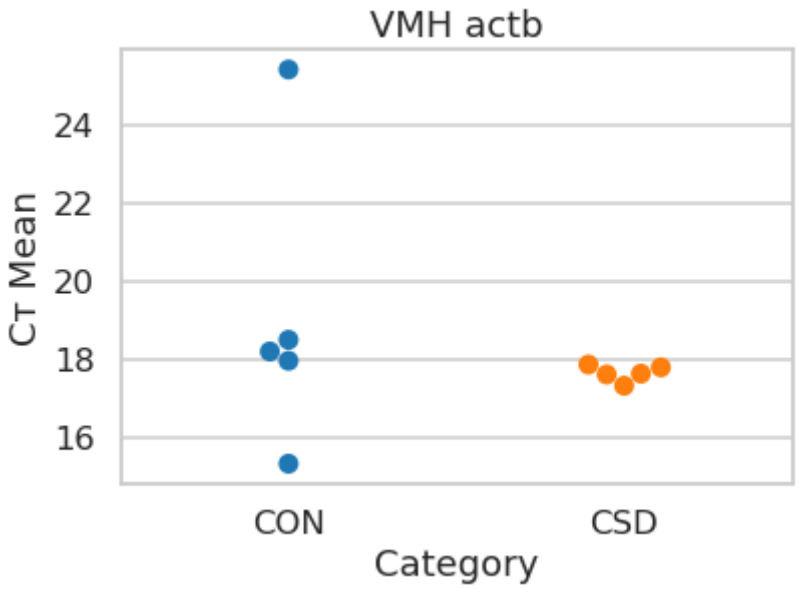


actb

In [21]:
```
plt.title("VMH actb")
sns.swarmplot(x='Category', y='Cт', data=df1[df1['Target Name'] ==
'actb'], size = 10)
plt.savefig("vmh_actb_ct.png",  bbox_inches = 'tight', dpi = 300)
```



In [22]:
```
plt.title("VMH actb")
sns.swarmplot(x='Category', y='Cт Mean', data=df1[df1['Target Name
'] == 'actb'].drop_duplicates(subset=['Sample Name']), size = 10)
plt.savefig("vmh_actb_ctmean.png",  bbox_inches = 'tight', dpi = 3
00)
```



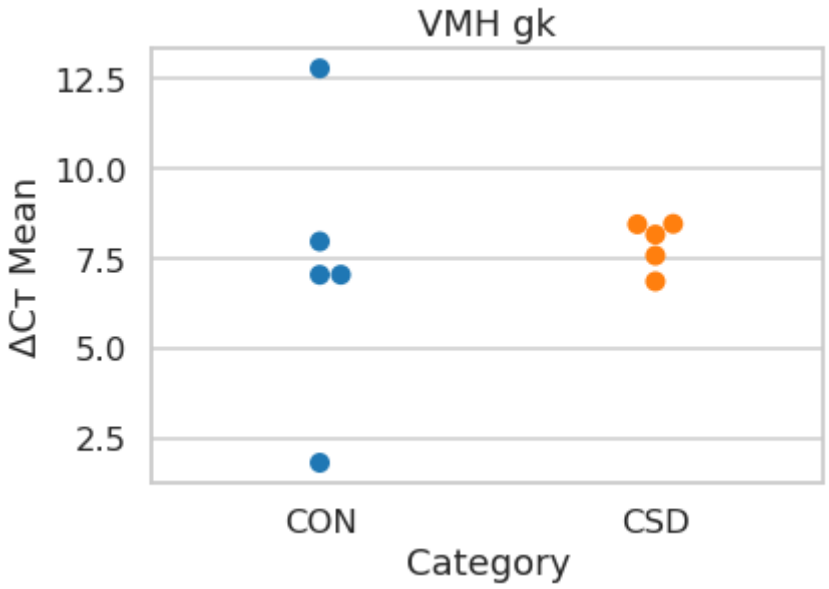
delta Ct

In [23]:
```
data_gk = df1[df1['Target Name'] == 'gk'].drop_duplicates(subset=
['Sample Name']).loc[:,['Sample Name','Cт Mean','Category']]
data_glut1 = df1[df1['Target Name'] == 'glut1'].drop_duplicates(su
bset=['Sample Name']).loc[:,['Sample Name','Cт Mean','Category']]
data_actb = df1[df1['Target Name'] == 'actb'].drop_duplicates(subs
et=['Sample Name']).loc[:,['Sample Name','Cт Mean','Category']]

data_actb1 = data_actb.drop(columns = ['Category']).rename({'Cт Me
an':'Cт Mean actb'}, axis = 'columns')

data_gk = data_gk.merge(data_actb1[['Sample Name', 'Cт Mean actb
']])
data_gk['ΔCт Mean'] = data_gk['Cт Mean'] -  data_gk['Cт Mean actb
']

data_glut1 = data_glut1.merge(data_actb1[['Sample Name', 'Cт Mean
actb']])
data_glut1['ΔCт Mean'] = data_glut1['Cт Mean'] -  data_glut1['Cт M
ean actb']
```

In [24]:
```
data_gk.to_csv('vhm_gk.csv')
data_glut1.to_csv('vhm_glut1.csv')
```

gk

In [25]:
```
plt.title("VMH gk")
sns.swarmplot(x='Category', y='ΔCт Mean', data=data_gk, size = 10)
plt.savefig("vmh_gk_delta_ctmean.png",  bbox_inches = 'tight', dpi
= 300)
```



glut1

In [26]:
```
plt.title("VMH glut1")
sns.swarmplot(x='Category', y='ΔCт Mean', data=data_glut1, size =
10)
plt.savefig("vmh_glut1_delta_ctmean.png",  bbox_inches = 'tight',
dpi = 300)
```