In [97]:
```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [98]:
```
cwd = os.getcwd()
cwd
```

Out[98]:
```
'/Users/ibergeland/Box Sync/UCSF Projects'
```

In [170]:
```
file1 = 'DPP feasibility.xlsx'
file2 = 'miR hits.xlsx'

x1 = pd.read_excel(file1)
x2 = pd.read_excel(file2)

print(x1.shape)
print(x2.shape)
```

```
(126, 5)
(66, 2)
```

In [171]:
```
x1.head()
```

Out[171]:

	Median	Mean	SD	discovery	FP CVD
hsa_mir_24_2_3p	0.000000	0.011628	0.152499	NaN	NaN
hsa_mir_320b_1_5p	0.000000	0.011628	0.152499	NaN	NaN
hsa_mir_375_5p	0.000000	0.019314	0.253295	NaN	NaN
hsa_mir_652_3p	9.834457	9.838846	0.331363	NaN	NaN
hsa_mir_326_5p	0.000000	0.061408	0.340012	NaN	NaN

In [172]:
```
x2.head()
```

Out[172]:

	DISCOVERY	FP CVD
a1	hsa_mir_652_3p	hsa_let_7d_5p
a1	hsa_mir_140_3p	hsa_let_7e_5p
a1	hsa_mir_24_3p	hsa_let_7g_5p
a1	hsa_mir_532_5p	hsa_let_7i_5p
a1	hsa_mir_20b_5p	hsa_miR_1_3p

In [173]:
```
####
DISCOVERY = x2.iloc[:,0].values
FP_CV = x2.iloc[:,1].values
```

In [174]:
```
#Create list of tuples of index values contained in both dataframes
testA = set(DISCOVERY) ## Discovery values
testB = set(FP_CV) ## FP_CV values
testC = set(x1.index.values) ##DPP index values

discovery_match = list(testC.intersection(testA)) ## returns values cont
ained in discovery column and DPP index
FPcv_match = list(testC.intersection(testB)) ## returns values contained
in FP_CV column and DPP index
```

In [175]:
```
print(testA)
```

```
{'hsa_mir_130b_3p', 'hsa_mir_342_3p', 'hsa_mir_326', 'hsa_mir_345_5p',
'hsa_mir_98_5p', 'hsa_mir_197_5p', 'hsa_mir_425_3p', 'hsa_mir_877_5p',
'hsa_mir_221_3p', 'hsa_mir_323a_3p', 'hsa_mir_874_3p', 'hsa_mir_23a_3
p', 'hsa_mir_186_5p', 'hsa_mir_374b_5p', 'hsa_let_7c_5p', 'hsa_mir_181c
_3p', 'hsa_mir_363_3p', 'hsa_mir_197_3p', 'hsa_mir_140_3p', 'hsa_mir_19
2_5p', 'hsa_mir_151a_5p', 'hsa_mir_106b_5p', 'hsa_mir_503_5p', 'hsa_mir
_422a', 'hsa_mir_330_3p', 'hsa_mir_424_5p', 'hsa_mir_92a_3p', 'hsa_mir_
133b', 'hsa_mir_532_5p', 'hsa_mir_296_5p', 'hsa_mir_203a_3p', 'hsa_mir_
20b_5p', 'hsa_mir_27a_3p', 'hsa_mir_342_5p', 'hsa_mir_320c', 'hsa_mir_2
3b_5p', 'hsa_mir_215_5p', 'hsa_mir_93_5p', 'hsa_mir_151b', 'hsa_mir_205
_5p', 'hsa_mir_151a_3p', 'hsa_mir_126_5p', 'hsa_mir_486_3p', 'hsa_let_7
f_5p', 'hsa_mir_30a_5p', 'hsa_mir_652_3p', 'hsa_mir_145_5p', 'hsa_mir_2
06', 'hsa_mir_24_3p', 'hsa_mir_379_5p', 'hsa_mir_136_3p', 'hsa_mir_1287
_5p', 'hsa_mir_126_3p', 'hsa_mir_144_3p', 'hsa_mir_29b_3p'}
```

In [176]:
```
print(testB)
```

```
{'hsa_miR_18a_5p', nan, 'hsa_miR_192_5p', 'hsa_miR_142_3p', 'hsa_miR_48
5_3p', 'hsa_miR_106b_5p', 'hsa_miR_26a_5p', 'hsa_miR_181b_5p', 'hsa_miR
_342_3p', 'hsa_miR_30a_5p', 'hsa_miR_328_3p', 'hsa_miR_199a_5p', 'hsa_m
iR_208b_3p', 'hsa_miR_320b', 'hsa_miR_126_3p', 'hsa_miR_26b_5p', 'hsa_m
iR_433_3p', 'hsa_miR_25_3p', 'hsa_miR_1_3p', 'hsa_miR_124_3p', 'hsa_miR
_103a_3p', 'hsa_miR_20b_5p', 'hsa_miR_122_5p', 'hsa_miR_29a_3p', 'hsa_m
iR_337_5p', 'hsa_miR_144_5p', 'hsa_miR_195_5p', 'hsa_miR_125b_5p', 'hsa
_miR_208a_3p', 'hsa_miR_15b_5p', 'hsa_miR_27b_3p', 'hsa_miR_155_5p', 'h
sa_miR_150_5p', 'hsa_miR_451a', 'hsa_miR_423_5p', 'hsa_let_7d_5p', 'hsa
_miR_133b', 'hsa_miR_199a_3p', 'hsa_miR_210_3p', 'hsa_let_7i_5p', 'hsa_
miR_505_5p', 'hsa_miR_93_5p', 'hsa_miR_375', 'hsa_miR_146a_5p', 'hsa_mi
R_145_5p', 'hsa_miR_16_5p', 'hsa_miR_194_5p', 'hsa_miR_214_3p', 'hsa_mi
R_28_5p', 'hsa_miR_590_5p', 'hsa_miR_29b_3p', 'hsa_miR_363_3p', 'hsa_mi
R_34a_5p', 'hsa_miR_499a_5p', 'hsa_miR_370_3p', 'hsa_let_7e_5p', 'hsa_m
iR_21_5p', 'hsa_miR_17_5p', 'hsa_miR_335_5p', 'hsa_miR_92a_3p', 'hsa_mi
R_125a_5p', 'hsa_let_7g_5p', 'hsa_miR_133a_3p', 'hsa_miR_19a_3p', 'hsa_
miR_486_5p', 'hsa_miR_27a_3p'}
```

In [177]:
```
print(testC)
```

```
{'hsa_mir_130b_3p', 'hsa_mir_370_3p', 'hsa_mir_208a_3p', 'hsa_mir_18a_5
p', 'hsa_mir_486_2_3p', 'hsa_mir_326_5p', 'hsa_mir_323a_3p', 'hsa_mir_1
33a_1_3p', 'hsa_mir_23a_3p', 'hsa_mir_186_5p', 'hsa_mir_320b_1_5p', 'hs
a_let_7c_5p', 'hsa_mir_197_3p', 'hsa_mir_151a_5p', 'hsa_let_7f_1_5p',
'hsa_mir_423_5p', 'hsa_mir_1_2_3p', 'hsa_mir_451a_5p', 'hsa_mir_181b_1_
5p', 'hsa_let_7d_5p', 'hsa_mir_194_2_5p', 'hsa_mir_103a_1_3p', 'hsa_mir
_199a_1_3p', 'hsa_let_7f_2_5p', 'hsa_mir_25_3p', 'hsa_let_7e_5p', 'hsa_
mir_379_5p', 'hsa_mir_136_3p', 'hsa_mir_485_3p', 'hsa_mir_17_5p', 'hsa_
let_7g_5p', 'hsa_mir_144_3p', 'hsa_mir_26a_1_5p', 'hsa_mir_342_3p', 'hs
a_mir_26b_5p', 'hsa_mir_590_5p', 'hsa_mir_98_5p', 'hsa_mir_320b_2_3p',
'hsa_mir_133b_3p', 'hsa_mir_197_5p', 'hsa_mir_194_1_5p', 'hsa_mir_221_3
p', 'hsa_mir_155_5p', 'hsa_mir_192_5p', 'hsa_mir_486_1_5p', 'hsa_mir_50
3_5p', 'hsa_mir_16_2_5p', 'hsa_mir_424_5p', 'hsa_mir_532_5p', 'hsa_mir_
375_5p', 'hsa_mir_320b_1_3p', 'hsa_mir_27a_3p', 'hsa_mir_214_3p', 'hsa_
mir_342_5p', 'hsa_mir_326_3p', 'hsa_mir_92a_2_3p', 'hsa_mir_199a_2_3p',
'hsa_let_7i_5p', 'hsa_mir_24_2_3p', 'hsa_mir_126_5p', 'hsa_mir_92a_1_3
p', 'hsa_mir_1_1_3p', 'hsa_mir_1287_5p', 'hsa_mir_103a_2_3p', 'hsa_mir_
29b_1_3p', 'hsa_mir_425_3p', 'hsa_mir_877_5p', 'hsa_mir_26a_2_5p', 'hsa
_mir_874_3p', 'hsa_mir_206_3p', 'hsa_mir_125a_5p', 'hsa_mir_199a_2_5p',
'hsa_mir_195_5p', 'hsa_mir_210_3p', 'hsa_mir_140_3p', 'hsa_mir_34a_5p',
'hsa_mir_19a_3p', 'hsa_mir_146a_5p', 'hsa_mir_505_5p', 'hsa_mir_375_3
p', 'hsa_mir_296_5p', 'hsa_mir_335_5p', 'hsa_mir_29a_3p', 'hsa_mir_205_
5p', 'hsa_mir_486_2_5p', 'hsa_mir_145_5p', 'hsa_mir_125b_2_5p', 'hsa_mi
r_16_1_5p', 'hsa_mir_28_5p', 'hsa_mir_30a_5p', 'hsa_mir_144_5p', 'hsa_m
ir_29b_2_3p', 'hsa_mir_345_5p', 'hsa_mir_451a_3p', 'hsa_mir_124_1_3p',
'hsa_mir_320c_1_3p', 'hsa_mir_27b_3p', 'hsa_mir_125b_1_5p', 'hsa_mir_37
4b_5p', 'hsa_mir_15b_5p', 'hsa_mir_181c_3p', 'hsa_mir_363_3p', 'hsa_mir
_181b_2_5p', 'hsa_mir_106b_5p', 'hsa_mir_328_3p', 'hsa_mir_199a_1_5p',
'hsa_mir_330_3p', 'hsa_mir_337_5p', 'hsa_mir_122_5p', 'hsa_mir_208b_3
p', 'hsa_mir_150_5p', 'hsa_mir_203a_3p', 'hsa_mir_20b_5p', 'hsa_mir_151
b_3p', 'hsa_mir_142_3p', 'hsa_mir_433_3p', 'hsa_mir_21_5p', 'hsa_mir_21
5_5p', 'hsa_mir_93_5p', 'hsa_mir_151a_3p', 'hsa_mir_652_3p', 'hsa_mir_2
4_1_3p', 'hsa_mir_486_1_3p', 'hsa_mir_126_3p', 'hsa_mir_499a_5p', 'hsa_
mir_23b_5p'}
```

In [178]:
```
print(discovery_match)
```

```
['hsa_mir_130b_3p', 'hsa_mir_342_3p', 'hsa_mir_345_5p', 'hsa_mir_98_5
p', 'hsa_mir_197_5p', 'hsa_mir_425_3p', 'hsa_mir_877_5p', 'hsa_mir_221_
3p', 'hsa_mir_323a_3p', 'hsa_mir_874_3p', 'hsa_mir_23a_3p', 'hsa_mir_18
6_5p', 'hsa_mir_374b_5p', 'hsa_let_7c_5p', 'hsa_mir_181c_3p', 'hsa_mir_
363_3p', 'hsa_mir_197_3p', 'hsa_mir_140_3p', 'hsa_mir_192_5p', 'hsa_mir
_151a_5p', 'hsa_mir_106b_5p', 'hsa_mir_503_5p', 'hsa_mir_330_3p', 'hsa_
mir_424_5p', 'hsa_mir_532_5p', 'hsa_mir_296_5p', 'hsa_mir_203a_3p', 'hs
a_mir_20b_5p', 'hsa_mir_27a_3p', 'hsa_mir_342_5p', 'hsa_mir_215_5p', 'h
sa_mir_93_5p', 'hsa_mir_144_3p', 'hsa_mir_205_5p', 'hsa_mir_126_5p', 'h
sa_mir_151a_3p', 'hsa_mir_30a_5p', 'hsa_mir_652_3p', 'hsa_mir_145_5p',
'hsa_mir_379_5p', 'hsa_mir_136_3p', 'hsa_mir_1287_5p', 'hsa_mir_126_3
p', 'hsa_mir_23b_5p']
```

In [179]:
```
print(FPcv_match)
```

```
['hsa_let_7e_5p', 'hsa_let_7i_5p', 'hsa_let_7d_5p', 'hsa_let_7g_5p']
```

In [180]:
```
DPP_val = list(x1.index.values)
DPP_val[0] == discovery_match
```

Out[180]:
```
False
```

In [193]:
```
x1_discovery = pd.DataFrame(x1.loc[discovery_match])
x1_discovery.discovery = 'X'

x1_FPcv = pd.DataFrame(x1.loc[FPcv_match])
x1_FPcv.iloc[:,4] = 'X'
```

In [194]:
```
x1_new = pd.concat([x1_discovery, x1_FPcv])
```

In [198]:
```
DPP = pd.concat([x1, x1_new])
```

In [207]:
```
DPP.tail(60)
```

Out[207]:

	Median	Mean	SD	discovery	FP CVD
hsa_mir_215_5p	5.044394	4.902801	1.409622	NaN	NaN
hsa_mir_133b_3p	0.000000	0.917203	1.412628	NaN	NaN
hsa_mir_1287_5p	4.087463	3.757894	1.429397	NaN	NaN
hsa_mir_16_2_5p	4.857981	4.590077	1.465999	NaN	NaN
hsa_mir_208b_3p	1.000000	1.367396	1.552496	NaN	NaN
hsa_mir_451a_3p	3.753897	3.490164	1.567592	NaN	NaN
hsa_mir_29b_2_3p	2.000000	1.894336	1.592837	NaN	NaN
hsa_mir_337_5p	3.807355	3.441441	1.608527	NaN	NaN
hsa_mir_206_3p	4.614409	4.344164	1.692182	NaN	NaN
hsa_mir_208a_3p	0.000000	1.335618	1.732603	NaN	NaN
hsa_mir_124_1_3p	3.459432	3.118031	1.798116	NaN	NaN
hsa_mir_379_5p	4.754888	4.457206	1.802512	NaN	NaN
hsa_mir_130b_3p	9.691742	9.683894	0.848375	X	NaN
hsa_mir_342_3p	10.721042	10.678653	0.927287	X	NaN
hsa_mir_345_5p	8.301480	8.225339	0.802721	X	NaN
hsa_mir_98_5p	8.834463	8.848783	0.568305	X	NaN
hsa_mir_197_5p	0.000000	0.296448	0.780443	X	NaN
hsa_mir_425_3p	6.672425	6.633236	0.570569	X	NaN
hsa_mir_877_5p	5.169925	4.958120	0.966119	X	NaN
hsa_mir_221_3p	13.909843	13.875052	0.510014	X	NaN
hsa_mir_323a_3p	7.071373	7.047099	1.015618	X	NaN
hsa_mir_874_3p	7.672408	7.676221	0.575017	X	NaN
hsa_mir_23a_3p	15.239448	15.215018	0.415213	X	NaN
hsa_mir_186_5p	11.908955	11.914718	0.383638	X	NaN
hsa_mir_374b_5p	9.645657	9.563443	1.146249	X	NaN
hsa_let_7c_5p	7.758219	7.773823	0.523720	X	NaN
hsa_mir_181c_3p	5.044394	4.906109	1.056421	X	NaN
hsa_mir_363_3p	9.839991	9.850411	0.989722	X	NaN
hsa_mir_197_3p	9.082147	9.111763	0.469464	X	NaN
hsa_mir_140_3p	12.417189	12.423036	0.418691	X	NaN
hsa_mir_192_5p	8.859527	8.903272	0.730156	X	NaN
hsa_mir_151a_5p	11.875747	11.882875	0.478920	X	NaN
hsa_mir_106b_5p	11.994740	12.017610	0.711228	X	NaN
hsa_mir_503_5p	6.658211	6.503219	1.090228	X	NaN

	Median	Mean	SD	discovery	FP CVD
hsa_mir_330_3p	8.475711	8.365700	0.888420	X	NaN
hsa_mir_424_5p	8.326422	8.286710	0.919006	X	NaN
hsa_mir_532_5p	9.538189	9.530813	0.381518	X	NaN
hsa_mir_296_5p	5.339740	5.248171	1.036744	X	NaN
hsa_mir_203a_3p	7.535144	7.472193	1.369227	X	NaN
hsa_mir_20b_5p	8.126684	8.112696	1.076793	X	NaN
hsa_mir_27a_3p	12.183460	12.121560	0.546866	X	NaN
hsa_mir_342_5p	6.087463	6.005990	0.951496	X	NaN
hsa_mir_215_5p	5.044394	4.902801	1.409622	X	NaN
hsa_mir_93_5p	13.523434	13.503092	0.507903	X	NaN
hsa_mir_144_3p	10.425740	10.449526	1.030141	X	NaN
hsa_mir_205_5p	8.614480	8.567617	1.039363	X	NaN
hsa_mir_126_5p	13.049422	12.942758	0.715203	X	NaN
hsa_mir_151a_3p	11.391774	11.381391	0.527569	X	NaN
hsa_mir_30a_5p	11.968654	11.955174	0.580959	X	NaN
hsa_mir_652_3p	9.834457	9.838846	0.331363	X	NaN
hsa_mir_145_5p	11.950102	11.975558	0.530356	X	NaN
hsa_mir_379_5p	4.754888	4.457206	1.802512	X	NaN
hsa_mir_136_3p	7.535228	7.464690	1.148878	X	NaN
hsa_mir_1287_5p	4.087463	3.757894	1.429397	X	NaN
hsa_mir_126_3p	13.890117	13.898219	0.416129	X	NaN
hsa_mir_23b_5p	5.022197	4.846298	1.113720	X	NaN
hsa_let_7e_5p	6.781360	6.762497	0.897283	NaN	X
hsa_let_7i_5p	11.887982	11.908423	0.415571	NaN	X
hsa_let_7d_5p	10.576010	10.623229	0.541547	NaN	X
hsa_let_7g_5p	12.531379	12.532410	0.481866	NaN	X

In [208]:
```
### Comparisons
```

In [215]:
```
discovery = x1_discovery.iloc[:,:3]
fpcv = x1_FPcv.iloc[:,:3]
```

In [222]:
```
fpcv.sort_values(by = 'SD', ascending = False)
```

Out[222]:

	Median	Mean	SD
hsa_let_7e_5p	6.781360	6.762497	0.897283
hsa_let_7d_5p	10.576010	10.623229	0.541547
hsa_let_7g_5p	12.531379	12.532410	0.481866
hsa_let_7i_5p	11.887982	11.908423	0.415571

In [223]:
```
discovery.sort_values(by = 'SD', ascending = False)
```

Out[223]:

	Median	Mean	SD
hsa_mir_379_5p	4.754888	4.457206	1.802512
hsa_mir_1287_5p	4.087463	3.757894	1.429397
hsa_mir_215_5p	5.044394	4.902801	1.409622
hsa_mir_203a_3p	7.535144	7.472193	1.369227
hsa_mir_136_3p	7.535228	7.464690	1.148878
hsa_mir_374b_5p	9.645657	9.563443	1.146249
hsa_mir_23b_5p	5.022197	4.846298	1.113720
hsa_mir_503_5p	6.658211	6.503219	1.090228
hsa_mir_20b_5p	8.126684	8.112696	1.076793
hsa_mir_181c_3p	5.044394	4.906109	1.056421
hsa_mir_205_5p	8.614480	8.567617	1.039363
hsa_mir_296_5p	5.339740	5.248171	1.036744
hsa_mir_144_3p	10.425740	10.449526	1.030141
hsa_mir_323a_3p	7.071373	7.047099	1.015618
hsa_mir_363_3p	9.839991	9.850411	0.989722
hsa_mir_877_5p	5.169925	4.958120	0.966119
hsa_mir_342_5p	6.087463	6.005990	0.951496
hsa_mir_342_3p	10.721042	10.678653	0.927287
hsa_mir_424_5p	8.326422	8.286710	0.919006
hsa_mir_330_3p	8.475711	8.365700	0.888420
hsa_mir_130b_3p	9.691742	9.683894	0.848375
hsa_mir_345_5p	8.301480	8.225339	0.802721
hsa_mir_197_5p	0.000000	0.296448	0.780443
hsa_mir_192_5p	8.859527	8.903272	0.730156
hsa_mir_126_5p	13.049422	12.942758	0.715203
hsa_mir_106b_5p	11.994740	12.017610	0.711228
hsa_mir_30a_5p	11.968654	11.955174	0.580959
hsa_mir_874_3p	7.672408	7.676221	0.575017
hsa_mir_425_3p	6.672425	6.633236	0.570569
hsa_mir_98_5p	8.834463	8.848783	0.568305
hsa_mir_27a_3p	12.183460	12.121560	0.546866
hsa_mir_145_5p	11.950102	11.975558	0.530356
hsa_mir_151a_3p	11.391774	11.381391	0.527569
hsa_let_7c_5p	7.758219	7.773823	0.523720

	Median	Mean	SD
hsa_mir_221_3p	13.909843	13.875052	0.510014
hsa_mir_93_5p	13.523434	13.503092	0.507903
hsa_mir_151a_5p	11.875747	11.882875	0.478920
hsa_mir_197_3p	9.082147	9.111763	0.469464
hsa_mir_140_3p	12.417189	12.423036	0.418691
hsa_mir_126_3p	13.890117	13.898219	0.416129
hsa_mir_23a_3p	15.239448	15.215018	0.415213
hsa_mir_186_5p	11.908955	11.914718	0.383638
hsa_mir_532_5p	9.538189	9.530813	0.381518
hsa_mir_652_3p	9.834457	9.838846	0.331363

In [ ]: