

20210326_Feature_Analysis_Number_Nuclei

March 29, 2021

1 Feature analysis by number of nuclei

We load the neccesary libraries

```
[1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

We load the dataset with all the features for all the proteins and add some utility columns.

```
[2]: df = pd.read_csv('../data/processed/2021-03-26_features.csv', index_col=0)
df.insert(0, 'protein', df.index.str.split("_").str[0])
df.insert(1, 'plate', df.index.str.split("_").str[1])
df.insert(2, 'stimulus', df.index.str.split("_").str[2])
df
```

```
[2]:      protein plate stimulus white_area_ratio area_mean \
NegsiRNA_C3c_CA NegsiRNA C3c CA 0.153746 16637.950000
NegsiRNA_C3b_CA NegsiRNA C3b CA 0.304616 7925.681159
NegsiRNA_E3_NO NegsiRNA E3 NO 0.297828 9152.662983
AKT1_B2_NO AKT1 B2 NO 0.330304 9813.732919
NegsiRNA_L2_NS NegsiRNA L2 NS 0.188984 16354.025641
...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
... ...
```

NegsiRNA_H3_NS	6005.232074	8918.0	3276.0	13.820058	183
NegsiRNA_E3_CA	6795.547574	9237.0	4849.0	13.660881	158
CTNNA1_A3_NO	15246.811242	8115.0	7508.0	12.918557	178
NUMB_E3_CA	5398.034067	8229.0	3936.0	13.871534	1946
ENAH_B3_NO	7062.259261	8996.0	4996.0	13.719685	45
... texture_std_filter_mode texture_entropy_filter_mean \					
NegsiRNA_C3c_CA	...	516.588501		3.294059	
NegsiRNA_C3b_CA	...	835.012390		3.826592	
NegsiRNA_E3_NO	...	696.241333		3.656484	
AKT1_B2_NO	...	1188.369629		4.154108	
NegsiRNA_L2_NS	...	625.615723		3.548006	
...	
NegsiRNA_H3_NS	...	803.342468		3.914721	
NegsiRNA_E3_CA	...	652.602112		3.663541	
CTNNA1_A3_NO	...	525.445984		3.328485	
NUMB_E3_CA	...	576.452637		3.543888	
ENAH_B3_NO	...	1232.572144		4.054212	
texture_entropy_filter_std texture_entropy_filter_median \					
NegsiRNA_C3c_CA		0.607975		3.287055	
NegsiRNA_C3b_CA		0.547050		3.841366	
NegsiRNA_E3_NO		0.564080		3.654402	
AKT1_B2_NO		0.523035		4.191269	
NegsiRNA_L2_NS		0.576589		3.554798	
...	
NegsiRNA_H3_NS		0.542904		3.918922	
NegsiRNA_E3_CA		0.588857		3.677247	
CTNNA1_A3_NO		0.597775		3.338016	
NUMB_E3_CA		0.565170		3.543282	
ENAH_B3_NO		0.530159		4.078297	
texture_entropy_filter_mad texture_entropy_filter_entropy \					
NegsiRNA_C3c_CA		0.430039		1.327159	
NegsiRNA_C3b_CA		0.387300		1.170749	
NegsiRNA_E3_NO		0.401522		1.216988	
AKT1_B2_NO		0.372221		1.086672	
NegsiRNA_L2_NS		0.407309		1.250064	
...	
NegsiRNA_H3_NS		0.389364		1.157889	
NegsiRNA_E3_CA		0.415929		1.278160	
CTNNA1_A3_NO		0.426792		1.301539	
NUMB_E3_CA		0.403005		1.220069	
ENAH_B3_NO		0.371750		1.118129	
texture_entropy_filter_mode global_entropy_discrete \					
NegsiRNA_C3c_CA		3.282318		4.790249	

NegsiRNA_C3b_CA		3.869591	5.144402
NegsiRNA_E3_NO		3.590049	4.918801
AKT1_B2_NO		4.354190	5.791014
NegsiRNA_L2_NS		3.542132	4.889136
...
NegsiRNA_H3_NS		4.026793	5.463957
NegsiRNA_E3_CA		3.870739	4.984301
CTNNA1_A3_NO		3.273900	4.569930
NUMB_E3_CA		3.634047	4.776272
ENAH_B3_NO		4.248230	5.609815

	global_entropy_kde	global_coeff_var
NegsiRNA_C3c_CA	4.802271	0.475106
NegsiRNA_C3b_CA	5.154480	0.500748
NegsiRNA_E3_NO	4.929815	0.507799
AKT1_B2_NO	5.803408	0.540478
NegsiRNA_L2_NS	4.899215	0.469468
...
NegsiRNA_H3_NS	5.476650	0.559494
NegsiRNA_E3_CA	4.994752	0.500610
CTNNA1_A3_NO	4.580187	0.455083
NUMB_E3_CA	4.785130	0.453597
ENAH_B3_NO	5.623074	0.575680

[282 rows x 279 columns]

The normalization must be done using the NegsiRNA samples with no stimuli as reference, so we filter the data in order to work with just that samples.

```
[3]: df_negsi_ns = df[(df['protein'] == 'NegsiRNA') & (df['stimulus'] == 'NS')]
df_negsi_ns
```

```
[3]:   protein plate stimulus white_area_ratio area_mean \
NegsiRNA_L2_NS NegsiRNA L2    NS      0.188984 16354.025641
NegsiRNA_G3_NS NegsiRNA G3    NS      0.239506 12123.135135
NegsiRNA_F3_NS NegsiRNA F3    NS      0.281612 12648.402985
NegsiRNA_D1_NS NegsiRNA D1    NS      0.250790 8580.611650
NegsiRNA_G1_NS NegsiRNA G1    NS      0.330879 6605.234310
NegsiRNA_C2_NS NegsiRNA C2    NS      0.307114 8603.752632
NegsiRNA_F2_NS NegsiRNA F2    NS      0.306993 12201.522388
NegsiRNA_A2_NS NegsiRNA A2    NS      0.225510 11074.236364
NegsiRNA_A2b_NS NegsiRNA A2b   NS      0.241458 10652.517857
NegsiRNA_E3_NS NegsiRNA E3    NS      0.211370 11009.532544
NegsiRNA_H2_NS NegsiRNA H2    NS      0.255573 9811.854749
NegsiRNA_E1_NS NegsiRNA E1    NS      0.211719 18056.135922
NegsiRNA_D3_NS NegsiRNA D3    NS      0.294313 12152.729927
NegsiRNA_I3_NS NegsiRNA I3    NS      0.260990 11035.069620
NegsiRNA_B3_NS NegsiRNA B3    NS      0.262637 8053.944444
```

NegsiRNA_B1_NS	NegsiRNA	B1	NS	0.243277	9202.711340
NegsiRNA_E2_NS	NegsiRNA	E2	NS	0.284126	12795.136364
NegsiRNA_I2_NS	NegsiRNA	I2	NS	0.246134	9312.005236
NegsiRNA_I1_NS	NegsiRNA	I1	NS	0.299052	8106.588235
NegsiRNA_D2_NS	NegsiRNA	D2	NS	0.263888	9188.904762
NegsiRNA_L3_NS	NegsiRNA	L3	NS	0.226791	11545.683544
NegsiRNA_G2_NS	NegsiRNA	G2	NS	0.309893	11228.724138
NegsiRNA_A1b_NS	NegsiRNA	A1b	NS	0.212412	12903.812500
NegsiRNA_F1_NS	NegsiRNA	F1	NS	0.291162	7888.471698
NegsiRNA_C3c_NS	NegsiRNA	C3c	NS	0.166232	18213.842593
NegsiRNA_H1_NS	NegsiRNA	H1	NS	0.293823	8127.180488
NegsiRNA_L1_NS	NegsiRNA	L1	NS	0.251836	6128.954861
NegsiRNA_B2_NS	NegsiRNA	B2	NS	0.320149	10843.964029
NegsiRNA_A1_NS	NegsiRNA	A1	NS	0.217278	14832.384615
NegsiRNA_C1_NS	NegsiRNA	C1	NS	0.267462	9048.539267
NegsiRNA_C3_NS	NegsiRNA	C3	NS	0.266316	9891.285714
NegsiRNA_C3b_NS	NegsiRNA	C3b	NS	0.269283	10201.005917
NegsiRNA_A3_NS	NegsiRNA	A3	NS	0.273287	7620.057778
NegsiRNA_H3_NS	NegsiRNA	H3	NS	0.287646	9603.737143

	area_std	area_median	area_mad	area_entropy	area_mode	\
NegsiRNA_L2_NS	11973.890243	14692.0	7801.0	14.286719	133	
NegsiRNA_G3_NS	6724.839819	11964.0	4811.0	14.120058	25070	
NegsiRNA_F3_NS	8348.251731	12314.5	5483.0	14.213718	4	
NegsiRNA_D1_NS	6151.951701	7528.5	4195.0	13.757380	4	
NegsiRNA_G1_NS	3941.027072	6294.0	3071.0	13.391345	5024	
NegsiRNA_C2_NS	6699.223730	7642.5	4340.0	13.586796	299	
NegsiRNA_F2_NS	7913.931587	11633.5	6091.0	13.923187	16108	
NegsiRNA_A2_NS	7234.436975	10136.0	4622.0	14.074884	5514	
NegsiRNA_A2b_NS	6125.591787	10230.0	3775.5	14.063475	69	
NegsiRNA_E3_NS	7945.905466	10004.0	5388.0	14.060565	82	
NegsiRNA_H2_NS	6847.282495	8618.0	4570.0	13.715774	7278	
NegsiRNA_E1_NS	15471.468437	16188.0	12616.0	13.809054	95	
NegsiRNA_D3_NS	8110.248734	11876.0	5961.0	14.076585	336	
NegsiRNA_I3_NS	6644.736169	10831.5	4897.5	13.984469	8356	
NegsiRNA_B3_NS	5300.844179	7139.5	3632.0	13.750485	6754	
NegsiRNA_B1_NS	5206.755148	9120.5	3713.0	13.858280	11489	
NegsiRNA_E2_NS	7040.560225	13526.5	5126.0	14.150015	69	
NegsiRNA_I2_NS	7931.011754	7919.0	6029.0	13.280509	8763	
NegsiRNA_I1_NS	5002.528808	7219.5	3163.5	13.848589	6283	
NegsiRNA_D2_NS	6068.328707	8424.0	3847.0	13.839825	14391	
NegsiRNA_L3_NS	14132.782812	6152.0	6013.5	12.345705	78	
NegsiRNA_G2_NS	8042.836038	10511.0	4622.0	13.982571	282	
NegsiRNA_A1b_NS	9260.934805	12086.0	6106.5	14.151729	1612	
NegsiRNA_F1_NS	4766.367290	7382.0	3231.0	13.790822	5116	
NegsiRNA_C3c_NS	13401.789463	15444.5	7711.5	14.550112	71	
NegsiRNA_H1_NS	5135.299471	7600.0	3569.0	13.735204	8711	

NegsiRNA_L1_NS	3980.164240	5596.0	2629.5	13.319315	430
NegsiRNA_B2_NS	6713.171666	11086.0	5034.0	13.846587	258
NegsiRNA_A1_NS	9837.703766	13374.0	6761.0	14.315463	205
NegsiRNA_C1_NS	4743.457069	8945.0	3072.0	13.857769	9194
NegsiRNA_C3_NS	6490.588598	8862.0	3723.0	13.948571	15978
NegsiRNA_C3b_NS	6332.172848	10094.0	4581.0	13.996158	6
NegsiRNA_A3_NS	6916.672056	6364.0	5063.0	12.978098	102
NegsiRNA_H3_NS	6005.232074	8918.0	3276.0	13.820058	183

	... texture_std_filter_mode	texture_entropy_filter_mean	\
NegsiRNA_L2_NS	... 625.615723	3.548006	
NegsiRNA_G3_NS	... 718.552246	3.727769	
NegsiRNA_F3_NS	... 722.297607	3.767866	
NegsiRNA_D1_NS	... 545.209656	3.337365	
NegsiRNA_G1_NS	... 625.229614	3.641625	
NegsiRNA_C2_NS	... 662.526245	3.540871	
NegsiRNA_F2_NS	... 773.727905	3.818430	
NegsiRNA_A2_NS	... 830.302002	3.853664	
NegsiRNA_A2b_NS	... 952.687744	3.920384	
NegsiRNA_E3_NS	... 627.693970	3.643499	
NegsiRNA_H2_NS	... 926.946533	3.840170	
NegsiRNA_E1_NS	... 856.864685	3.805060	
NegsiRNA_D3_NS	... 864.600952	3.995174	
NegsiRNA_I3_NS	... 521.851685	3.500026	
NegsiRNA_B3_NS	... 814.747070	3.893659	
NegsiRNA_B1_NS	... 543.660339	3.427511	
NegsiRNA_E2_NS	... 897.456848	3.982647	
NegsiRNA_I2_NS	... 1149.334717	4.140892	
NegsiRNA_I1_NS	... 828.977722	3.897173	
NegsiRNA_D2_NS	... 856.769043	3.834476	
NegsiRNA_L3_NS	... 695.771912	3.528654	
NegsiRNA_G2_NS	... 811.739990	3.874558	
NegsiRNA_A1b_NS	... 491.947327	3.486550	
NegsiRNA_F1_NS	... 747.367981	3.851476	
NegsiRNA_C3c_NS	... 435.136047	3.312429	
NegsiRNA_H1_NS	... 1150.170654	4.030219	
NegsiRNA_L1_NS	... 428.135498	3.152497	
NegsiRNA_B2_NS	... 818.546753	3.898607	
NegsiRNA_A1_NS	... 1128.855469	3.961820	
NegsiRNA_C1_NS	... 800.925659	3.842814	
NegsiRNA_C3_NS	... 774.164246	3.745677	
NegsiRNA_C3b_NS	... 703.533936	3.702349	
NegsiRNA_A3_NS	... 741.770386	3.576088	
NegsiRNA_H3_NS	... 803.342468	3.914721	
	texture_entropy_filter_std	texture_entropy_filter_median	\
NegsiRNA_L2_NS	0.576589	3.554798	

NegsiRNA_G3_NS	0.585909	3.736361
NegsiRNA_F3_NS	0.610388	3.771776
NegsiRNA_D1_NS	0.603501	3.332474
NegsiRNA_G1_NS	0.555086	3.626953
NegsiRNA_C2_NS	0.559543	3.527349
NegsiRNA_F2_NS	0.575995	3.829004
NegsiRNA_A2_NS	0.544865	3.856772
NegsiRNA_A2b_NS	0.549168	3.920960
NegsiRNA_E3_NS	0.581148	3.649028
NegsiRNA_H2_NS	0.531972	3.841366
NegsiRNA_E1_NS	0.620179	3.843931
NegsiRNA_D3_NS	0.561773	4.017999
NegsiRNA_I3_NS	0.576899	3.484495
NegsiRNA_B3_NS	0.600451	3.905555
NegsiRNA_B1_NS	0.594228	3.425236
NegsiRNA_E2_NS	0.574562	4.001739
NegsiRNA_I2_NS	0.516526	4.167897
NegsiRNA_I1_NS	0.553714	3.910956
NegsiRNA_D2_NS	0.563589	3.854734
NegsiRNA_L3_NS	0.555682	3.545647
NegsiRNA_G2_NS	0.564003	3.872705
NegsiRNA_A1b_NS	0.680161	3.464541
NegsiRNA_F1_NS	0.568455	3.868102
NegsiRNA_C3c_NS	0.631270	3.278902
NegsiRNA_H1_NS	0.509701	4.045447
NegsiRNA_L1_NS	0.565519	3.121224
NegsiRNA_B2_NS	0.568767	3.918922
NegsiRNA_A1_NS	0.599597	3.994627
NegsiRNA_C1_NS	0.590561	3.851371
NegsiRNA_C3_NS	0.536341	3.750256
NegsiRNA_C3b_NS	0.550546	3.714841
NegsiRNA_A3_NS	0.567480	3.582391
NegsiRNA_H3_NS	0.542904	3.918922

	texture_entropy_filter_mad	texture_entropy_filter_entropy \
NegsiRNA_L2_NS	0.407309	1.250064
NegsiRNA_G3_NS	0.411526	1.272247
NegsiRNA_F3_NS	0.442338	1.323828
NegsiRNA_D1_NS	0.424929	1.316325
NegsiRNA_G1_NS	0.403233	1.187092
NegsiRNA_C2_NS	0.392757	1.203993
NegsiRNA_F2_NS	0.418175	1.241040
NegsiRNA_A2_NS	0.385118	1.167518
NegsiRNA_A2b_NS	0.384791	1.177964
NegsiRNA_E3_NS	0.411083	1.261344
NegsiRNA_H2_NS	0.372277	1.134429
NegsiRNA_E1_NS	0.431008	1.341901

NegsiRNA_D3_NS	0.407636	1.198382
NegsiRNA_I3_NS	0.418966	1.245573
NegsiRNA_B3_NS	0.423569	1.299940
NegsiRNA_B1_NS	0.408690	1.296574
NegsiRNA_E2_NS	0.416457	1.229191
NegsiRNA_I2_NS	0.359380	1.079359
NegsiRNA_I1_NS	0.390719	1.184749
NegsiRNA_D2_NS	0.402235	1.211260
NegsiRNA_L3_NS	0.375640	1.195292
NegsiRNA_G2_NS	0.409674	1.208634
NegsiRNA_A1b_NS	0.505528	1.472987
NegsiRNA_F1_NS	0.409815	1.220676
NegsiRNA_C3c_NS	0.467934	1.359910
NegsiRNA_H1_NS	0.361890	1.065560
NegsiRNA_L1_NS	0.396549	1.215180
NegsiRNA_B2_NS	0.409488	1.220697
NegsiRNA_A1_NS	0.415603	1.293800
NegsiRNA_C1_NS	0.421004	1.277670
NegsiRNA_C3_NS	0.380570	1.144272
NegsiRNA_C3b_NS	0.392230	1.180999
NegsiRNA_A3_NS	0.391521	1.229254
NegsiRNA_H3_NS	0.389364	1.157889

	texture_entropy_filter_mode	global_entropy_discrete \
NegsiRNA_L2_NS	3.542132	4.889136
NegsiRNA_G3_NS	3.773858	5.055952
NegsiRNA_F3_NS	3.700548	5.271409
NegsiRNA_D1_NS	3.198723	4.685021
NegsiRNA_G1_NS	3.619654	5.020604
NegsiRNA_C2_NS	3.552567	4.857666
NegsiRNA_F2_NS	4.036623	5.236222
NegsiRNA_A2_NS	3.929126	5.330096
NegsiRNA_A2b_NS	3.855077	5.515296
NegsiRNA_E3_NS	3.641267	4.910865
NegsiRNA_H2_NS	3.945087	5.264074
NegsiRNA_E1_NS	4.024760	5.231173
NegsiRNA_D3_NS	4.149089	5.526948
NegsiRNA_I3_NS	3.394771	4.787890
NegsiRNA_B3_NS	3.952438	5.498966
NegsiRNA_B1_NS	3.462031	4.947426
NegsiRNA_E2_NS	4.026258	5.520210
NegsiRNA_I2_NS	4.206283	5.760896
NegsiRNA_I1_NS	3.871614	5.459870
NegsiRNA_D2_NS	4.093087	5.256749
NegsiRNA_L3_NS	3.607759	4.731224
NegsiRNA_G2_NS	3.855763	5.398465
NegsiRNA_A1b_NS	3.284930	5.133406

NegsiRNA_F1_NS	4.036214	5.296394
NegsiRNA_C3c_NS	3.052355	4.776217
NegsiRNA_H1_NS	4.231466	5.611062
NegsiRNA_L1_NS	2.950857	4.300187
NegsiRNA_B2_NS	4.148959	5.384783
NegsiRNA_A1_NS	4.188291	5.606467
NegsiRNA_C1_NS	3.837719	5.447795
NegsiRNA_C3_NS	3.790999	5.018499
NegsiRNA_C3b_NS	3.648828	5.043743
NegsiRNA_A3_NS	3.508055	4.825808
NegsiRNA_H3_NS	4.026793	5.463957

	global_entropy_kde	global_coeff_var
NegsiRNA_L2_NS	4.899215	0.469468
NegsiRNA_G3_NS	5.069138	0.544023
NegsiRNA_F3_NS	5.285017	0.549777
NegsiRNA_D1_NS	4.695801	0.484134
NegsiRNA_G1_NS	5.029001	0.455166
NegsiRNA_C2_NS	4.869074	0.493571
NegsiRNA_F2_NS	5.247883	0.517509
NegsiRNA_A2_NS	5.342982	0.529134
NegsiRNA_A2b_NS	5.530317	0.584109
NegsiRNA_E3_NS	4.925221	0.510195
NegsiRNA_H2_NS	5.276570	0.512042
NegsiRNA_E1_NS	5.246148	0.580449
NegsiRNA_D3_NS	5.539703	0.555338
NegsiRNA_I3_NS	4.797593	0.460566
NegsiRNA_B3_NS	5.514775	0.602573
NegsiRNA_B1_NS	4.959982	0.534868
NegsiRNA_E2_NS	5.534404	0.564693
NegsiRNA_I2_NS	5.773547	0.552322
NegsiRNA_I1_NS	5.469479	0.510708
NegsiRNA_D2_NS	5.269466	0.536048
NegsiRNA_L3_NS	4.739318	0.454978
NegsiRNA_G2_NS	5.410961	0.535431
NegsiRNA_A1b_NS	5.145887	0.560308
NegsiRNA_F1_NS	5.306806	0.495210
NegsiRNA_C3c_NS	4.786464	0.492888
NegsiRNA_H1_NS	5.622157	0.510829
NegsiRNA_L1_NS	4.304584	0.407190
NegsiRNA_B2_NS	5.396349	0.515614
NegsiRNA_A1_NS	5.622126	0.575093
NegsiRNA_C1_NS	5.462683	0.574510
NegsiRNA_C3_NS	5.029440	0.482483
NegsiRNA_C3b_NS	5.054879	0.492560
NegsiRNA_A3_NS	4.836290	0.486360
NegsiRNA_H3_NS	5.476650	0.559494

```
[34 rows x 279 columns]
```

For each feature in the dataset, we plot the feature value against the number of nuclei in the corresponding NegsiRNA NS image, and also fit a linear regression model to the samples (they're a lot of features)

```
[4]: df_negsi_ns = df_negsi_ns.sort_values(by='nuclei_n')

data = df_negsi_ns.iloc[:, 3: ].drop(columns='nuclei_n')
has_nan = (data.isna().sum() > 0)
data = data[has_nan[~has_nan].index]
x = np.expand_dims(np.array(df_negsi_ns['nuclei_n']), axis=1)

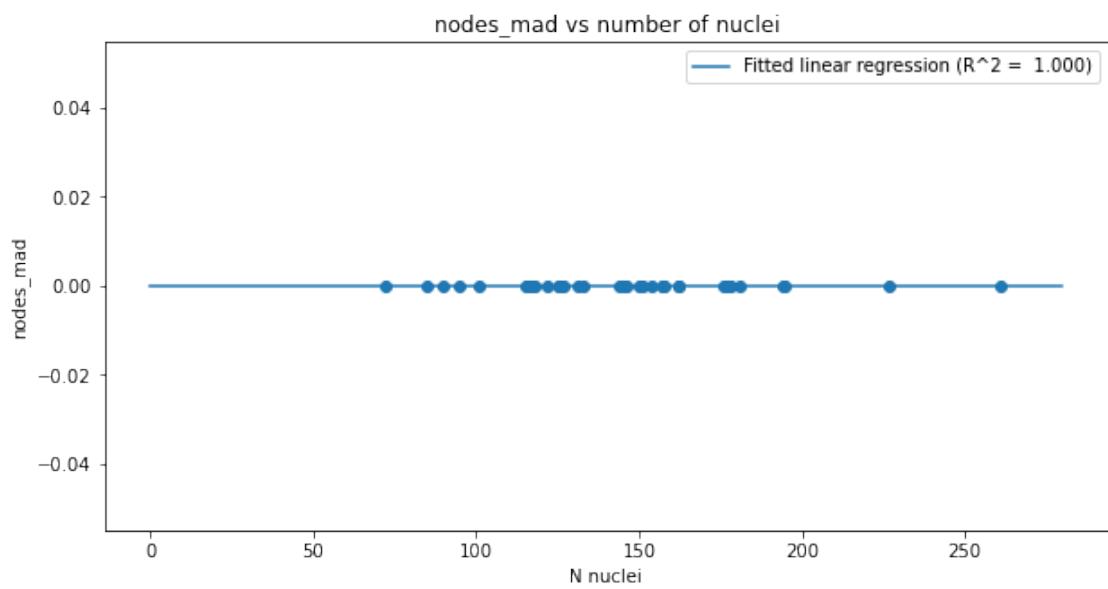
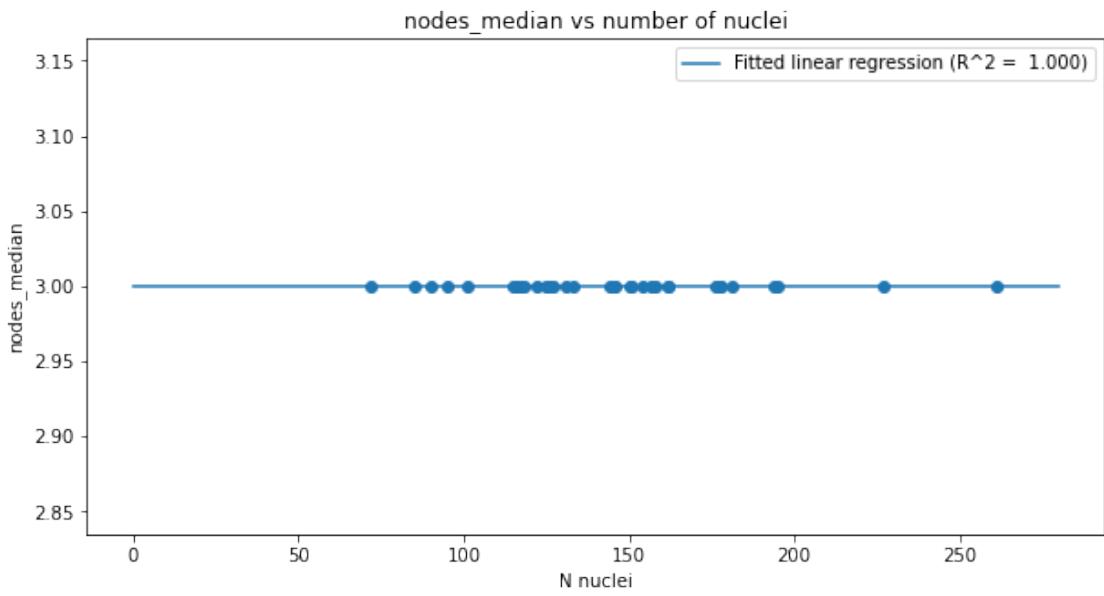
regrs = {}

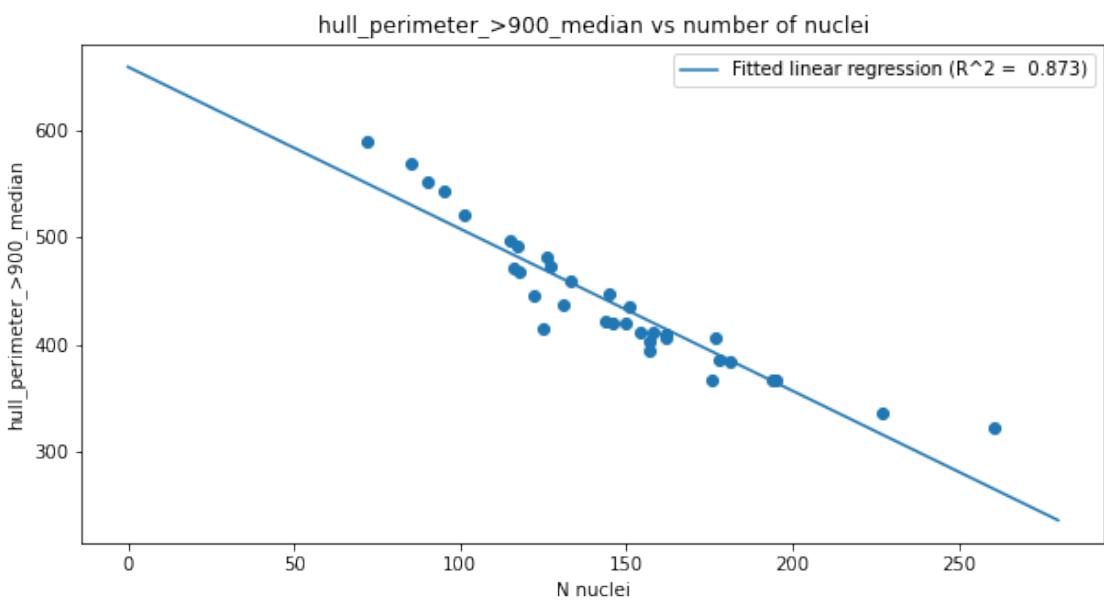
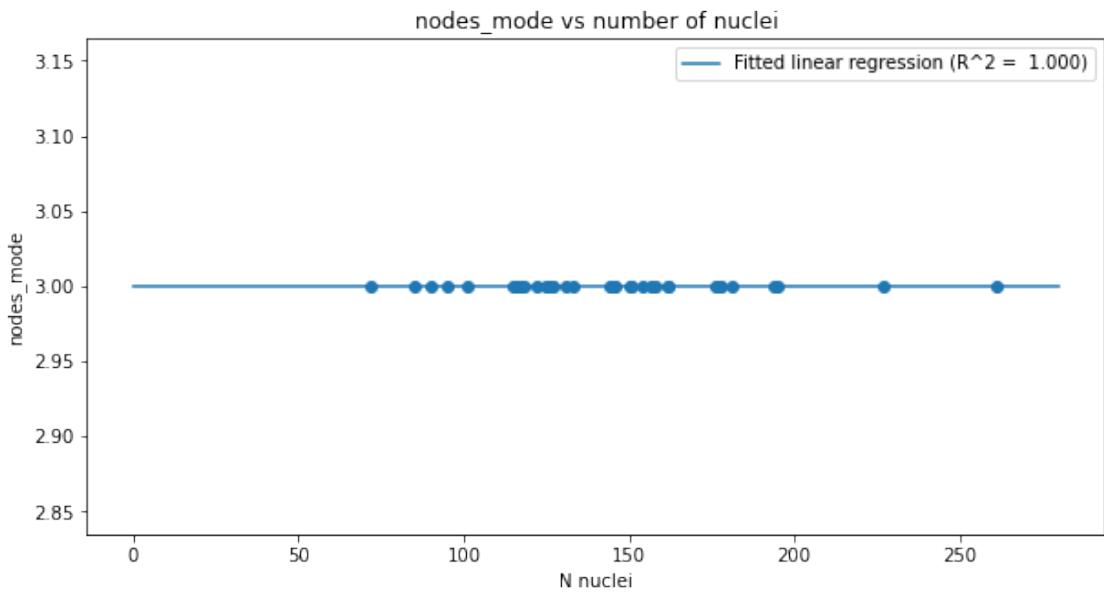
for column in data.columns:
    y = data[column]
    regr = LinearRegression()
    regr.fit(x, y)
    y_predicted = regr.predict(x)
    r2 = r2_score(y, y_predicted)
    regrs[column] = [r2, regr]

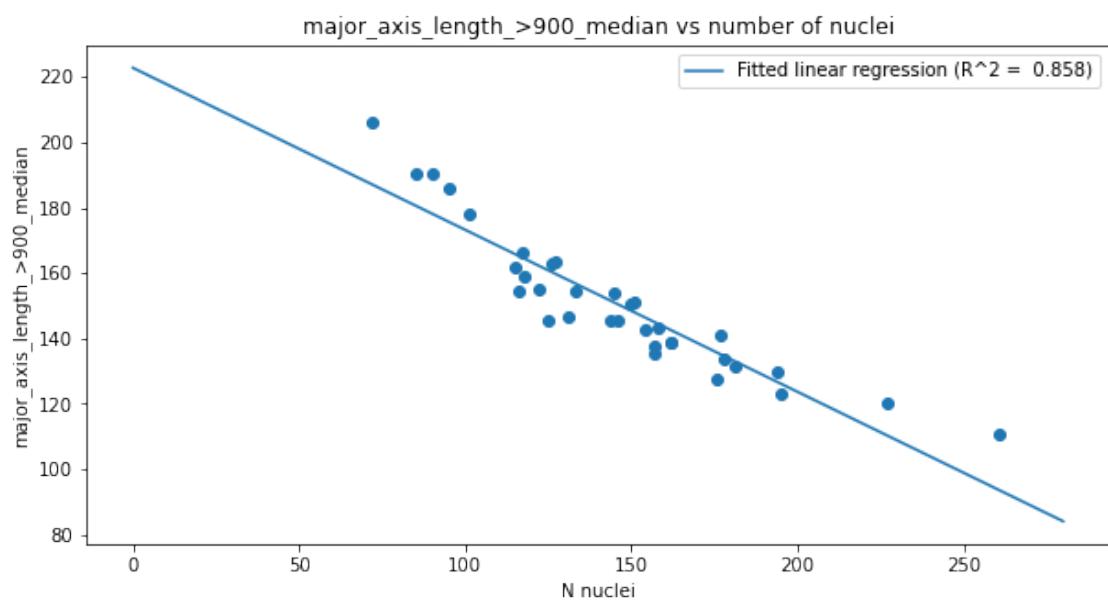
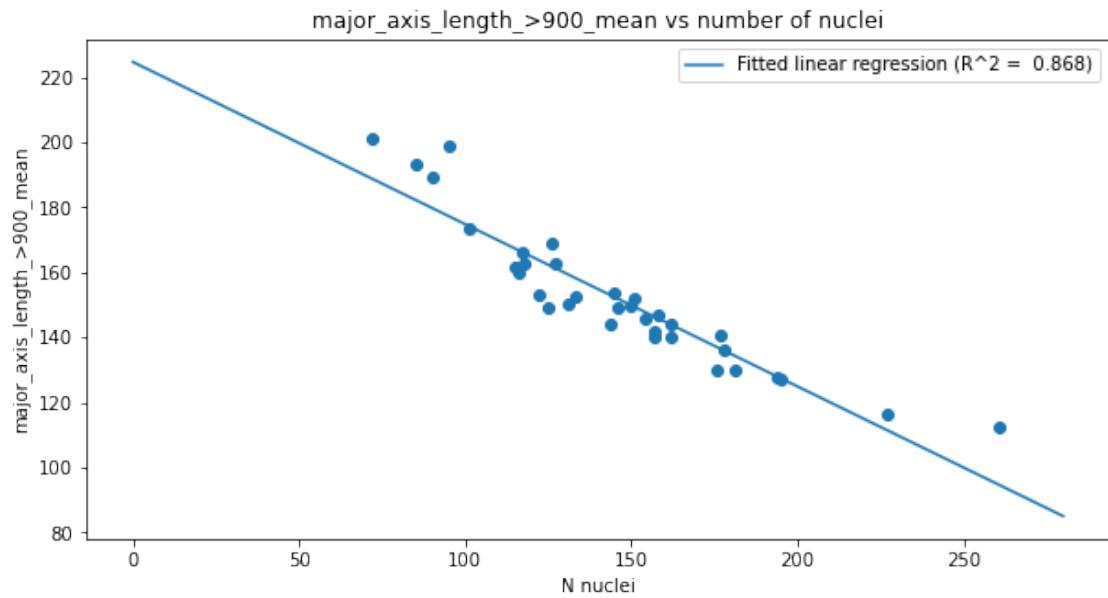
regrs = {k: v for k, v in sorted(regrs.items(), reverse=True, key=lambda item: item[1][0])}
```

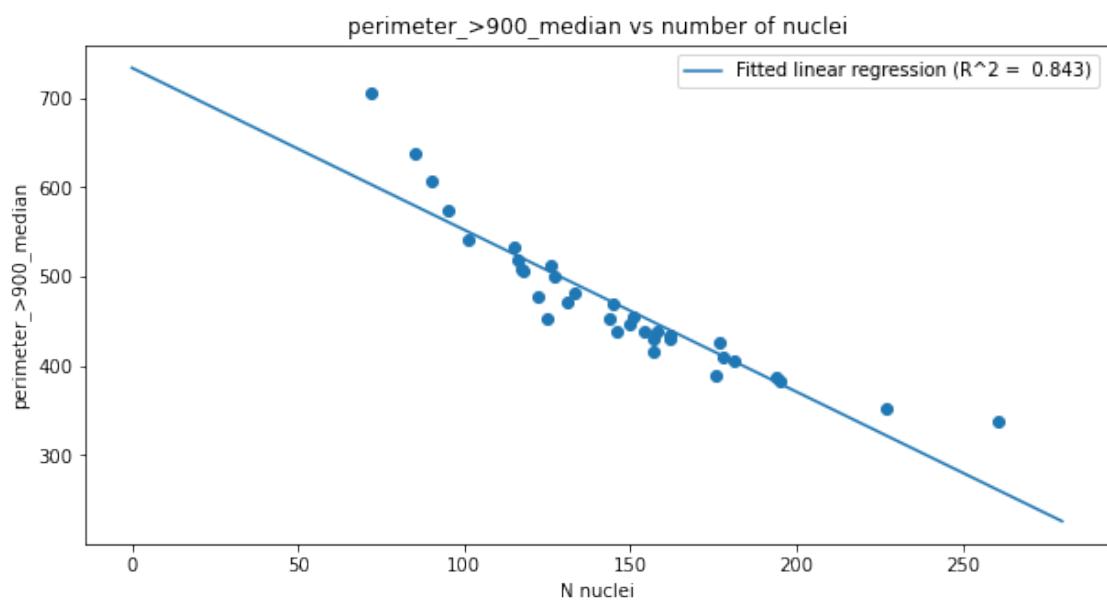
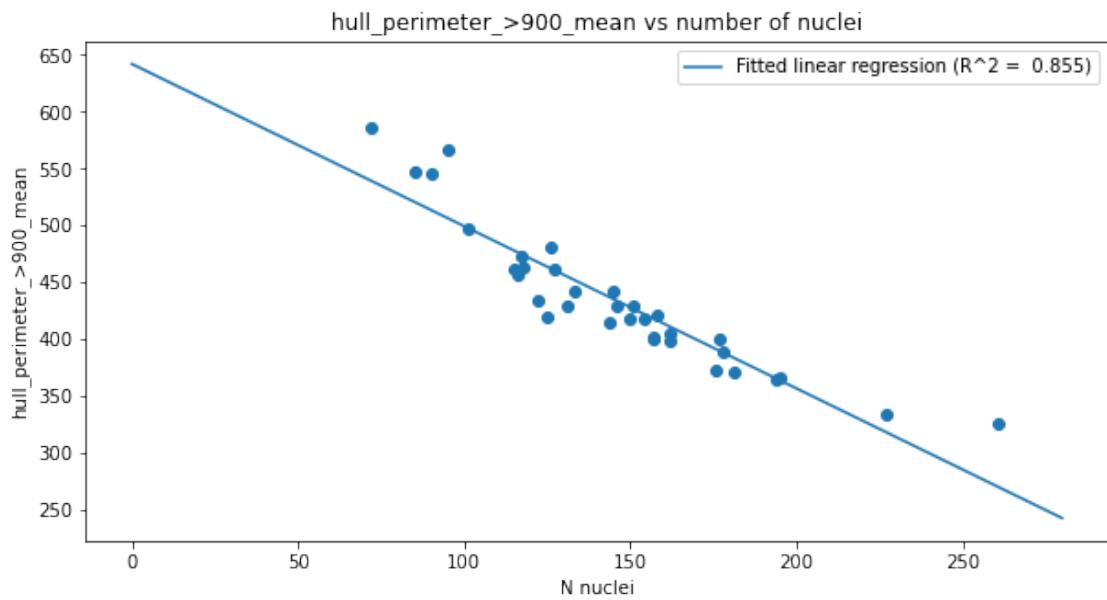
```
[6]: x_dummy = np.expand_dims(np.linspace(0, 280), axis=1)

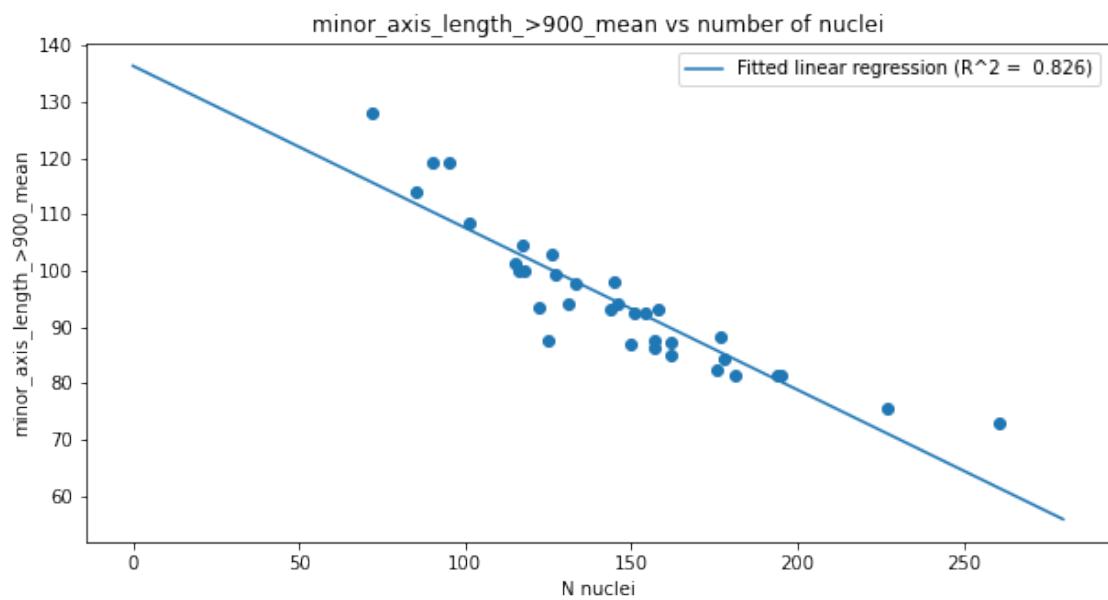
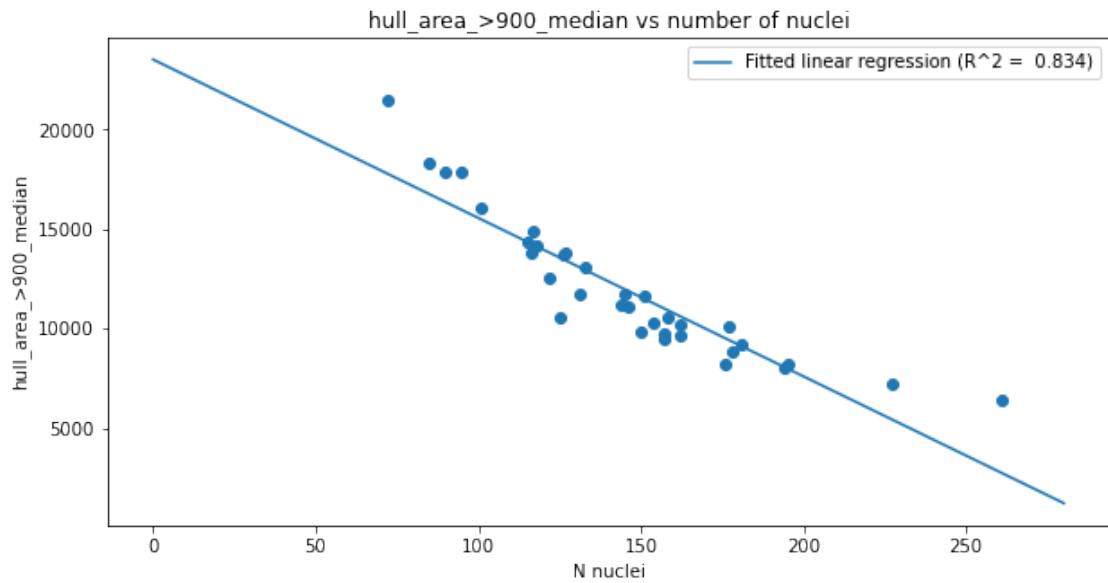
for feature_name, (r2, linear_regr) in regrs.items():
    y = data[feature_name]
    y_dummy = linear_regr.predict(x_dummy)
    plt.figure(figsize=(10, 5))
    plt.scatter(x, y)
    plt.plot(x_dummy, y_dummy, label=f'Fitted linear regression (R^2 = {r2:.3f})')
    plt.title(f'{feature_name} vs number of nuclei')
    plt.ylabel(feature_name)
    plt.xlabel('N nuclei')
    plt.legend()
    plt.show()
    plt.close()
```

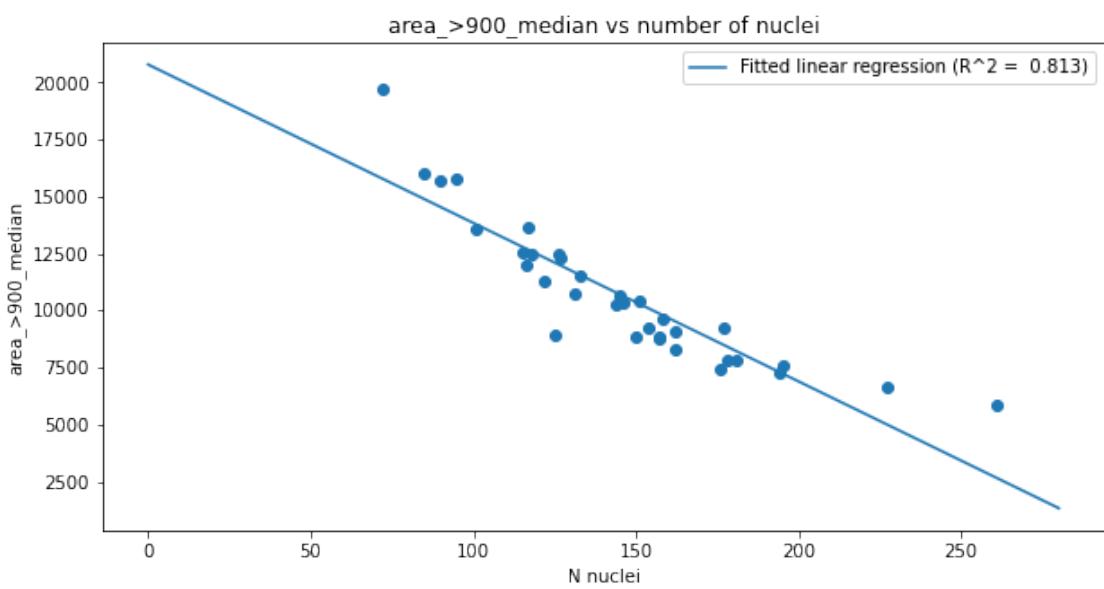
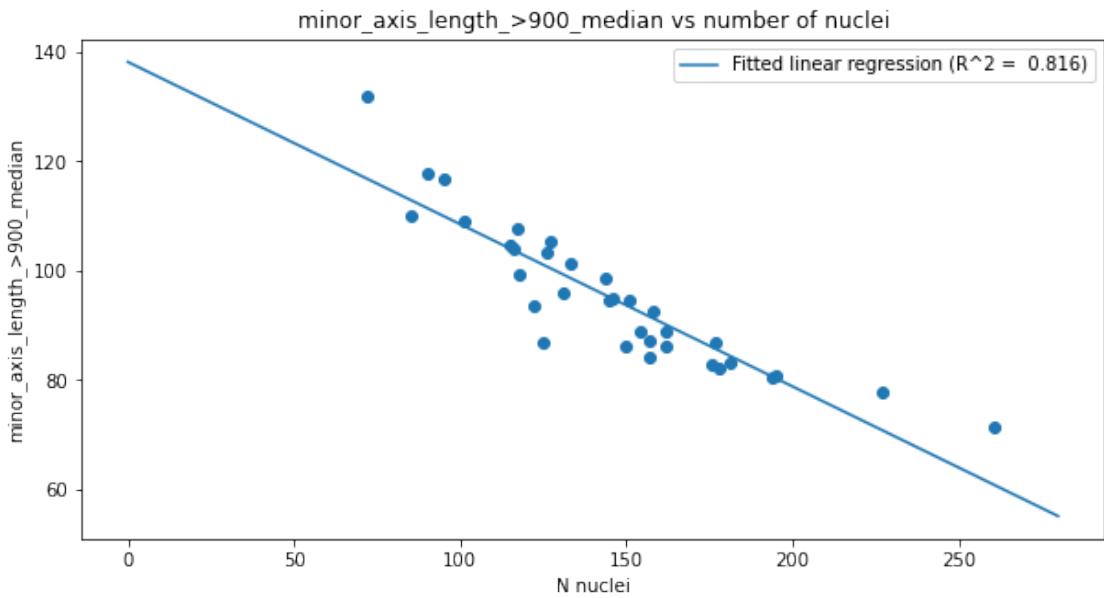




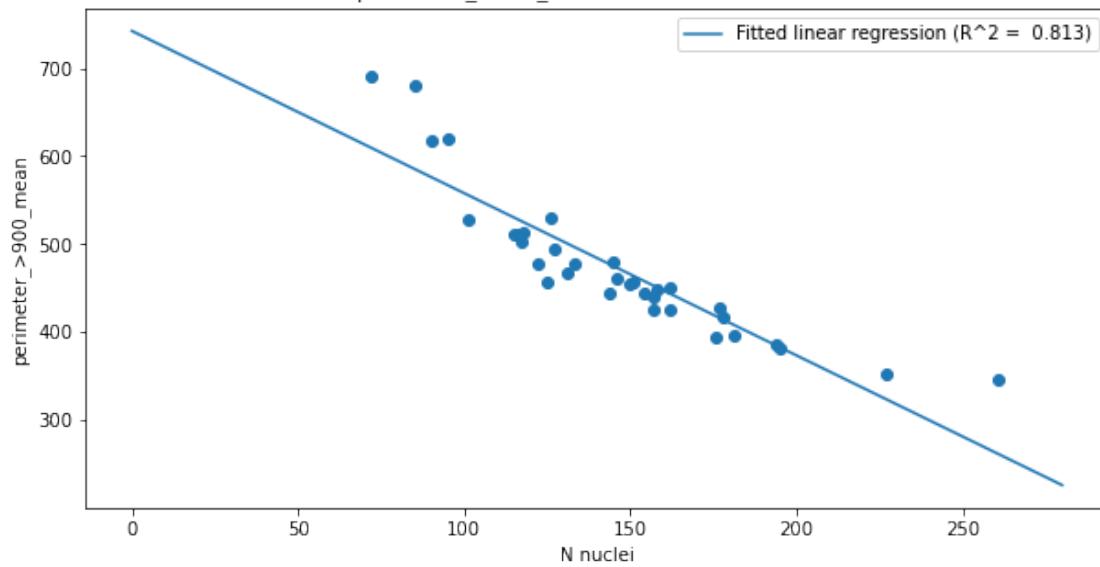




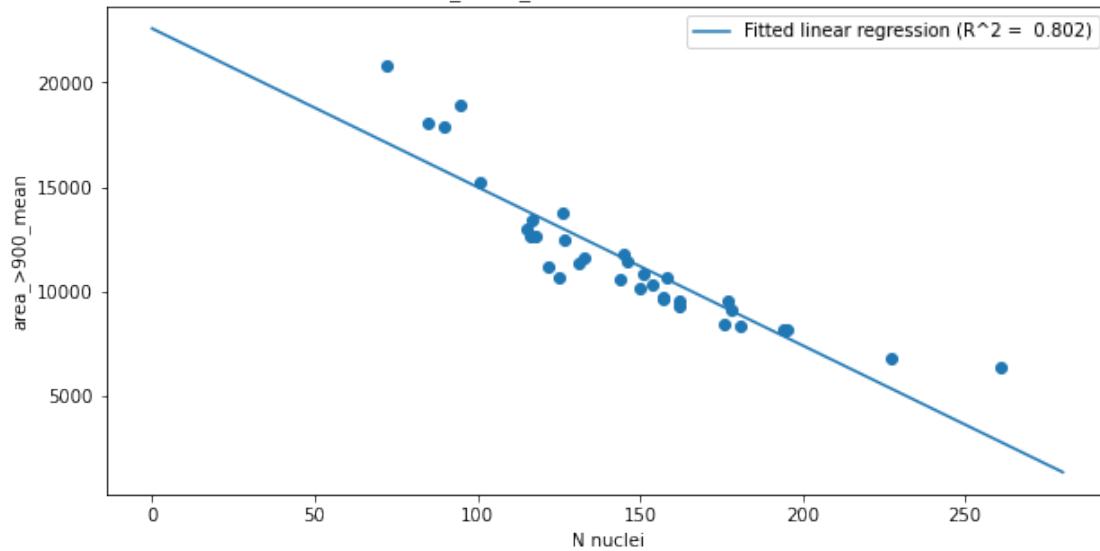


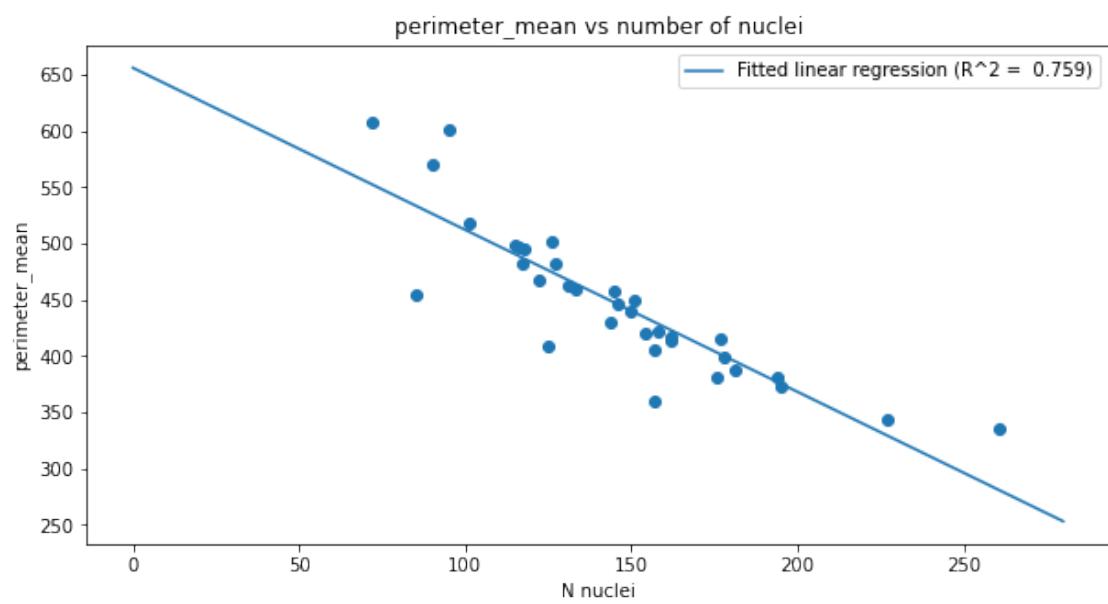
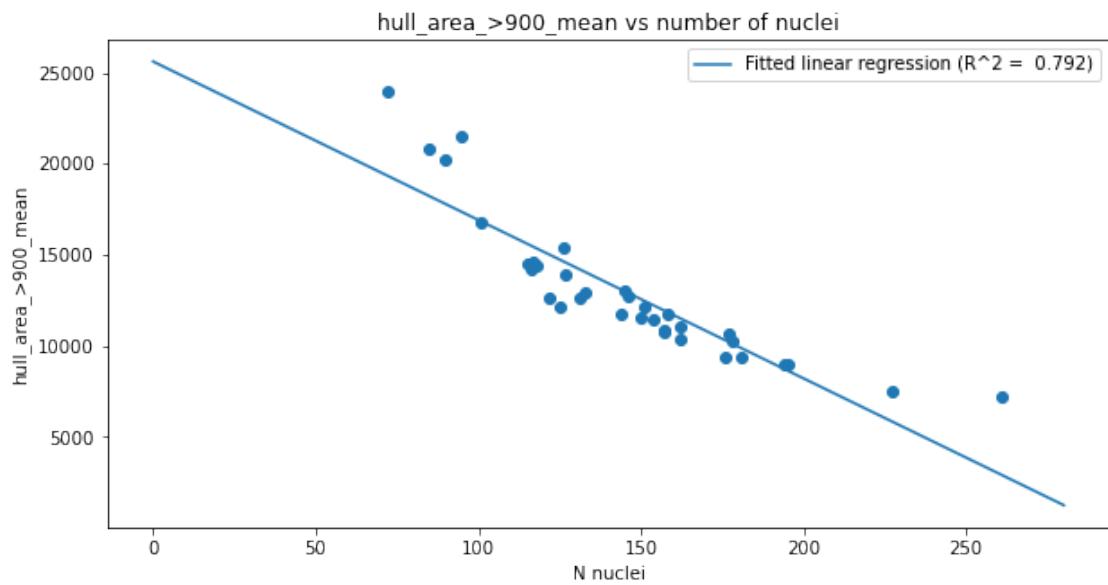


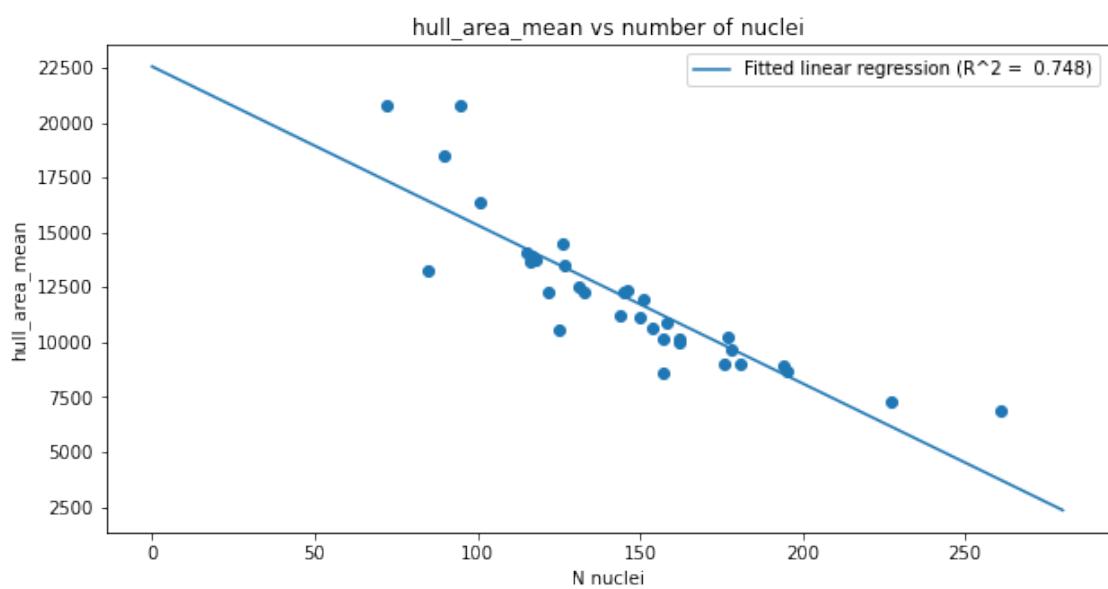
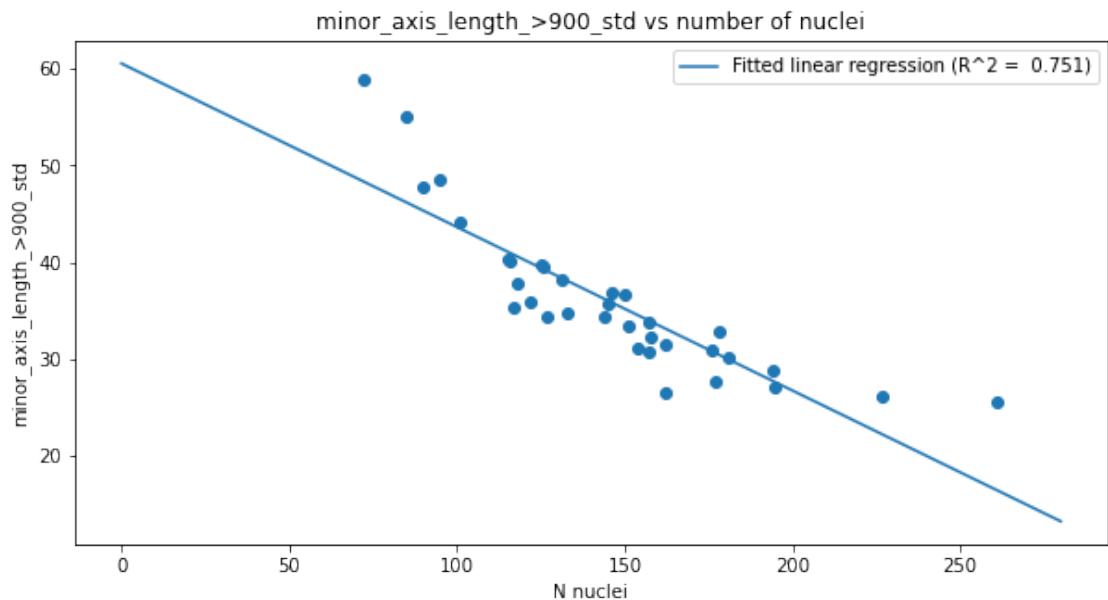
perimeter_>900_mean vs number of nuclei

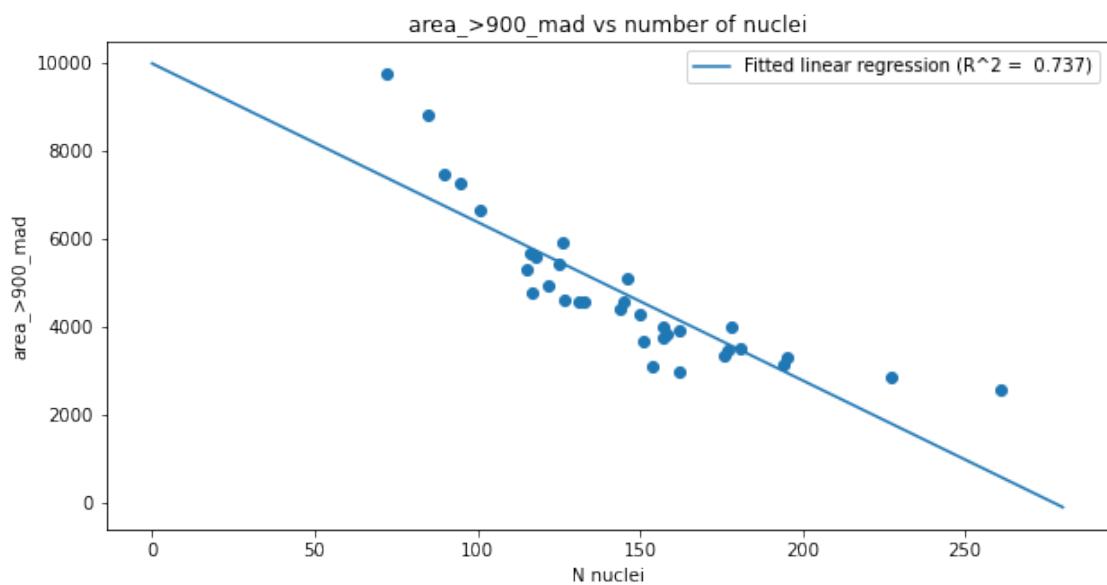
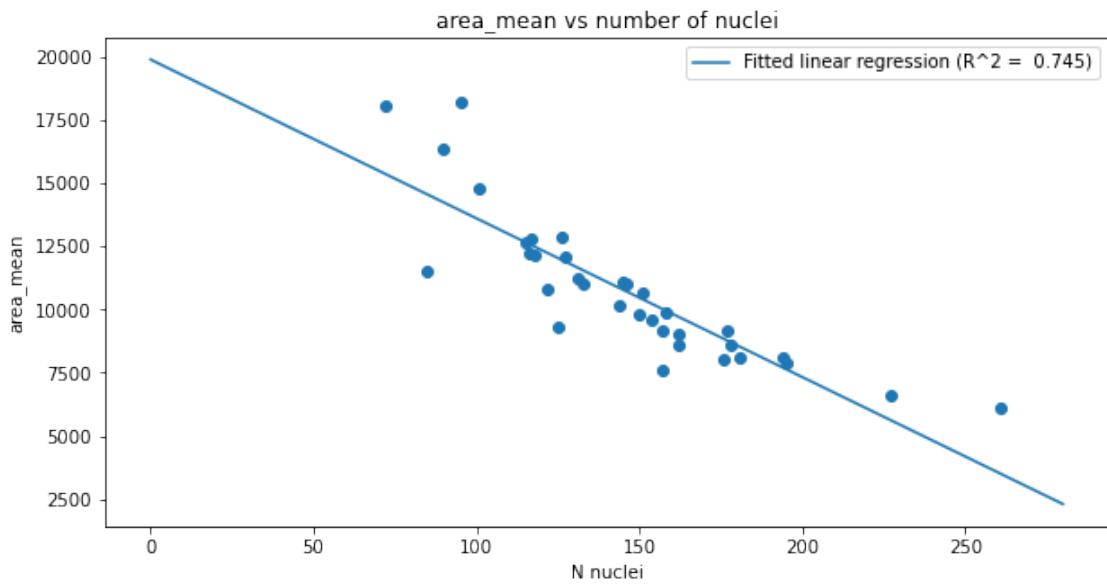


area_>900_mean vs number of nuclei

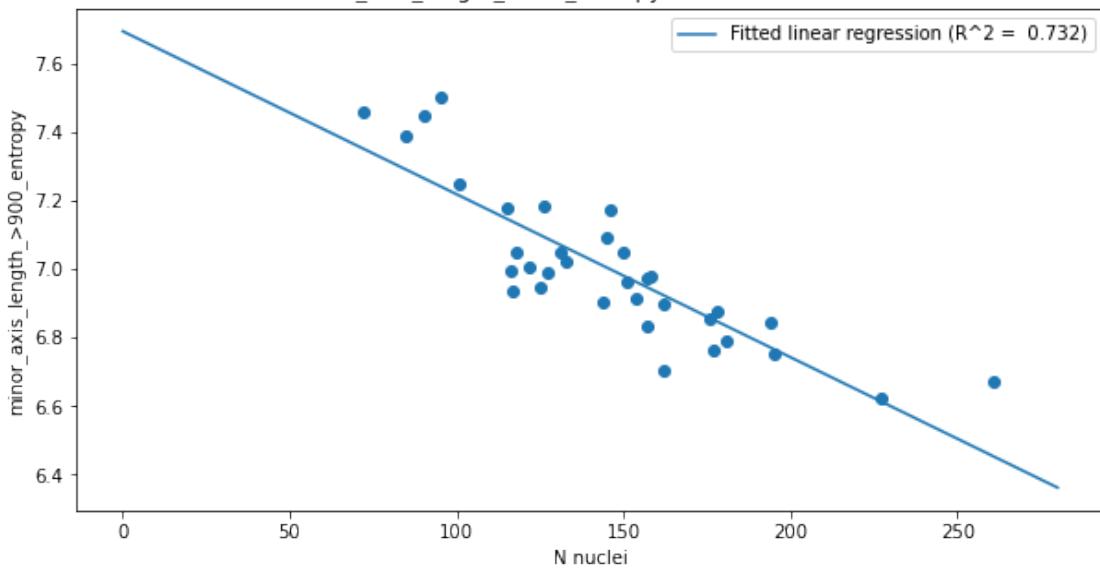




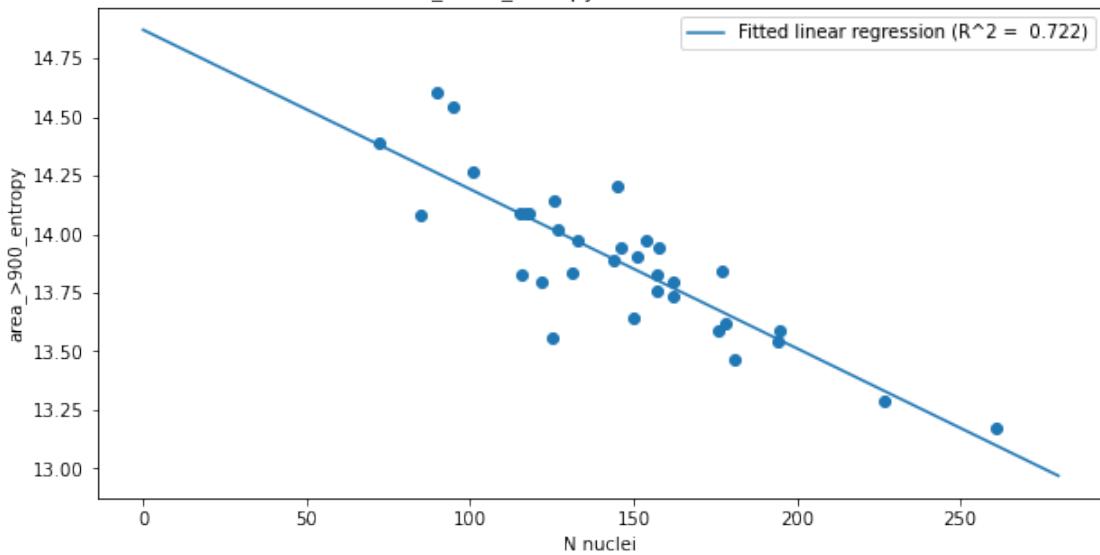


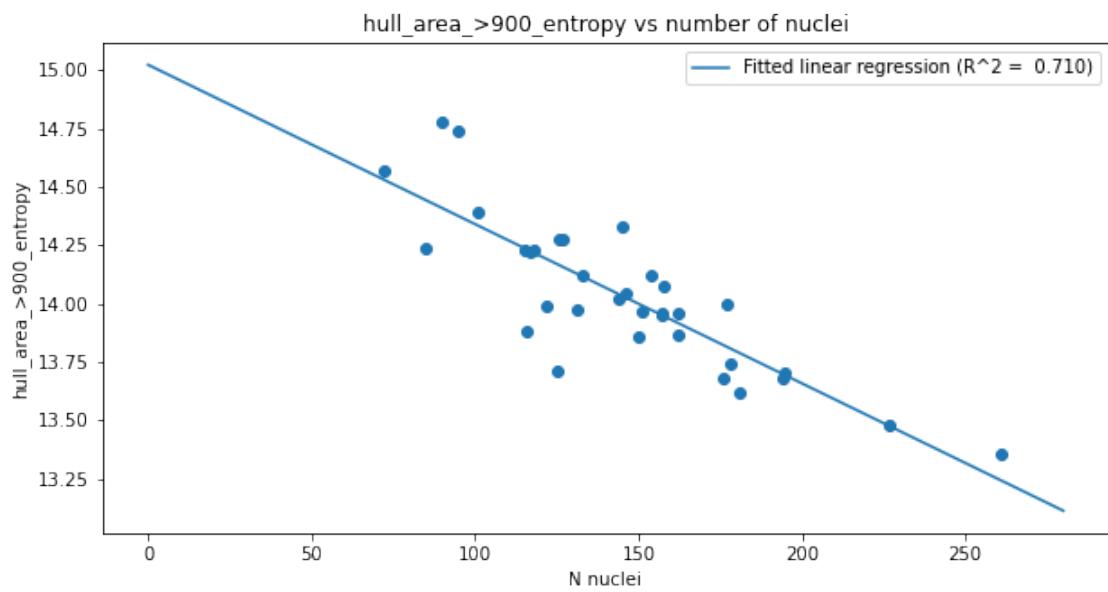
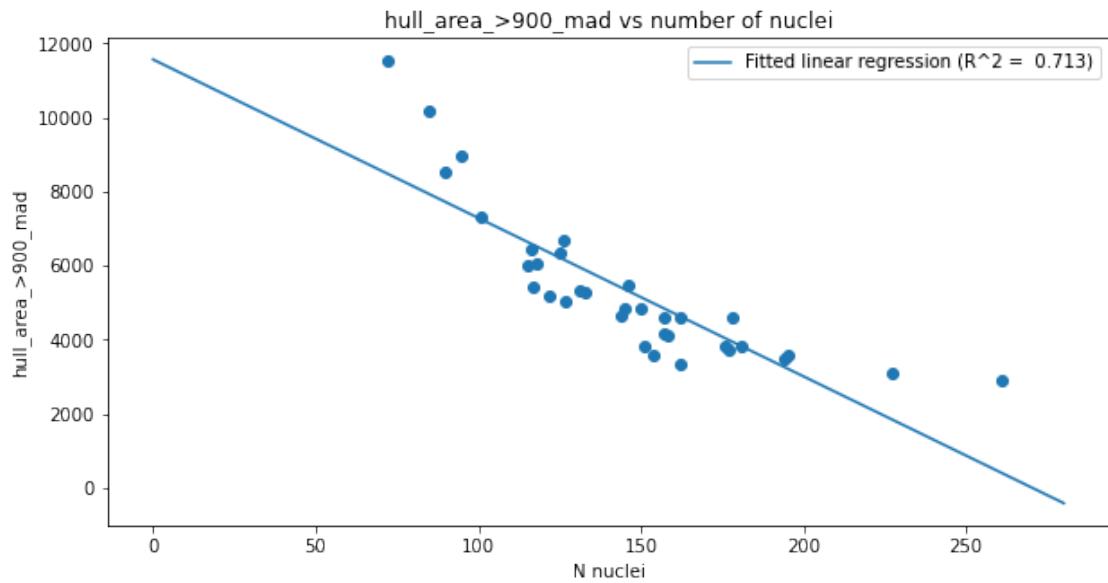


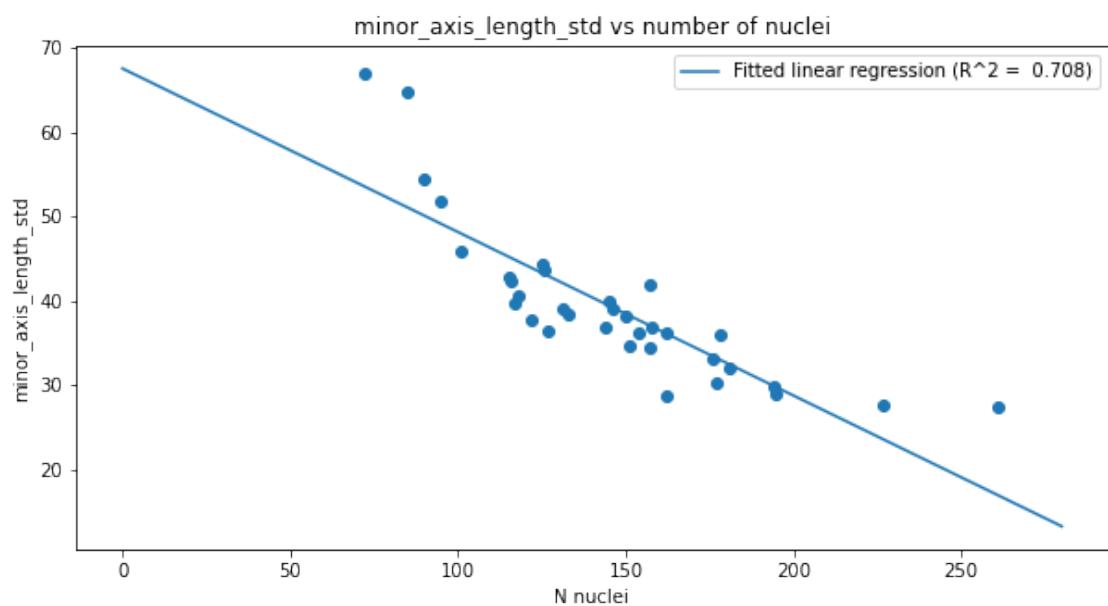
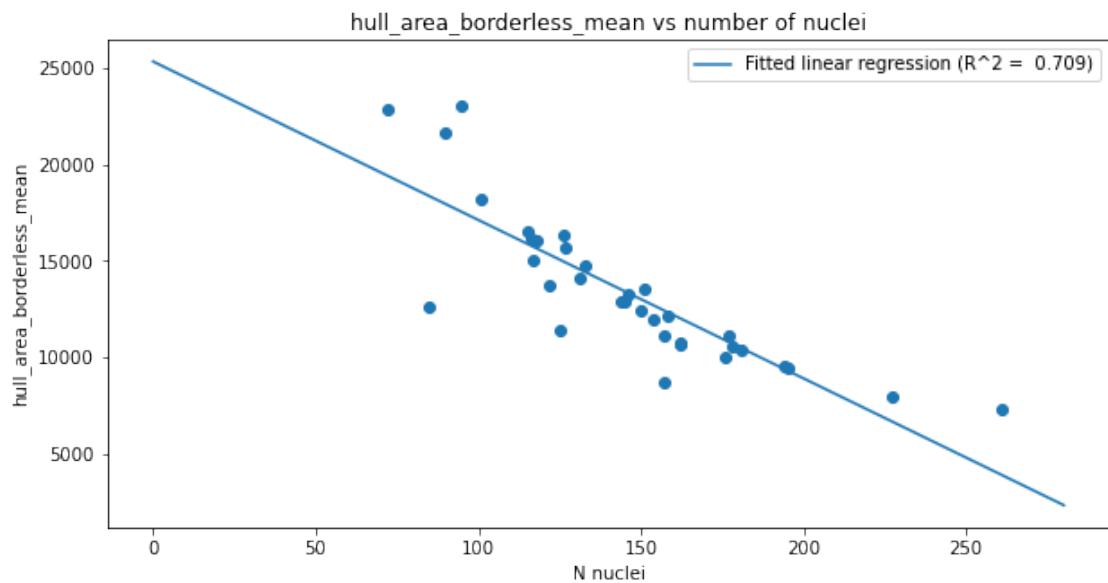
minor_axis_length >900_entropy vs number of nuclei

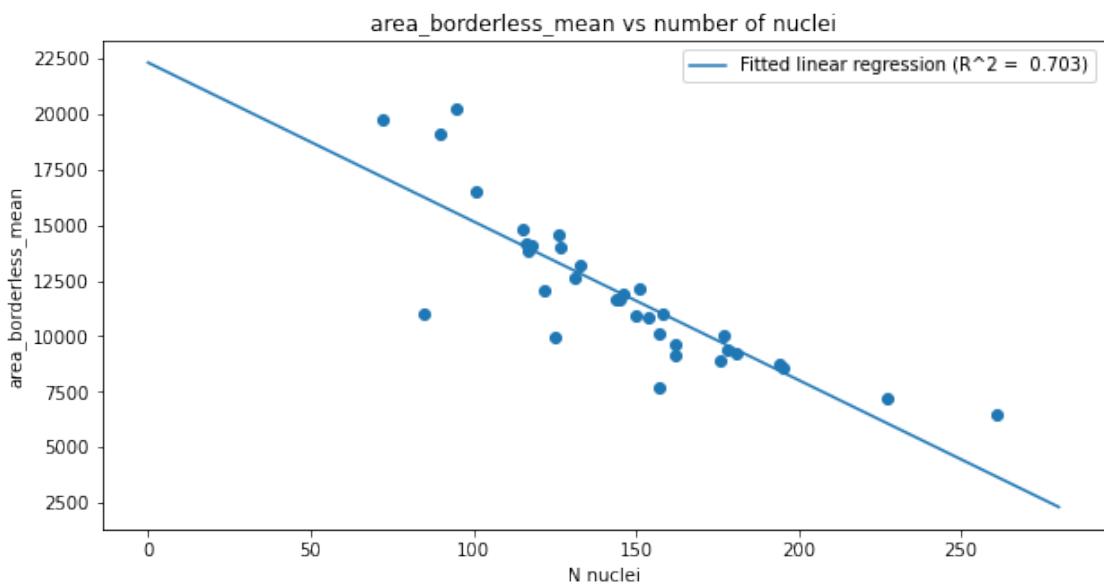
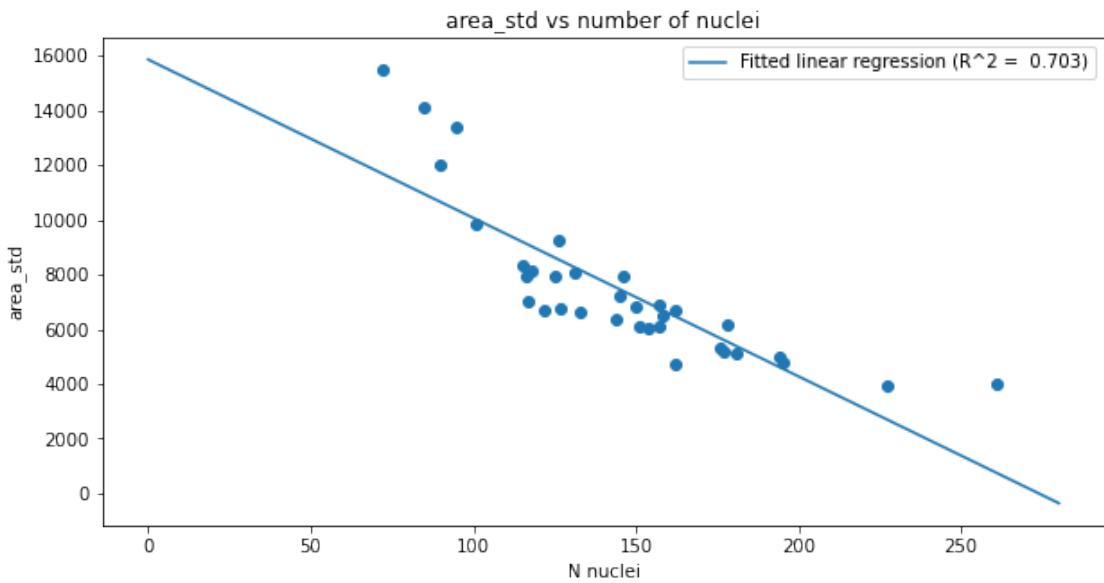


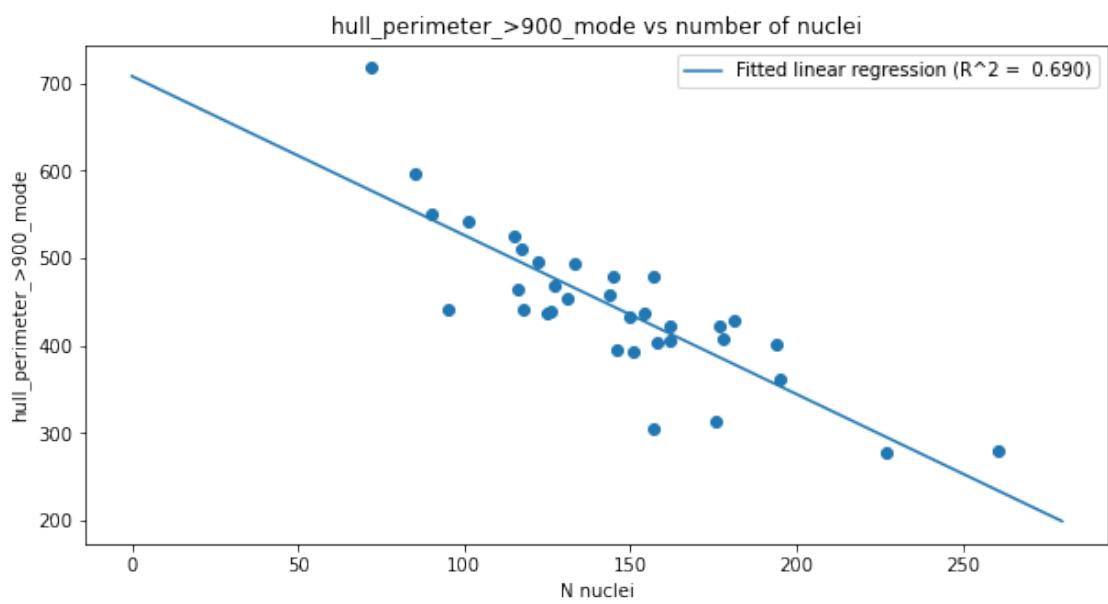
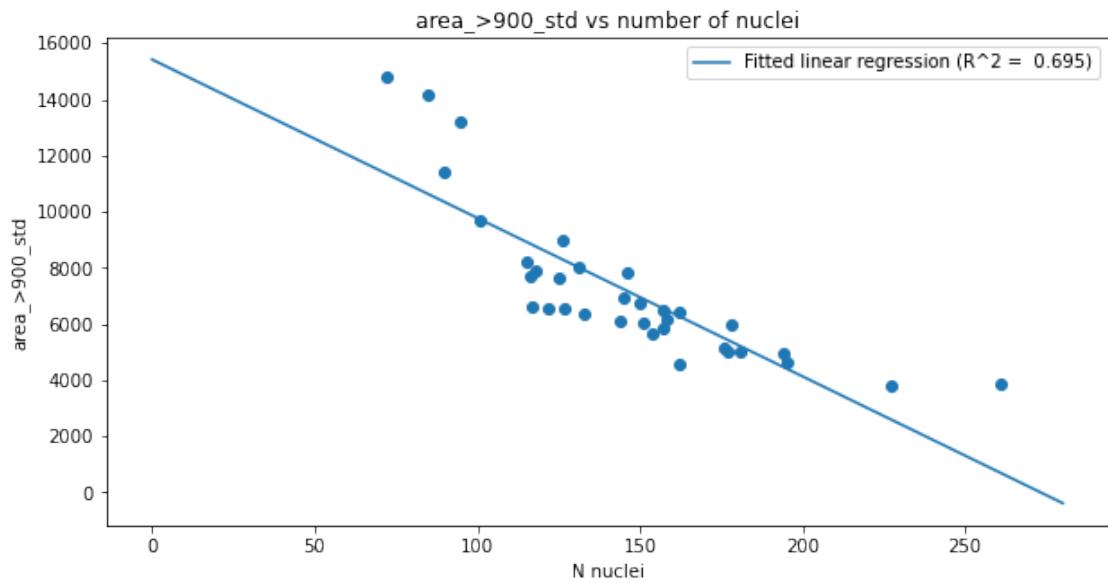
area >900_entropy vs number of nuclei

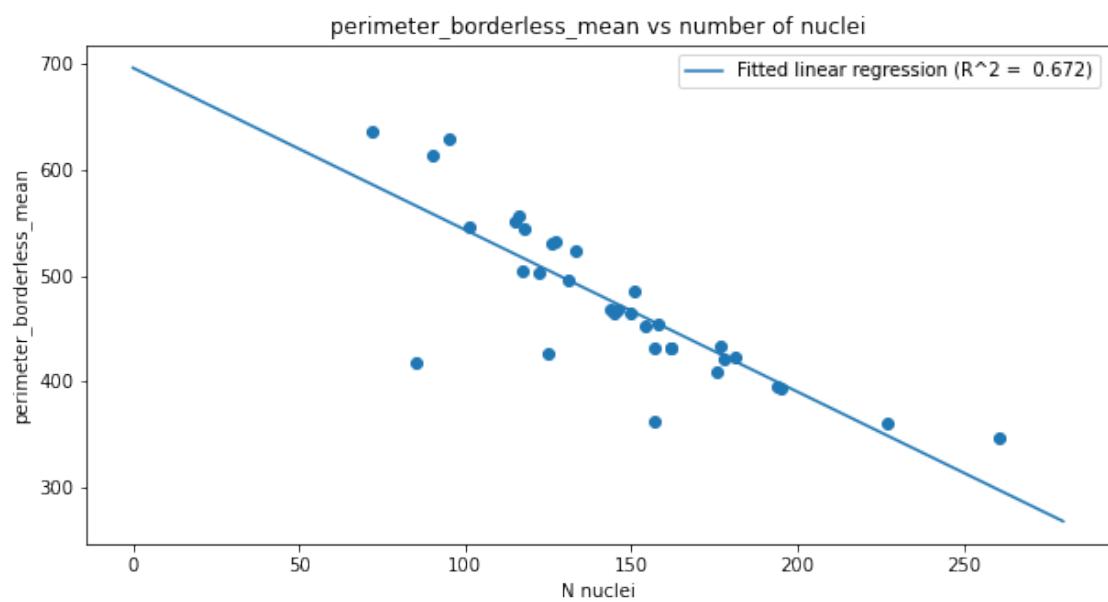
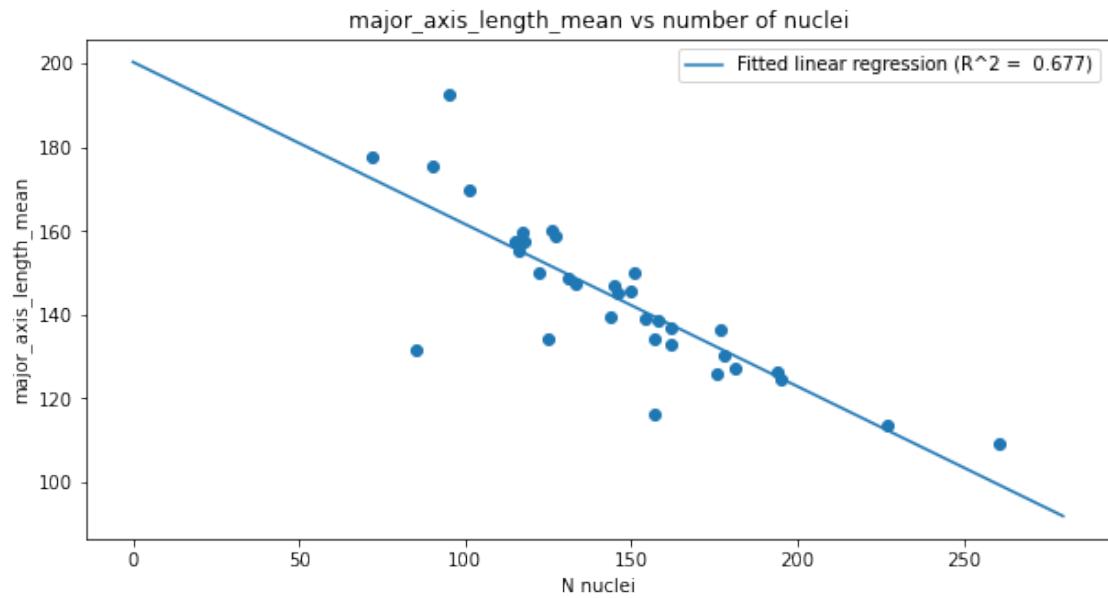


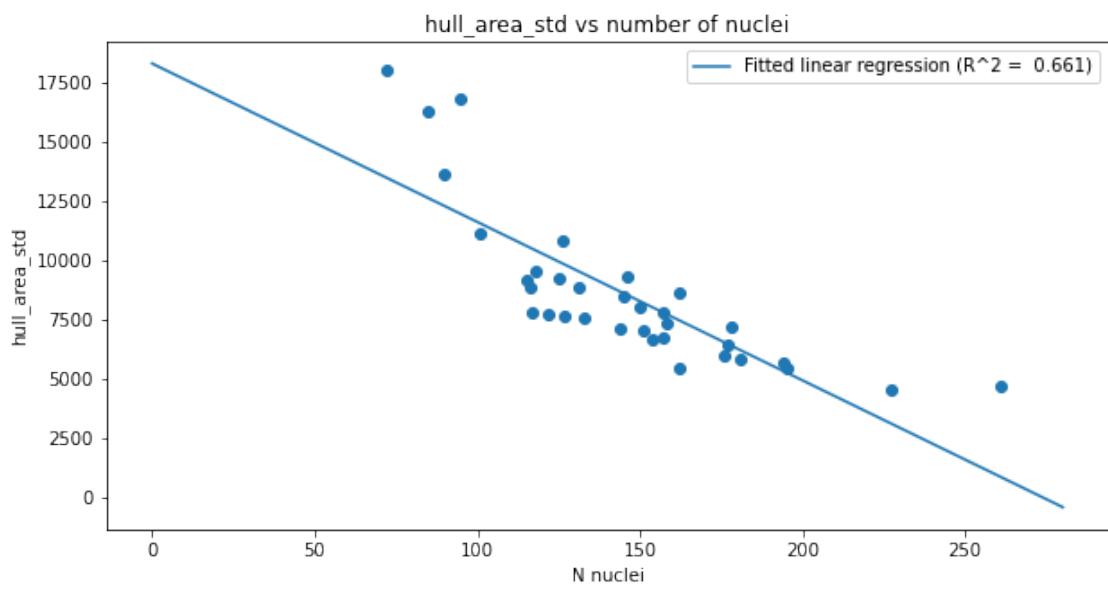
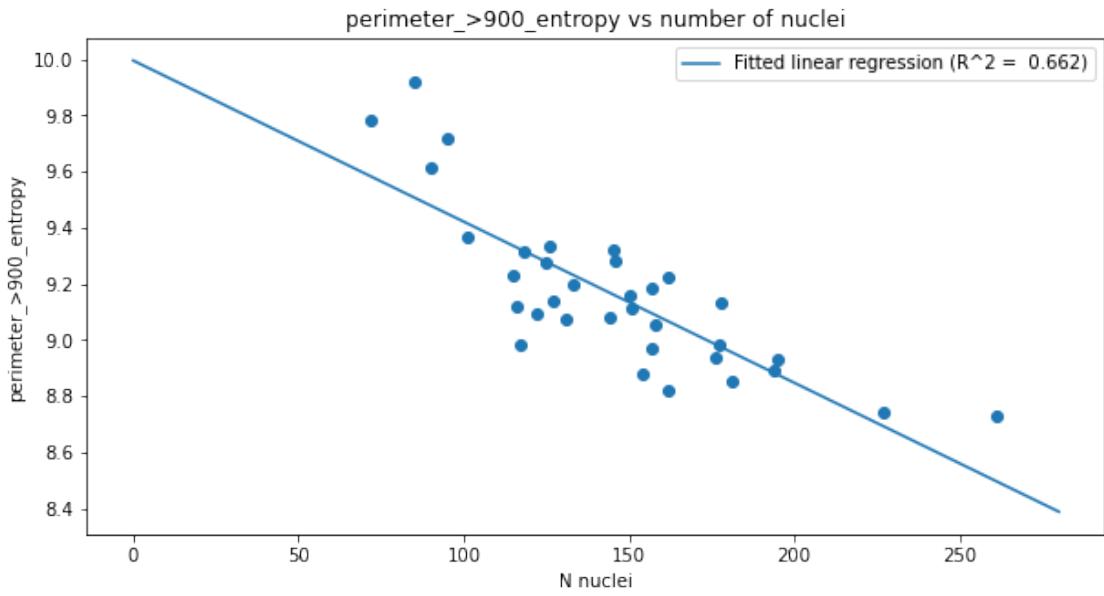


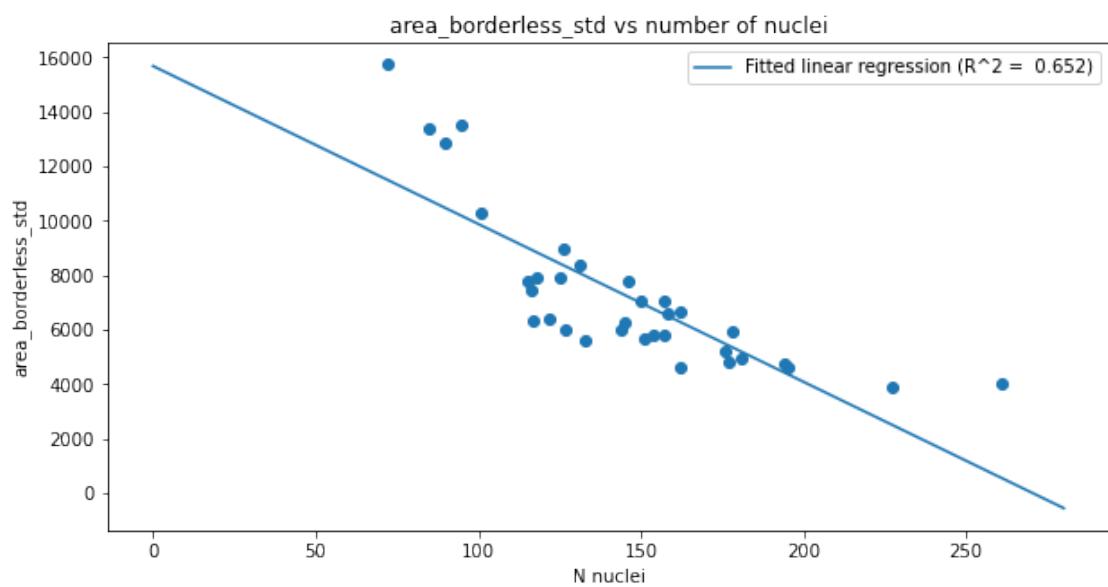
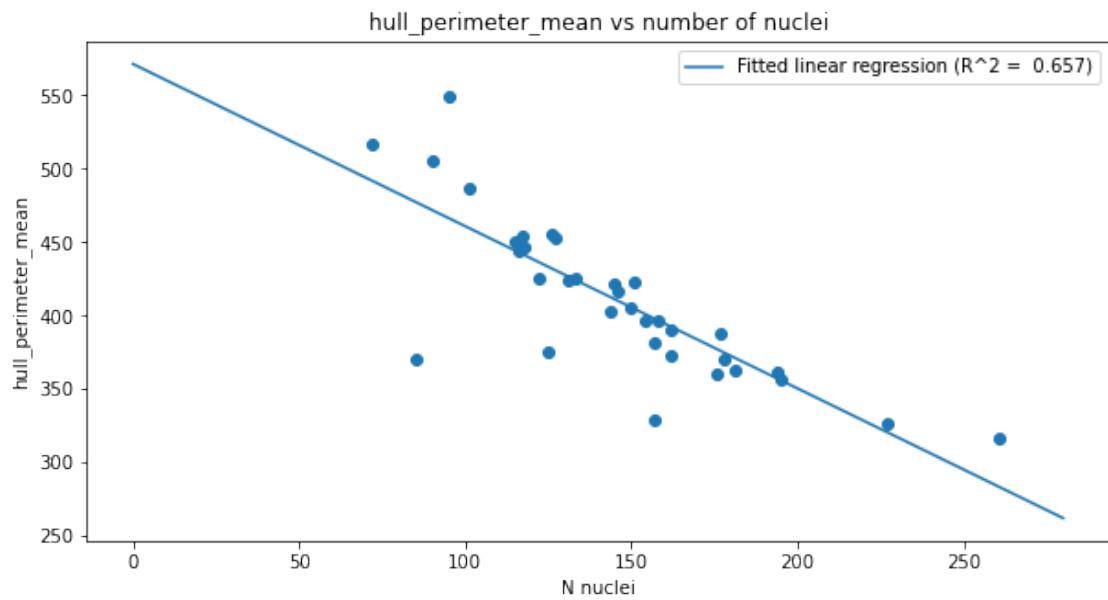


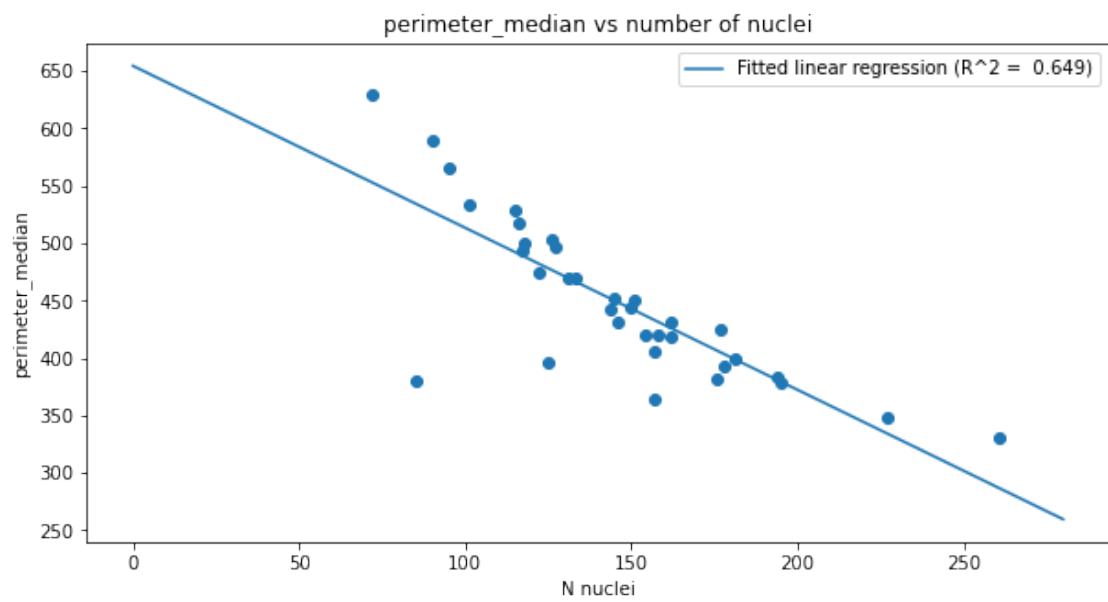
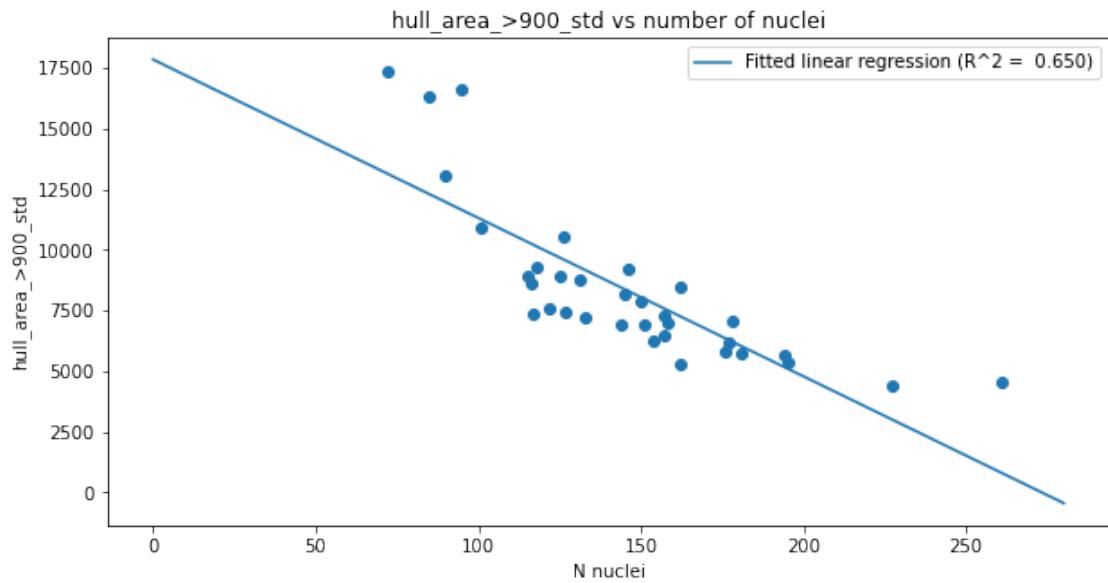


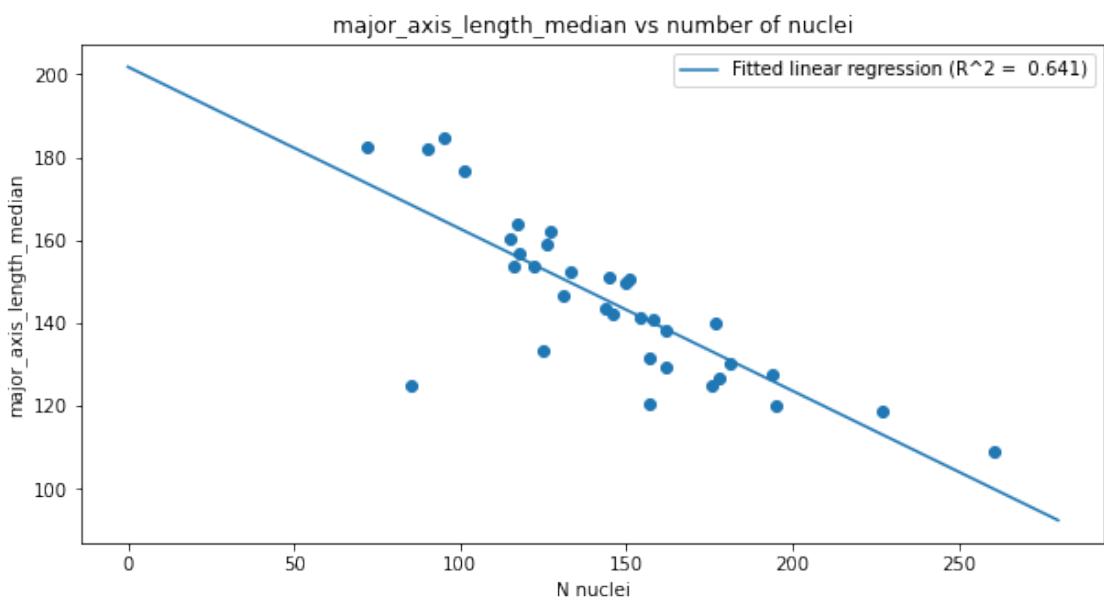
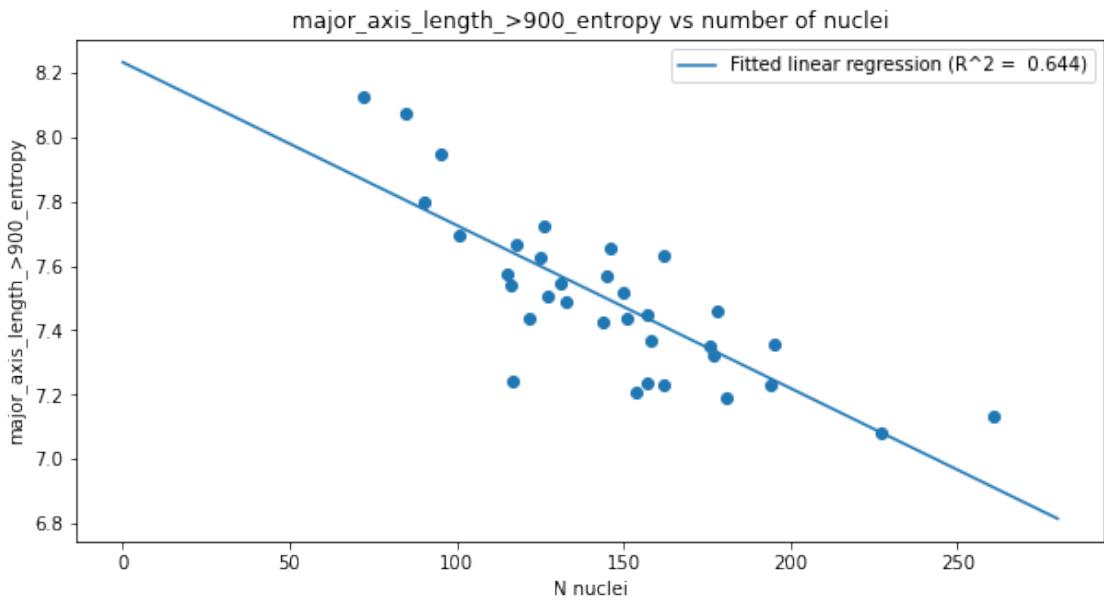


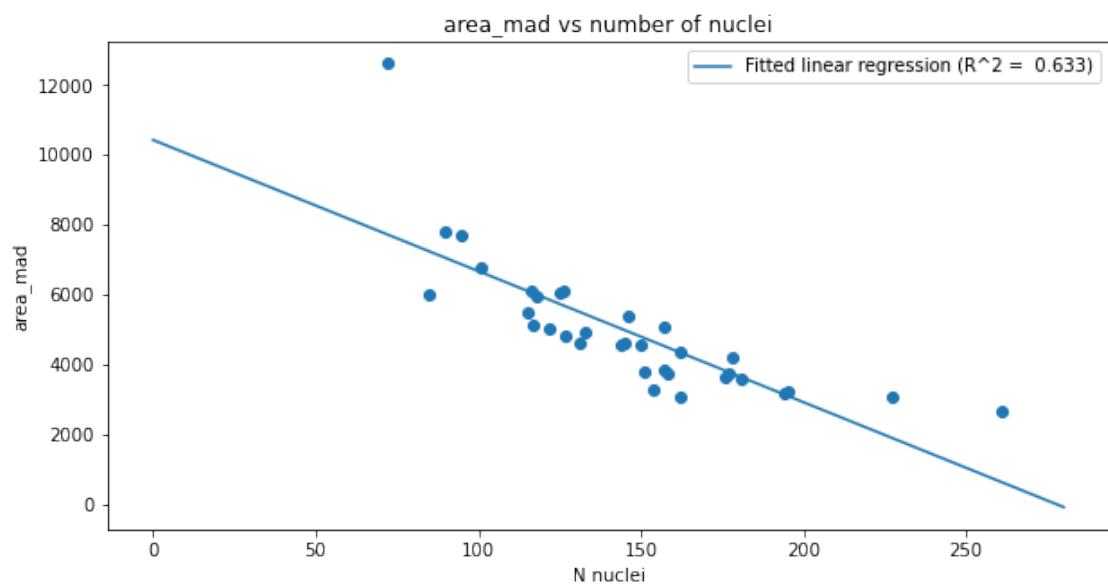
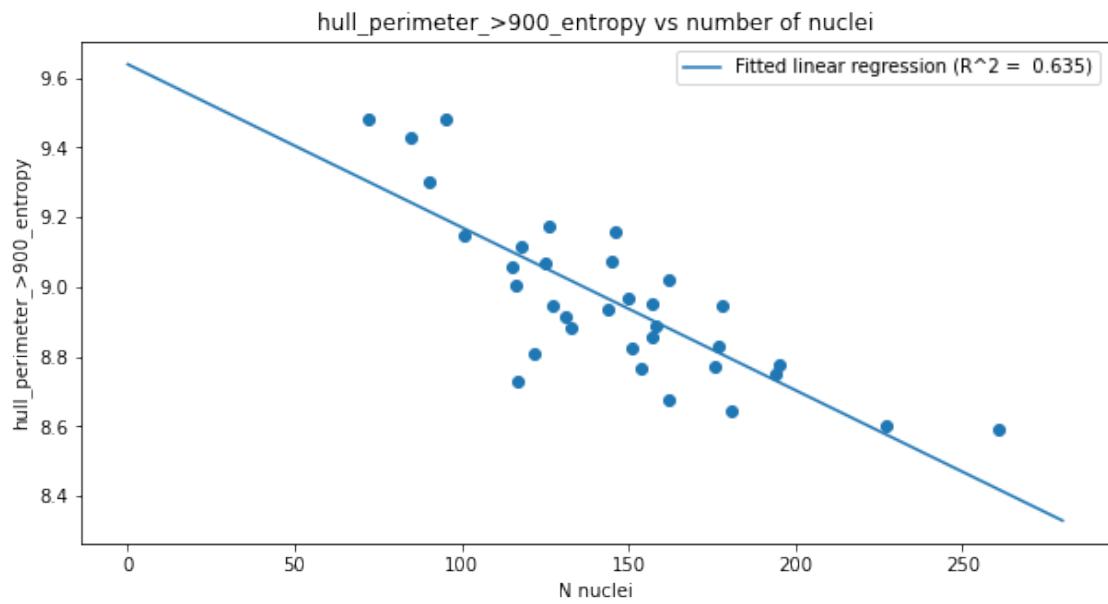


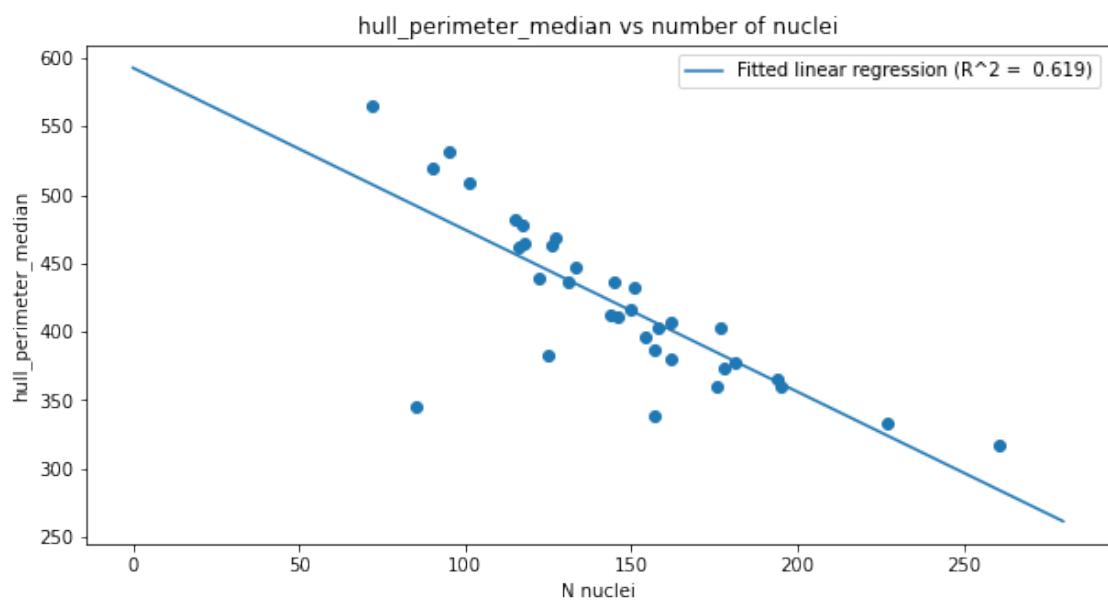
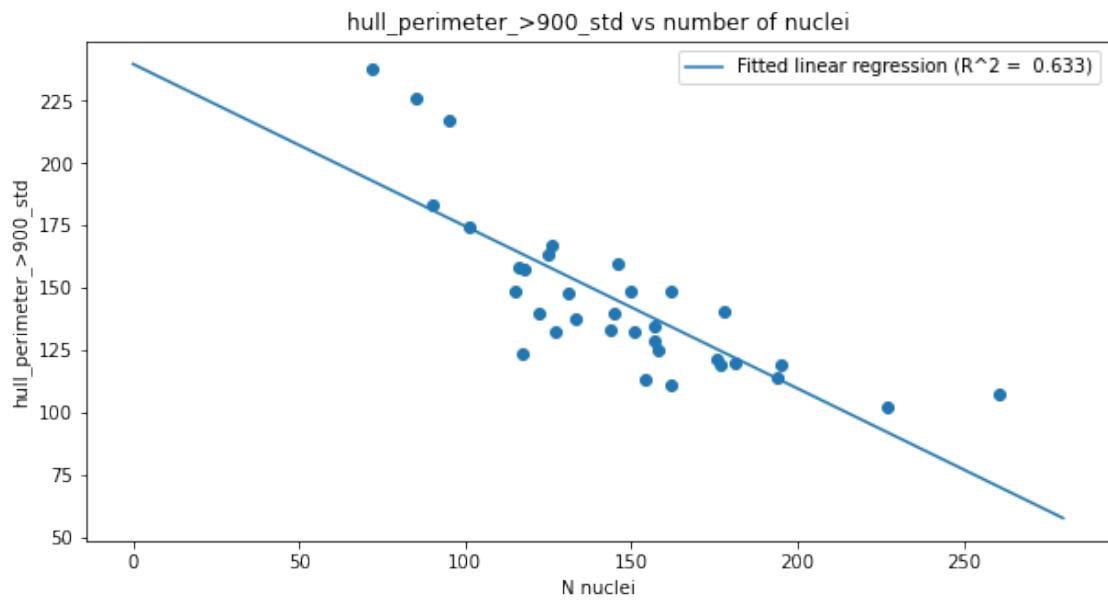


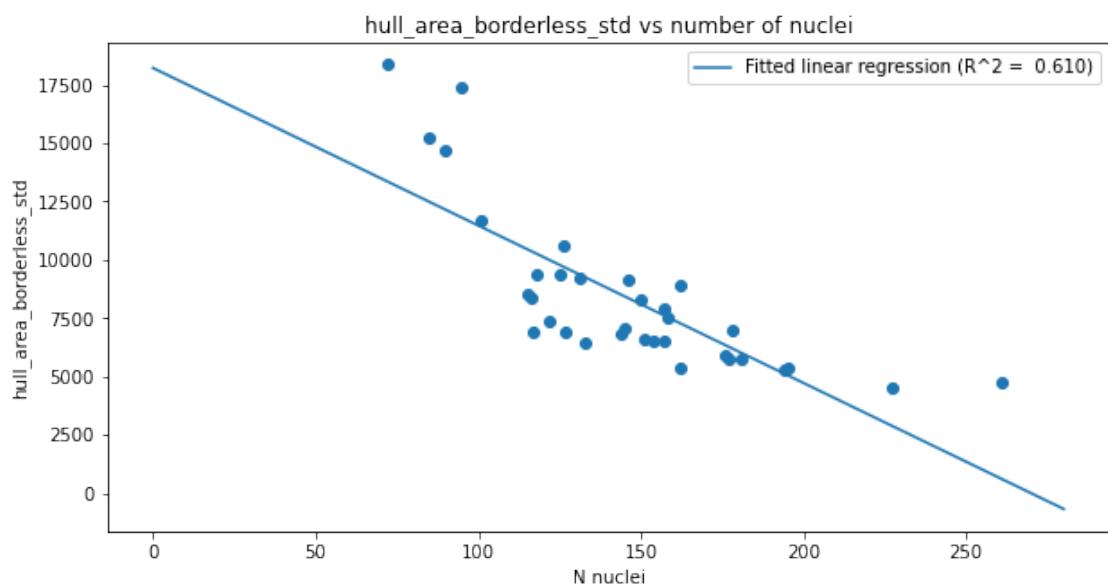
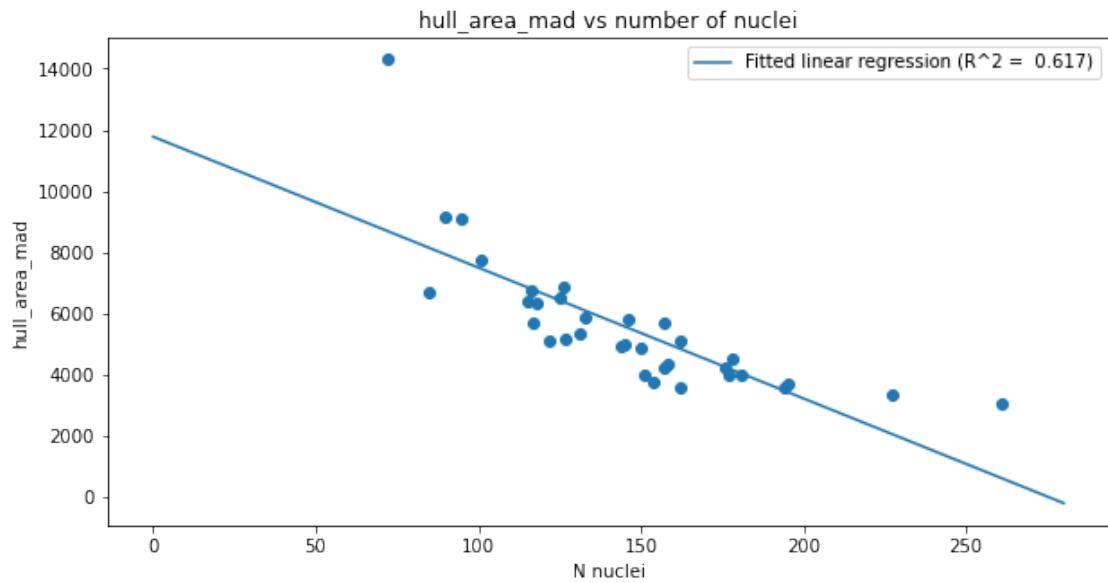


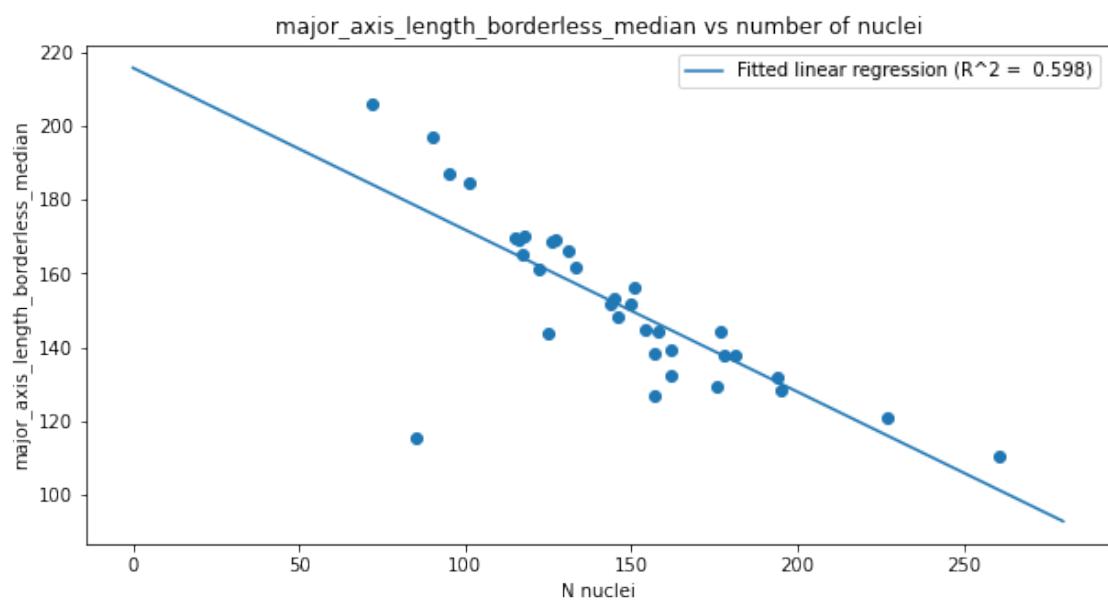
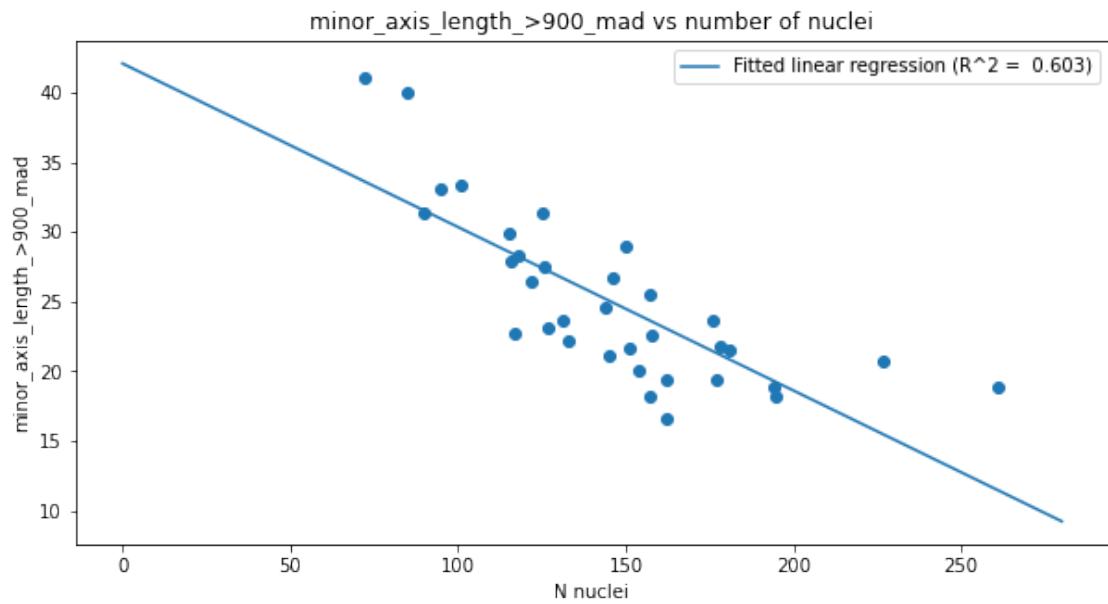


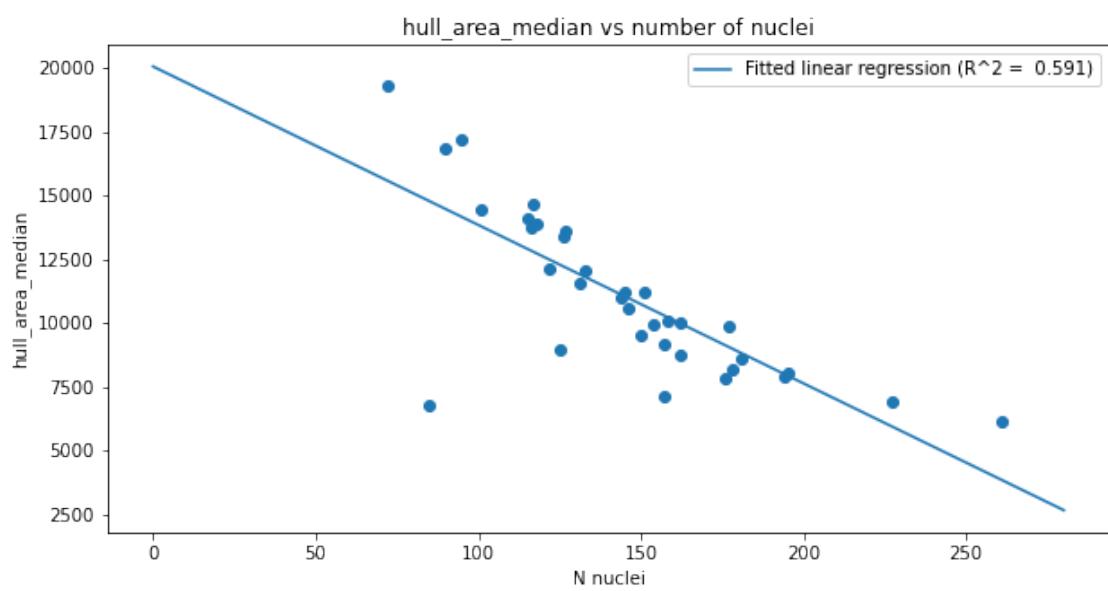
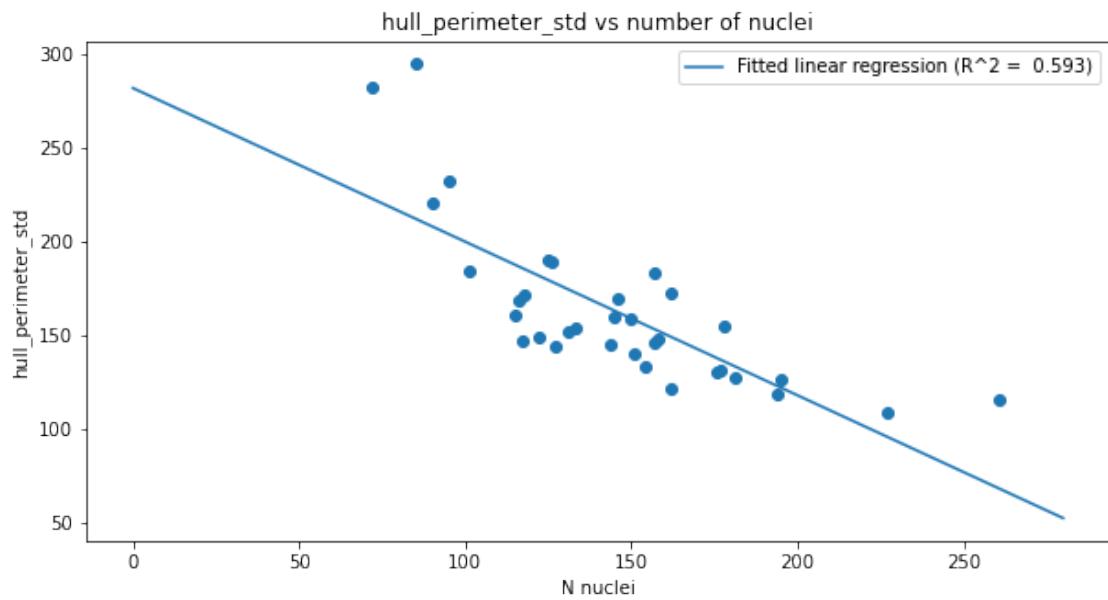


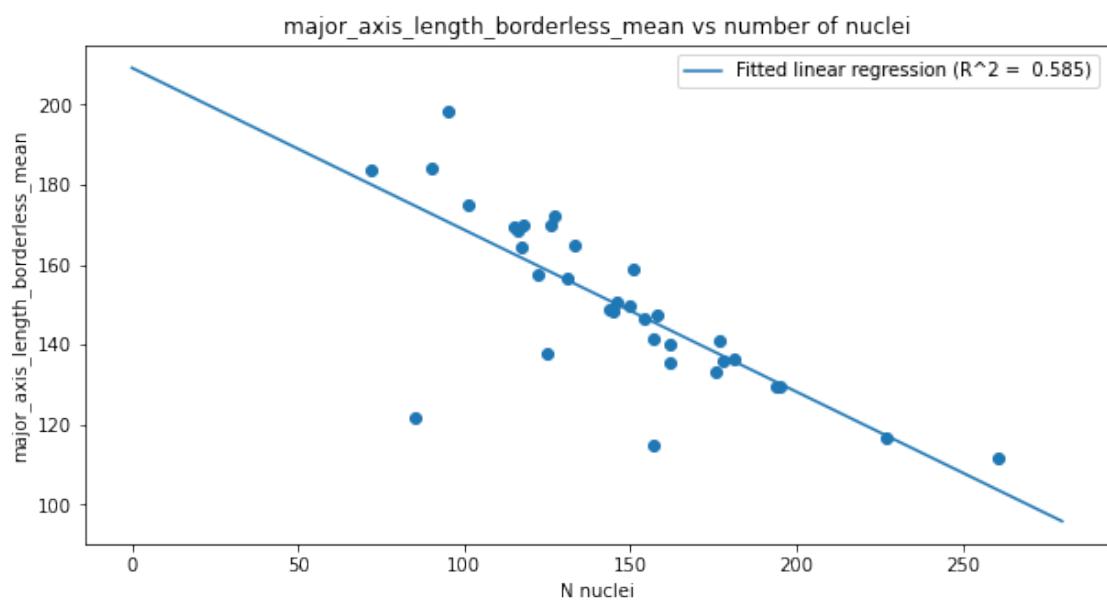
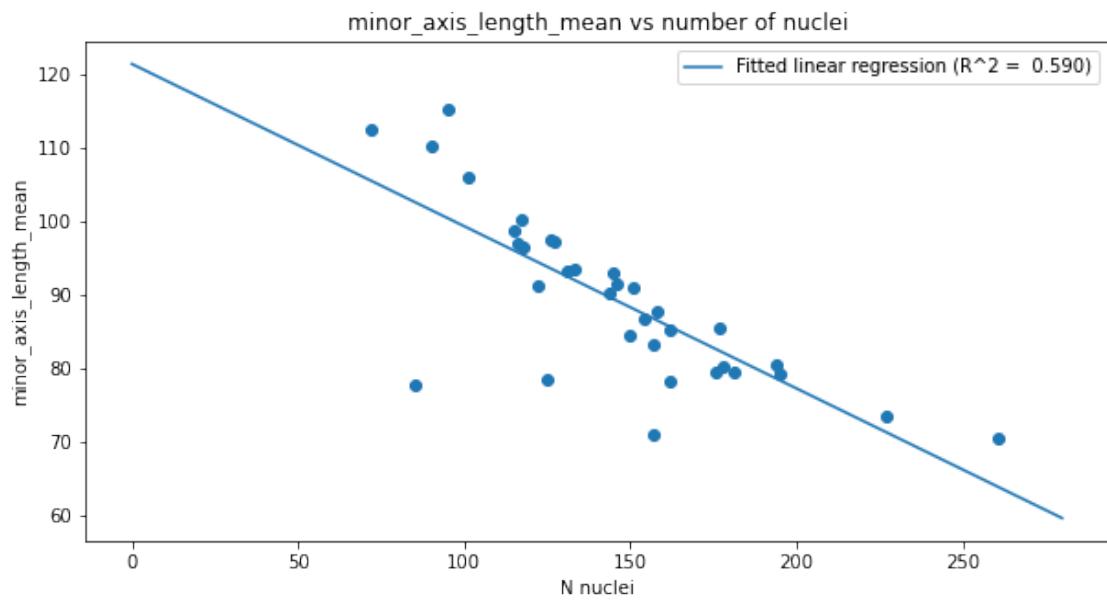


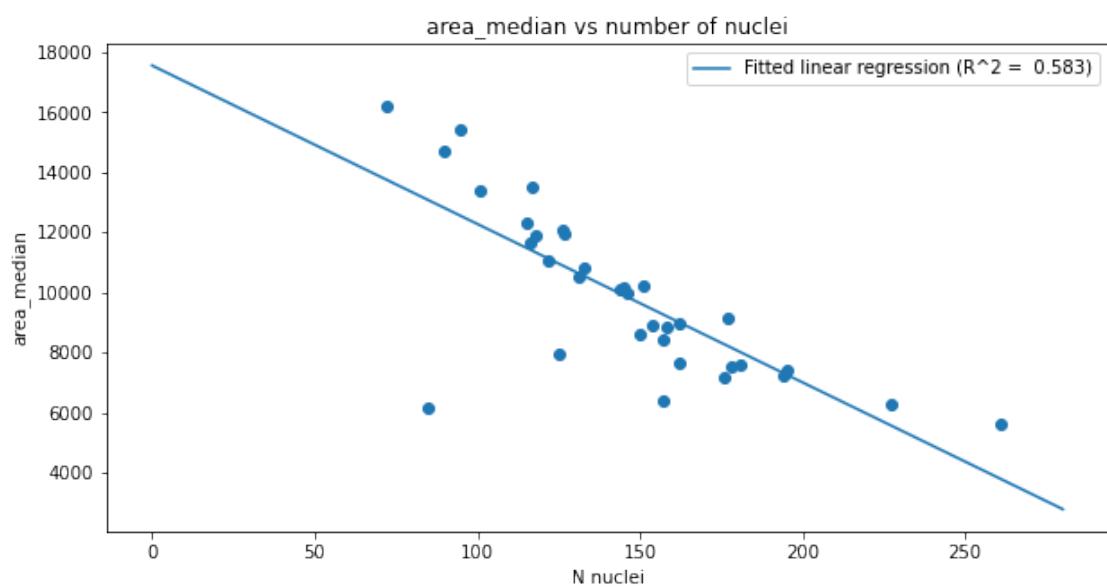
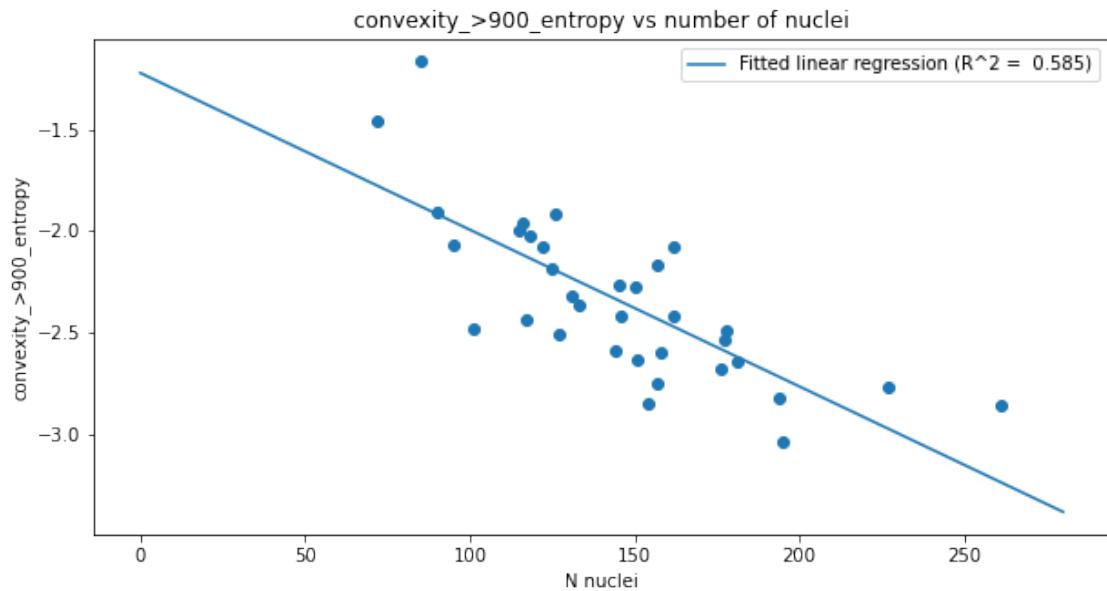


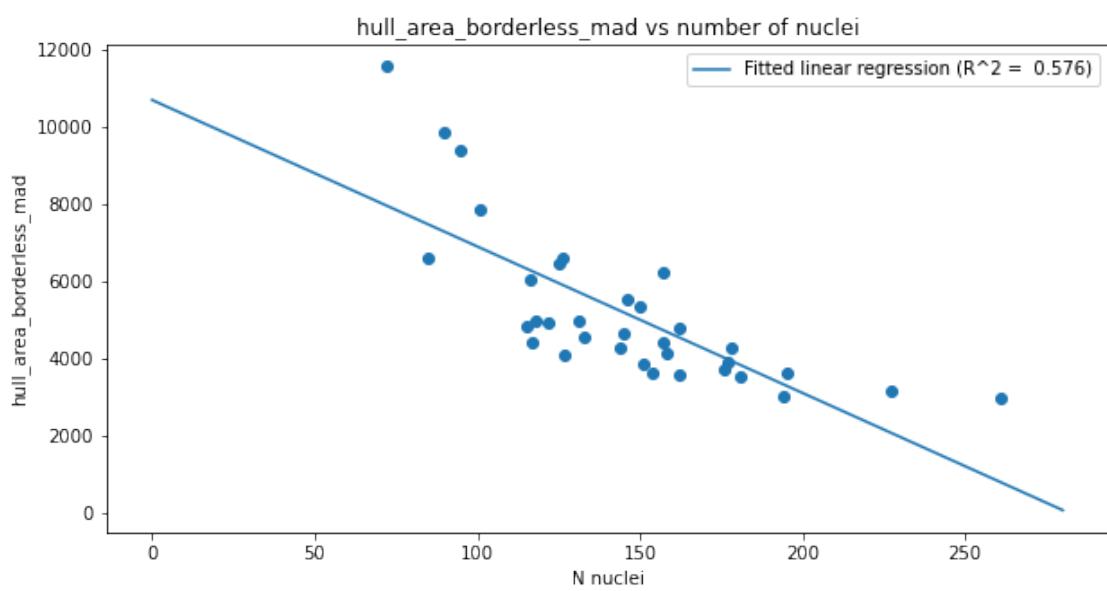
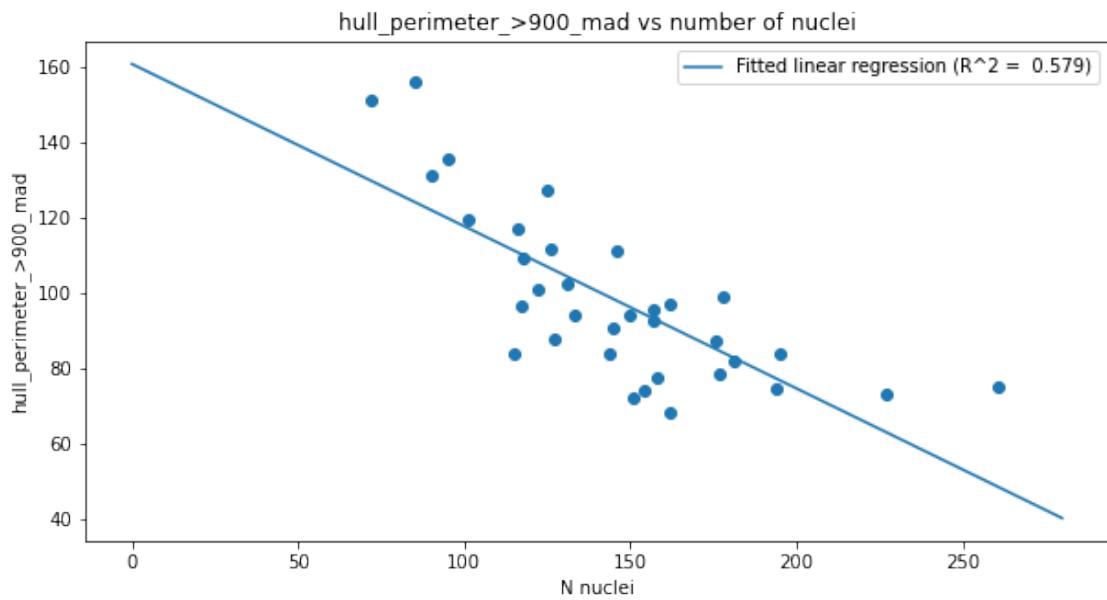


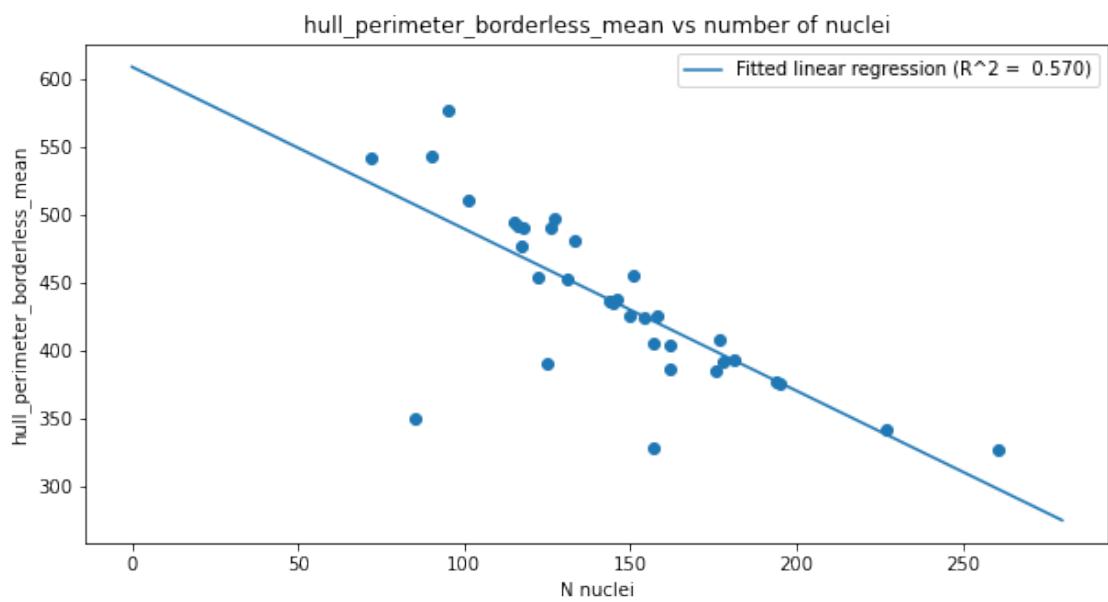
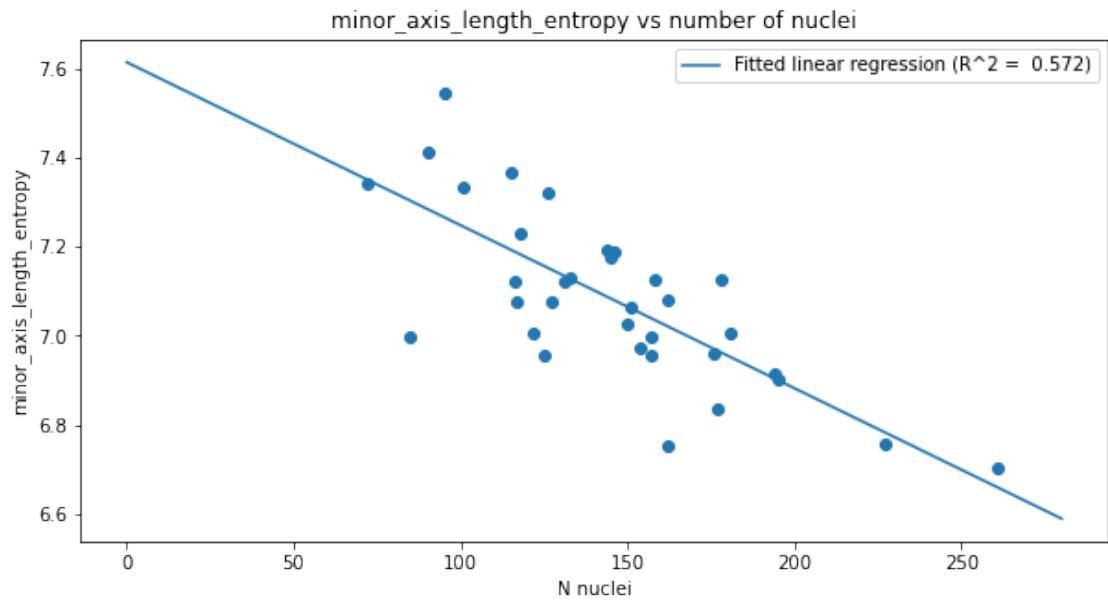


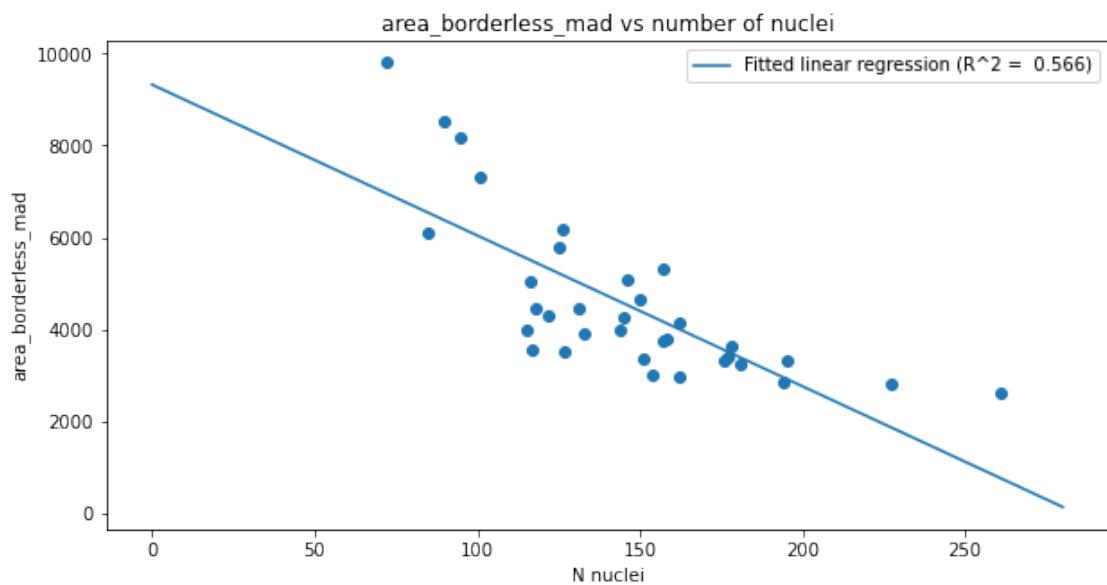
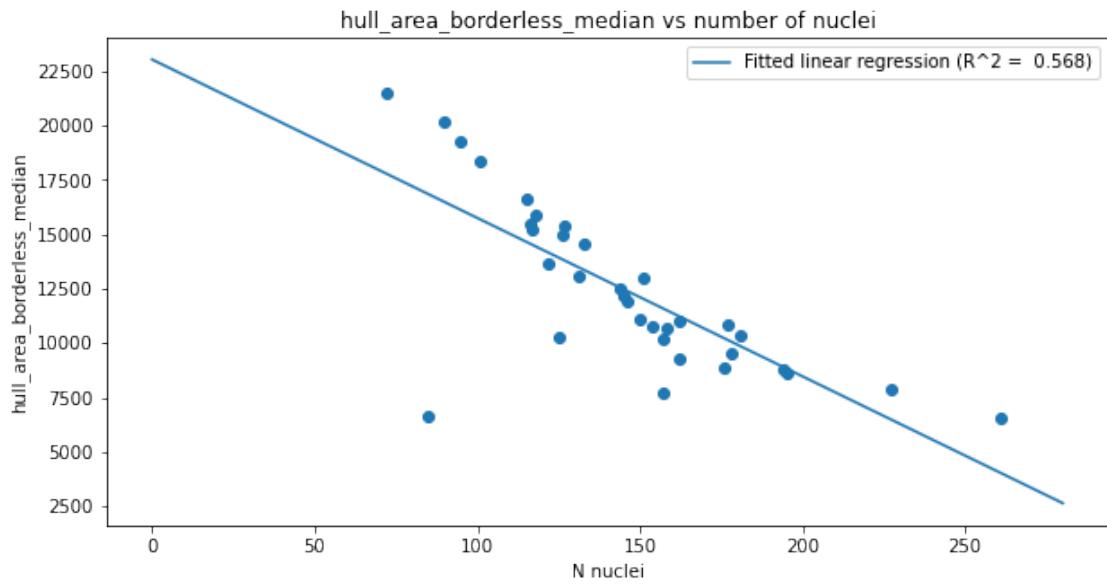


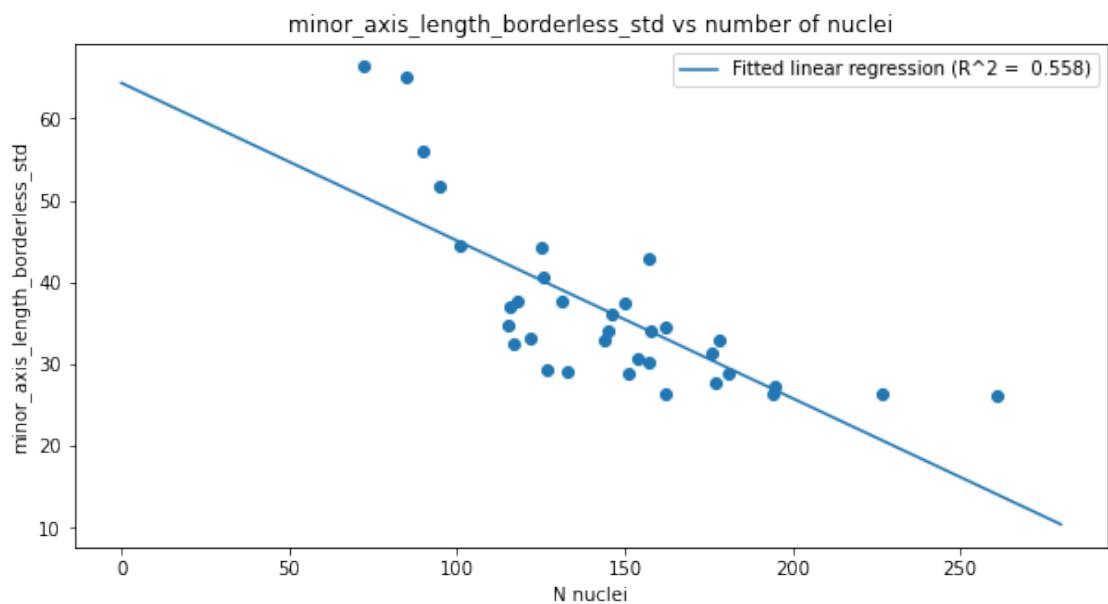
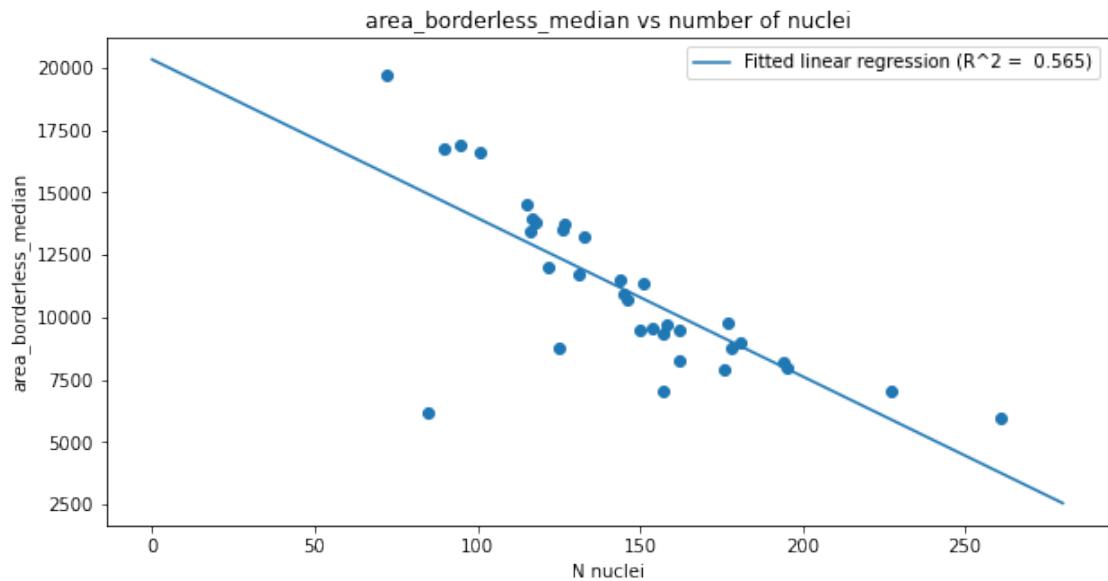


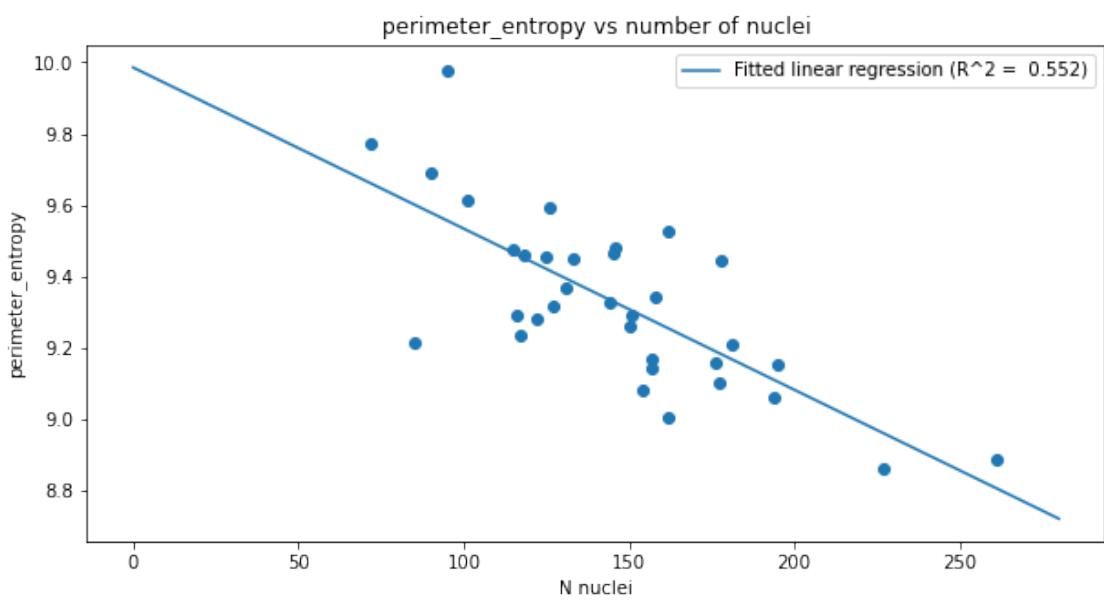
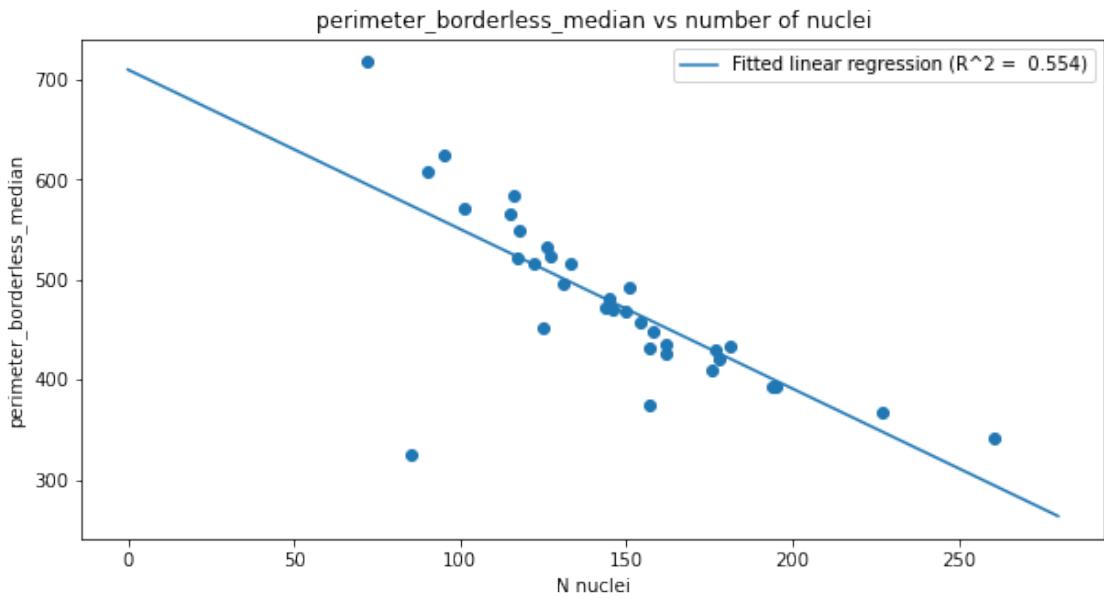


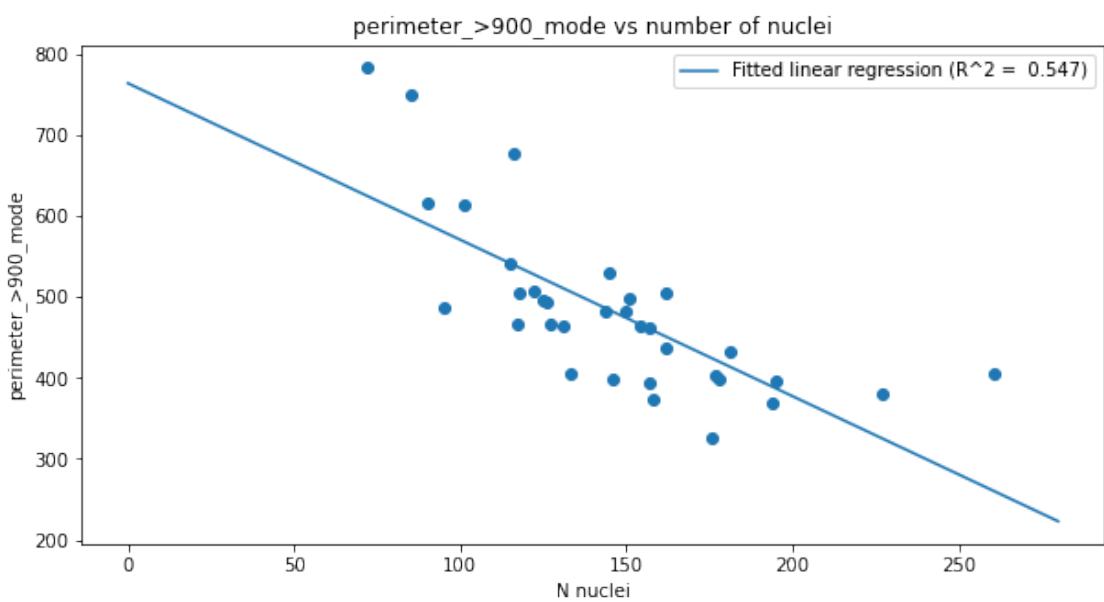
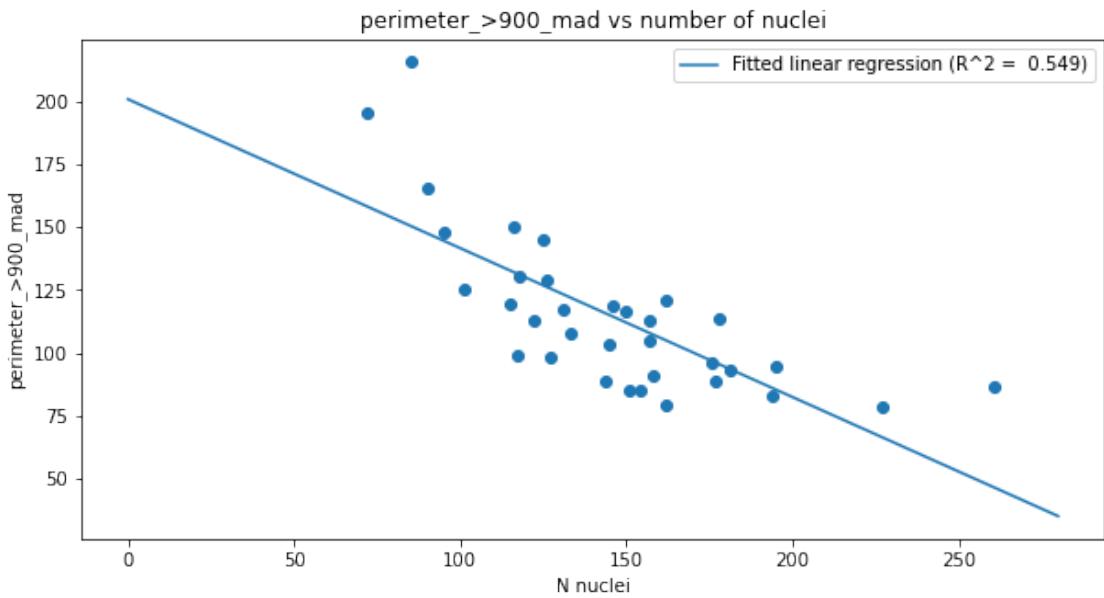


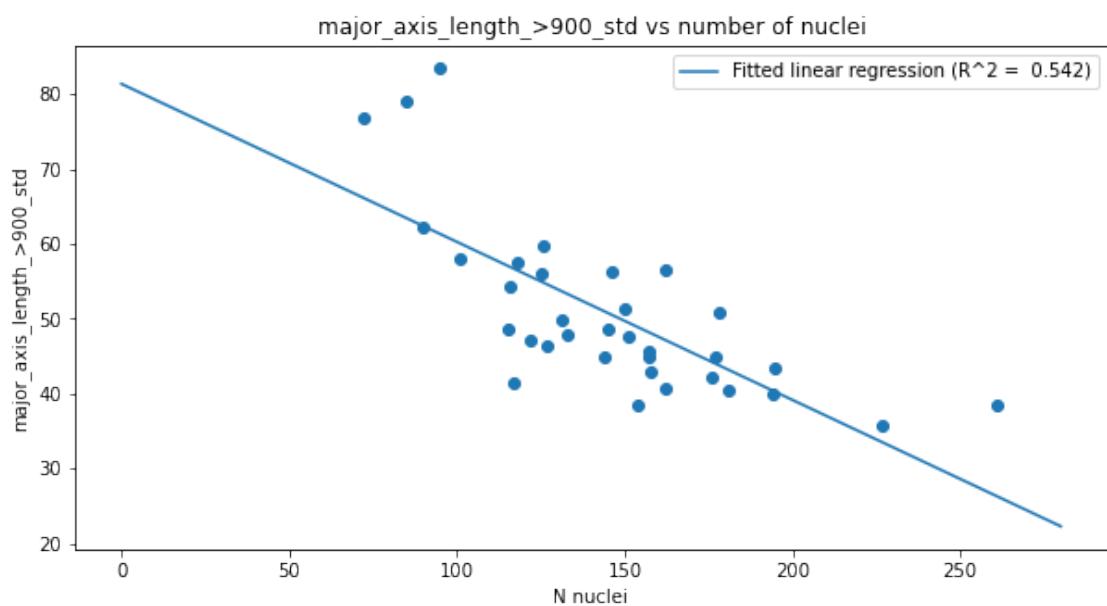
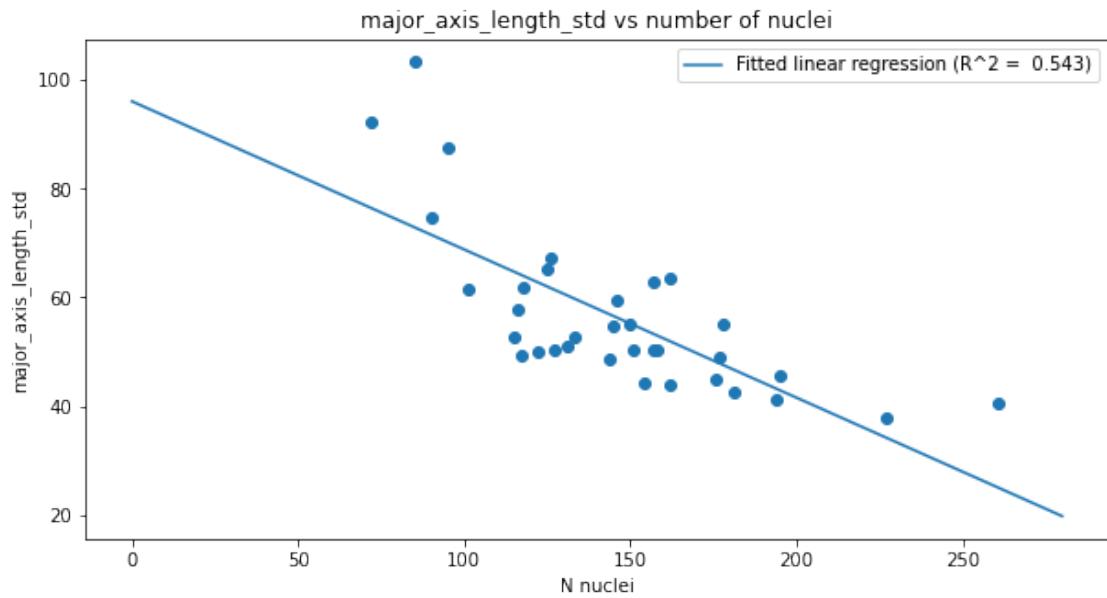


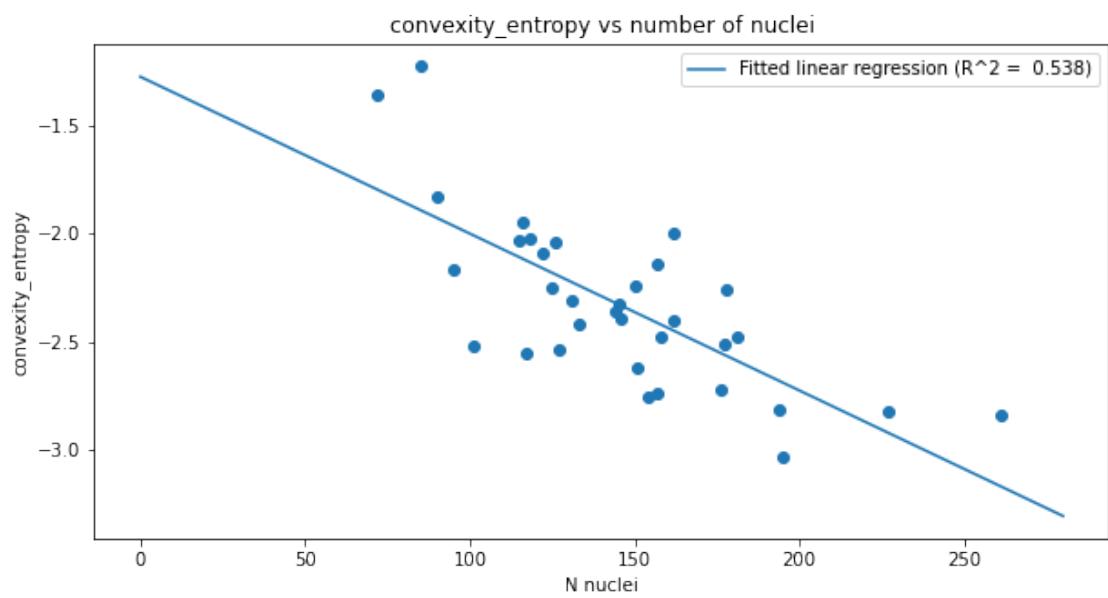
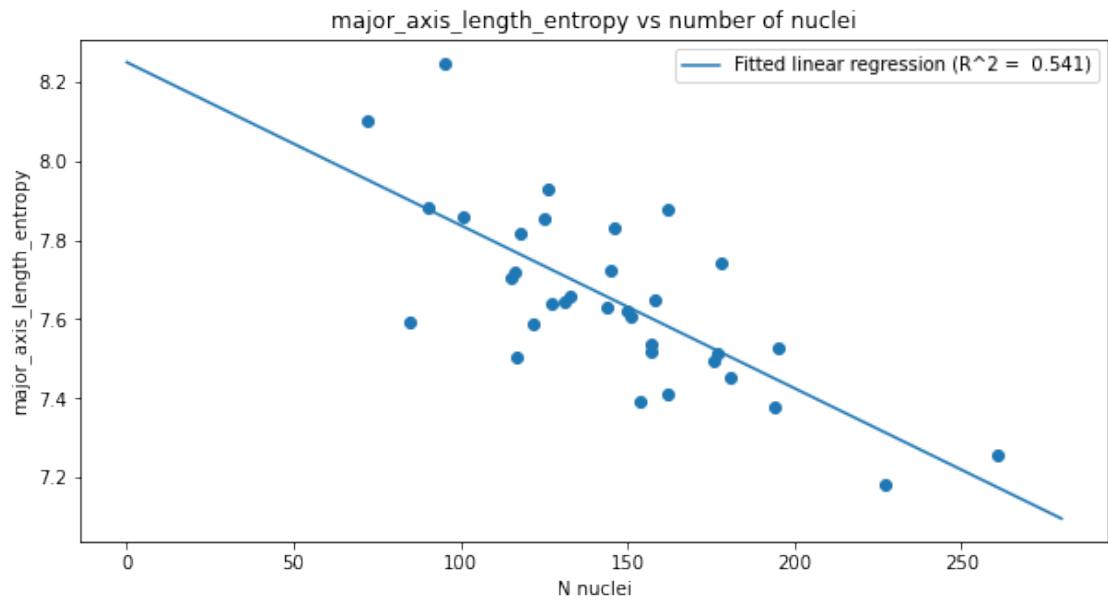


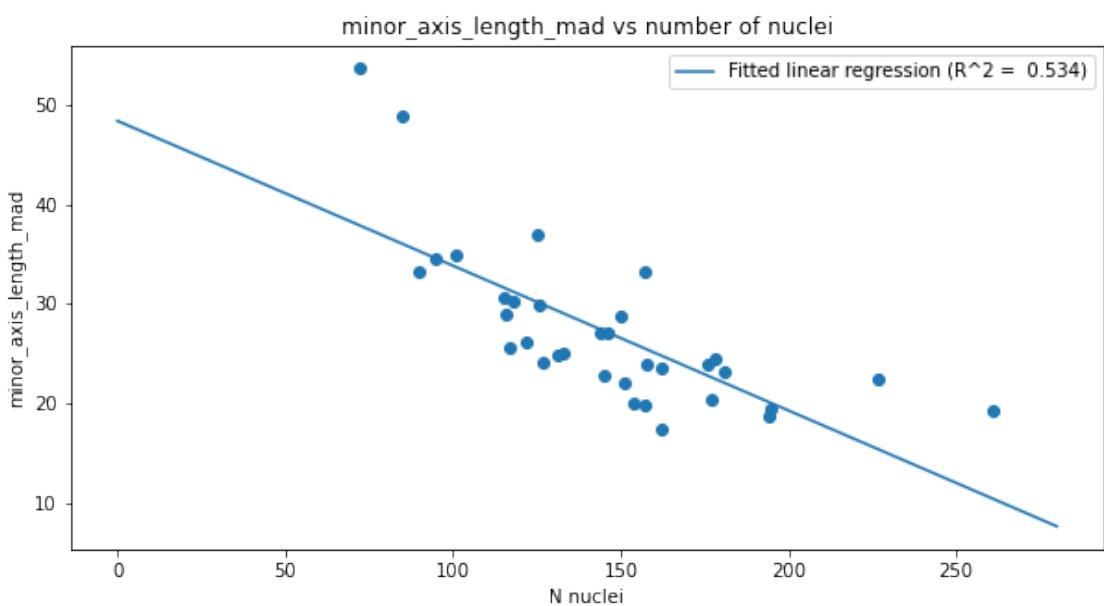
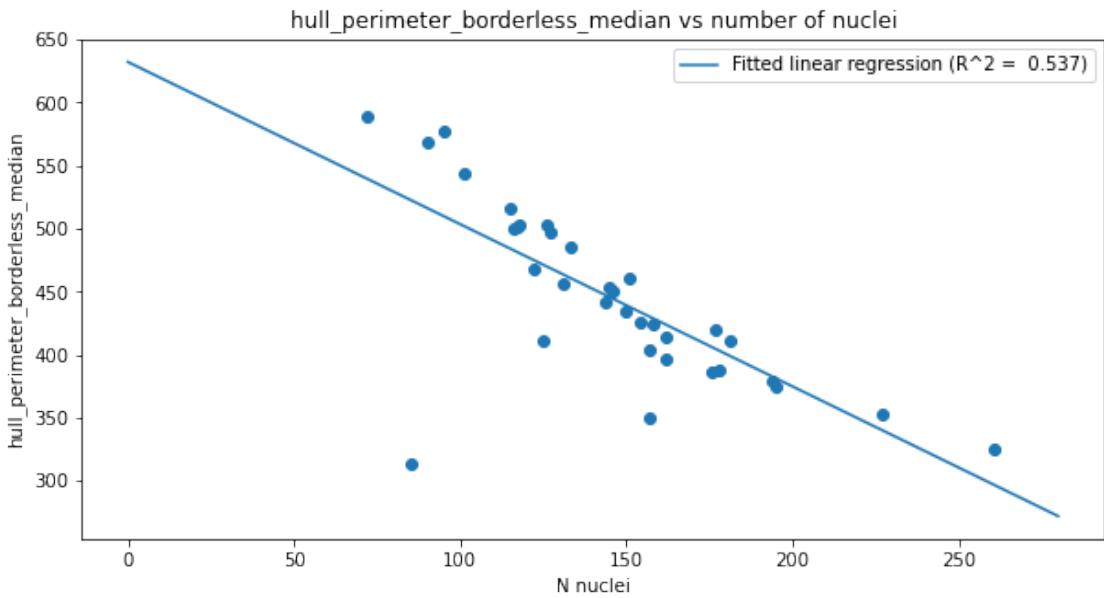


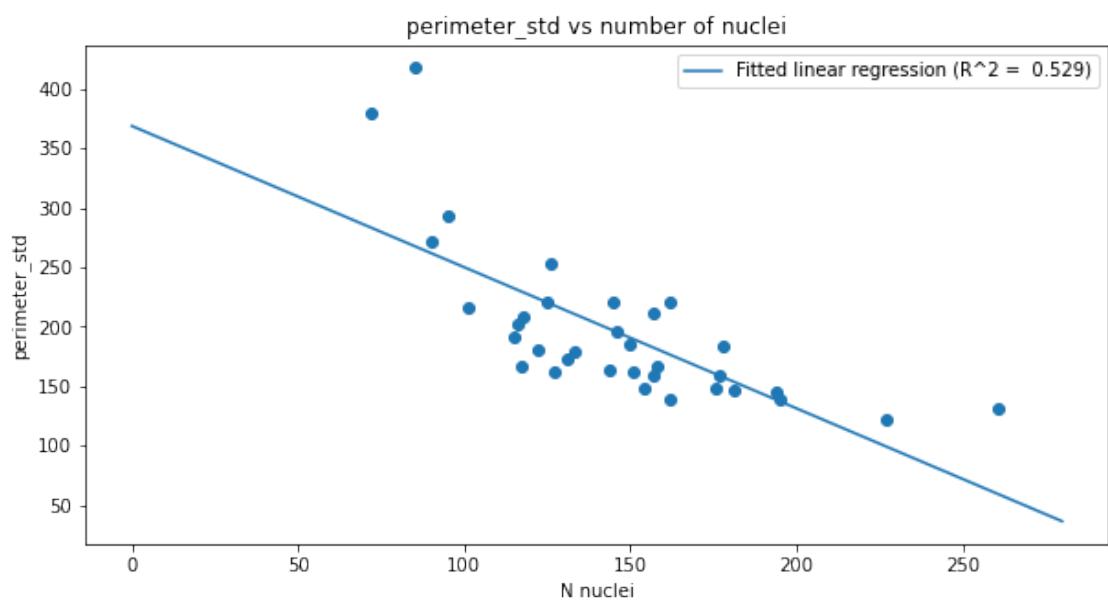
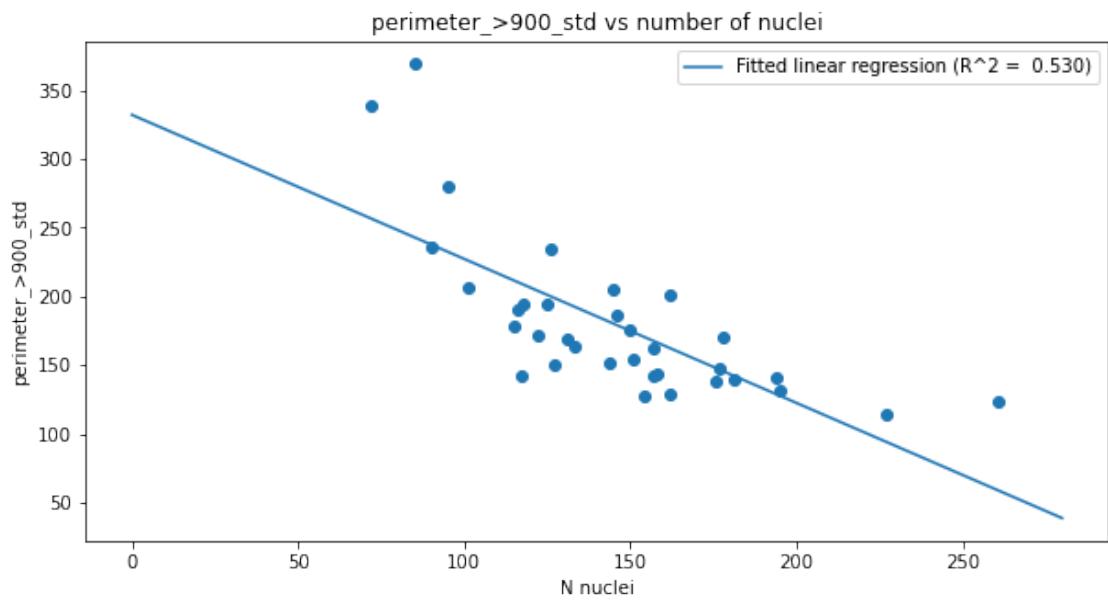


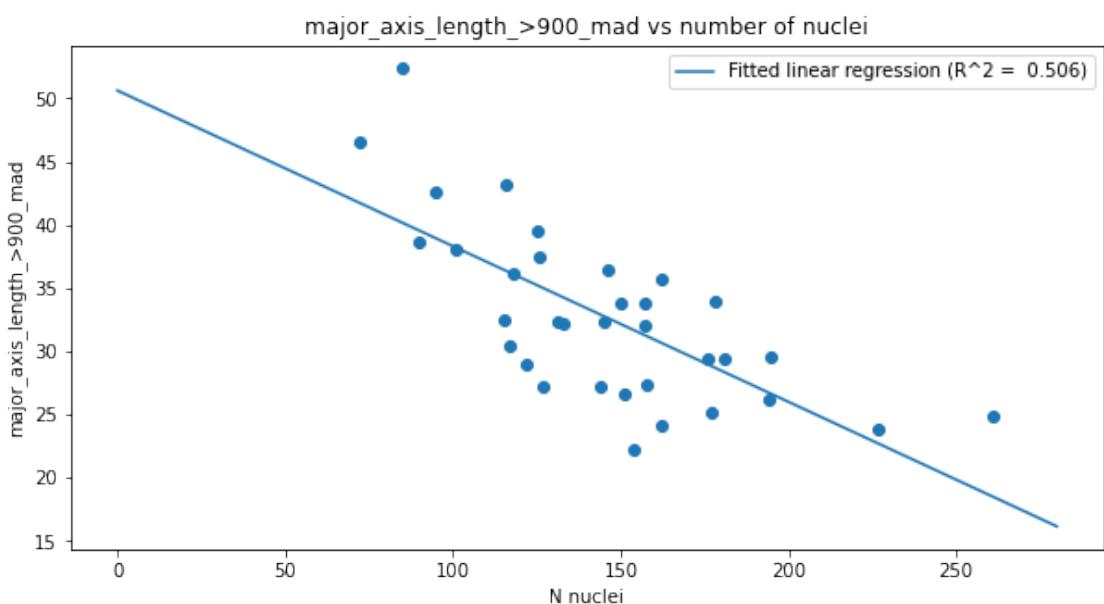
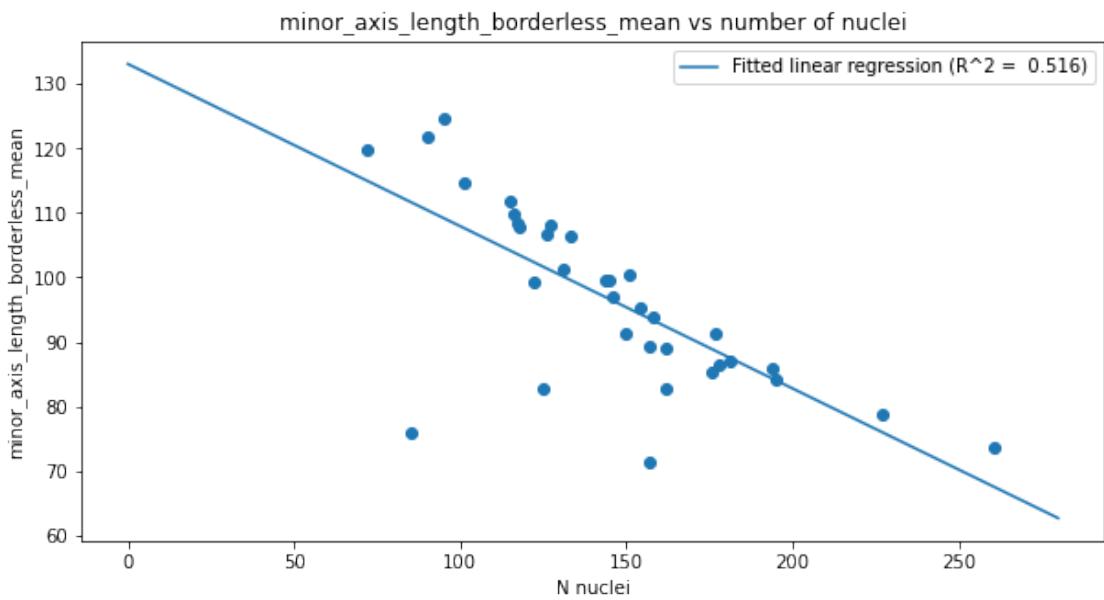


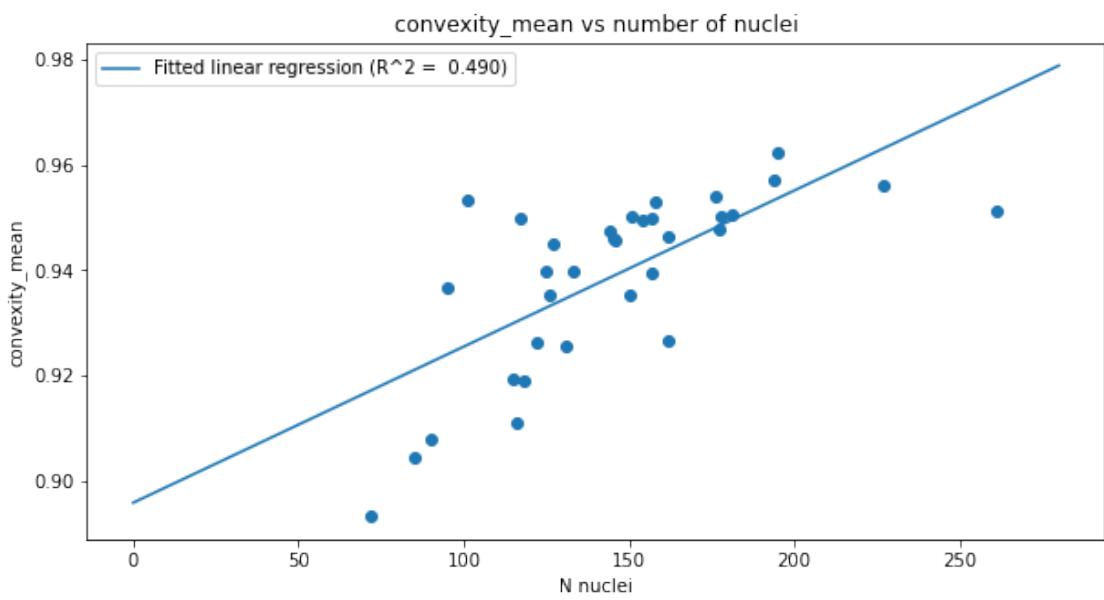
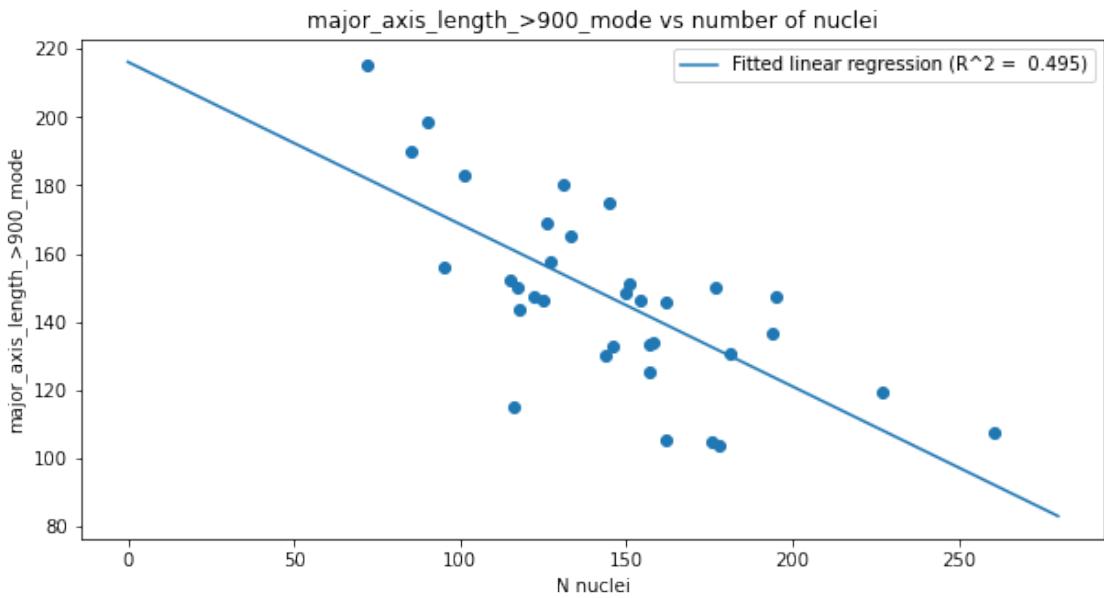


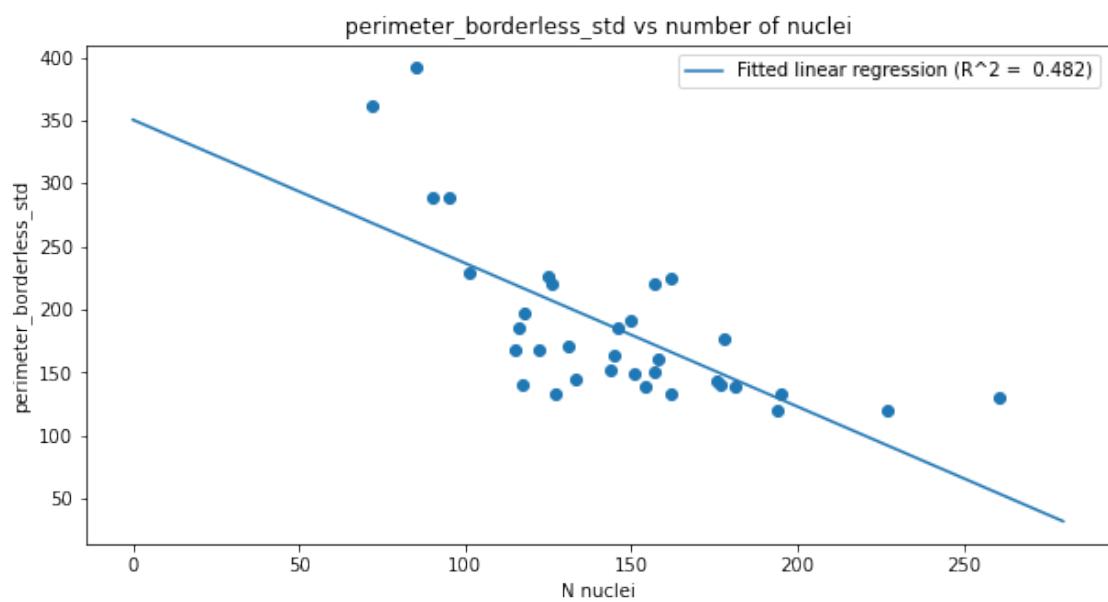
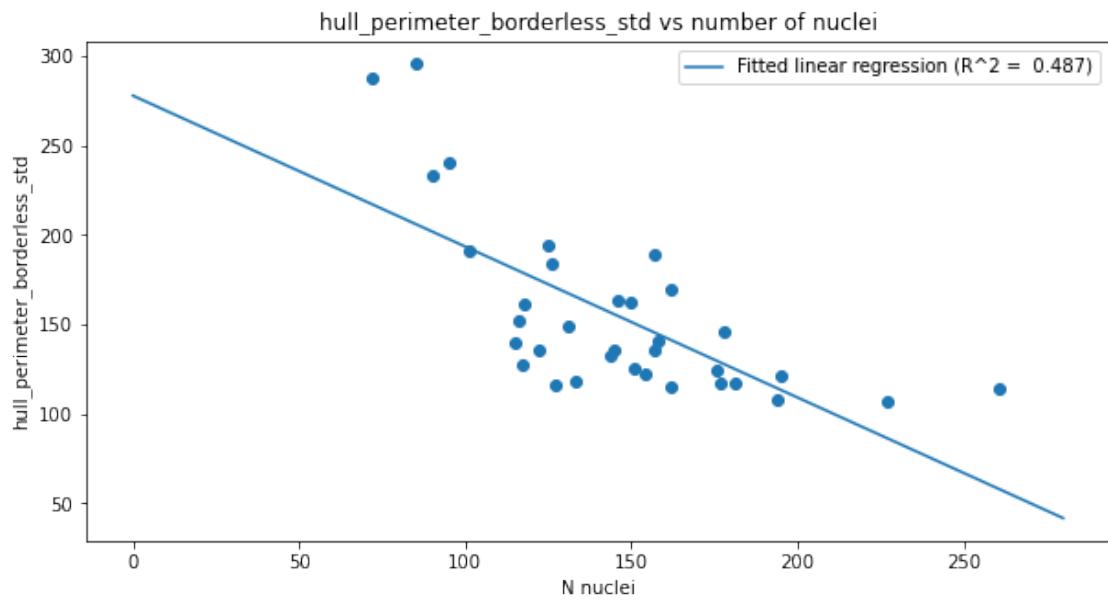


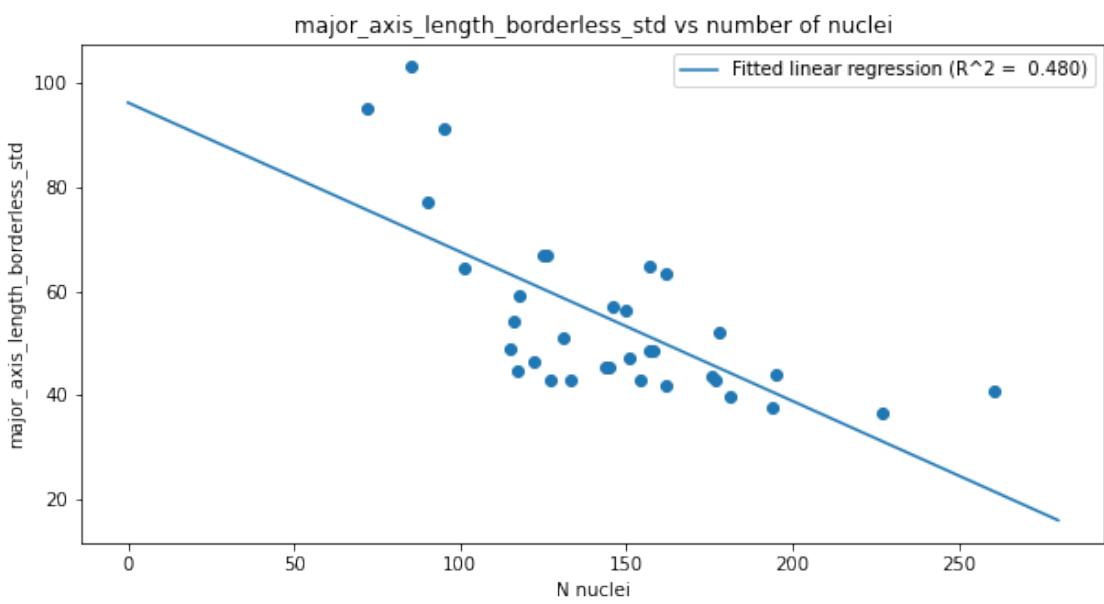
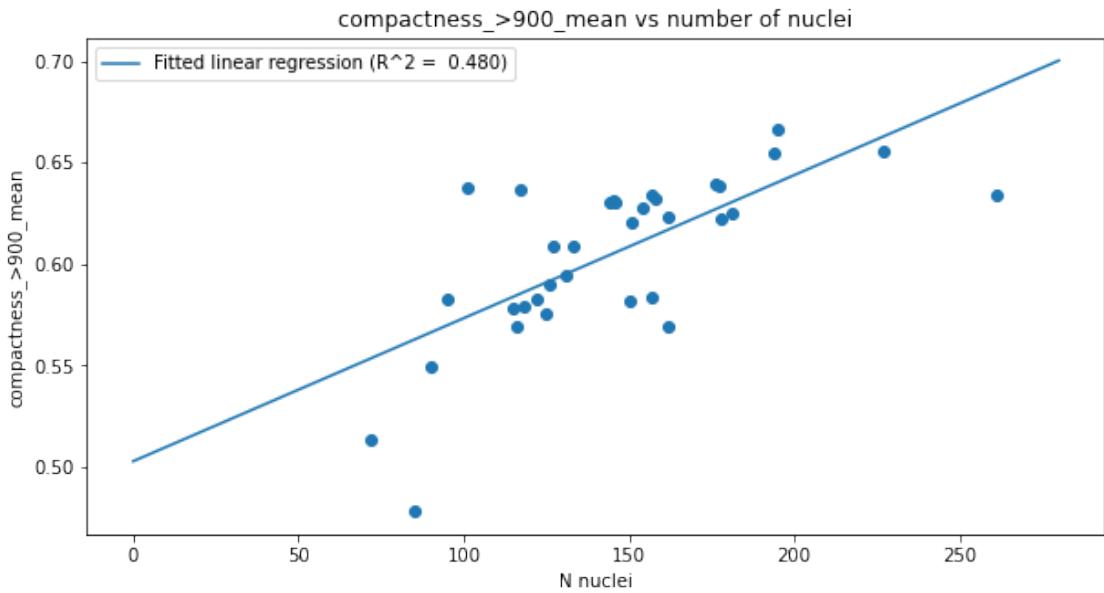


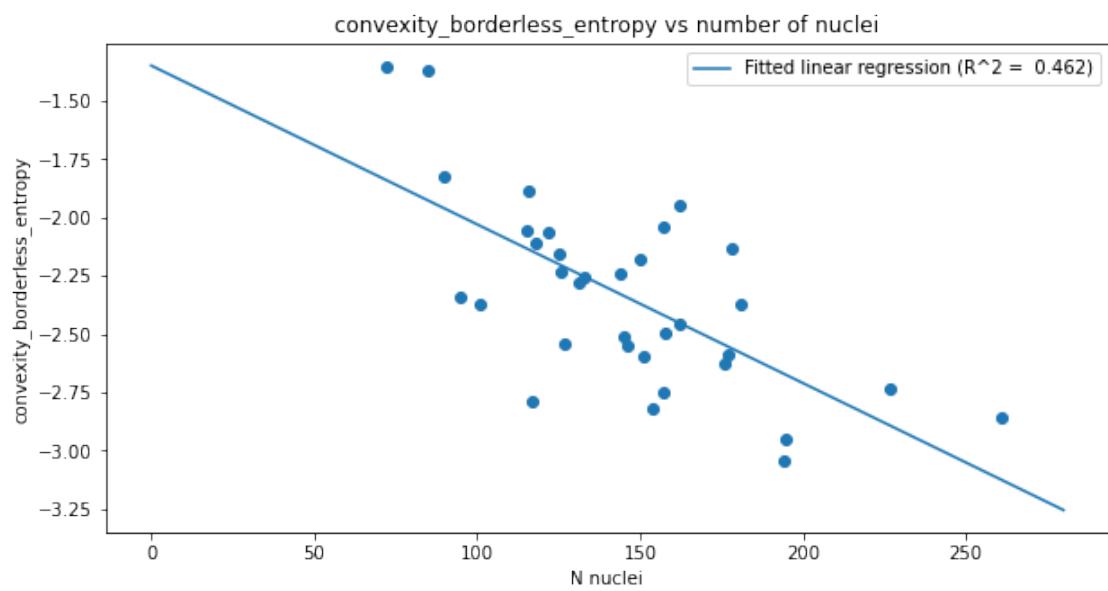
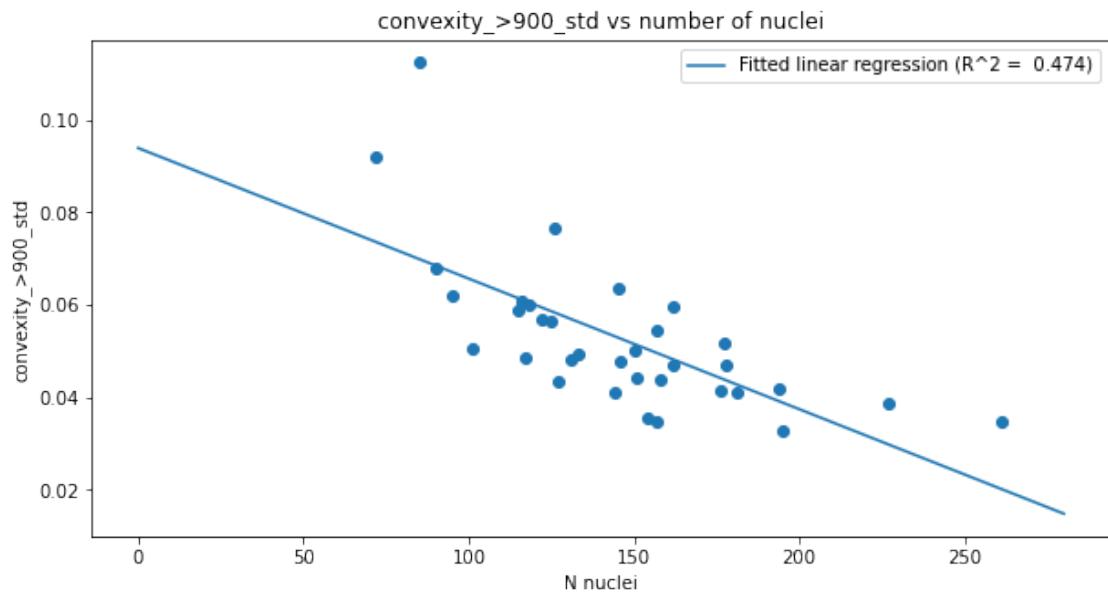


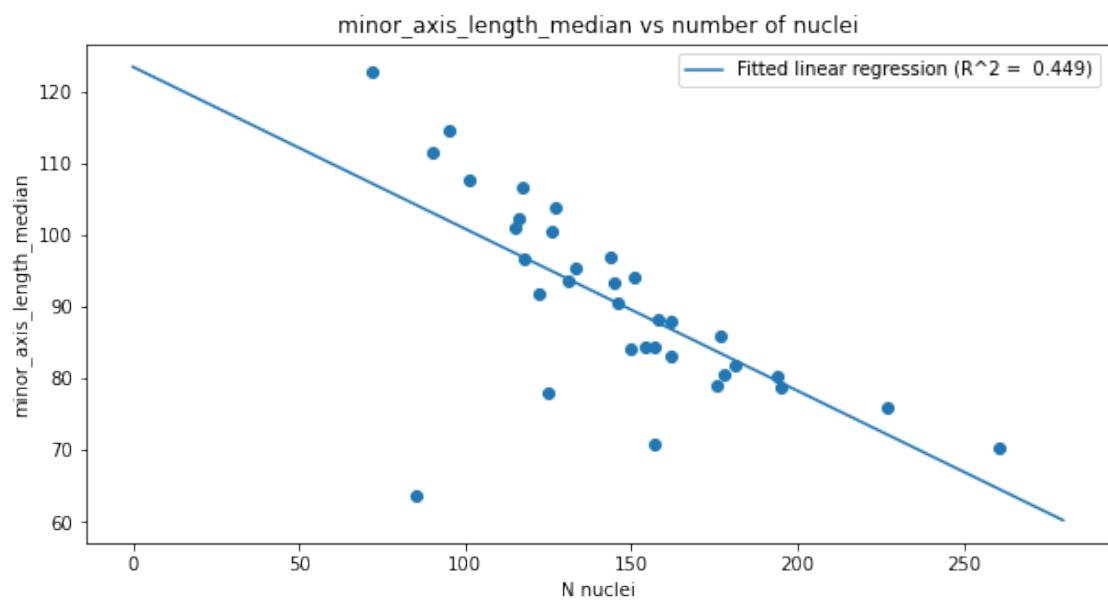
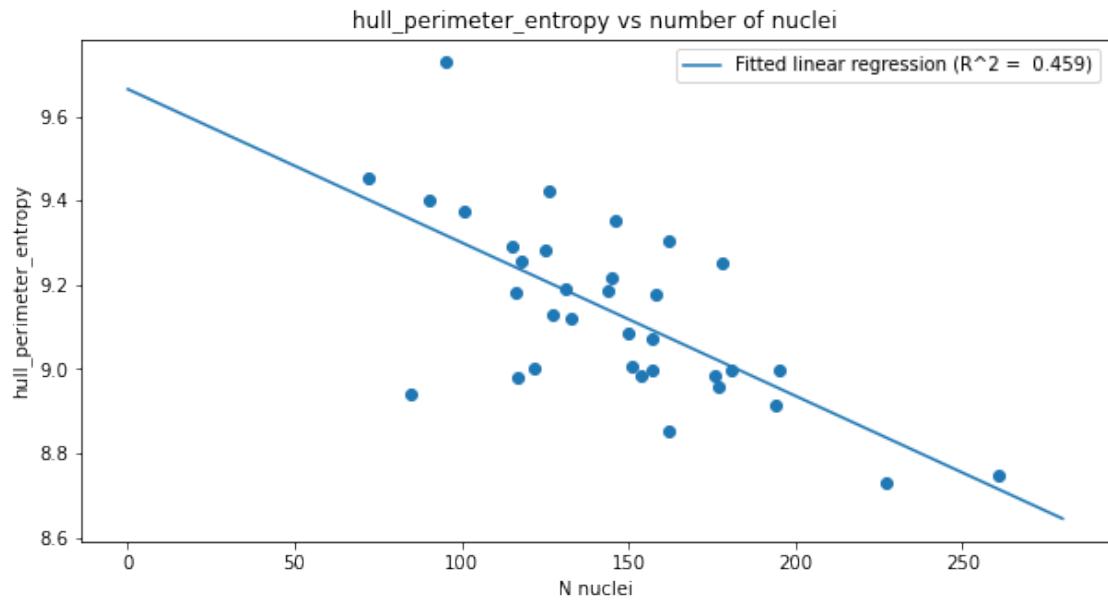


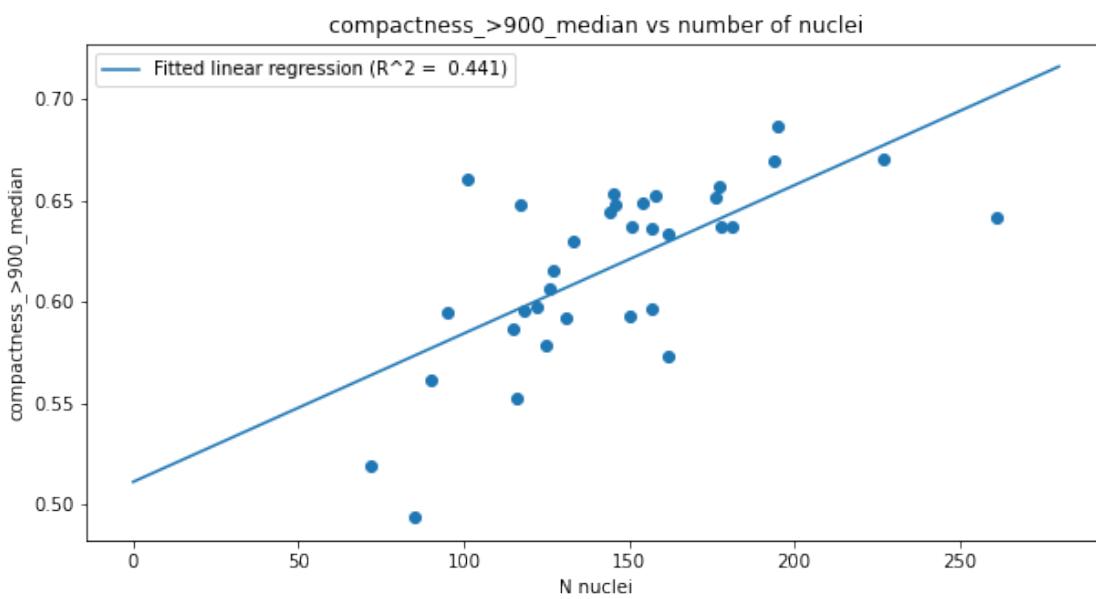
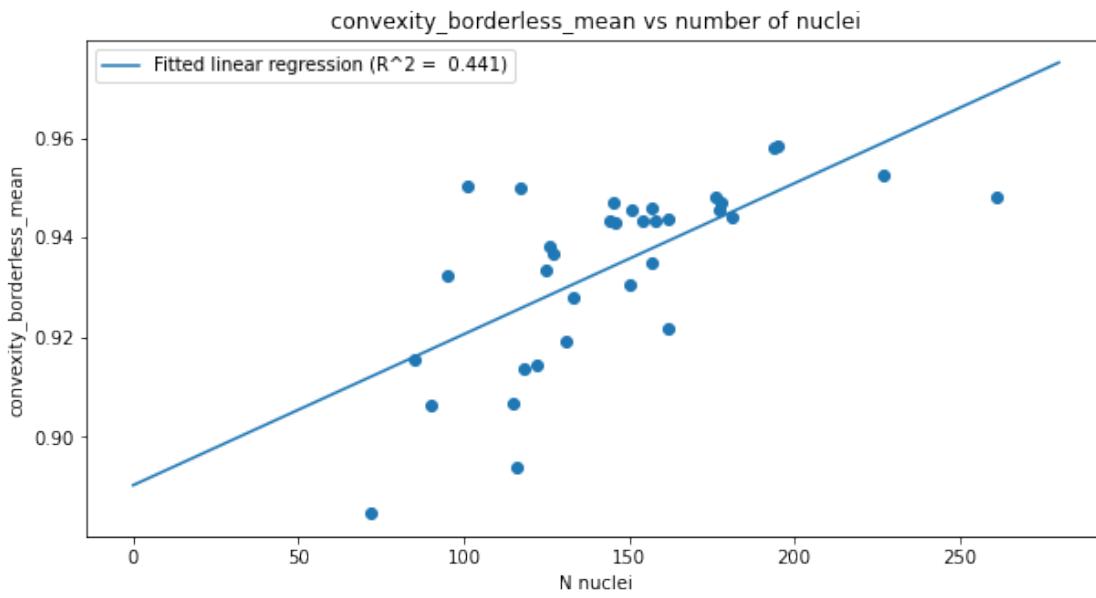


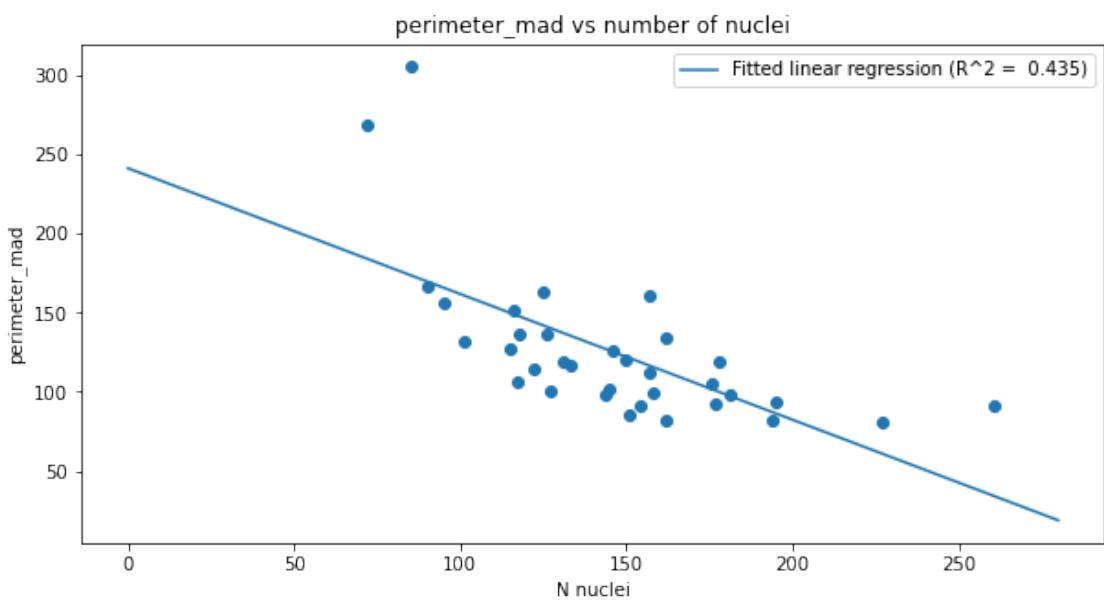
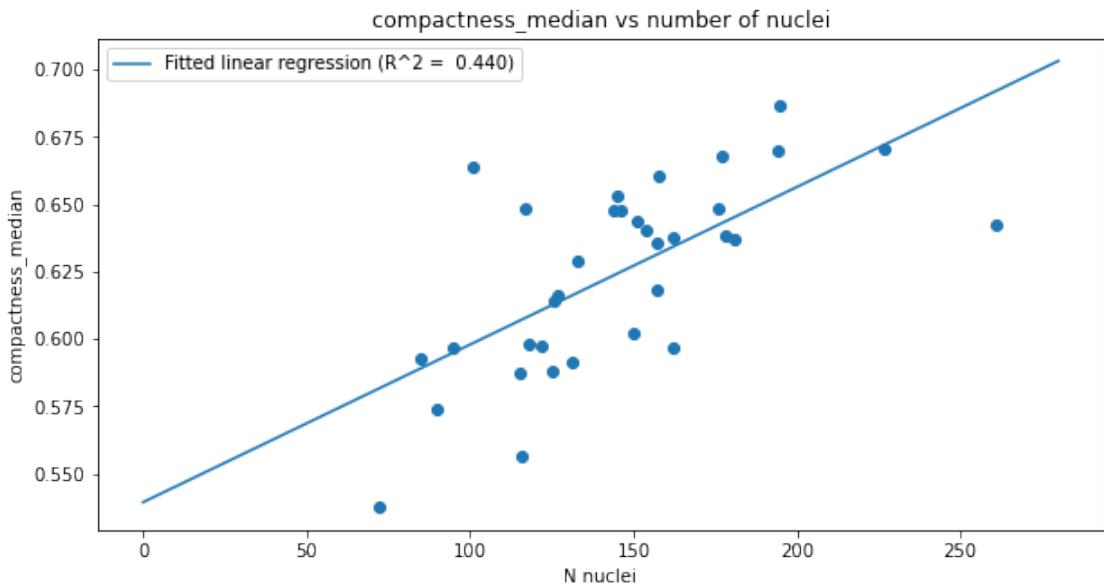


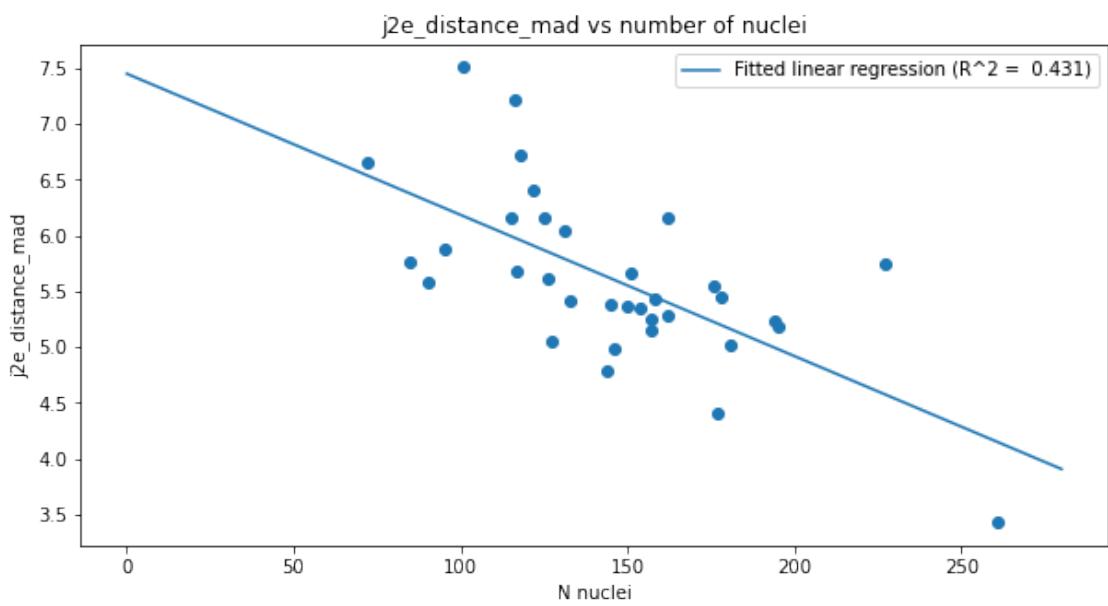
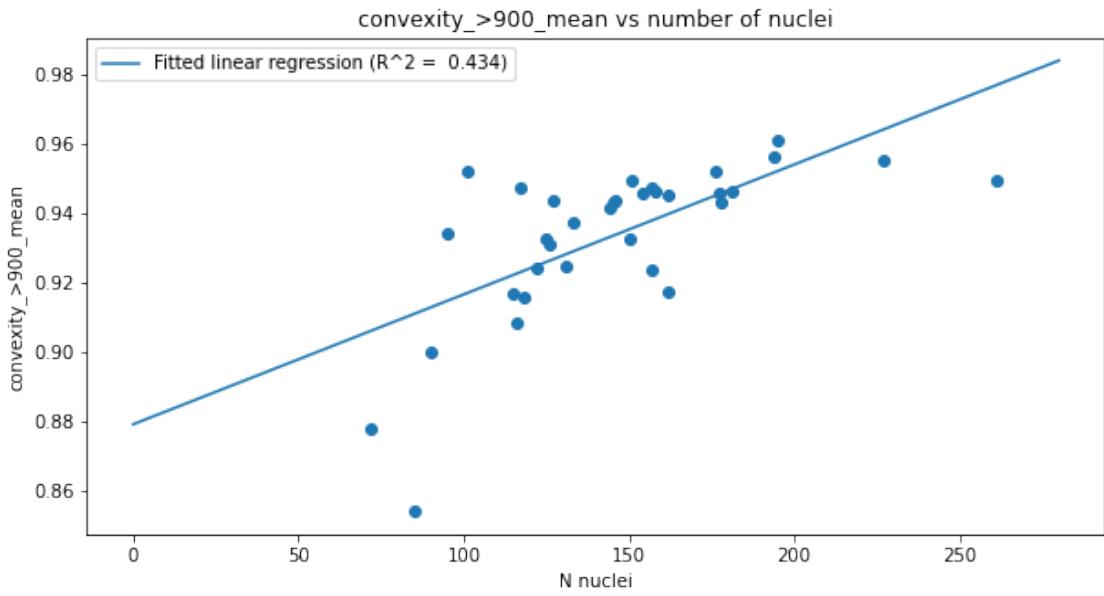


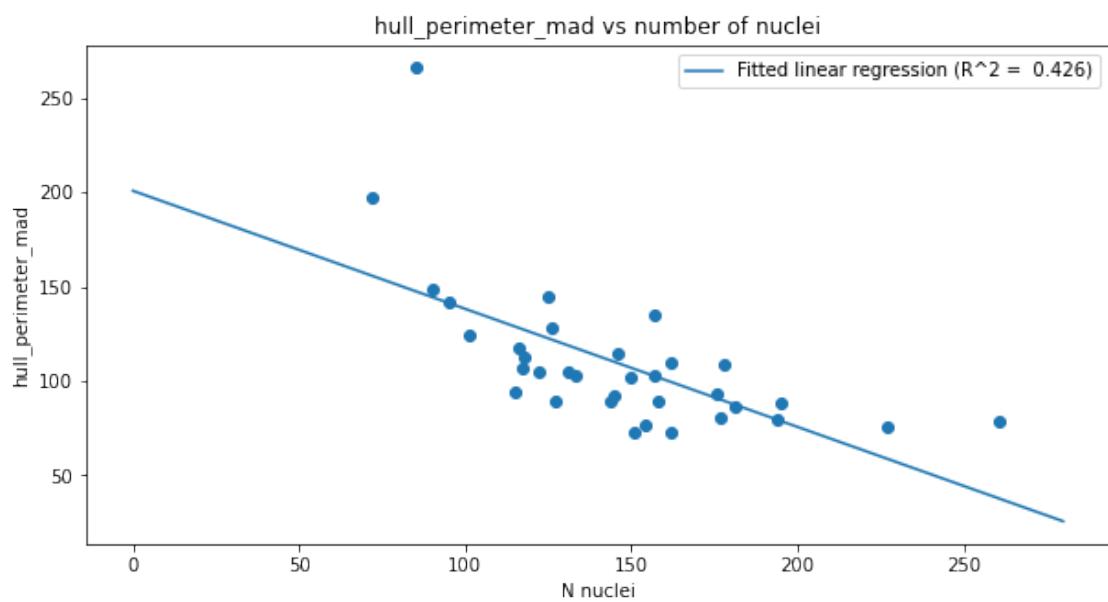
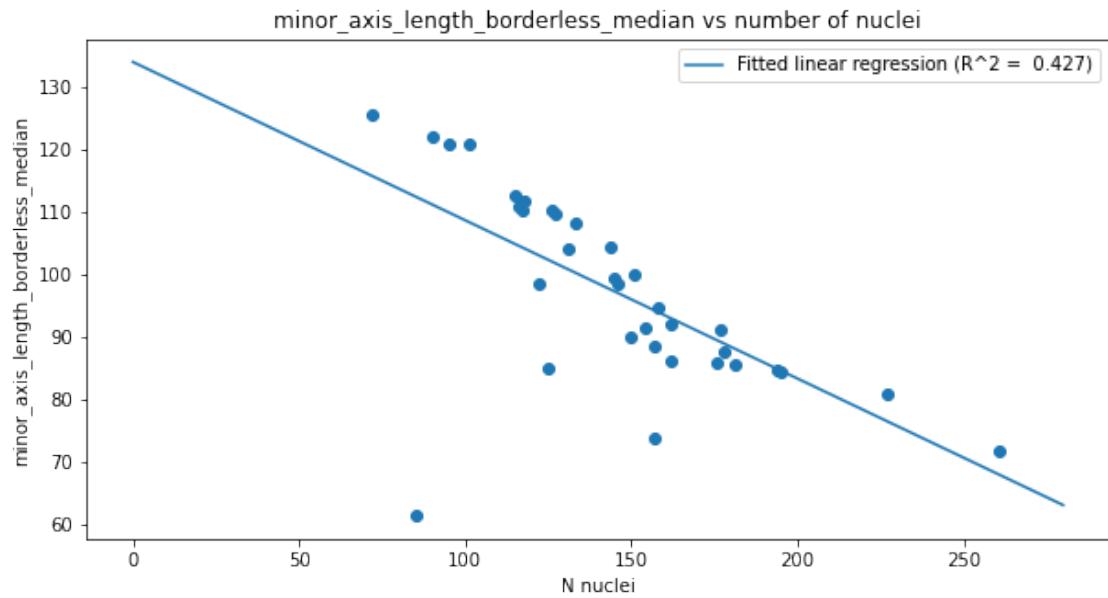


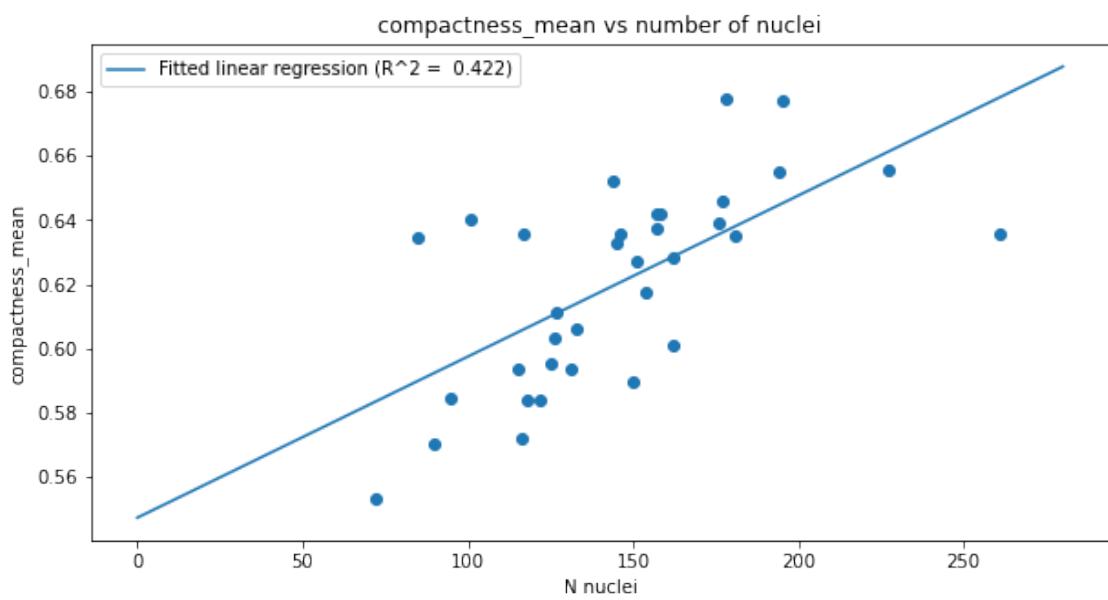
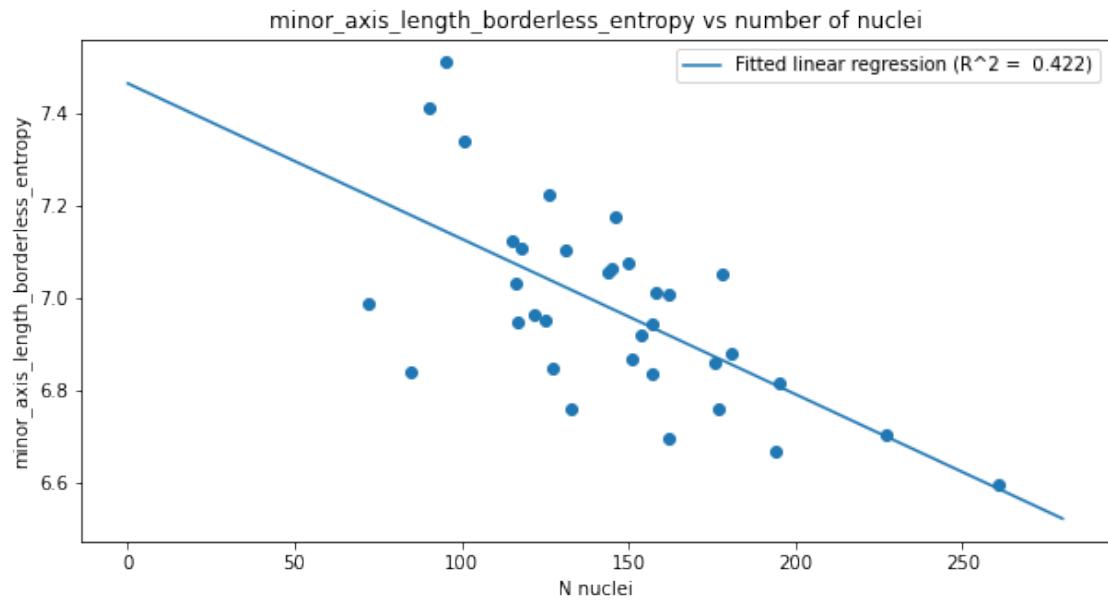


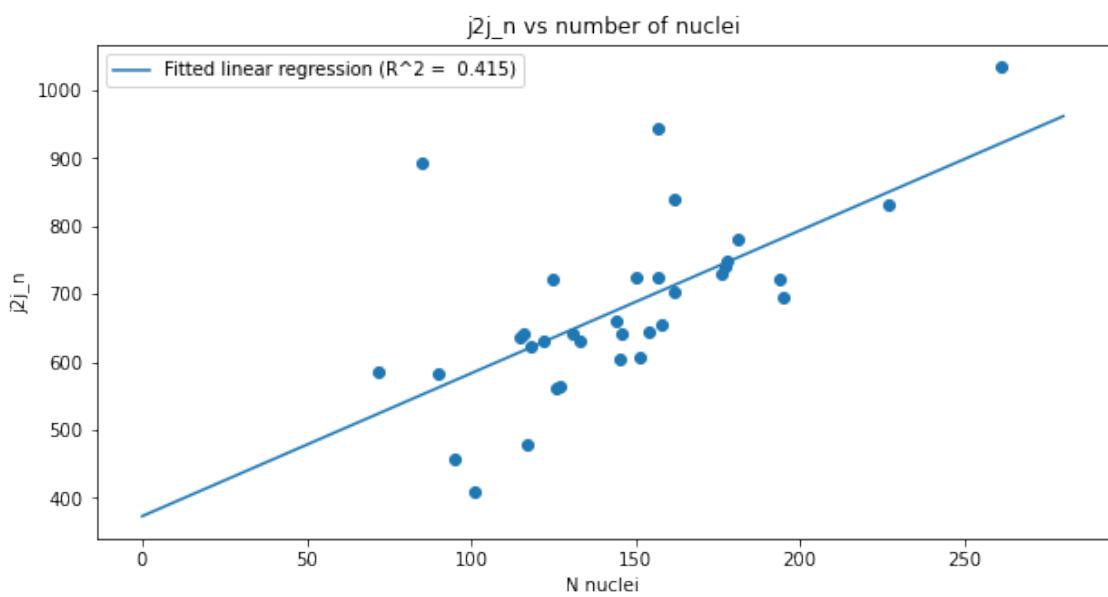
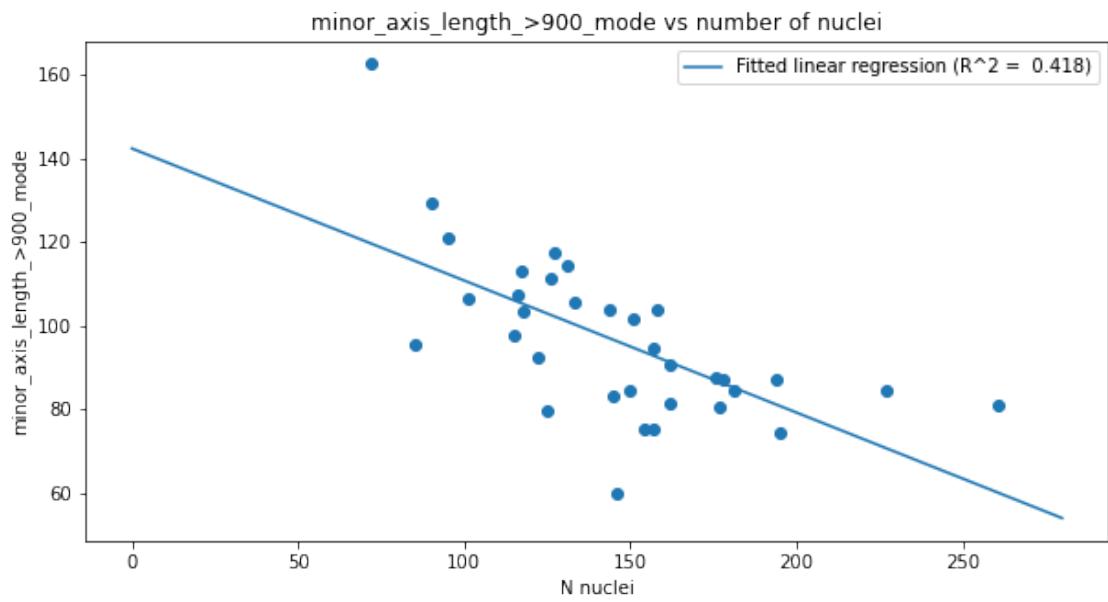


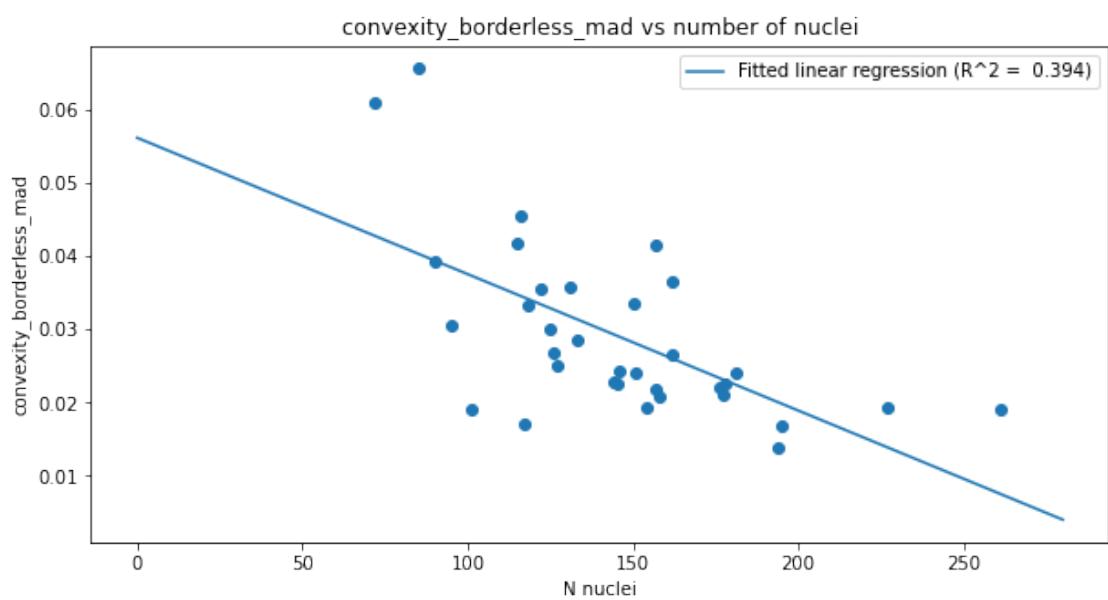
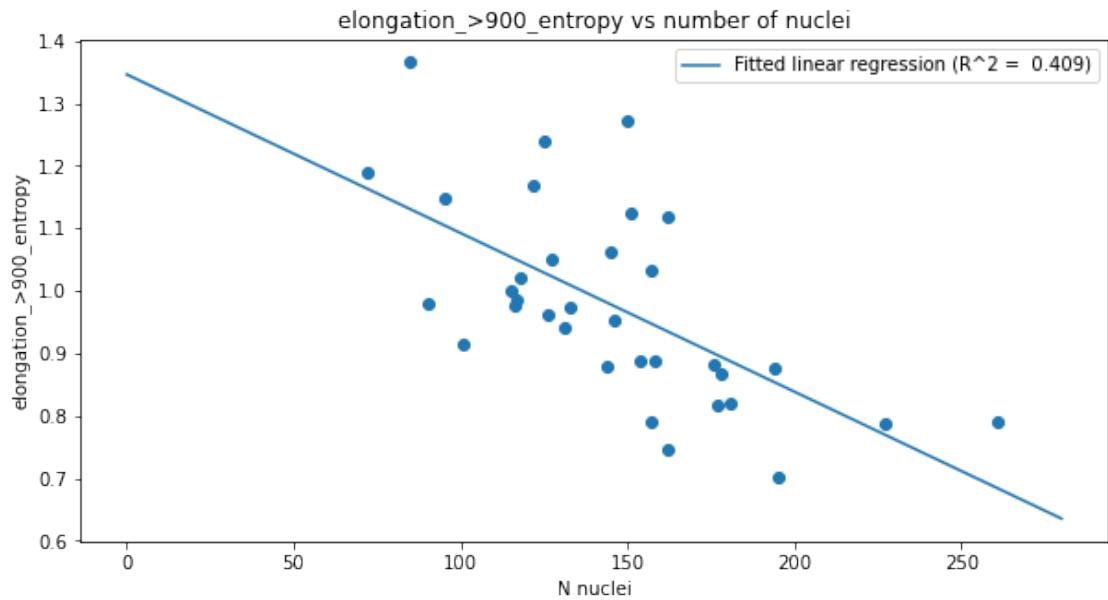


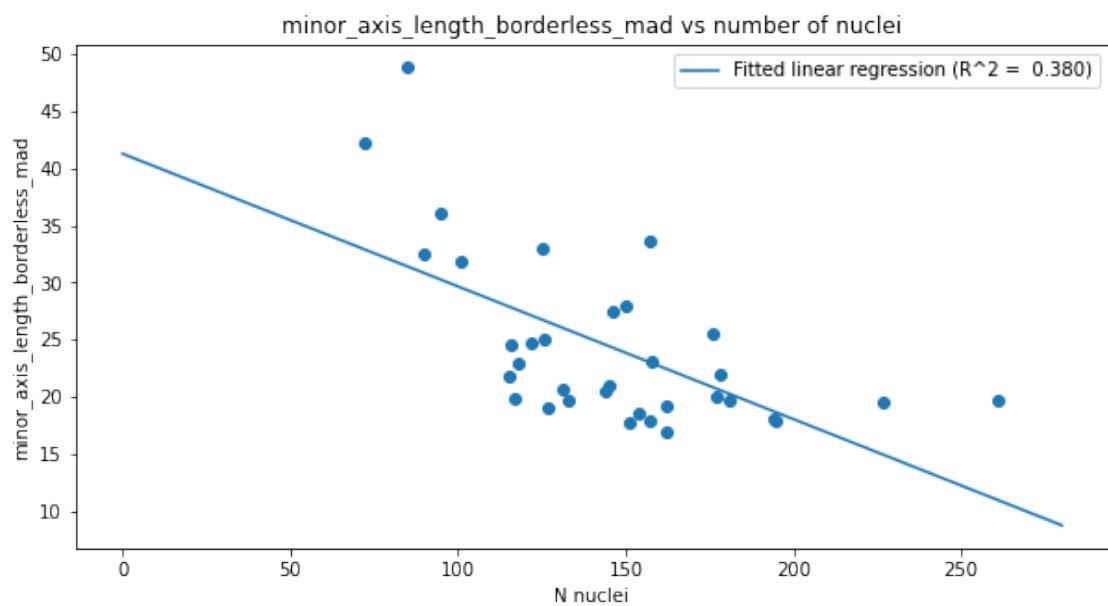
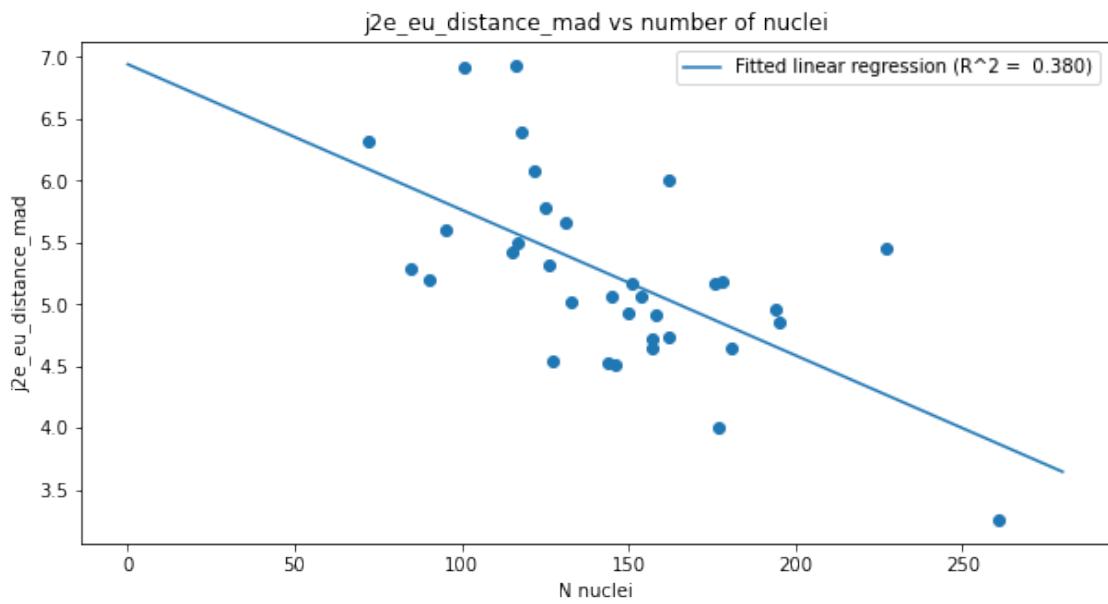


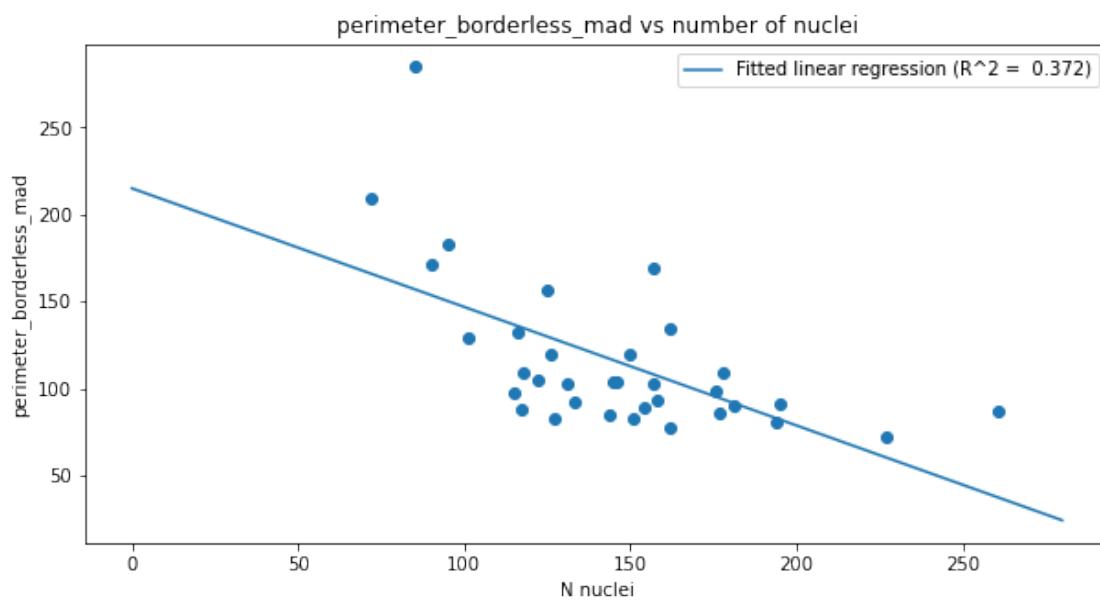
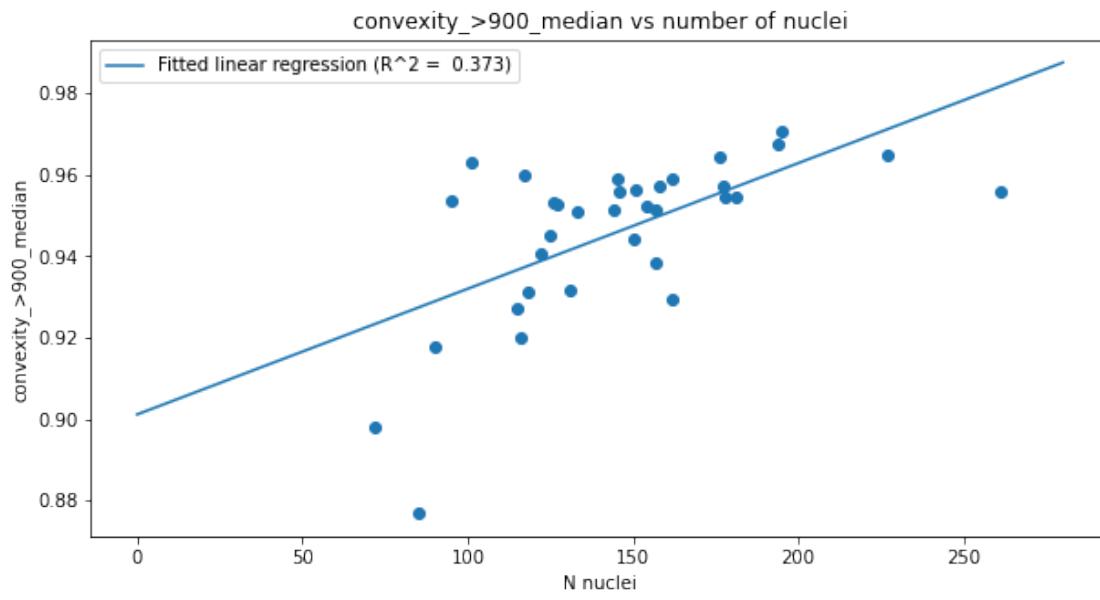


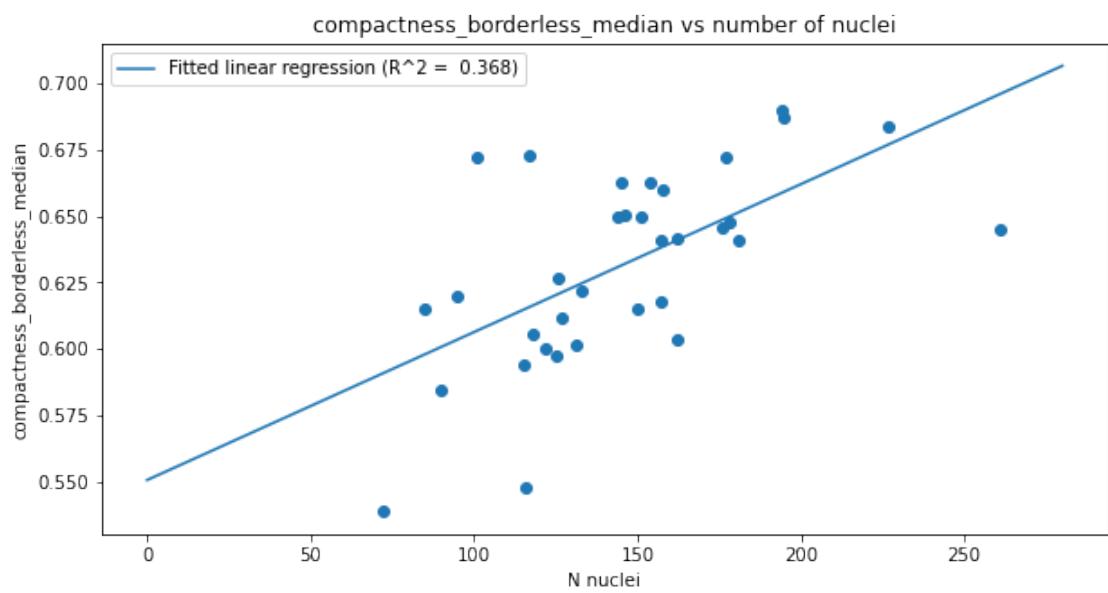
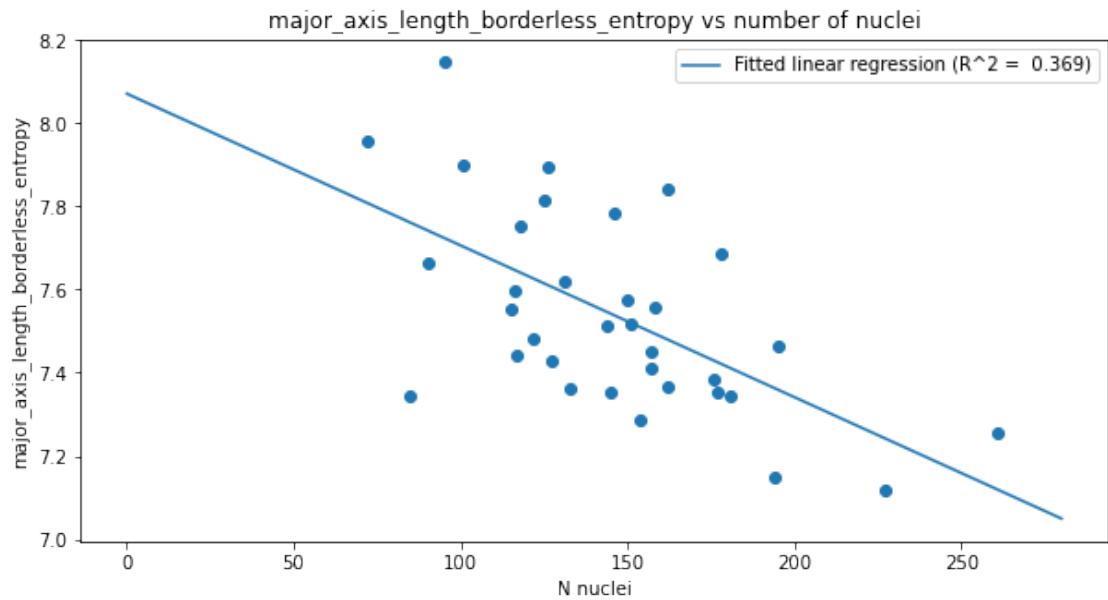


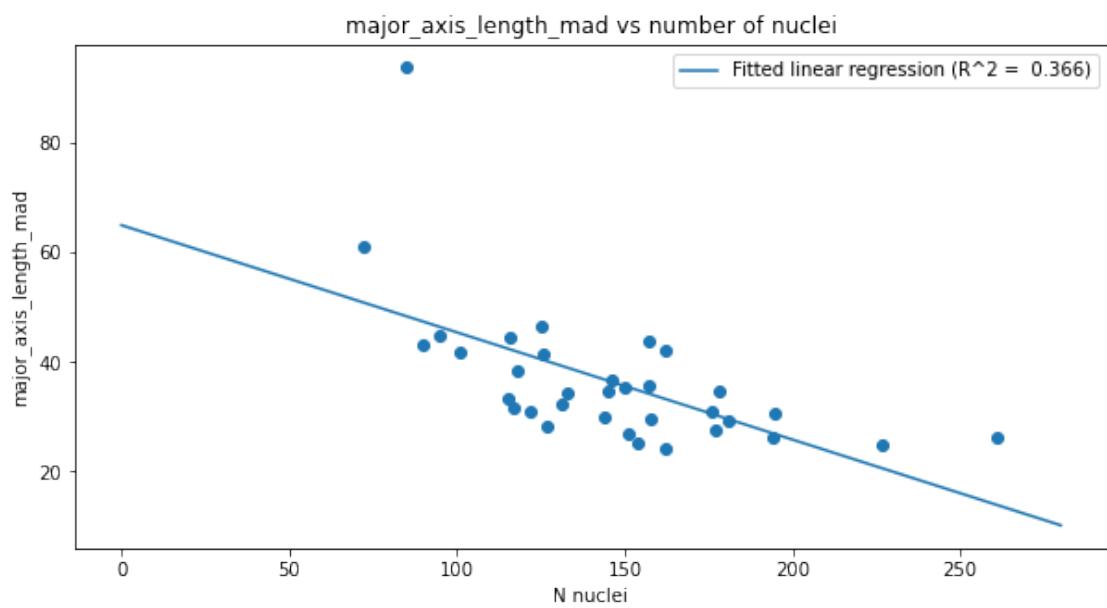
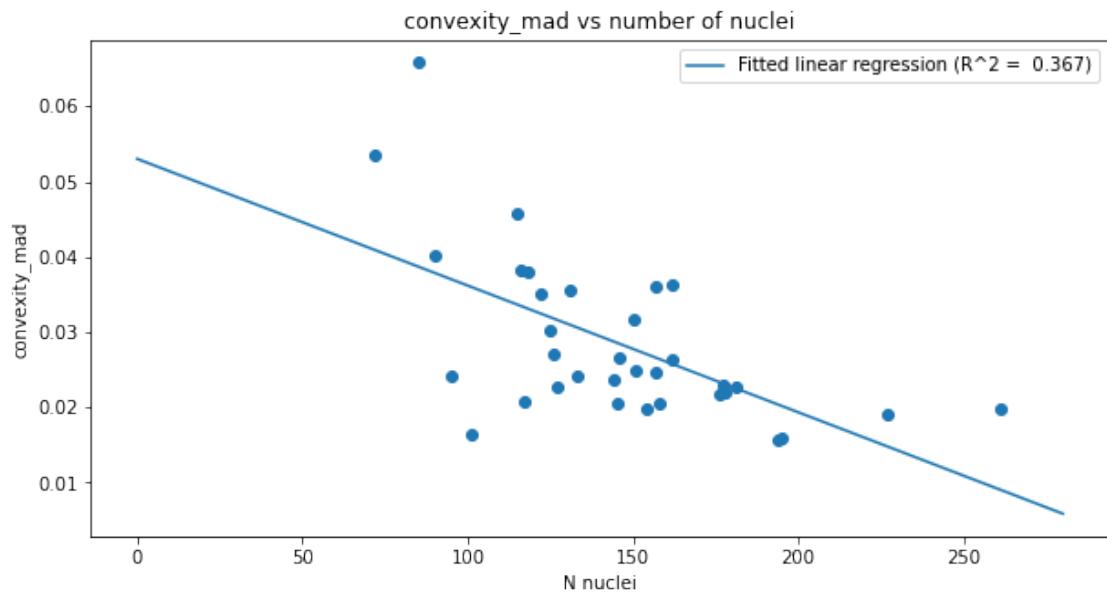


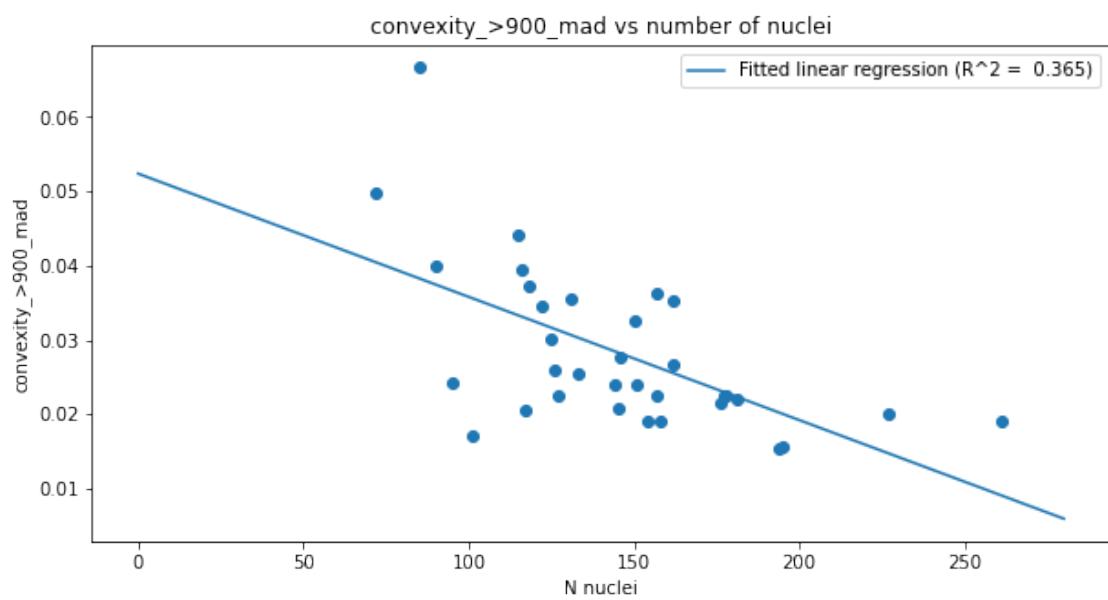
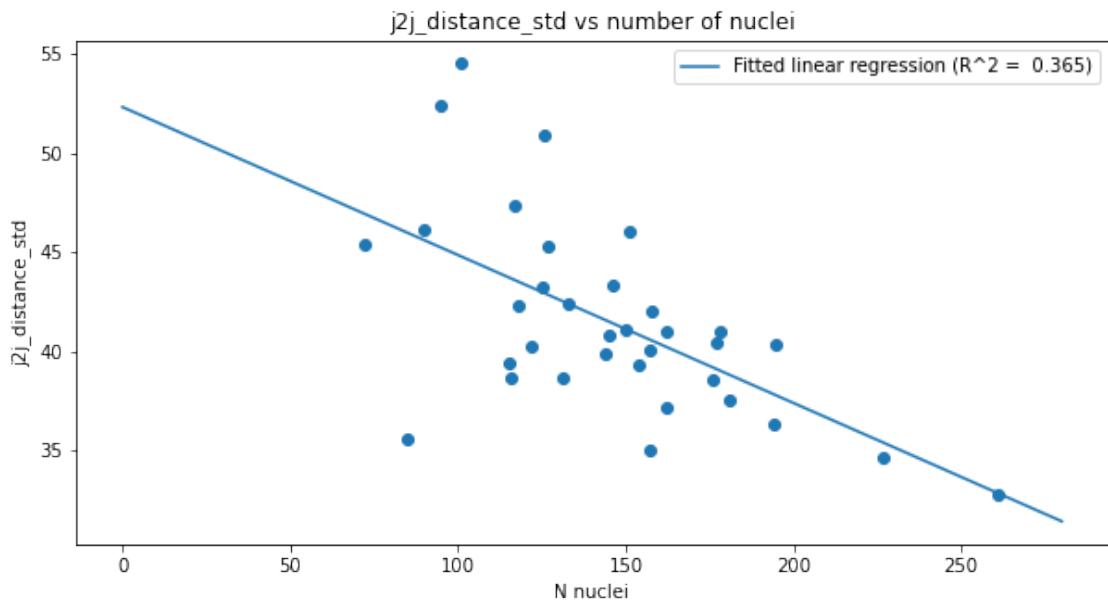


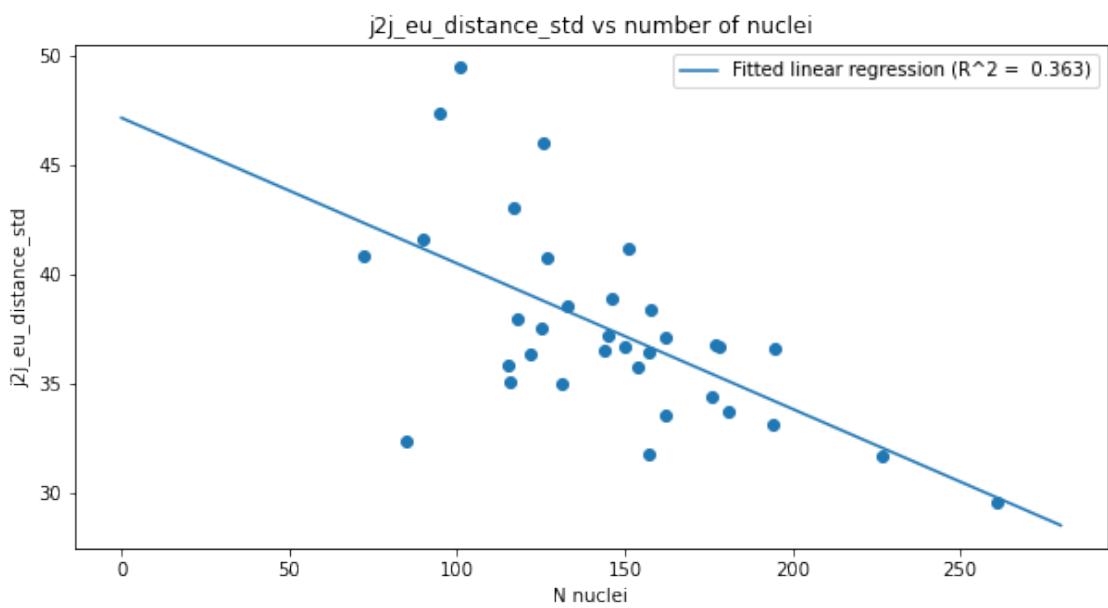
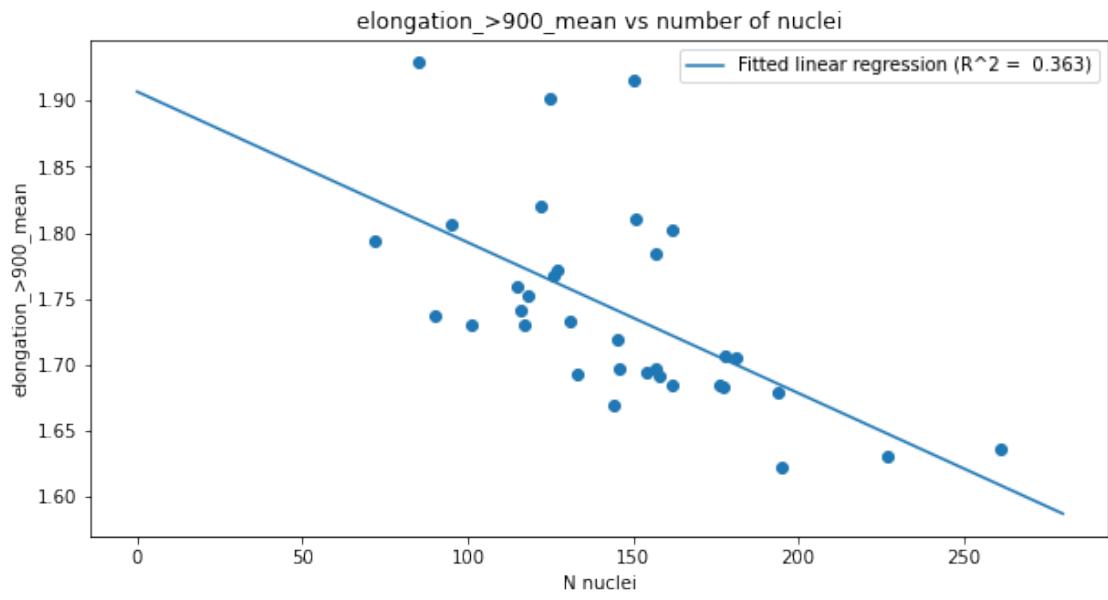


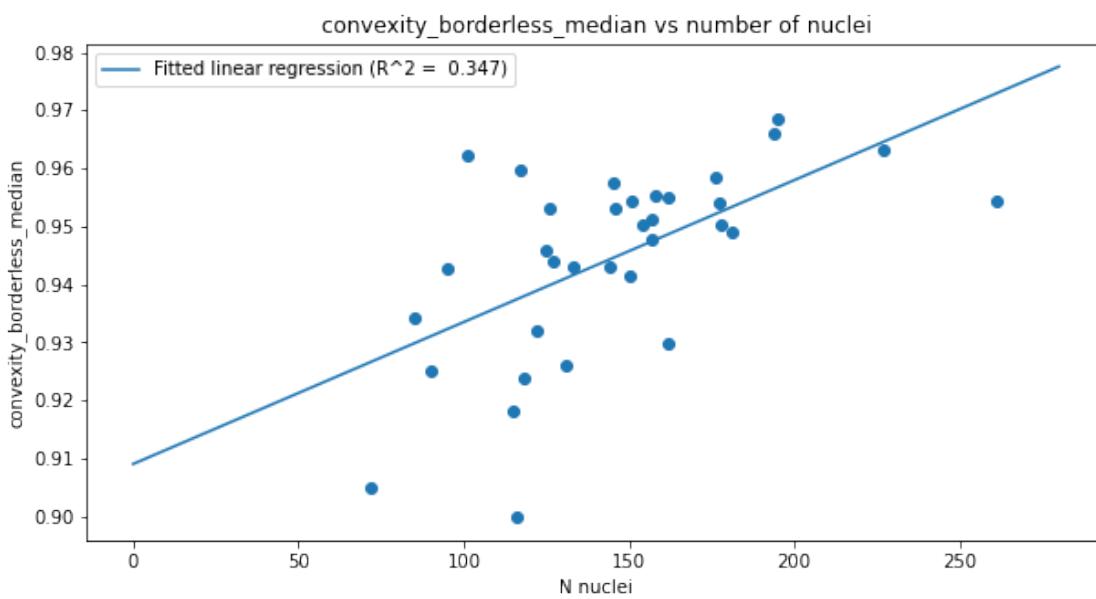
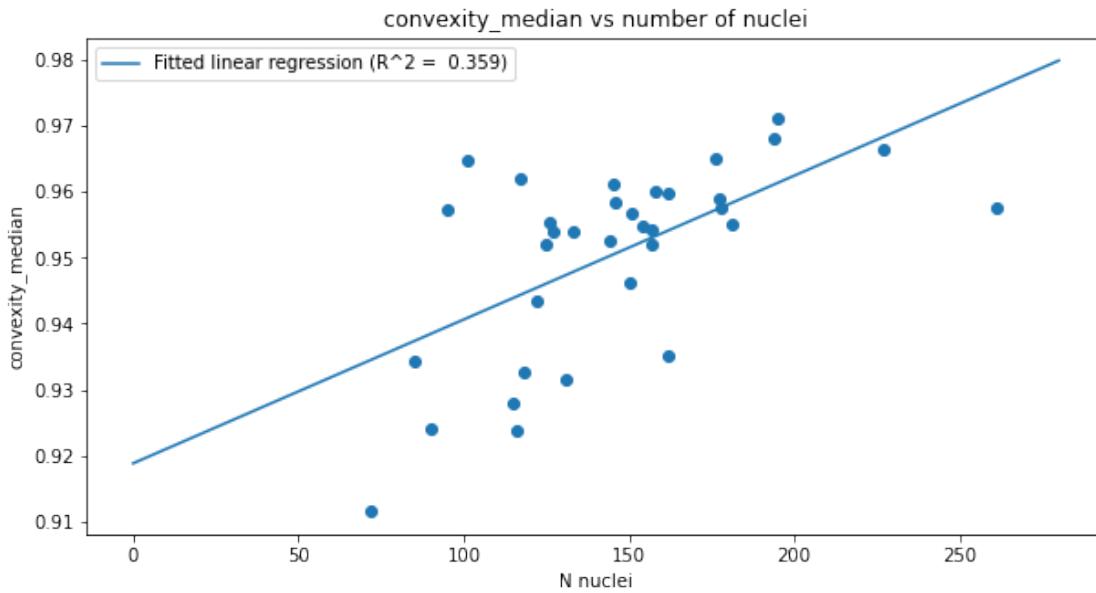


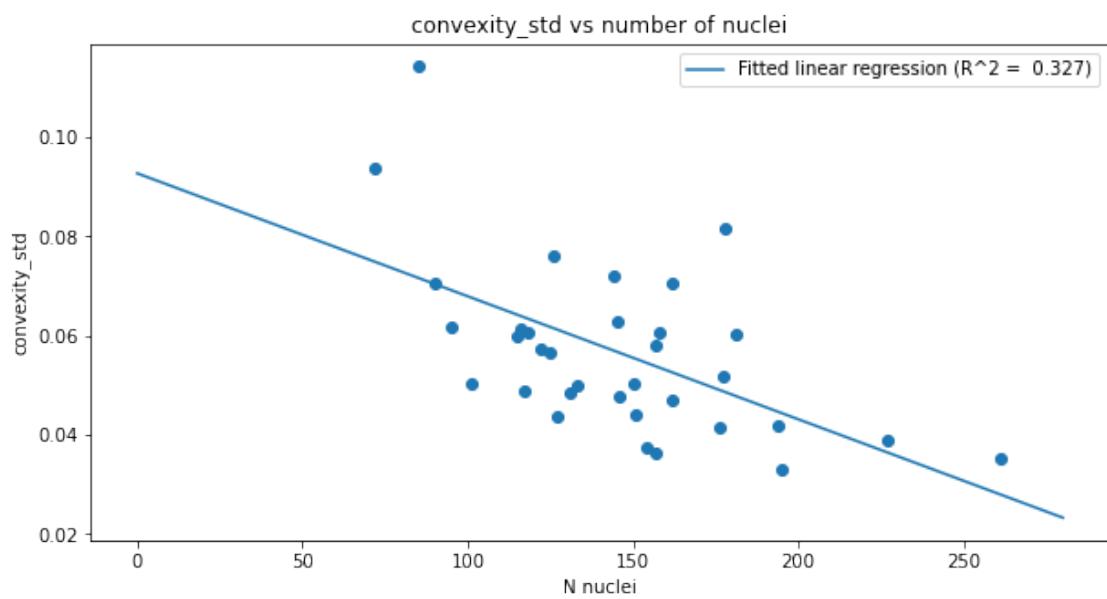
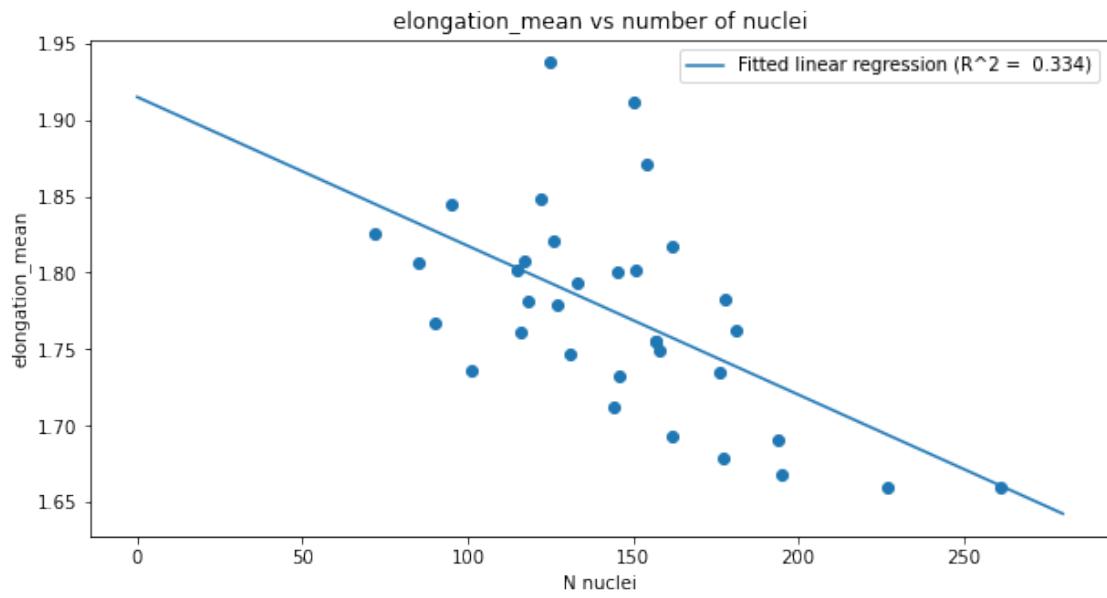


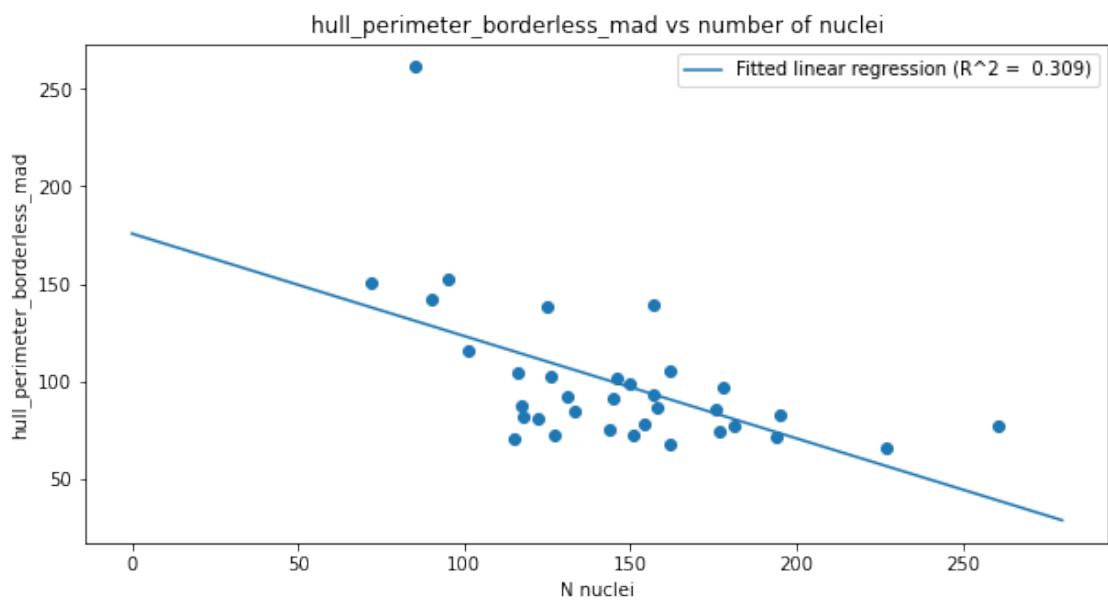
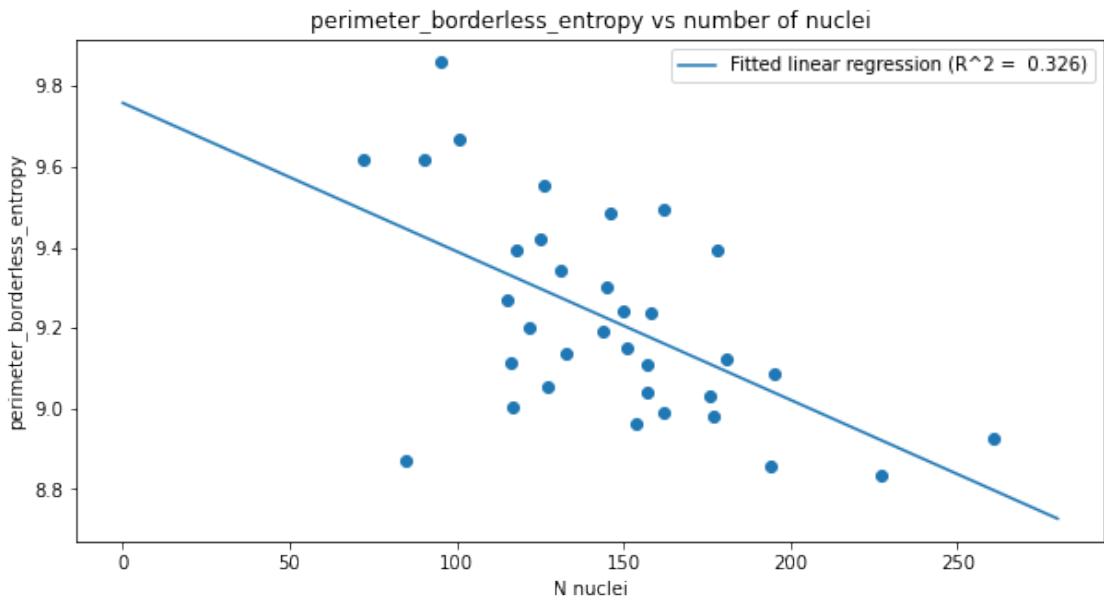


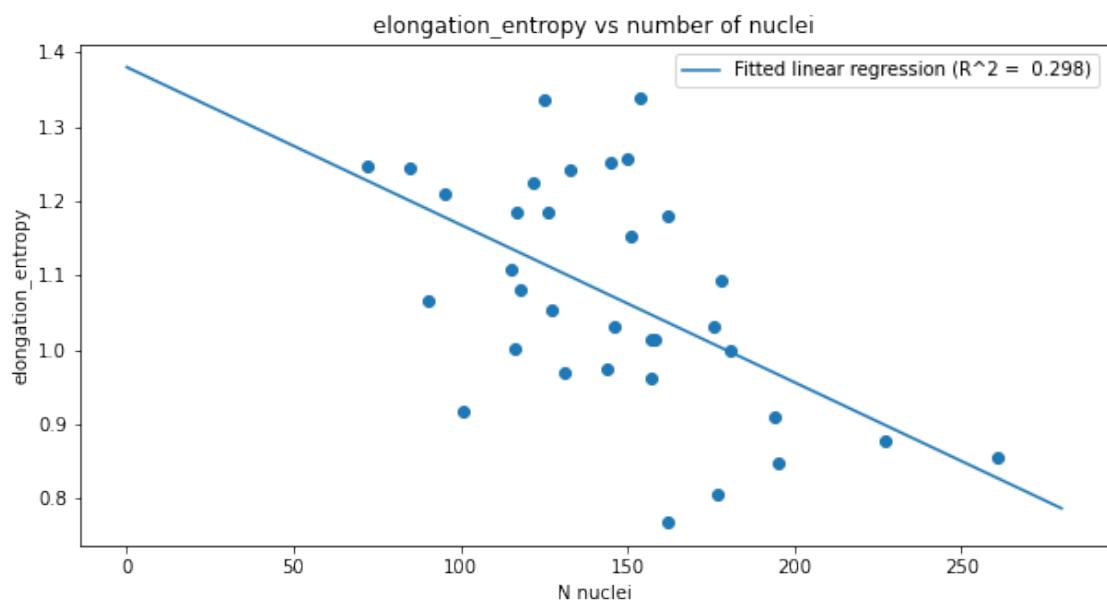
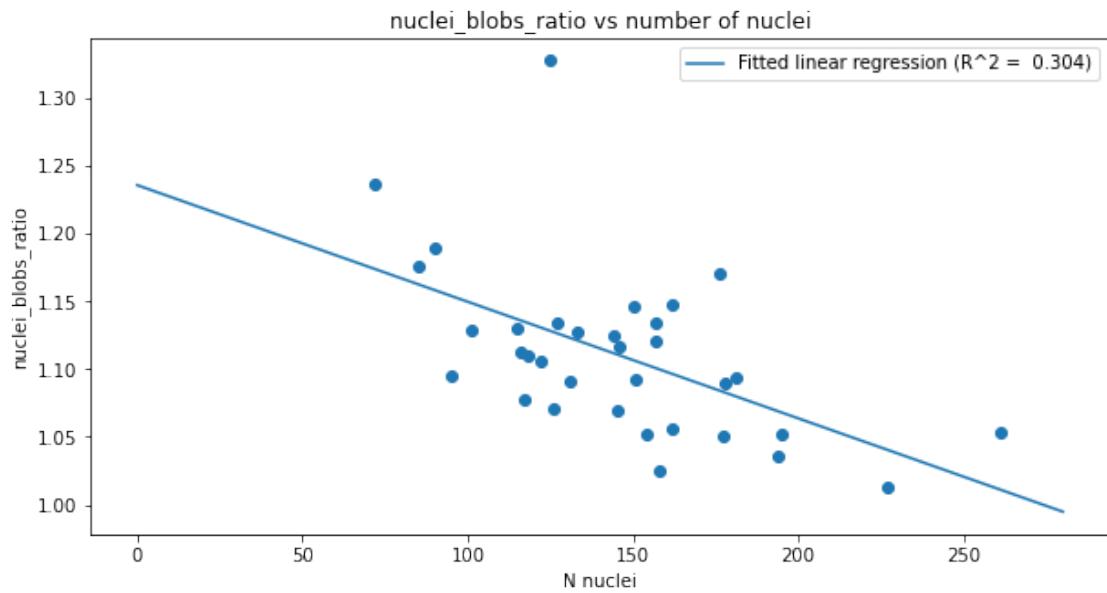


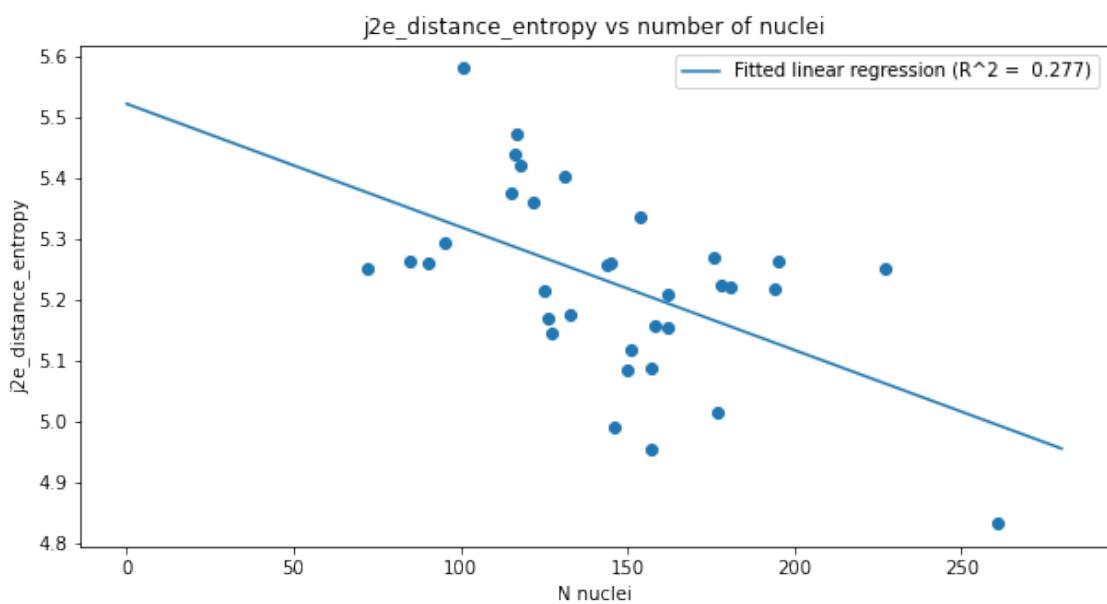
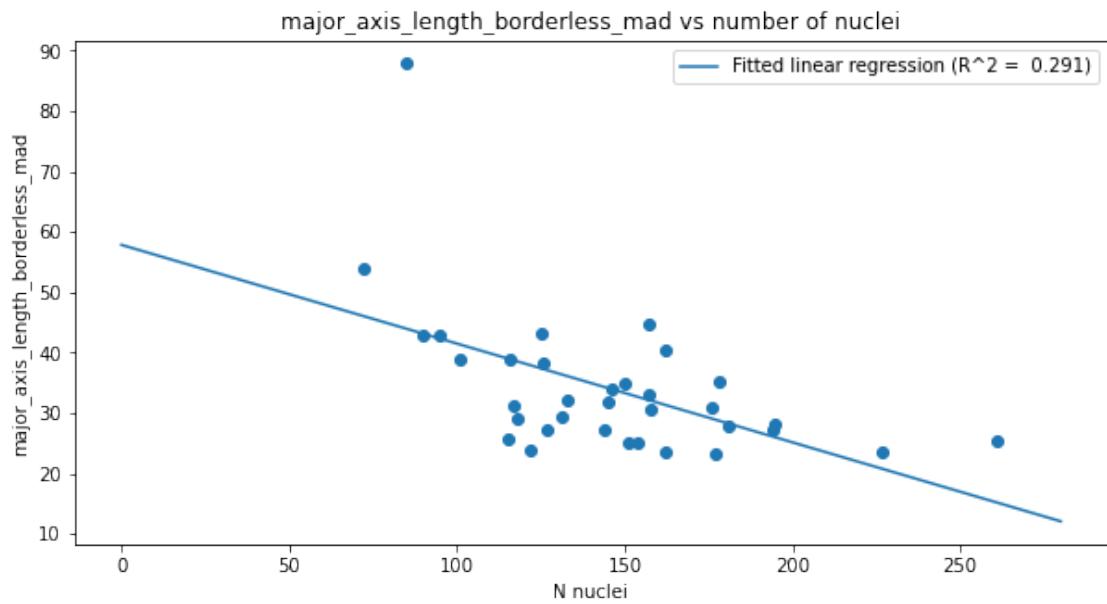


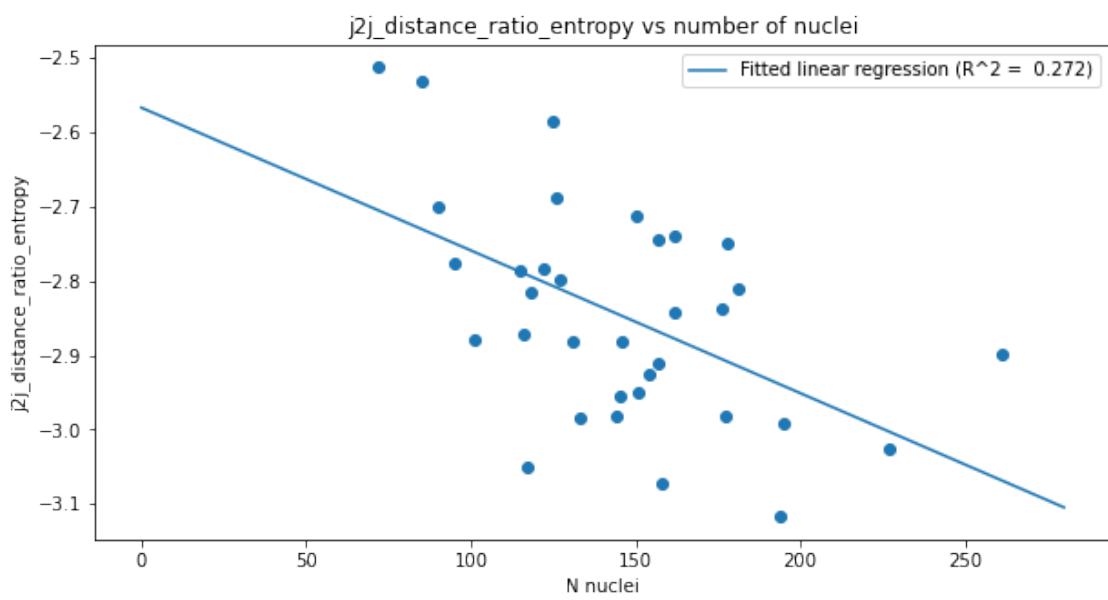
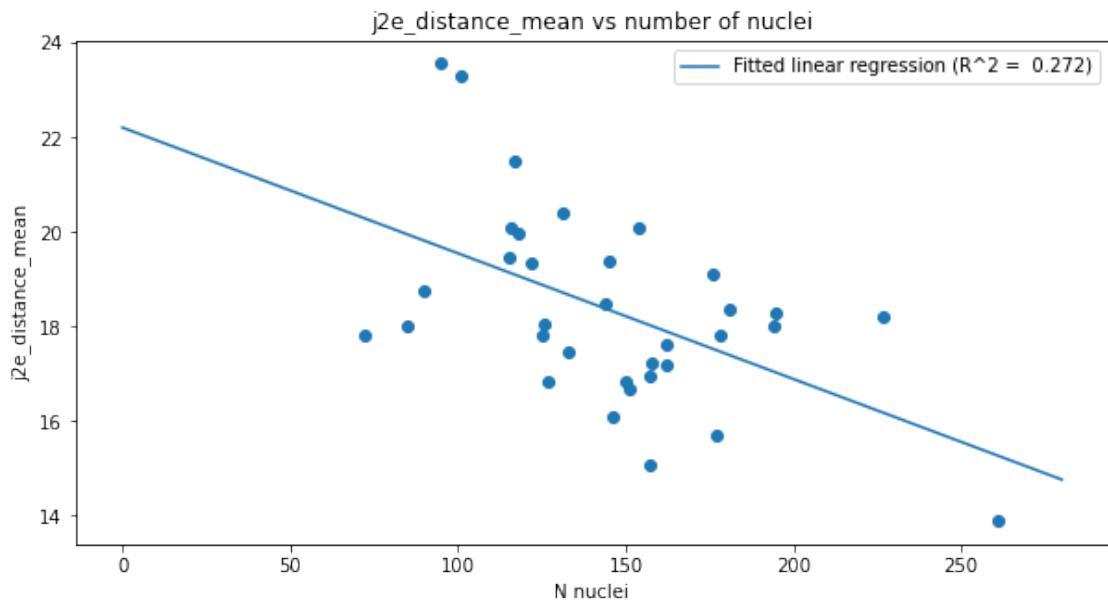


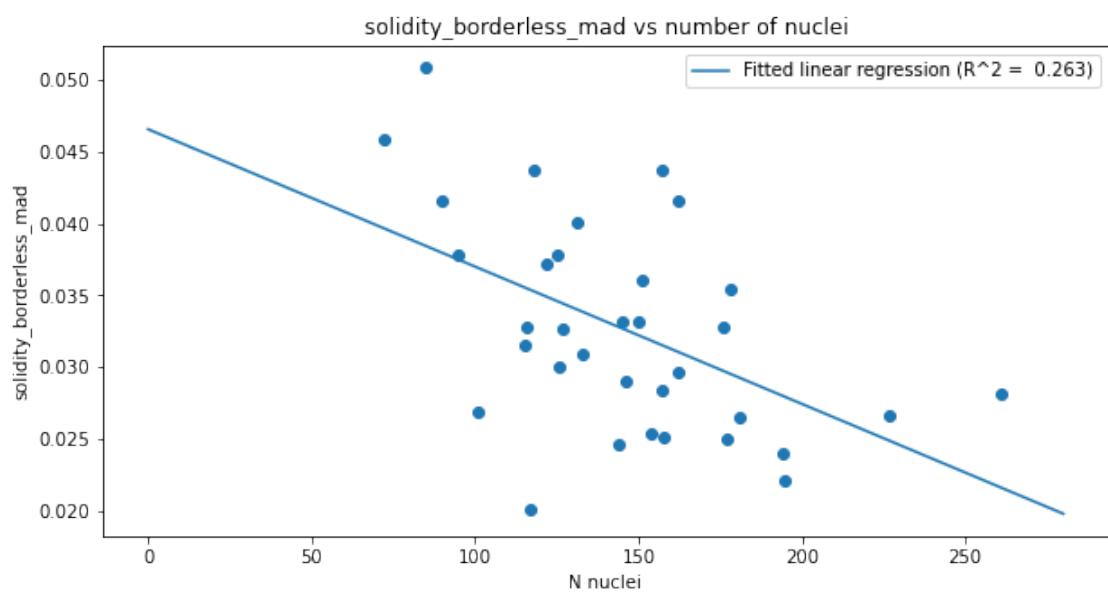
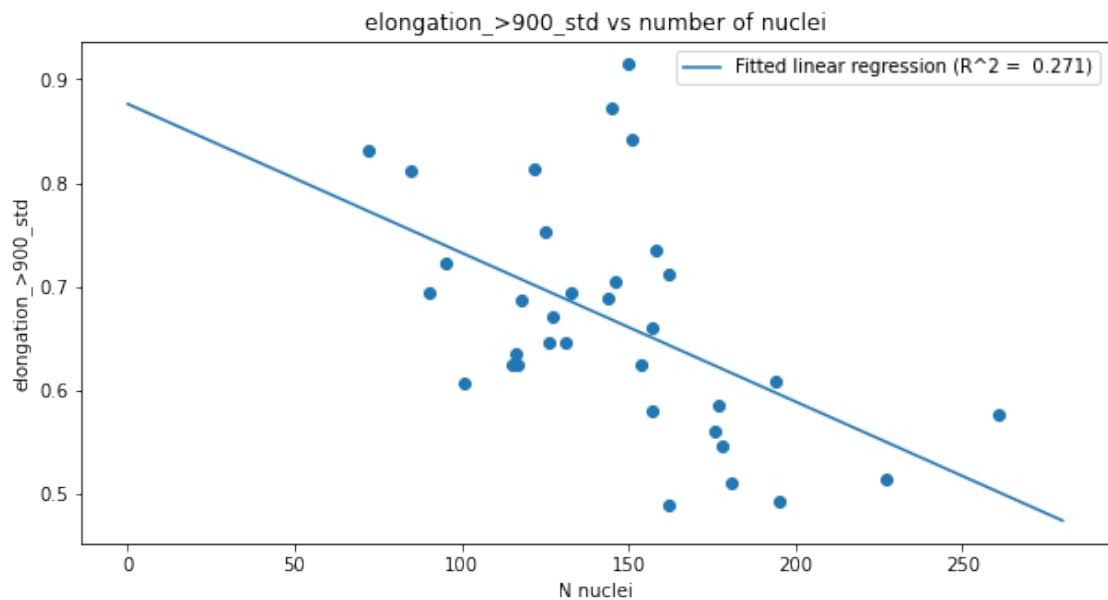


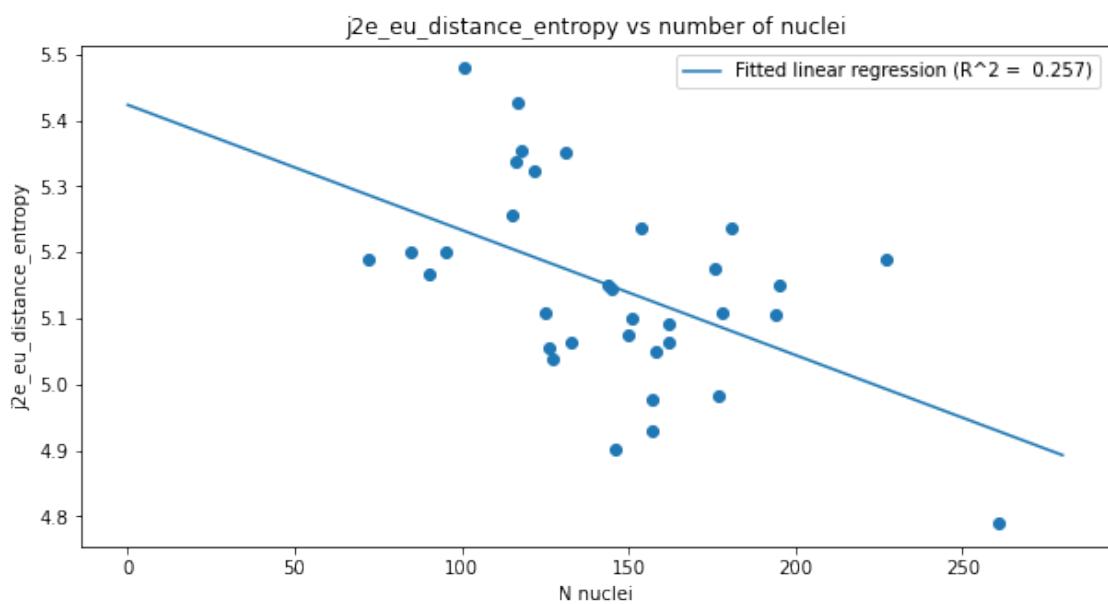
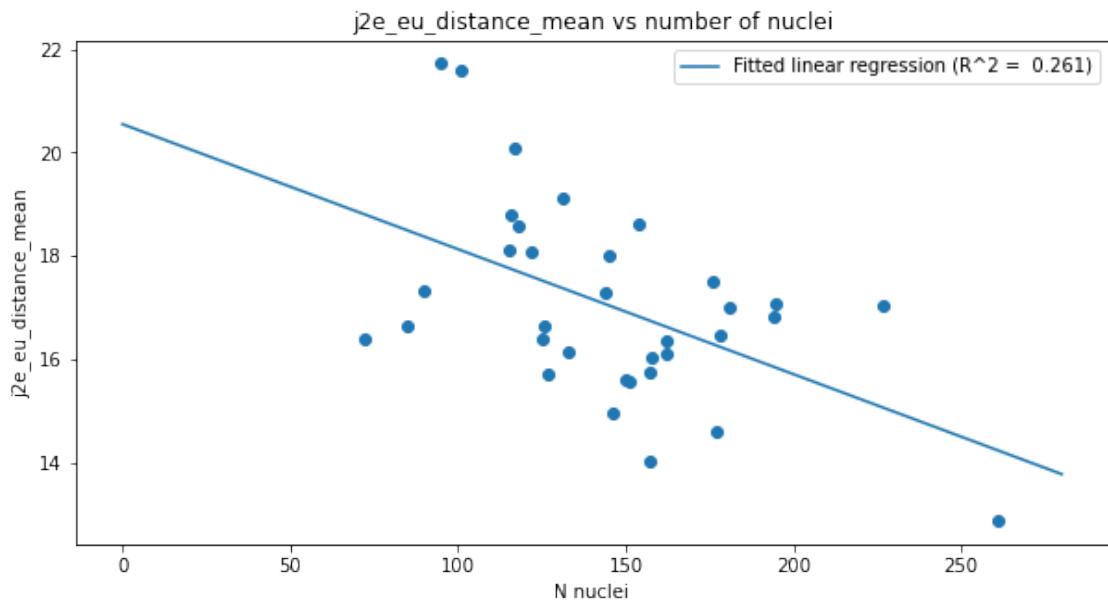


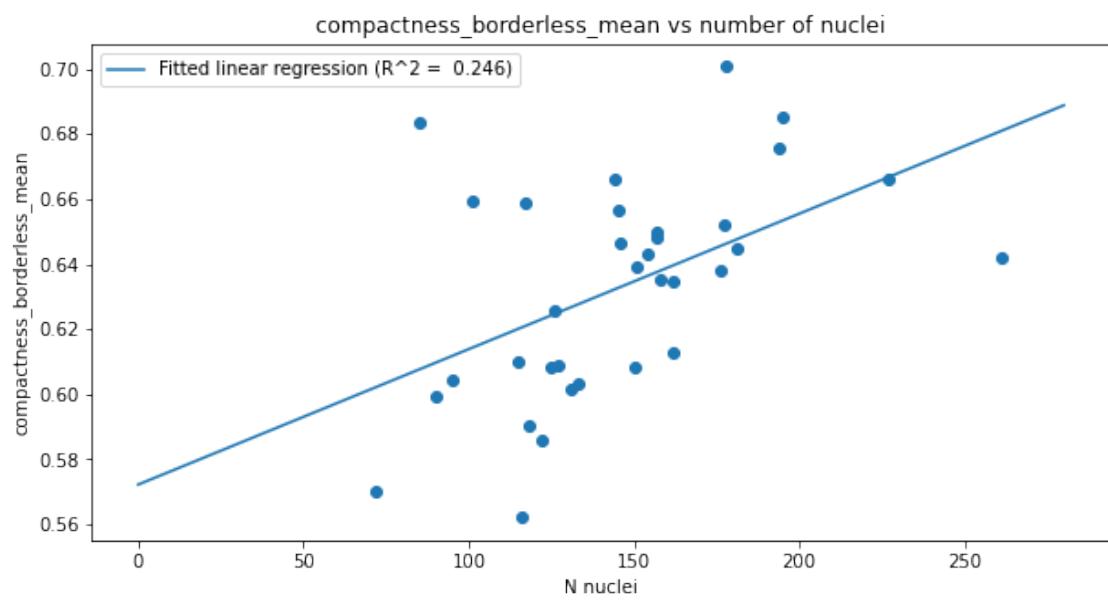
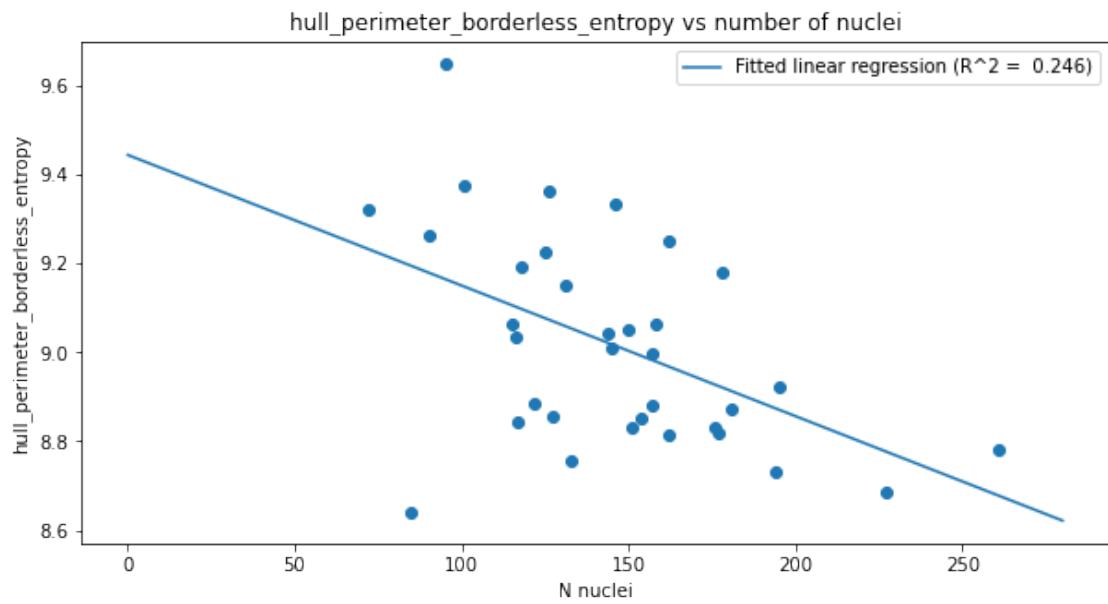


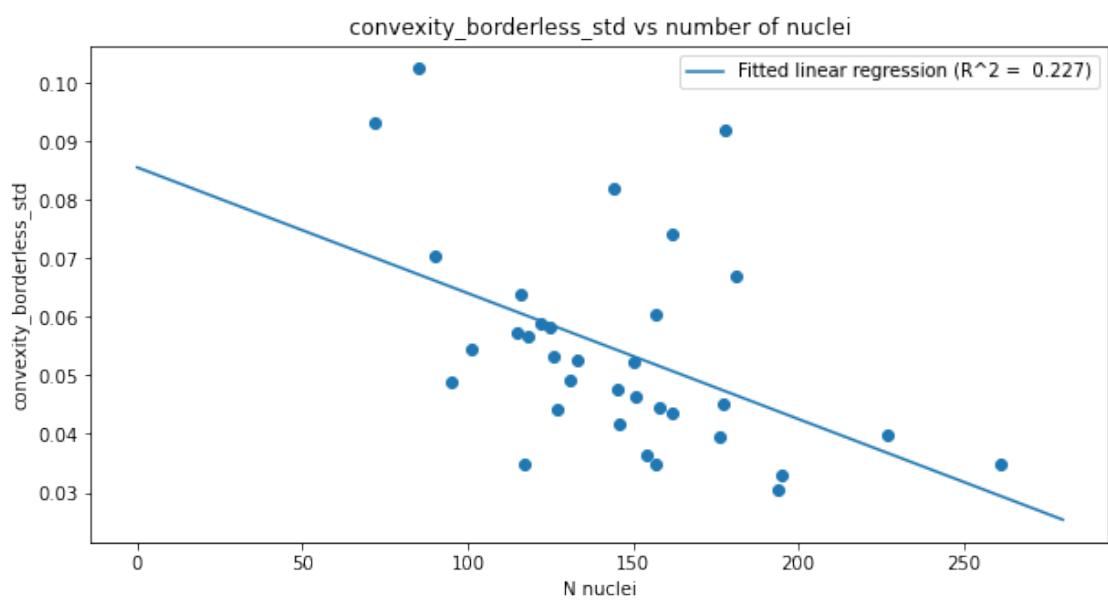
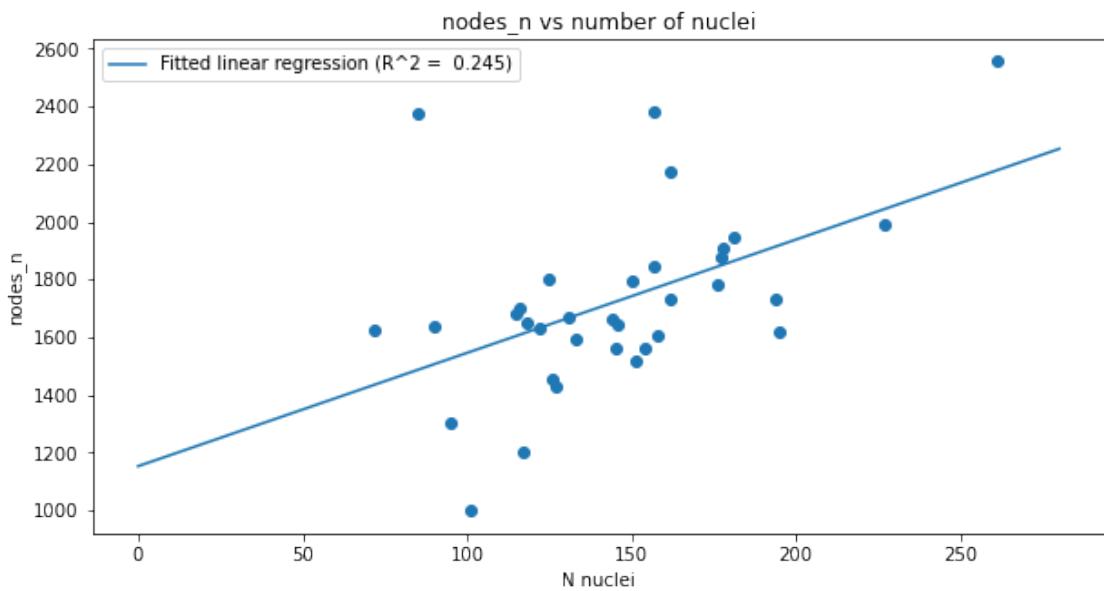




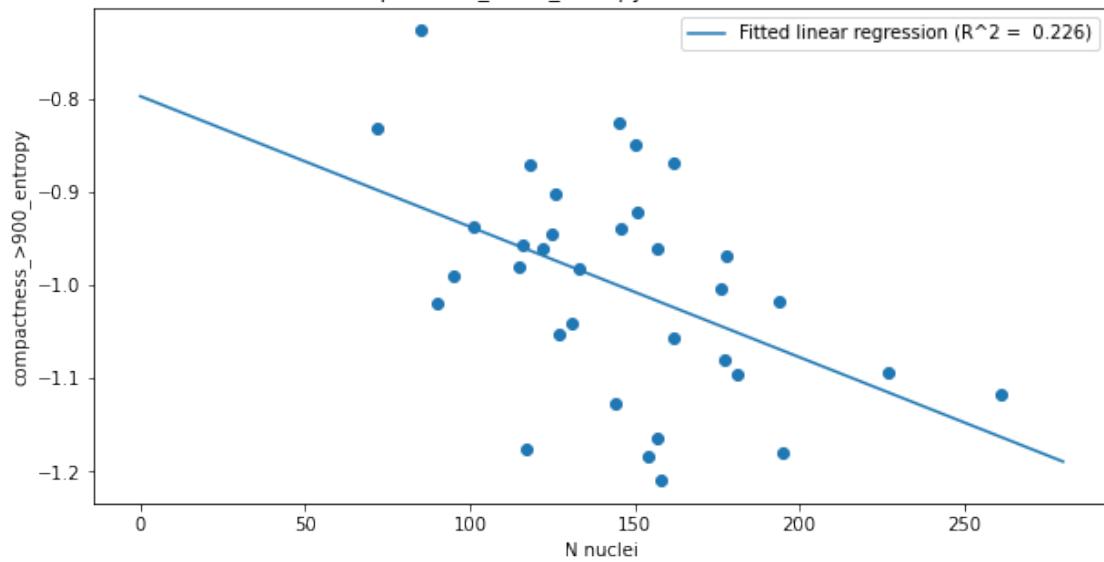




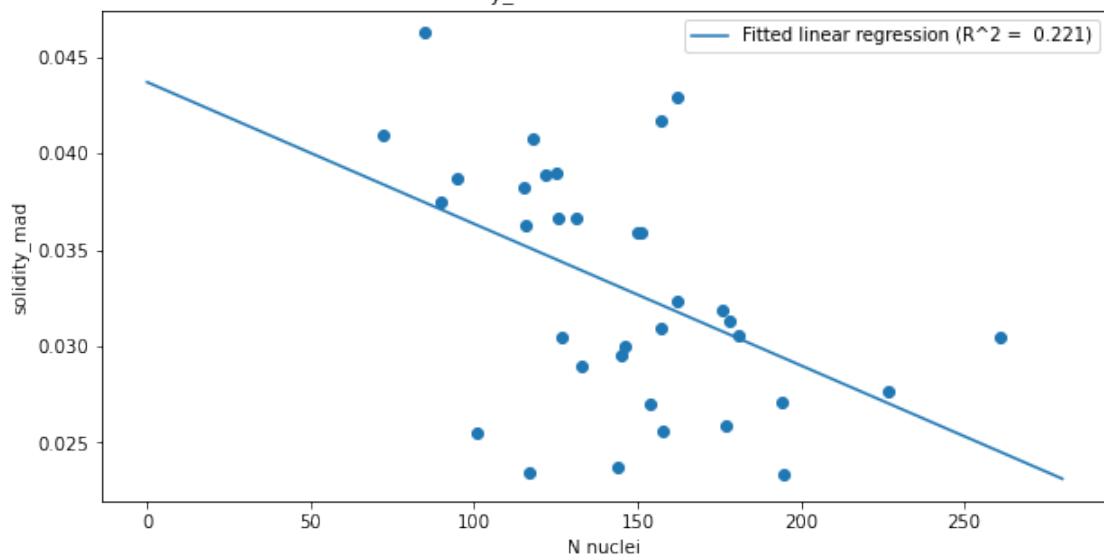


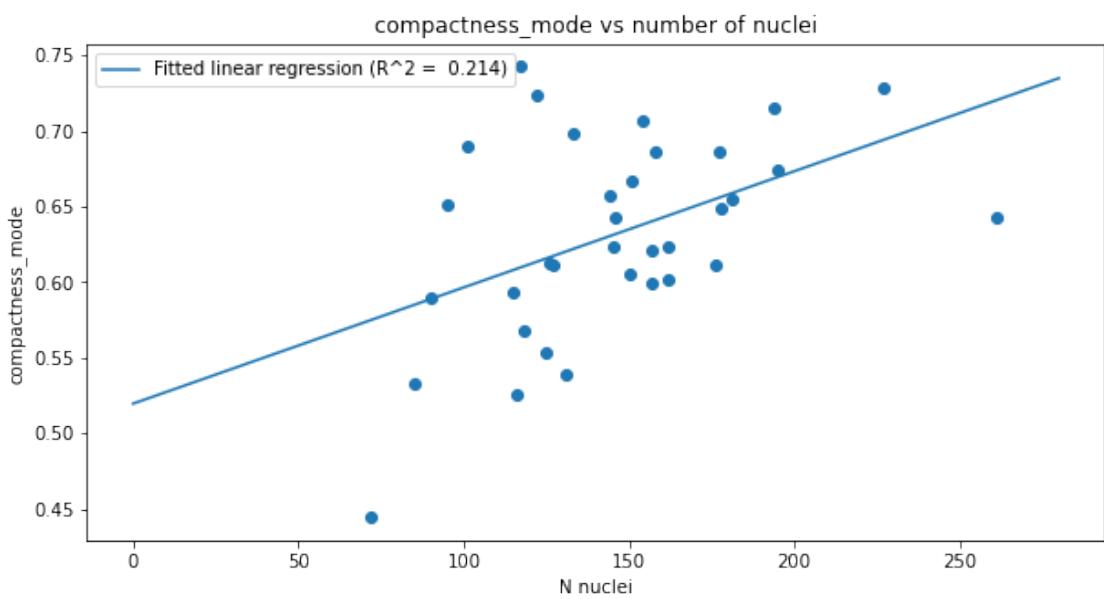
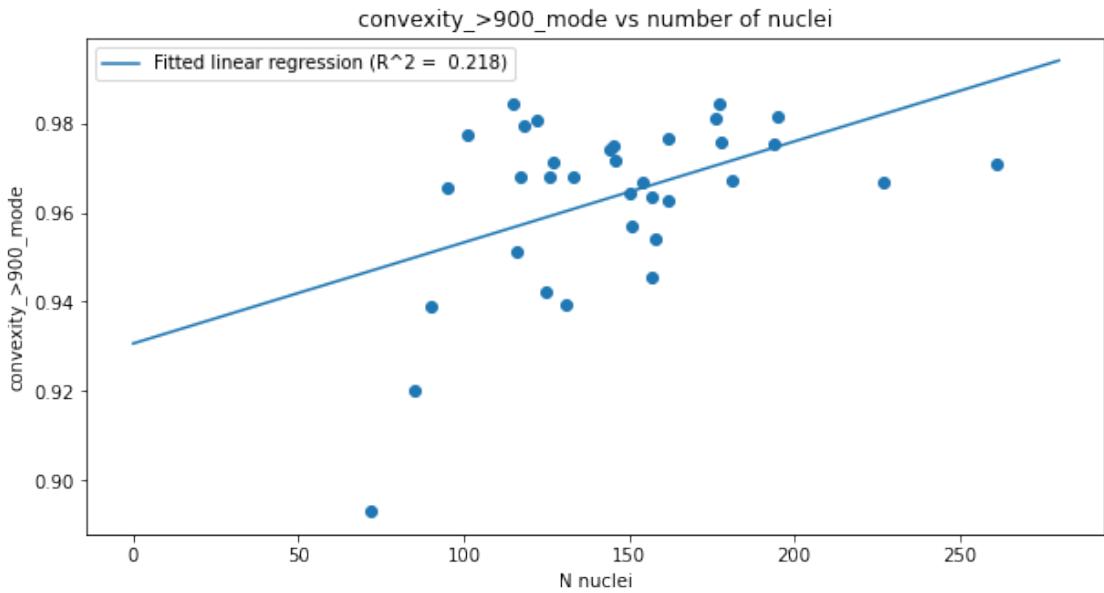


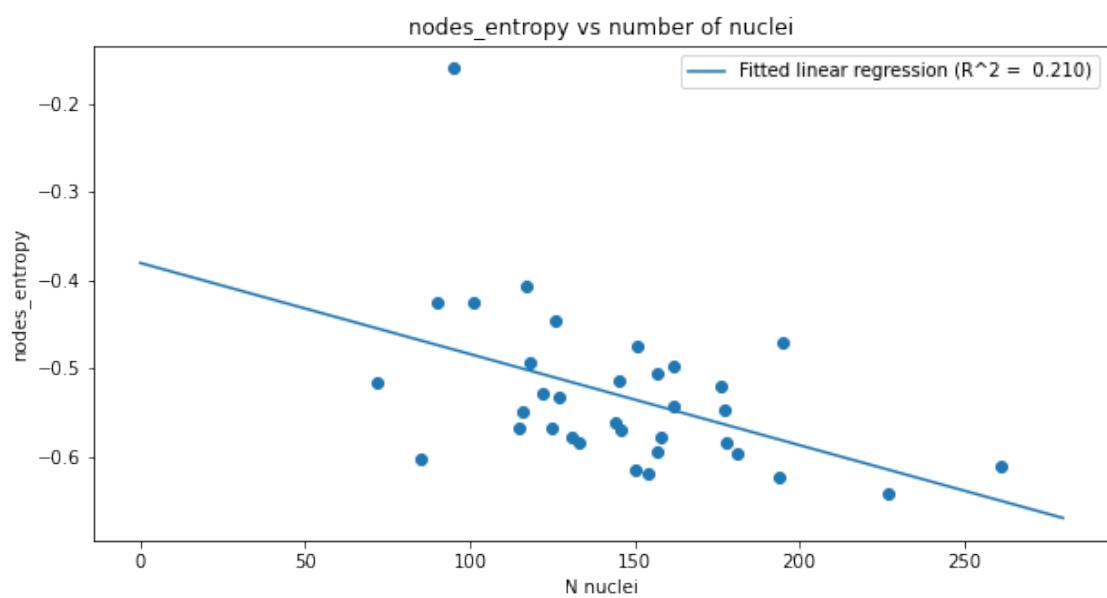
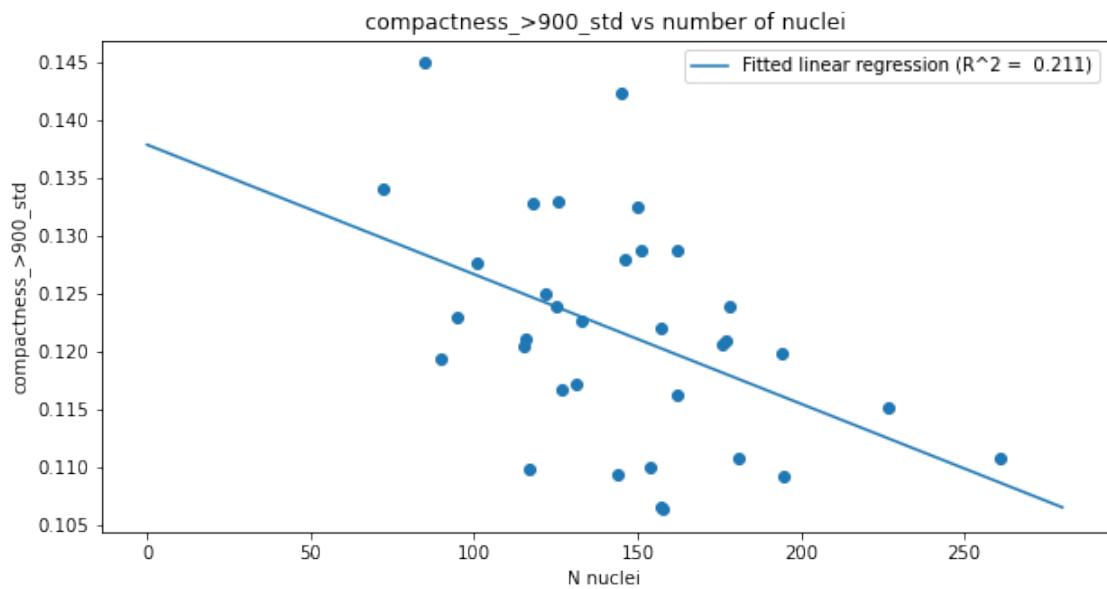
compactness_>900_entropy vs number of nuclei

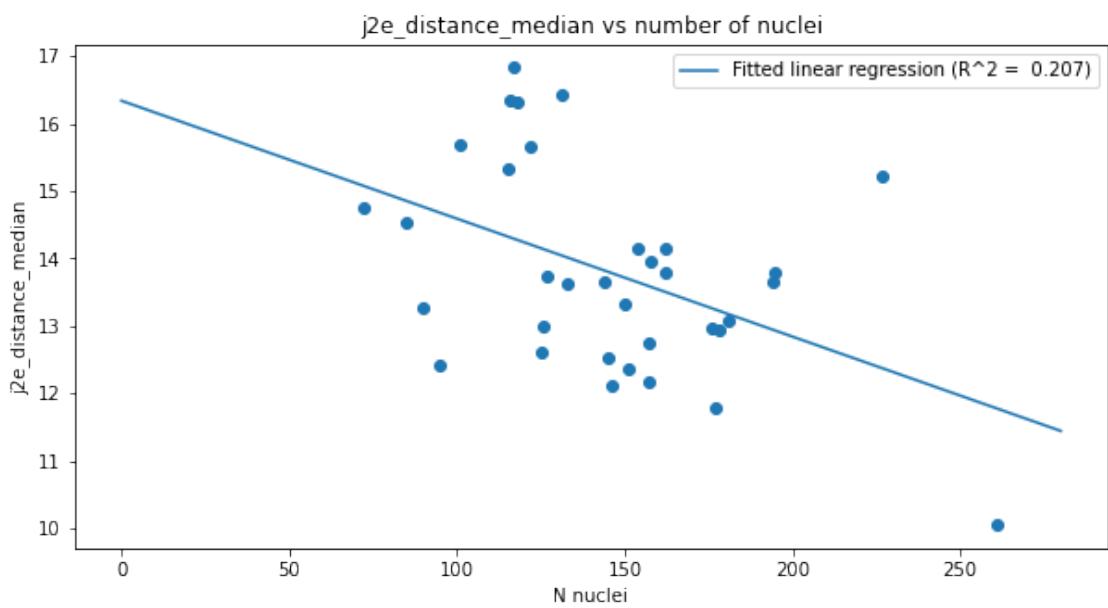
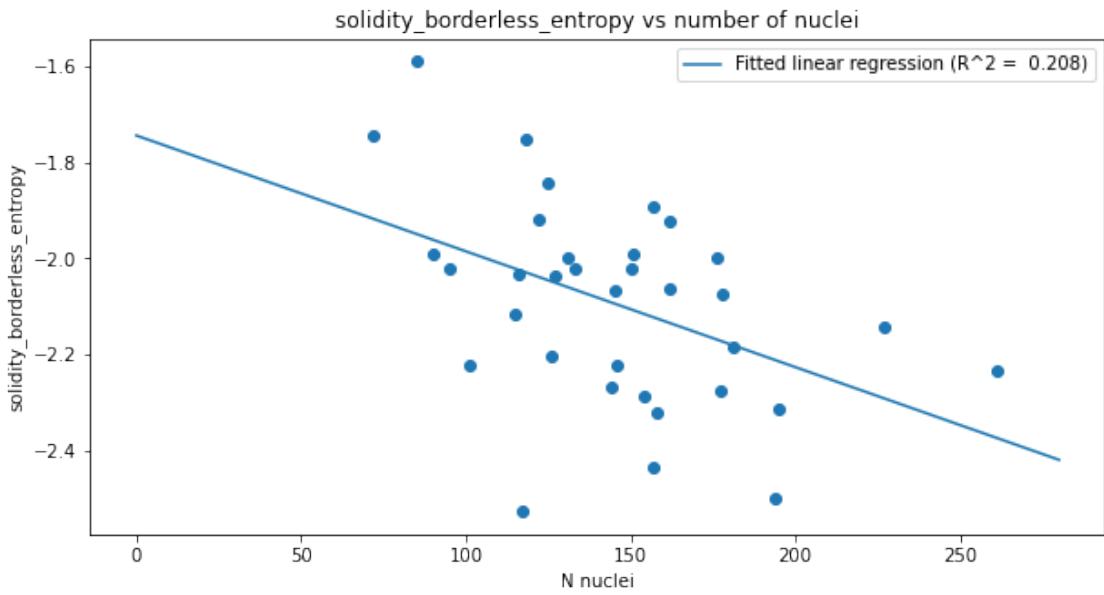


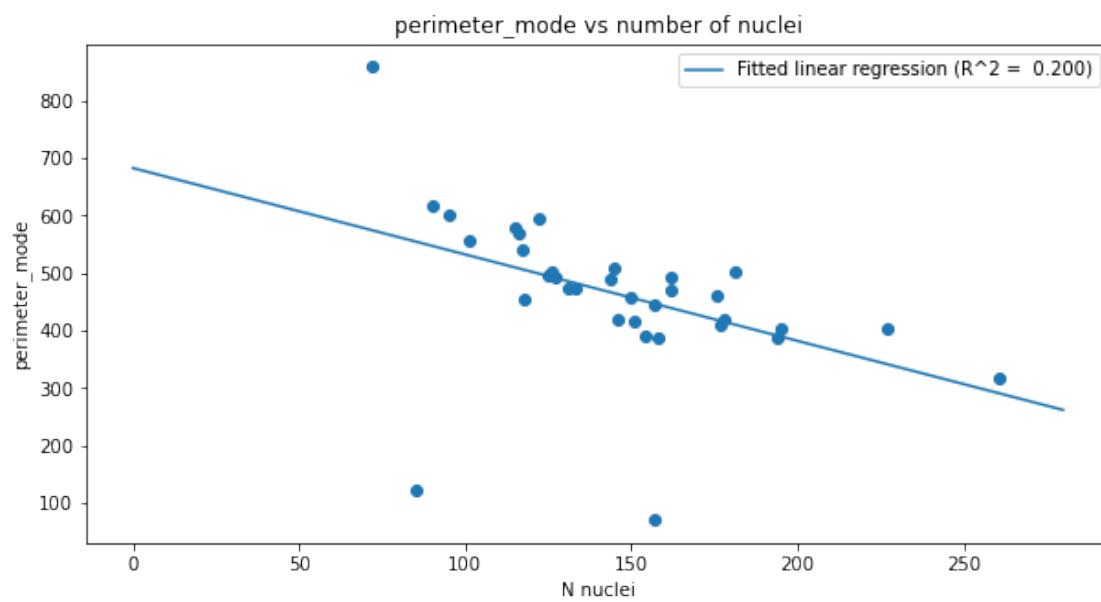
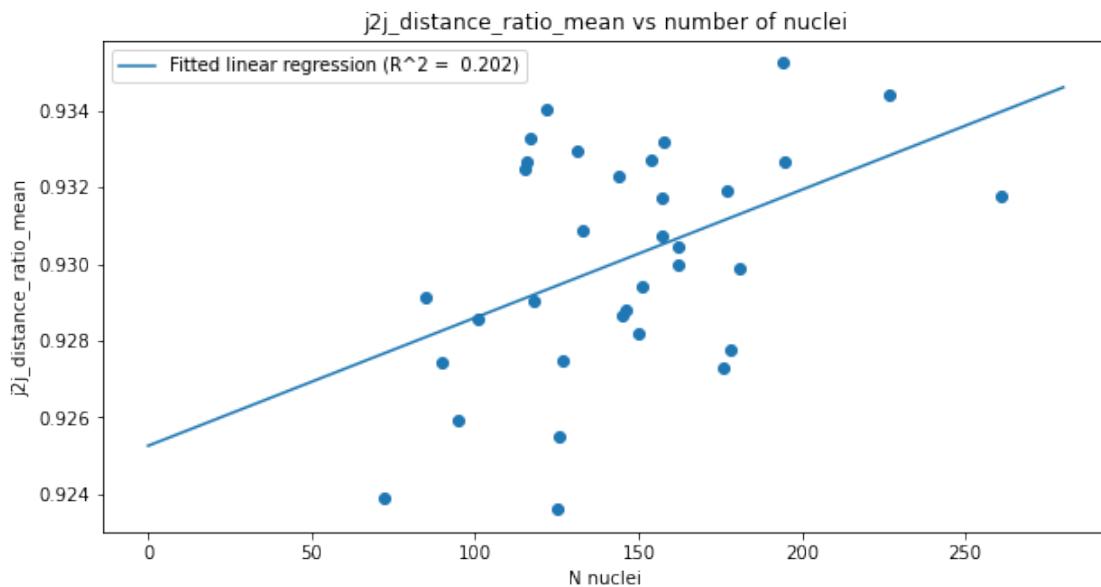
solidity_mad vs number of nuclei

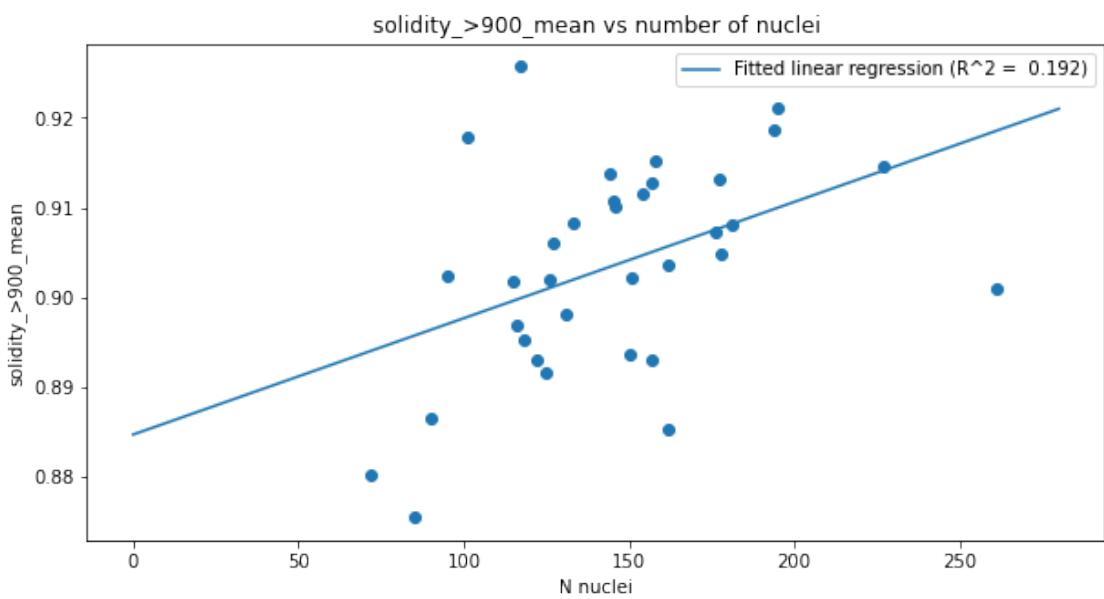
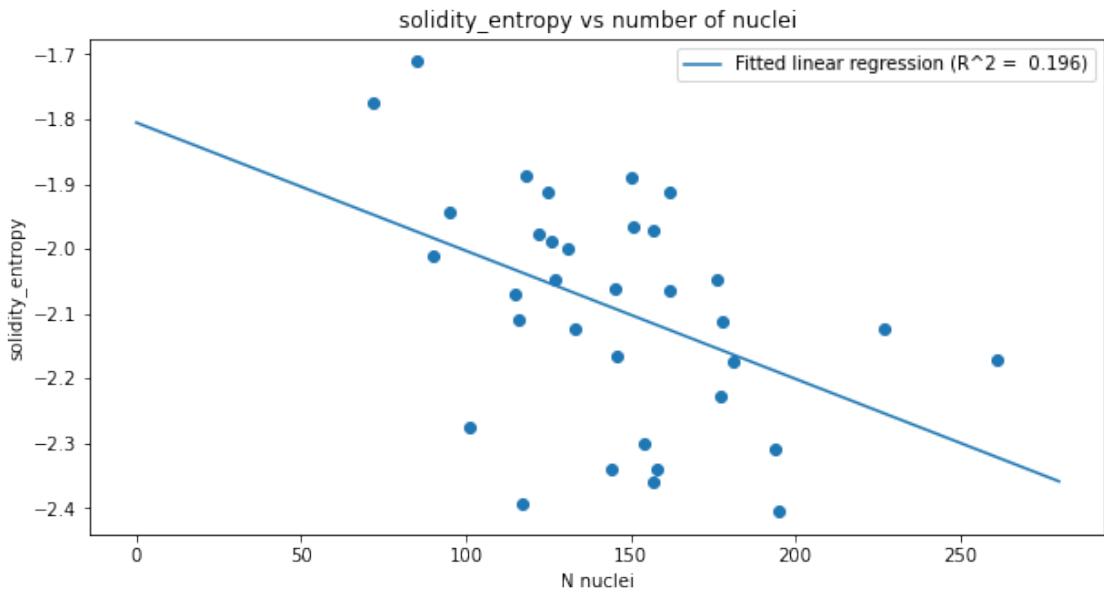


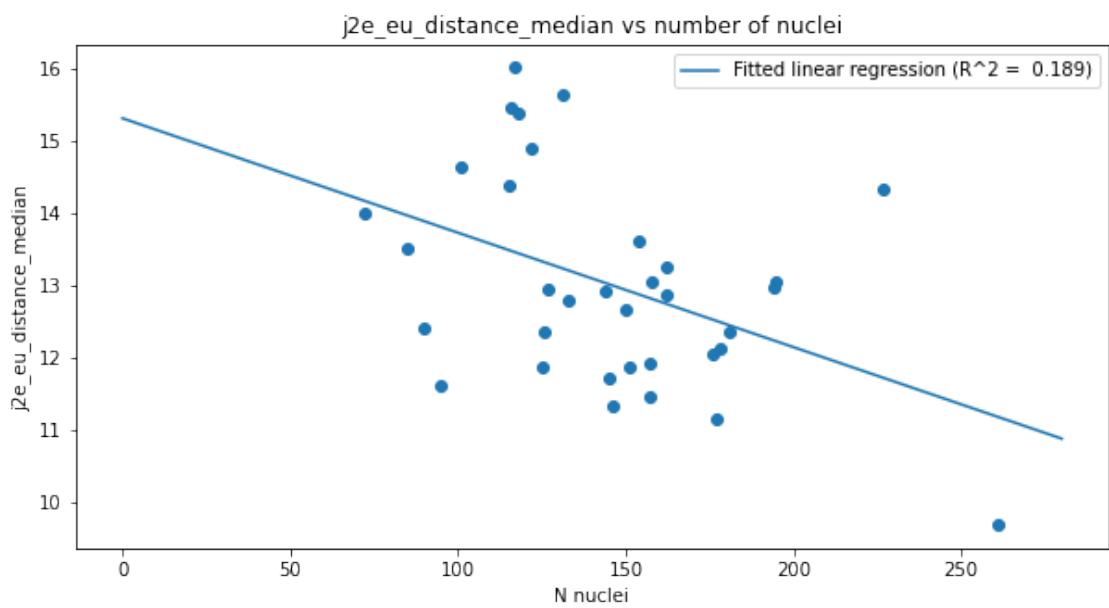
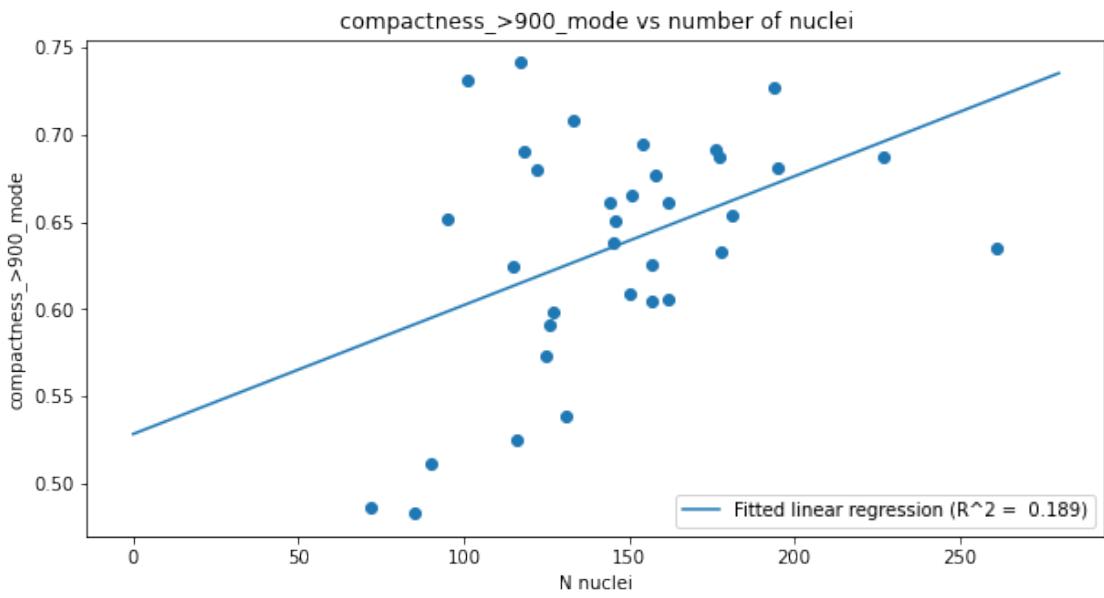


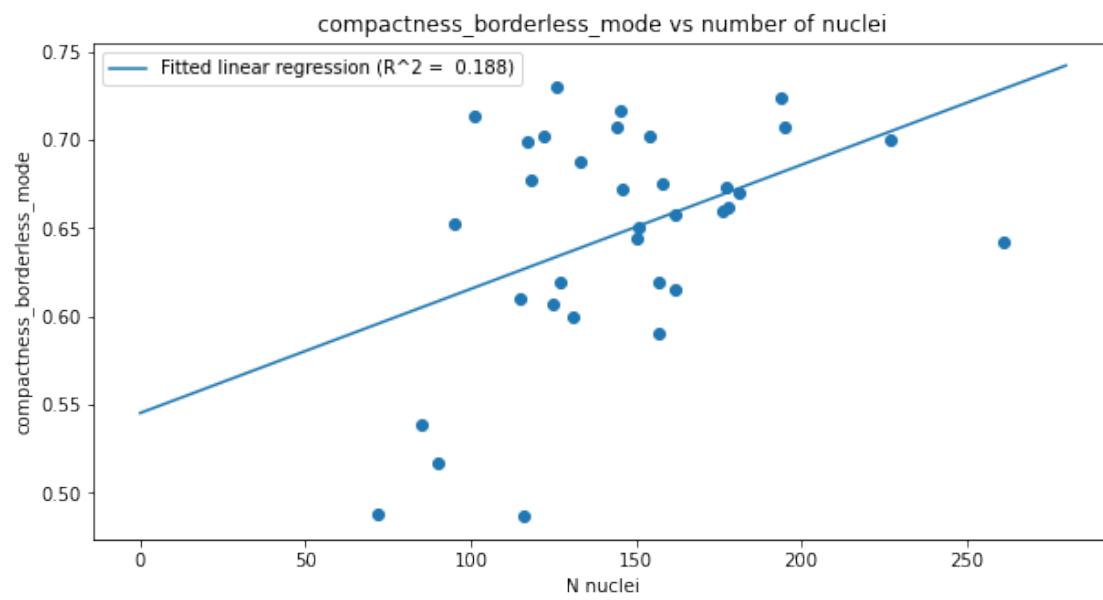
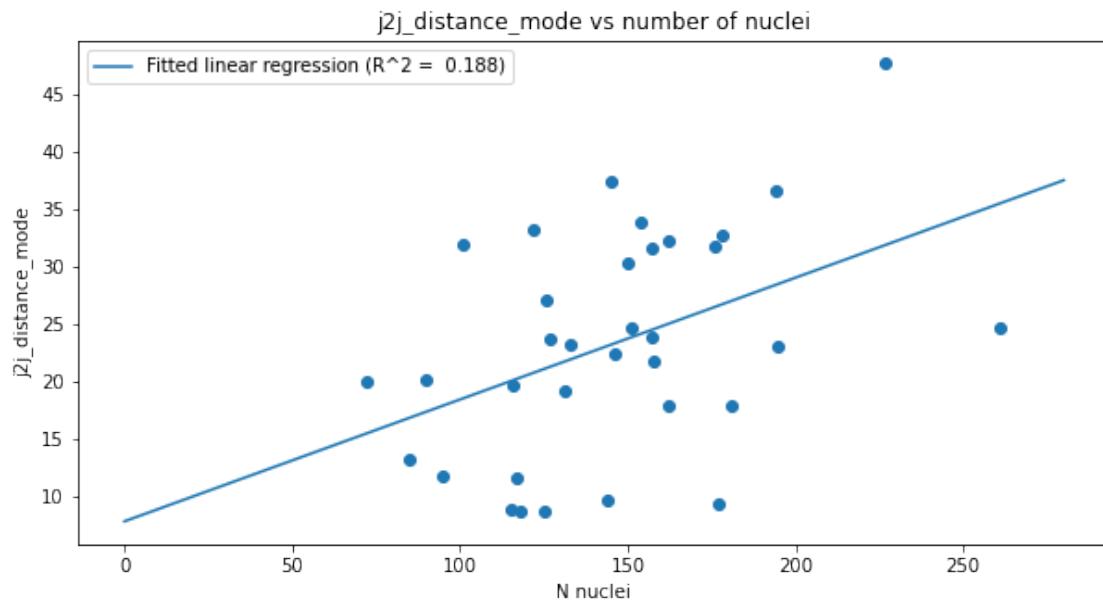


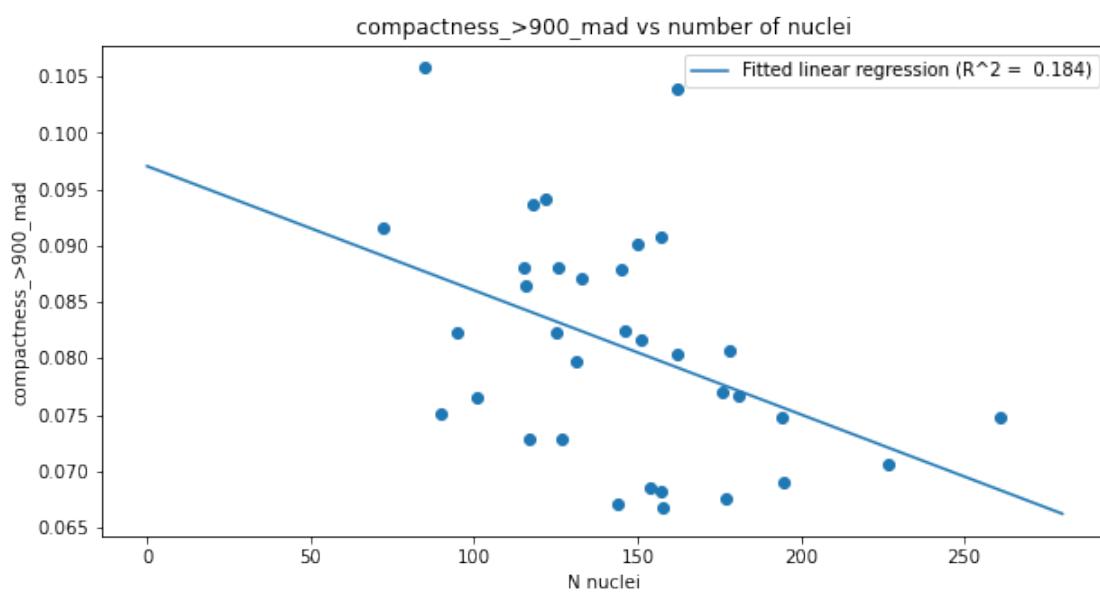
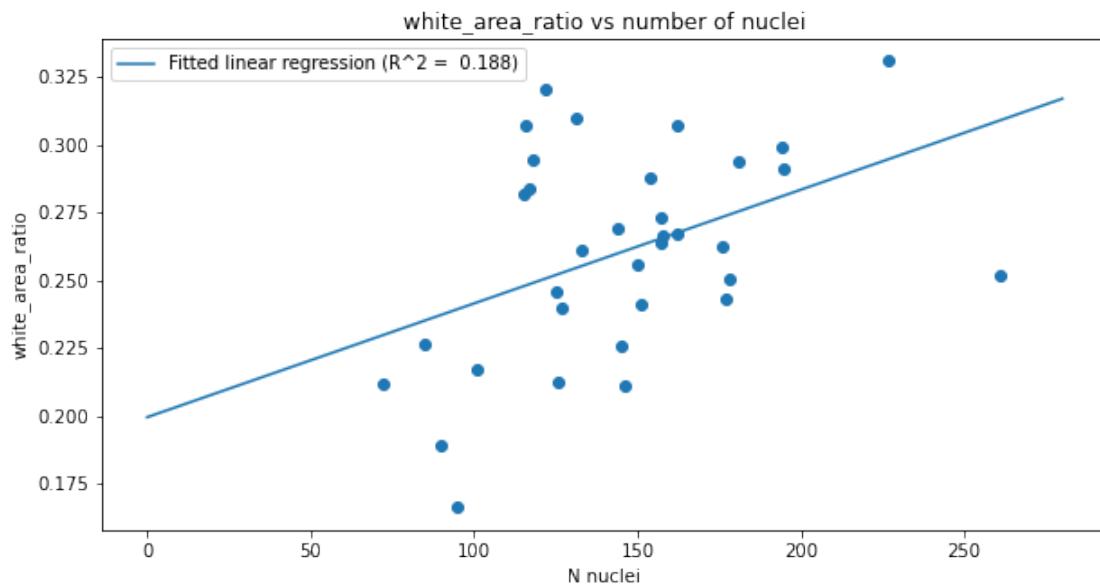


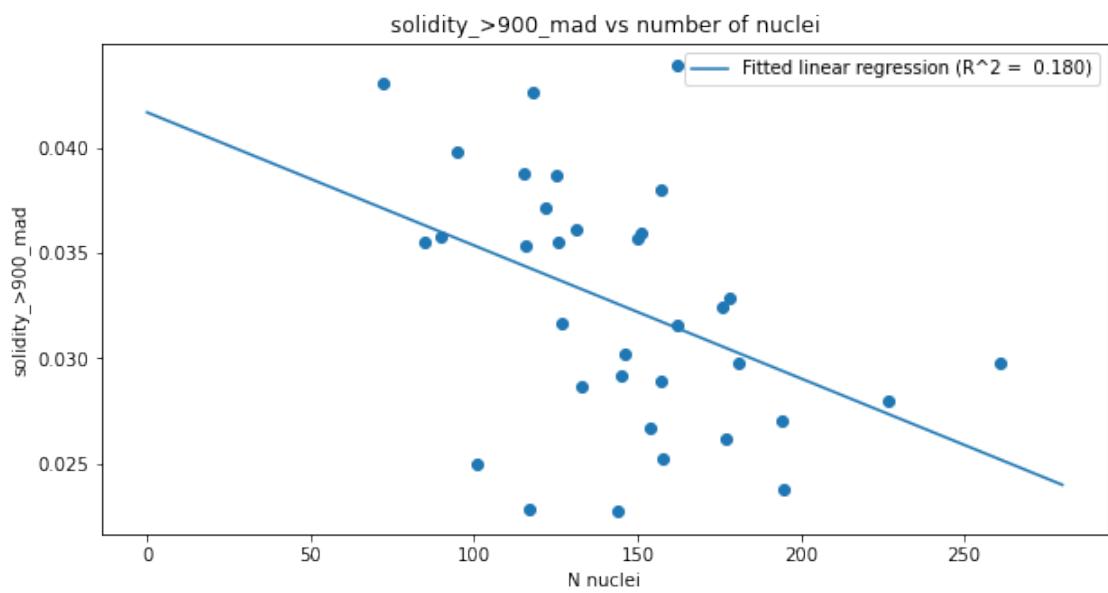
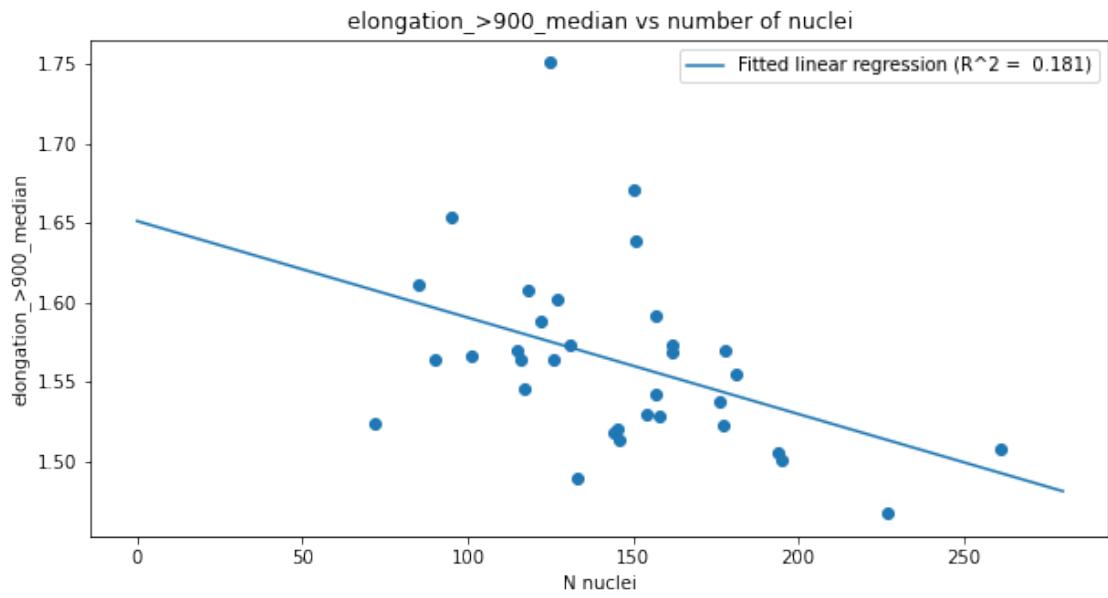


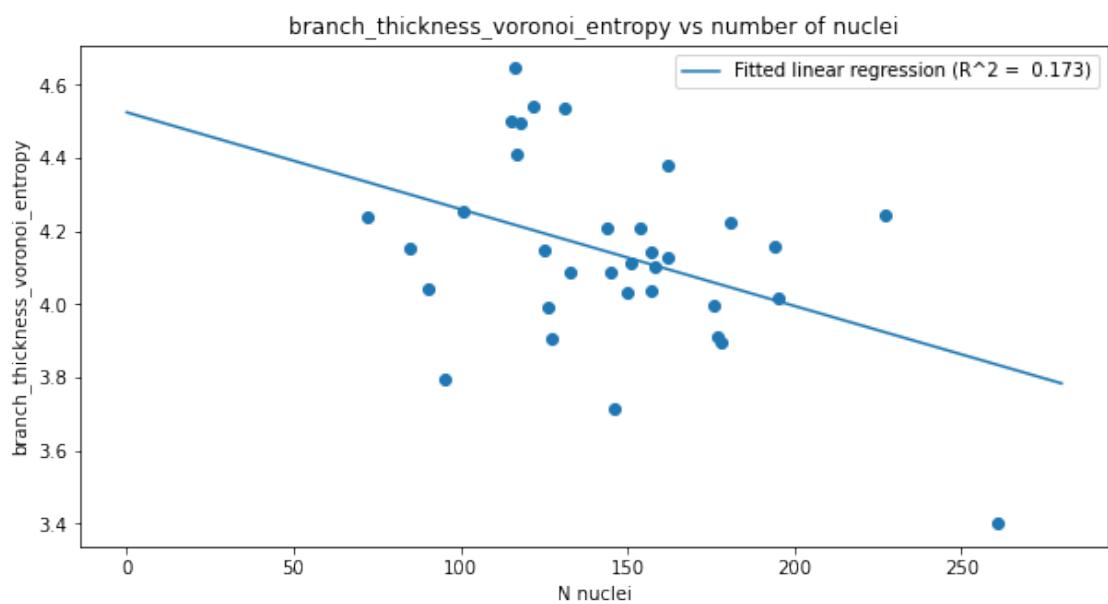
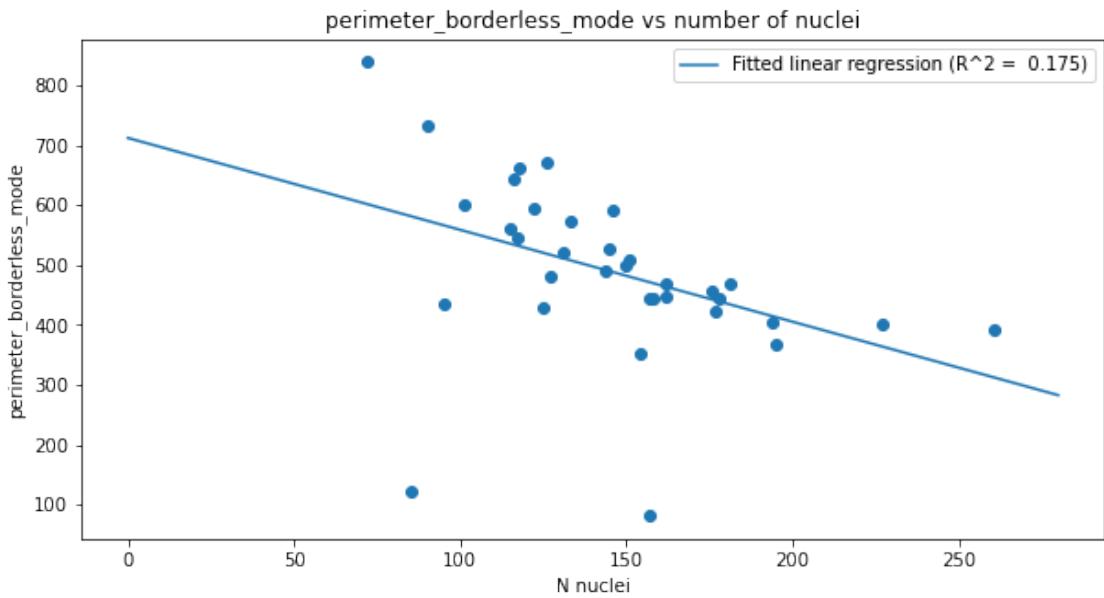


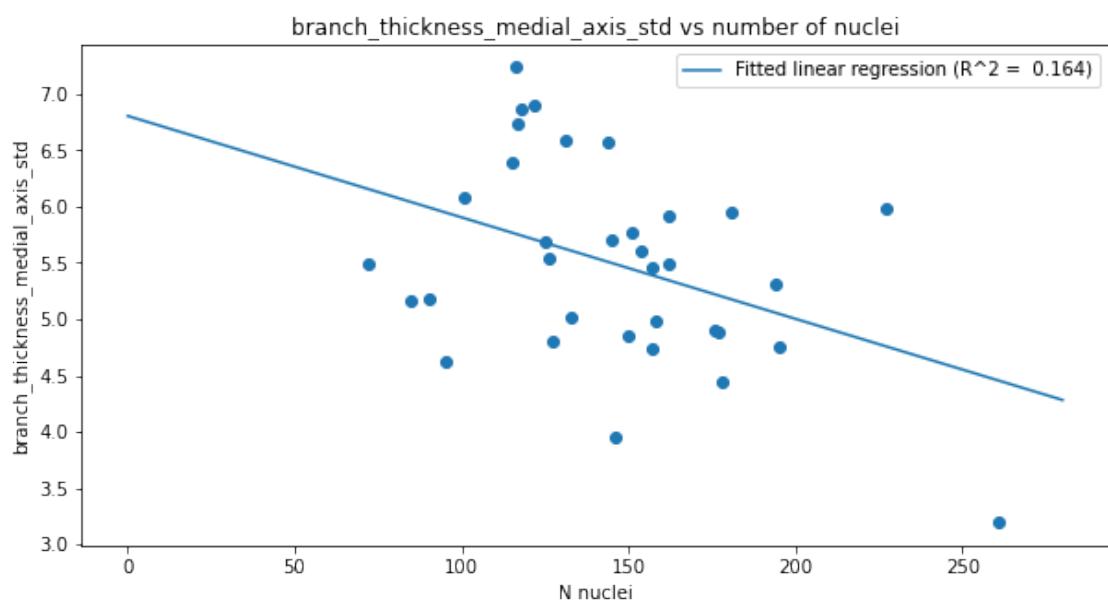
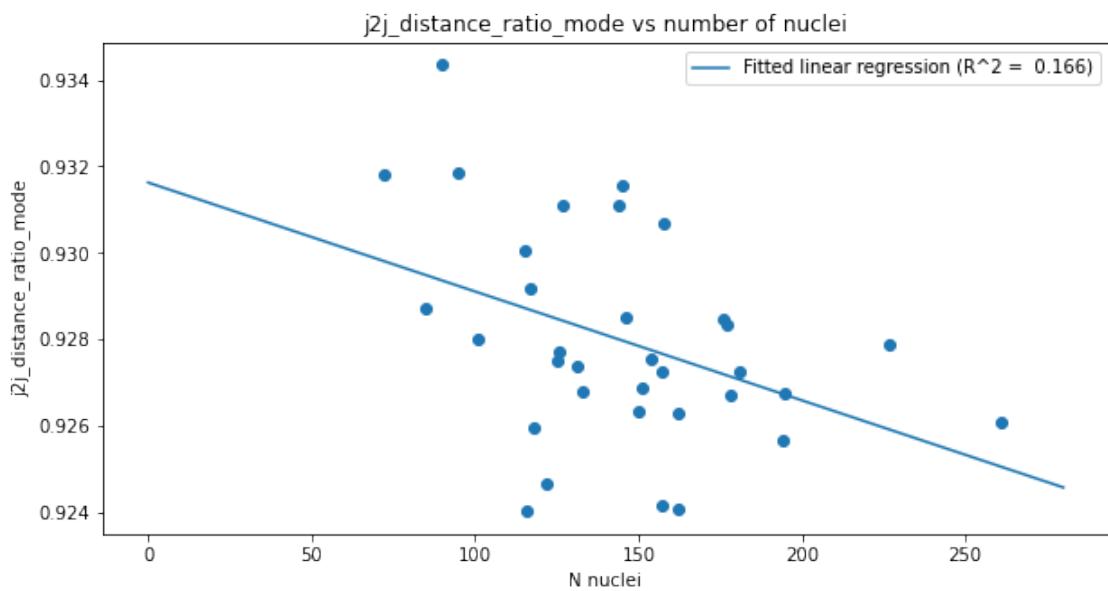


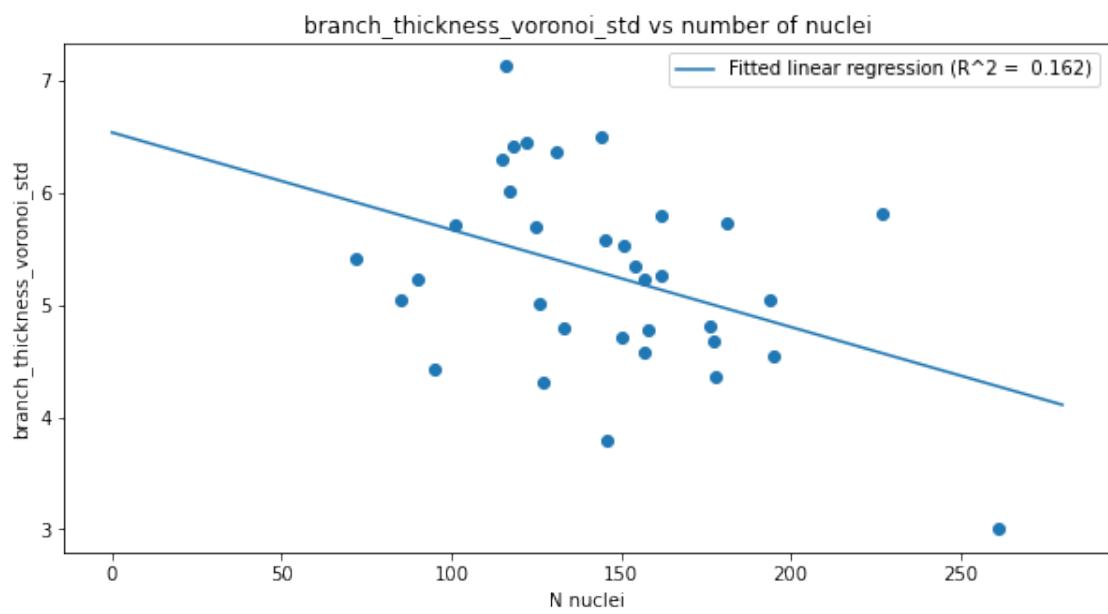
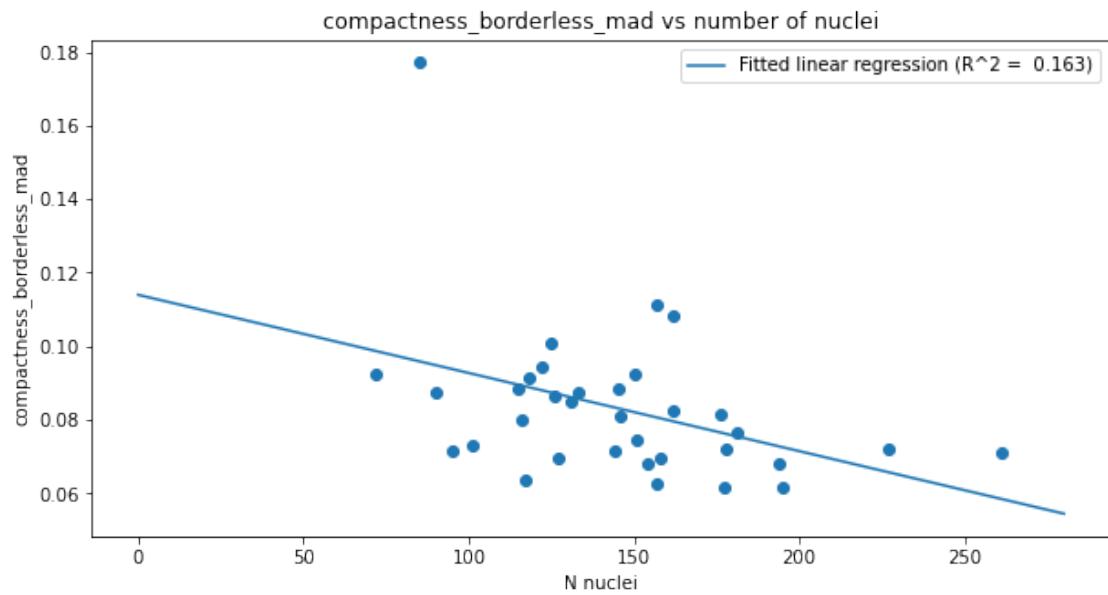


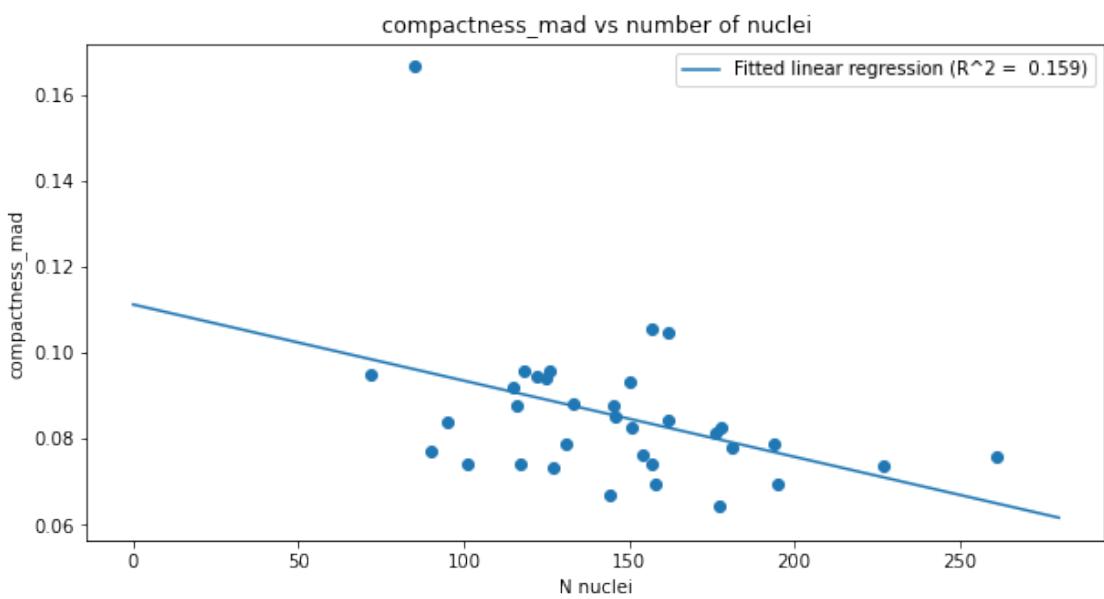
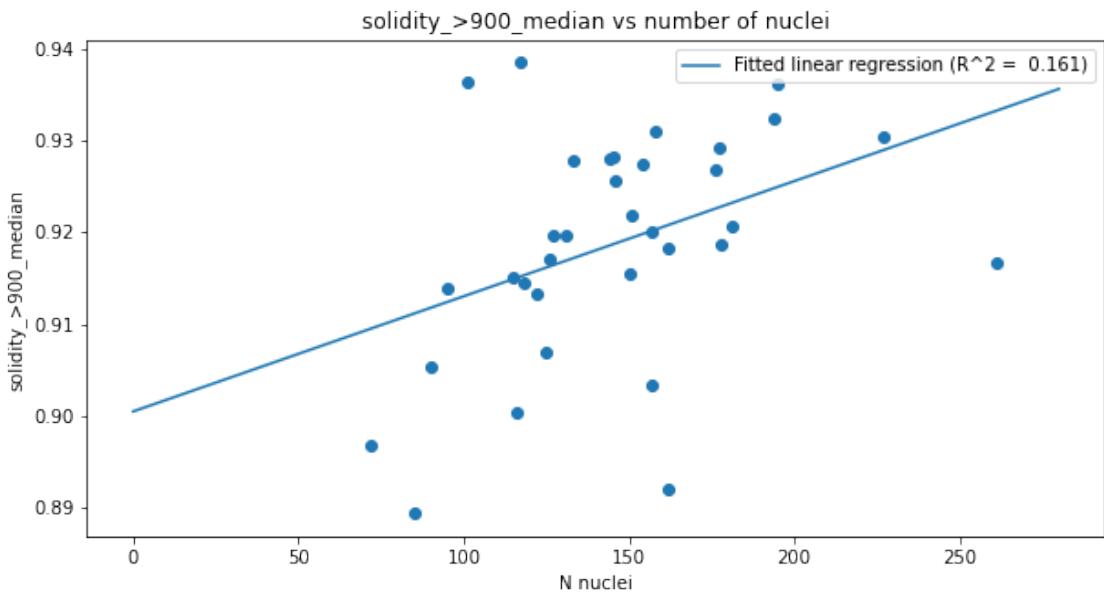


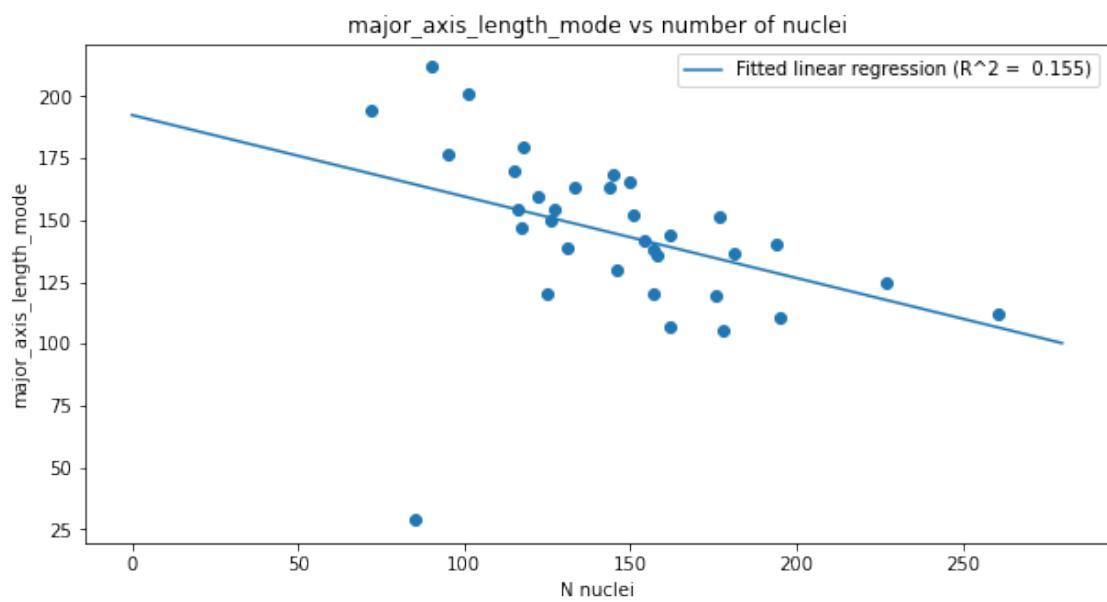
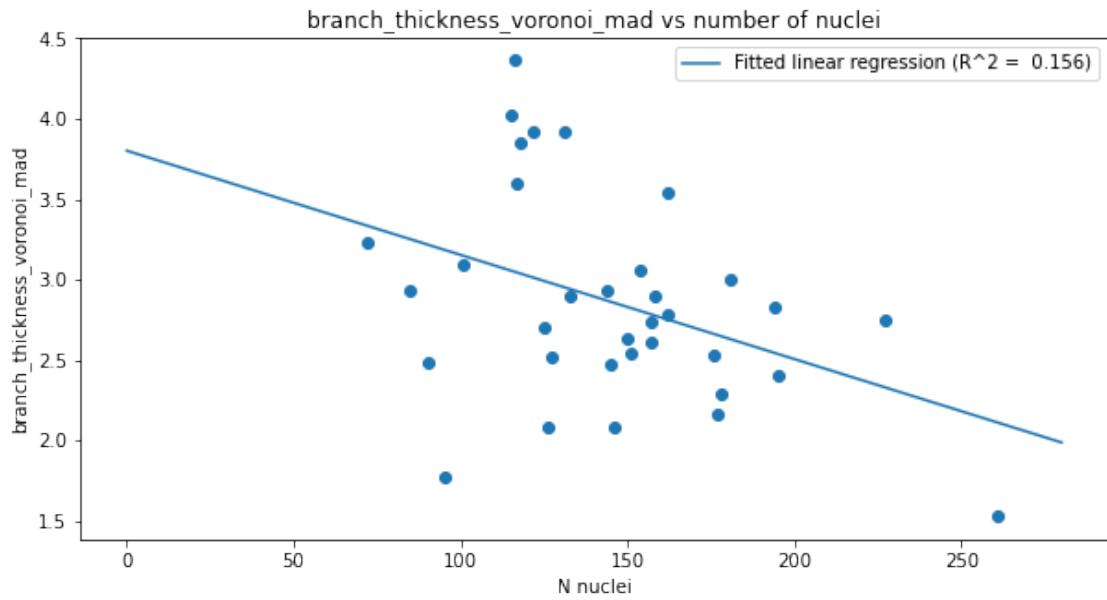


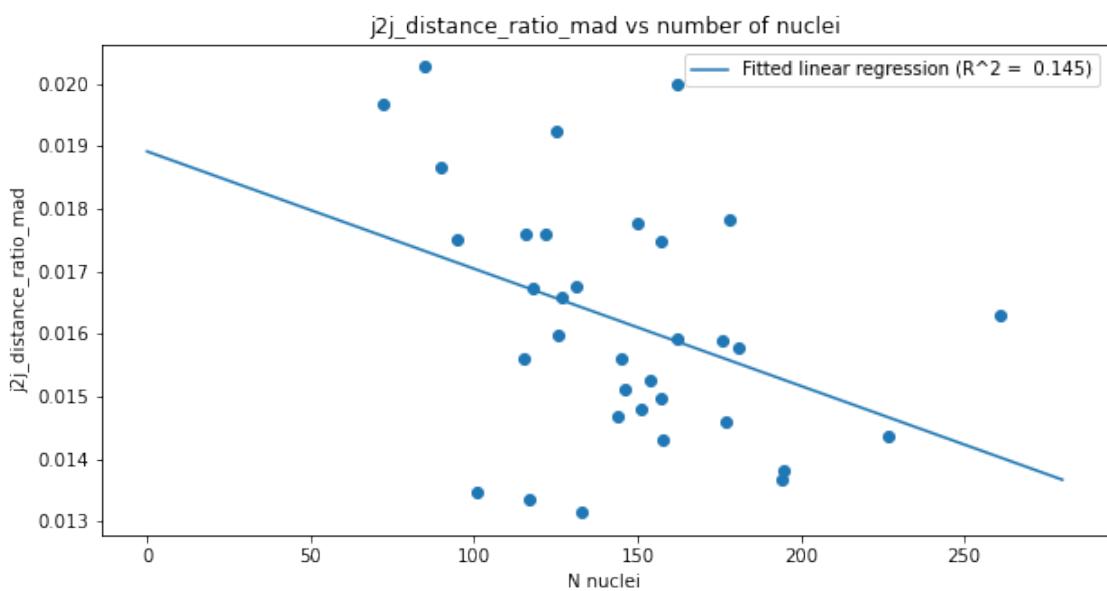
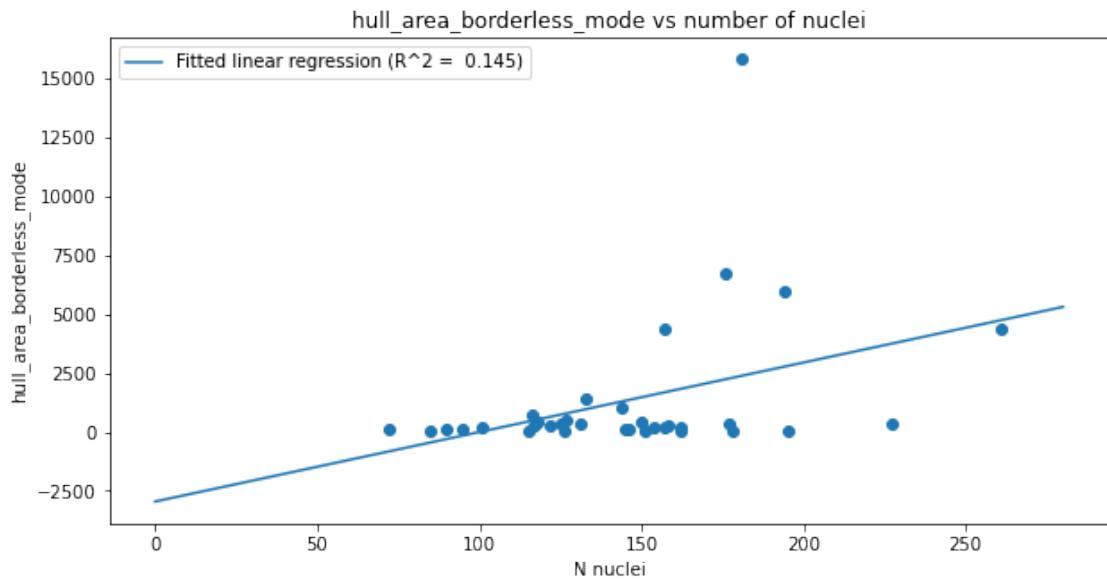


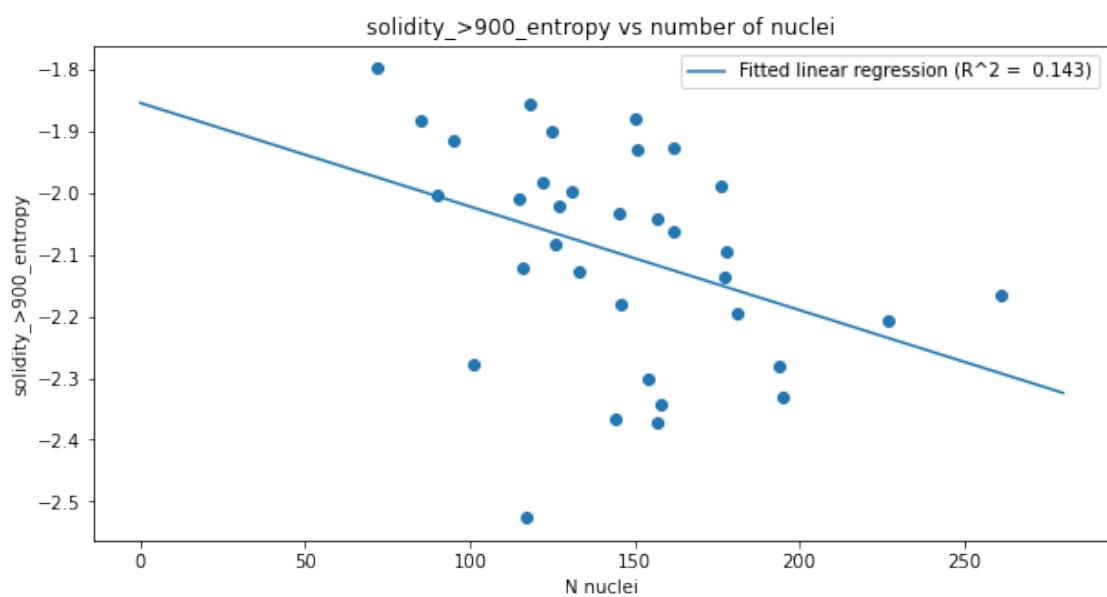
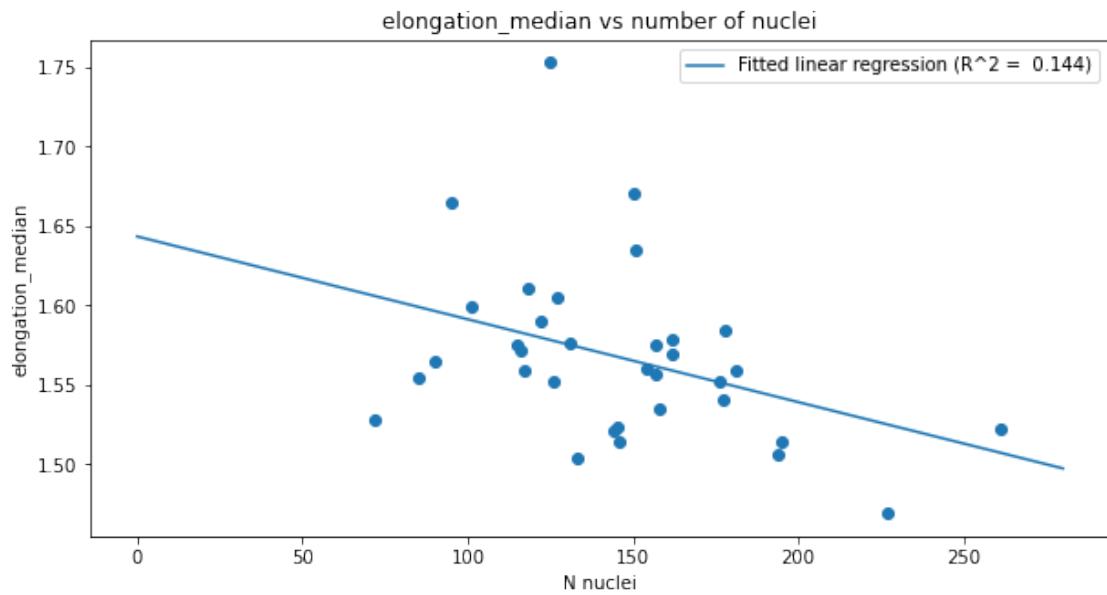


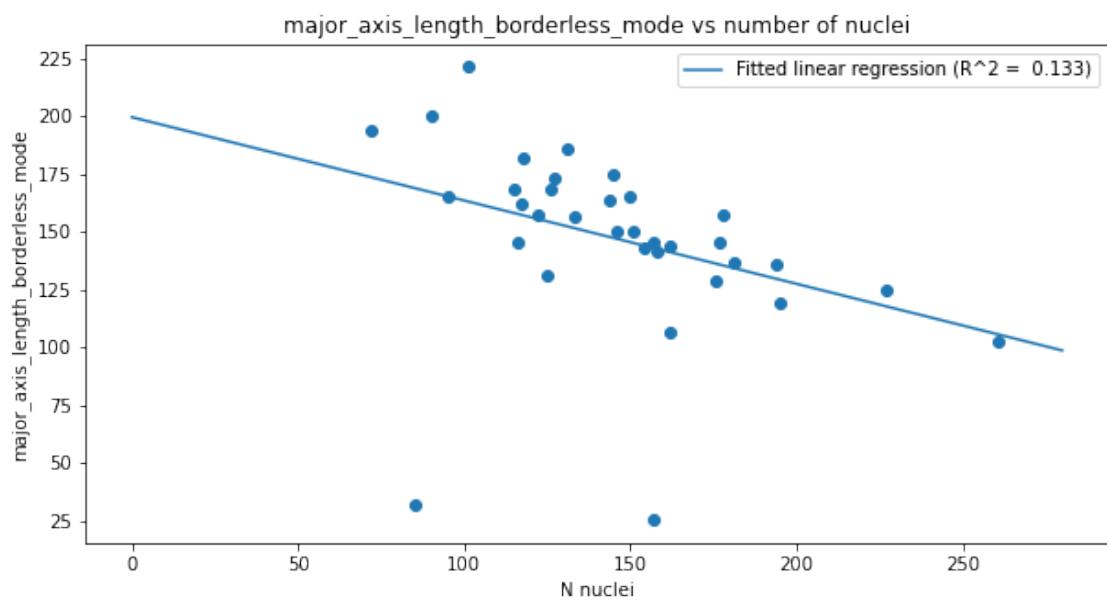
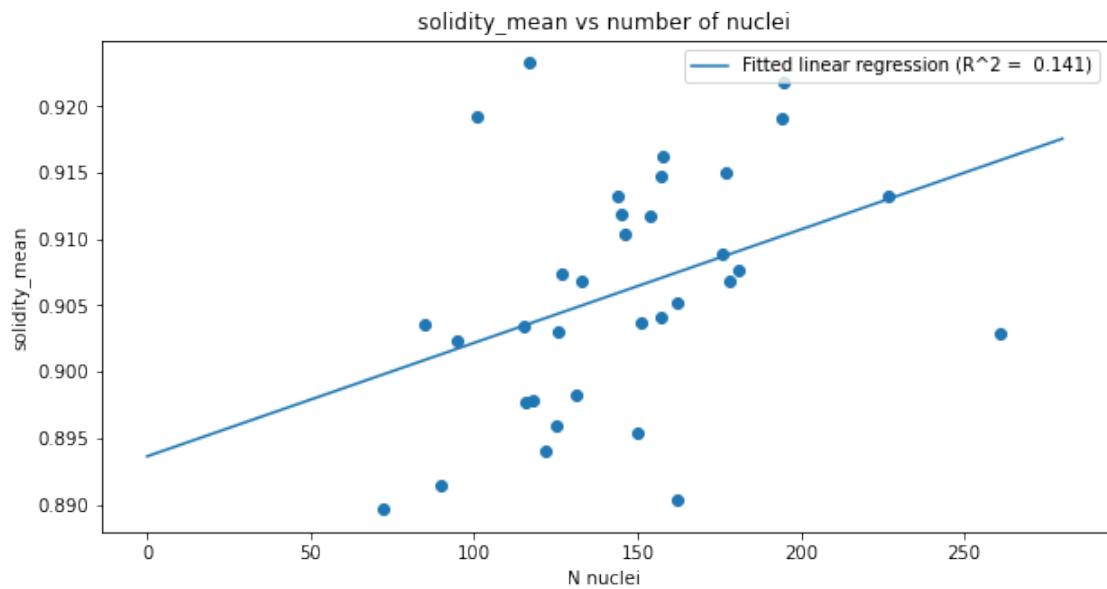


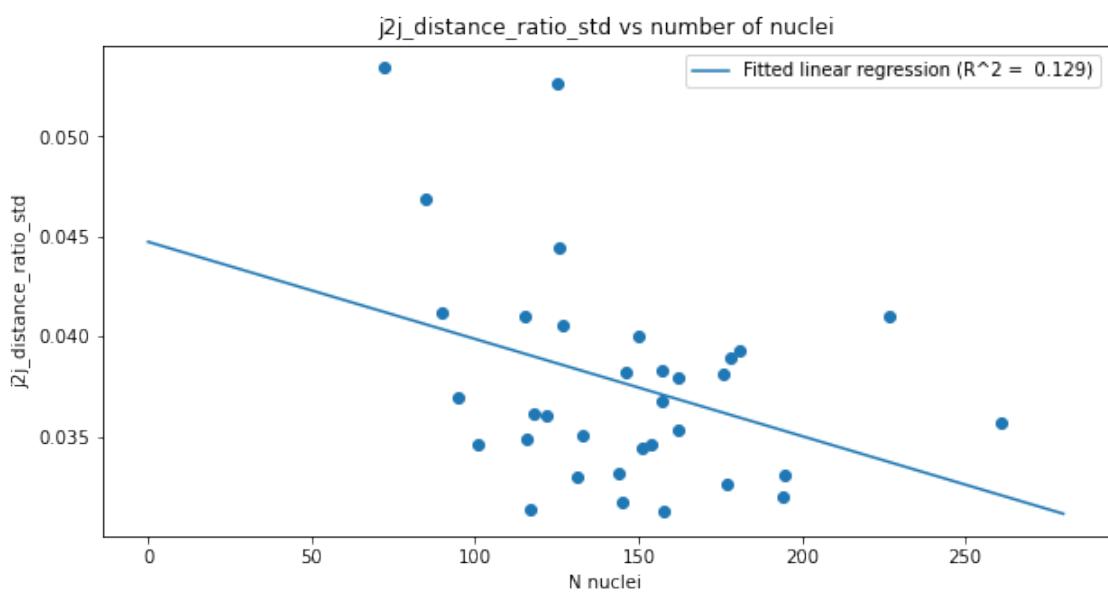
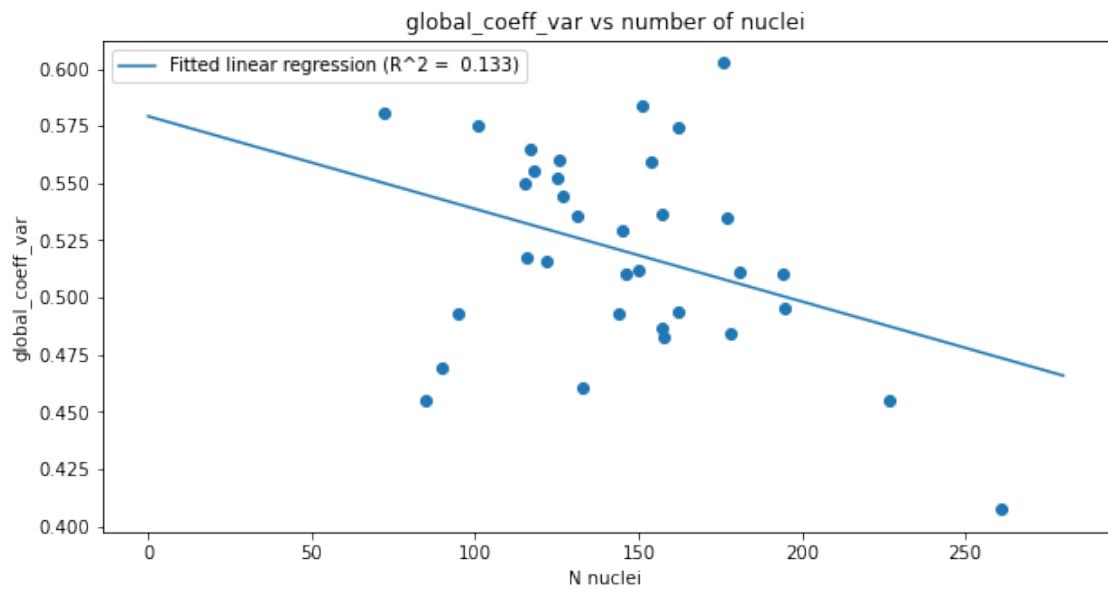


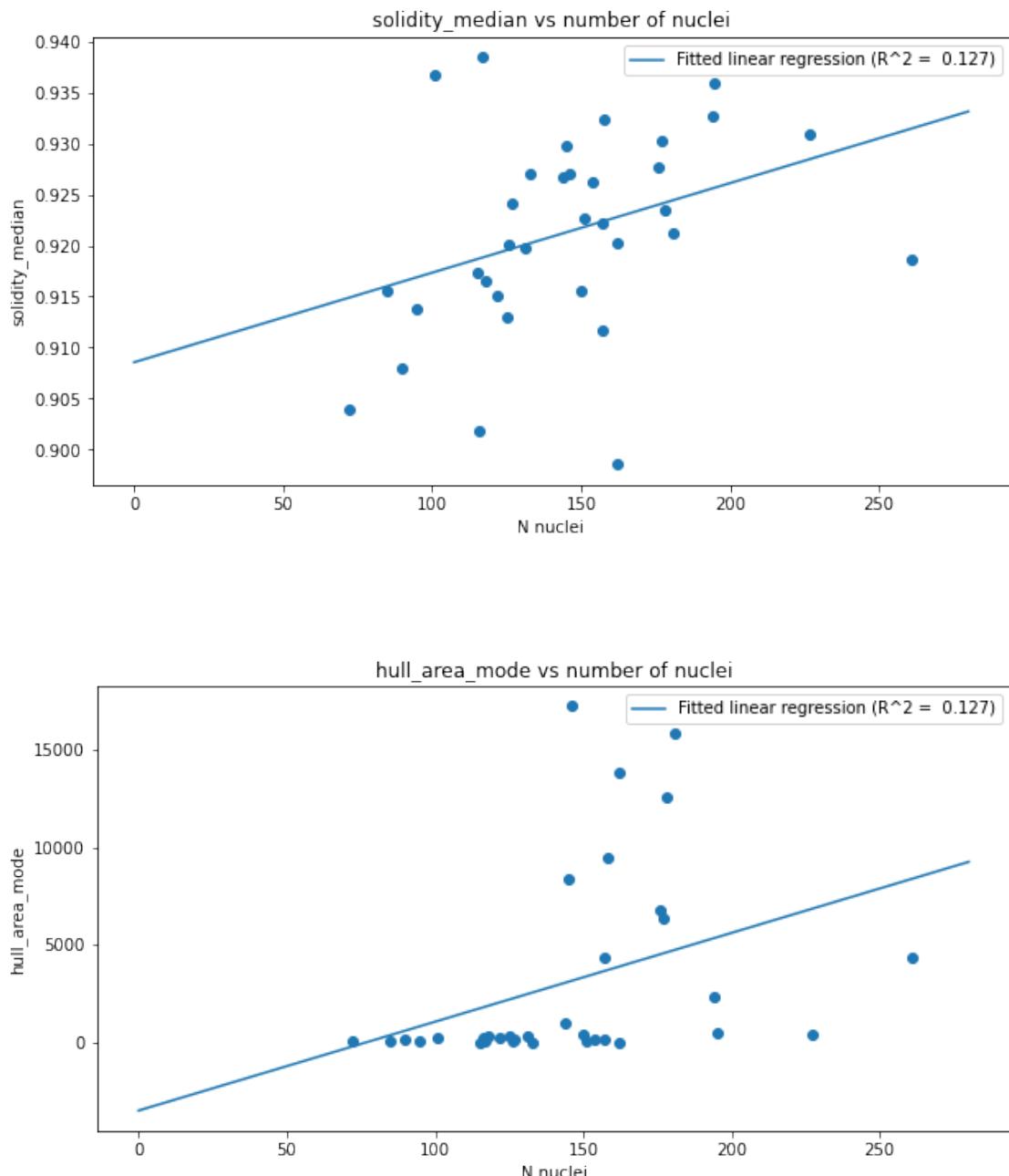




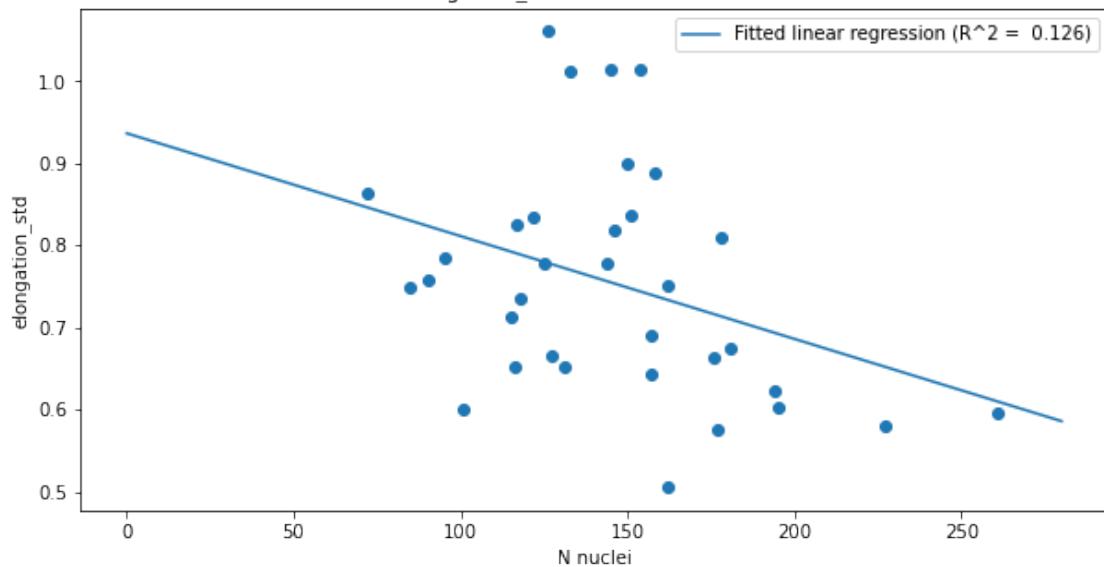




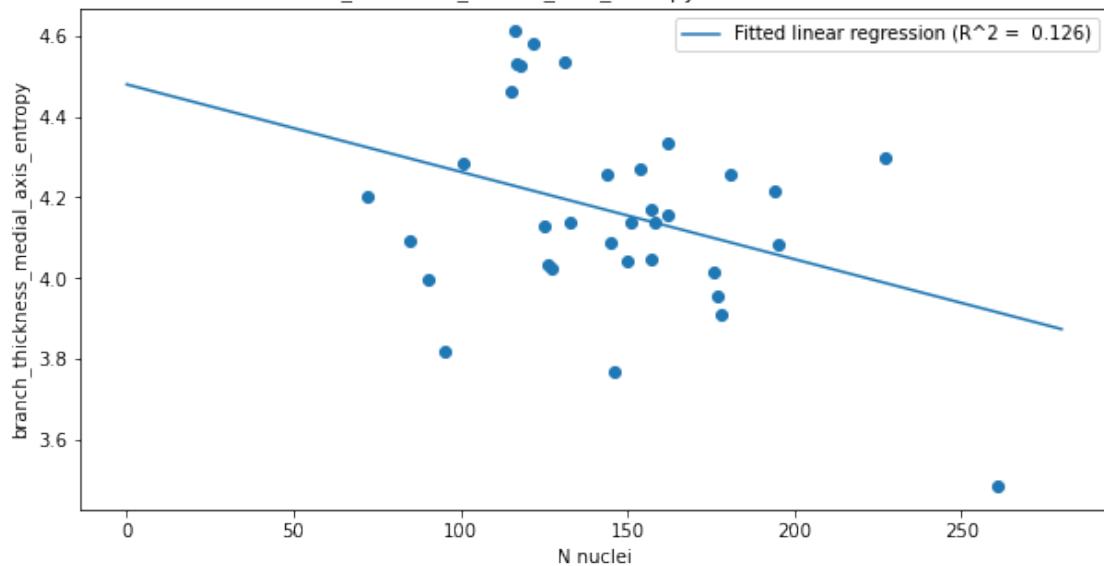


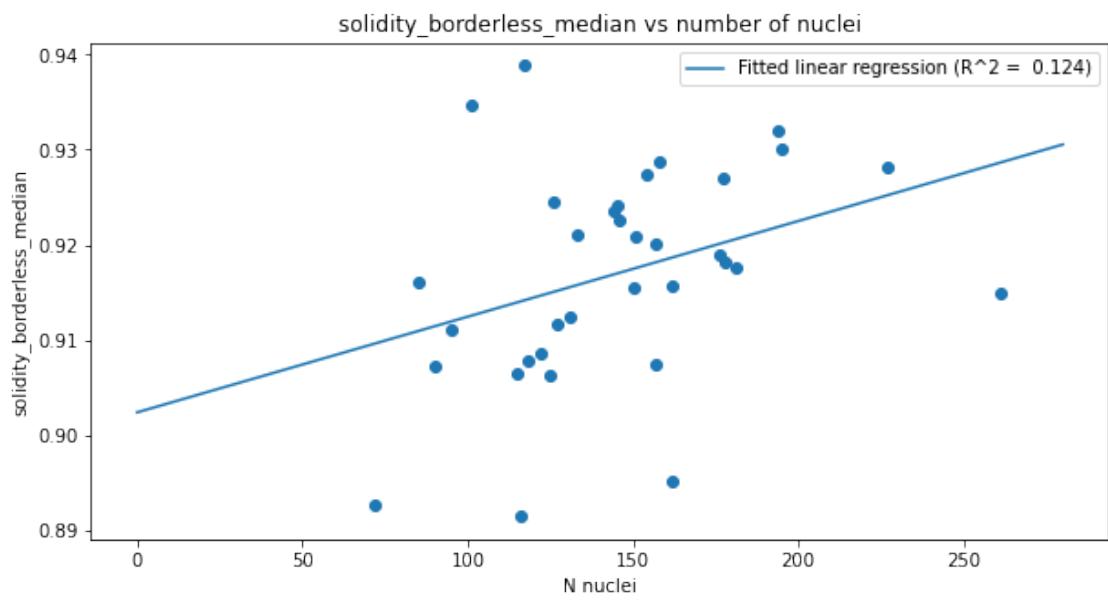
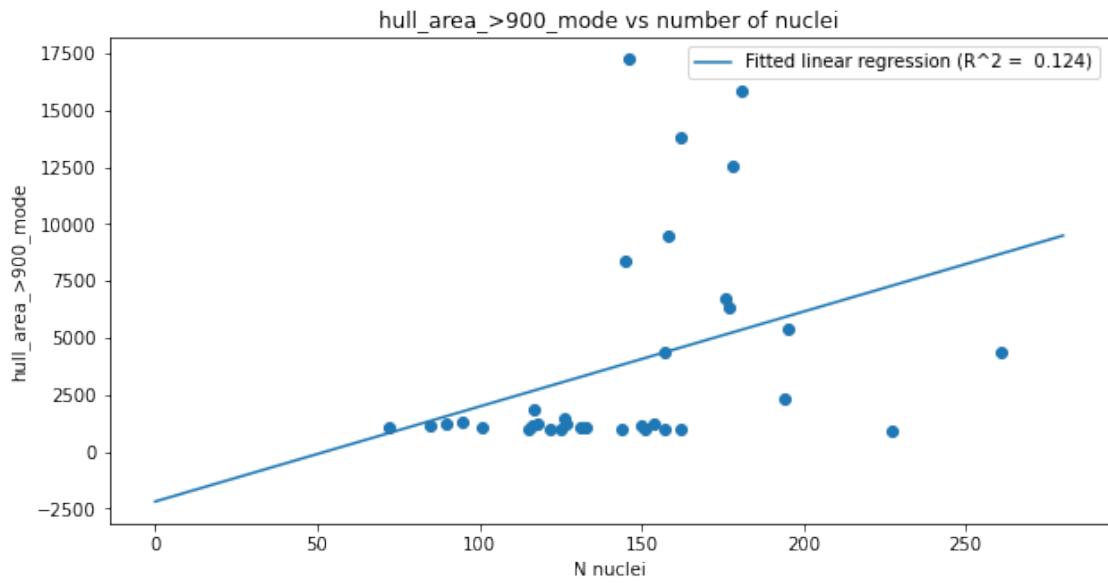


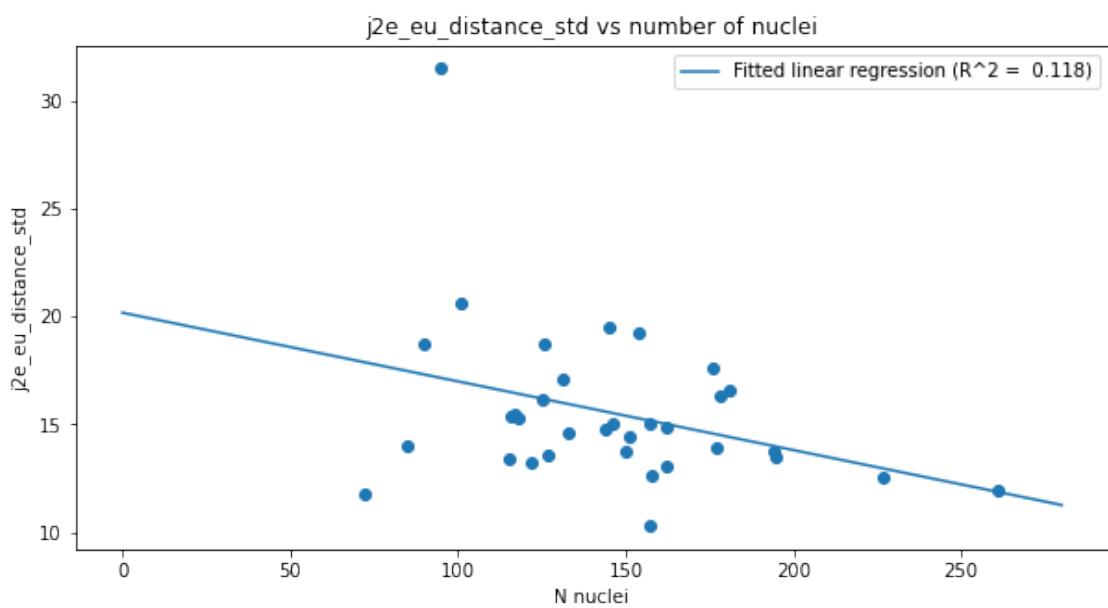
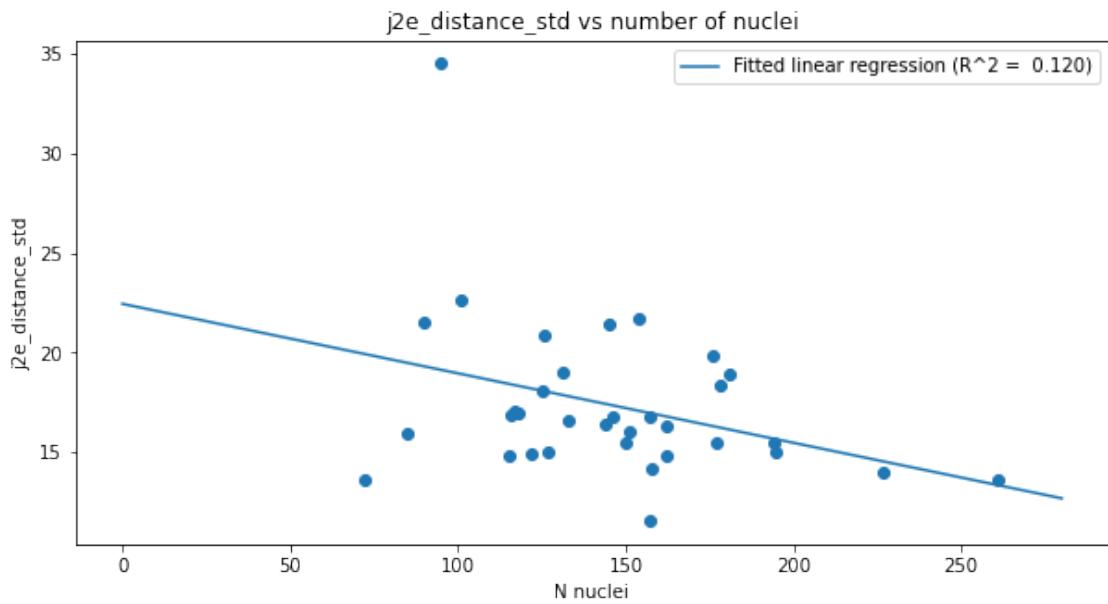
elongation_std vs number of nuclei

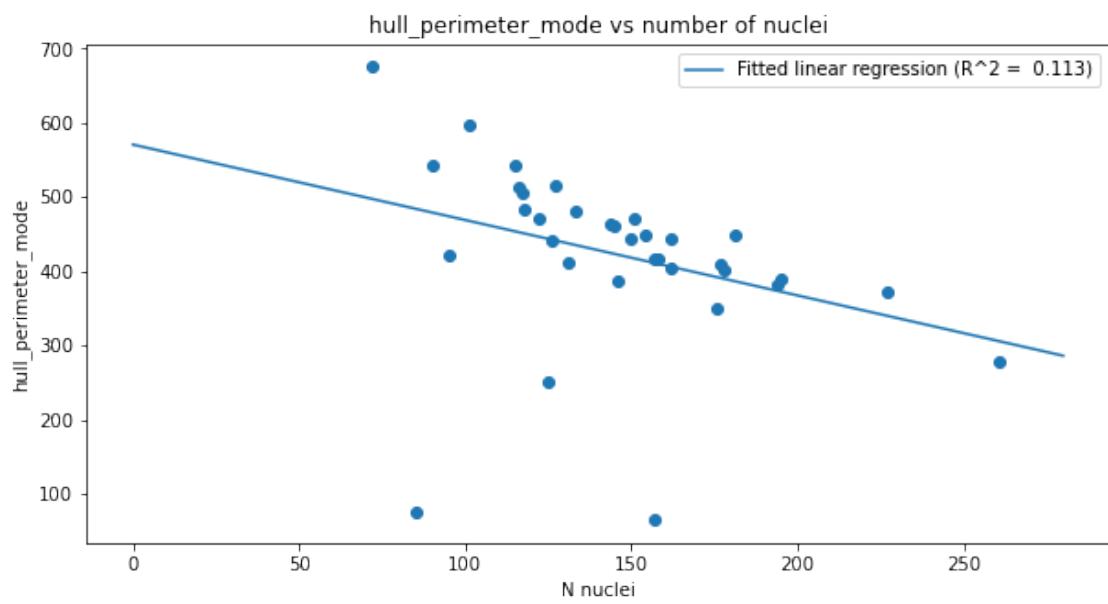
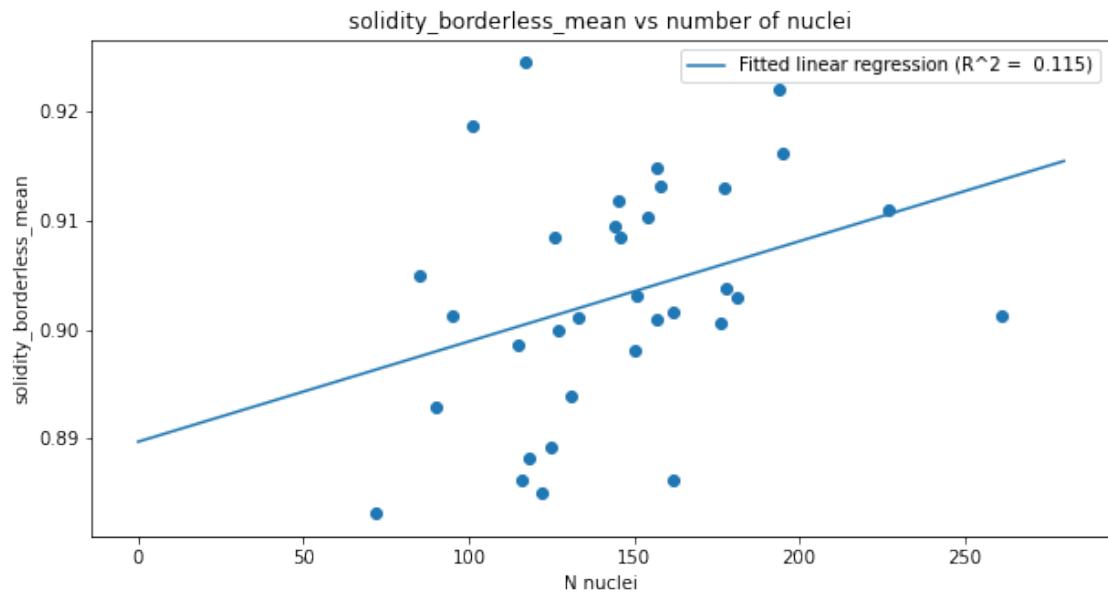


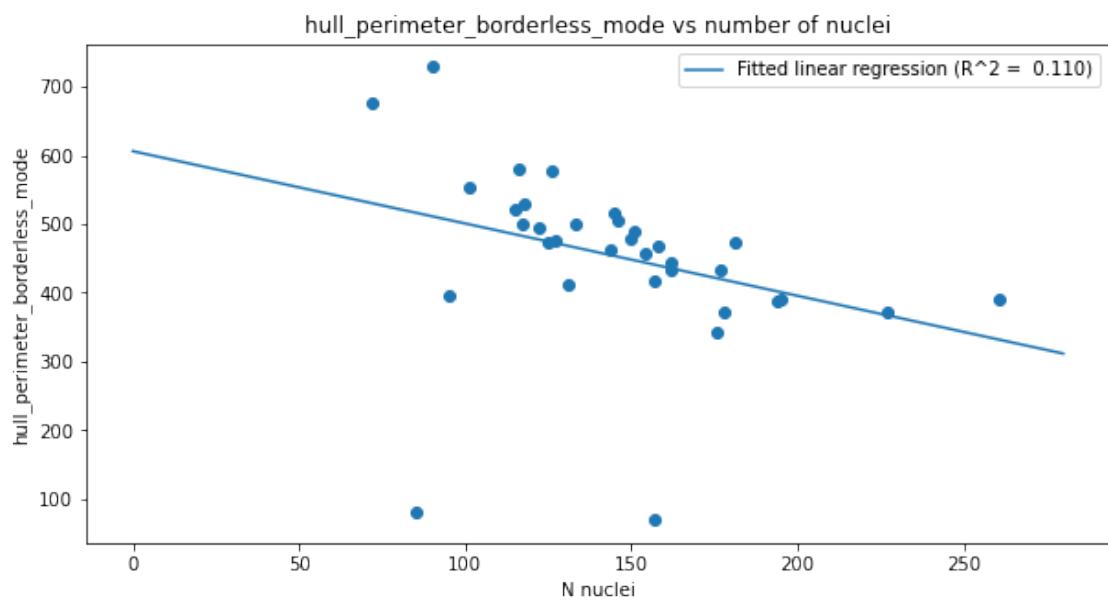
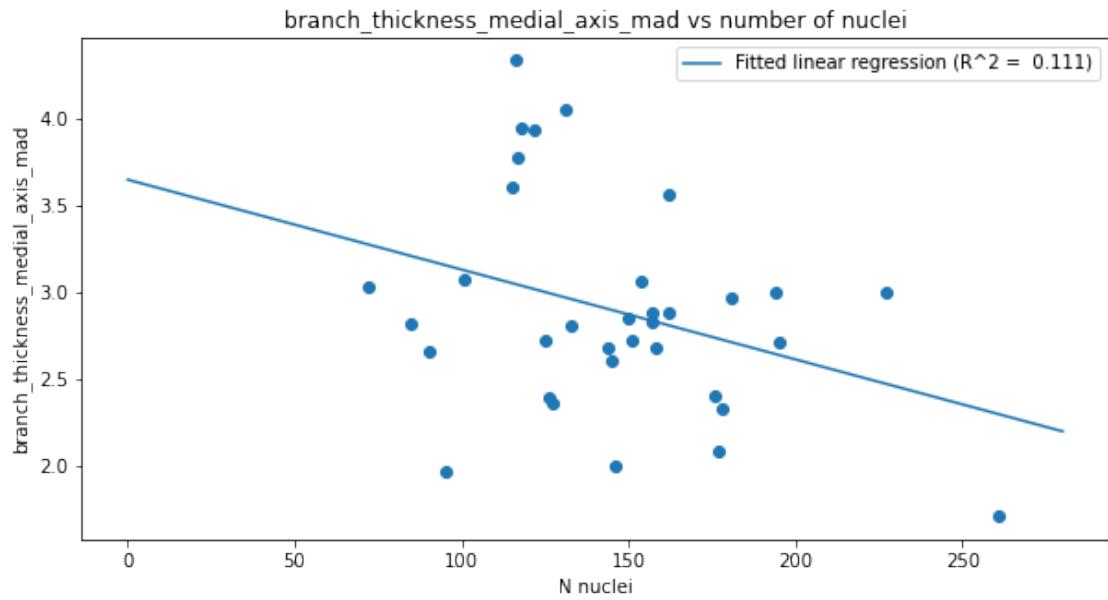
branch_thickness_medial_axis_entropy vs number of nuclei

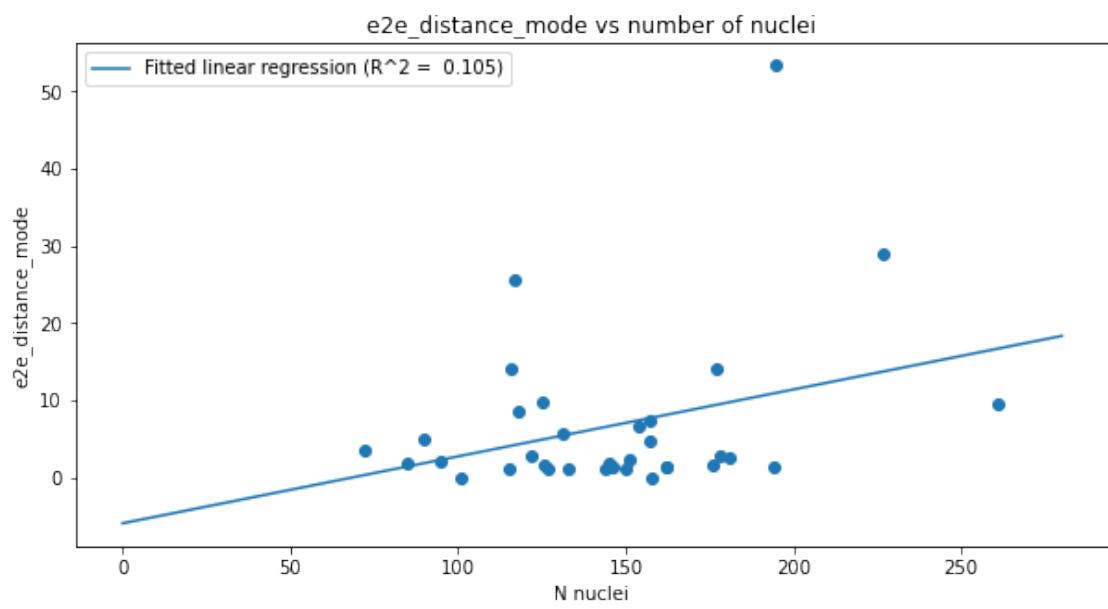
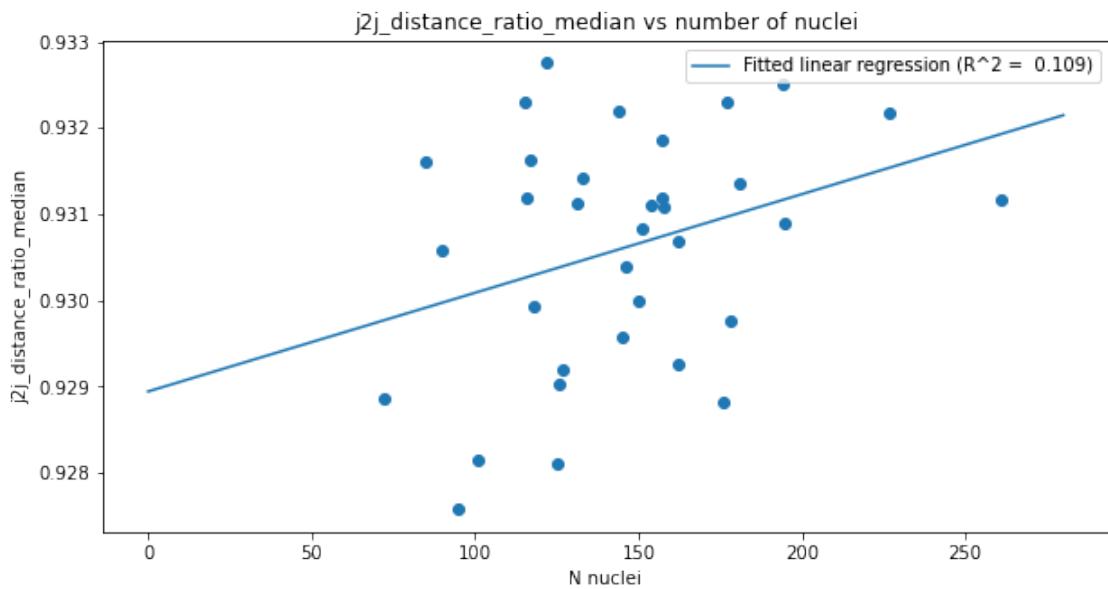


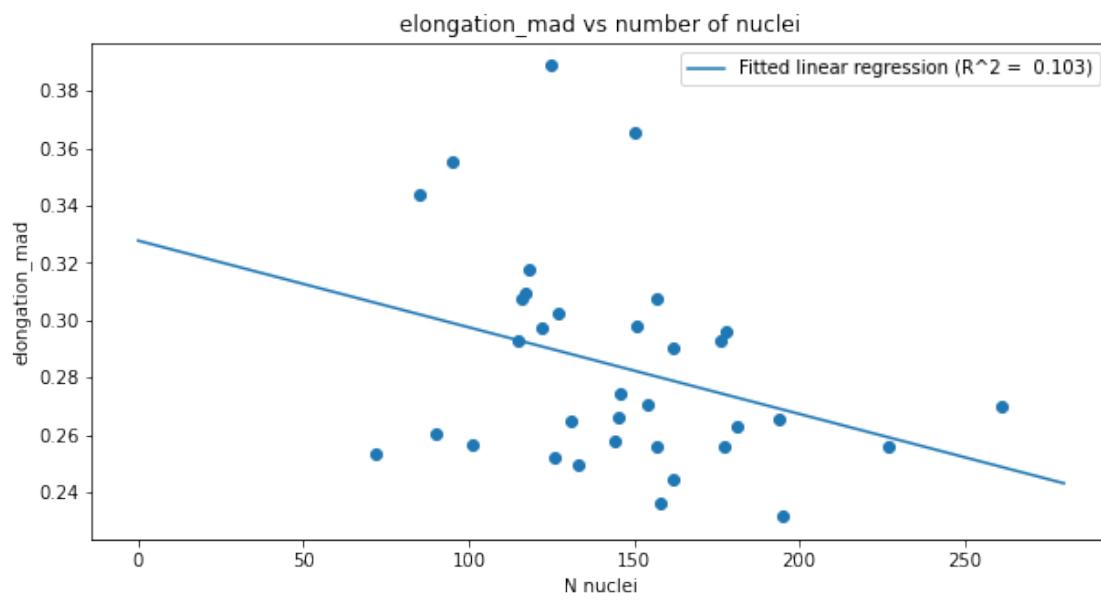
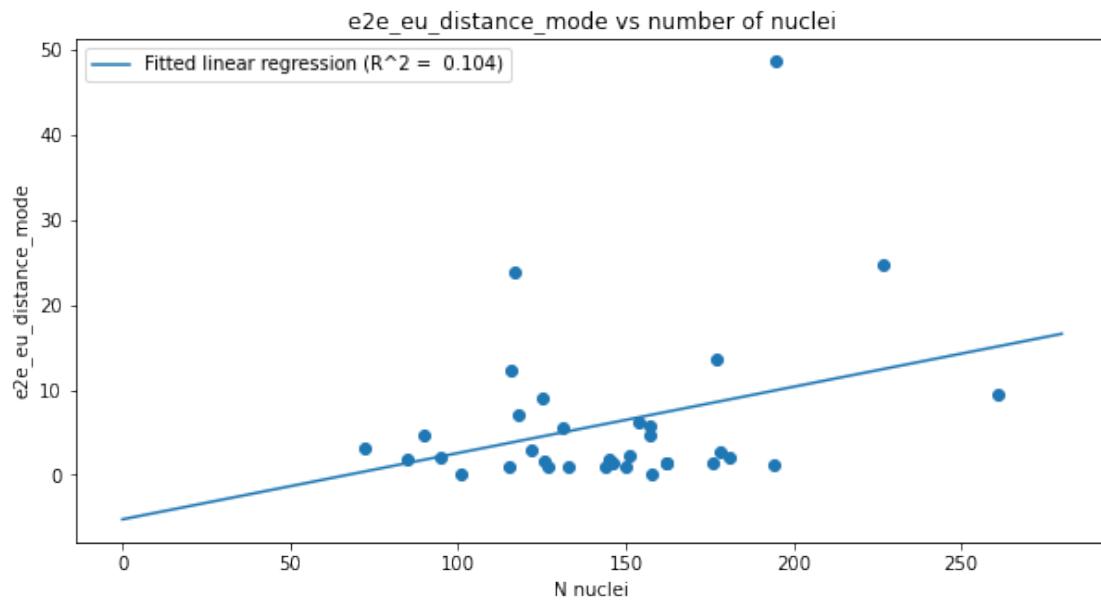


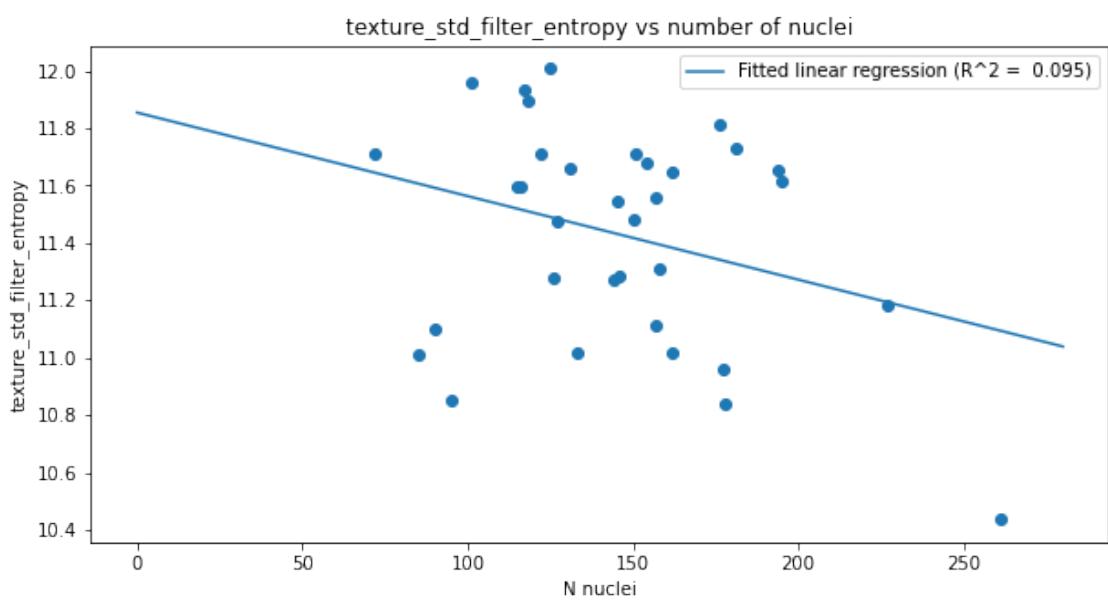
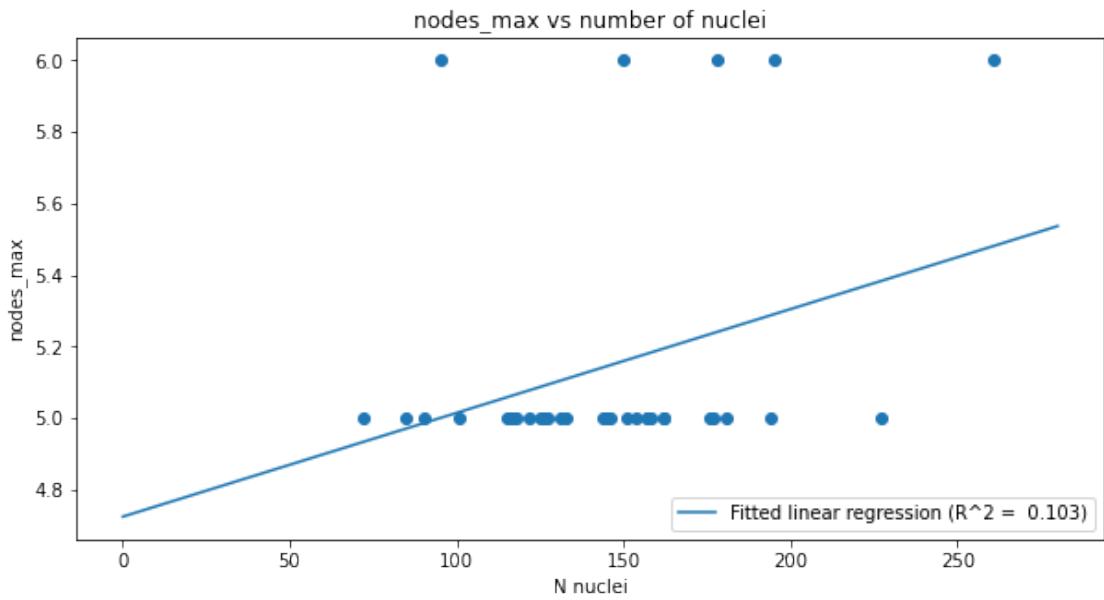


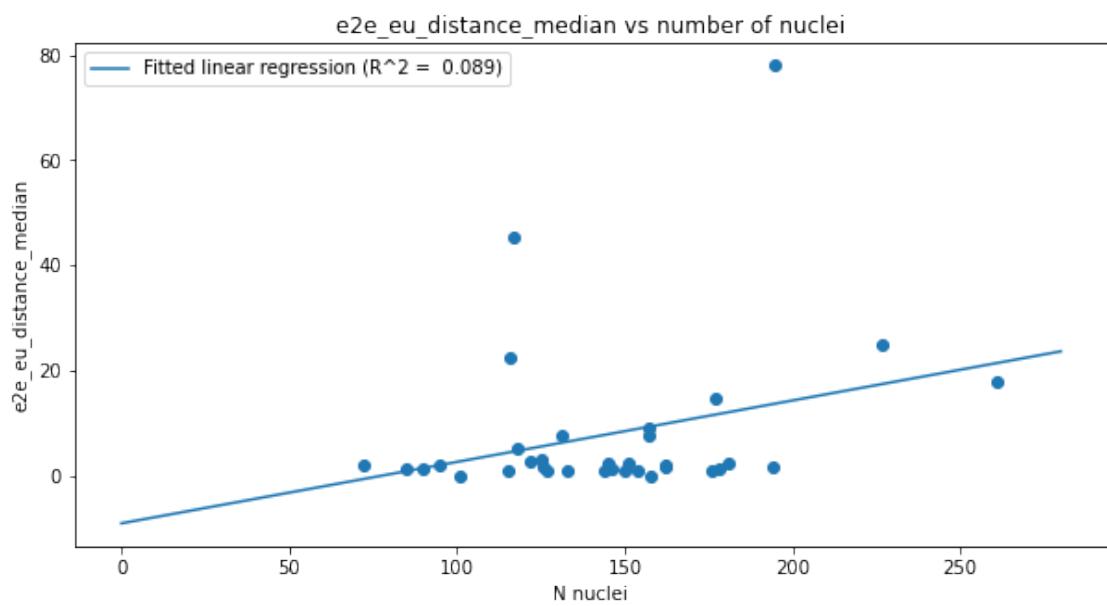
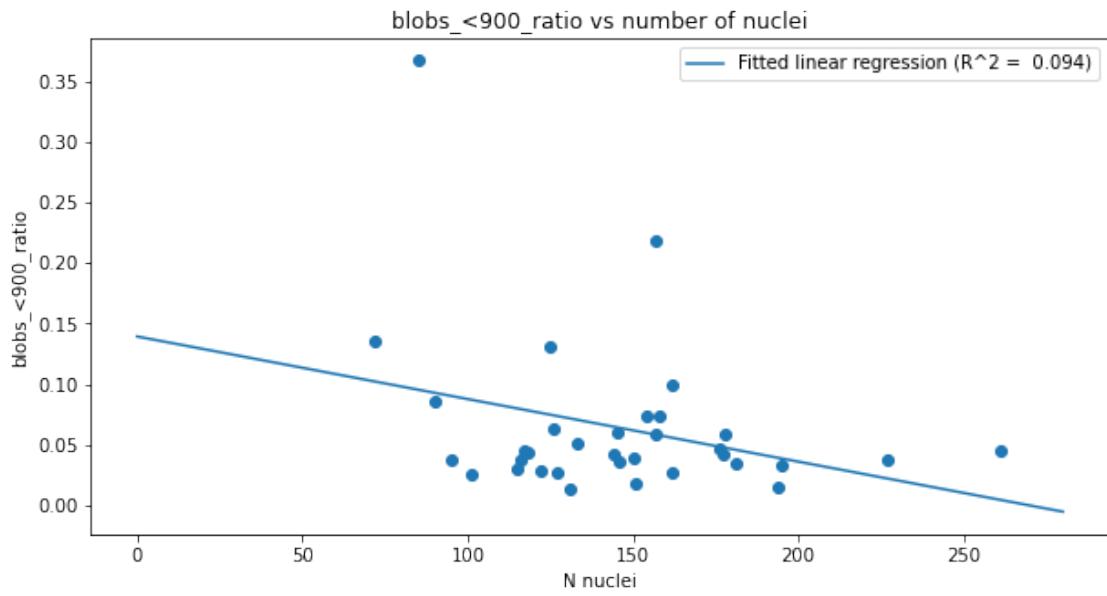


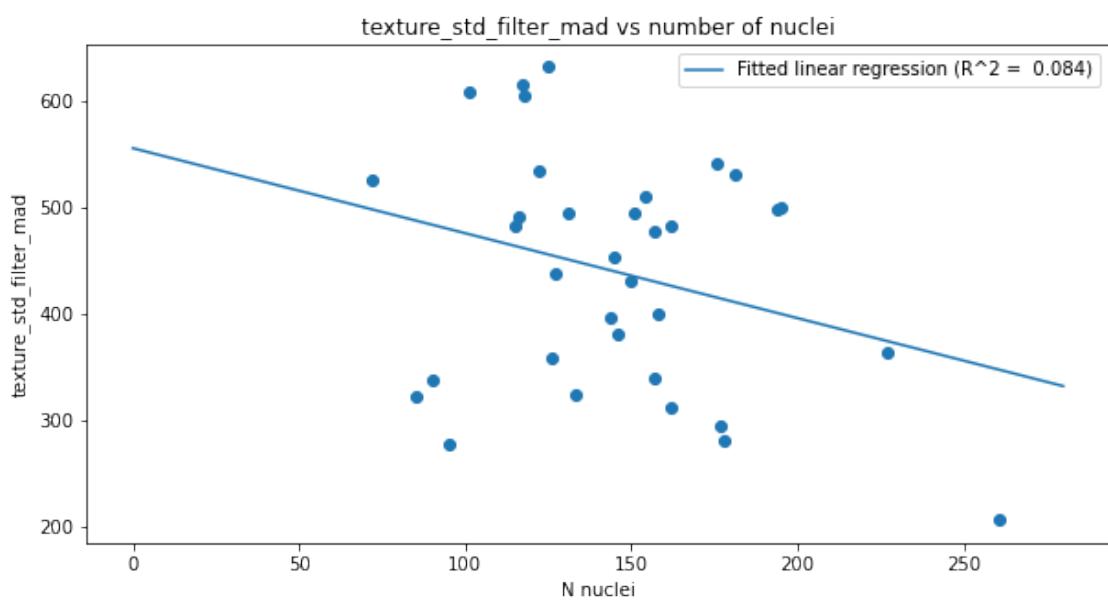
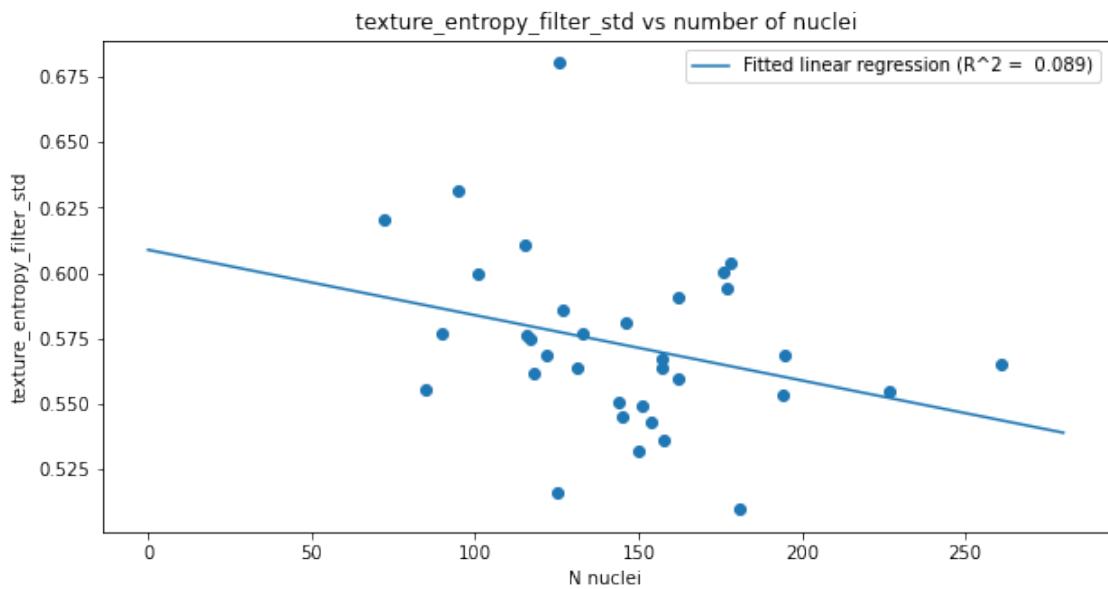


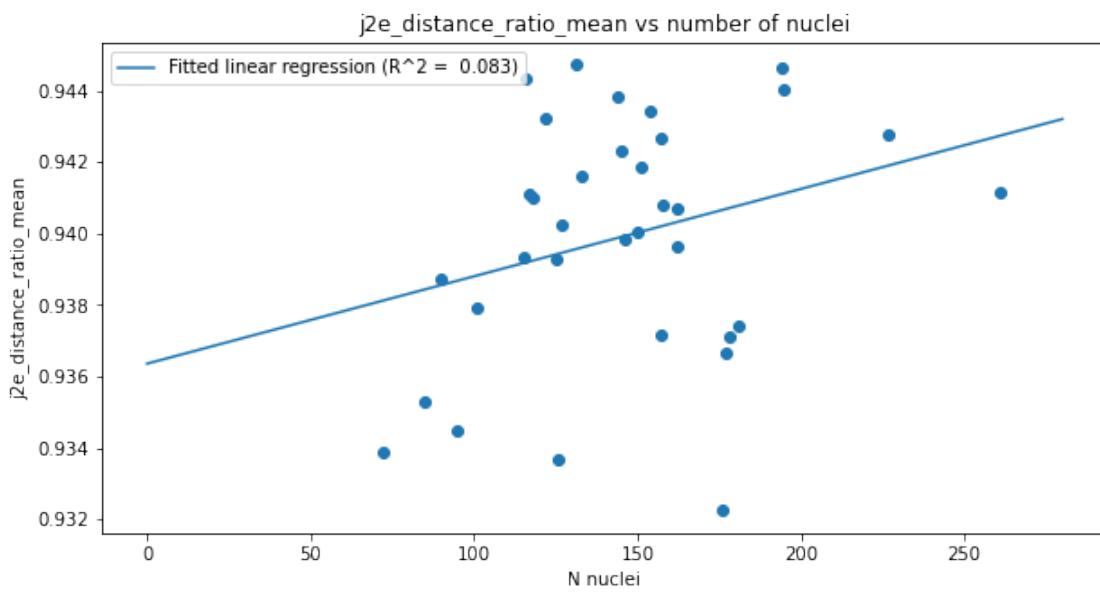
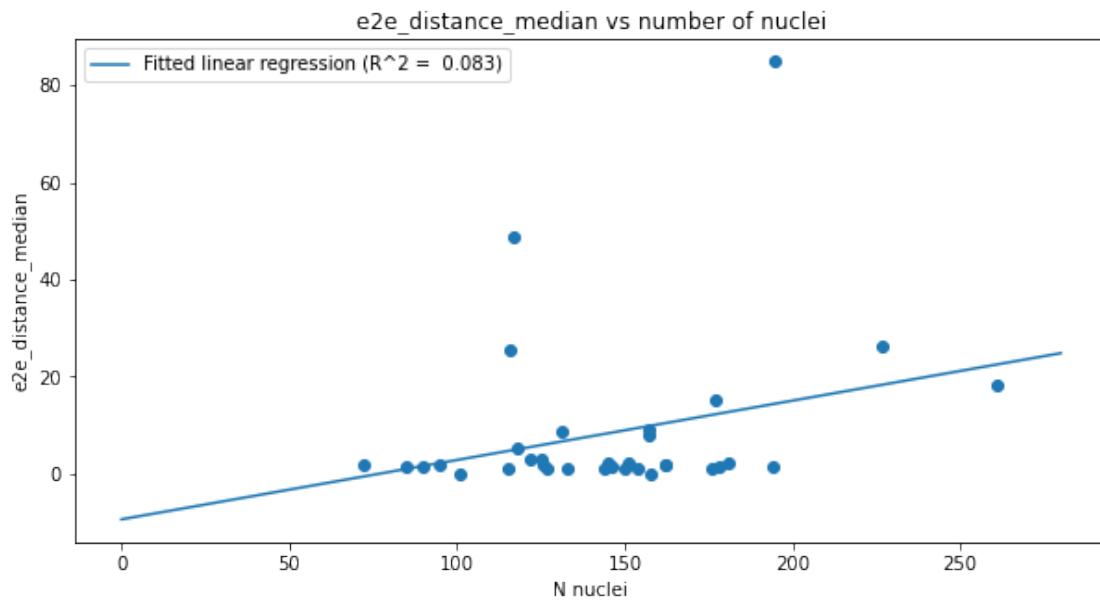




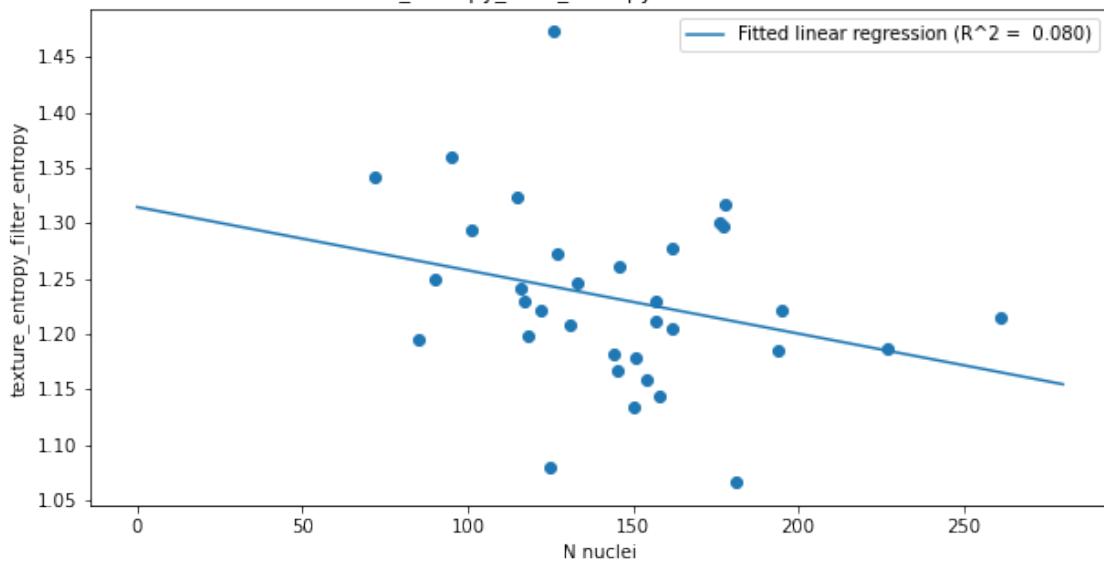




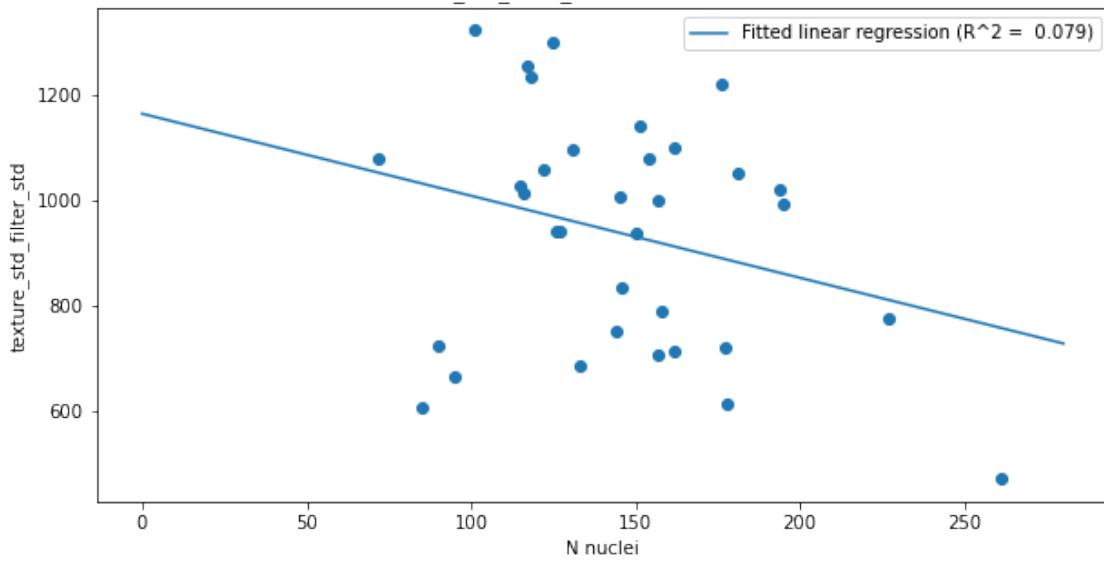


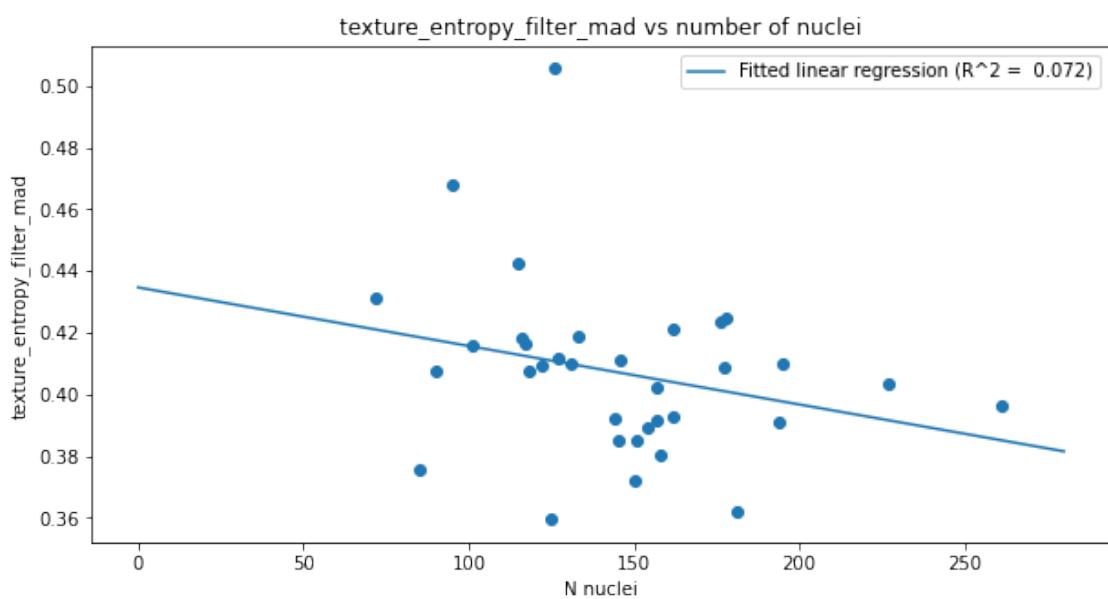
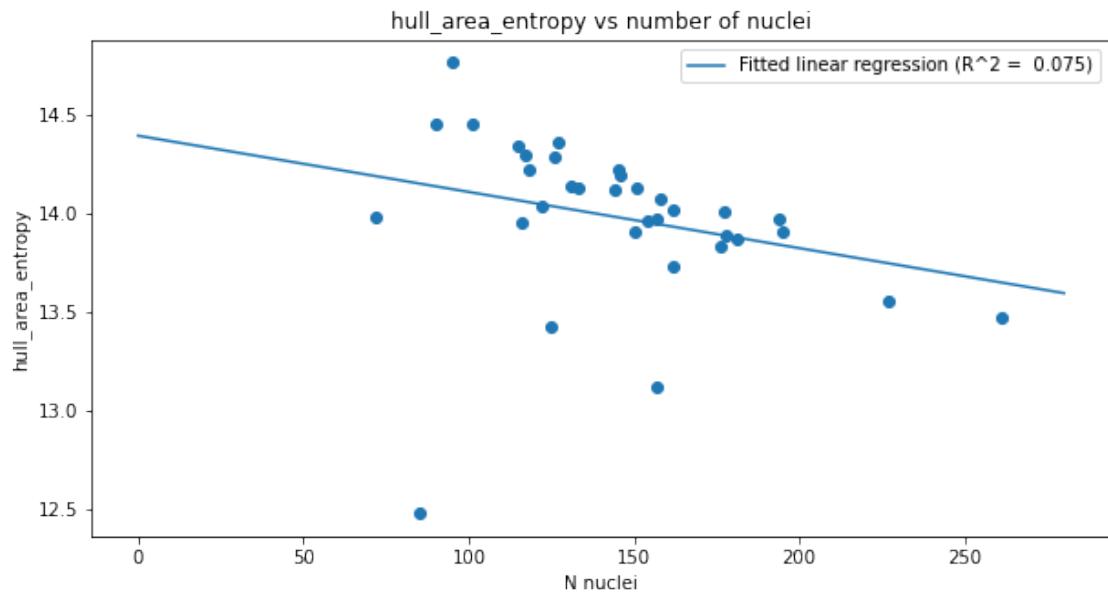


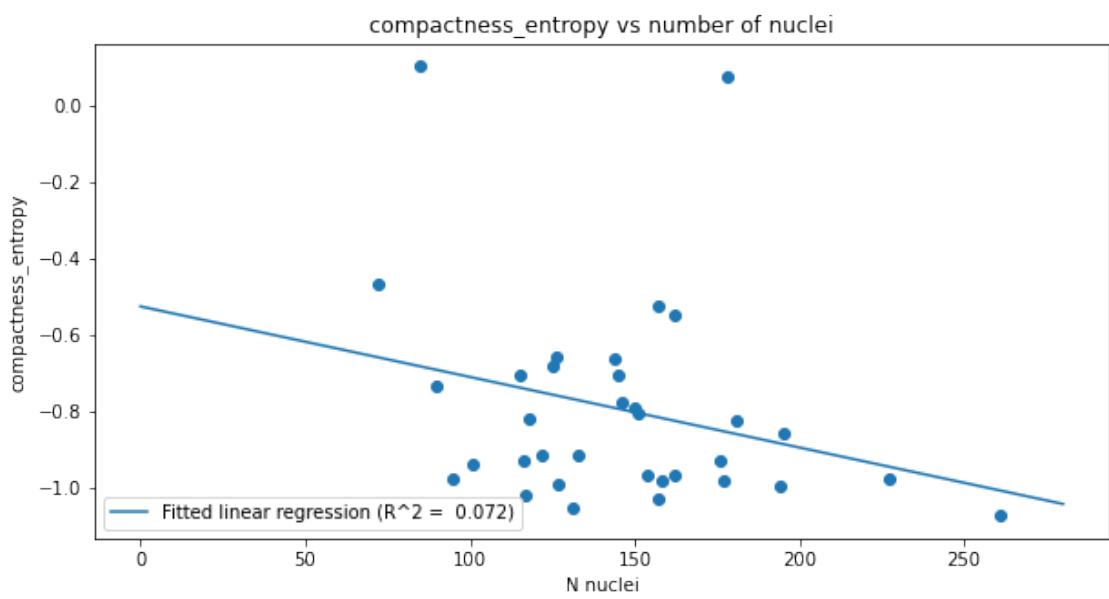
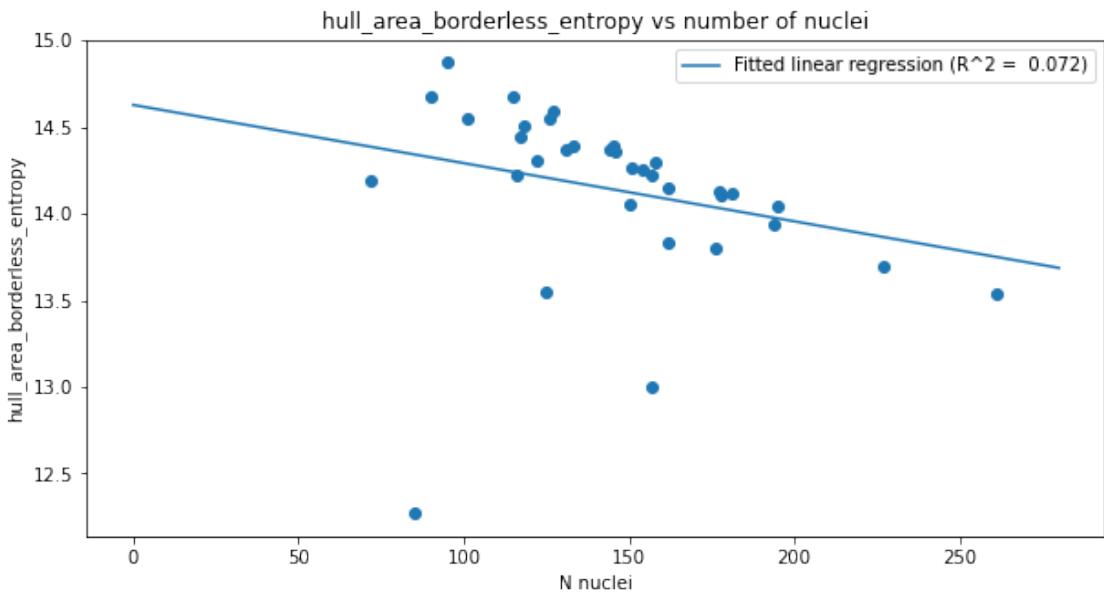
texture_entropy_filter_entropy vs number of nuclei

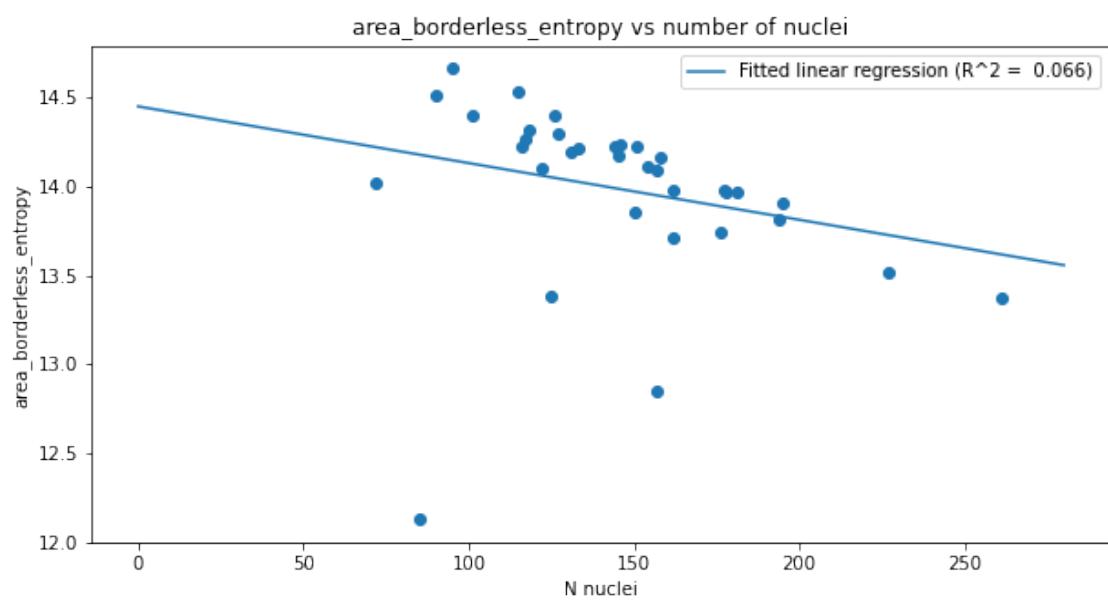
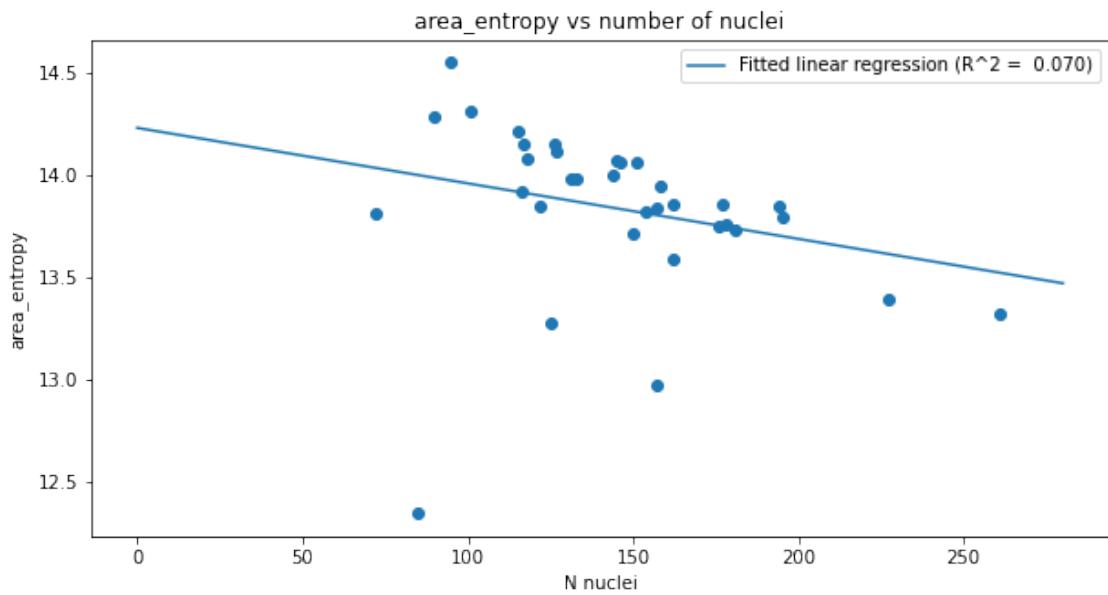


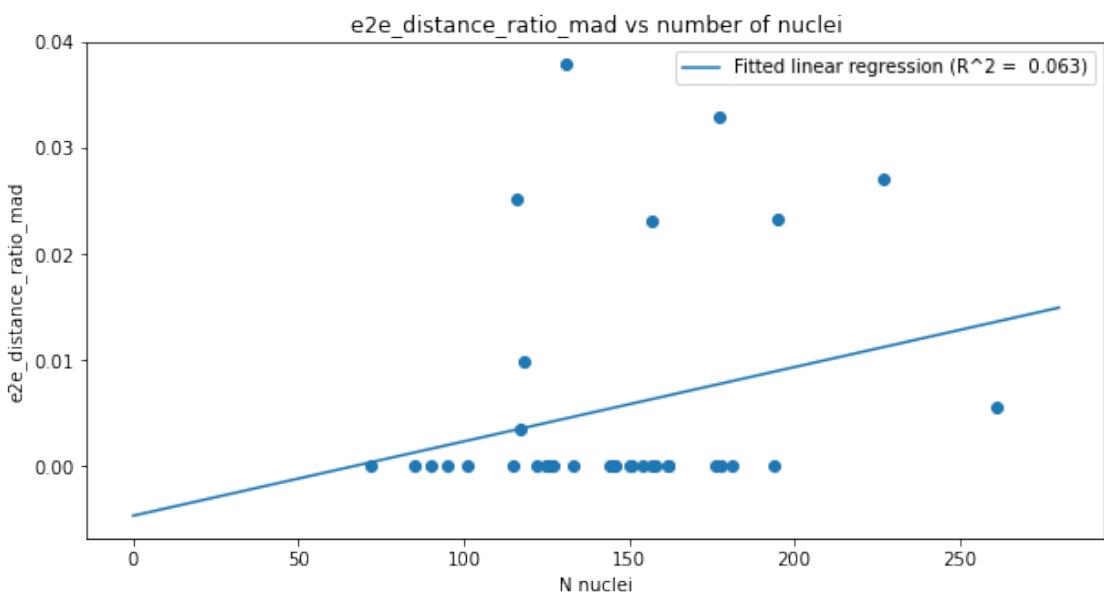
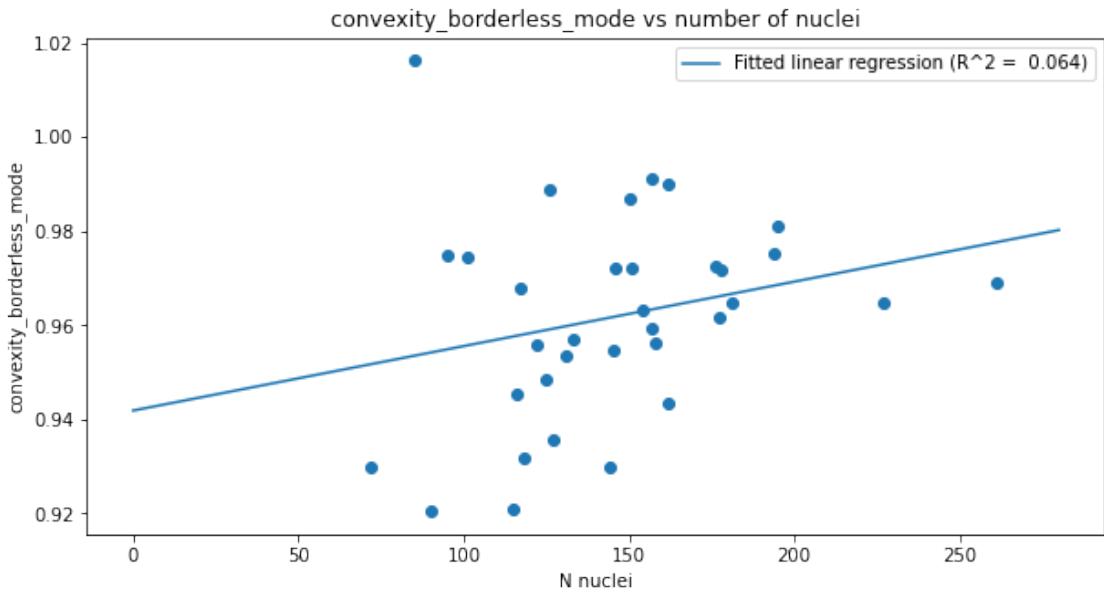
texture_std_filter_std vs number of nuclei

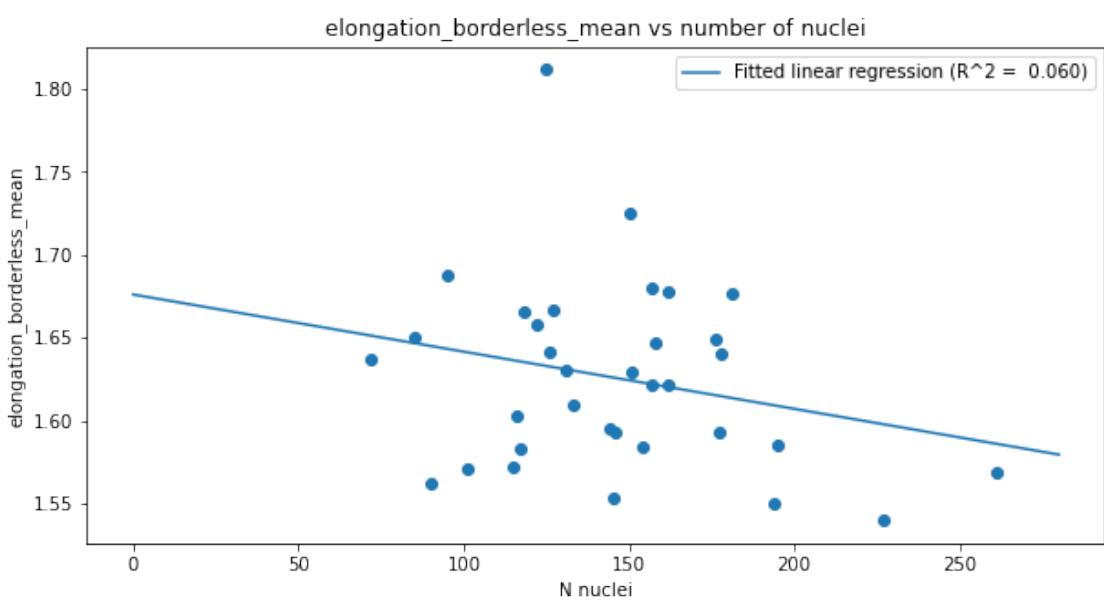
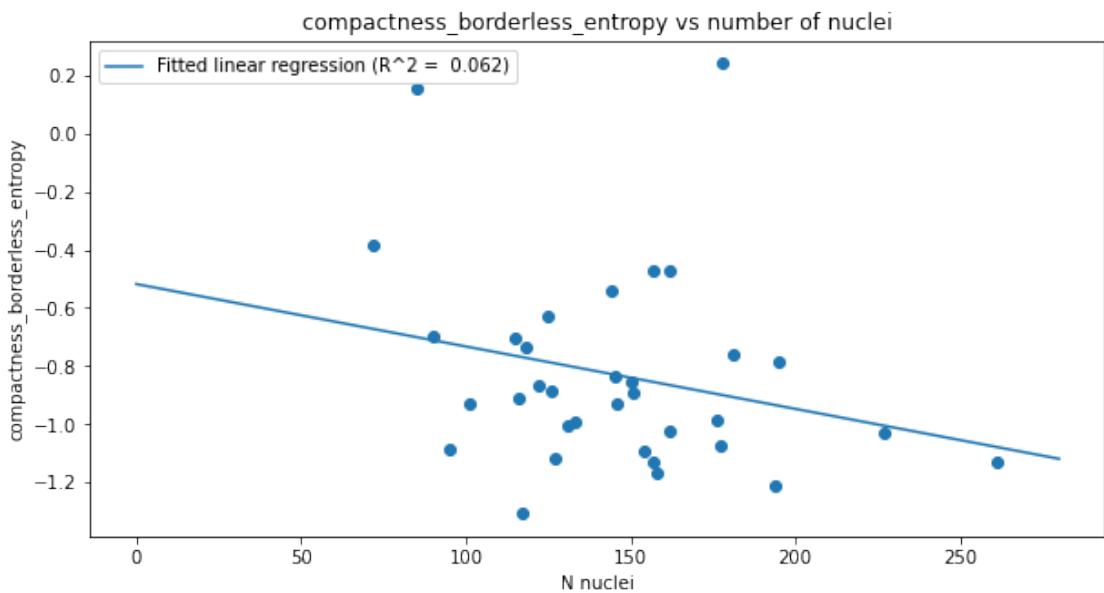


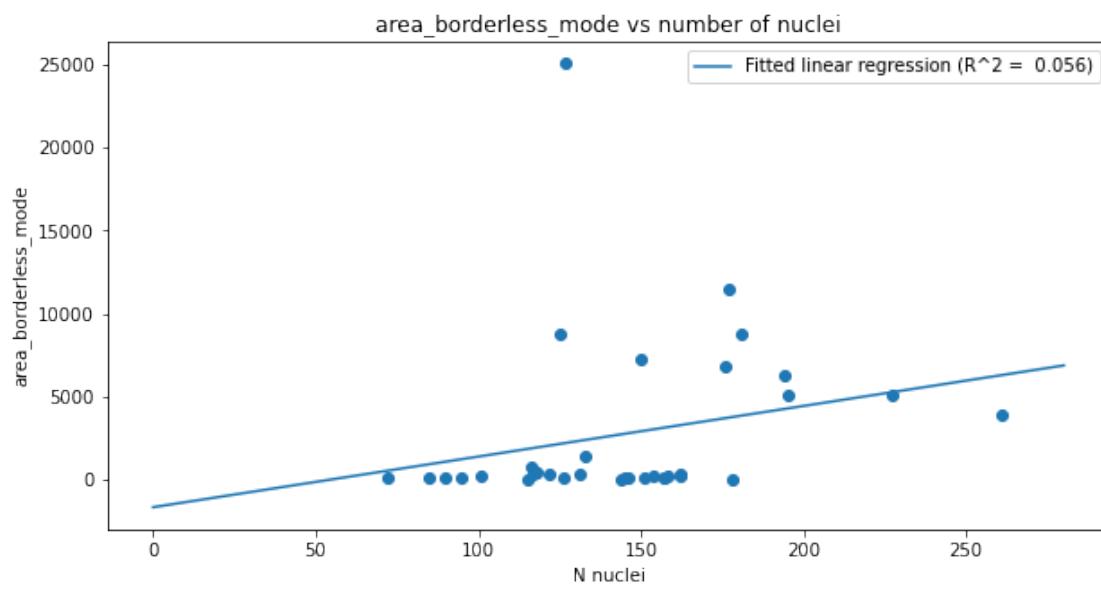
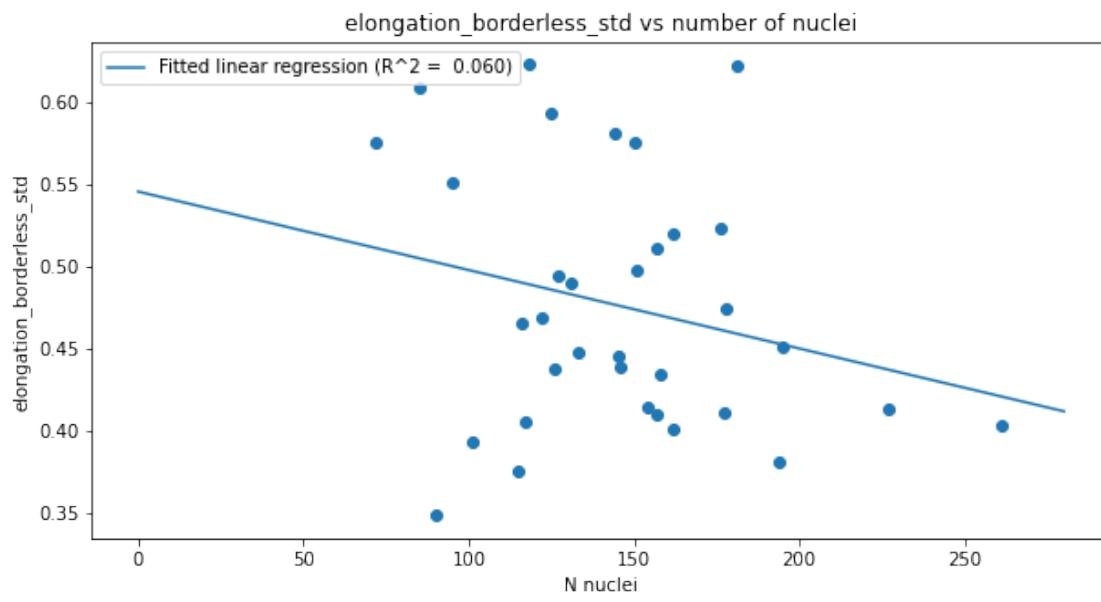


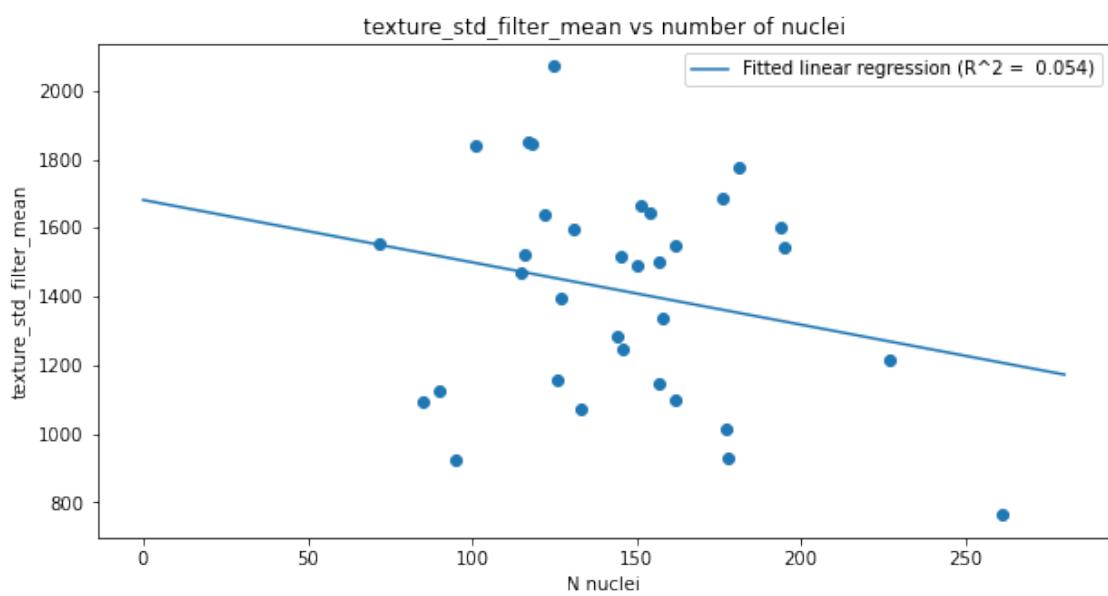
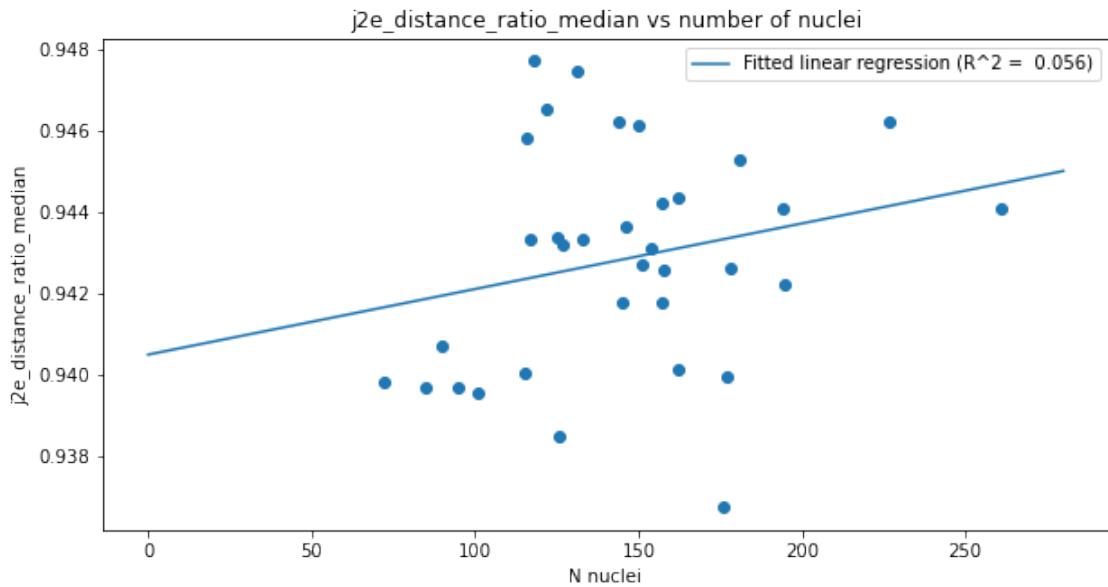


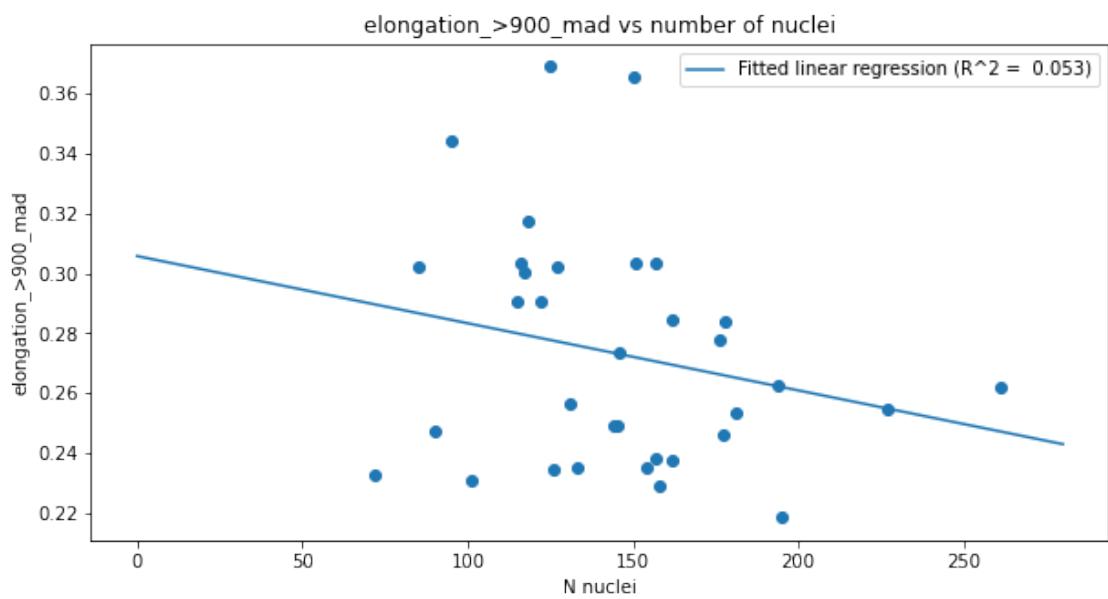
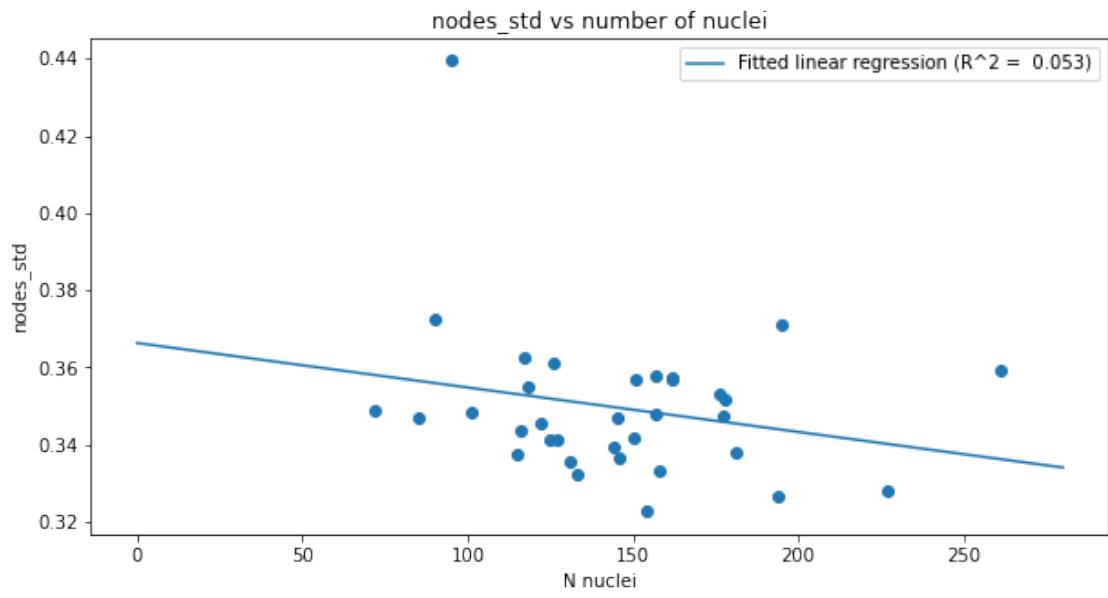




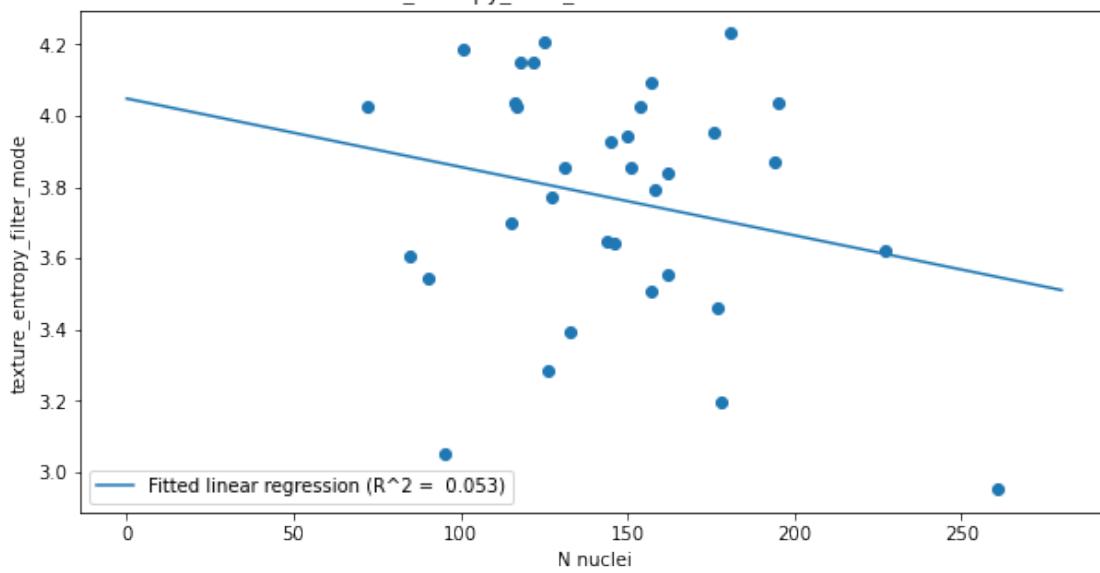




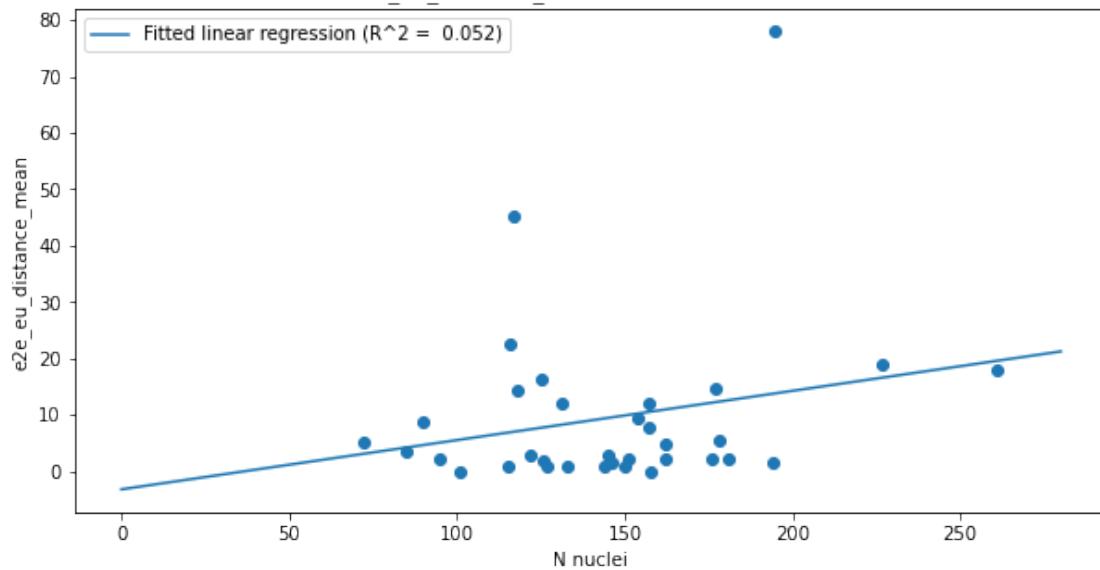


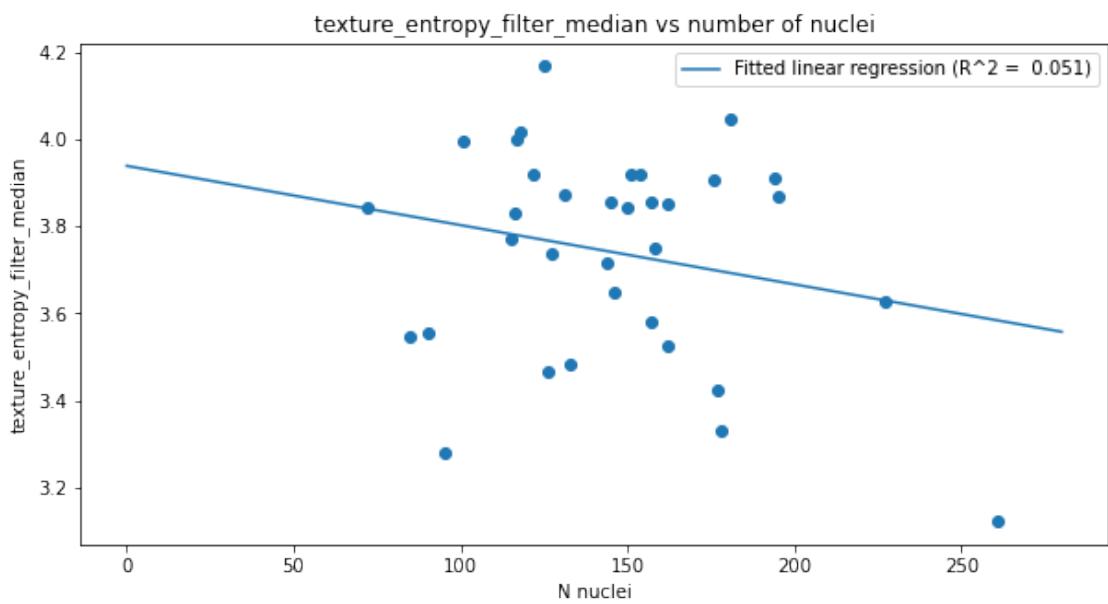
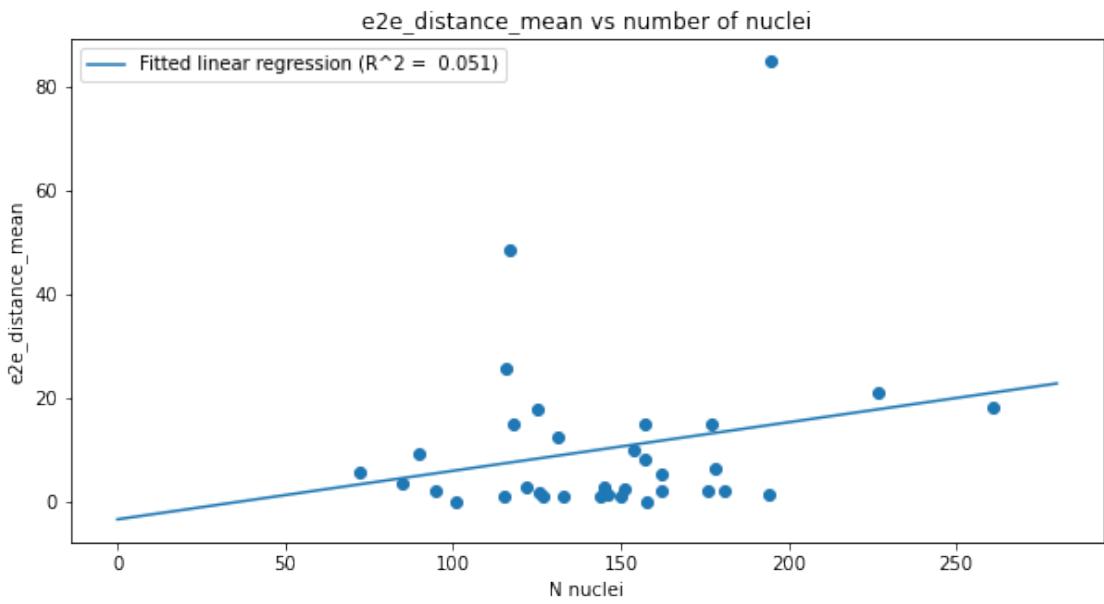


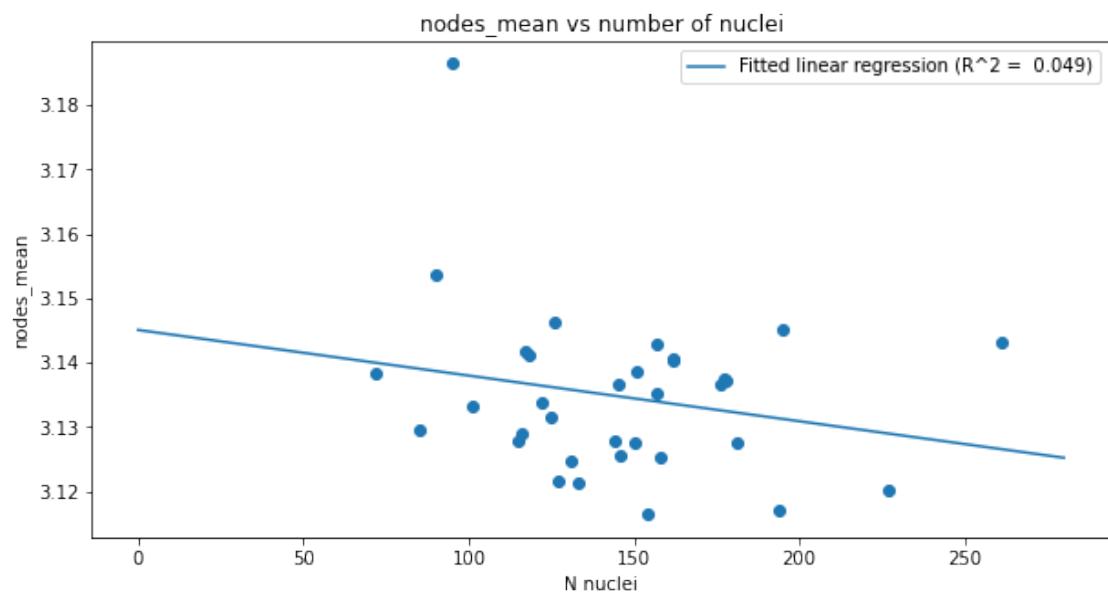
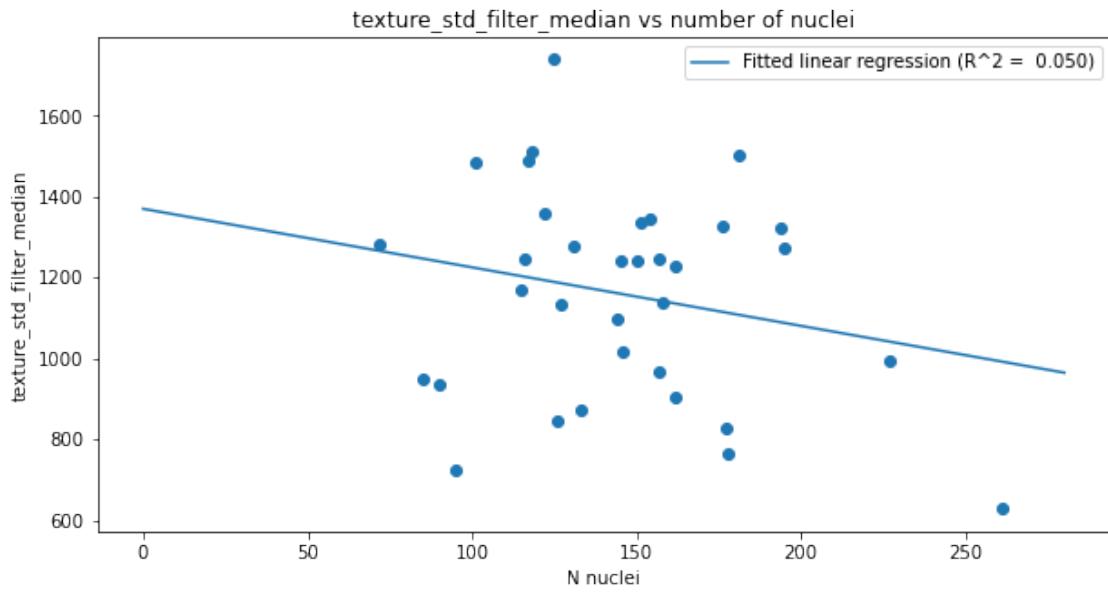
texture_entropy_filter_mode vs number of nuclei

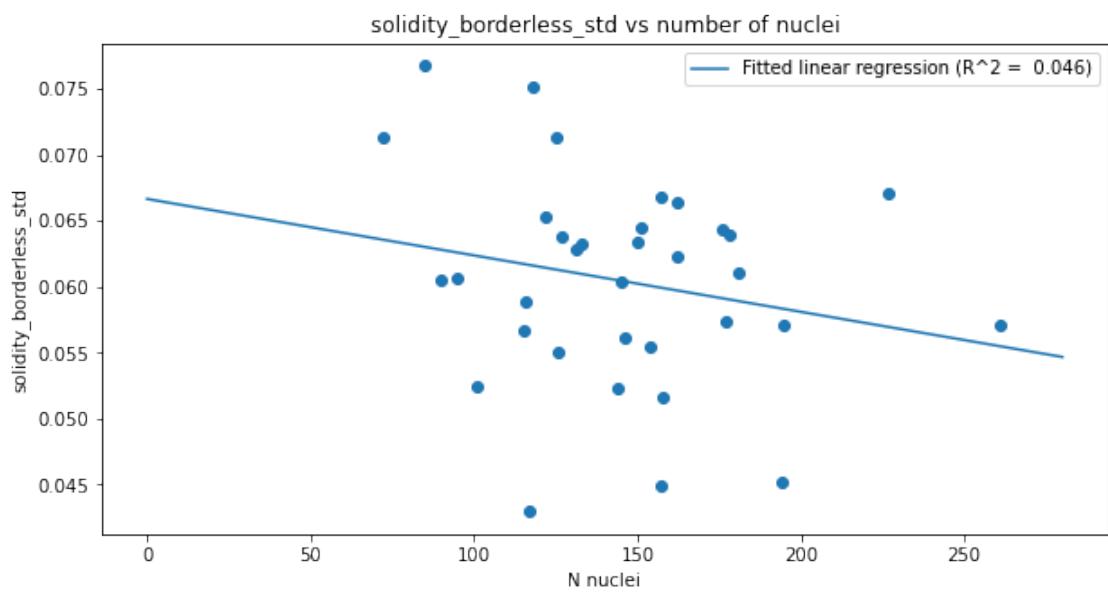
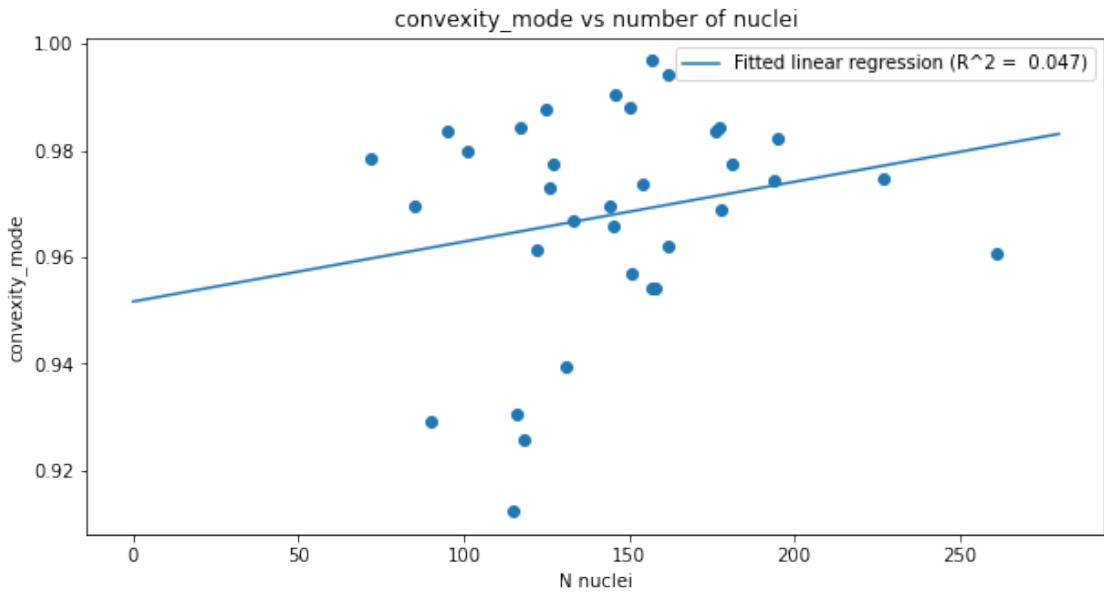


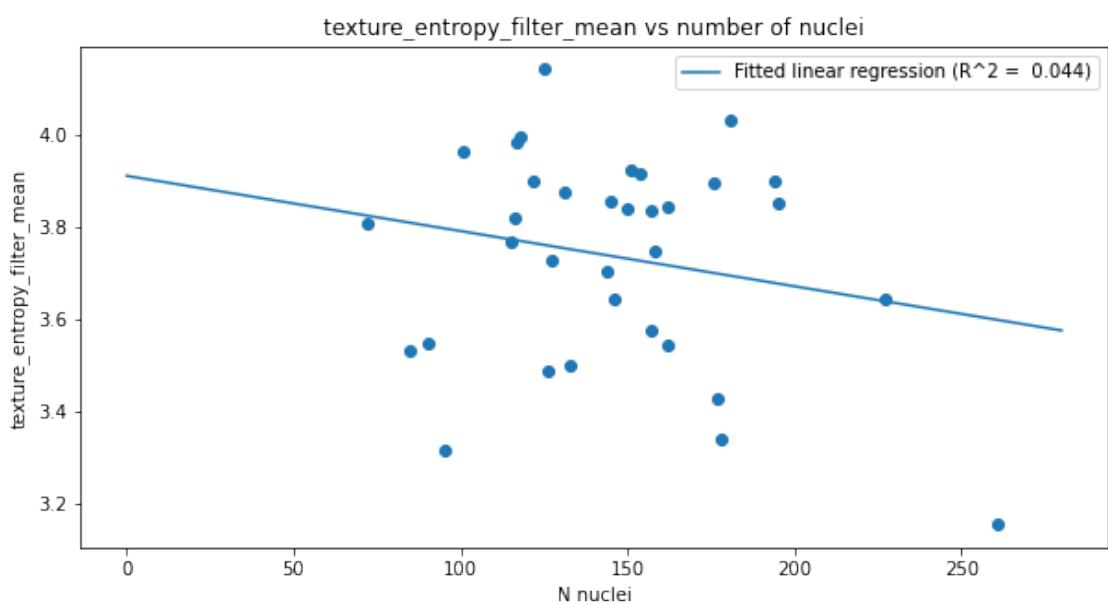
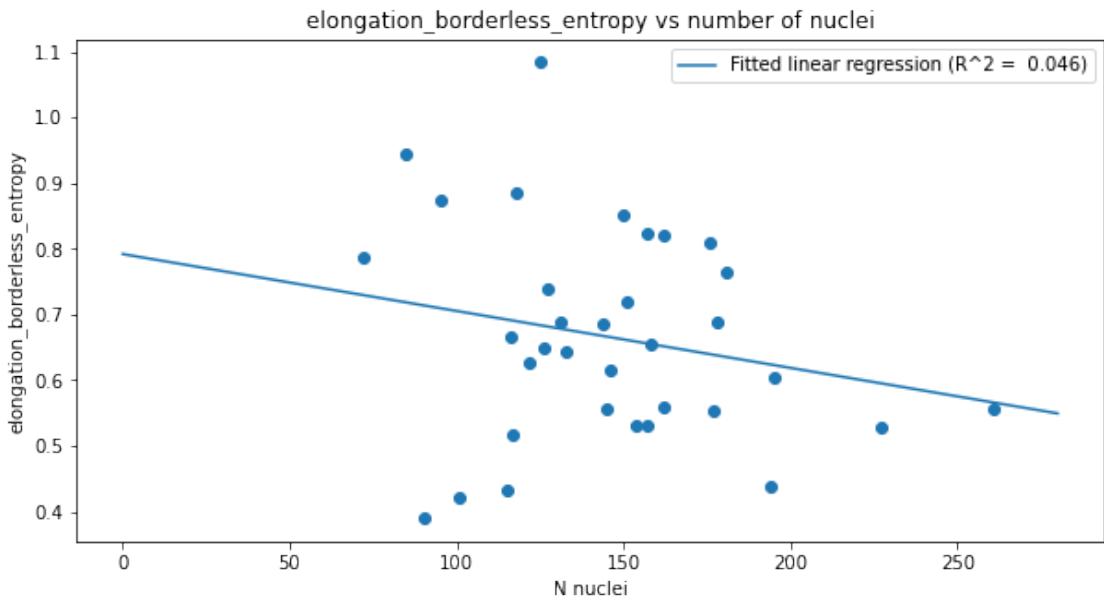
e2e_eu_distance_mean vs number of nuclei

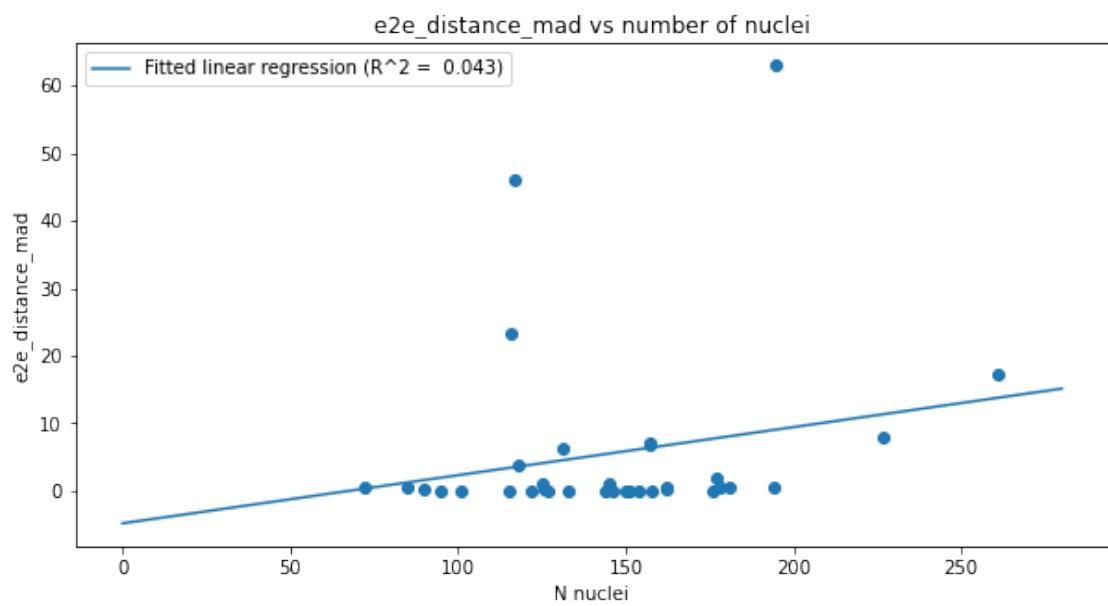
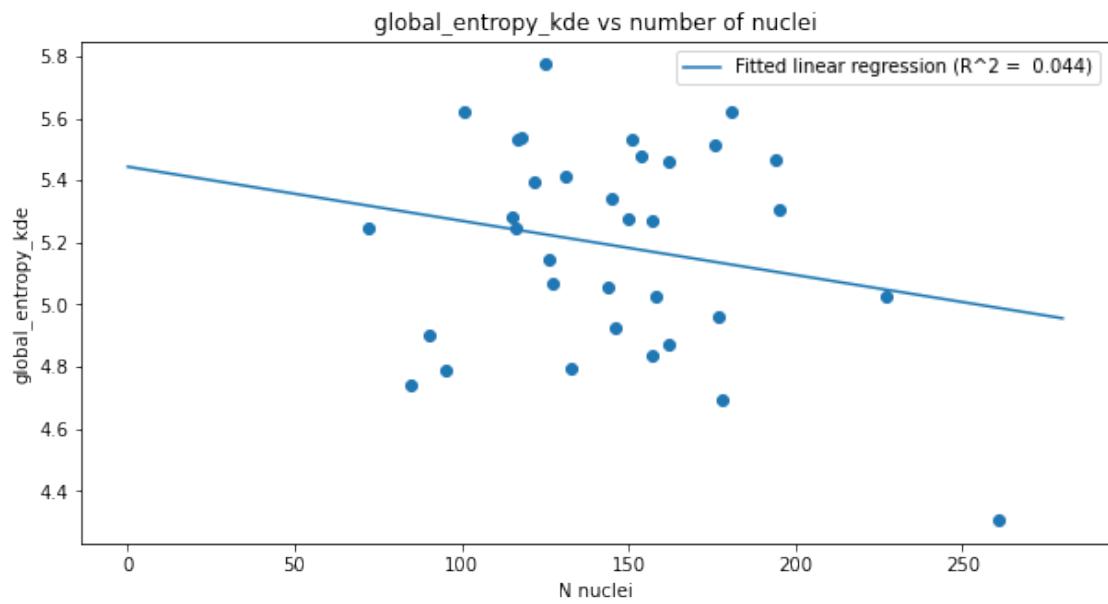


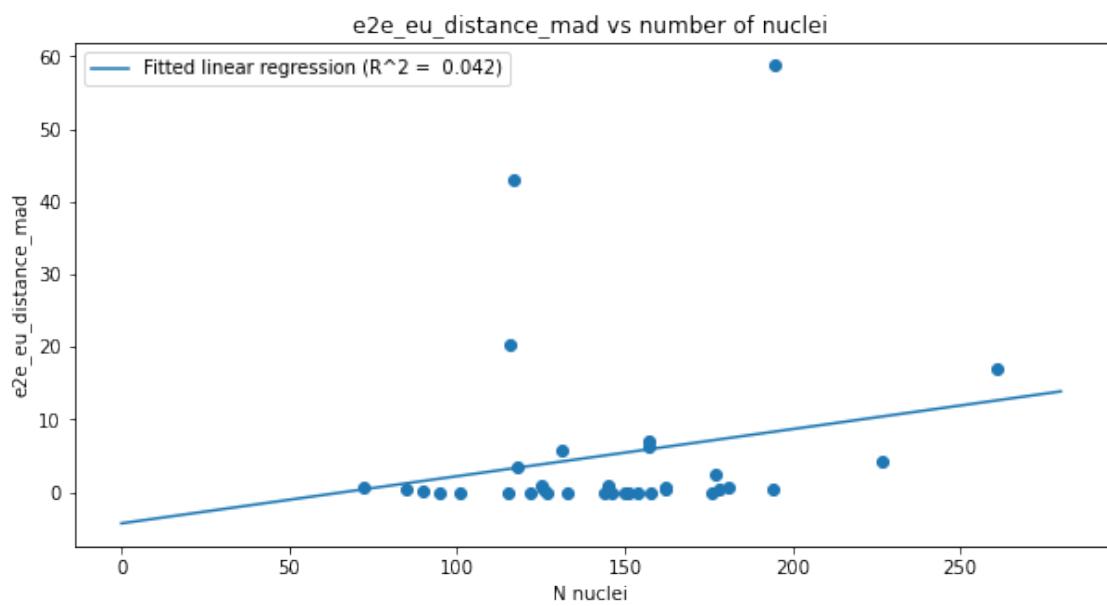
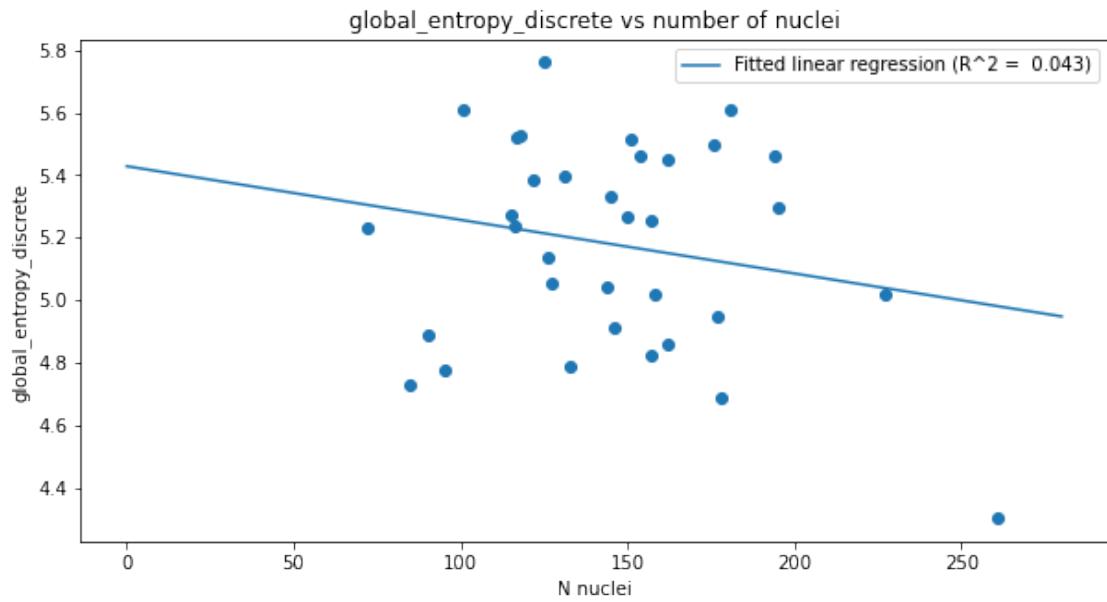


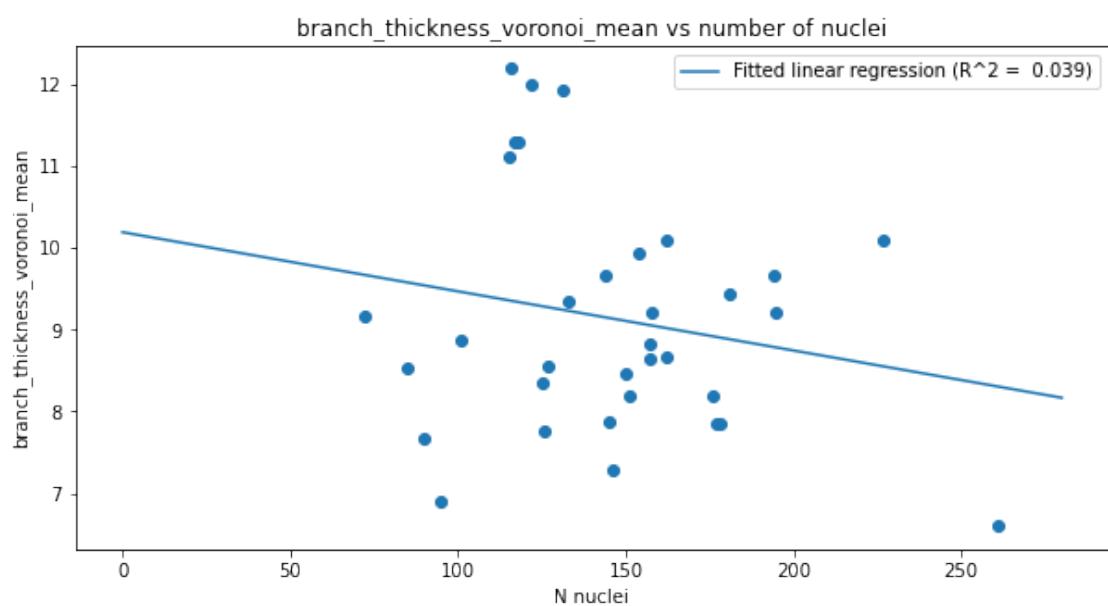
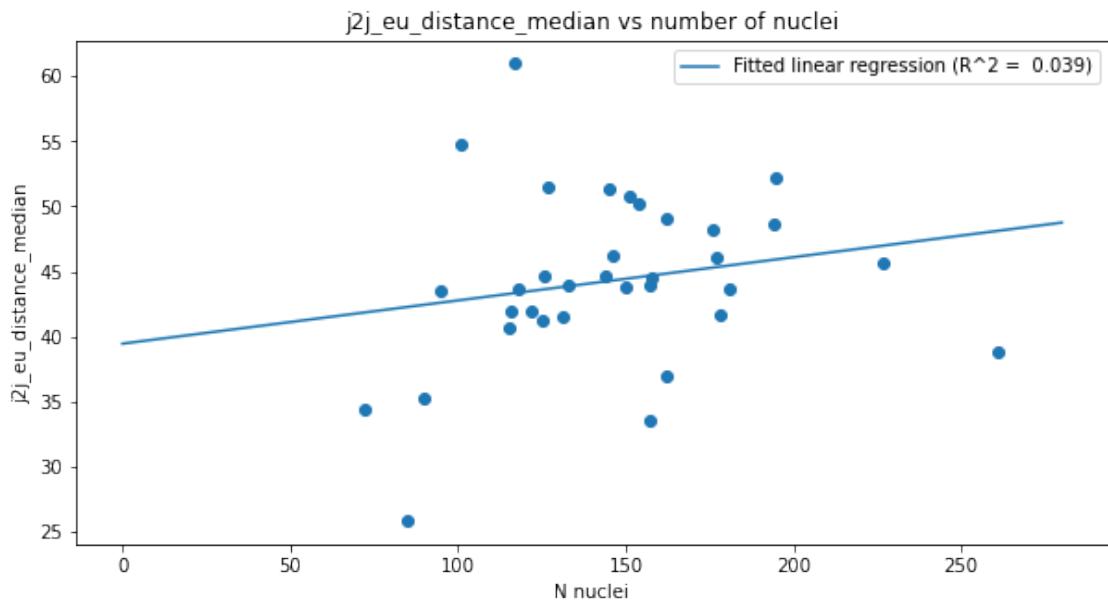


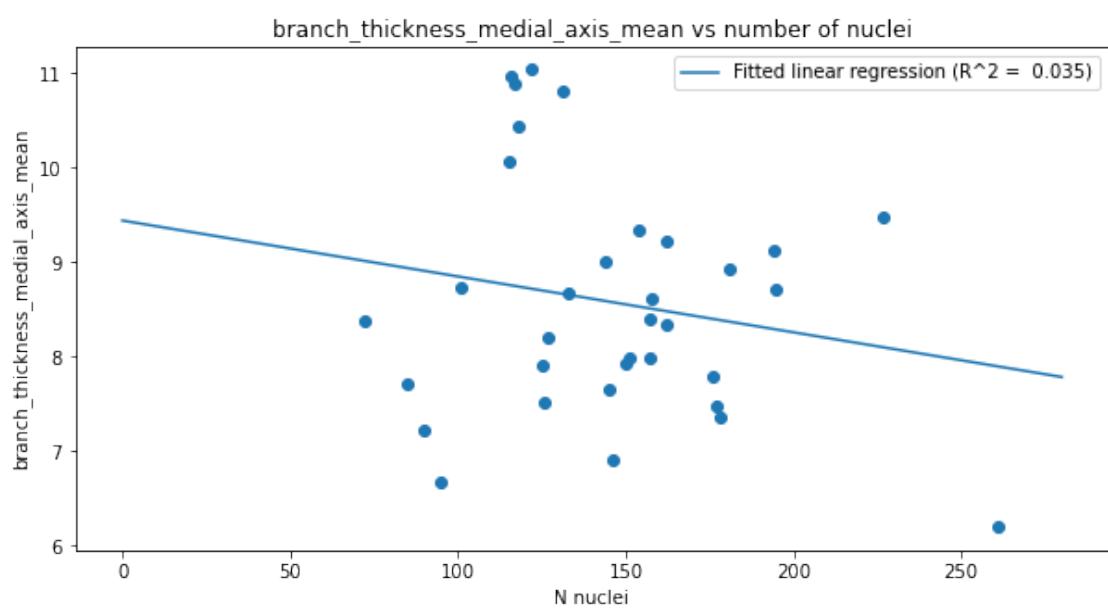
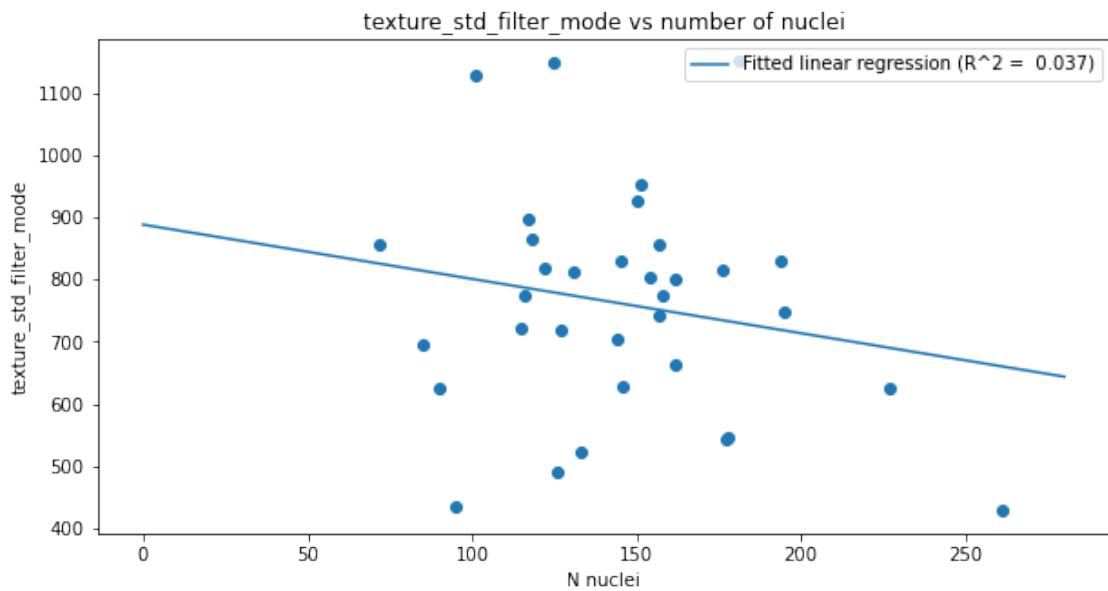


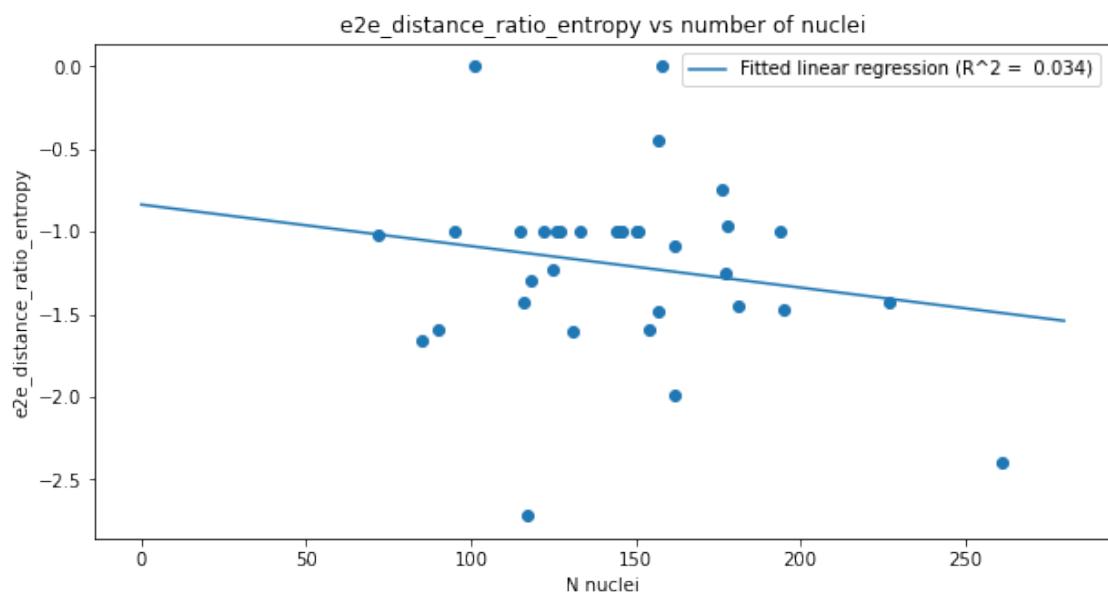
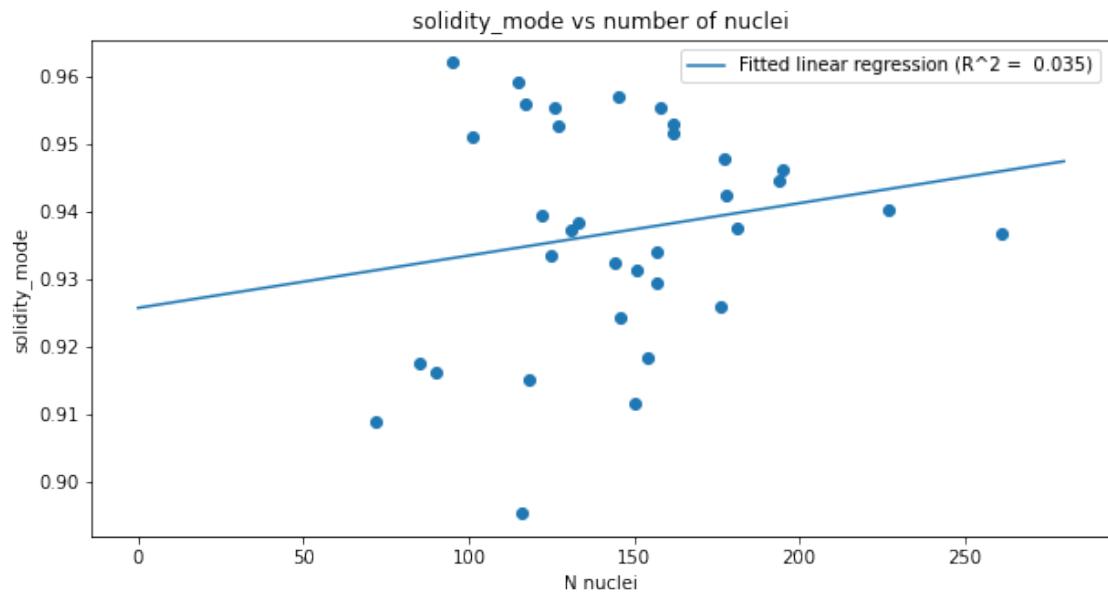


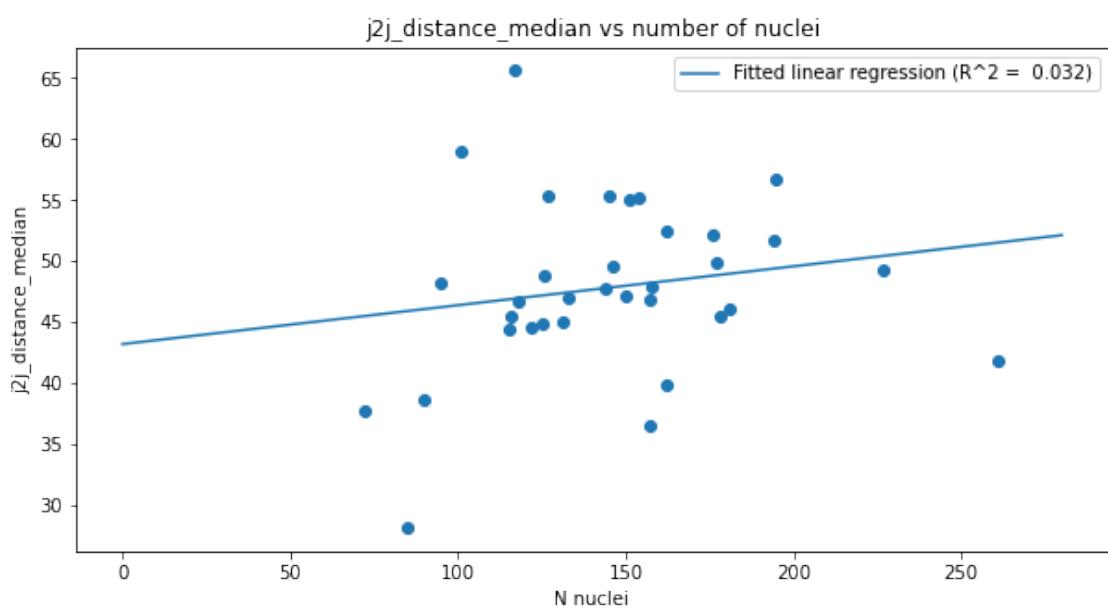
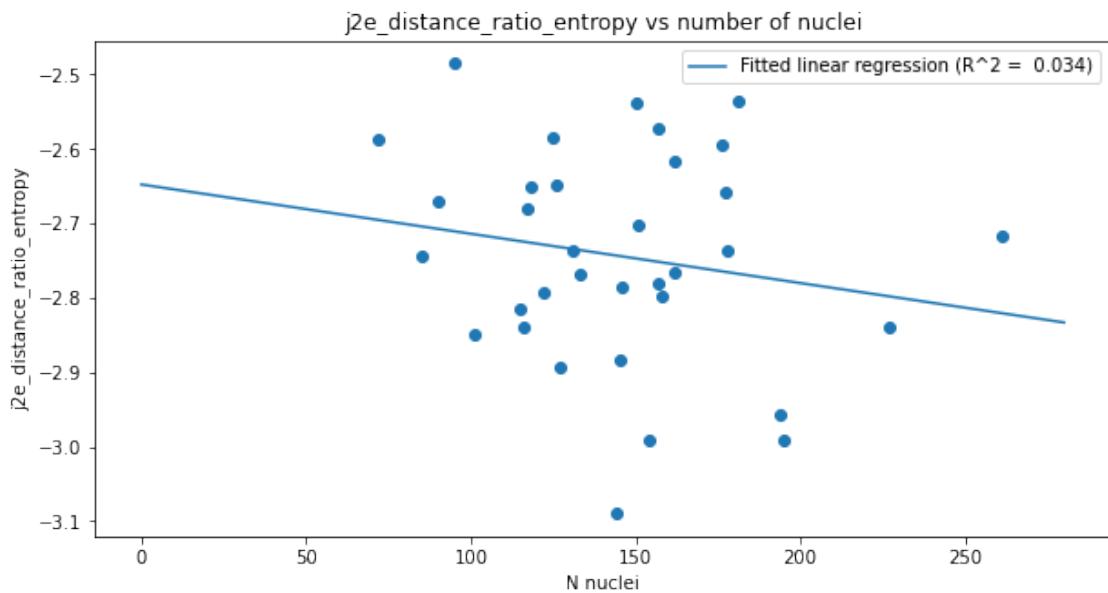


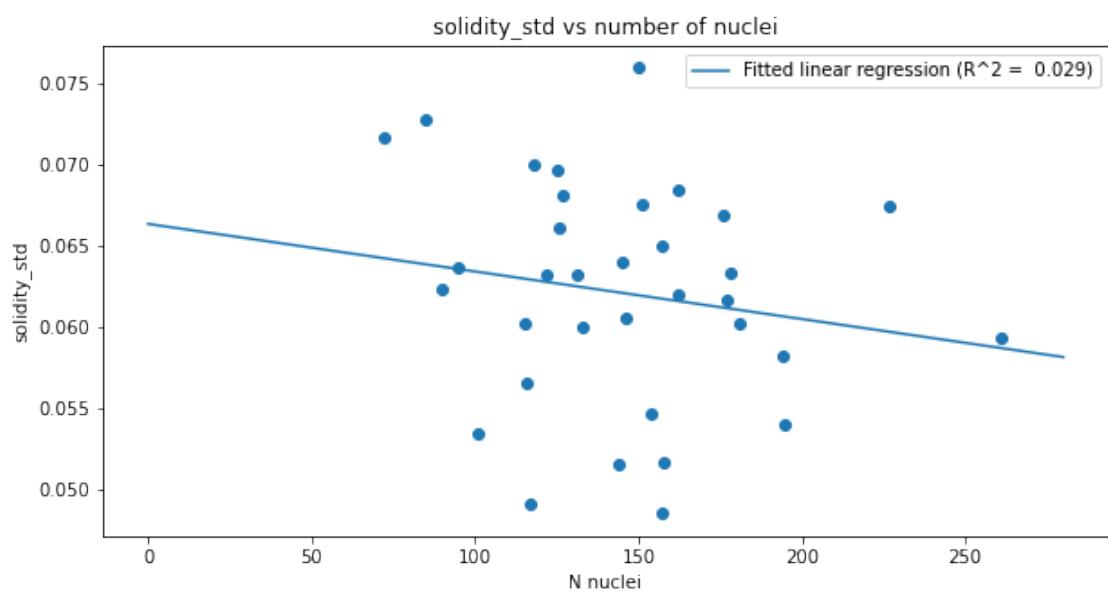
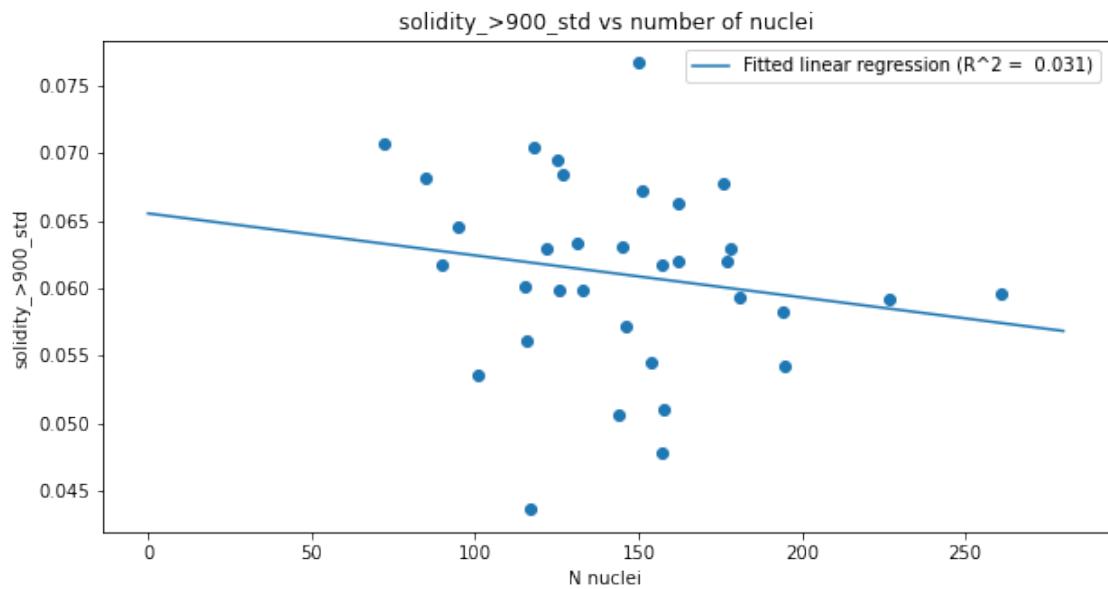


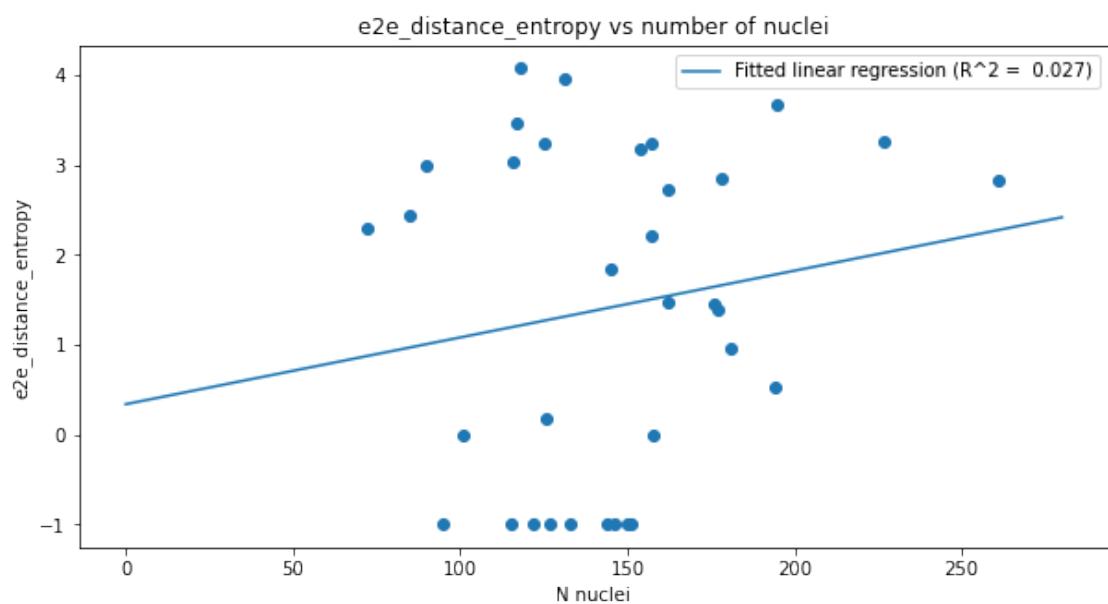
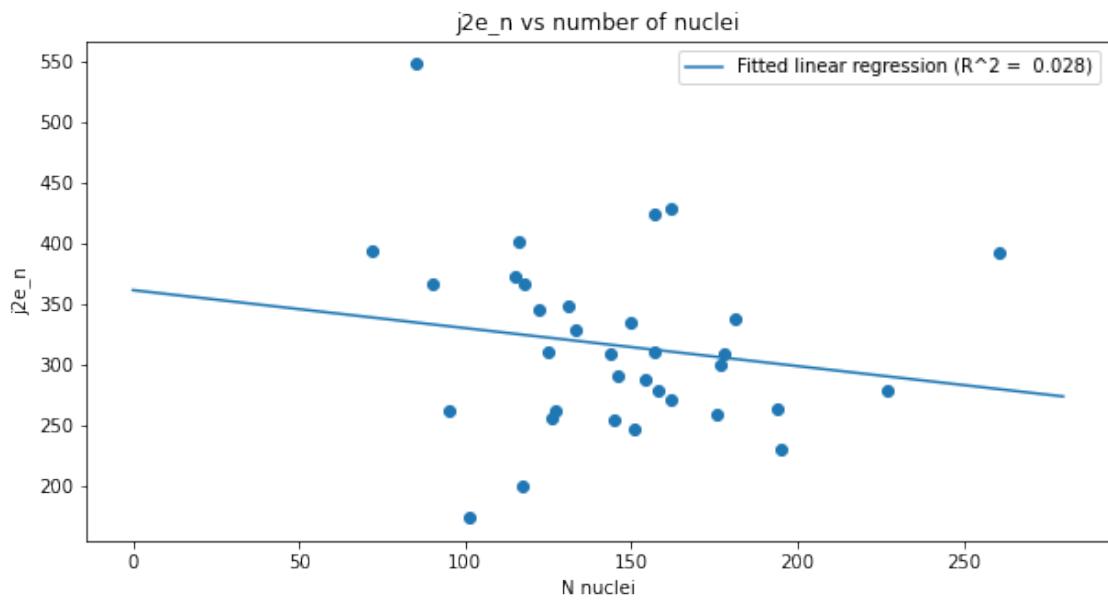


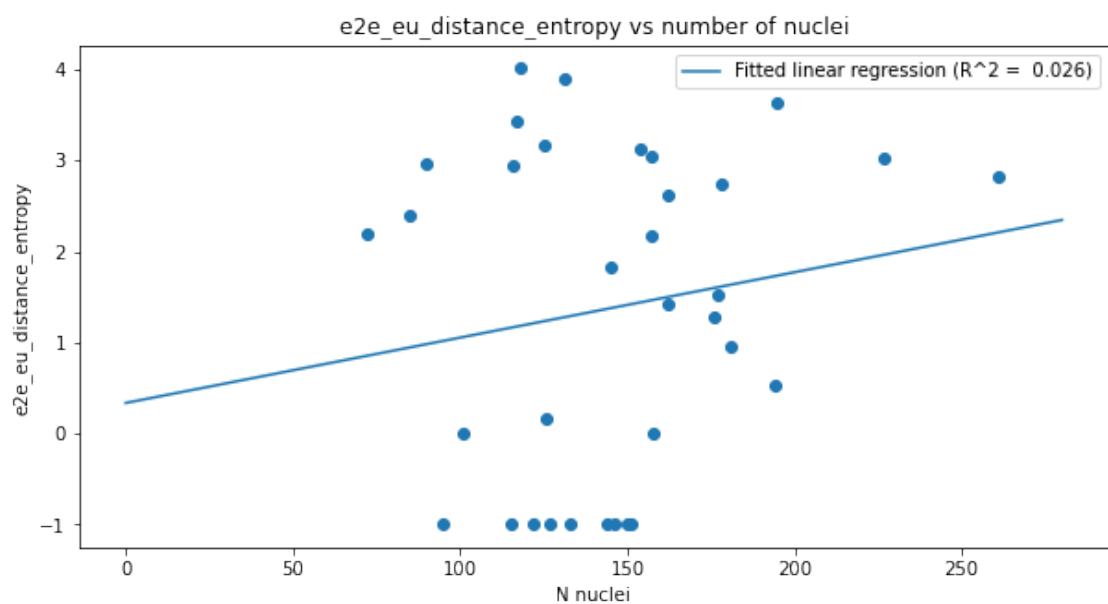
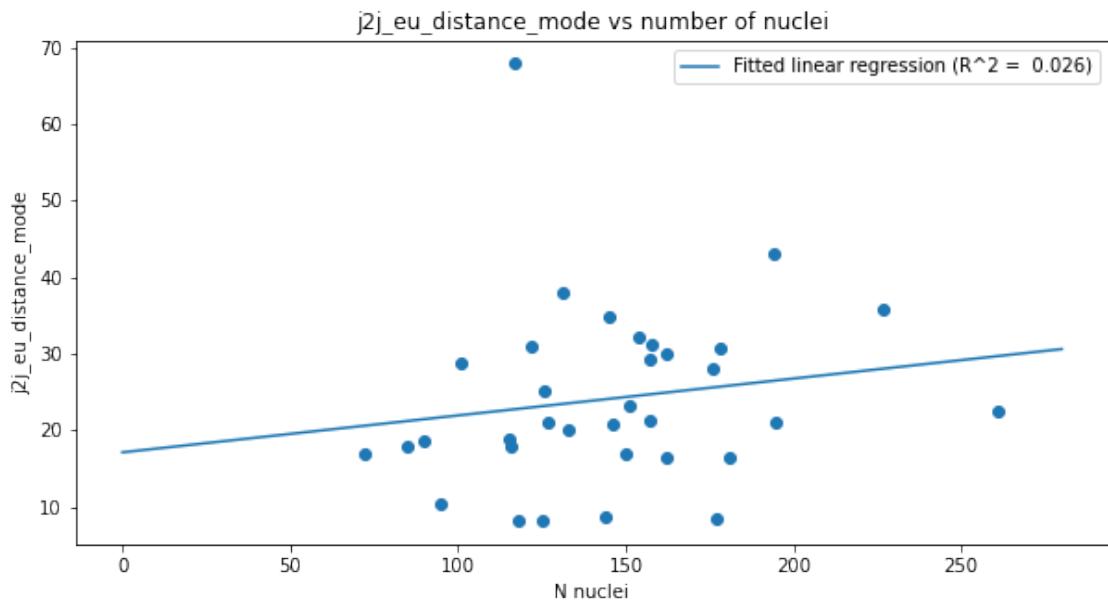


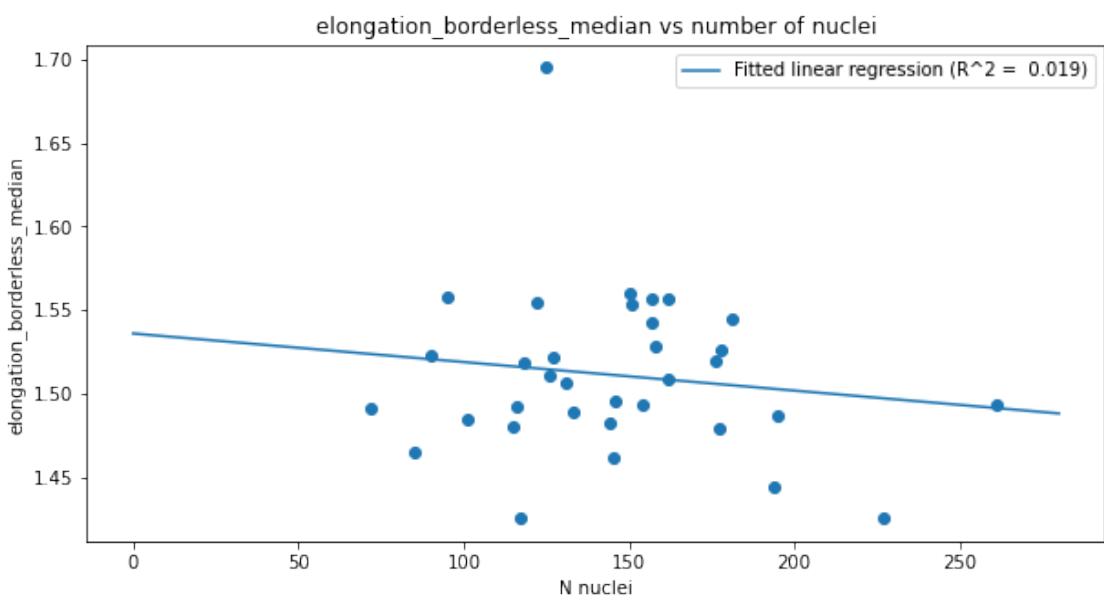
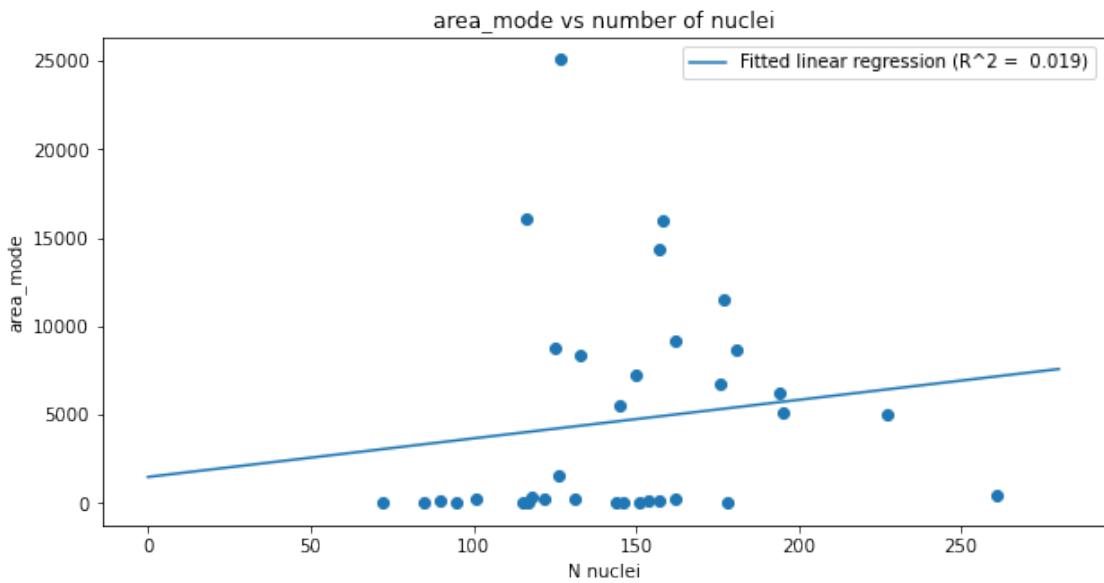


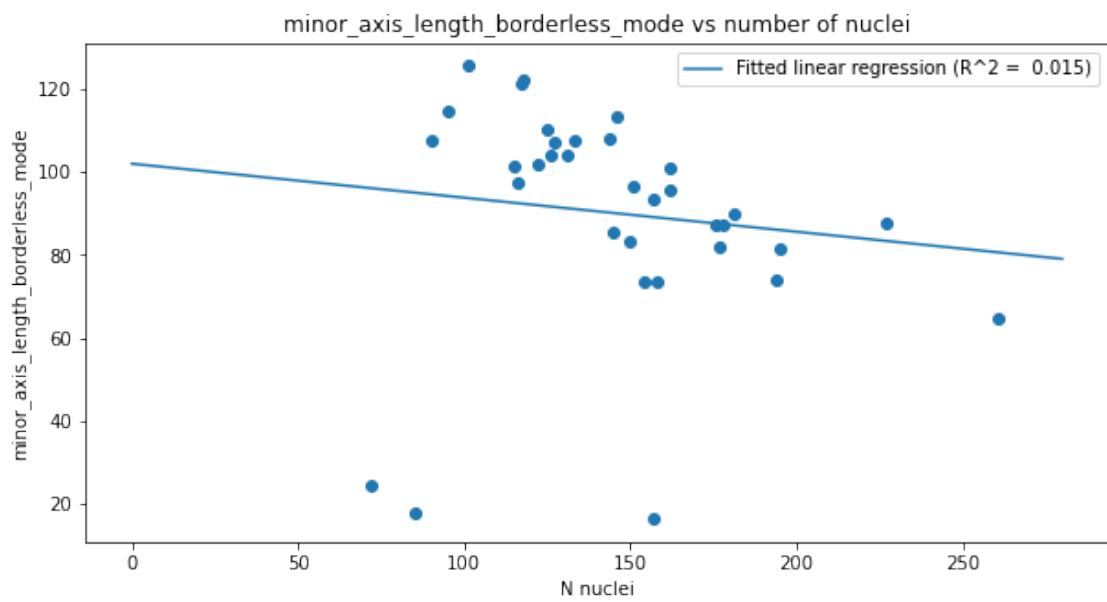
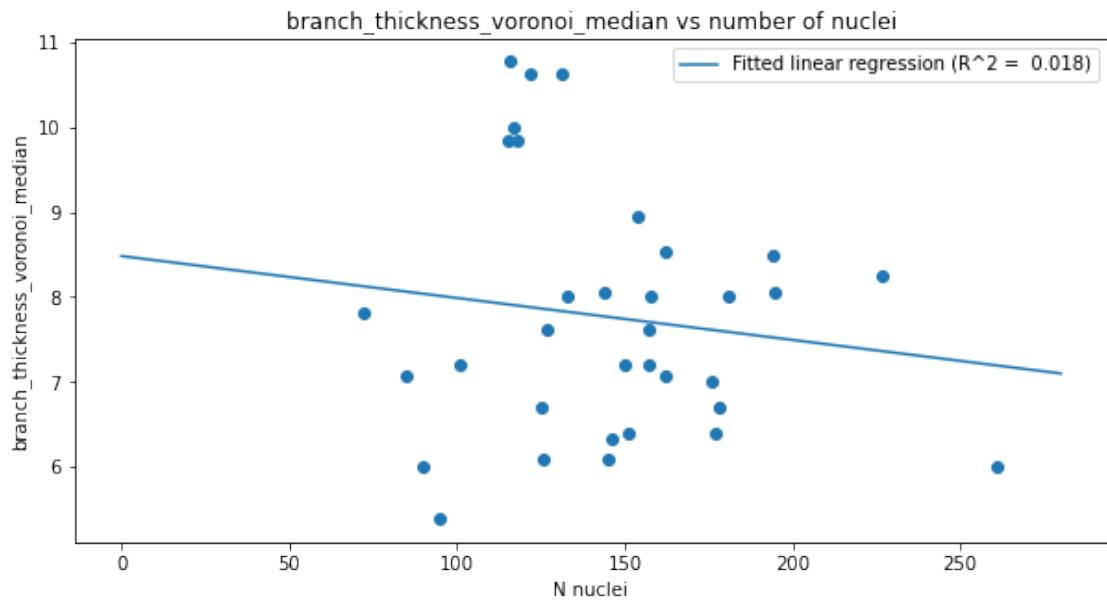




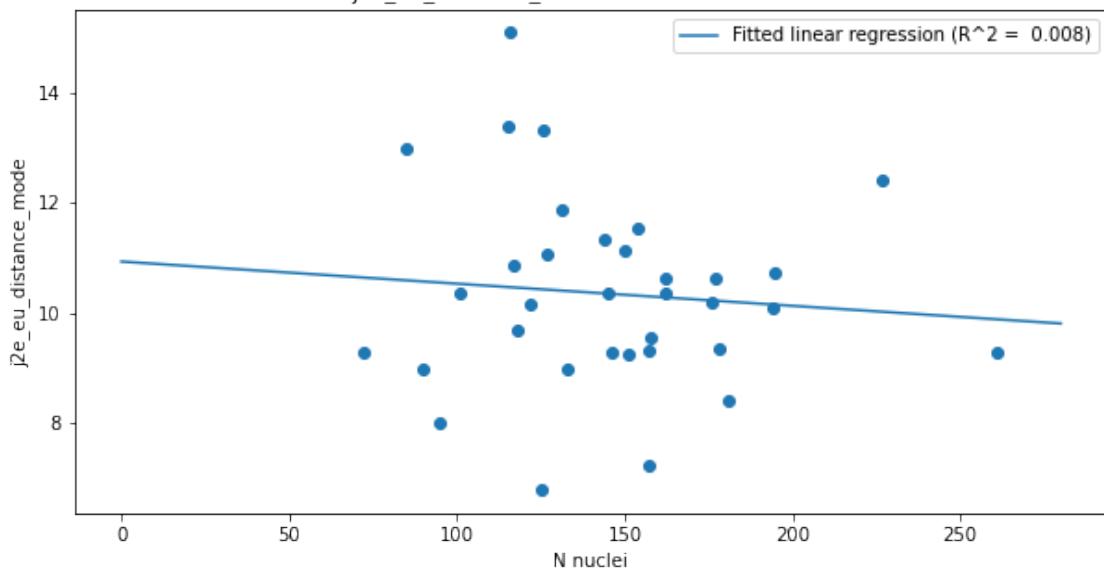




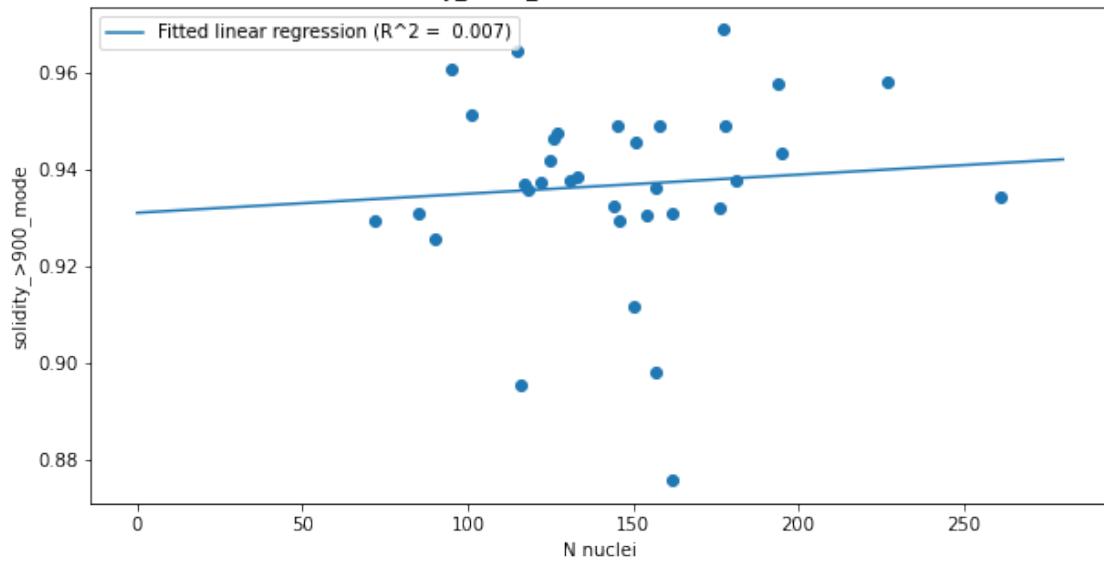


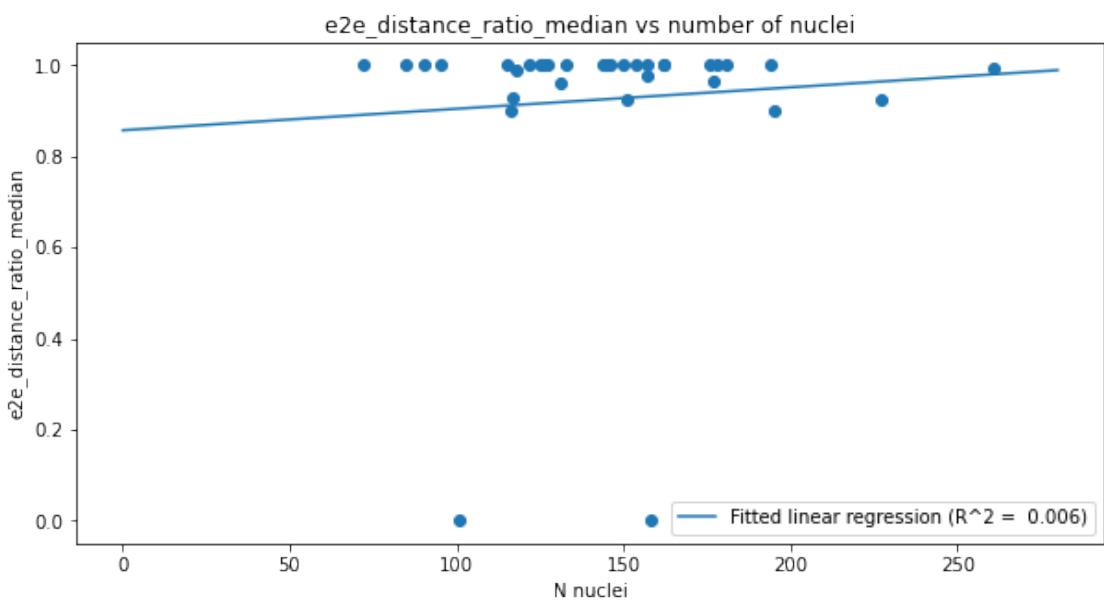
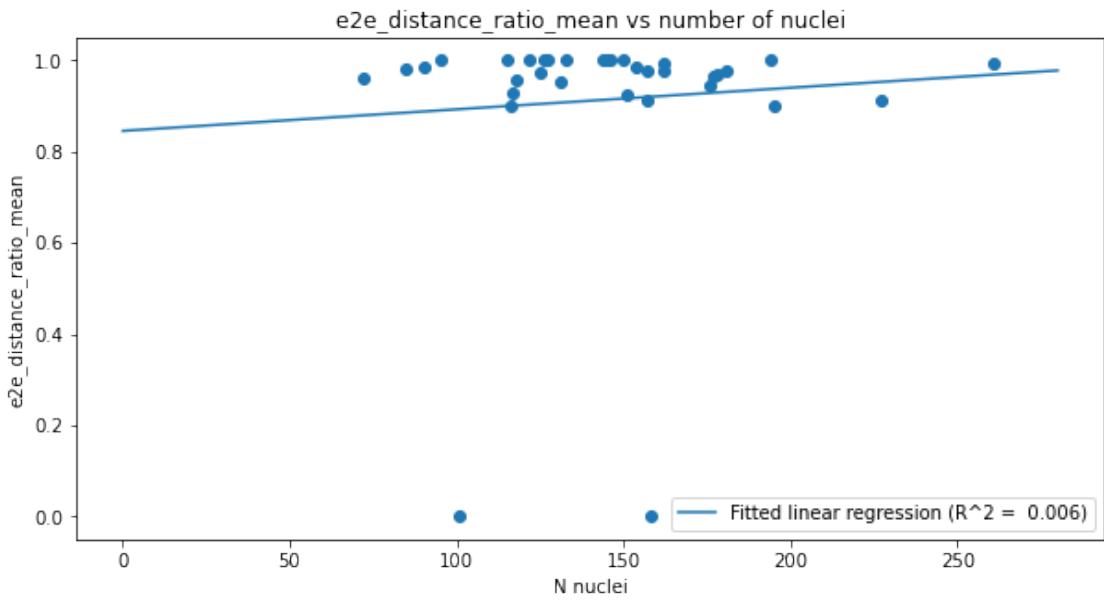


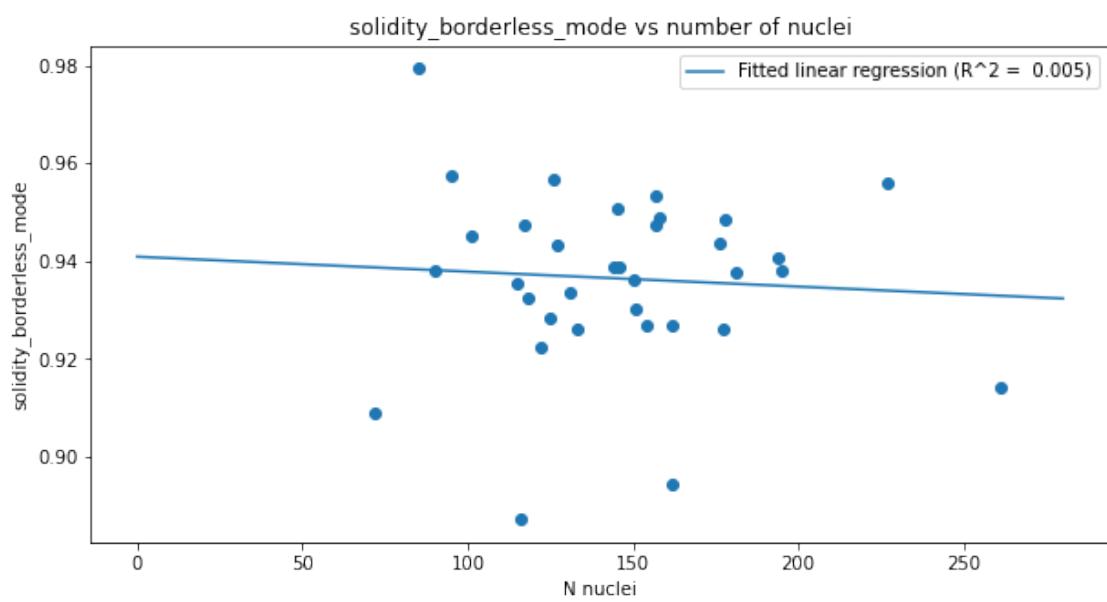
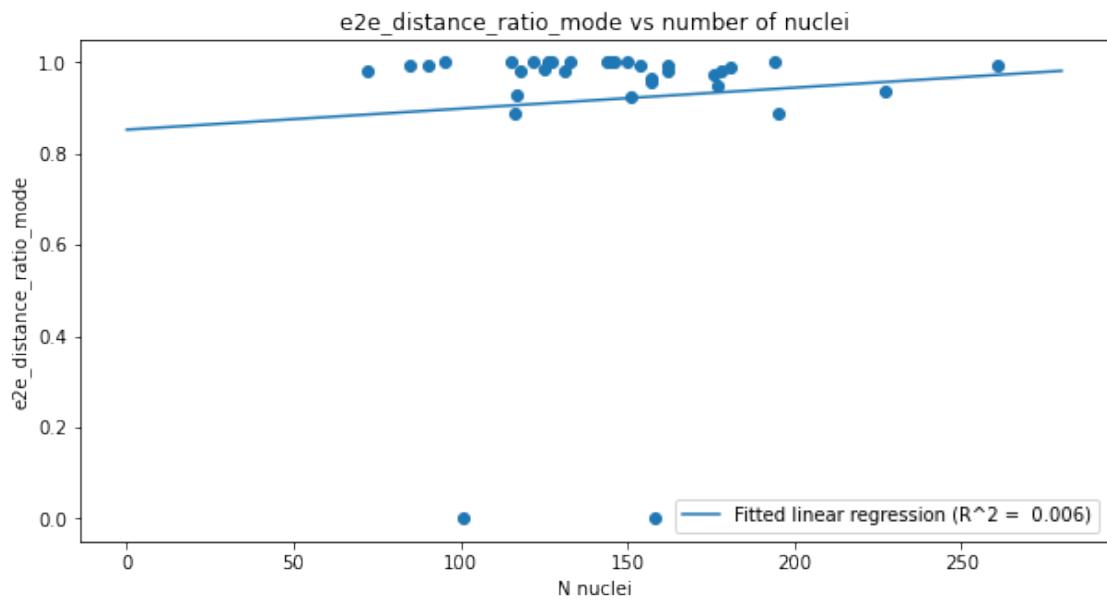
j2e_eu_distance_mode vs number of nuclei



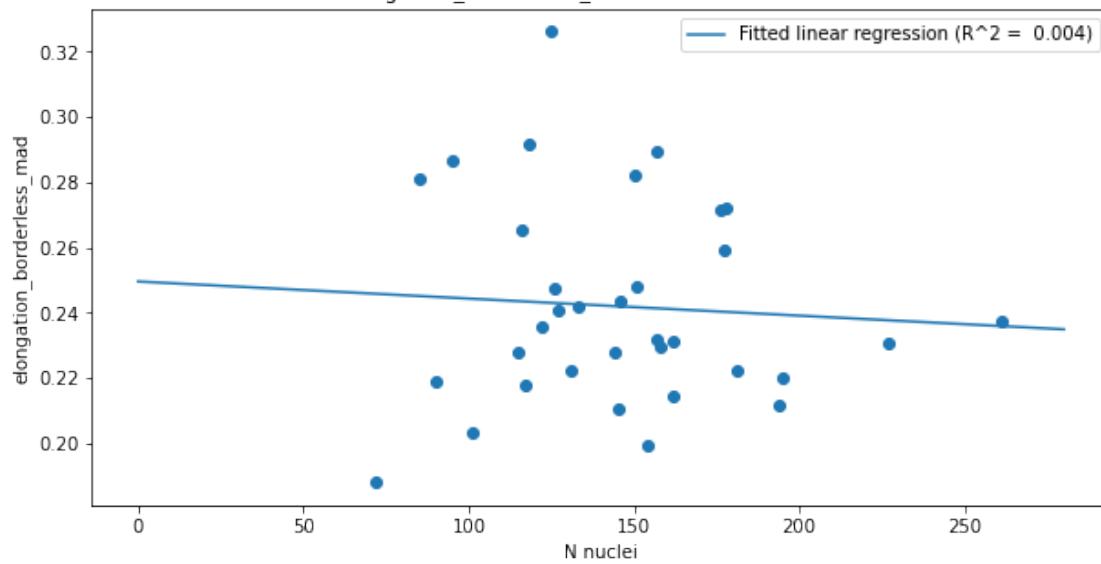
solidity_>900_mode vs number of nuclei



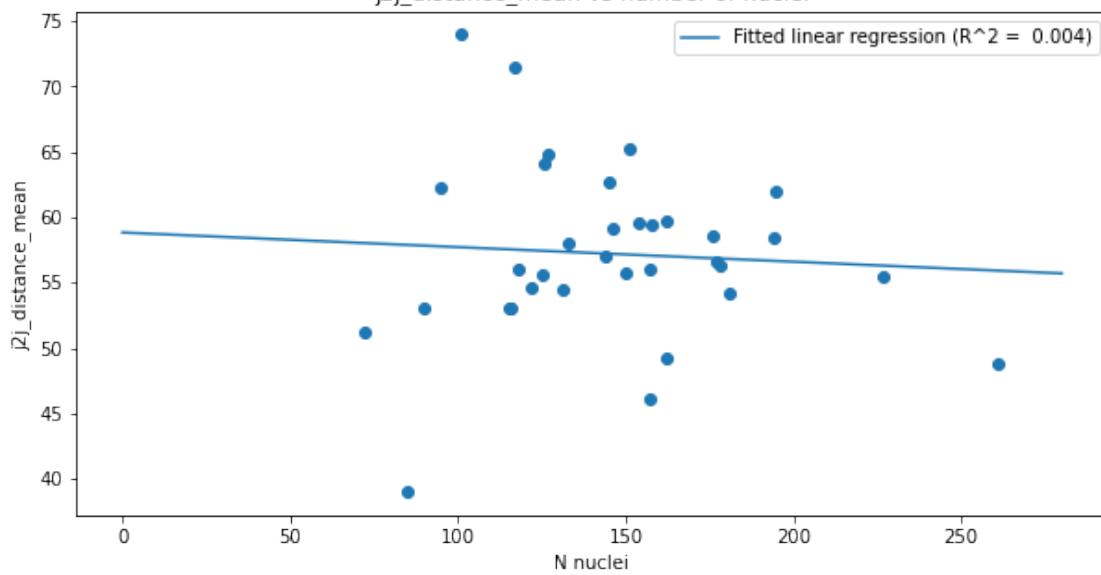


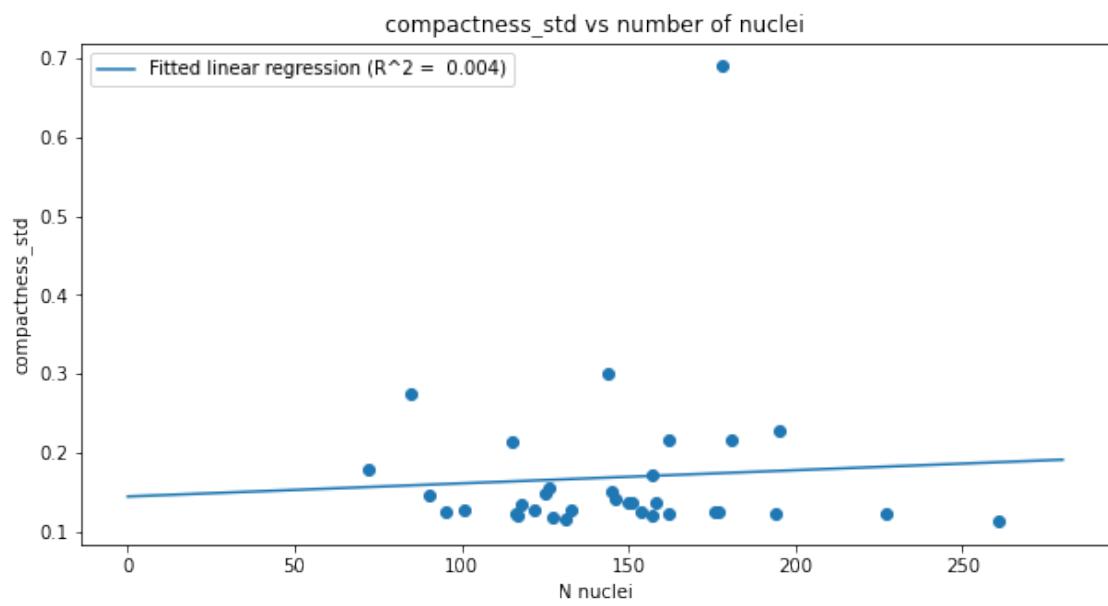
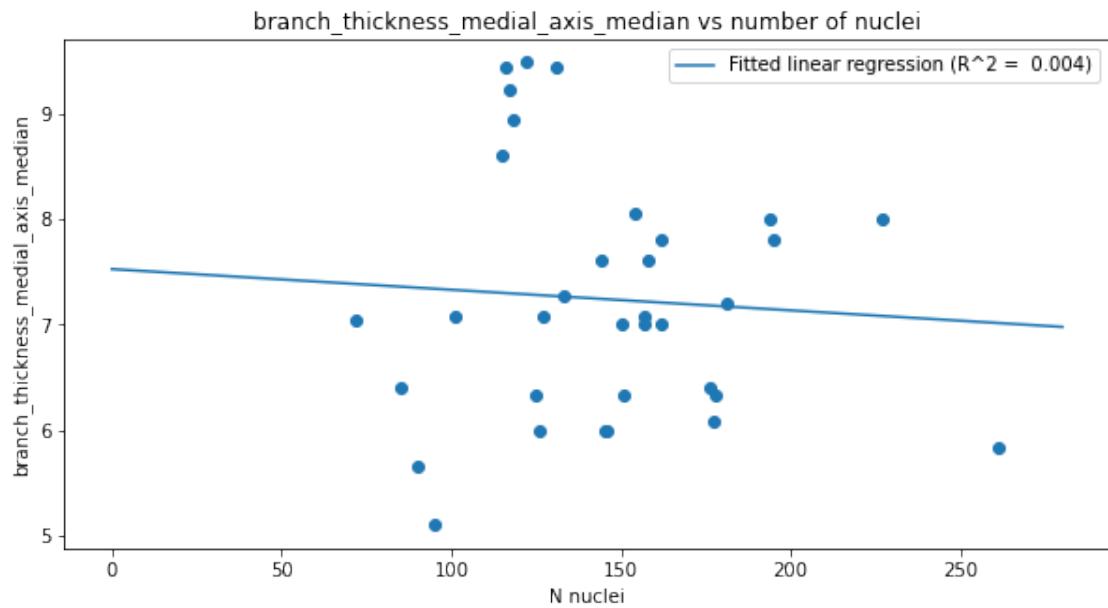


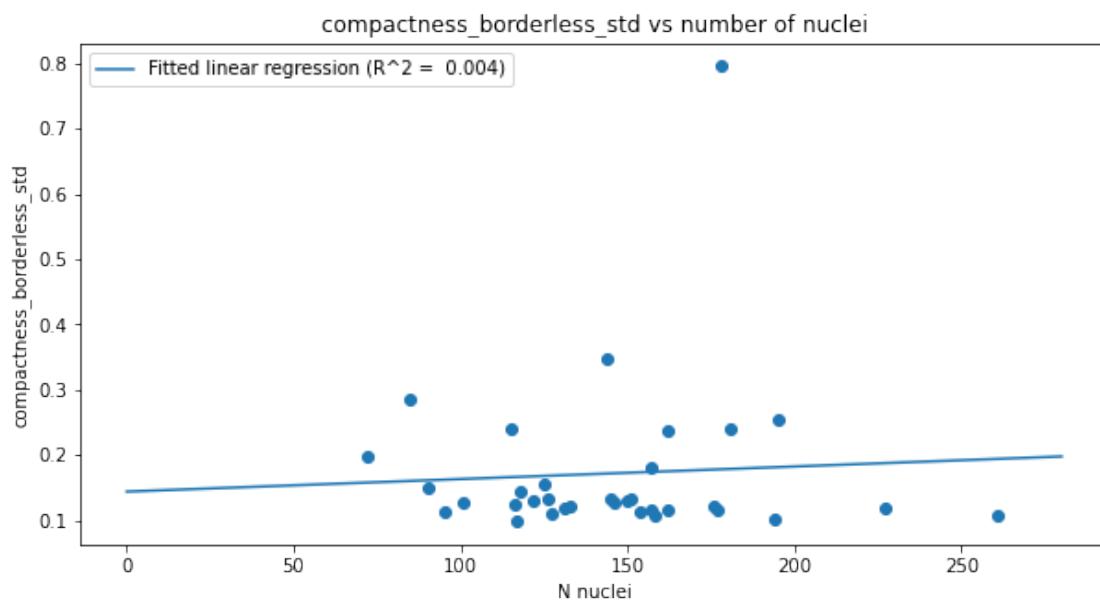
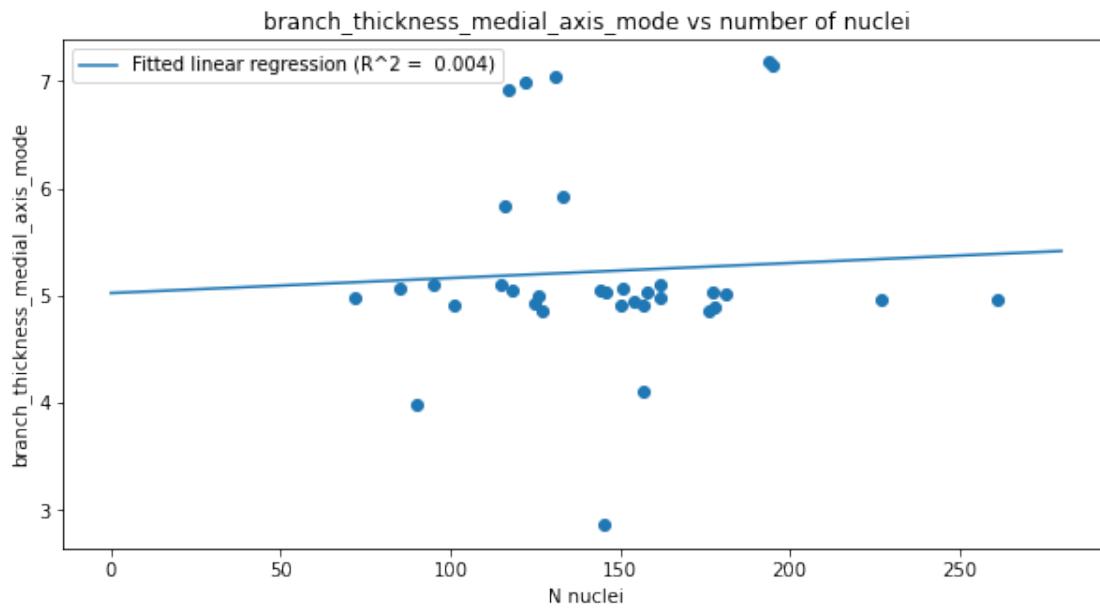
elongation_borderless_mad vs number of nuclei

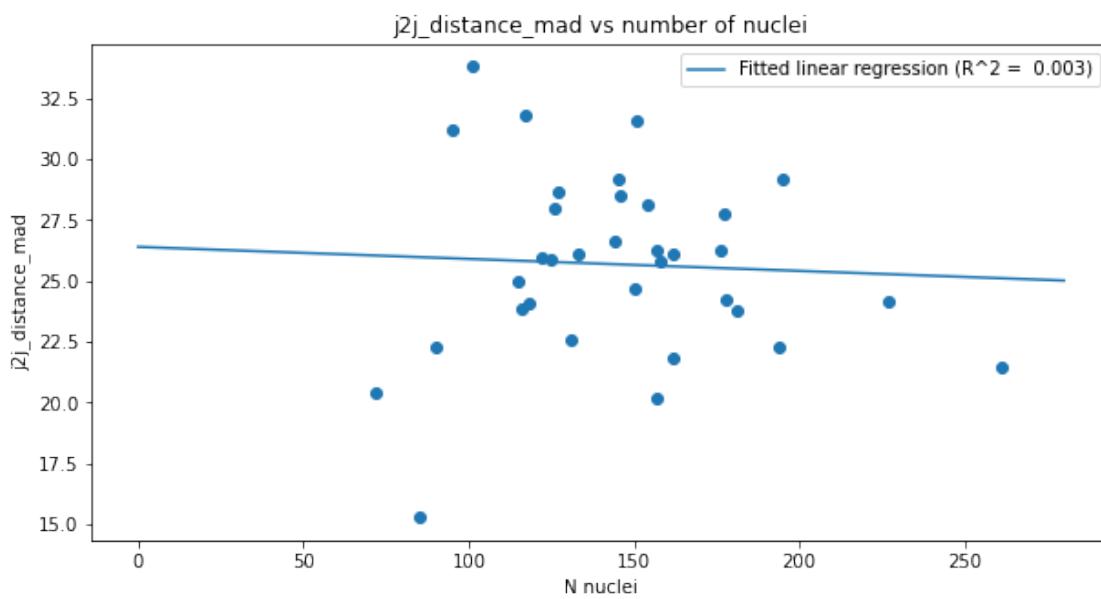
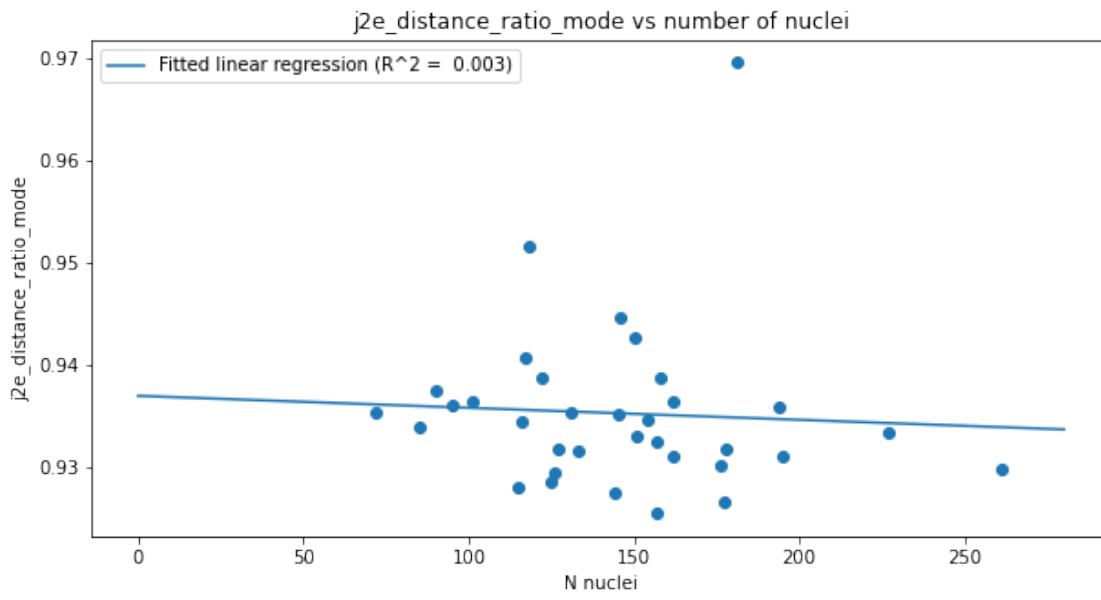


|2j_distance_mean vs number of nuclei

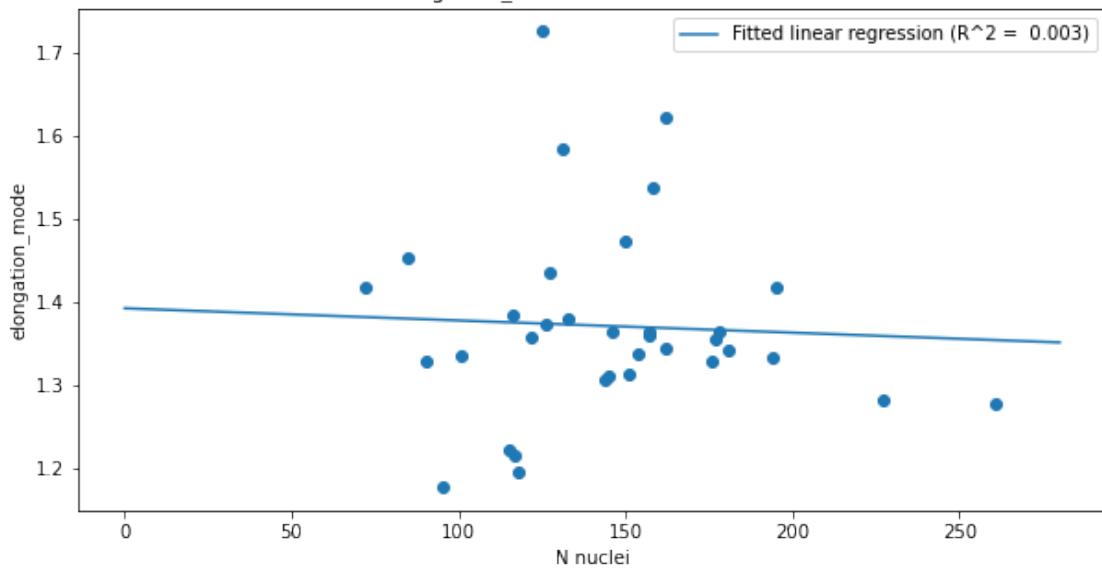




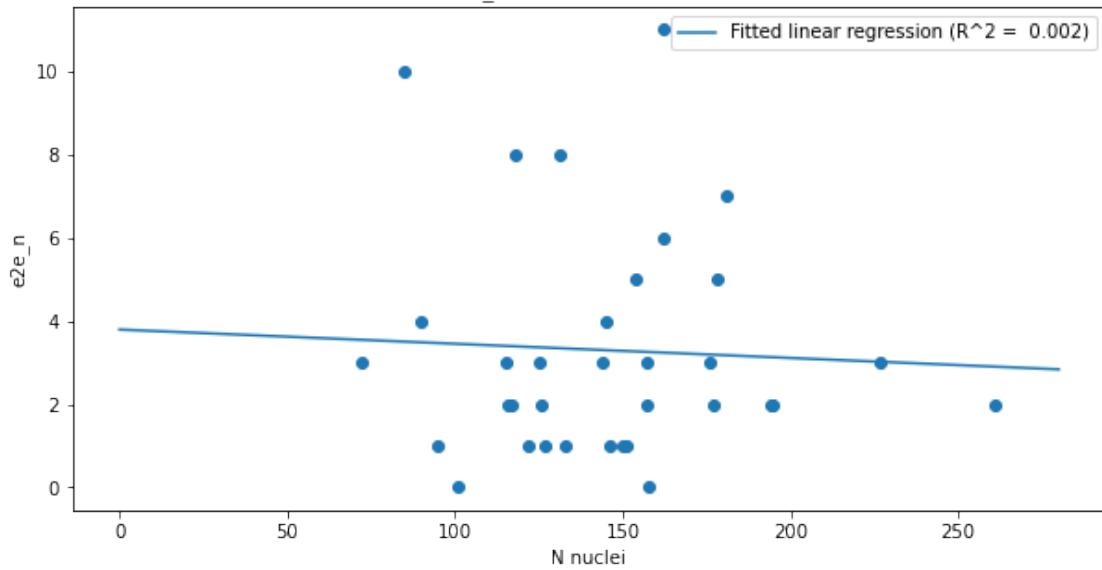


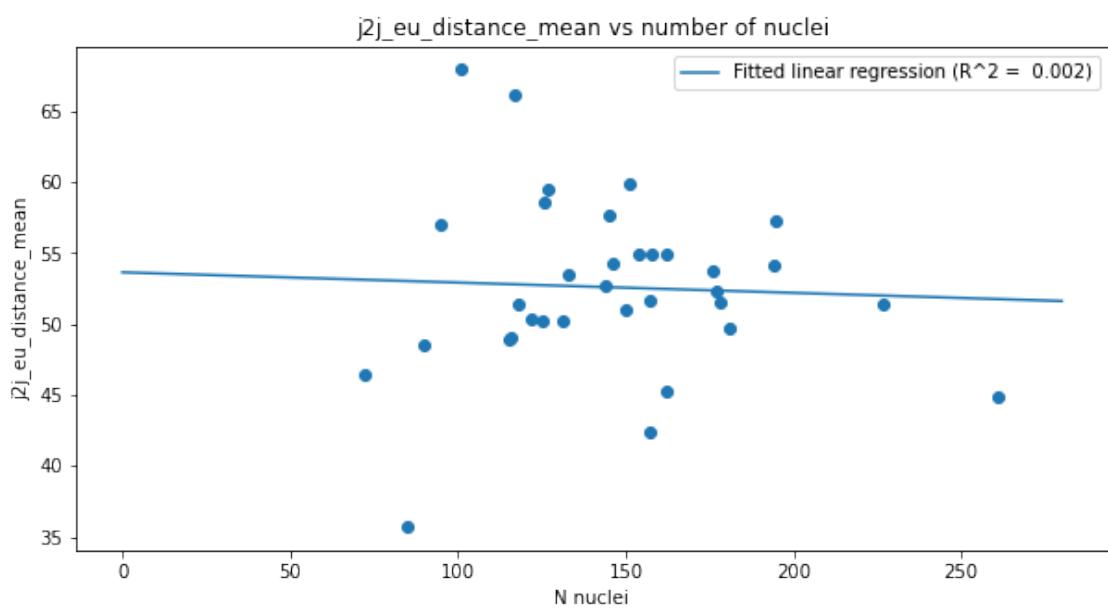
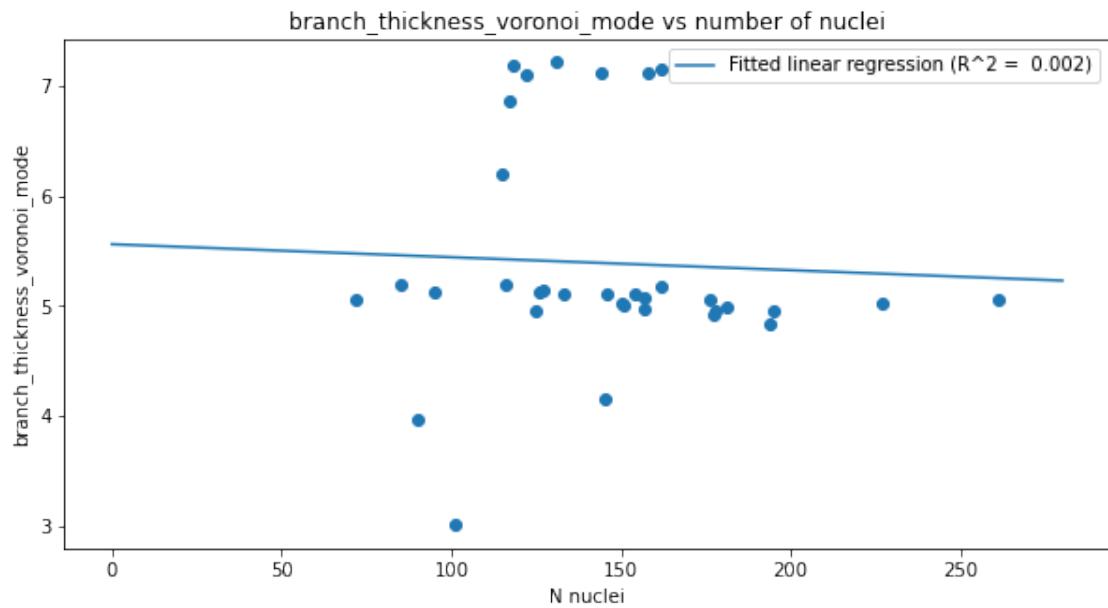


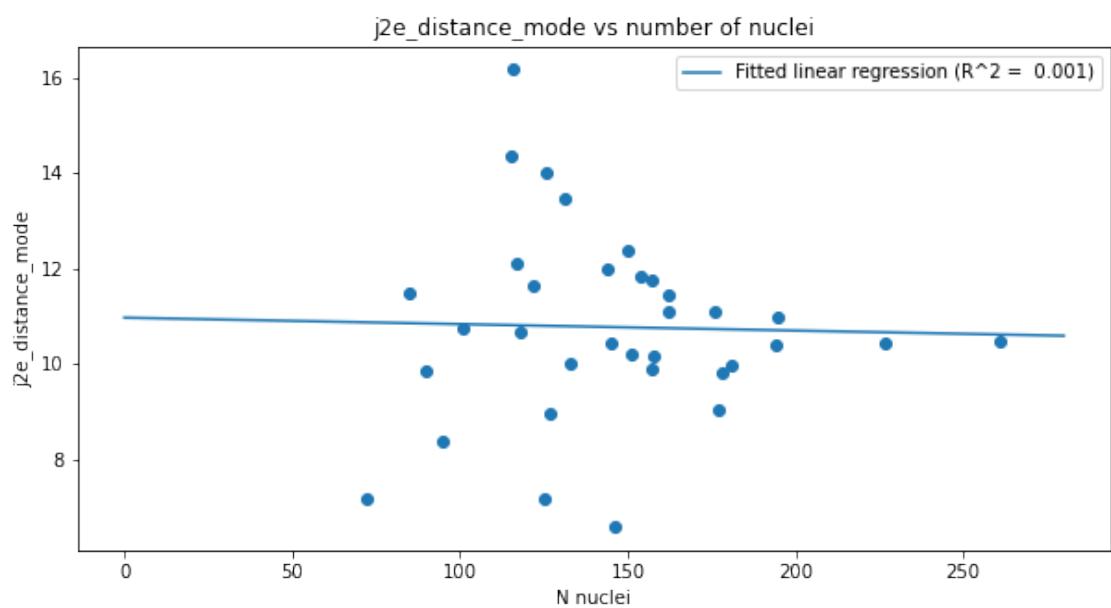
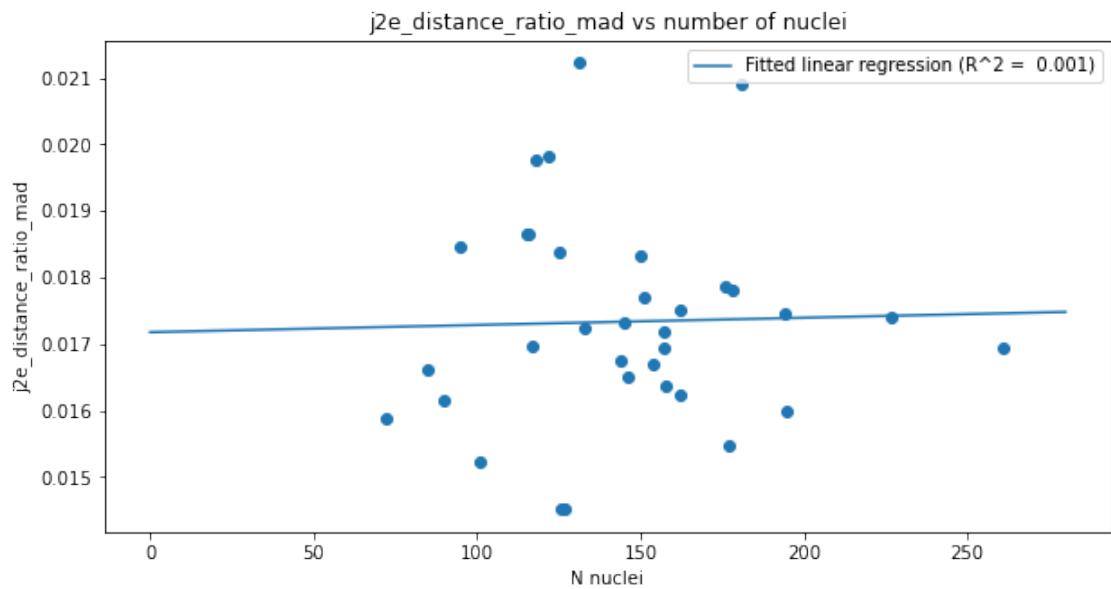
elongation_mode vs number of nuclei



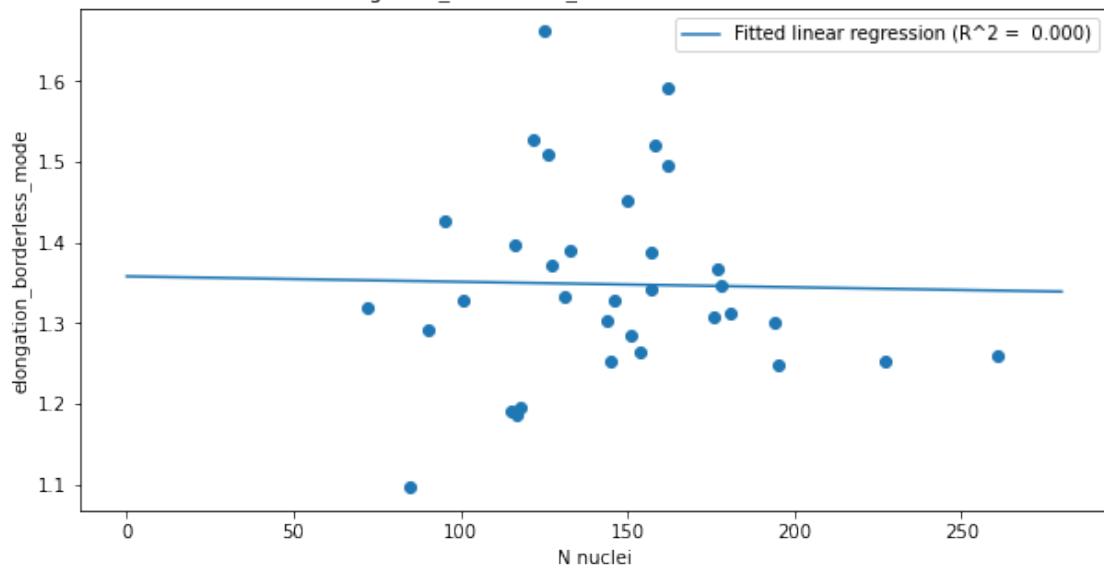
e2e_n vs number of nuclei



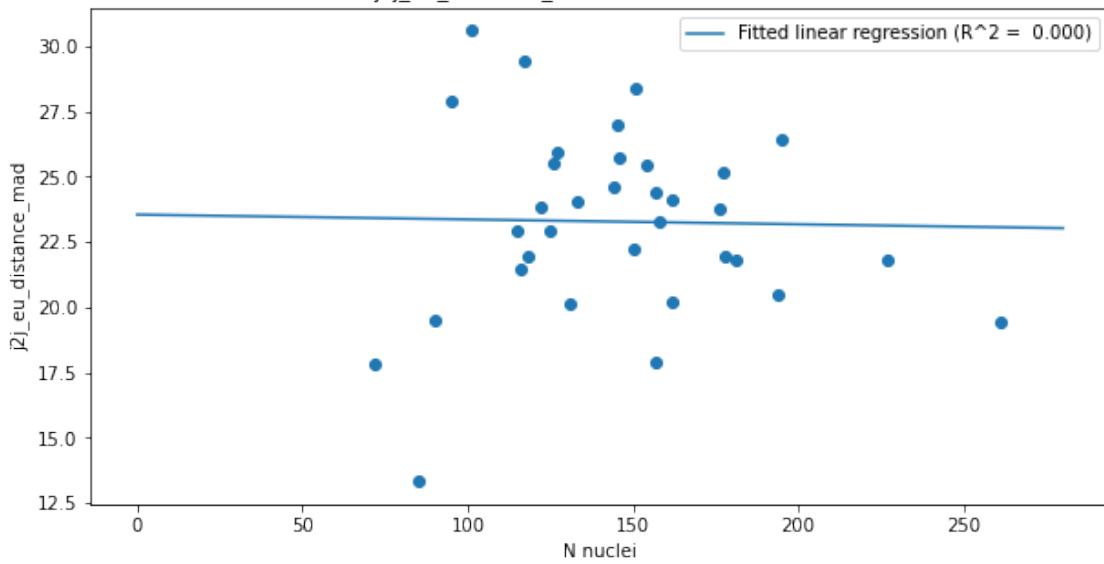


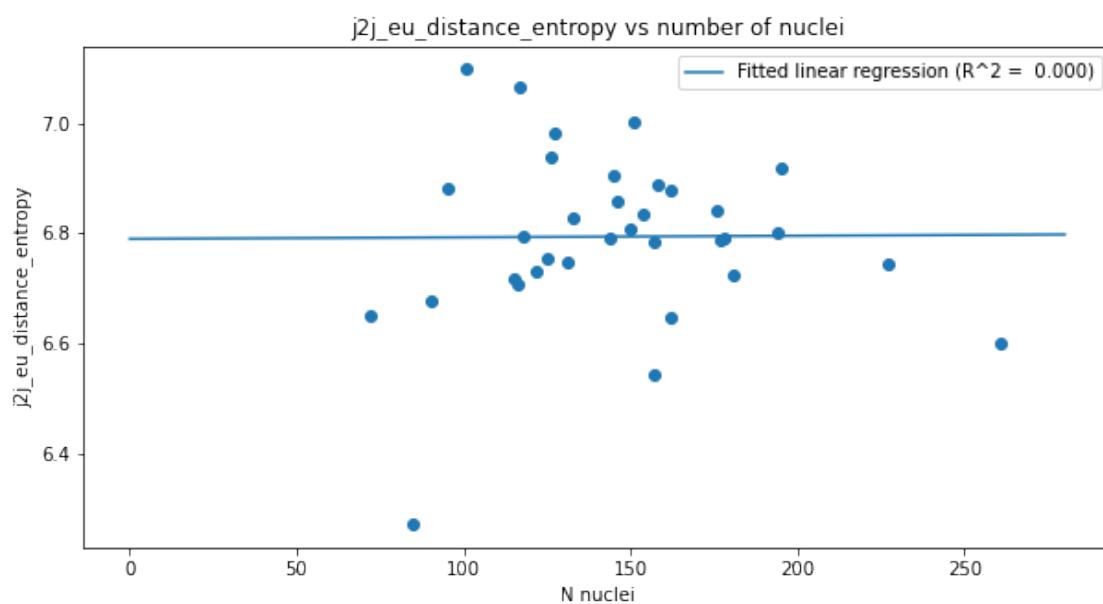
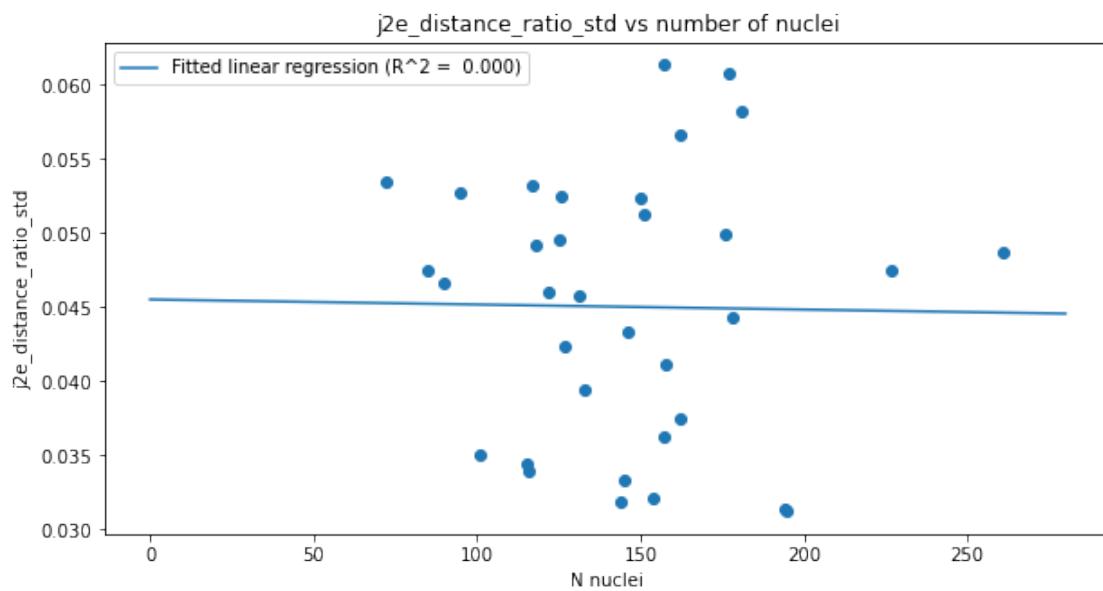


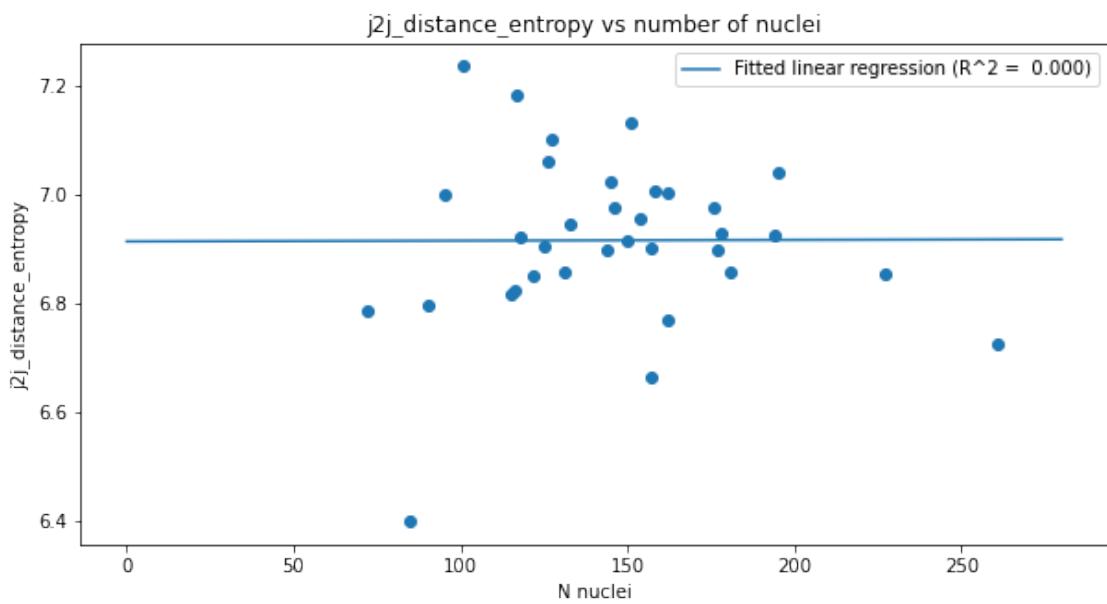
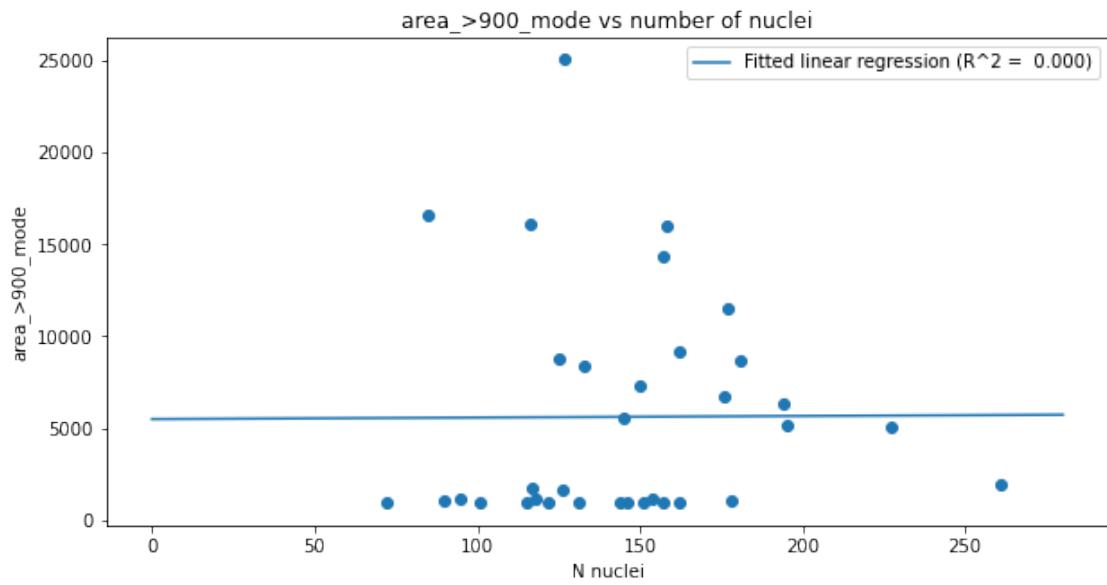
elongation_borderless_mode vs number of nuclei

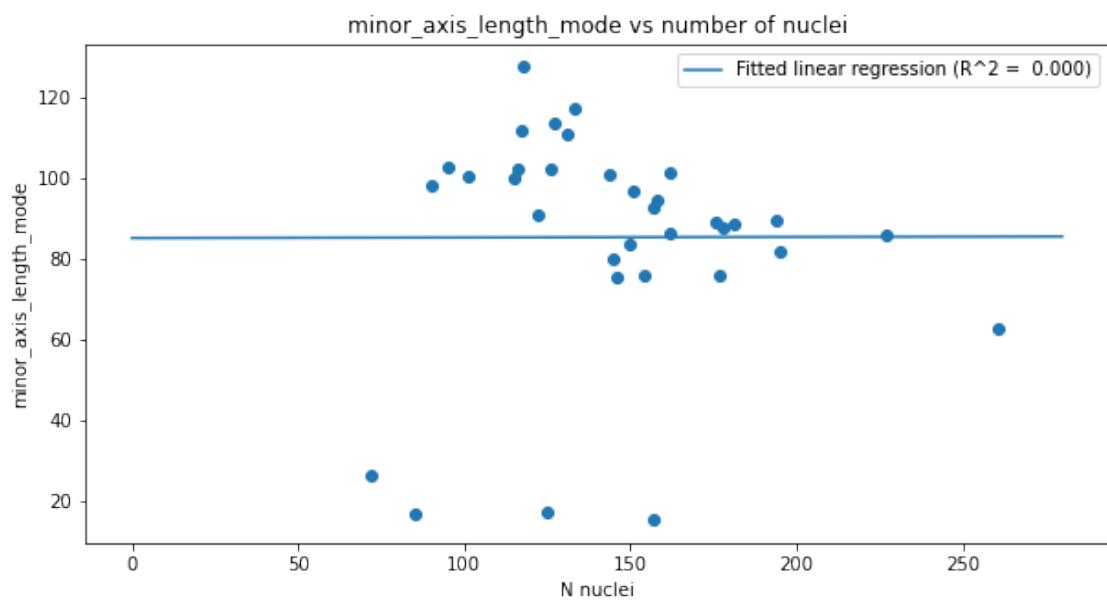
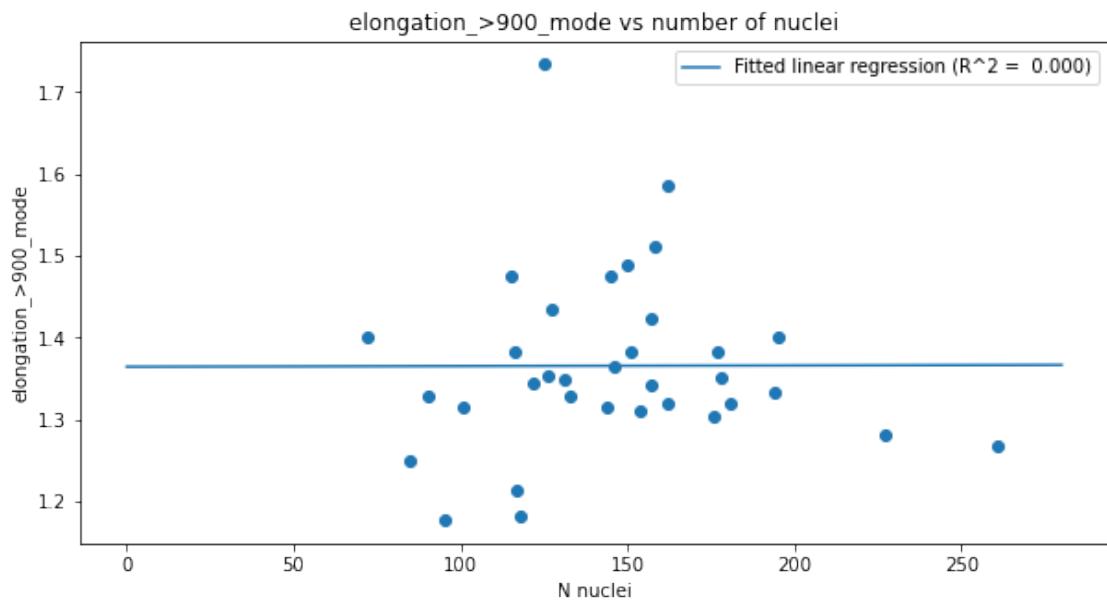


j2j_eu_distance_mad vs number of nuclei









[]: