Lecture 9: Building Functions

Eugen Buehler

November 14, 2016

R Functions are objects

R is a functional programming language. This means that functions are "objects", just like data frames, vectors, and other things that are assigned to variables and passed to other functions.

A rose by any other name

The name of a functions is actually the name of a variable that contains the function, in the same way that the

```
log
```

```
## function (x, base = exp(1))  .Primitive("log")
```

This means that we can create a copy of a function by assigning its value to a new variable.

```
myLogFunction <- log
myLogFunction
```

```
## function (x, base = exp(1))  .Primitive("log")
```

Functions are a kind of data and have a class

```
myNumber <- 7
class(myNumber)
```

```
## [1] "numeric"
```

```
class(log)
```

```
## [1] "function"
```

Creating a function

We can create a new function using the word "function" followed by the functions arguments and one or more R statements.

```
myDumbFunction <- function() 42
myDumbFunction()
```

```
## [1] 42
```

Creating a multi-statement function

If there is more than one statement in a function, they should be enclosed in curly brackets:

```r
doubleIt <- function(x) {
  myResult <- x * 2
  myResult
}
doubleIt(5)
```

```
## [1] 10
```

The last statement within the curly brackets will be the value returned by the function.

Functions live in their own little world

Inside a function, variables that existed in your environment can be used and even changed. However, any changes made, including changing data stored in variables and creating new variables, happens solely within the function. Your environment stays the same.

```
exists("myResult")
```

```
## [1] FALSE
```

```
myResult <- 1000
doubleItOutput <- doubleIt(2)
myResult
```

```
## [1] 1000
```

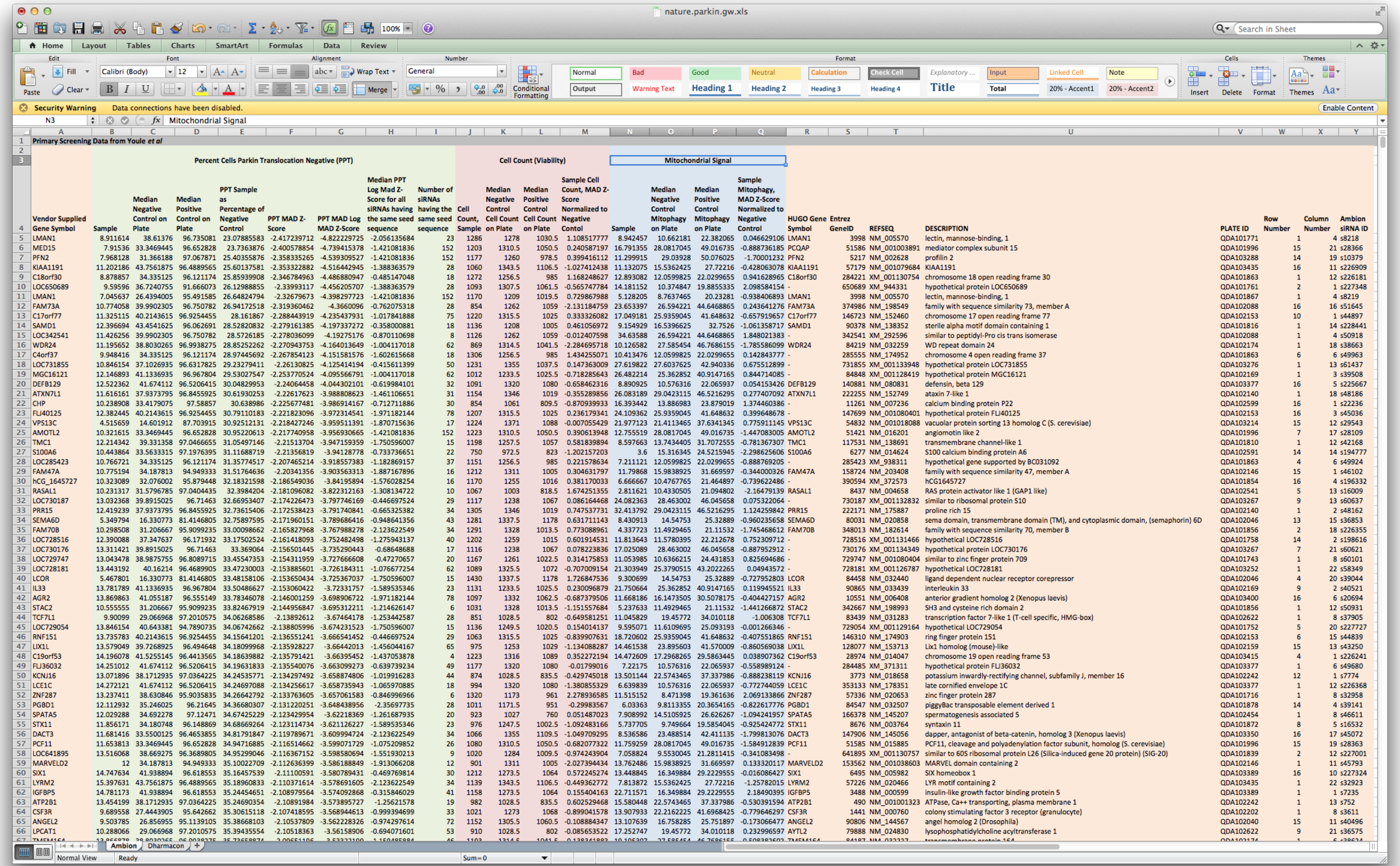
Example Data Set

The data set used in today's lecture comes from an siRNA screen that we published a few years ago. The screen looked for genes that influence parkin translocation.

High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. Hasson SA, Kane LA, Yamano K, Huang CH, Sliter DA, Buehler E, Wang C, Heman-Ackah SM, Hessa T, Guha R, Martin SE, Youle RJ. Nature. 2013.

The data set will be available for download from the lectures portion of the class web page.

Preview the Data



nature.parkin.gw.xls

Security Warning Data connections have been disabled.

N3 *fx* Mitochondrial Signal

Primary Screening Data from Youle *et al*

	Percent Cells Parkin Translocation Negative (PPT)								Cell Count (Viability)				Mitochondrial Signal											
Vendor Supplied Gene Symbol	Sample	Median Negative Control on Plate	Median Positive Control on Plate	PPT Sample as Percentage of Negative Control	PPT MAD Z-Score	PPT MAD Log MAD Z-Score	Median PPT Log Mad Z-Score for all siRNAs having the same seed sequence	Number of siRNAs having the same seed sequence	Cell Count, Sample	Median Negative Control Cell Count on Plate	Median Positive Control Cell Count on Plate	Sample Cell Count, MAD Z-Score Normalized to Negative Contol	Sample	Median Negative Control Mitophagy on Plate	Median Positive Control Mitophagy on Plate	Sample Mitophagy, MAD Z-Score Normalized to Negative Control	HUGO Gene Symbol	Entrez GeneID	REFSEQ	DESCRIPTION	PLATE ID	Row Number	Column Number	Ambion siRNA ID
LMAN1	8.911614	38.61376	96.735081	23.07885583	-2.417239712	-4.822229725	-2.056135684	23	1286	1278	1030.5	1.108517777	8.942457	10.662181	22.382065	0.046629106	LMAN1	3998	NM_005570	lectin, mannose-binding, 1	QDA101771	1	4	s8218
MED15	7.91536	33.3469445	96.652828	23.7363876	-2.400578854	-4.739415378	-1.421081836	152	1203	1310.5	1050.5	0.240587197	16.791355	28.0817045	49.016735	-0.888736185	PCQAP	51586	NM_001003891	mediator complex subunit 15	QDA101996	15	21	s28366
PFN2	7.968128	31.366188	97.067871	25.40355876	-2.358335265	-4.539309527	-1.421081836	152	1177	1260	978.5	0.399416112	11.299915	29.03928	50.076025	-1.70001232	PFN2	5217	NM_002628	profilin 2	QDA103288	14	19	s10379
KIAA1191	11.202186	43.7561875	96.4889565	25.60137581	-2.353322882	-4.516442945	-1.388363579	28	1060	1343.5	1106.5	-1.027412438	11.132075	15.5362425	27.72216	-0.428063078	KIAA1191	57179	NM_001079684	KIAA1191	QDA103435	16	11	s226909
C18orf30	8.878857	34.335125	96.121174	25.85939908	-2.346784963	-4.486880947	-0.485147048	18	1272	1256.5	985	1.168248627	12.893082	12.0599825	22.0299655	0.941628965	C18orf30	284221	XM_001130754	chromosome 18 open reading frame 30	QDA101863	1	12	s226181
LOC650689	9.59596	36.7240755	91.666073	26.12988855	-2.33993117	-4.456205707	-1.388363579	28	1093	1307.5	1061.5	-0.565747784	14.181152	10.374847	19.8855335	2.098584154	-	650689	XM_944331	hypothetical protein LOC650689	QDA101761	2	1	s227348
LMAN1	7.045637	26.4394005	95.491585	26.64824794	-2.32679673	-4.398297723	-1.421081836	152	1170	1209	1019.5	0.729867988	5.128205	8.7637465	20.23281	-0.938406893	LMAN1	3998	NM_005570	lectin, mannose-binding, 1	QDA101867	1	4	s8219
FAM73A	10.774058	39.9902305	96.750782	26.94172518	-2.319360462	-4.3660096	-0.762075318	28	854	1262	1059	-2.131184759	23.653397	26.594221	44.6468865	0.243641276	FAM73A	374986	NM_198549	family with sequence similarity 73, member A	QDA102088	16	16	s51645
C17orf77	11.325115	40.2143615	96.9254455	28.161867	-2.288443919	-4.235437931	-1.017841888	75	1220	1315.5	1025	0.333326082	17.049181	25.9359045	41.648632	-0.657919657	C17orf77	146723	NM_152460	chromosome 17 open reading frame 77	QDA102153	10	1	s44897
SAMD1	12.396694	43.4541625	96.062691	28.52820832	-2.279161385	-4.197337272	-0.358000881	18	1136	1208	1005	0.461056972	9.154929	16.5396625	32.7526	-1.061358717	SAMD1	90378	NM_138352	sterile alpha motif domain containing 1	QDA101816	1	14	s228441
LOC342541	11.426256	39.9902305	96.750782	28.5726185	-2.278036099	-4.19275176	-0.870110698	8	1126	1262	1059	-0.012407598	34.63588	26.594221	44.6468865	1.848021383	-	342541	XM_292596	similar to peptidyl-Pro cis trans isomerase	QDA102088	1	4	s50918
WDR24	11.195652	38.8030265	96.9938275	28.85252262	-2.270943753	-4.164013649	-1.004117018	62	869	1314.5	1041.5	-2.284695718	10.126582	27.585454	46.7686155	-1.785586099	WDR24	84219	NM_032259	WD repeat domain 24	QDA102174	1	18	s38663
C4orf37	9.948416	34.335125	96.121174	28.97445692	-2.267854123	-4.151581576	-1.602615668	18	1306	1256.5	985	1.434255071	10.413476	12.0599825	22.0299655	0.142843777	-	285555	NM_174952	chromosome 4 open reading frame 37	QDA101863	6	6	s49963
LOC731855	10.846154	37.1026935	96.6317825	29.23279411	-2.26130825	-4.125414194	-0.415611399	50	1231	1355	1037.5	0.147363009	27.619822	27.6037625	42.940336	0.675512899	-	731855	XM_001133948	hypothetical protein LOC731855	QDA103276	1	13	s61437
MGC16121	12.146893	41.1336935	96.967804	29.53027547	-2.253770524	-4.095566791	-1.004117018	62	1012	1233.5	1025.5	-0.718285643	26.482214	25.362852	40.9147165	0.844714085	-	84848	XM_001128419	hypothetical protein MGC16121	QDA102169	1	3	s39508
DEFB129	12.522362	41.674112	96.5206415	30.04829953	-2.24064458	-4.044302101	-0.619984101	32	1091	1320	1080	-0.658462316	8.890925	10.576316	22.065937	0.054153426	DEFB129	140881	NM_080831	defensin, beta 129	QDA103377	16	5	s225667
ATXN7L1	11.616161	37.9373795	96.8455925	30.61930253	-2.22617623	-3.988808623	-1.461106651	31	1154	1346	1019	-0.355289856	26.083189	29.0423115	46.5216295	0.277407092	ATXN7L1	222255	NM_152749	ataxin 7-like 1	QDA102140	1	18	s48186
CHP	10.238908	33.4179075	97.58857	30.638986	-2.225677481	-3.986914167	-0.712711886	30	854	1061	809.5	-0.870939933	16.393442	13.886983	23.879019	1.374460386	-	11261	NM_007236	calcium binding protein P22	QDA102599	16	1	s22236
FLJ40125	12.382445	40.2143615	96.9254455	30.79110183	-2.221823096	-3.972314541	-1.971182144	78	1207	1315.5	1025	0.236179341	24.109362	25.9359045	41.648632	0.399648678	-	147699	NM_001080401	hypothetical protein FLJ40125	QDA102153	16	3	s45036
VPS13C	4.515659	14.601912	87.703915	30.92512131	-2.218427246	-3.959511391	-1.870715636	17	1224	1371	1088	-0.007055429	21.977123	21.4113465	37.6341345	0.775911145	VPS13C	54832	NM_001018088	vacuolar protein sorting 13 homolog C (S. cerevisiae)	QDA103214	15	12	s29543
AMOTL2	10.321615	33.3469445	96.652828	30.95220613	-2.217740958	-3.956930665	-1.421081836	152	1223	1310.5	1050.5	0.390613948	12.755519	28.0817045	49.016735	-1.447083005	AMOTL2	51421	NM_016201	angiomotin like 2	QDA101996	7	17	s28109
TMC1	12.214342	39.331358	97.0466655	31.05497146	-2.21513704	-3.947159359	-1.750596007	15	1198	1257.5	1057	0.581839894	8.597663	13.7434405	31.7072555	-0.781367307	TMC1	117531	NM_138691	transmembrane channel-like 1	QDA101810	1	12	s42168
S100A6	10.443864	33.5633315	97.1976395	31.11688719	-2.21356819	-3.94128778	-0.733736651	22	750	972.5	823	-1.202157203	3.6	15.316345	24.5215945	-2.298625606	S100A6	6277	NM_014624	S100 calcium binding protein A6	QDA102591	14	14	s194777
LOC285423	10.766721	34.335125	96.121174	31.35774517	-2.207465214	-3.918557383	-1.182869157	37	1151	1256.5	985	0.221578634	7.211121	12.0599825	22.0299655	-0.888769205	-	285423	XM_938311	hypothetical gene supported by BC031092	QDA101863	4	6	s49924
FAM47A	10.775194	34.187813	94.949333	31.51764636	-2.20341356	-3.903563313	-1.887167896	16	1212	1311	1005	0.304631797	11.79868	15.9838925	31.669597	-0.344000326	FAM47A	158724	NM_203408	family with sequence similarity 47, member A	QDA102146	15	1	s46102
hCG_1645727	10.323089	32.076002	95.879448	32.18321598	-2.186549036	-3.84195894	-1.576028254	16	1170	1255	1016	0.381170033	6.666667	10.4767765	21.464897	-0.739622486	-	390594	XM_372573	hCG1645727	QDA101854	16	4	s196332
RASAL1	10.231317	31.5796785	97.0404435	32.3984204	-2.181096082	-3.822312163	-1.308134722	10	1067	1003	818.5	1.674251355	2.811621	10.4330505	21.094802	-2.16479139	RASAL1	8437	NM_004658	RAS protein activator like 1 (GAP1 like)	QDA102541	5	13	s16009
LOC730187	13.032368	39.8915025	96.71463	32.66953407	-2.174226473	-3.797746169	-0.446697524	29	1117	1238	1067	0.086164468	24.082363	28.463002	46.045658	0.075322064	-	730187	XM_001132832	similar to ribosomal protein S10	QDA103267	9	13	s60637
PRR15	12.419239	37.9373795	96.8455925	32.73615406	-2.172538423	-3.791740841	-0.665325382	34	1305	1346	1019	0.747537731	32.413792	29.0423115	46.5216295	1.124259842	PRR15	222171	NM_175887	proline rich 15	QDA102140	1	2	s48162
SEMA6D	5.349794	16.330773	81.4146805	32.75897595	-2.171960151	-3.789686416	-0.948641356	43	1281	1337.5	1178	0.631711143	8.430913	14.54753	25.32889	-0.960235658	SEMA6D	80031	NM_020858	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6D	QDA102046	13	15	s36853
FAM70B	10.298508	31.206667	95.9099235	33.00098662	-2.165827968	-3.767988278	-2.123622549	34	1291	1328	1013.5	0.773088961	4.337723	11.4929465	21.11532	-1.745468612	FAM70B	348013	NM_182614	family with sequence similarity 70, member B	QDA101856	2	18	s226355
LOC728516	12.390088	37.347637	96.171932	33.17502524	-2.161418093	-3.752482498	-1.275943137	40	1202	1259	1015	0.601914531	11.813643	11.5780395	22.212678	0.752309712	-	728516	XM_001131466	hypothetical LOC728516	QDA101758	14	2	s198616
LOC730176	13.311421	39.8915025	96.71463	33.369064	-2.156501445	-3.735290443	-0.68648688	17	1116	1238	1067	0.078223836	17.025089	28.463002	46.045658	-0.887952912	-	730176	XM_001134349	hypothetical protein LOC730176	QDA103267	7	21	s60621
LOC729747	13.043478	38.9875755	96.8089715	33.45547353	-2.154311959	-3.727666608	-0.47270657	20	1167	1261	1022.5	0.314175853	11.053985	10.6366215	24.431853	0.825694686	-	729747	NM_001080404	similar to zinc finger protein 709	QDA101743	1	8	s60101
LOC728181	13.443192	40.16214	96.4689905	33.47230003	-2.153885601	-3.726184311	-1.076677254	62	1089	1325.5	1072	-0.707009154	21.303949	25.3790515	43.2022265	0.04943572	-	728181	XM_001126787	hypothetical LOC728181	QDA103252	1	22	s58349
LCOR	5.467801	16.330773	81.4146805	33.48158106	-2.153650434	-3.725367037	-1.750596007	15	1430	1337.5	1178	1.726847536	9.300699	14.54753	25.32889	-0.727952803	LCOR	84458	NM_032440	ligand dependent nuclear receptor corepressor	QDA102046	4	20	s39044
IL33	13.781789	41.1336935	96.967804	33.50486627	-2.153060422	-3.72331757	-1.589535346	23	1131	1233.5	1025.5	0.230096879	21.750664	25.362852	40.9147165	0.119945521	IL33	90865	NM_033439	interleukin 33	QDA102169	9	2	s40521
AGR2	13.869863	41.055187	96.555149	33.78346078	-2.146001259	-3.698906722	-1.971182144	78	1097	1332	1062.5	-0.687379506	11.668186	16.1473505	30.5078175	-0.404427157	AGR2	10551	NM_006408	anterior gradient homolog 2 (Xenopus laevis)	QDA103400	16	6	s20694
STAC2	10.555555	31.206667	95.9099235	33.82467919	-2.144956847	-3.695312211	-1.214626147	6	1031	1328	1013.5	-1.151557684	5.237633	11.4929465	21.11532	-1.441266872	STAC2	342667	NM_198993	SH3 and cysteine rich domain 2	QDA101856	1	12	s50931
TCF7L1	9.90099	29.066968	97.2010575	34.06268586	-2.13892612	-3.67464178	-1.253442587	28	851	1028.5	802	-0.649581251	11.045829	19.45772	34.010118	-1.006308	TCF7L1	83439	NM_031283	transcription factor 7-like 1 (T-cell specific, HMG-box)	QDA102622	1	8	s37905
LOC729054	13.846154	40.643381	94.7890735	34.06742662	-2.138805996	-3.674231523	-1.750596007	15	1136	1249.5	1020.5	0.154014137	9.595071	11.6109695	25.093193	-0.001266346	-	729054	XM_001129164	hypothetical LOC729054	QDA101752	5	20	s227727
RNF151	13.735783	40.2143615	96.9254455	34.15641201	-2.136551241	-3.666541452	-0.446697524	29	1063	1315.5	1025	-0.839907631	18.720602	25.9359045	41.648632	-0.407551865	RNF151	146310	NM_174903	ring finger protein 151	QDA102153	6	15	s44839
LIX1L	13.579049	39.7268925	96.494648	34.18099968	-2.135928227	-3.66442013	-1.456044167	65	975	1253	1029	-1.134088287	14.461538	23.895603	41.570009	-0.860569038	LIX1L	128077	NM_153713	Lix1 homolog (mouse)-like	QDA102159	15	13	s43250
C19orf53	14.196078	41.5255145	96.4413565	34.18639882	-2.135791421	-3.66395452	-1.437053878	4	1223	1316	1089	0.352272194	14.472609	17.2968265	29.5863445	0.038907362	C19orf53	28974	NM_014047	chromosome 19 open reading frame 53	QDA103415	4	1	s226241
FLJ36032	14.251012	41.674112	96.5206415	34.19631833	-2.135540076	-3.663099273	-0.639739234	49	1177	1320	1080	-0.01799016	7.22175	10.576316	22.065937	-0.558989124	-	284485	XM_371311	hypothetical protein FLJ36032	QDA103377	1	6	s49680
KCNJ16	13.071896	38.1712935	97.0364225	34.24535771	-2.134297492	-3.658874806	-1.019916283	44	874	1028.5	835.5	-0.429745018	13.501144	22.5743465	37.337986	-0.888238119	KCNJ16	3773	NM_018658	potassium inwardly-rectifying channel, subfamily J, member 16	QDA102242	12	1	s7774
LCE1C	14.272121	41.674112	96.5206415	34.24697088	-2.134256617	-3.658735943	-1.065970885	18	994	1320	1080	-1.380855329	6.639839	10.576316	22.065937	-0.772744059	LCE1C	353133	NM_178351	late cornified envelope 1C	QDA103377	1	12	s226368
ZNF287	13.237411	38.630846	95.9035835	34.26642792	-2.133763605	-3.657061583	-0.846996966	6	1320	1173	961	2.278936585	11.515152	8.471398	19.361636	2.069133866	ZNF287	57336	NM_020653	zinc finger protein 287	QDA101716	1	8	s32958
PGBD1	12.112932	35.246025	96.21645	34.36680307	-2.131220251	-3.648438956	-2.35697735	28	1011	1171.5	951	-0.29983567	6.03363	9.8113355	20.3654165	-0.822617776	PGBD1	84547	NM_032507	piggyBac transposable element derived 1	QDA101878	14	4	s39141
SPATA5	12.029288	34.692278	97.12471	34.67425229	-2.123429954	-3.62218369	-1.261687935	20	923	1027	760	0.051487023	7.908992	14.5105925	26.626267	-1.094241957	SPATA5	166378	NM_145207	spermatogenesis associated 5	QDA102454	1	8	s46611
STX11	11.856171	34.180748	96.148869	34.68669264	-2.123114734	-3.621126227	-1.589535346	23	976	1247.5	1002.5	-1.092483166	5.737705	9.749664	19.5854045	-0.925424772	STX11	8676	NM_003764	syntaxin 11	QDA101872	8	5	s16532
DACT3	11.681416	33.5500125	96.4653855	34.81791847	-2.119789671	-3.609994724	-2.123622549	34	1066	1355	1109.5	-1.049709295	8.536586	23.488514	42.411135	-1.799813076	DACT3	147906	NM_145056	dapper, antagonist of beta-catenin, homolog 3 (Xenopus laevis)	QDA103350	16	17	s45072
PCF11	11.653813	33.3469445	96.652828	34.94716885	-2.116514662	-3.599071729	-1.075209852	26	1080	1310.5	1050.5	-0.682077322	11.759259	28.0817045	49.016735	-1.584912839	PCF11	51585	NM_015885	PCF11, cleavage and polyadenylation factor subunit, homolog (S. cerevisiae)	QDA101996	15	19	s28363
LOC641895	13.516068	38.669275	96.3689805	34.95299046	-2.116367152	-3.598580694	-1.551930213	9	1020	1284	1009.5	-0.974243904	7.058824	9.5530045	21.2811415	-0.341083498	-	641895	XM_001130757	similar to 60S ribosomal protein L26 (Silica-induced gene 20 protein) (SIG-20)	QDA101839	2	12	s227001
MARVELD2	12	34.187813	94.949333	35.10022709	-2.112636399	-3.586188849	-1.913066208	12	901	1311	1005	-2.027394434	13.762486	15.9838925	31.669597	0.133320117	MARVELD2	153562	NM_001038603	MARVEL domain containing 2	QDA102146	1	11	s45793
SIX1	14.747634	41.938894	96.618553	35.16457539	-2.11100591	-3.580789431	-0.469769814	30	1212	1273.5	1064	0.572245274	13.448845	16.349884	29.2229555	-0.016086427	SIX1	6495	NM_005982	SIX homeobox 1	QDA103389	16	10	s227324
LYRM2	15.397631	43.7561875	96.4889565	35.18960833	-2.110371614	-3.578691605	-2.123622549	34	1139	1343.5	1106.5	-0.449362772	7.813872	15.5362425	27.72216	-1.25782015	LYRM2	57226	NM_020466	LYR motif containing 2	QDA103435	1	22	s32923
IGFBP5	14.781173	41.938894	96.618553	35.24454651	-2.108979564	-3.574092868	-0.315846029	41	1158	1273.5	1064	0.155404163	22.711571	16.349884	29.2229555	2.18490395	IGFBP5	3488	NM_000599	insulin-like growth factor binding protein 5	QDA103389	1	1	s7235
ATP2B1	13.454199	38.1712935	97.0364225	35.24690354	-2.10891984	-3.573895727	-1.25621578	19	982	1028.5	835.5	0.602529468	15.580448	22.5743465	37.337986	-0.530391594	ATP2B1	490	NM_001001323	ATPase, Ca++ transporting, plasma membrane 1	QDA102242	1	13	s752
CSF3R	9.689558	27.4443905	95.642662	35.30615118	-2.107418595	-3.568944613	-0.999394699	33	1021	1273	1068	-0.899041578	13.907933	22.2162225	41.6968425	-0.779646297	CSF3R	1441	NM_000760	colony stimulating factor 3 receptor (granulocyte)	QDA102202	1	8	s3611
ANGEL2	9.503785	26.856955	95.1139105	35.38668103	-2.10537809	-3.562228326	-0.974297614	72	1152	1305.5	1060.5	-0.108884347	13.107639	16.758285	25.751897	-0.173066477	ANGEL2	90806	NM_144567	angel homolog 2 (Drosophila)	QDA102040	15	11	s40496
LPCAT1	10.288066	29.066968	97.2010575	35.39435554	-2.10518363	-3.56158906	-0.694071601	53	910	1028.5	802	-0.085653522	17.252747	19.45772	34.010118	0.232996597	AYTL2	79888	NM_024830	lysophosphatidylcholine acyltransferase 1	QDA102622	9	21	s36575

Import the data

```
ambion <- read.csv("nature.parkin.gw.ambion.csv", stringsAsFactors = FALSE)
str(ambion)
```

```
## 'data.frame':    65196 obs. of  25 variables:
##  $ Vendor.Supplied.Gene.Symbol                                    : chr
##  $ Sample                                                         : num
##  $ Median.Negative.Control.on.Plate                               : num
##  $ Median.Positive.Control.on.Plate                               : num
##  $ PPT.Sample.as.Percentage.of.Negative.Control                   : num
##  $ PPT.MAD.Z.Score                                                : num
##  $ PPT.MAD.Log.MAD.Z.Score                                        : num
##  $ Median.PPT.Log.Mad.Z.Score.for.all.siRNAs.having.the.same.seed.sequence: num
##  $ Number.of.siRNAs.having.the.same.seed.sequence                 : int
##  $ Cell.Count..Sample                                             : int
##  $ Median.Negative.Control.Cell.Count.on.Plate                    : num
##  $ Median.Positive.Control.Cell.Count.on.Plate                    : num
##  $ Sample.Cell.Count..MAD.Z.Score.Normalized.to.Negative.Contol   : num
##  $ Sample.1                                                       : num
##  $ Median.Negative.Control.Mitophagy.on.Plate                     : num
##  $ Median.Positive.Control.Mitophagy.on.Plate                     : num
```

Check for missing data

```r
apply(ambion, 2, function(x) sum(is.na(x)))
```

```
##                                           Vendor.Supplied.Gene.Symbol
##                                                                     0
##                                                                Sample
##                                                                   441
##                                          Median.Negative.Control.on.Plate
##                                                                   441
##                                          Median.Positive.Control.on.Plate
##                                                                   441
##                              PPT.Sample.as.Percentage.of.Negative.Control
##                                                                   441
##                                                         PPT.MAD.Z.Score
##                                                                   441
##                                                     PPT.MAD.Log.MAD.Z.Score
##                                                                   441
## Median.PPT.Log.Mad.Z.Score.for.all.siRNAs.having.the.same.seed.sequence
##                                                                   441
##                              Number.of.siRNAs.having.the.same.seed.sequence
##                                                                   441
```

Investigate Missing Data

```
ambion[is.na(ambion[,1]),][1,]
```

```
##    Vendor.Supplied.Gene.Symbol Sample Median.Negative.Control.on.Plate
## NA                        <NA>     NA                               NA
##    Median.Positive.Control.on.Plate
## NA                               NA
##    PPT.Sample.as.Percentage.of.Negative.Control PPT.MAD.Z.Score
## NA                                           NA              NA
##    PPT.MAD.Log.MAD.Z.Score
## NA                      NA
##    Median.PPT.Log.Mad.Z.Score.for.all.siRNAs.having.the.same.seed.sequence
## NA                                                                      NA
##    Number.of.siRNAs.having.the.same.seed.sequence Cell.Count..Sample
## NA                                             NA                 NA
##    Median.Negative.Control.Cell.Count.on.Plate
## NA                                          NA
##    Median.Positive.Control.Cell.Count.on.Plate
## NA                                          NA
##    Sample.Cell.Count..MAD.Z.Score.Normalized.to.Negative.Contol Sample.1
## NA                                                           NA       NA
```

Eliminate Missing Data

```
ambion <- ambion[! is.na(ambion[,2]), ]
apply(ambion, 2, function(x) sum(is.na(x)))
```

```
##                                          Vendor.Supplied.Gene.Symbol
##                                                                    0
##                                                               Sample
##                                                                    0
##                                      Median.Negative.Control.on.Plate
##                                                                    0
##                                      Median.Positive.Control.on.Plate
##                                                                    0
##                          PPT.Sample.as.Percentage.of.Negative.Control
##                                                                    0
##                                                      PPT.MAD.Z.Score
##                                                                    0
##                                              PPT.MAD.Log.MAD.Z.Score
##                                                                    0
## Median.PPT.Log.Mad.Z.Score.for.all.siRNAs.having.the.same.seed.sequence
##                                                                    0
##                            Number.of.siRNAs.having.the.same.seed.sequence
```

Simplify the Data

Often it will be helpful to create a new data frame with only the data we wish to analyze.

```
ambion.simple <- ambion[,c(19,25,1,21,7,13,17)]
ambion.simple[1,]
```

```
##   Entrez.GeneID Ambion.siRNA.ID Vendor.Supplied.Gene.Symbol
## 1          3998           s8218                       LMAN1
##                DESCRIPTION PPT.MAD.Log.MAD.Z.Score
## 1 lectin, mannose-binding, 1                -4.82223
##   Sample.Cell.Count..MAD.Z.Score.Normalized.to.Negative.Contol
## 1                                                      1.108518
##   Sample.Mitophagy..MAD.Z.Score.Normalized.to.Negative.Control
## 1                                                    0.04662911
```

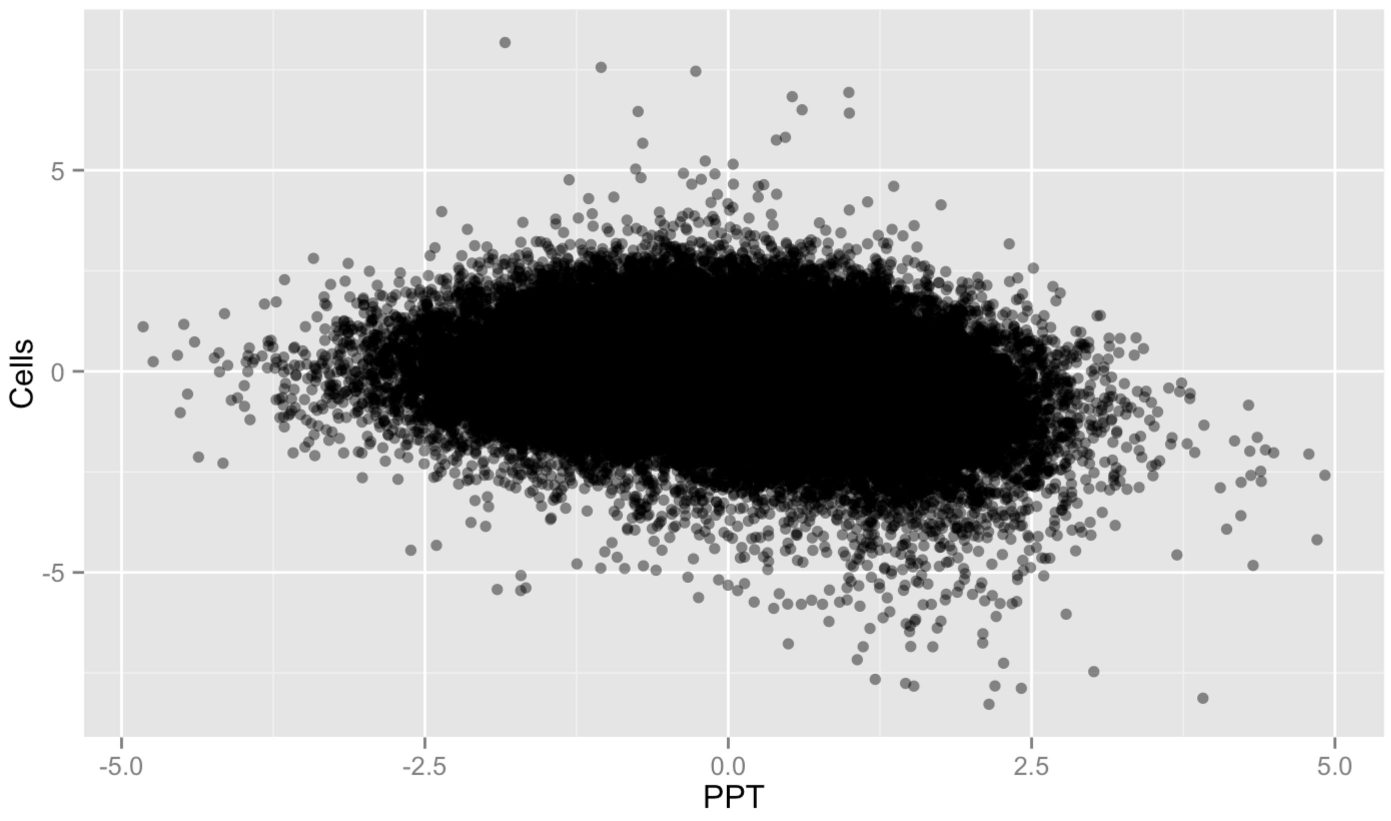
Simplify our Column Names

```r
library(knitr, quietly = TRUE)
colnames(ambion.simple) <- c("GeneID","siRNA","Symbol","Description",
                             "PPT","Cells","Mitophagy")
kable(head(ambion.simple, n=4), format = "markdown")
```

GeneID	siRNA	Symbol	Description	PPT	Cells	Mitophagy
3998	s8218	LMAN1	lectin, mannose-binding, 1	-4.822230	1.1085178	0.0466291
51586	s28366	MED15	mediator complex subunit 15	-4.739415	0.2405872	-0.8887362
5217	s10379	PFN2	profilin 2	-4.539309	0.3994161	-1.7000123
57179	s226909	KIAA1191	KIAA1191	-4.516443	-1.0274124	-0.4280631

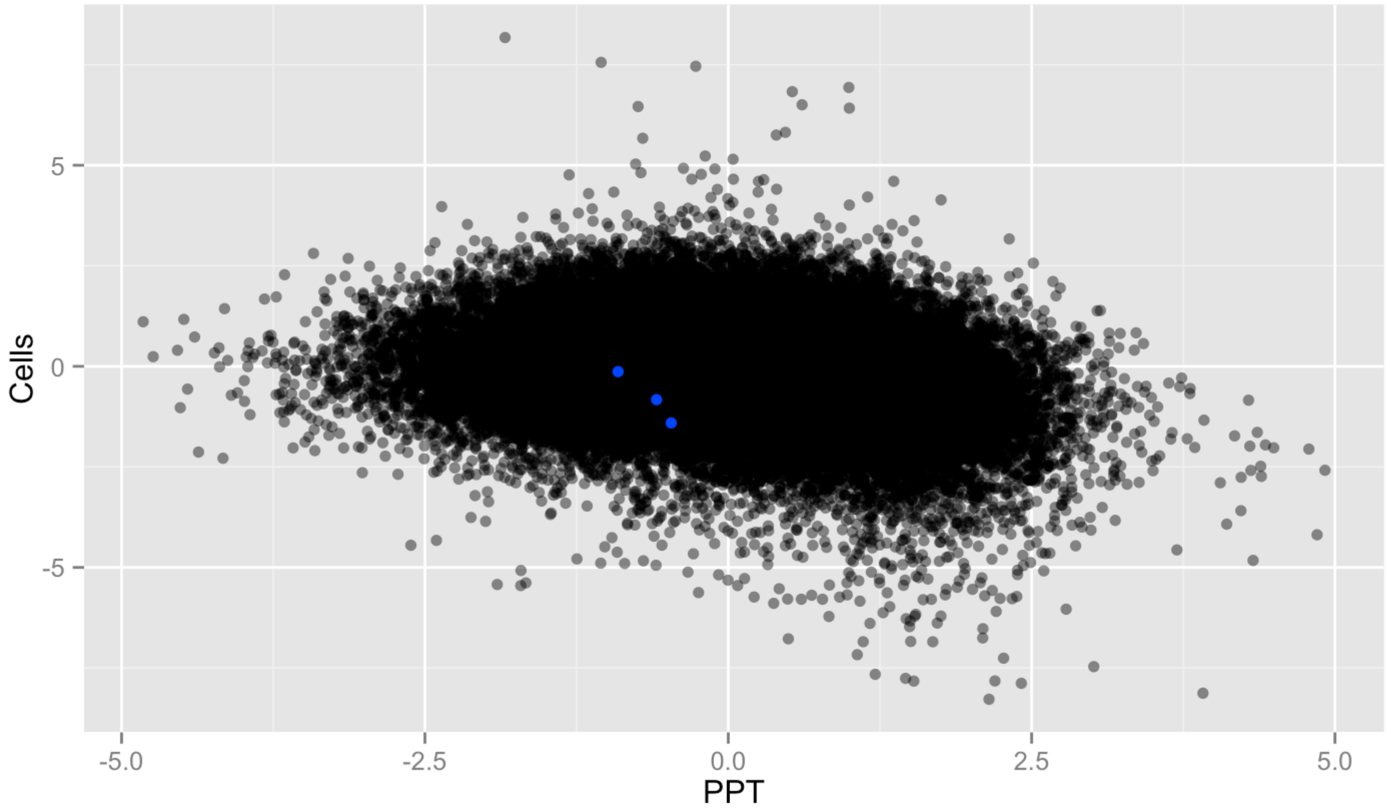
Evaluate how our variables interact

```
library(ggplot2, quietly = TRUE)
ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_point(alpha=0.5)
```



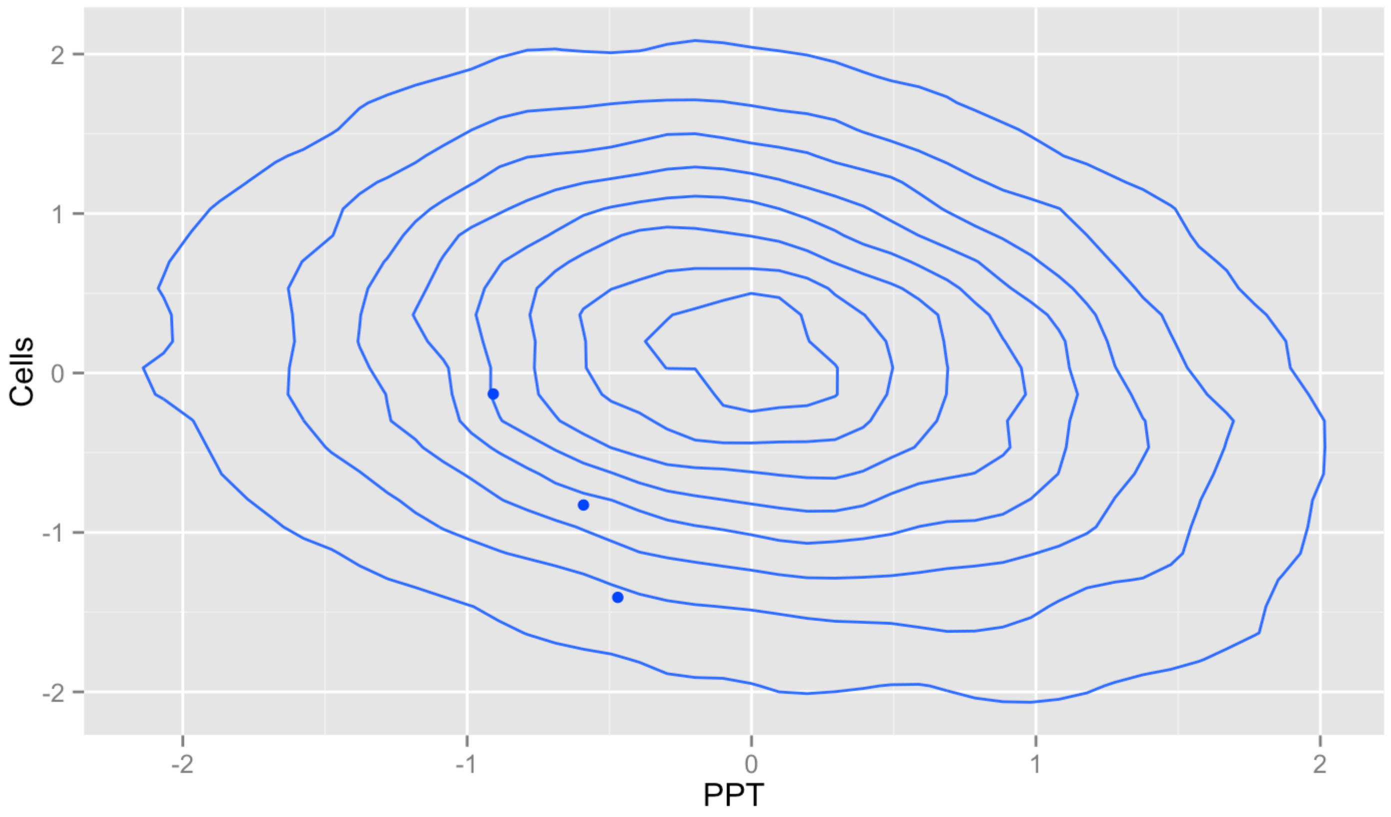
Evaluate how our variables interact

```
ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_point(alpha=0.5) +
    geom_point(data = ambion.simple[ambion.simple$Symbol == "PARK2",],
               color="blue")
```



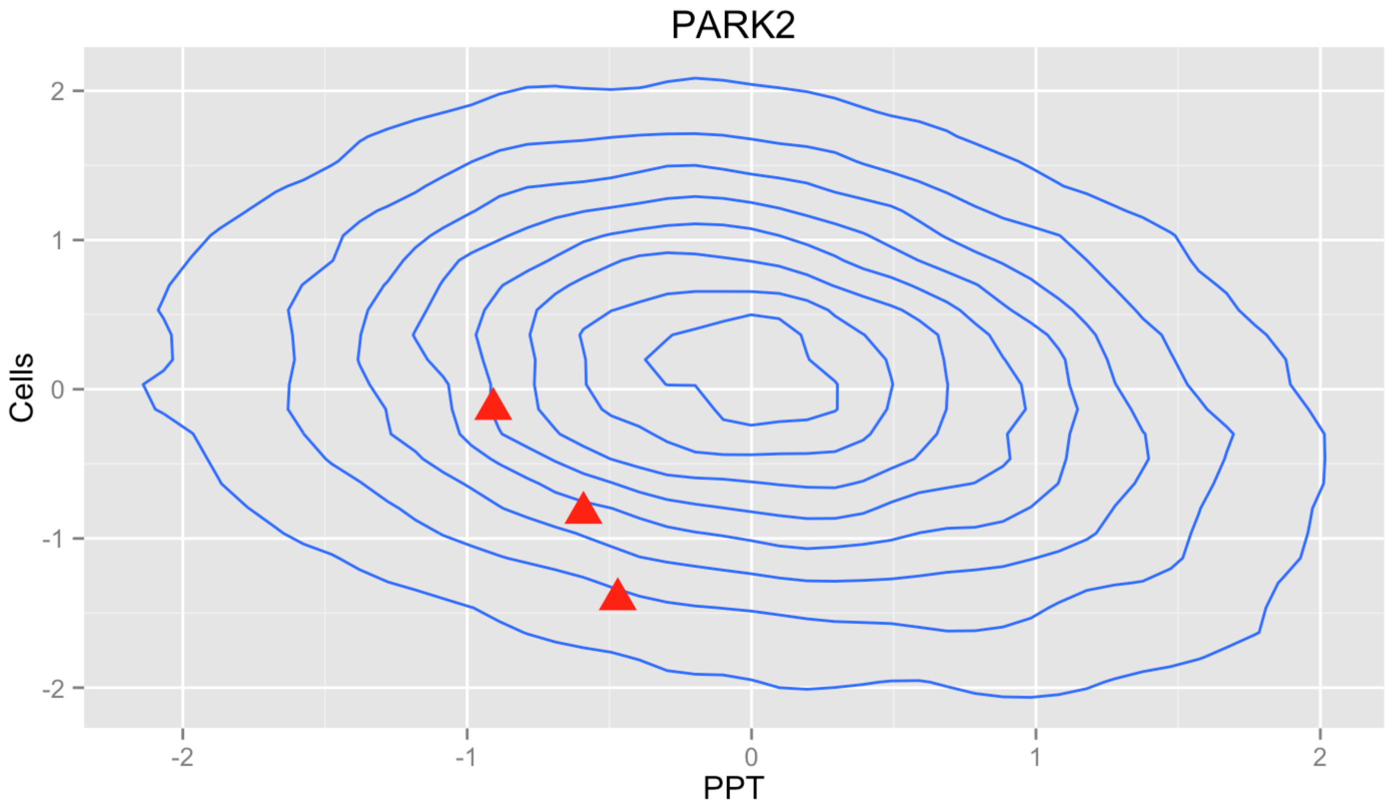
Refine the plot

```
ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_density2d() +
    geom_point(data = ambion.simple[ambion.simple$Symbol == "PARK2",],
               color="blue")
```



Further refine the plot

```
ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_density2d() +
    geom_point(data = ambion.simple[ambion.simple$Symbol == "PARK2",],
               color="red", shape=17, size =5) +
        ggtitle("PARK2")
```



Adding gene description

```
description <- ambion.simple$Description[ambion.simple$Symbol == "PARK2"][1]
description
```

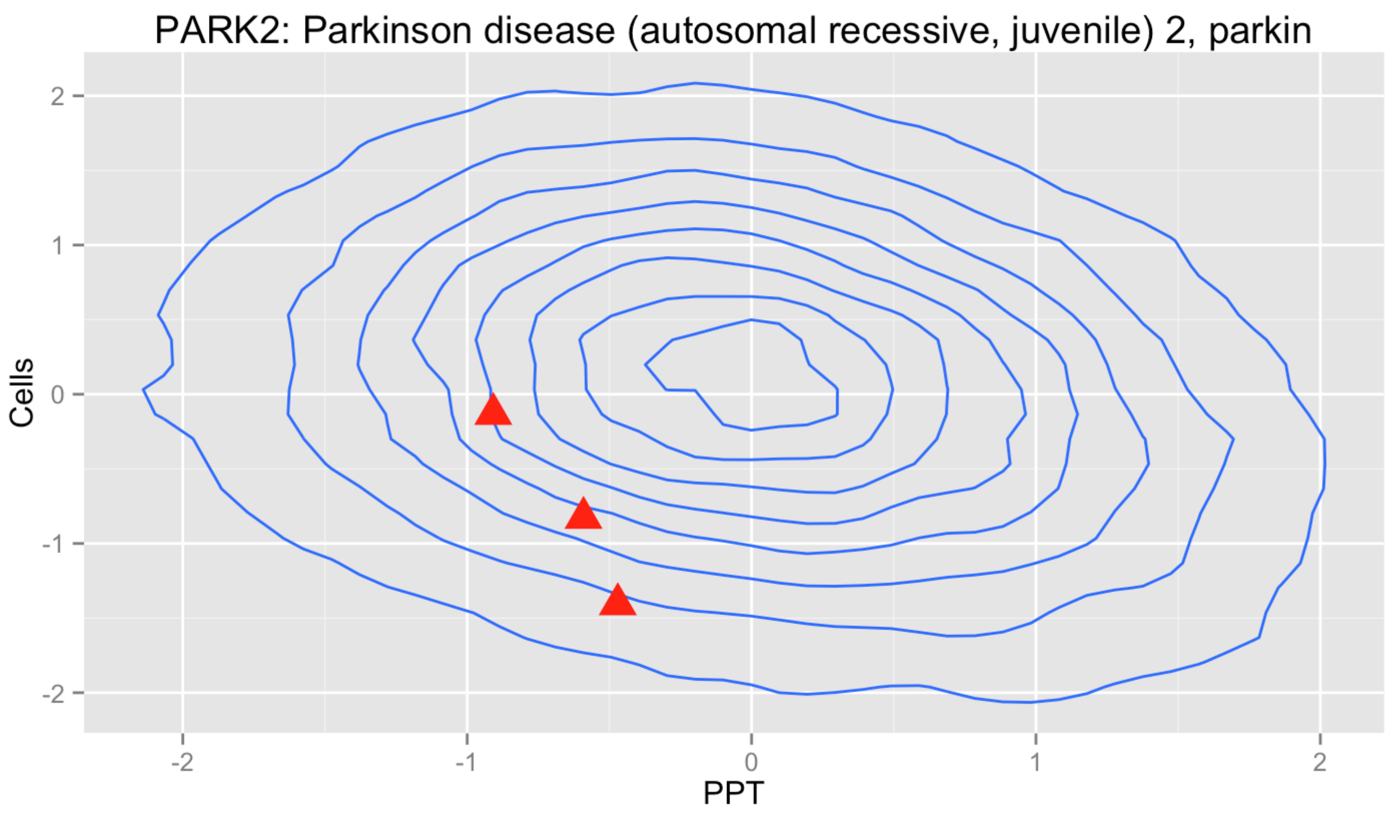
```
## [1] "Parkinson disease (autosomal recessive, juvenile) 2, parkin"
```

```
myTitle <- paste("PARK2",description,sep=": ")
myTitle
```

```
## [1] "PARK2: Parkinson disease (autosomal recessive, juvenile) 2, parkin"
```

Final version of plot

```
ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_density2d() +
    geom_point(data = ambion.simple[ambion.simple$Symbol == "PARK2",],
               color="red", shape=17, size =5) +
        ggtitle(myTitle)
```



Making the refined plot into a function

Now that we have our custom plot looking right, we would like to be able to do the same for other genes but without so much typing. First, make a new R Script in RStudio:

Constructing a new function from your history

Frequently, making a function will simply be a function of selecting the right parts of your history and hitting the "to source" button.

Function with PARK2 hard coded

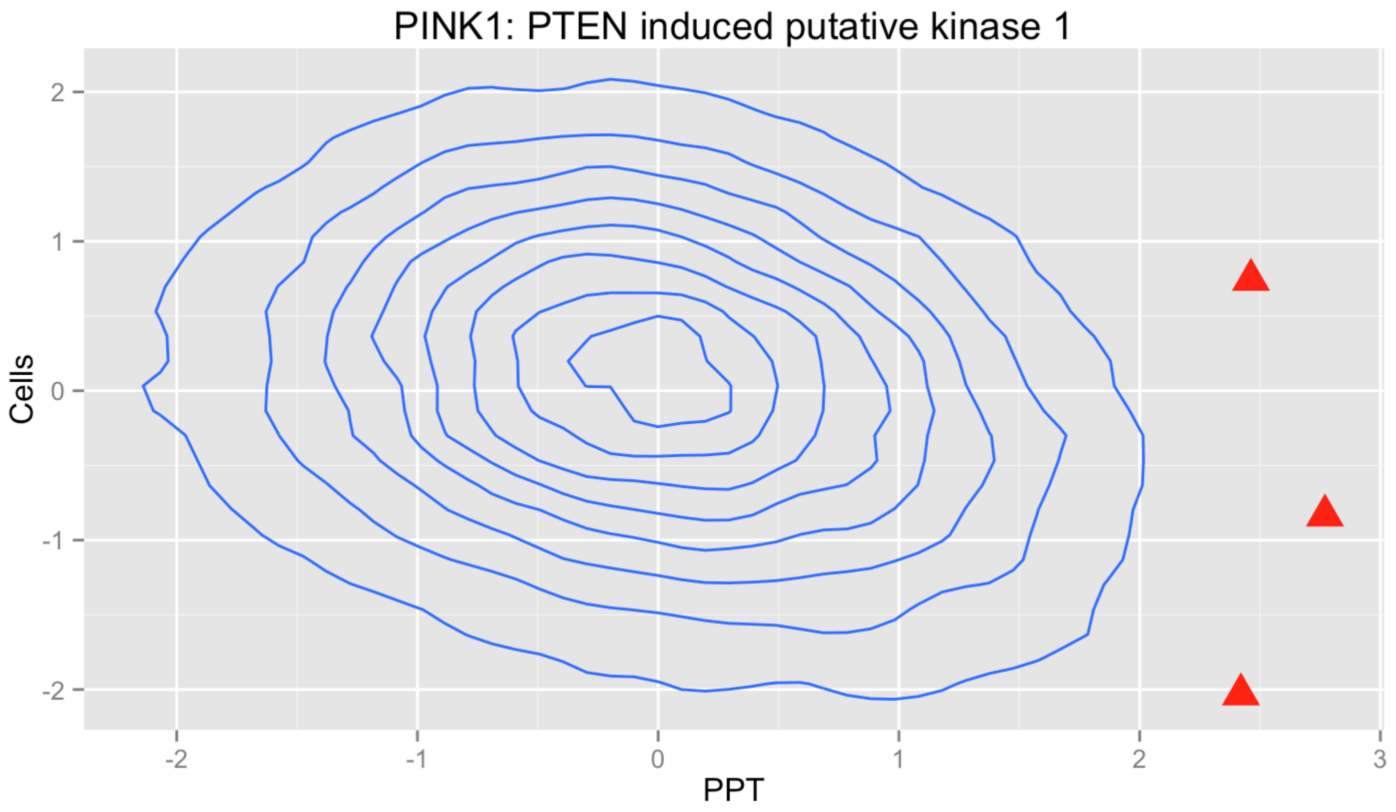
```r
graphGene <- function(gene) {
  description <- ambion.simple$Description[ambion.simple$Symbol == "PARK2"][1]
  myTitle <- paste("PARK2",description,sep=": ")
  ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_density2d() +
    geom_point(data = ambion.simple[ambion.simple$Symbol == "PARK2",],
               color="red", shape=17, size =5) +
    ggtitle(myTitle)
}
```

Function made generic

```r
graphGene <- function(gene) {
  description <- ambion.simple$Description[ambion.simple$Symbol == gene][1]
  myTitle <- paste(gene,description,sep=": ")
  ggplot(ambion.simple, aes(x=PPT, y=Cells)) + geom_density2d() +
    geom_point(data = ambion.simple[ambion.simple$Symbol == gene,],
               color="red", shape=17, size =5) +
    ggtitle(myTitle)
}
```

Our function in action

```
graphGene("PINK1")
```



Default values for function arguments

```
pdfGene <- function(gene, file=paste(gene,".pdf", sep="")) {
    pdf(file, width=5, height=5)
    graphGene(gene)
    dev.off()
}
```

Passing on extra arguments to our function

We can use the ellipse notation (…) to indicate that extra arguments to our function should be passed on to a function that is inside our function (in this case pdf).

```
pdfGene <- function(gene, file=paste(gene,".pdf", sep=""), ...) {
    pdf(file, ...)
    graphGene(gene)
    dev.off()
}
pdfGene("PINK1", width=10, height=10)
```

```
## quartz_off_screen
##                  2
```

Control of Flow: If/Else

We can decide whether something happens in our function using "if" and "if/else".

```
sillyFunction <- function(x) {
  if (x < 5) {
    returnValue <- x
  }
  else {
    returnValue <- x / 2
  }
  returnValue
}
sillyFunction(12)
```

```
## [1] 6
```

Control of Flow: For

```
pdfGenes <- function(genes, file=paste(gene,".pdf", sep=""), ...) {
    pdf(file, ...)
    for (gene in genes) {
      graphGene(gene)
    }
    dev.off()
}
pdfGenes(c("PLK1","PINK1","BRCA1"))
```