knit: (function(inputFile, encoding) { rmarkdown::render(inputFile, encoding = encoding, output_file = file.path(dirname(inputFile), paste0(Sys.Date(), "_Elys.NPC.pdf"))) })

1. Elys ChIP-seq peaks (Capelson) and Nup98 nucleoplasmic (nuc) and nuclear pore-associated (npc) HMM domains

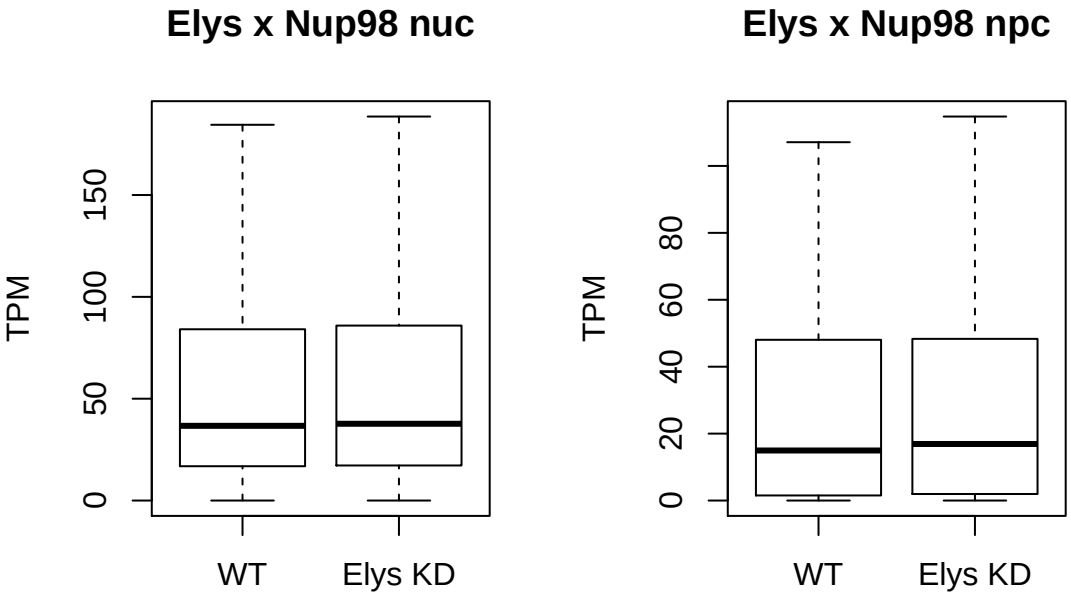
Elys ChIP-seq peaks from S2 cells overlap with Nup98 nuc HMM domains non-randomly (p-value $< 10^{-4}$). Overlap with Nup98 npc domains is mostly random (p-value = 0.8588).

There are **5981** Elys ChIP-seq peaks that overlap with nucleoplasmic Nup98 domains. Summary of these domains lengths:

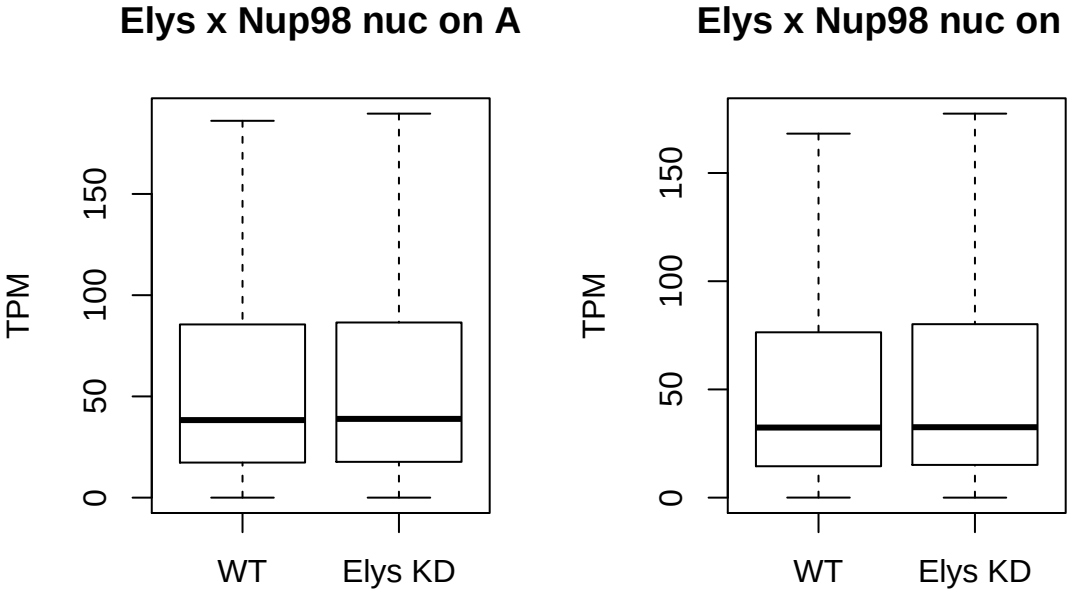
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
135	423	728	953.9632	1239	8630

2. Genes which promoters reside in Elys peaks defined above

There are **5216** genes, which TSSs are in Elys peaks that intersect with Nup98 nuc and **1040** genes which TSSs overlap with peaks that overlap with Nup98 npc. **812** genes are in both groups, so they were excluded. Expression of genes in either of groups above doesn't significantly change upon Elys KD:



For genes which TSSs reside in Elys peaks that overlap with nucleoplasmic Nup98 domains expression level doesn't change whether we take into account those that located on the X-chromosome or autosomes

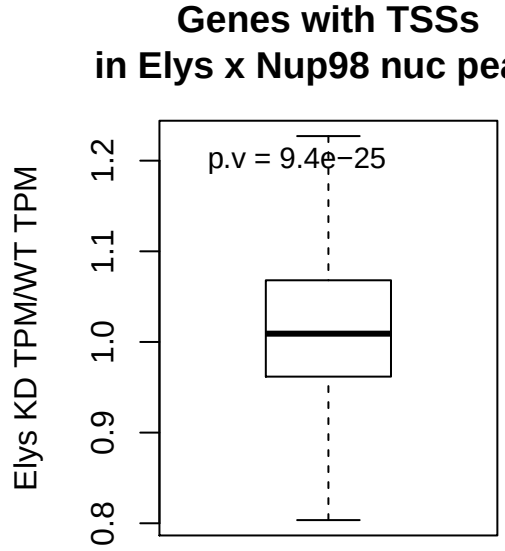


3. Genes with different levels of expressions and their expression level change upon Elys KD

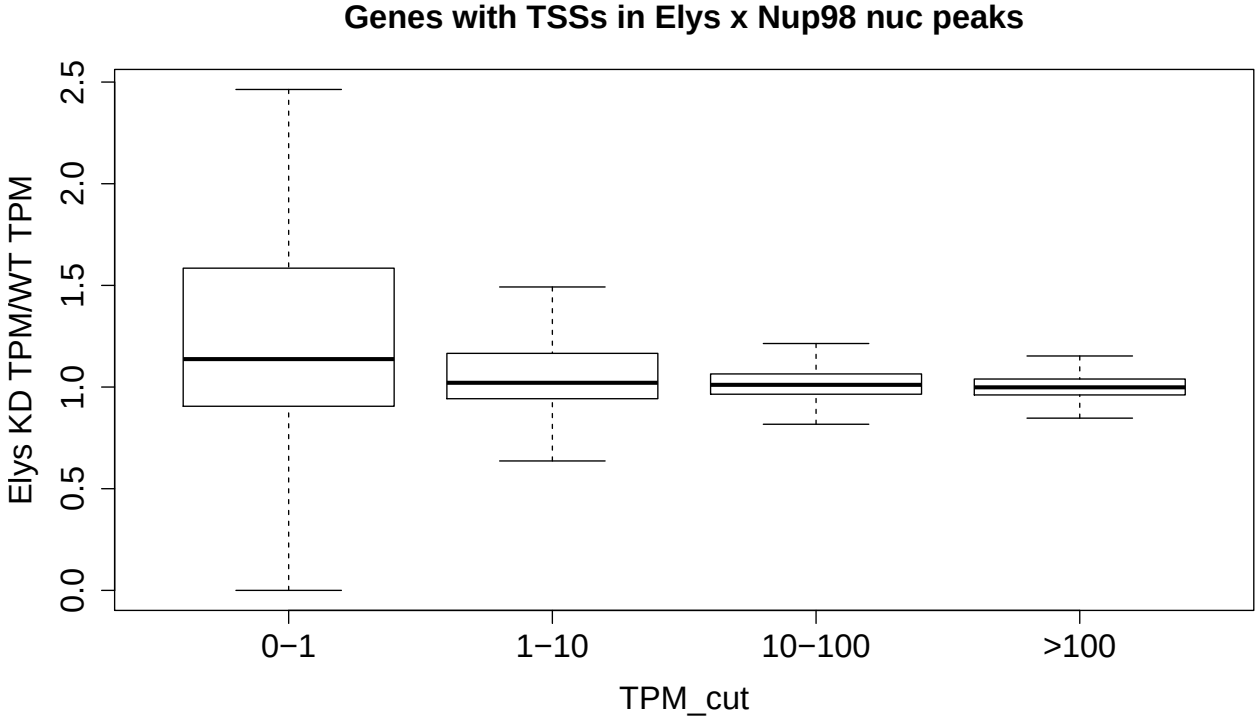
Genes were split based on their expression into 4 groups: **0-1** TPM, **1-10** TPM, **10-100** TPM and **>100** TPM. In either of these groups expression changes between WT and KD are insignificant:

TPM_cut	**ElysKD_TPMav**	**WT_TPMav**	**EKD_norm_WTav**	**n**	**pv**
0-1	0.40	0.34	1.28	150	0.343
1-10	6.15	5.66	1.10	448	0.0767
10-100	40.91	39.81	1.02	2592	0.227
>100	691.79	697.04	1.00	848	0.752

4. The same only for ratios of TPMs in KD vs WT



For genes split by expression level:



TPM_cut	ElysKD_TPMav	WT_TPMav	EKD_norm_WTav	n	pv
0-1	0.40	0.34	1.28	150	3.94e-06
1-10	6.15	5.66	1.10	448	3.85e-08
10-100	40.91	39.81	1.02	2592	1.23e-21
>100	691.79	697.04	1.00	848	0.493

5. HMM3 for Elys and Nup98 domains; HMM2 for LADs

YS decided that we should re-identify Elys and NUP98 domains using HMM with 3 states (using only the highest score category as domains). In the case of Elys he wanted to identify HMM3 domains for each of the biological replica and then use only their intersection. However, he wasn't pleased with the results and decided to stick with the previously made HMM3 domains for merged replicas. In the case of Nup resulted domains were predictably more narrow than HMM2 ones:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Sum
NPC HMM2	0	1470.25	2108.0	2830.472	3600.00	35123	20056726
NPC HMM3	**261**	**1204.00**	**1812.5**	**2269.014**	**2809.50**	**16817**	**20112539**
NUC HMM2	10	1795.00	3643.0	5886.832	7621.00	82029	22787927
NUC HMM3	**261**	**1502.00**	**2672.0**	**3745.508**	**4820.25**	**49517**	**21004807**

6. H3K27Ac, GAF and pBAP/BAP domains

Next, I had to retrieve some additional data to compare with Nup98 domains. In the case of H3K27Ac ChIP-on-chip I recalculated coverage of peaks using 300nt bins, because R was crashing when tried parsing original data, and used HMM3. pBAP/BAP and GAF peaks were already given. Then I used permutation tests to assess non-randomness of overlaps but noticed some differences when I was using tests based on length of overlaps of randomly shuffled domains and when was using Jaccard index. So I decided to compile a table with results received from both flavours of permutation test:

NUP	**Interact**	**ix**	**%NUP**	**%Interact**	**p.v.width**	**p.v.jac**
NPC	H3K27Ac	1913467	0.095	0.158	0.314	0.575
NUC	H3K27Ac	9178672	0.437	0.757	0.000	0.000
NPC	GAF	492161	0.024	0.229	0.000	0.000
NUC	GAF	1642387	0.078	0.765	0.000	0.000
NPC	pBAP/BAP	190281	0.009	0.178	0.002	1.000
NUC	pBAP/BAP	314680	0.015	0.294	0.000	1.000

7. Overlap of embryonic Elys domains with Nup98 domains

In this table I compiled data of Elys domains' overlap with Nup98 HMM3 domains. Though we decided to use HMM3 domains for merged replicas of Elys, I also used ones which were made based on intersection of domains obtained for each of the replicas, but the difference is not very dramatic:

HMM3	**NUP**	**ix**	**%NUP**	**%Elys**	**p.v.width**	**p.v.jac**
merged	NPC	10777194	0.536	0.194	0	0
merged	NUC	16318921	0.777	0.294	0	0
by each rep	NPC	4694295	0.233	0.165	0	0
by each rep	NUC	11863264	0.565	0.416	0	0

8. New approach for Nup98 domains segregation

YS found out that there is a substantional overlap between Nup98 NPC and nucleoplasmic domains. Based on this, we decided to throw out regions in both of those domain types that overlap with each other. Resulted domains are designated Nup98 NPC* and Nup98 Nuc*. Also, it was decided not to use Jaccard tests.

In the case of H3K27ac, GAF and pBAP/BAP, Nuc* domains showed non-randomness for all of the proteins/protein complexes, unlike NPC* domains:

NUP	Interact	ix	%NUP length	%Interact length	p.v.width	%NUP domains N	%Interact domains N
NPC*	H3K27Ac	447241	0.026	0.037	1.00	0.052	0.120
NUC*	H3K27Ac	7712446	0.434	0.636	0.00	0.522	0.769
NPC*	GAF	118866	0.006	0.055	1.00	0.022	0.076
NUC*	GAF	1269092	0.060	0.591	0.00	0.220	0.680
NPC*	pBAP/BAP	137412	0.007	0.128	0.67	0.049	0.137
NUC*	pBAP/BAP	261811	0.012	0.245	0.00	0.120	0.264

The same goes for the intersection with Elys, except for embryos' domains where replicas were merged:

Tissue	HMM3	NUP	ix	%NUP length	%Elys length	p.v.width	%NUP domains N	%Elys domains N
S2	NA	NPC*	463157	0.027	0.044	1	0.140	0.076
S2	NA	NUC*	4376692	0.246	0.415	0	0.668	0.353
Embryo	merged	NPC*	5708689	0.338	0.207	0	0.596	0.434
Embryo	merged	NUC*	8251511	0.464	0.299	0	0.641	0.204
Embryo	by each rep	NPC*	2945591	0.175	0.103	1	0.310	0.363
Embryo	by each rep	NUC*	10114560	0.569	0.355	0	0.636	0.396

Then it was decided to fill in Nup98 NPC* and NUC* domains all the gaps that are less than 900 bp