

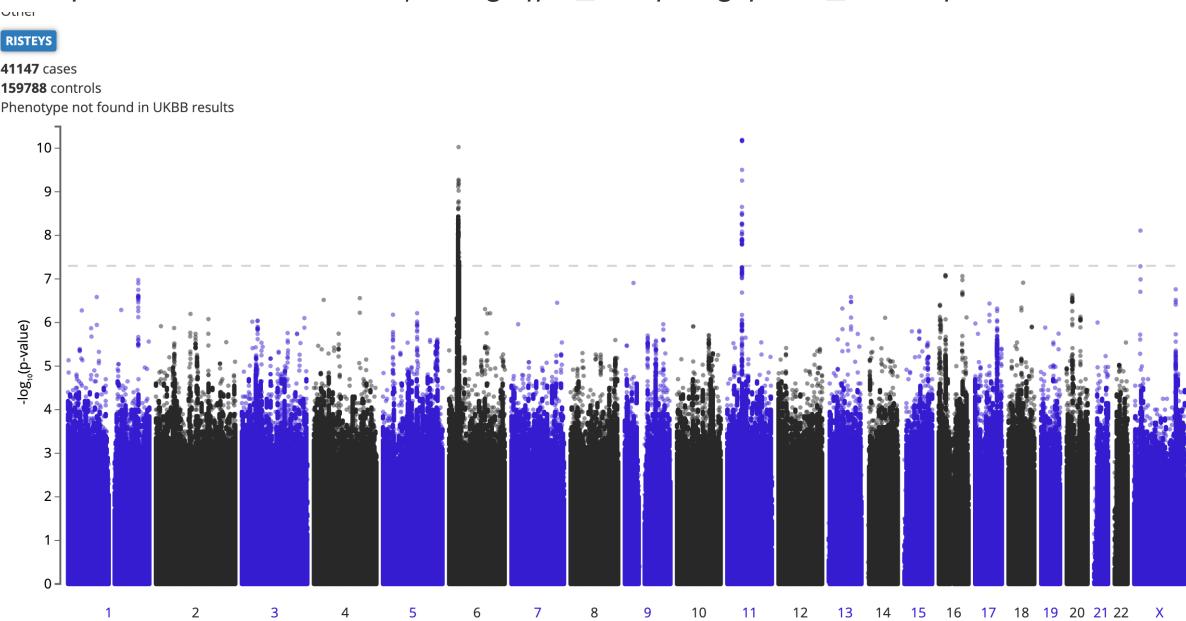
Drug purchase : SSRI R5

SSRI (N06AB) vs Lithium vs NDRI

R log: /Users/kumMar/Dropbox (Partners

HealthCare)/Daly_Lab/drugResponseAnalysis/SSRIanalysis.R

From previous version of dev, and gs://r5_data/drugs/case_control/



chr	pos	ref	alt	locus	rsid	nearest gene	consequence	INFO	FIN enrichment	af	af cases	af controls	OR
11	46035522	T	C	locus	rs12364441	PHF21A	intron	0.998	1.17	0.194	0.203	0.192	1.08
6	29978172	G	A	locus	rs3132685	HLA-A	intron	1.00	0.675	0.0760	0.0693	0.0778	0.89
6	28828079	A	T	locus	rs3135296	TRIM27	intergenic	0.999	0.605	0.0515	0.0459	0.0529	0.88
6	29367377	T	C	locus	rs9257793	OR5V1	intron	0.998	0.511	0.0554	0.0496	0.0569	0.88
6	28235278	T	G	locus	rs13205211	ZSCAN9	downstream gene	1.00	0.614	0.0517	0.0462	0.0531	0.89
X	20353947	G	A	locus	rs140012985	RPS6KA3	intergenic	0.945	10.0	0.0365	0.0406	0.0354	1.13
16	21709216	T	G	locus	rs2003622	OTOA	intron	0.991	1.55	0.102	0.0978	0.103	0.92
16	72889551	G	A	locus	rs78069158	ZFHX3	intron	0.991	0.793	0.170	0.163	0.171	0.94
1	213927843	A	T	locus	rs2011374	PROX1	intron	0.994	1.15	0.577	0.587	0.575	1.05
18	45690590	T	C	locus		SLC14A2	intron	0.678	inf	0.000385	0.000628	0.000323	4.63

Item	Number SSRI	Number SSRI (with VNR)	with VNR all antidepressants and Lithium (N06A and N05AN01)

Number of data points	84,477,978		
Number of drug purchases points	42,776,151		
Number of SSRI purchases	897,901	291,620	1,849,497 (1,849,422 with VNR) only 75 missing
Number of purchases (unique)	56,519	33,022	81,096 (81,096 with VNR)
EVENT_AGE Mean (SD)	44.31 ± 16.28	45.62	50.90
Freq purchase median	7 (Q1: 2, Q3: 21, Max: 288)	4 (Q1: 2, Q3: 10, Max: 196)	
Freq purchase mean	15.89	8.23	
summary(data.purchases.d rug.bursts\$daysPassed.diff)	#Min. 1st Qu. Median Mean 3rd Qu. Max. #0.0 40.0 89.0 15 0.6 111.0 8516.0	#Min. 1st Qu. Median Mean 3rd Qu. Max. # 0.0 35.0 84.0 193.2 112.0 8724.0	Min. 1st Qu. Median Mean 3rd Qu. Max. 0.0 28.0 63.0 125.4 103.0 8582.0

1. Ideas from biobank data group meeting?

1. Are there comorbidities, other symptoms, phenotype correlation

2. is there a schema for VNR? -- yes, and cleaned
 1. Getting updated version from Tuomo
 3. What is the recommended dosage? DDD
https://www.who.int/medicines/regulation/medicines-safety/toolkit_ddd/en/
-

Number of purchases for the different drugs?

1. All VNR codes are 6 digits (https://wiki.vnr.fi/?page_id=36), except a few which are 7 because
 1. *"There are some common drugs but the formulation is weird. For example ursodeoxycholic acid is commonly used in powder and suspended in water before drinking it but never seen it being prescribed as ready-made suspension. The suspension is erityislupavalmiste, not the powder."* --
https://www.kela.fi/documents/10180/3612716/Korvattavat%20potilasko_htaiset%20erityislupavalmisteet%20%28pdf%29/02594f8d-ded2-4d71-92cc-3f3e03402394
2. 727 VNRO (9 are missing, sent to Ari): 142 unique drugs (regardless of package sizes and dosages)
 1. mg * package size / number of days = dose per day (follows DDD and not as crazy as before)

 R5.SSRI.vnr.SSRI.summary.pdf

14 kB

1. Abnormally high dosages could be potentially from end of year purchase to use the purchase quota or fast metabolizers (Jaana)
1. Drug purchased regardless of size and package
 1. Different companies likely for financial reasons? Look at Lexapro 4 and 8 of two different companies
 2. Can I merge some of these since they are the same drug but from different brands?
 1. Sensible for the questions that we are asking, different formulations, different uptake? How quickly or slowly the uptake? Not worry about that now.
 2. Jaana suggests just looking at dose and nothing else

3. This policy has created competition between generic drugs as well and resulted in lowering of the medication costs. See e.g. http://www.apteekkariliitto.fi/media/3-apteekkariliitto.fi/englanti/annual-reviews/afp_annualreview_2016.pdf
3. What are the dates of capture? When is the first date and when is the last date?
 - Usage vs non-usage period: typical time, defining SD, and then that would tell us general usage -- pre2dup modeling (Ari)

MDD frequency? Initial diagnosis and outcomes?

1. Severe outcome and what drugs are most bought in different categories
2. Relevant parameters that can be accessed.
3. If there are improvements or decrease of things!
4. Depression - controls : cases = 13053 : 6514

so what i did in epilepsy was just a quick check on how many individuals with a given epilepsy drug prescribed might actually have any epilepsy diagnosis that way i discovered the drugs with the largest discrepancies
Quick table

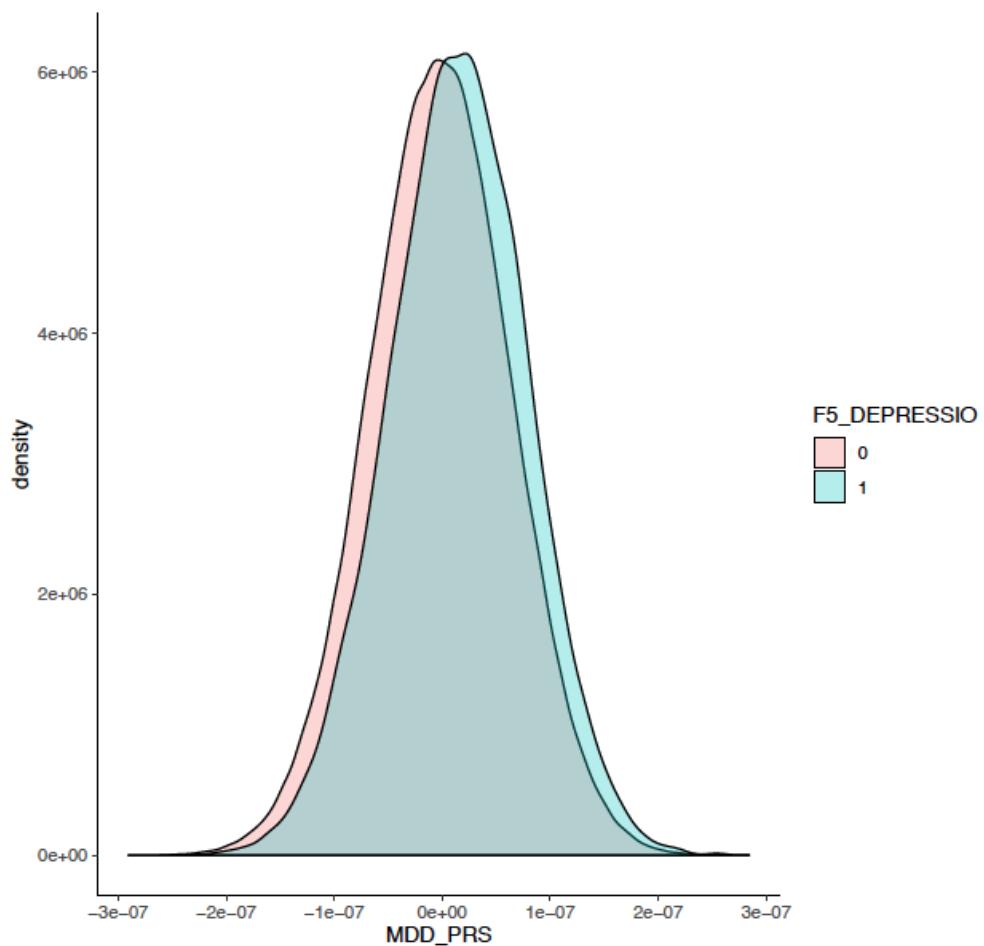
1. Psychoth - controls : cases = 17928 : 1342
2. From Paavo: SSRIs are very widely used for numerous other psychiatric and non-psychiatric indications in addition to MD:
<https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2010.1024#EX1>
3. MDD PRS: gs://r5_data/PRS/scores/MDD2018_ex23andMe.19fields.sscore

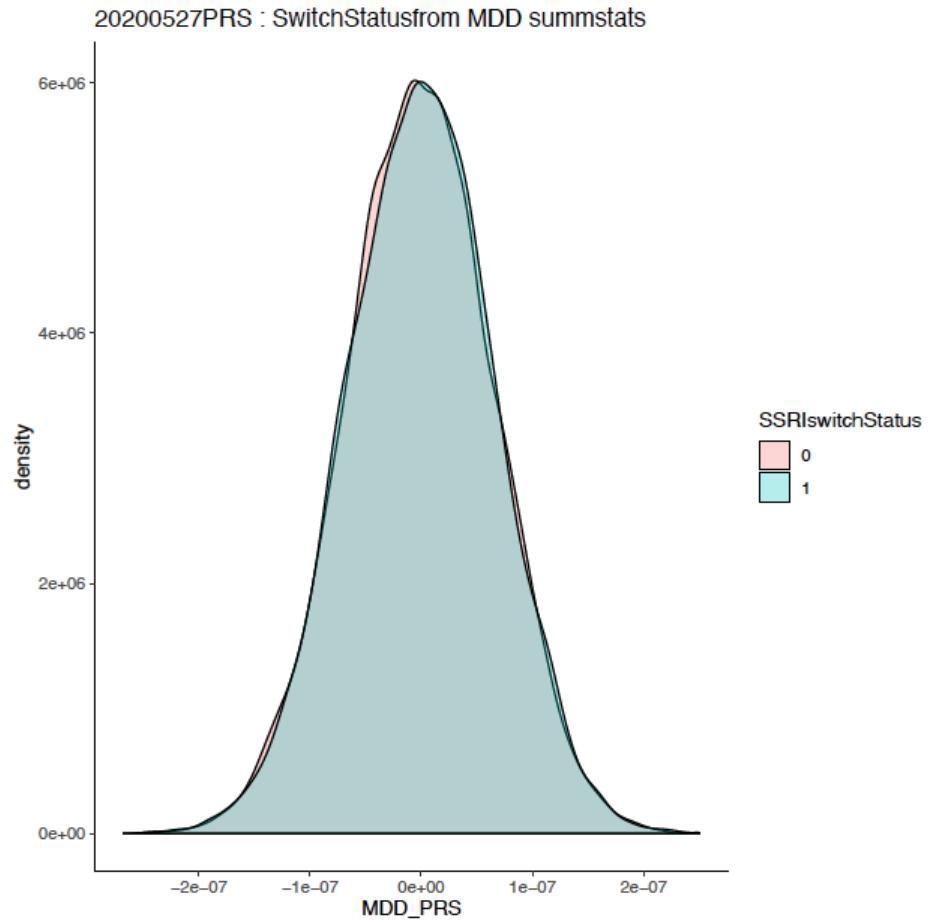
SummStats from Naomi Wray's 2018 paper but have asked Pietero to update it to David Howard's updated paper in 2019

Snapshots from /Daly_Lab/drugResponseAnalysis/PRS/20200527_PRS.pdf

Raymond and Ben : Something new, I don't think that there is any new space here?
MR Preso paper -- Ben and Chia Yen
Looking at the correlations
Fitting data is overfitted

20200527PRS : F5_DEPRESSIONfrom MDD summstats





PRS corrs

https://github.com/FINNGEN/PRS_analysis/tree/master/scripts <- how to run

https://github.com/FINNGEN/PRS_analysis

```
gs://r5_data/PRS/corr/finngen_R5_prs_pheno_corr.tsv
```

```
gsutil cat gs://r5_data/PRS/corr/finngen_R5_prs_pheno_corr.tsv
| awk '$1 ~ "F5" || $1~"ANTIDEPRESSANTS" {print $0}' >
finngen_R5_prs_pheno_corr.F5only.tsv
```

Genetic Correlations

rg list

```
gs://r5_data/ldsc/ldsc_r5_1000_cases_prs_traits.txt
```

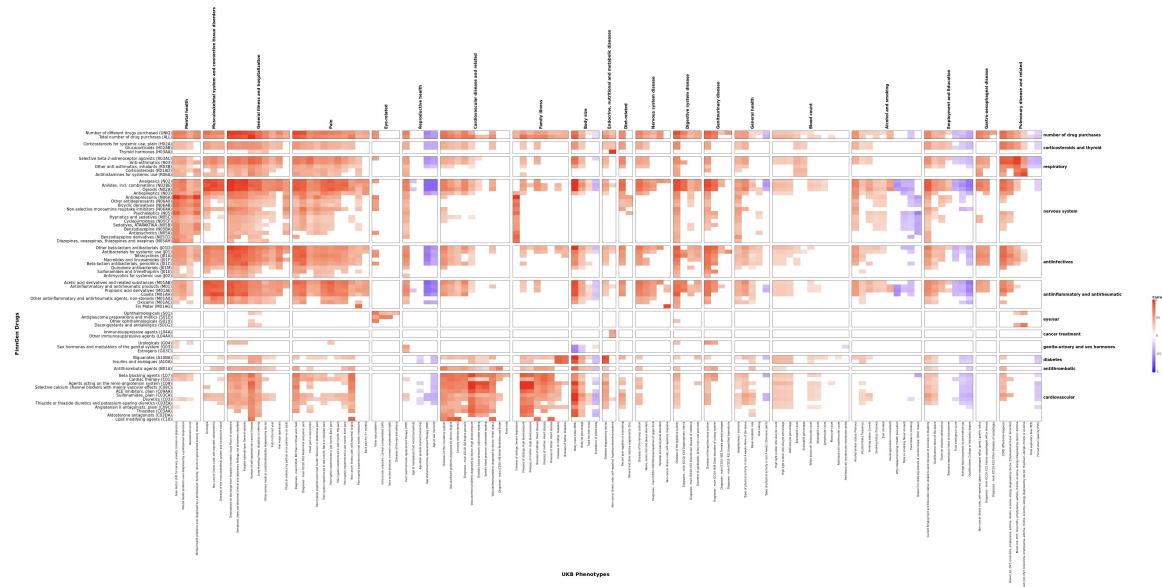
<https://github.com/FINNGEN/LDSC/tree/master/wdl> <- how to run

```
gsutil cat gs://r5_data/ldsc/ldsc_r5_1000_cases_prs_traits.txt  
| awk '($1 ~ "F5" || $1 ~ "ANTIDEPRESS") && $2 ~ "MDD2018"  
{print $0}' > ldsc_r5_1000_cases_prs_traits.F5_MDD_2018.txt  
gs://r4_data/ldsc/FG.ldsc_rg.txt.bgz | R4 phenotype to phenotype Rg
```

rgList from Kristin on UKBB to drugs

<https://github.com/FINNGEN/LDSC>

The data from below available here: fg_ukb_rg_matrix_pthres.xlsx



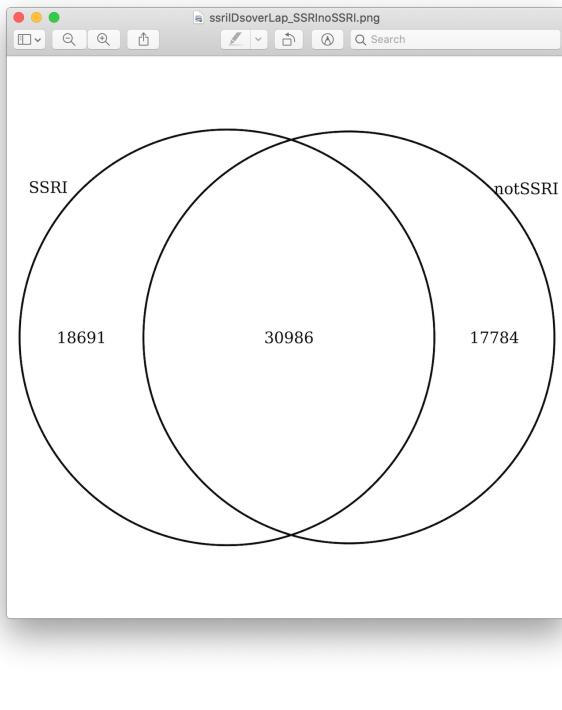
SSRI specific Rg -- F5_SSRI_phenotypes.txt

Switching drugs

1. What event makes someone switch drugs?
 1. Previously on an antidepressant, looking at an event and see what did they switch?
2. UK Guidelines are similar to

Finland <https://www.nice.org.uk/guidance/cg90/chapter/1-Guidance#sequencing-treatments-after-initial-inadequate-response>

3. Create a binary of switch vs no switch



4. What is the interesting event there? How are you measuring the switching of the drug? What is the common drug switches?
5. CYP metabolizing enzymes and the variants? Or fast/slow metabolizers
6. Hospital discharge / visits --
 1. bullema and OCD dosage is higher than normal, or anxiety disorders is lower
 2. psychotic depression or suicidal ideation

Run on May 20th, 2020

20200519: da3153fd-e066-47a6-88e2-9d967876aa42

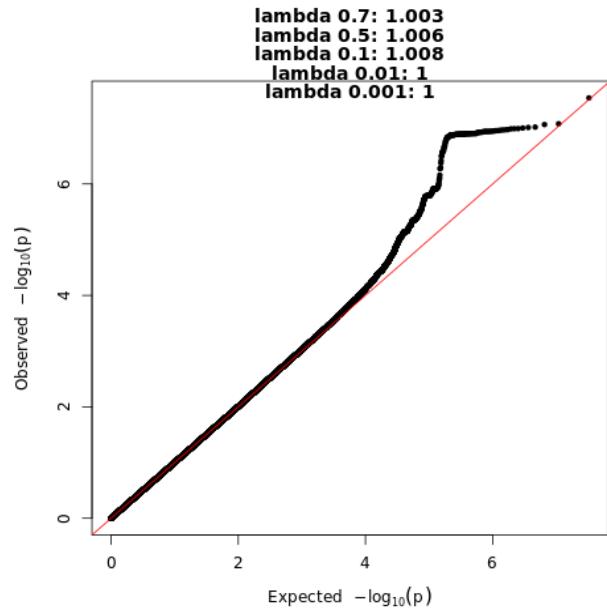
Results are here:

```
gsutil cp gs://fg-cromwell/saige/da3153fd-e066-47a6-88e2-  
9d967876aa42/call-test_combine/shard-  
102/sub.test_combine/9f558179-4c83-4ba4-b3be-  
82f2c72d6e69/call-combine/attempts-2/
```

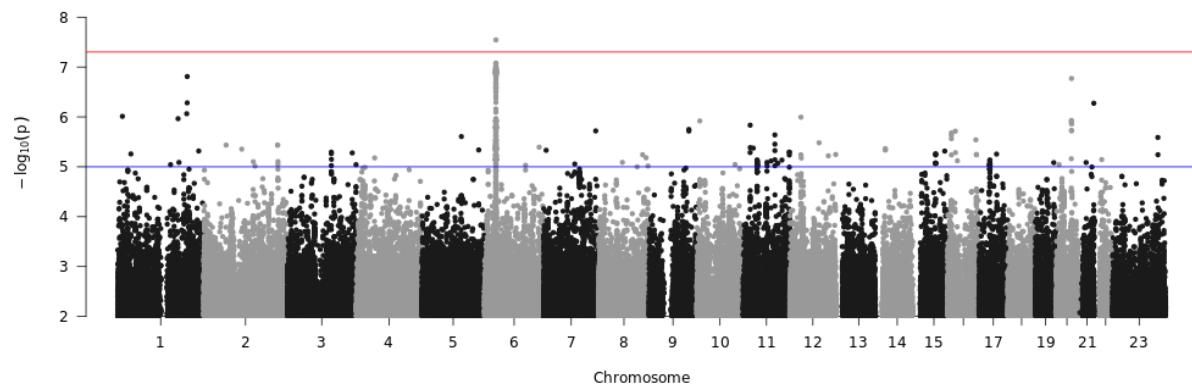
And now here:

vm: /home/kumar/workDir/SAIGE_analysis/20200520_SSRIswitch

QQ plot



loglog Manhattan plot



```
zcat SSRISwitchStatus.pheweb.gz | awk 'NR==1 || $5 < 1e-07  
{print $0}' | less -S
```


Autoreporting results:

locus_id	chr	start	end	enrichment	lead_pv	lead_beta	lead_maf	lead_ma_f_cases	lead_ma_f_controls	most_severe_gene	most_severe_c
chr6_32_706_117_C_T	6	312 081 35	329 137 25	1.38	2.86 E-08	-0.0 989	0.52 1	0.50 7	0.53	MT CO3 P1 (pseudo gene)	drownstream gene variant

HLA panel? Thought from Lea

strict (credible sets)

found associations relaxed (r^2, ld)

systemic scleroderma|0.577;

rheumatoid arthritis|0.577;

primary biliary cirrhosis|0.516;

biliary liver cirrhosis|0.516;

Sarcoidosis|0.451;

autism spectrum disorder|0.421;

schizophrenia|0.421;

multiple sclerosis|0.349;

oligoclonal band measurement|0.349;

mosquito bite reaction itch intensity measurement|0.31;

type I diabetes mellitus|0.288;

C-peptide measurement|0.288;

complement C4 measurement|0.284;

Oral ulcer|0.276;

hepatocellular carcinoma|0.269;

nodular sclerosis Hodgkin lymphoma|0.262;

heel bone mineral density|0.25;

body height|0.246;
blood protein measurement|0.243;
parental longevity|0.213;
low density lipoprotein cholesterol measurement|0.195;
C-reactive protein measurement|0.195;
Henoch-Schoenlein purpura|0.194;
IGA glomerulonephritis|0.194;
high altitude adaptation|0.194;
urinary albumin to creatinine ratio|0.164;
cervical carcinoma|0.162;
response to thiopurine|0.15;
thiopurine immunosuppressant-induced pancreatitis|0.15;
inflammatory bowel disease|0.15;
Seropositive Rheumatois Arthritis|0.139;
neoplasm of mature B-cells|0.139;
asthma|0.137;pemphigus vulgaris|0.126;anti-neutrophil antibody associated vasculitis|0.126;temporal arteritis|0.125;Parkinson's disease|0.125;Parkinson's disease|0.125;type II diabetes mellitus|0.122;hepatitis C induced liver cirrhosis|0.112;ulcerative colitis|0.11;Epstein Barr virus nuclear antigen 1 IgG measurement|0.11;Drugs used in diabetes use measurement|0.106;latent autoimmune diabetes in adults|0.106;chronic lymphocytic leukemia|0.106;Inflammatory Bowel Disease|0.106;granulocyte count|0.105;neutrophil count|0.105;basophil count|0.105;myeloid white cell count|0.105;Sjogren syndrome|0.105;eosinophil count|0.105;wheat allergic reaction|0.105;lymphoma|0.0994;allergy|0.0976;lymphocyte count|0.0962;mathematical ability|0.0916;Eczema|0.0885;allergic rhinitis|0.0885;systemic lupus erythematosus|0.0858;body fat percentage|0.084;blood metabolite measurement|0.084;urate measurement|0.0798;susceptibility to Mycobacterium tuberculosis infection measurement|0.0596;unipolar depression|0.0596;Churg-Strauss syndrome|0.0587;wet macular degeneration|0.0553;age-related macular degeneration|0.0553;atrophic macular degeneration|0.0553;complement factor B measurement|0.0553

found_associations_relaxed: NA

credible_set_variants: NA

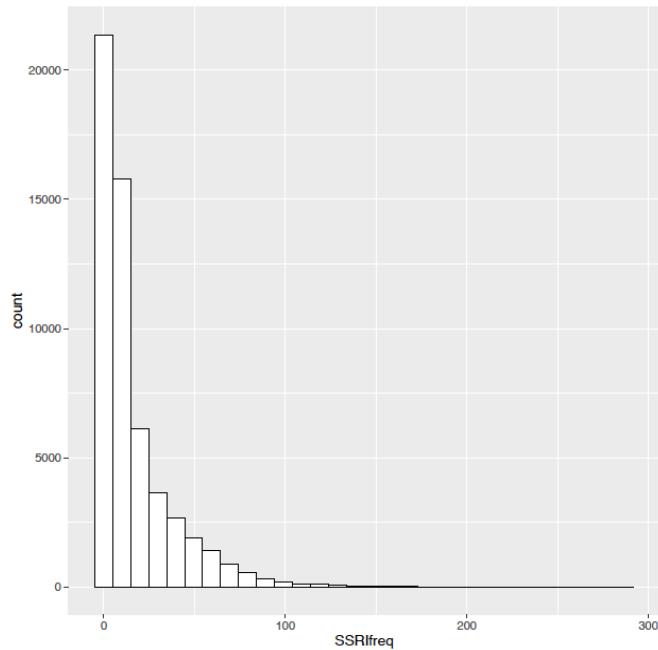
functional_variants_strict:

chr6_31628105_C_A|missense_variant|0.084;
chr6_31633567_T_C|missense_variant|0.0825;
chr6_31659746_C_T|missense_variant|0.0813;
chr6_31946403_G_A|missense_variant|0.0553;
chr6_31960529_A_G|missense_variant|0.0553;
chr6_32222613_T_G|missense_variant|0.0596;
chr6_32369909_A_G|missense_variant|0.0859;
chr6_32661978_G_A|missense_variant|0.113;
chr6_32666522_A_ACC|pLoF|0.116;
chr6_32666523_GTA_G|pLoF|0.115
functional_variants_relaxed: NA
specific_efo_trait_associations_strict: NA
specific_efo_trait_associations_relaxed: NA

-
1. SSRI 1 vs SSRI 2
 1. more constant: anti hypertensive, thyroid replacements
 2. less constant: painkillers

Purchase bursts

1. number of purchases --
 1. 56519 purchasers
 2. Median: 7 (Q1: 2, Q3: 21, Max: 288)
2. From Juha: <35 years have been filtered out as the drug registry starts from 1994. *_IRN is inverse rank normalized*



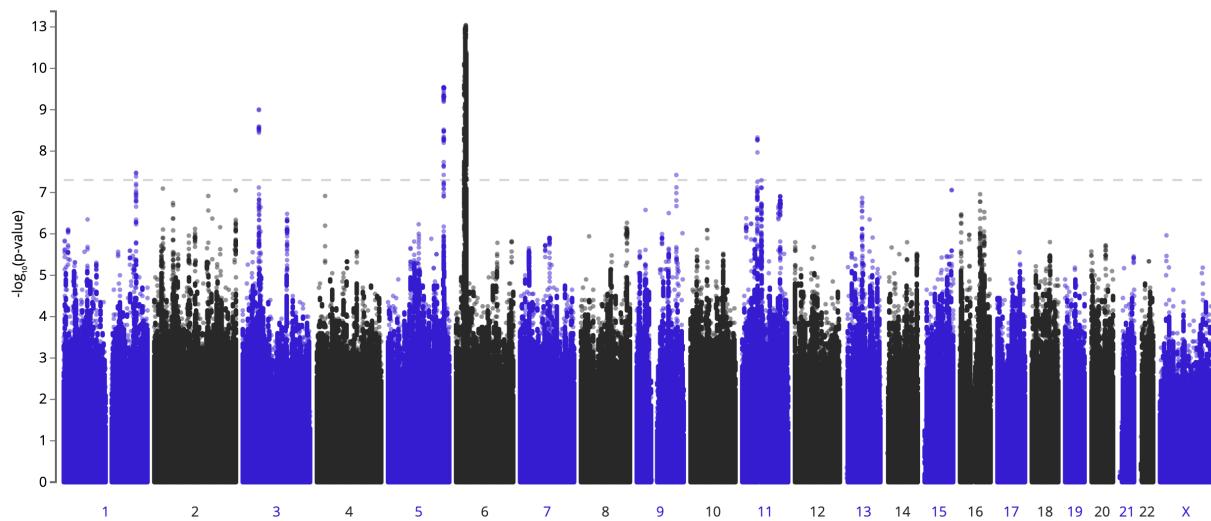
From Juha

Bicyclic derivatives

ATC

51048 samples

Phenotype not found in UKBB results



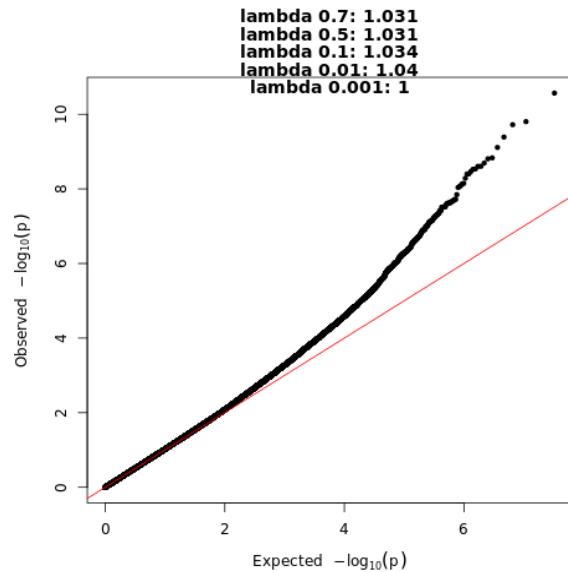
Lead variants

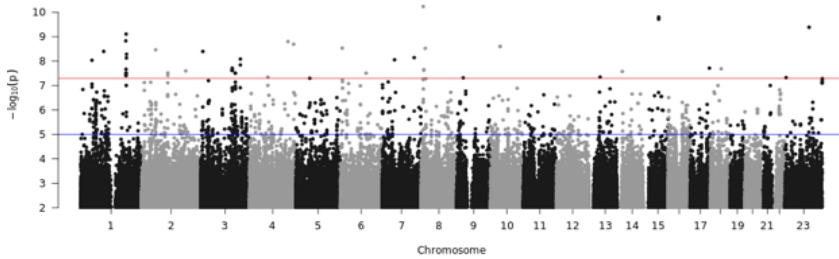
chr	pos	ref	alt	locus	rsid	nearest gene	consequence	INFO FG	FIN enrichment	af	OR	p-value
6	30070870	T	C	locus	rs9261290	RNF39	3 prime UTR	1.00	0.657	0.0665	0.97	1.9e-13
6	28872131	G	A	locus	rs3118365	TRIM27	intergenic	0.997	0.600	0.0522	0.96	3.1e-13
6	29374998	G	A	locus	rs3749971	OR12D3,OR5V1	missense	1.00	0.625	0.0556	0.96	4.7e-13
6	28355925	G	A	locus	rs13198809	ZKSCAN3,ZSCA...	intron	0.999	0.602	0.0507	0.96	1.2e-12
5	165056066	G	T	locus	rs4339373	MAT2B	intron	0.996	1.12	0.592	1.01	2.8e-10
3	48972137	C	G	locus	rs73082337	ARIH2	intron	0.997	1.31	0.162	0.98	9.9e-10
11	46035522	T	C	locus	rs12364441	PHF21A	intron	0.998	1.17	0.195	1.02	4.7e-9
6	25873518	C	A	locus	rs9467626	SLC17A3	intron	1.00	0.577	0.0488	0.97	1.9e-8
1	213928283	A	G	locus	rs1019969	PROX1	intron	0.994	0.856	0.429	0.99	3.3e-8
9	117728182	A	G	locus	rs183126703	TLR4	intergenic	0.964	2.06	0.0107	0.94	3.8e-8

20200522: 7cc22c2b-5e14-4d20-ab7f-bb1d4c914a74

cw metadata 7cc22c2b-5e14-4d20-ab7f-bb1d4c914a74 --summary

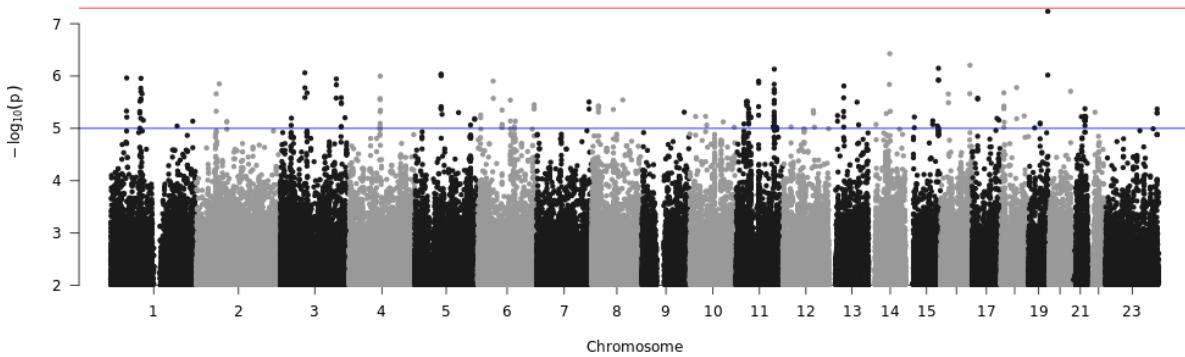
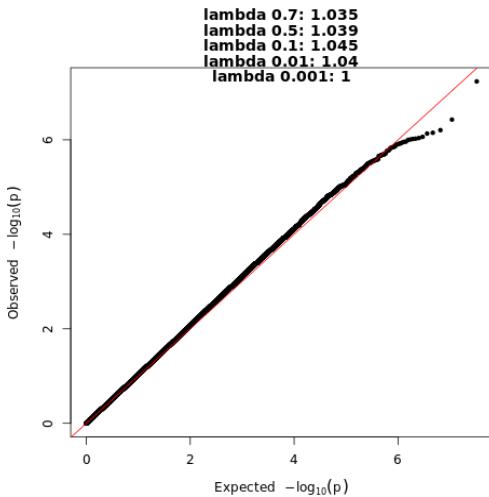
When NA=0.0 and IRN





Job ID: 61403335-108f-471a-98c1-3540b1c24cc1

When I NA-ed out the 0 purchases.



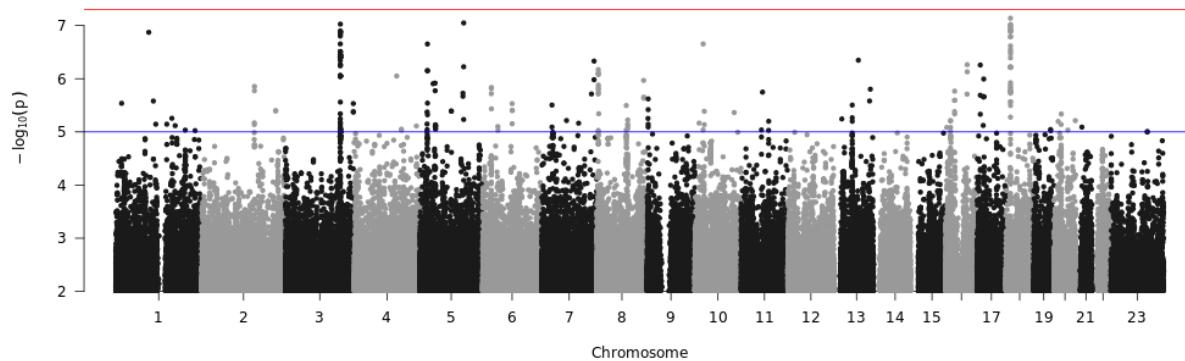
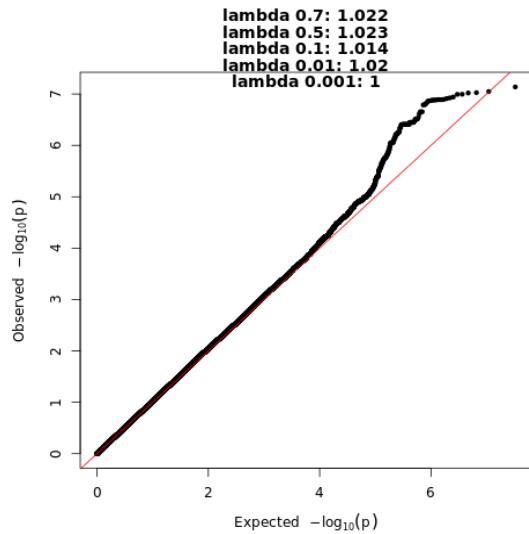
#chrom	pos	ref	alt	pval	beta	sebeta	maf
3	70450827	G	A	8.676e-07	-0.5016	0.1020	9.170e-04
5	76200273	A	G	9.852e-07	0.0321	0.0066	2.599e-01
5	76200661	A	G	9.464e-07	0.0322	0.0066	2.600e-01
5	76203246	C	G	9.180e-07	0.0322	0.0066	2.599e-01

11	110307680	T	C	7.379e-07	0.0827	0.0167	3.070e-02
14	61391205	AT	A	3.739e-07	0.9067	0.1784	3.509e-04
15	95124714	G	C	7.099e-07	0.1759	0.0355	6.689e-03
16	84256586	T	A	6.244e-07	0.2674	0.0537	2.918e-03
19	55418672	G	A	9.606e-07	-0.2485	0.0507	3.435e-03
19	55429846	G	A	5.805e-08	-0.3090	0.0570	2.651e-03
--SHISA7 (rs186648667)							

How about using dose per unit measurement of dose computed / DDD

 [SSRIunits.perDrug.preNorm.pdf](#) 10 kB

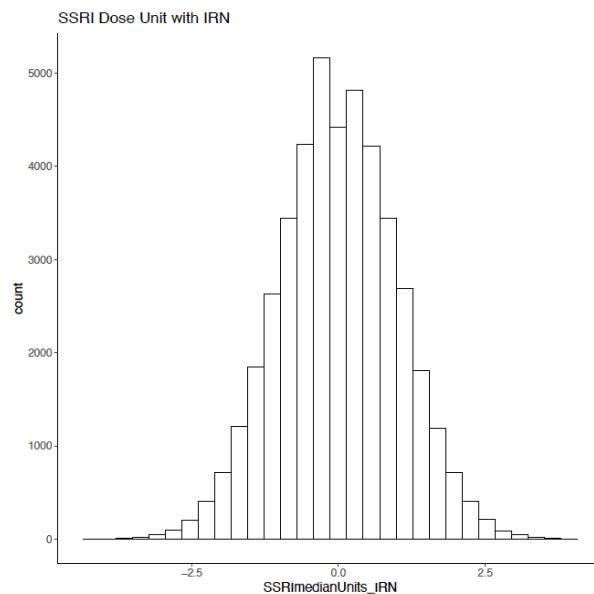
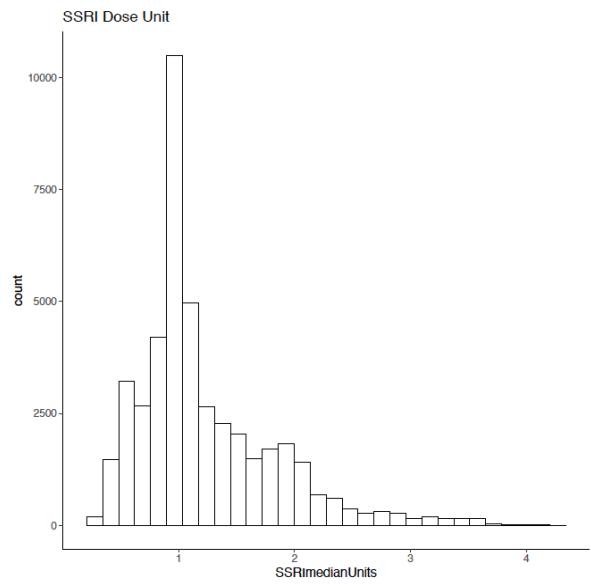
 [20200531_doseUnitDist.SSRI.pdf](#) 6 kB



#chrom	pos	ref	alt	pval	beta	sebeta	maf
3	158567491	T	C	9.413e-08	-0.0355	0.0067	5.581e-01
5	126731733	T	C	8.953e-08	-0.5842	0.1093	1.029e-03
18	13560569	A	G	9.561e-08	-0.0422	0.0079	7.766e-01
18	13562009	T	C	7.282e-08	-0.0426	0.0079	7.770e-01

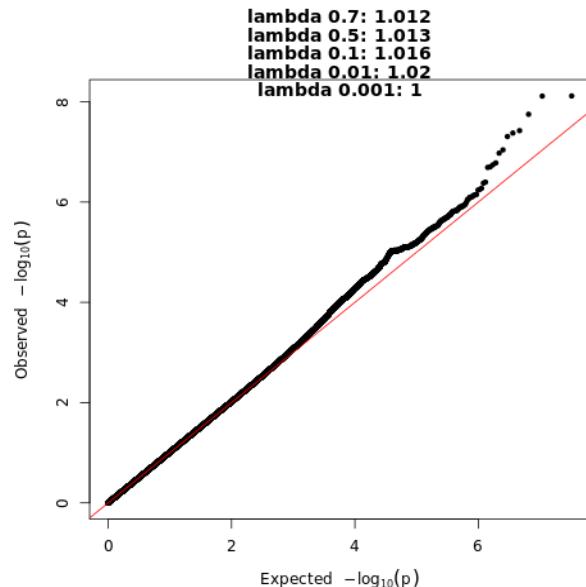
1. Can I bin these into a specific unit difference and then run the analysis separately?
2. Bin decrease as one, bin increase as one, and see medians and quantiles to see trajectories
 1. How about the 0?
 2. What are the extremes?

3. Yes, but in a second run, remove any increases that are close to no increase. And equally, remove any samples with only slight decreases. So only keep the extremes. Sometimes that works, sometimes not.



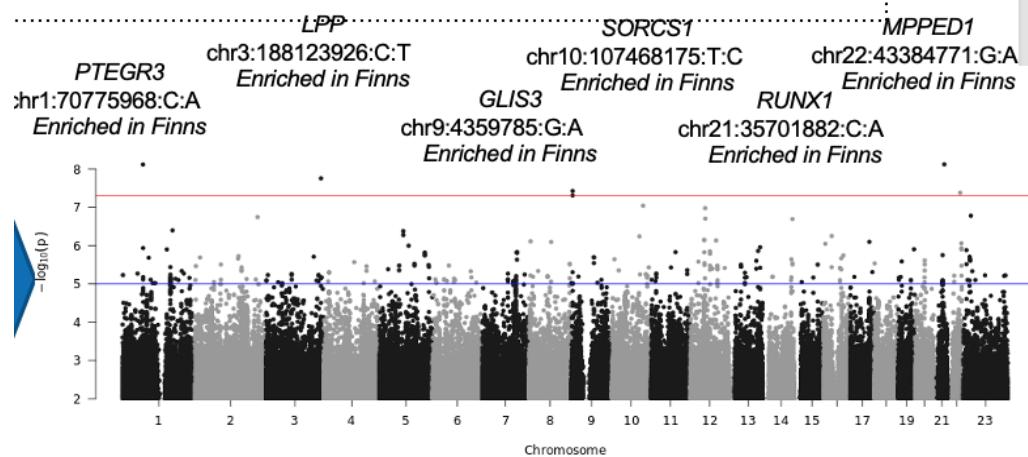
20200623_SSRIunits.perDrug.pre... 10 kB

Refer to the non-IRN files because the run was run with IRN = false on the wdl



o add text

N = 44,119
Median: 1.03 (0.25-4.26)



- Check FinnGen associations
- Check for gaps
- 0.01 as a filter, look fishy to him

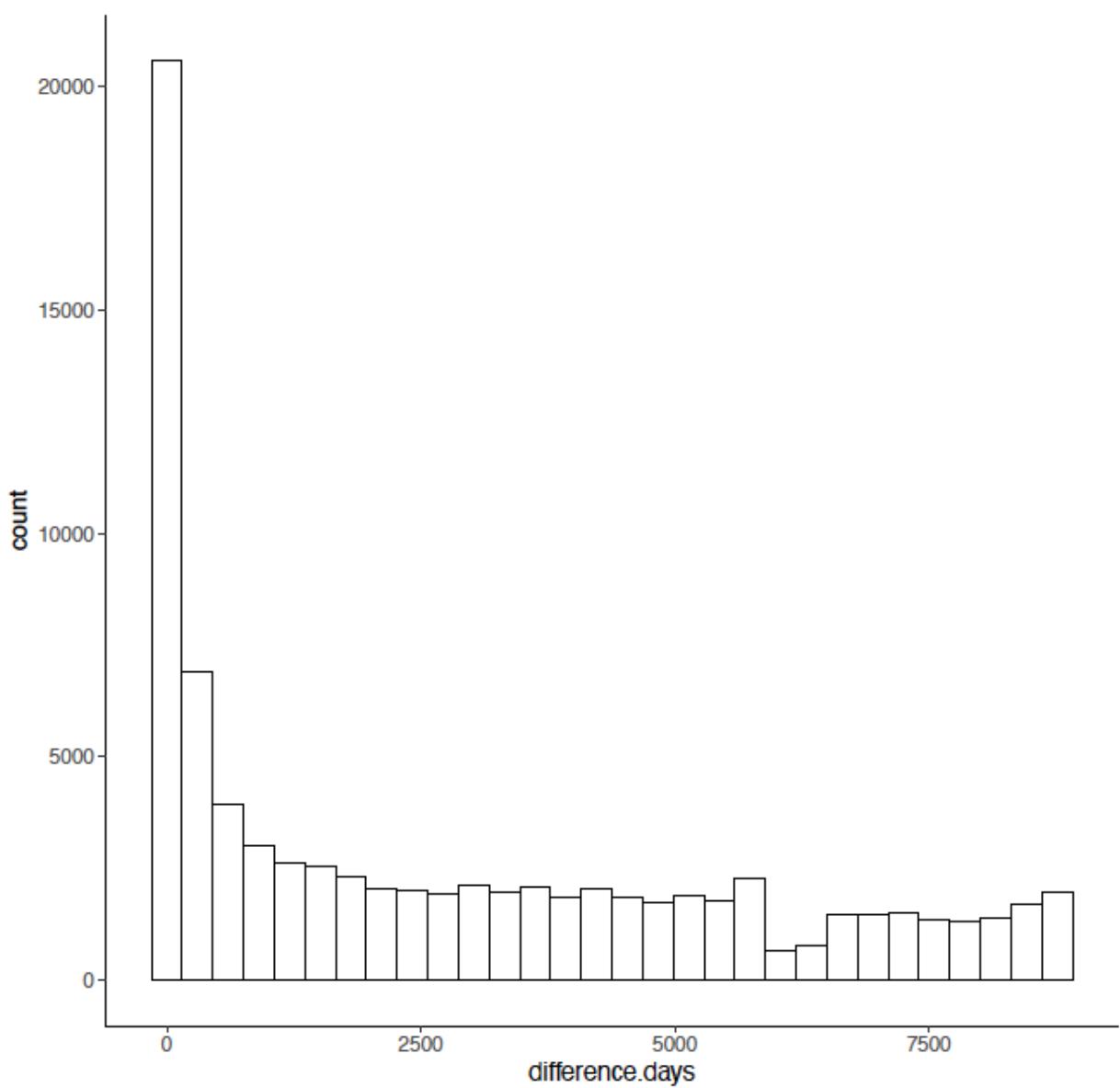
locus_id	chr	start	enrichmen t	lead_pval	lead_be ta	lead_seb eta	lead_ma f	most_severe_g ene	most_severe_c onsequence
chr1_707 75968_C _A	1	7032637 8	3.16	7.71E-09	1.23	0.212	0.00010 3	RP5-952N6.1	intron_variant

chr10_10										
7468175	10	1071990 56	0.801	9.11E-08	0.953	0.178	0.00014 4	NA		intergenic_variant
_T_C										
chr21_35										
701882_C_A	21	3478643 8	1.00E+06	7.64E-09	0.622	0.108	0.00042 3	RUNX1		intron_variant
_G_A										
chr22_4										
3384771	22	4333374 2	0.24	4.20E-08	0.558	0.102	0.00052 7	NA		intergenic_variant
_G_A										
chr3_188										
123926_C_T	3	1872138 07	1.00E+06	1.77E-08	0.345	0.0613	0.00111	RP11-430L16.1		intron_variant
_G										
chr9_43										
60353_A_G	9	4335111	3.12	3.76E-08	0.64	0.116	0.00031 3	NA		intergenic_variant

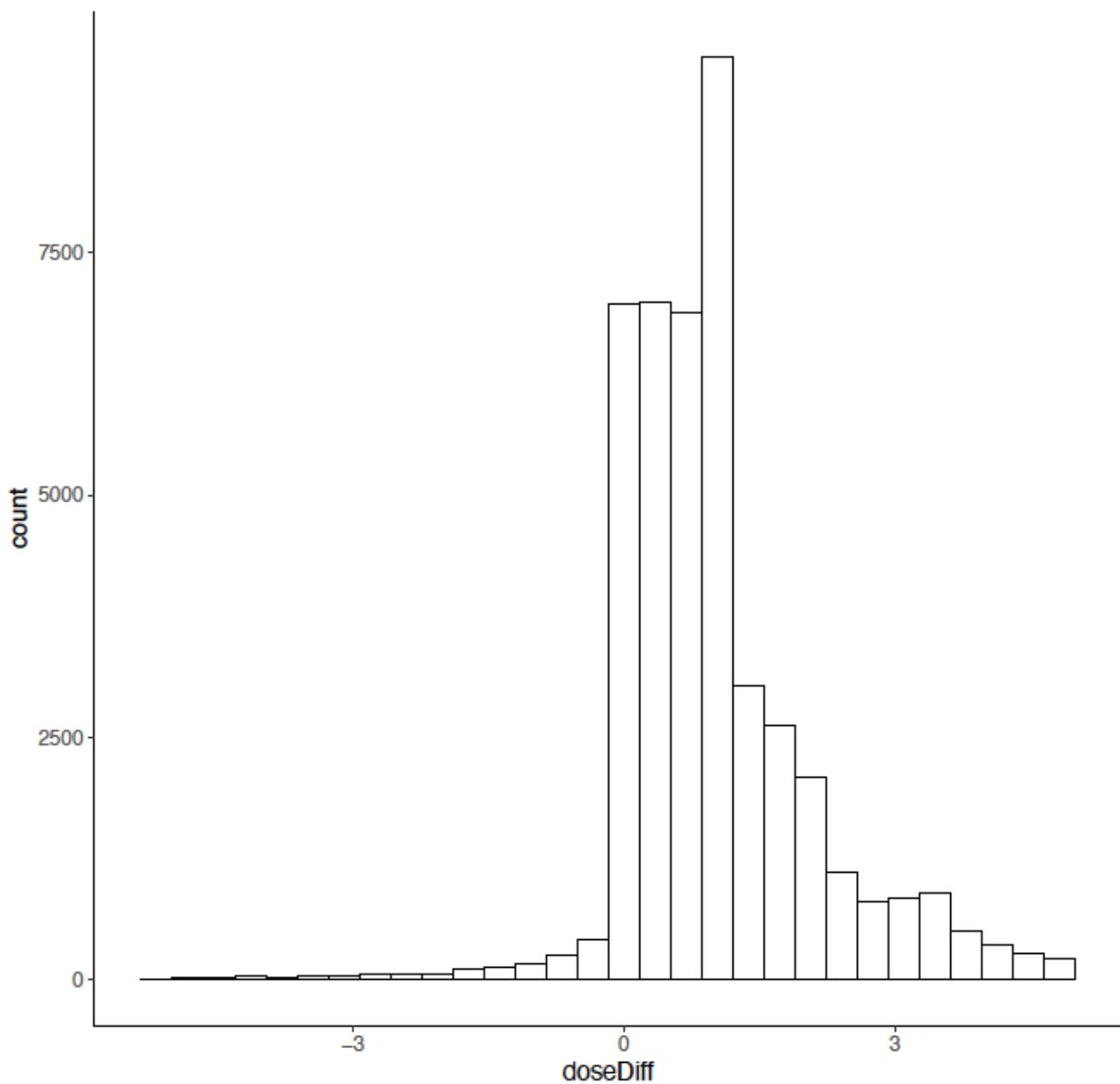
Check with Mark about these variants and LD : are these real? These are really rare -- from Aarno

Survival Analysis --

20200609 distPurchaseDates.surv



20200611 distPurchaseDoses.surv

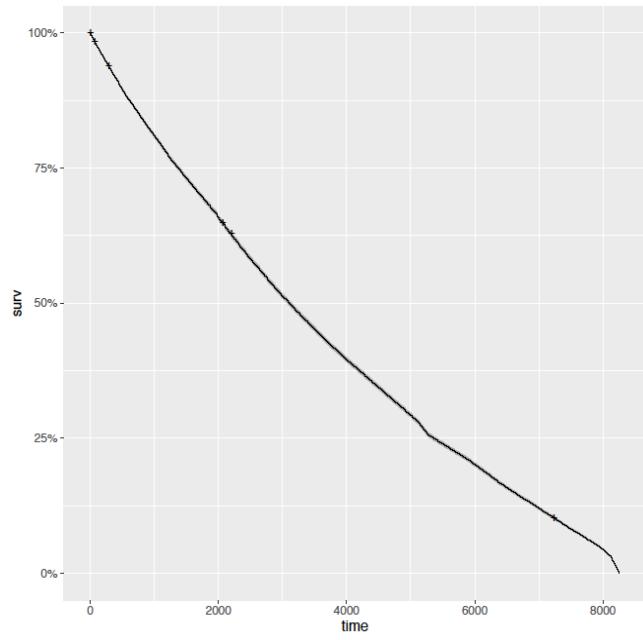


Notes from Mark:

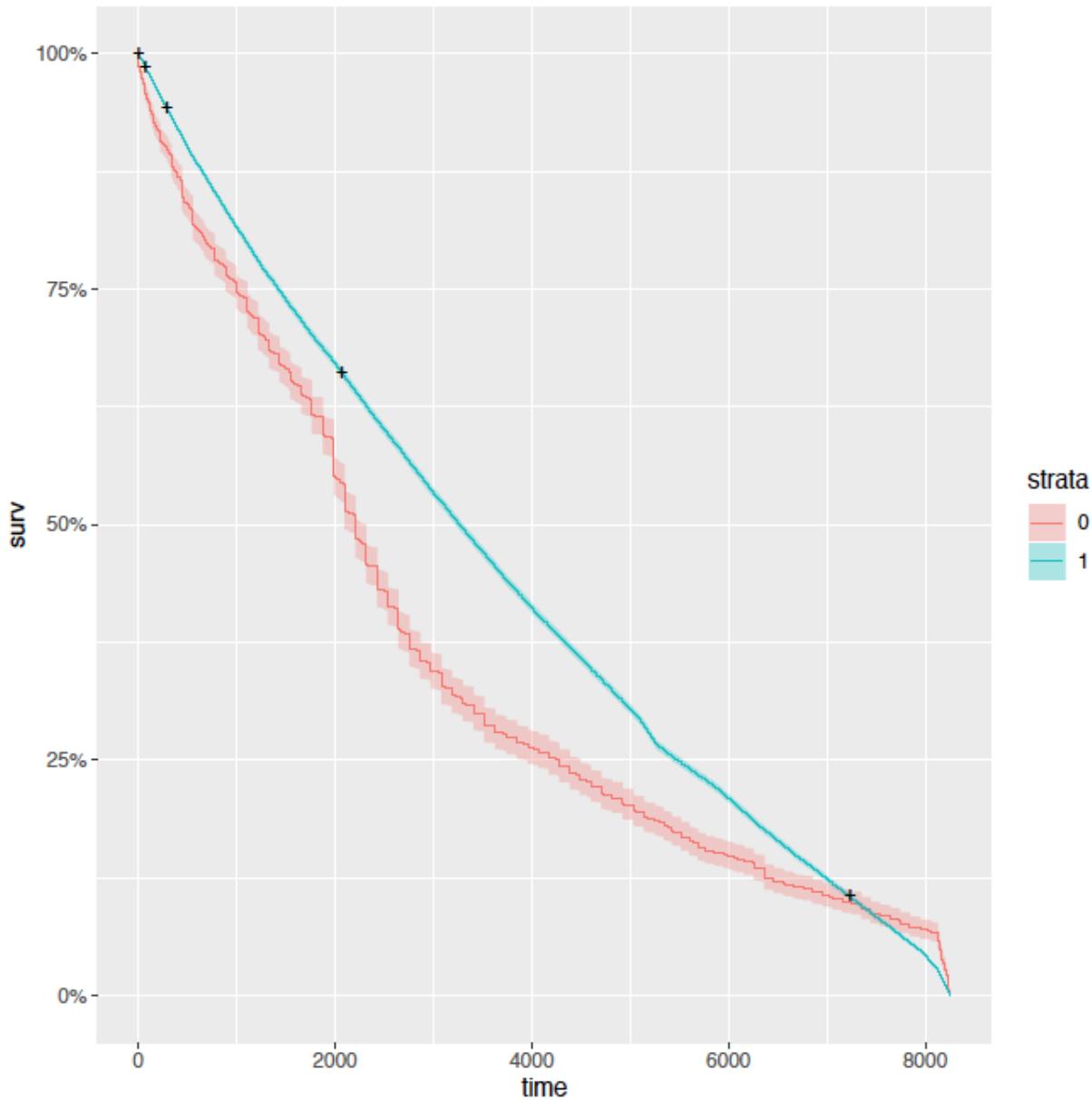
What is the tally of people at 0?

Cox ph of people going up without markers and only adjusting for sex, age, PCs

Almost completely random predictions because concordance is <0.55 for both



without cox ph

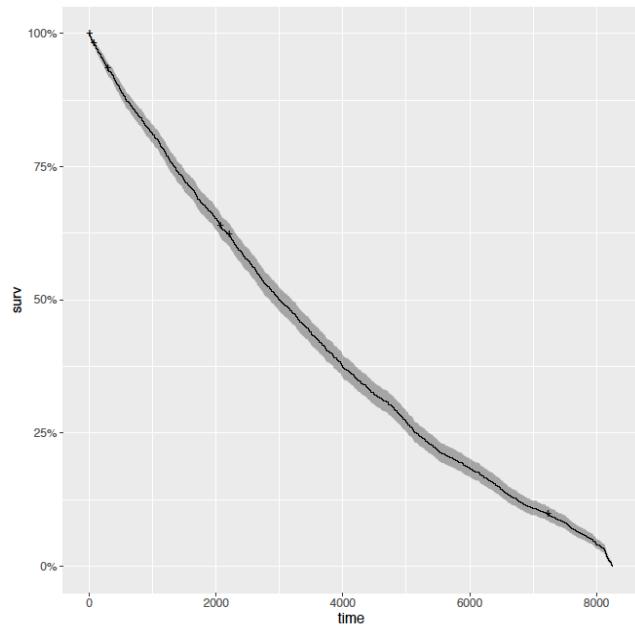


0 = censoring curve

1 = survival curve, after how long time that they increased med dose

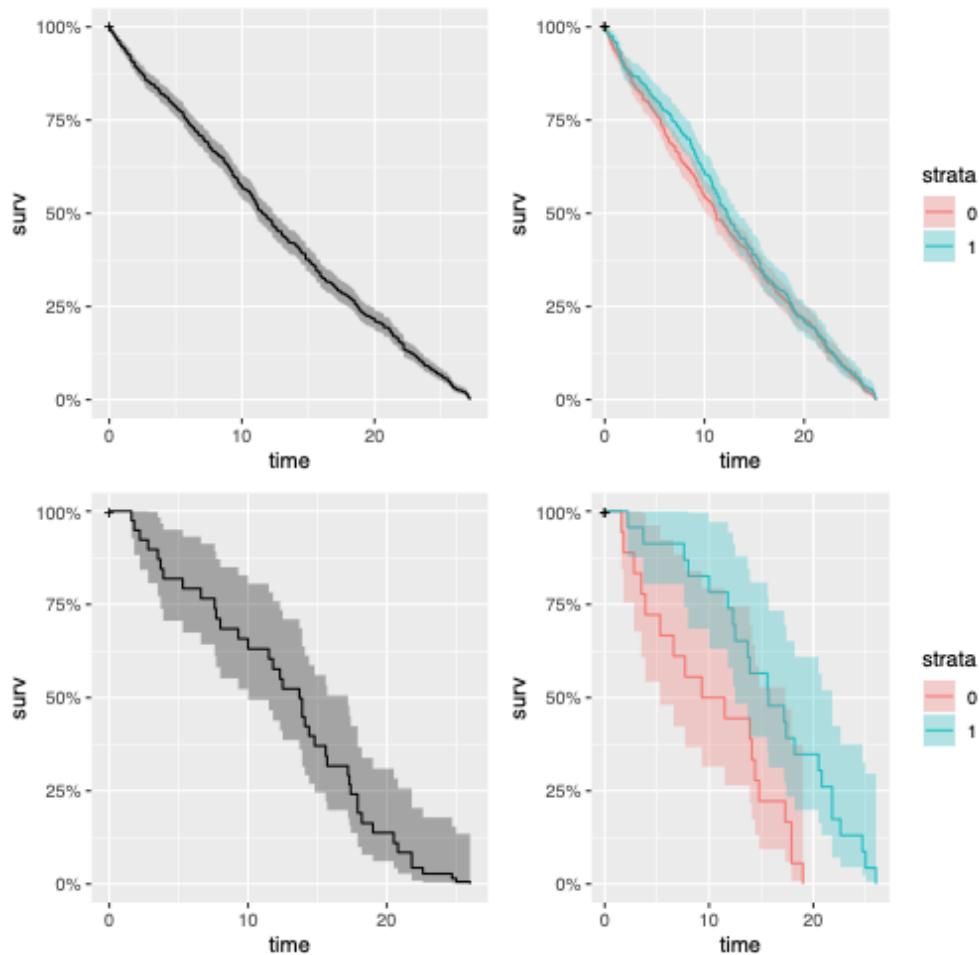
How many people go down?

Cox ph of people going down without markers and only adjusting for sex, age, PCs

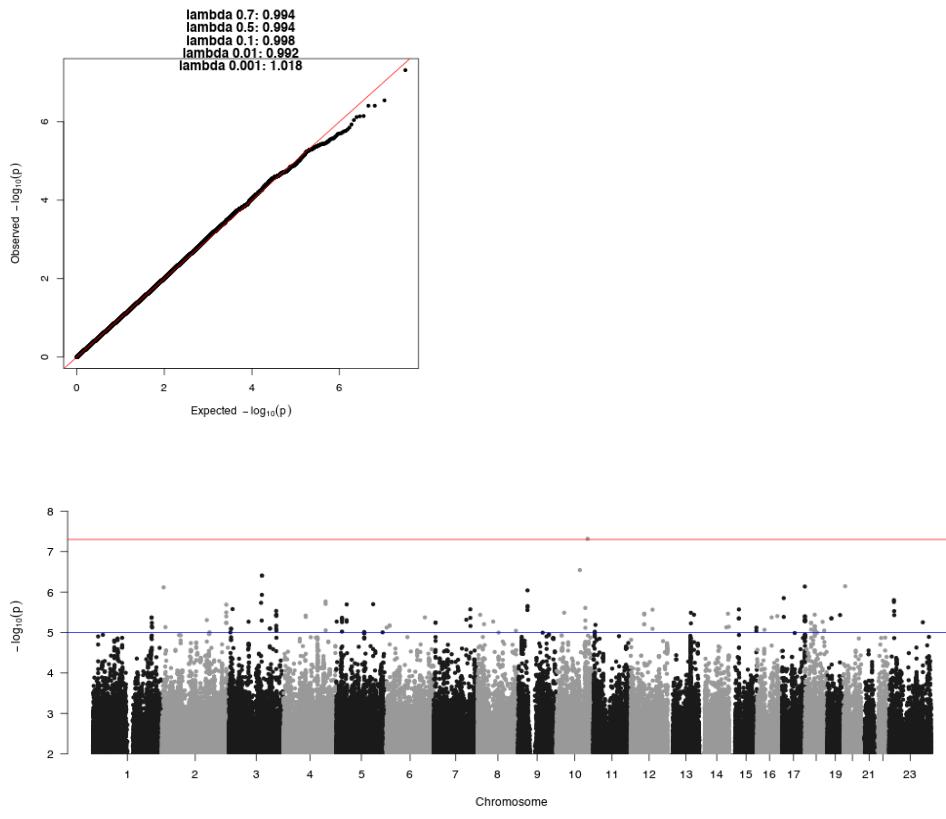


Binned this by months, and then took the first tenth of the data

20200611_lessThan27.28.months_coxPH.dosage

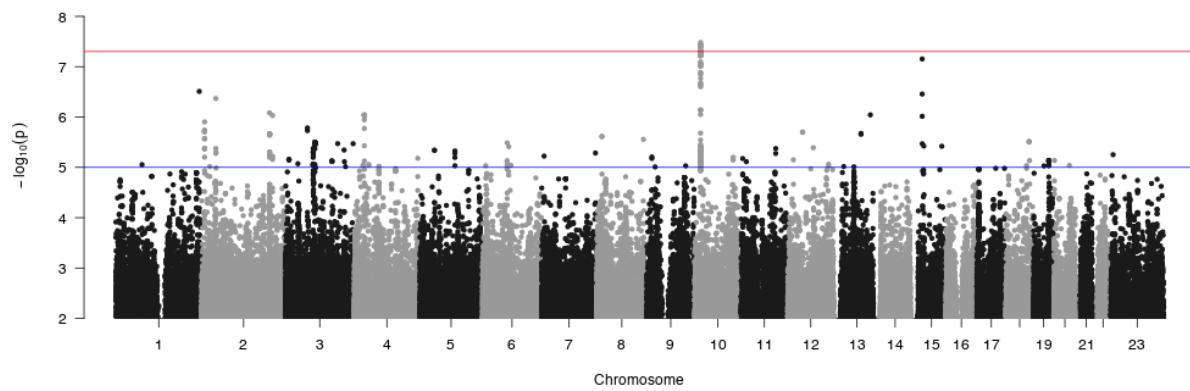
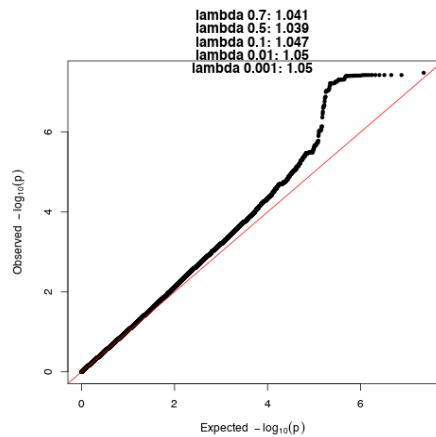


20200611_lessThan272.8.months_... 97 kB

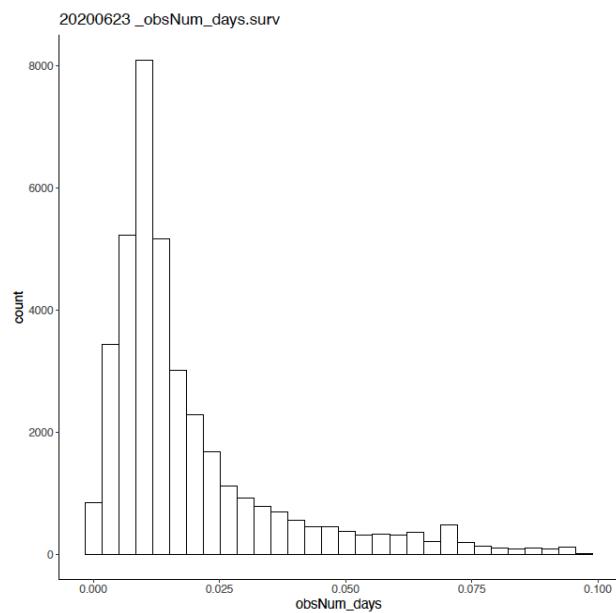


10 113421453 G A 4.82738102246969e-08 HABP2

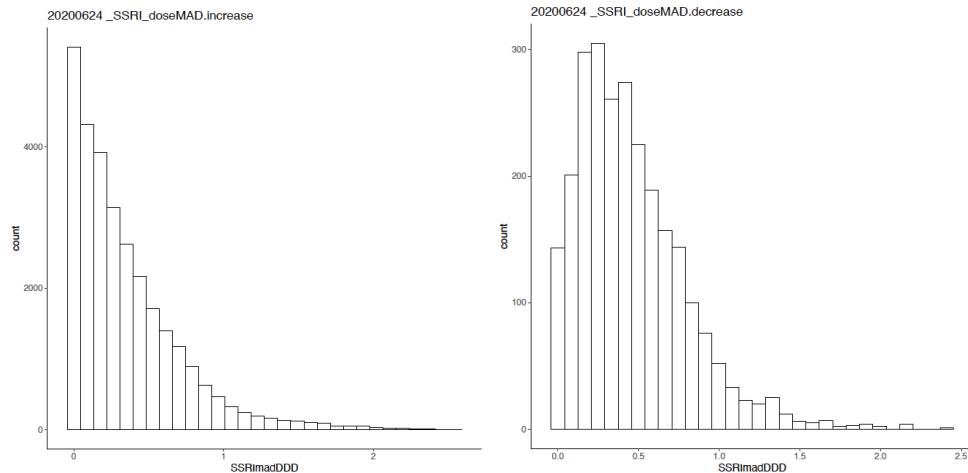
Decrease in dosage



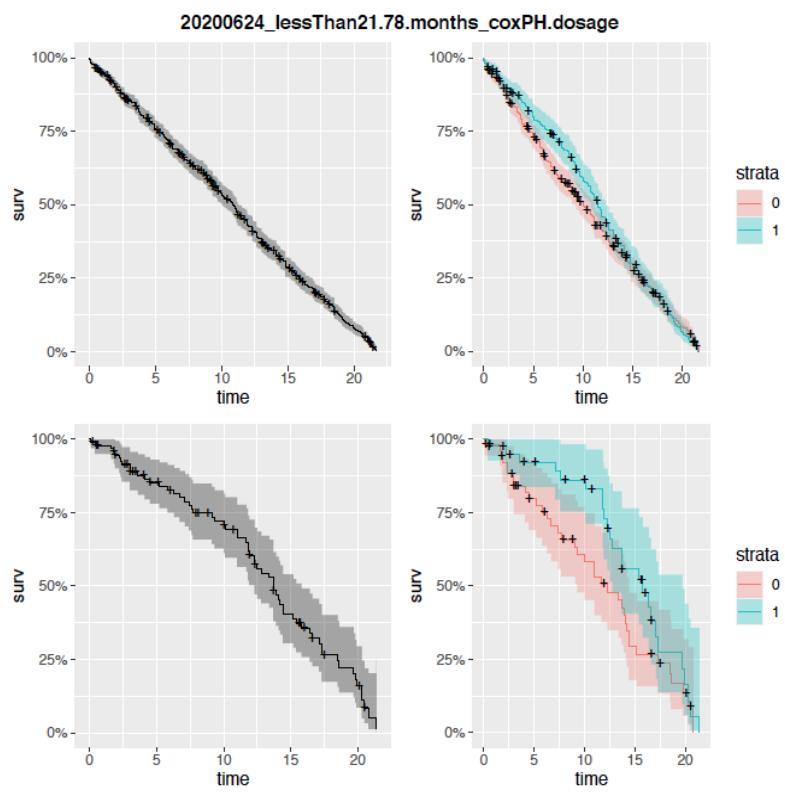
June 23rd, 2020 -- dosage increase / decrease per unit values



Using MAD to remove outliers MAD >2.0

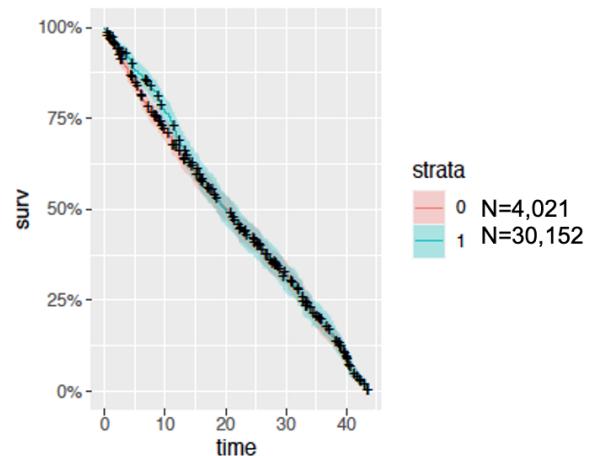


20200624_distPurchaseDates.surv... 8 kB



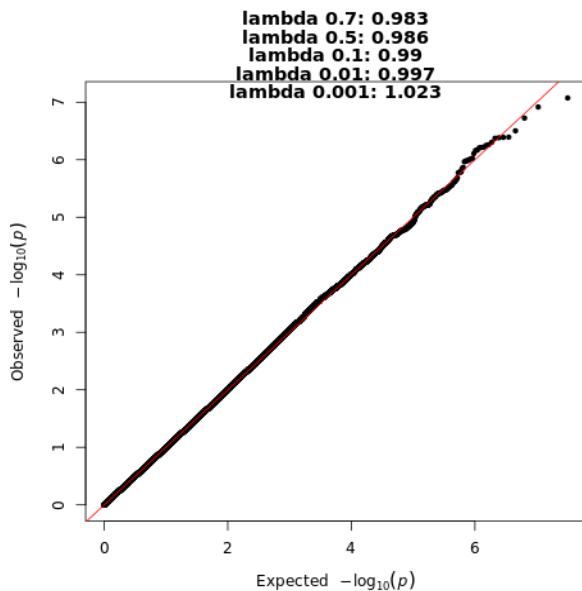
decrease: 0=1041; 1=2592

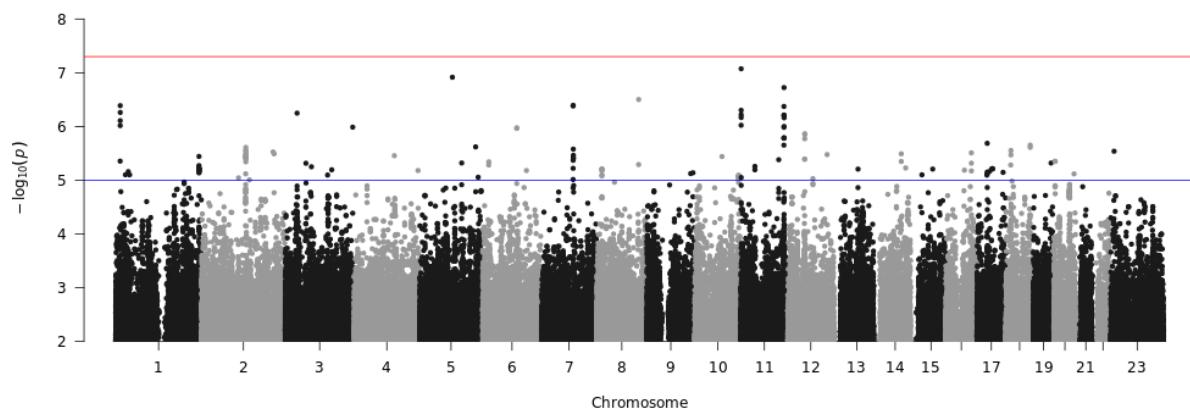
- Tracking dosage per unit of SSRI in months (median= 40.10)
 - Censored (0) = no change over time (± 0.5)
 - Events (1) = dosage increase
- Chr11:476394:G:A ($p < 8.43 \times 10^{-8}$)
 - [PTDSS2](#) ($AF_{FINN} = 0.012$)
 - [FinnGen phewas](#): Obstructed labour due to maternal pelvic abnormality ($p < 10^{-5}$)
 - High affinity for docosahexaenoic acid (DHA) : SSRI can be augmented with DHA



- SSRI start with low doses
- First dose - maximum dose, time to event

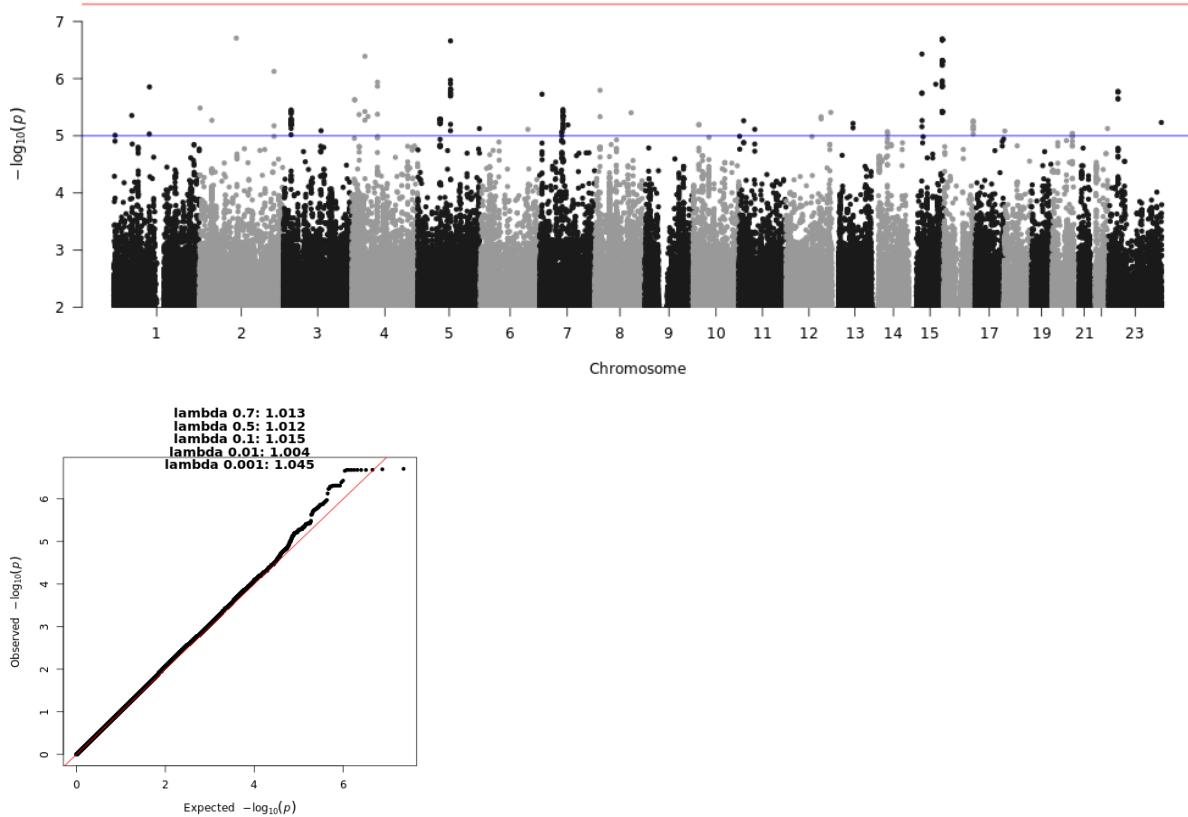
increase





locus_id	chr	start	end	enrichment	lead_pval	lead_beta	lead_sebet	lead_maf	lead_fca	lead_controls	lead_imputation	mosaic_segregation	most_severe	functional_annotation
chr1														chr11_61072
1_4														4_C_T missense_variation
763	11	298	1830	0.779	8.43E-08	0.22	0.041	0.011	0.01	0.01	0.99	PTD_SS2	intron_ss2	63
94_G_A		920	224			4	9	9	21	06	4			ant[0.4]

decrease

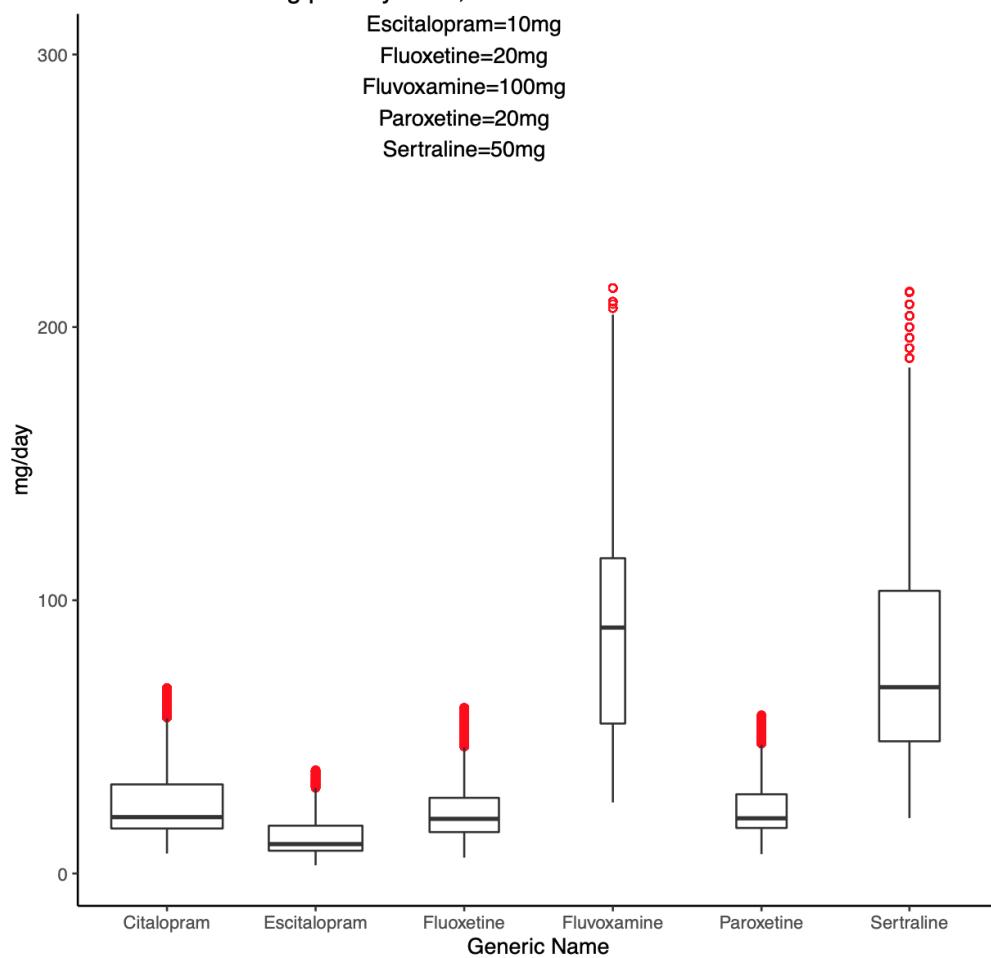


Can PRS be used to predict the dosage outcomes?
Track what is last constant dosage.

SSRI generic name switches?

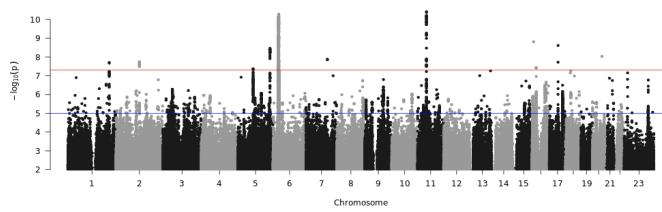
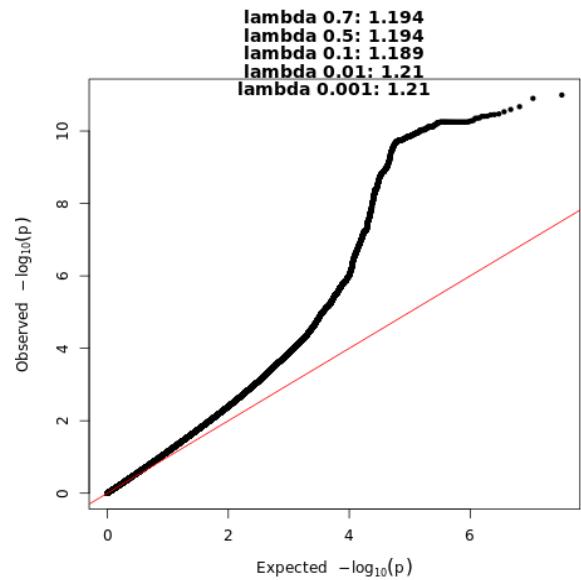
From [20200528_purchaseDistPerGeneric.SSRI.clean.pdf](#)

SSRI Generic Drug per day dose, clean for <10% > 90%



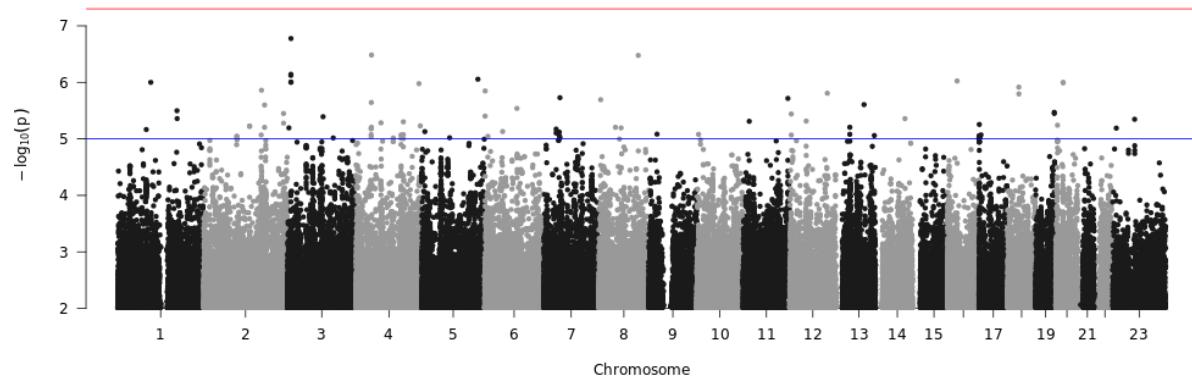
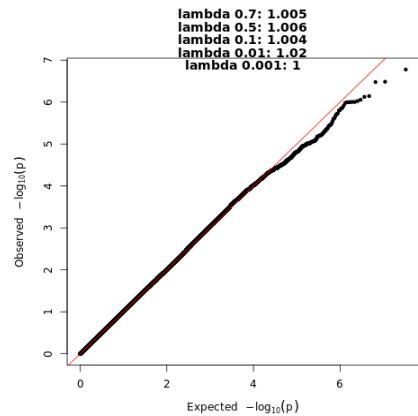
20200529: aad046fe-0af6-499b-aa80-676ad5b43091

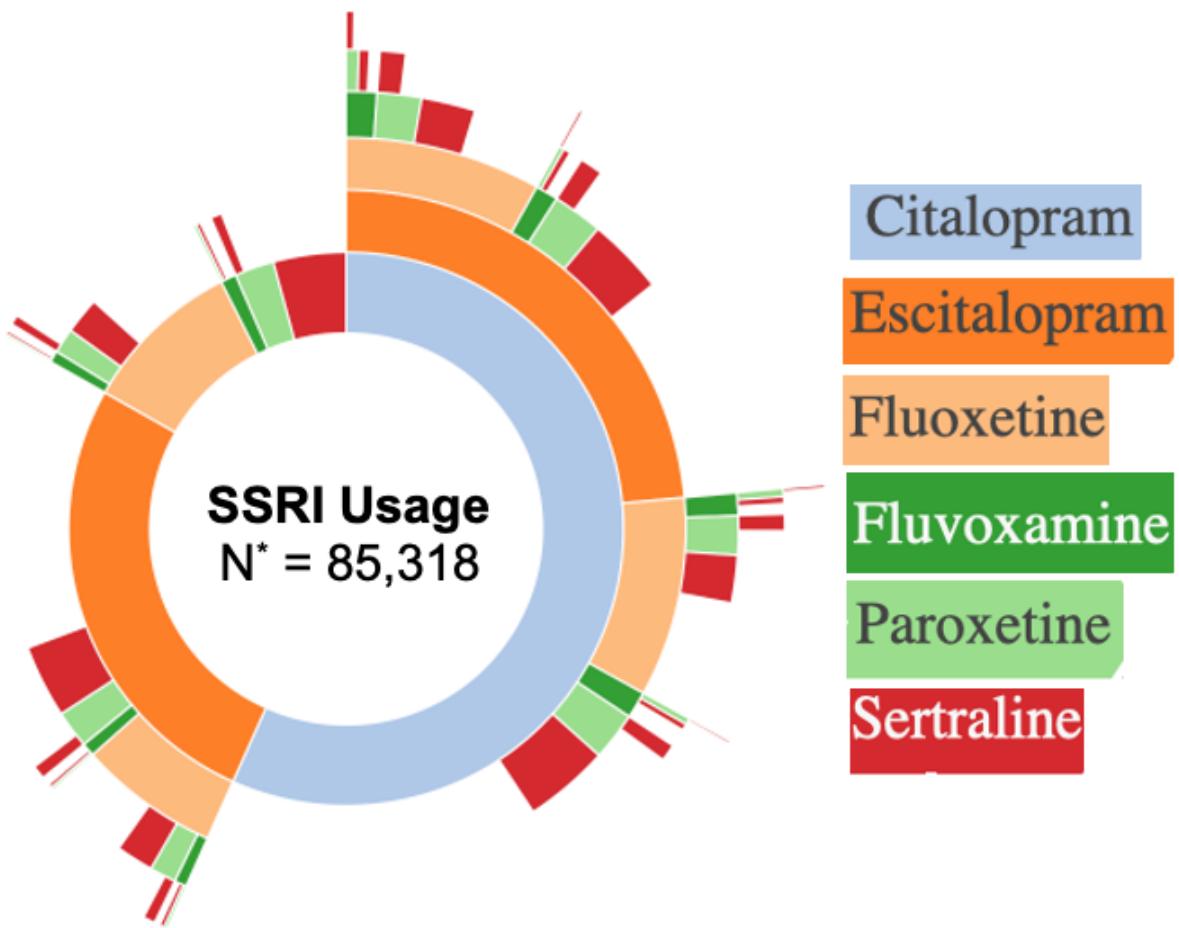
Highly inflated!!! Perhaps because the NAs were 0ed



Job ID: 61403335-108f-471a-98c1-3540b1c24cc1

Going to redo this with NAs as NA





* 28,799 points consume > 1 SSRI

** number of people = 20,430

[🔗 R5_SSRI_sunburst.html](#)

454 kB

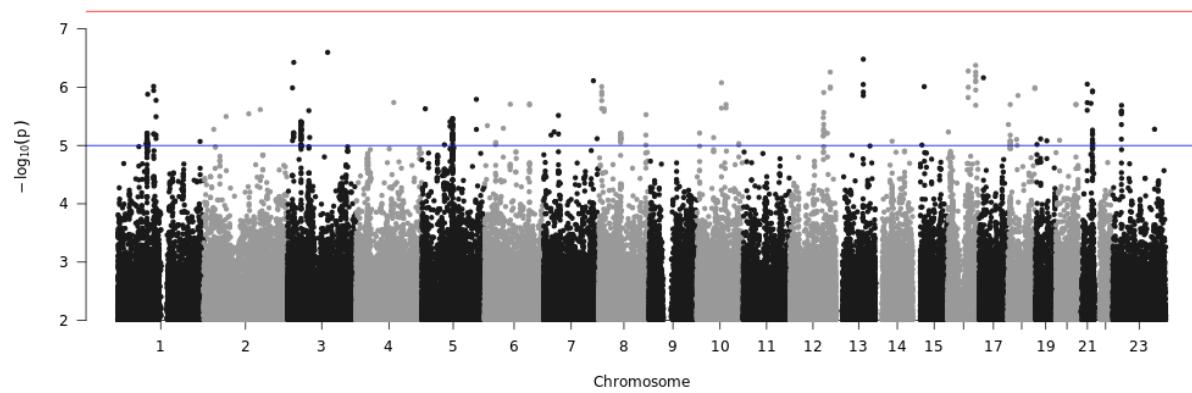
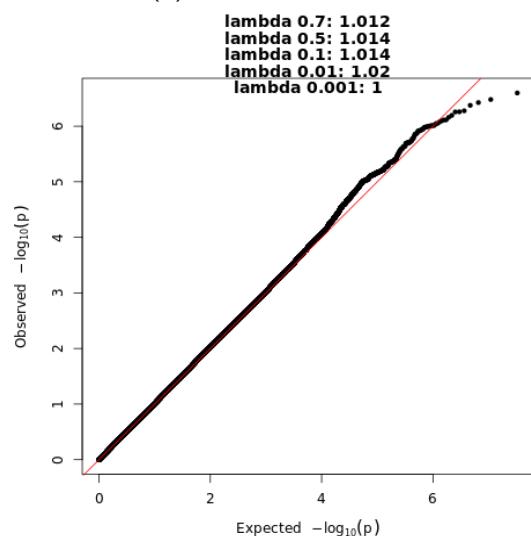
SSRI	Number of FinnGen IDs	Final count from latest purchases
Citalopram	27,545	18,052
Escitalopram	24,118	20,053
Fluoxetine	13,981	6,791

Fluvoxamine	2,282	933
Paroxetine	5,655	2,957
Sertraline	11,737	7,733
TOTAL	85,318	56,519

Binary trait for switching of SSRI from one type to another

didn't switch (0): 35178

switched (1): 19980



May need some data cleaning!

Augmenting or using drugs at the same time?

TODO

This week,

1) I'd also like to test out what differences are observed in responders of specific SSRI groups.

-- dosage normalization between SSRIs

2) Are there differences in depression as the outcome if I were just to subset for SSRI users?

```
summary(phenos.cov[phenos.cov$SSRIfreq>0,]$F5_DEPRESSIO)
  0      1  NA's
34248 19036 165508
summary(phenos.cov$F5_DEPRESSIO)
  0      1  NA's
192220 23424   3148
```

-- F5 phenotypes

Sent!

-- Rgs instead

-- set up phenotype for drug Rgs

3) Survival analysis on the SSRI users (Wei has to debug a few things before she can confidently point me to the GitHub repo).

-- meeting with Wei on Tuesday to talk about how to run it and what are the phenotype coding for it

-- survival analysis set up

4. autoregressor

1. Suggest autoregressor edits for requirement.txt

2.

For next week, I was thinking that I could summarize some of the work that I've been doing as an ASHG abstract and talk to you during our Friday AM meeting? What do you think, Mark? It would encompass the first set of hits from Juha's scan, and be complemented with whatever work that followed it since then.