Supplemental_Fig_S1. Changes compared to reference S288C genome. Diagrams show the transposon interruption of SRD1 in FY4 (**A**) and the integration of the reporter in the ancestral euploid *GAP1* CNV reporter strain (**B**). Topology diagrams for evolved strains indicating CNV breakpoints, orientations, and the occurrence of transposon events (**C-I**) and most likely mechanism of action TY: transposon-yeast event; ODIRA: origin dependent inverse triplication; NHEJ: non-homologous end joining; HR: homologous recombination

Supplemental_Fig_S2. Intersectional set of CNV amplified genes. Schematic diagram showing the CNV amplified genes at the GAP1 locus (**A**). ComQuad is entirely contained within ComTrip (shown as nesting brackets), which is in turn, entirely contained within Trip2, etc. Trip1 (dotted line) spans from YKR002W to YKR049C and so incompletely overlaps with ComQuad (missing YKR050W), ComTrip, Trip2, and Sup. The identical overlap of CNV amplified genes at the GLC7 locus is also shown for Trip1 and ComSup (**B**).

Supplemental_Fig_S3. There is no relationship between CNV size and relative fitness. A-B) The fitness of evolved strains containing *GAP1* CNVs was determined by pairwise competition experiments with a nonfluorescent, unevolved reference strain in glutamine-limited chemostats. Relative fitness was not found to significantly correlate with Additional Bases (**A**) (Adj.R-squared = -0.06, p.value = 0.46) or the number of genes amplified (**B**) (Adj.R-squared = 0.21, p.value = 0.17). **C-D**) Average growth rate of *GAP1* CNVs relative to the ancestral, euploid strain in YPGal batch culture. Relative growth rate was also found to not significantly correlate with Additional Bases (**C**) (Adj.R-squared = -0.19, p.value = 0.86) or the number of genes amplified (**D**) (Adj.R-squared = -0.2, p.value = 0.93).

Supplemental_Fig_S4. Copy-number corrected gene expression between the seventeen genes amplified in every CNV. Seventeen genes comprise the core set of genes that are amplified in every *GAP1* CNV included in this study. As such, their amplification may represent a commonality between the strains. Because they are not all amplified to the same degree either within or between strains (**Supplemental_Table_S4)** we correct for the copy-number of the CNV amplification to compare their expression relative to each other (See Methods). We find the strains themselves have different medians of expression over the core set, ranging from -0.2 to 0.55 Log2FC, consistent with strain differences in expression. We also observe that the median expression for each gene is different as well, ranging from -0.27 to 0.76 Log2FC. Only 5 genes have expression significantly different from expected given their copy-number, with DID2, SHB17, and YKR045C being significantly higher than expected and SET3 and UTH1 being significantly less.

Supplemental_Fig_S5. The number of unique insertion sites scales with the number of reads sequenced. The total number of unique insertion sites identified per library increases with the total number of reads sequenced (using all methods and sequencing runs).

Supplemental_Fig_S6. Weak correlation between CDS length and median transposon insertions.
The frequency of normalized insertions only weakly correlates with CDS length (Adj. R-squared = 0.198, p-value < 0.01). This suggests that length could be a confounding factor in tests that sought to compare CDSs of different lengths. Importantly, we are only comparing the CDSs of genes between strains, and as such the CDS lengths are identical in all cases.

In terms of detection limitations, the smallest non-dubious CDSs identified in the majority of samples are only 78 (RPL41A, RPL41B, and YOR302W) nucleotides long. Only one non-dubious CDS, YJR151W-A, is shorter (51 nucleotides) and was identified in 4 samples. The smallest CDS with identified insertions in all samples is the 87 nucleotide long YJL077W-A. This suggests that our detection ability is not limited to a subset of long CDSs.

Supplemental_Fig_S7. Empirical estimation of false negative rate by binned CDS length.
To estimate the occurrence of false negative zero counts in our analysis we calculated the frequency of false negatives between the two Euploid replicates.

We can evaluate the frequency of false zeroes by considering any event where one Euploid replicate is zero and the other is above the lower count threshold (ie. 50). Genome wide we find this happens rarely (0.003) however, this skews strongly by size as can be seen when categorizing the false negative rates by CDS size. Notably, however, we find that no 100 nucleotide category has a false negative rate of over 0.041, suggesting that the FDR for false negatives is below a 0.05 cutoff for even shorter CDS lengths.

Supplemental_Fig_S8. The number of unique insertion sites in the non-essential and essential genes of each strain. We find significantly fewer unique insertions in the essential genes relative to the non-essential genes in each strain (Mann-Whitney U, p-value <= 0.0001).

Supplemental_Fig_S9. A metagene analysis showing the mean unnormalized distribution of unique insertions relative to CDS length.

A) Representative distributions with genes separated by essential and non-essential designations as previously annotated by deletion and measurement of growth on rich media (yeast peptone dextrose) (Winzeler et al. 1999). We find noticeable separation in most strains **B)** As the essential and non-essential designation may be condition dependent we also compared metagene insertion distributions using a four quartile relative fitness measure made using rich media with 2% galactose from 3704 viable deletion mutant strains and 782 temperature-sensitive (TS) alleles (Costanzo et al. 2021). The first quartile (Q1, red) contains genes whose deletion causes the greatest measurable fitness defects, with relative fitness between 0.053 and 0.896. There was no relative fitness obtained for 21 genes (presumably there was no growth), these are marked NA (gray). We find distinct separation between the Q1 and NA and the other quartiles in all but one strain.

Supplemental_Fig_S10. Transposon insertions in non-amplified genes. A) Boxplots of unique insertion sites per gene, with individual genes plotted as points, for essential (red) and non-essential (blue) genes (Winzeler et al. 1999). All genes on Chromosome XI that are not within the CNV boundaries are shown. P-values from Welch's t-test are indicated by the following: ns: $p > 0.01$; ****: $p < 0.0001$. **B)** Linear regression was used to fit the normalized insertions per non-amplified gene in CNV strains (y-axis) to the mean number of normalized insertions per gene in the euploid replicates (x-axis), genome-wide. Adjusted p-values and slope from linear regression are annotated.

A.
Eu Ancestor Rep. 2 (Normalized Tn Hits)
Eu Ancestor Rep.1 (Normalized Tn Hits)
Adj R2: 0.83
p-value < 0.01
Slope: 0.93
Gene
Sig. Outlier (n = 198)
B.
Aneu | Aneu | DGY1728 (Normalized Tn Hits)
Ancestor (Normalized Tn Hits)
Adj R2: 0.71
p-value < 0.01
Slope: 1.03
Adj R2: 0.54
p-value < 0.01
Slope: 0.70
Gene
CNV Gene (CN=2, n = 251)
Sig. Outlier (S.residual 2)
42 of 251 CNVs are Sig. Outliers
88 of 4211 non-CNVs are SO
9.4-fold Enrichment
(FET, p-value < 0.01)
C.
Trip1 | ComSup | DGY1734 (Normalized Tn Hits)
Ancestor (Normalized Tn Hits)
Adj R2: 0.59
p-value < 0.01
Slope: 1.41
Adj R2: 0.72
p-value < 0.01
Slope: 0.86
Gene
CNV Gene (CN=2, n = 87)
CNV Gene (CN=4, n = 20)
Sig. Outlier (S.residual 2)
73 of 107 CNVs are Sig. Outliers
81 of 4355 non-CNVs are SO
133-fold Enrichment
(FET, p-value < 0.01)
D.
Trip2 | trip4 | DGY1736 (Normalized Tn Hits)
Ancestor (Normalized Tn Hits)
Adj R2: 0.80
p-value < 0.01
Slope: 1.77
Adj R2: 0.63
p-value < 0.01
Slope: 0.70
Gene
CNV Gene (CN=3, n = 69)
Sig. Outlier (S.residual 2)
39 of 69 CNVs are Sig. Outliers
80 of 4393 non-CNVs are SO
70-fold Enrichment
(FET, p-value < 0.01)

Supplemental_Fig_S11. R-squared linear regression of Tn insertions between CNV strains and Euploid ancestor.

A R-squared linear regression was performed for all CNV strains relative to the Euploid ancestor. Linear regressions were made for all genes (black circles, black line) or for CNV associated genes (red circles copy number CN= 2, blue triangles CN=3, green diamonds CN = 4, red line). Significant outliers are genes with standardized residuals greater than 2 (red squares). The enrichment of significant outliers in CNV associated genes relative to their occurrence in non-CNV associated genes was calculated using Fisher exact test. In each strain the CNV associated genes were significantly enriched in outliers (FET, p-value < 0.01).

Supplemental_Fig_S12. Heatmap of the ratio of unique normalized transposon insertions in the CNV containing strain relative to the Eu ancestor for Chromosome XI with a map of the CNV regions. This figure shows a heatmap representation of the ratio of unique normalized insertions in the CNV containing strain (each row) relative to the ancestor. Here we see a substantial clustering of higher insertions occuring within the CNVs (colored bars), consistent with SFig 11.

Supplemental_Fig_S13. Boxplot of Transposon insertion fold-change categorized by copy-number.
Here we show the log2FC of insertions in ChrXI across all strains. Insertions are separated by gene copy-number. Similar to Supplemental Fig_S6OLD, we find a near linear relationship between gene copy number and the increase in the median number of insertions observed. Notably, there is a compression of the median with increasing copy number, from 97% expectation for cn = 1 to 77% for cn = 4, which is most likely due to the saturation of unique hits.

Supplemental_Fig_S14. Enriched GO terms identified using Gene Set Enrichment Analysis (GSEA). GSEA was applied to a ranked gene list based on log2 fold changes in insertions, obtained from differential analysis comparing each CNV insertion profile to the euploid insertion profiles, with the false discovery rate (Q-valueFDR, circle size) for enriched terms set to 0.05. Positive enrichment scores (red) indicate functions that have increased insertions in the CNV strain. Negative enrichment scores (blue) indicate functions that have decreased insertion frequencies in the CNV strain. ComSup had no significant enrichment of any gene sets.

Supplemental_Fig_S15. Genetic interactions of CNV strains. A) Genes that have significantly different insertions in CNV strains versus euploid. Genes which were significant for at least one CNV strain, from differential analysis comparing each CNV insertion profile to the euploid insertion profiles. Positive log2FoldChange values have more insertions in CNV strains than euploid strains, while negative log2FoldChange have fewer insertions in CNV strains than euploid strains. If a gene is amplified the copy number is annotated. P-values adjusted with the Benjamini and Hochberg method: *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$. **B)** Additive genetic interaction for CNV and *BMH1* double mutants. Calculated from growth rates in Supplemental_File_4.

Supplemental_Fig_S16. mRNA expression of amplified genes is highly correlated with euploid expression. For each CNV, the subset of genes within the CNV boundaries are shown. Pearson's correlation coefficient and corresponding p-value are annotated.

Supplemental_Fig_S17. Insertions not correlated with mRNA expression of amplified gene expression. It could be the case that genes with higher levels of transcription may have a higher fitness cost than genes with lower levels of expression. To address this we calculated the correlation between the normalized transposon insertion (log2 fold-change relative to Euploid) and the normalized mRNA abundance (log2 fold-change relative to Euploid). For each CNV, the subset of genes within the CNV boundaries are shown (**A**). Taking the set of all non-CNV associated genes we see that there is a significant but negligible negative correlation (slope: -0.04, Adj. R-squared: 8.8e-4, p-value: 7.2e-10), (**B**). After correcting for copy number we also combined CNV amplified genes and found a similar negligible negative correlation although not significant (slope: -0.10608, Adj. R-squared: 2.9e-3, p-value: 0.075. This suggests that whatever fitness cost may be induced by the additional transcriptional burden of CNV amplified genes it is minor compared to other factors.

Supplemental_Fig_S18. Separation of changes in expression in relation to adaptation from those due to amplification. Here, a schematic using TIF1 / YKR059W (eIF1A) represents the estimation of CNV and adaptation effects on gene expression. In order to separate changes in expression due to CNV amplification and adaptation to glutamine-limited media we first separate genes by copy number (CN) with all CNVs being collected together. The TPM normalized RNA abundances are then copy-number corrected and a ratio is calculated relative to the Euploid ancestor. The two categories are then compared using Mann-Whitney U (MWU).

Supplemental_Fig_S19. Pearson correlation between CNV strains and Torres 2007 aneuploids grown in batch culture for ESR genes. Log2 fold change in mRNA expression comparing CNV or aneuploid strain to euploid strain. The data from Torres is the mean for all aneuploid strains measured.

Supplemental_Fig_S20. Pearson correlation between CNV strains and Torres 2007 aneuploids grown in chemostats for ESR genes. Log2 fold change in mRNA expression comparing CNV or aneuploid strain to euploid strain. The data from Torres is the mean for all aneuploid strains measured.

Supplemental_Fig_S21. Pearson correlation between CNV strains and Tsai 2019 aneuploids for CAGE genes. Log2 fold change in mRNA expression comparing CNV or aneuploid strain to euploid strain.

Supplemental_Fig_S22. Pearson correlation between Torres 2007 aneuploids and Tsai 2019 aneuploids for CAGE genes. The data from Torres is the mean for all aneuploid strains measured.

Supplemental_Fig_S23. Genes with significantly different mRNA expression from the euploid in all strains that are not on chromosome XI. Genes with positive log2FoldChange have higher expression in the CNV strain than the euploid strain.

Supplemental_Fig_S24. Select Log_2 FC transcript abundances across multiple aneuploid studies. A) Log2 mRNA expression for 436 genes (rows) significantly differentially expressed in at least one CNV strain versus the euploid strain. Data corresponding to genes from **(A)** from Torres et al. 2007 aneuploids in batch (**B**), Torres et al. 2007 aneuploids in chemostat (**C**), Tsai et al. 2019 (**D**), Hose et al. 2020 wild aneuploid strains (**E**), and Hose et al. wild aneuploid strains with ssd1 deleted (**F**). The first five (**A-E**) are compared to closely related euploids, while the aneuploids with ssd1 deletions are compared to their wild-type aneuploid counterparts (**F**).

Supplemental_Fig_S25. DESeq and GSEA differences between Trip1, ComQuad and Aneu

To better understand potential differences between the Trip1, ComQuad and Aneu strains we also performed DESeq to identify genes with significantly different mRNA abundances between these strains. We found that the most significant outlier in expression between these strains was INH1, a regulatory inhibitor of mitochondrial function, and SFT1, a INH1 paralog both associated with CCCP sensitivity (Ichikawa et al. 1990). Both of these are significantly higher in the Aneu strain (3.6 log2FC and 1.5 log2FC, respectively) and other CCCP resistant strains than in the CCCP sensitive strains. To help characterize how these strains may have distinct system level differences, we next identified genes with significantly different transcript abundances (DESeq2, adj.p-value <= 0.05) between each Aneu, Trip1, and ComQuad strain and the BMH1 insertion sensitive strains. A GSEA performed on these significantly different genes found that both ComQuad and Trip1 are enriched in suppressed ‘respiratory chain complex’, ‘respirasome / mitochondrial respirasome’ and ‘inner mitochondrial membrane protein complexes’. Intriguingly, Aneu alone is enriched in activating ‘mitochondrial protein-containing complexes’. This is suggestive of large scale differences in mitochondrial function particular to these strains

.

Supplemental_Fig_S26. Effect of CCCP treatment on CNV strain growth rate. Average and standard deviation (error bars) final optical density (OD) relative to the ancestral, euploid strain in YPGal batch culture in either control condition or with 25 μM CCCP. Strains Trip1 and ComQuad have greatly (>75%) reduced growth under CCCP treatment suggesting these strains may have a mitochondrial disorder.

Supplemental Tables: All Supplementary tables are publicly available here:

https://github.com/pspealman/CNV_essentiality/tree/main/Supplementary_Tables/

Supplemental_Table_S1. Strain characteristics. CNV type based on long and short read sequencing and genome assembly from (Spealman et al. 2022). More information about SNPs/indels including reference sequence and mutant sequence can be found in (Lauer et al. 2018) S10 Table https://doi.org/10.1371/journal.pbio.3000069.s027.

Strain name	Strain name new	Gresham Lab Name	*GAP1* hypothesized CNV type	Clone ID in Lauer et al. 2018	Generation Isolated	SNPs/indels
eu	Eu	DGY1657	NA	NA	NA	NA
aneu	Aneu	DGY1728	Aneuploid	gln_01_c1	150	YNL284C-B missense variant; YPL232W (SSO1) disruptive in frame deletion
trip1	ComSup	DGY1734	Complex: supernumerary chromosome and ODIRA	gln_02_c3	250	YHL002W (HSE1) missense variant; Chr XIV:96555 non-coding variant; Chr XIV:96603 non-coding variant
trip2	Trip1	DGY1747	ODIRA	gln_08_c2	150	YMR129W (POM152) missense variant; Chr V:431779 non-coding variant; Chr XII:915075 non-coding variant
trip3	ComTrip	DGY1751	Complex: ODIRA and homologous recombination	gln_09_c3	250	YOL103W-A missense variant; YNR031C (SSK2) stop gained
trip4	Trip2	DGY1736	ODIRA	gln_03_c2	250	YJR152W (DAL5) stop lost & splice region variant & conservative inframe deletion; Chr V:55180 non-coding variant; Chr X:524178 non-coding variant; Chr X:745685 non-coding variant
iso	Sup	DGY1744	Supernumerary chromosome	gln_07_c1	250	YMR171C (EAR1) missense variant; YJL128C (PBS2) missense variant; Chr XV:594618 non-coding variant
quad	ComQuad	DGY1740	Complex: ODIRA + transposon	gln_05_c1	150	YOL077C (BRX1) missense variant; YNL338W frameshift_variant

Supplemental_Table_S2. Evaluation of the 19 single nucleotide variants identified in the CNV containing strains. Table contains each SNV identified in the evolved strains and not inherited from the ancestor. Details include locus, gene or nearest-neighbor gene potentially affected, and potential effect.

Supplemental_Table_S3. Evaluation 5 nucleotide variants within CDS identified in the CNV containing strains.

Lab Name	Strain name	Systematic name	type	Gene	Gene ontology	Gene associated pathway	Gene interaction
DGY1728	Aneu	YPL232W	disruptive in frame deletion	SSO1	nitrogen compound transport [GO:0071705]		SLT2 / YHR030C
DGY1747	Trip1	YMR129W	missense_ variant	POM152	nitrogen compound transport [GO:0071705]		
DGY1751	ComTrip	YNR031C	stop gained	SSK2	nitrogen compound transport [GO:0071705]; osmosensory signaling MAPK cascade [GO:0000161]	HOG pathway	CYT1 / YOR065W; SLT2 / YHR030C
DGY1736	Trip2	YJR152W	stop lost & splice region variant & conservative in frame deletion	DAL5	nitrogen compound transport [GO:0071705]		CYT1 / YOR065W; SLT2 / YHR030C
DGY1744	Sup	YJL128C	missense variant	PBS2	osmosensory signaling MAPK cascade [GO:0000161]	HOG pathway	CYT1 / YOR065W; SLT2 / YHR030C

Supplemental_Table_S4. Change in genome size relative to ancestor.

Name	Strain	CNV associated (bp)	rDNA locus (bp)	Total Change (bp)
Aneu	DGY1728	666816	-1149393	-482577
ComSup	DGY1734	372950	-687962	-315012
Trip2	DGY1736	321710	307378	629088
ComQuad	DGY1740	79151	-344556	-265405
Sup	DGY1744	396338	-544994	-148656
Trip1	DGY1747	191518	-464468	-272950
ComTrip	DGY1751	168878	-66842	102036

Supplemental_Table_S5. Calculation of rDNA locus copy-number. Each row contains the depth of the rDNA feature relative to the genome depth. The median for these features is calculated and treated as the copy number in that strain. This estimated copy-number is then used to calculate the total number of nucleotides lost or gained.

	DGY1657	DGY1728	DGY1734	DGY1747	DGY1751	DGY1736	DGY1744	DGY1740
ETS1-1	83.5	36.3	61.0	65.5	82.0	94.4	65.0	73.5
ETS1-2	69.2	32.8	44.1	50.8	61.2	70.8	50.9	57.5
ETS2-1	65.9	19.9	33.3	42.8	45.3	59.5	44.9	39.0
ETS2-2	84.5	32.8	62.4	68.5	82.9	95.5	66.0	71.2
RDN18-1	80.9	32.0	58.9	64.7	78.6	89.7	61.8	70.8
RDN18-2	71.7	30.2	51.3	57.2	68.2	76.9	54.6	62.9
RDN25-1	68.7	26.6	47.0	53.2	61.6	70.7	50.5	55.2
RDN25-2	82.4	29.9	61.0	66.1	80.9	89.4	63.8	67.5
RDN37-1	73.0	28.5	51.4	58.5	70.7	79.0	55.1	63.0
RDN37-2	78.3	30.0	55.6	63.1	77.0	86.0	60.4	66.2
RDN5-1	84.3	33.8	63.1	69.7	83.8	93.9	65.8	73.3
RDN5-2	65.7	17.9	28.9	40.7	42.3	56.0	43.1	35.7
RDN58-1	76.8	29.6	53.3	62.1	74.3	85.2	59.0	67.2
RDN58-2	77.7	29.2	55.0	63.3	76.3	86.5	60.5	67.4
Median	77.3	29.9	54.1	62.6	75.3	85.6	59.7	66.7
Percentage	1.000	0.387	0.701	0.810	0.975	1.108	0.773	0.863
Size	1351834	202441	663872	887366	1284992	1659212	806840	1007278
Bp change	0	-1149393	-687962	-464468	-66842	307378	-544994	-344556

Supplemental_Table_S6. Median Relative Depth of DNA per Gene, CNV, and Gene Copy Number
This table contains the median long-read sequencing depth per gene relative to the genome depth, this is resolved into copy number using structural breakpoint information.

Supplemental_Table_S7. ncRNA enrichment. This table contains the results of the ncRNA enrichment analysis wherein frequency of ncRNA elements within CNVs was tested for enrichment relative to the global background frequency. Only tRNA was found to be significantly enriched and even then only in the *GLC7*-YER137C CNV.

Supplemental_Table_S8. All strains Insertions per Gene, not normalized. Raw unique count of insertions per gene (CDS) per strain per replicate.

Supplemental_Table_S9. All strains Insertions per Gene, normalized. Normalized unique count of insertions per gene (CDS) per strain per replicate.

Supplemental_Table_S10. All strains Insertions per Gene, normalized, median. Median of replicate abundance of normalized unique counts of insertions per gene (CDS) per strain.

Supplemental_Table_S11. Hermes mutagenesis library characteristics for uniquely identified insertion sites.

Sample	Total sites	Minimum reads per position	Maximum reads per position	Mean reads per position	Median reads per position
Eu_1	172384	1	4761	20.09	8
Eu_2	136167	1	2966	14.56	5
Aneu	301220	1	26598	22.45	4
ComSup	95152	1	2722	15.80	4
Trip1	85327	1	2071	10.82	3
ComTrip	122326	1	8531	23.86	6
Trip2	329624	1	10567	18.73	5
Sup	126562	1	6620	23.58	6
ComQuad	221218	1	8455	17.87	4

Supplemental_Table_S12. Pearson correlation of insertions per gene for different sequencing runs.

	Euploid_1_nyc2	Euploid_2_nyc1	Euploid_2_nyc2
Euploid_1_nyc1	0.979	0.918	0.903
Euploid_1_nyc2		0.906	0.896
Euploid_2_nyc1			0.966
	Aneu_bgi2	Aneu_nyc1	
Aneu_bgi1	0.987	0.94	
Aneu_bgi2		0.94	
	ComSup_nyc2		
ComSup_nyc1	0.957		
	Trip1_nyc2		
Trip1_nyc1	0.904		
	ComTrip_nyc2		
ComTrip_nyc1	0.966		
	Trip2_bgi2	Trip_nyc1	
Trip2_bgi1	0.974	0.928	
Trip2_bgi2		0.937	

	Sup_nyc2		
Sup_nyc1	**0.98**		
	ComQuad_bgi2	**ComQuad_nyc1**	
ComQuad_bgi1	**0.977**	**0.867**	
ComQuad_bgi2		**0.855**	

Supplemental_Table_S13. Summary of R-squared Outliers. Tab-delimited table containing Normalized Tn abundances, copy-number, and standardized residuals for each gene in each strain relative to the euploid strain.

Supplemental_Table_S14. Calculation of number of genes exceeding R-squared significance threshold Tab-delimited table containing the results of the DSG evaluation, namely, 'cnv_hits' is the times a CNV associated gene met the significance criteria (standardized residual > 2 and copy_number_corrected_log2FC > 1) , 'cnv_miss' for when it failed those criteria. 'Non_hits' and 'non_miss' are the same test applied to non-CNV associated genes.

Supplemental_Table_S15. Summary of Insert Outliers and Proportional Covariants. Conceivably, some genes may significantly deviate from insert frequency expectations (significant outliers, **Supplemental_Table_S14**) because of compensatory changes in other genes. To determine if this is the case we looked for genes with high rates of proportional correlation across strains, for example a 2-fold increase in one gene sees a 2-fold change in the other, and this proportionality is consistent across strains. We performed this check across all genes that were significant outliers in frequency of insertions. We empirically derived an estimated FDR by calculating the background frequency using genes that were not significant outliers. While we did find a small number of significant outliers that had proportionality with other genes no strain had more than would be expected at random.

Supplemental_Table_S16. Genes with no insertions in euploid replicates. These are genes that have no insertions in either replicate of the euploid strain 1657. If they were previously annotated as essential, they are labeled "yes" (Winzeler et al. 1999). Relative fitness for some of these genes on media with galactose was previously measured (column "Galactose") (Costanzo et al. 2021), and are labeled as "low fitness galactose" if that relative fitness measure was less than one.

Supplemental_Table_S17. Genes with length normalized and copy-number corrected insertions, with z-score values. Table contains CDS length normalized and copy-number corrected insertion abundances from (**Supplemental_Table_S9)**, along with the cross sample global z-score.

Supplemental_Table_S18. Results from differential analysis of number of insertions per CNV strain compared to euploid replicates. Table contains genes with significantly differently abundant numbers of insertions as calculated by DESeq2 using (**Supplemental_Table_S9**).

Supplemental_Table_S19. Gene set enrichment analysis of log2 fold change number of insertions per CNV strain compared to euploid replicates. Table contains the Revigo reduced set of GO terms from the significantly enriched gene set generated using clusterprofiler (**Supplemental_Table_S18.xlsx)**

Supplemental_Table_S20. Pearson correlation of RNA abundance for different sequencing runs.

	Euploid_2	Euploid_3
Euploid_1	0.989	0.995
Euploid_2	1	0.984
	Aneu_2	Aneu_3
Aneu_1	0.996	0.955
Aneu_2	1	0.956
	ComSup_2	ComSup_3
ComSup_1	0.981	0.937
ComSup_2	1	0.974
	Trip_3	
Trip_1	0.991	
	ComTrip_2	
ComTrip_1	0.999	
	Trip2_2	Trip_3
Trip2_1	0.999	0.999
Trip2_2		1

	Sup_2	**Sup_3**
Sup_1	**0.998**	**1**
Sup_2		**0.999**
	ComQuad_2	**ComQuad_3**
ComQuad_1	**0.237**	**1**
ComQuad_2		**0.228**

Supplemental_Table_S21. RNAseq read counts table for each strain. Tab delimited results of BedTools coverage run of each sample, gene name corrected to Standard Name (SGD), only counting protein coding genes (Y*).

Supplemental_Table_S22. TPM normalized RNAseq abundances table for each strain. The TPM normalized values of **Supplemental_Table_S21.** Because DESeq2 requires unnormalized read matrices but other analyses require normalization, this file was generated to provide support for the latter.

Supplemental_Table_S23. Results from DESeq2 of counts per gene from RNAseq for each CNV strain compared to euploid. Table contains genes with significantly differently abundant mRNA abundances as calculated by DESeq2 using (**Supplemental_Table_S21**)

Supplemental_Table_S24. Mann-Whitney U test for Log2FoldChange of gene expression. This table shows the result of a Mann-Whitney U test comparing the log2 transformed TPM normalized RNA-seq abundances between the evolved CNV containing strains and the ancestral euploid strains. This test is conducted on the CNV associated genes and copy number normal genes. We find that the mean of CNV associated genes is significantly higher in the CNV strains than the euploid ancestor, with the CNV strains on average being 1.28 FC higher. Conversely non-CNV associated genes show no consistent FC across the CNV strains, with an average of 1.01 FC higher.

CN_state	Strain	U_statistic	Evolved_Median	Ancestral_Median	Evo_over_Anc_ratio	p-value
CNV	DGY1728	544154	5.960782	5.133703	1.161108	1.86E-16
CNV	DGY1734	90467	5.921132	4.926497	1.201895	2.35E-10
CNV	DGY1736	42994	6.737724	4.964693	1.357128	1.34E-26
CNV	DGY1740	2498.5	6.808703	4.594912	1.481792	0.000249
CNV	DGY1744	71967.5	6.239757	4.892394	1.2754	3.27E-18
CNV	DGY1747	10934	6.419402	4.725971	1.358324	5.78E-08
CNV	DGY1751	18748.5	5.819294	5.040278	1.154558	0.00013
Non_CNV	DGY1728	1.36E+08	4.803955	4.917162	0.976977	1.85E-08
Non_CNV	DGY1734	1.54E+08	5.040854	4.931223	1.022232	0.000134
Non_CNV	DGY1736	1.64E+08	5.230348	4.930947	1.060719	1.72E-32
Non_CNV	DGY1740	1.73E+08	5.355736	4.932343	1.08584	3.58E-70
Non_CNV	DGY1744	1.55E+08	5.046671	4.932278	1.023193	6.25E-05
Non_CNV	DGY1747	65623251	4.939313	5.06196	0.975771	1.73E-08
Non_CNV	DGY1751	64253171	4.678993	4.905237	0.953877	1.17E-13

Supplemental_Table_S24. Expected expression values. Tab delimited file with CNV expected expression as calculated by the ancestral mRNA abundances (**Supplemental_Table_S21**). Multiplied by the copy-number from (**Supplemental_Table_S6).**

Supplemental_Table_S25. DESeq2 results from Observed versus Expected values. Results from differential analysis of expected counts per gene **Supplemental_Table_S24)** from RNAseq for each CNV strain compared to observed abundances (**Supplemental_Table_S21).**

Supplemental_Table_S26. Table of FET analysis of DESeq2 results of CNV expression rates for both Observed and Expected values. Observed data is the evolved strain compared to the ancestor with no copy number correction and we find significantly higher expression in CNV associated genes in each strain, compared to the ancestor, consistent with gene amplification models. Expected data is expression in the CNV strain compared to copy number corrected ancestor expression. We do not find a significant difference in expression of CNV associated genes from what is expected given their copy number, suggesting there is no dosage compensation specific to CNVs.

Supplemental_Table_S27. To evaluate the change in expression (Log2FC relative to euploid) observed between CNV amplified genes and their copy number normal (CNN) counterparts we performed a series of pairwise Wilcoxon comparisons between strains. This table contains the CNV strain, the CNN strain, the number of genes being compared, the median CNV Log2FC expression, the median CNN log2FC expression, the difference and Wilcoxon calculated p-value.

Supplemental_Table_S28. To separate the observed gene expression into expression changes (relative to the euploid and glutamine-limited growth condition naive ancestor) due to CNV gene amplification versus adaptation of expression to glutamine-limited growth conditions we first performed binning by copy-number (**Supplemental_Table_S6)** then compared TPM normalized expression distributions (**Supplemental_Table_S22)** between bins using Mann-Whitney U. Table contains gene name, ratio, and MWU p-value.

Supplemental_Table_S29. Gene set enrichment analysis of log2 fold change counts per gene from RNAseq for each CNV strain compared to euploid. Table contains clusterprofiler output using **Supplemental_Table_S23**

Supplemental_Table_S30. Results of hypergeometric test for over-representation of GO terms in clustered, differentially expressed genes. Table derived from hypergeometric test performed on **Supplemental_Table_S29.**

Supplemental_Table_S31. Results of hypergeometric test for over-representation of GO terms in clustered, differentially expressed genes excluding genes on chromosome XI. Table derived from hypergeometric test performed on **Supplemental_Table_S29** with chromosome XI filtered out**.**

Supplemental_Table_S32. Oligos and Primers used in this work.

Oligo Name	Purpose	Sequence
Hermes_F	Inverse PCR to generate PCR product used for sequencing at BGI	TGATTCATCGACACTCGG
Hermes_R	Inverse PCR to generate PCR product used for sequencing at BGI	TCATAAGTAGCAAGTGGCGC
Nextera hermes_enrichment	Amplify hermes containing fragments and add an i5 adaptor	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNtcataagtagcaagtggcgc
Nextera i7_enrichment	Amplify hermes containing fragments	GTCTCGTGGGCTCGG
i5_amp	i5 end amplification	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG

Supplemental Methods:

Yeast Strains

The euploid ancestral GAP1 CNV reporter and the evolved GAP1 CNV strains were previously described and characterized in Lauer et al. 2018, and are haploid derivatives of the reference strain S288C (more specifically, FY4/5) with a constitutively expressed mCitrine gene and KanMX G418-resistance cassette inserted 1,118 base pairs upstream of GAP1. This construct is referred to as the GAP1 CNV reporter.

Evaluation of SNVs identified in CNV strains.

Strains were sequenced and SNVs identified as described in Lauer et al. (2018). Here we evaluated the potential impact each SNV may have on the organism (**STable 2**) using Ensembl's VEP (McLaren et al. 2016). Each SNV is also evaluated for significant (DESeq2, adj.p-value <= 0.05) changes in transposon insertion abundance (DESEQ_insertions_log2_FC_relative_to_Eu, DESEQ_insertions_padj) and transcript abundance (DESEQ_mRNA_log2_FC_relative_to_Eu, DESEQ_mRNA_padj). If the SNV occurs inside a gene then the insertion and transcript abundances are calculated for that gene, if it occurs in a non-coding or intergenic region it is evaluated for all proximal (500 nucleotides) CDSs.

No CDSs associated with SNVs have significant differences in insertion or transcript abundance except for a Gag Protein (YNL284C-B) in the Aneu strain with a 1.74-fold increase in mRNA. Importantly, these are not direct tests of the effect of a variant on the fitness of an organism and the ultimate effect they have may not be visible using only insertions and transcript abundance.

To further evaluate the potential effects these genes had, we separated the SNVs into low probability severity and high probability severity groups (**STable 3**). The majority (14 out of

19) of SNVs were categorized as low severity, being in non-coding regions, transposons, tRNA, telomeres, mis-sense mutations with high (>0.05) SIFT scores (Kumar et al. 2009), or dubious ORFs. For the remaining 5 SNVs (each of which is isolated in their own strain) we then evaluated the associated genes for shared ontologies, pathways, and interactions. We found 4 genes (*SSO1*, *POM152*, *SSK2*, *DAL5*) with nitrogen compound transport [GO:0071705] ontologies. We found 2 genes (*SSK2* and *PBS2*) with osmosensory signaling MAPK cascade [GO:0000161], also known as the HOG pathway. We also found that 4 genes (*SSO1*, *SSK2*, *DAL5*, *PBS2*) that interact with SLT2, a Serine/threonine MAP kinase known to play a role in stress-induced mitophagy that parallels the *HOG1* stress-induced mitophagy pathway (Mao et al. 2011), (Feng et al. 2014). We also identified 3 genes (*SSK2*, *DAL5*, *PBS2*) that interact with *CYT1*, a subunit of the mitochondrial respiratory complex III.

Growth analysis in batch culture

To evaluate growth rates under the same conditions as those used to induce transposon mutagenesis we performed growth rate analysis in YPGal batch cultures. For each experiment, we inoculated three colonies per strain into 3-5 mL YPGal, and grew them overnight at 30℃. In triplicate per original colony, we back diluted 5 µL of culture into 195 µL fresh YPGal or YPGal with 25 µM carbonyl-cyanide 3-chlorophenylhydrazone (CCCP) in a Costar Round Bottom 96 well plate (Ref 3788). We treated the lid with 0.05% Triton X-100 in 20% ethanol to prevent condensation (Brewster 2003). We collected OD600 data over approximately 48 hours using a Tecan Spark with the following parameters: Temperature control: On; Target temperature: 30 [°C]; Kinetic Loop; Kinetic cycles: 530; Interval time: Not defined; Mode: Shaking; Shaking (Double Orbital) Duration: 240 [s]; Shaking (Double Orbital) Position: Current; Shaking (Double Orbital) Amplitude: 2 [mm]; Shaking (Double Orbital) Frequency: 150 [rpm]; Mode: Absorbance; Measurement wavelength: 600 [nm]; Number of flashes: 10; Settle time: 50 [ms]; Mode: Fluorescence Top Reading; Excitation: Monochromator; Excitation wavelength: 497 [nm];

ExcitationBandwidth: 30 [nm]; Gain: Calculated From: B5 (50%); Mirror: AUTOMATIC; Number of flashes: 30; Integration Time: 40 [μs]; Lag time: 0 [μs]; Settle time: 0 [μs]; Z-Position mode: From well B5.

Transposon mutagenesis

A single transformant for each strain was used to inoculate a 30 mL YPD + 200 μg/mL Hygromycin B, and incubated approximately 24 hours at 30℃ with agitation, until OD5. To induce transposition, the culture was then diluted to OD = 0.05 in YPGalactose + 200 μg/mL Hygromycin B to a final volume of 50 mL, and incubated 24 hours at 30℃ with agitation. The culture was diluted to 0.05 in 50 mL YPGalactose + 200 μg/mL Hygromycin B and incubate 24 hours three more times, for a total of four serial transfers in YPGalactose + 200 μg/mL Hygromycin B. The culture was pelleted by centrifugation for five minutes at 4000 rpm, the supernatant removed, then resuspended to OD0.5 in 50 mL YPD and incubated 24 hours at 30℃ with agitation, then diluted again to OD0.5 in 50 mL YPD and incubated 24 hours at 30℃ with agitation, to release selection to maintain pSG36_HygMX. The cultures were then diluted to OD = 0.5 in 100 mL YPD + 200 μg/mL Hygromycin B and incubated 24 hours at 30℃ with agitation to select for cells with the transposon in the genome. The final culture was pelleted by centrifugation for five minutes at 4000 rpm, the supernatant removed, resuspended with 1 mL sterile water, split into four 250 μL aliquots, and pelleted for two minutes at 8000 rpm. The supernatant was removed and cell pellets were frozen at -20℃ for storage until DNA extraction was performed.

Insertion site sequencing

DNA was extracted from cell pellets using the MasterPure™ Yeast DNA Purification Kit (Lucigen, cat #MPY80200), with an additional initial incubation with zymolyase at 37°C to enhance cell lysis, and using a Glycogen/Sodium Acetate/Ethanol DNA precipitation(Green and

Sambrook 2016). For each sample, 2 µg of DNA was digested with 50 units of DpnII and 5 µL NEBuffer™ DpnII (NEB, cat #R0543L), in a total volume of 50 µL; and 2 µg of DNA was digested with 50 units of NlaIII and 5 µL CutSmart® Buffer (NEB, cat #R0125L), in a total volume of 50 µL, for 16 hours at 37℃. The reactions were heat inactivated, then circularized by ligation in the same tube with 25 Weiss units T4 Ligase and 40 µL T4 ligase buffer (Thermo Scientific cat #EL0011) for 6 hr at 22°C, in a volume of 400 µL. Circularized DNA was precipitated using a Glycogen/Sodium Acetate/Ethanol DNA precipitation (Green and Sambrook 2016). Inverse PCRs for each sample and digestion were performed with primers Hermes_F and Hermes_R with 0.5 µL of each circularized DNA sample per reaction. PCR was performed with DreamTaq (ThermoFisher cat #EP0701), with the following program: 2 min at 95°C followed by 32 cycles of 30 s at 95°C, 30 s at 57.6°C, 3 min at 72°C, and a final extension step of 10 min at 72°C. The PCRs products were confirmed on 2% agarose gels, and the concentration was quantified using Qubit™ dsDNA BR Assay Kit.

Library preparation and sequencing were performed using two different library preparation methods and sequencing set ups as follows.

For each sample (1728, 1736, and 1740) and digestion, 35 PCR reactions using primers (**STable 32**) each with 0.5 µL of each circularized DNA were performed as described above and the PCR products were pooled and cleaned using a Glycogen/Sodium Acetate/Ethanol DNA precipitation (Green and Sambrook 2016). For each sample, at least 6 µg at minimum 30 ng/µl was then sent to the BGI (Beijing Genomics Institute) for library preparation and sequenced using a paired-end (2 x 100) protocol on an Illumina Hi-Seq 4000 or DNBseq platform.

For each sample and digestion 4 PCR reactions were performed as described above and the PCR products were pooled by sample and cleaned using a Glycogen/Sodium Acetate/Ethanol DNA precipitation (Green and Sambrook 2016). Five ng of each PCR product pool was used as input into a modified Nextera XT library preparation. To increase library complexity, for each

sample, two tagmentation reactions were performed. PCR to enrich for fragments with hermes sequence and add an i5 adaptor were performed on the tagmented DNA using NPM Buffer, primers Nextera_hermes_enrichment and Nextera_i7_enrichment, and the following program: 3 min at 72°C, then 30 s at 95°C, followed by 9 cycles of 10 s at 95°C, 30 s at 55°C, 30 s at 72°C, and a final extension step of 5 min at 72°C. The reactions were pooled by sample, cleaned using AmPure XP beads, and resuspended in 20 µL of molecular grade water, which was used as input for an indexing and library amplification PCR. Each sample was indexed with an i7 index from the Nextera XT kit, and amplification of the i5 end was performed with primer i5_amp (which contains no i5 index), using the 2X KAPA PCR master mix (Roche cat. #KK2611), and the same program described for the PCR after tagmentation.PCR cleanup and size selection was performed with AmPure XP beads. The fragment size of each library was measured with an Agilent TapeStation 2200 and qPCR was performed to determine the library concentration. The libraries were pooled at equimolar concentrations, and sequenced using a single-end (1 x 150) protocol on an Illumina NextSeq 500. Libraries were prepared once, but sequenced in two consecutive sequencing runs for increased coverage.

Transposon insertion sequencing site identification and annotation

Using cutadapt v1.16 (Martin 2011) with the expected Hermes TIR sequence on the 5' end were identified, and the TIR was trimmed. If the TIR was followed by plasmid sequence, these reads were discarded. For reads sequenced at BGI (paired end sequencing), the read with the TIR sequence was identified and its mate was discarded. For reads sequenced at NYC (Nextera based prep, single end sequencing), Nextera transposase sequences were identified and removed. Reads with a length less than 20 bases after all cleaning steps were discarded, and the remaining reads were checked for quality using fastqc v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the modified reference genome using bwa mem v.0.7.15 (Li and Durbin 2010) and BAMs were generated

with samtools v1.9 (Li et al. 2009). Samples prepared and sequenced by more than one method had high Pearson correlations (0.85-0.94) in the number of unique insertions identified per gene (**STable 8**), and therefore were combined into a single BAM file before performing downstream analysis. For the majority of the analyses, BAMs were combined by sample, for ease of processing and to prevent redundant insertion site identification. BAMs were parsed with a custom python script which identifies the first base of the read as the position of the insertion. The script output all unique insertion positions and the number of reads per insertion position. Positions were annotated using bedtools v2.26.0 (Quinlan and Hall 2010) and a custom GFF containing amended annotations for the custom genome (Supplemental File 1). All analyses use unique insertion positions, and do not take into account the number of reads per unique insertion position. Uniquely identified insertion sites are supported by an average of 18.6 sequencing reads. The libraries have between 85,327 and 329,624 unique insertion sites identified, with an average of 176,664 insertion sites, corresponding to approximately one insertion per 69 bases in the yeast genome (NCBI R64 assembly**; STable 4**). We normalize for differences in sequencing depth by calculating insertions per million: number of unique insertion sites per feature/(total unique insertion sites/1,000,000) (Levitan et al. 2020) and require a minimum of 50 insertions per million per feature for all comparisons. We do not normalize for gene length, as we are comparing genes between strains, not within strains. The hermes transposon method does show an insertion preference for nucleosome-free regions, which tend to be right before and right after genes. However, as we consider only coding regions and do not perform comparisons between genes within the genome, differential nucleosome occupancy is unlikely to impact our analysis.

Transposon insertion frequency and false negative rates

The median rate of transposon insertion per nucleotide within CDS regions genome wide for all samples is 0.115 per nucleotide or ~1 insertion per 12 nucleotides. If we only consider the

Euploid replicates this is virtually unchanged at 0.118 per nucleotide. Note that these are lower frequencies than that calculated when including intergenic regions (~1 insert per 7 nucleotides) but it allows us to set a conservative lower bound of 50 normalized unique insertions when testing for significant differences in CDS insertion frequencies.

The frequency of normalized insertions only weakly correlates with ORF length (Adj. R^2 = 0.285, p-value < 0.01, **SFig 6**). The smallest ORF identified in the majority of samples was 78 nucleotides long and the smallest ORF identified in all samples was 87 nucleotides long.

We estimated the false negative rate in our analysis using the two Euploid replicates, which show good agreement (Adj. R-squared: 0.83, p-value <0.01, **SFig 11A**), and determining the frequency with which one Euploid replicate is below the count threshold of 50 and the other replicate is above. Genome wide we find this happens rarely (0.3%) however, this skews strongly by size as can be seen when categorizing the false negative rates by CDS size (**SFig 7**). However, no 100 nucleotide size category has a false negative rate over 0.041.

Proportional covariation of insert frequencies between genes.

The insertion frequency of a given gene may significantly deviate from expectations (significant outliers, **STable 14**) because of compensatory changes in other genes. To determine if this is the case we looked for genes with high rates of proportional correlation across strains, for example a 2-fold increase in one gene sees a 2-fold change in the other, and this proportionality is consistent across strains. We performed this check across all genes that were significant outliers in frequency of insertions. We empirically derived an estimated FDR by calculating the background frequency using genes that were not significant outliers. While we identify a small number in each strain (**STable 15**) this was never above the rate expected at random.

RNA sequencing

For RNA sequencing, we grew overnight cultures from three replicate colonies per strain in 5 mL YPGal, then 2 mL (euploid, ComTrip) or 5 mL (other strains) of overnight culture was pelleted and subsequently resuspended in 5 mL fresh YPGal. The cultures were allowed to grow for three hours in fresh YPGal before harvesting cells by vacuum filtration and fixing immediately in liquid nitrogen, so that all cultures were harvested while cells were proliferating. RNA was extracted and purified using a hot acid phenol/chloroform and Phase Lock Gels as described in (Neymotin et al. 2014). Samples were enriched for polyadenylated RNA using the Lexogen Poly(A) RNA Selection Kit V1.5 (cat. # 157.96) and stranded RNAseq libraries were prepared using the Lexogen CORALL Total RNA-Seq Library Prep Kit (cat. # 095.96) according to the manufacturer's protocol. The libraries were pooled at equimolar concentrations, and sequenced using a paired-end (2 x 150) protocol on an Illumina NextSeq 500. The resulting FASTQs were trimmed, aligned, and UMI deduplicated, and coverage per feature was calculated using an in-house pipeline which can be found at https://greshamlab.bio.nyu.edu/wp-content/uploads/2021/11/Windchime_pipeline.nb_.html. Coverage per feature correlation between replicates was high, with the exception of one replicate of ComQuad, which was excluded from further analysis (**STable 8**). Trip1 and ComTrip also only had two replicates, as library preparation failed for one replicate in each.

Effect of amplification on mRNA abundance of transcription factors and targets

To determine what effect the amplification of transcription factors (TF) had on each strain we first used YeastTract (Feng et al. 2014; Teixeira et al. 2023) and SGD (Feng et al. 2014; Teixeira et al. 2023; Cherry et al. 1998) to identify all amplified TFs and their regulatory targets (both manually curated and high-throughput identified (**STable 7**)). We used DESeq2 on the observed (ie. not copy-number corrected) transcript abundances of the TFs to determine if they had significantly higher transcript abundance than the ancestor (DESeq2, p-value <= 0.05).

Most CNV amplified TFs also had significantly different transcript abundances (61%), however many of these TFs that were not associated with CNVs also had significantly different transcript abundances leading to a large disconnect between the two groups (median Jaccard score = 0.32).

To determine if the amplification of TFs led to significant differences in the gene expression of TF targets, we used DESeq2 on the expected (ie. copy-number corrected) transcript abundances of the TF targets to determine if they were significantly different in transcript abundance, relative to the ancestor (DESeq2, adj. p-value <=0.05). We binned these targets as being associated with CNV amplified TFs or not, and significantly differentially expressed or not, and then used Fisher's Exact Test to evaluate if there was a significant contingency between these categories. We found only one instance, *IXR1* in the Aneu strain, where an increase in TF copy number had a significant enrichment (FET, 1.67-fold higher, p-value = 2×10^{-4}) in significantly differentially expressed targets. This same procedure was carried out using TFs with significantly different transcript abundances (DESeq2, p-value <= 0.05) instead of copy-number. Using this approach we found additional instances of agreement, such as *IXR1* in Trip2, Trip1, Sup, Aneu (FET, 2.28-fold higher, p-value = 2×10^{-12}), *MSN4* in Trip1 and ComQuad (FET, 2.0-fold higher, p-value 5×10^{-5}), and *ABF1* in Trip1, Sup, Aneu (FET, 1.53-fold higher, p-value 1×10^{-3}). Taken together, this suggests that evolved strains with significantly different mRNA abundances of transcription factors can exhibit increased numbers of significantly differentially expressed targets - but that increased TF copy-numbers do not tightly correspond with increased TF mRNA abundances.

Calculation of length normalized copy-number corrected insertion frequencies

In order to compare insertion frequencies between genes, we first must normalize the insertions by gene length, to prevent bias by short length genes we apply a minimum gene length threshold of 500 nucleotides. To compare between strains with different copy-numbers we further divide this by the copy-number of the gene in the relevant background. This length

normalized copy-number corrected value is calculated for all genes in all backgrounds and then used to derive a global z-score. Z-scores with an absolute value greater than 2.58 are considered statistically significant (**STable 17**).

Defining copy number effects on gene expression.

All CNV strains have undergone adaptation to glutamine-limited conditions over hundreds of generations (Lauer et al. 2018). As such, the expression of CNV associated genes may differ from the Euploid ancestor due to increased copy number or as a consequence of other heritable variation. To separate changes in transcript abundance due to the gene being amplified within the CNV (i.e. “direct”) from the effect of other variation (i.e. “indirect”) we classified each gene in each strain as being either CNV amplified or not (**SFig 18**). We corrected TPM normalized gene expression for the appropriate copy number and calculated the ratio to the median Euploid ancestor expression. The CNV and non-CNV distributions were compared using a Mann-Whitney U (MWU) test.

Comparison of gene transcript abundances to previous aneuploid stress response studies

Previous studies (Torres et al. 2007; Terhorst et al. 2020) found large aneuploid amplifications significantly correlated with reduced growth rate and stronger yeast environmental stress responses (ESR) (Gasch et al. 2000). In order to compare the results of (Torres et al. 2007; Terhorst et al. 2020) to our own, we first subset the ESR genes from both our datasets (798 of 868, for which we had complete data). We found 41 ESR genes amplified in the Aneu strain, 12 were amplified in at least one CNV, and only one was found to be amplified in all the CNV strains, *FMP46*, which encodes a mitochondria associated protein with no known biological function. Because Torres et al. 2007 used strains with aneuploid amplifications for nearly all chromosomes and may also contain additional smaller CNVs and structural rearrangements, we

did not perform copy number correction for any strain. We calculated the log_2 fold-change in mRNA for each of our evolved strains relative to our euploid ancestor and calculated the mean log_2 fold-change in mRNA for each aneuploid strain from Torres et al. 2007 relative to their euploid strain. We then calculated the Pearson coefficient using the log_2 fold-changes observed in this study in relation to the mean log_2 fold-change reported by Torres et al. Notably, we found these data often showed a negative correlation. Because ESR had previously been shown to be more pronounced in slower growing strains we also evaluated how the growth rates of our strains structured the correlation with the Torres data. We found that our slowest growing strains had the largest anti-correlation with the Torres data.

A similar approach was performed for Taise et al. 2019 which had found aneuploidy inducing a hypo-osmotic like stress response involving 222 genes they termed the common aneuploidy gene-expression or CAGE genes. We found 12 CAGE genes were amplified in our Aneu strain, 3 were amplified in at least one CNV strain, and none were present in all CNV strains. Similar to the process described for Torres et al 2007, we subset the CAGE genes (215 of the 222, we had complete data for) in both sets and then compared the log_2 fold-change of mRNA abundance of each of our evolved strains relative to the euploid ancestor and the log_2 fold-change of mRNA abundance of each Tsai et al. aneuploid relative to their euploid ancestor. We then calculated the Pearson coefficient for each of our evolved strains relative to the Tsai data. We found weak, but positive correlations, between the two sets. Evaluating these in regards to growth rate suggests that the Pearson coefficient of the sets is independent of the growth rates.

Gene copy number determination and transcript abundance copy number correction.

The determination of copy number for each gene in each strain (**STable 6**) was performed using the reconstructed CNV topologies using hybrid long-read and short read sequencing (Spealman

et al. 2022);(Spealman et al. 2023), ODIRA containing CNVs were resolved as described previously (Spealman et al. 2020).

These copy numbers were then used to make an expected mRNA abundance estimate, or copy number corrected estimate. In order to evaluate dosage compensation of CNVs we sought to have an accurate null model. This expected expression model assumes no dosage compensation, and as such, the expected expression of a CNV associated gene would be equal to the euploid expression multiplied by however many copies of the gene are present in any given strain (**STable 24**). The difference between the observed and expected expression can then be evaluated using DESeq2 (**STable 25**), as described above. In the event of CNV dosage compensation one would expect the observed value to be significantly less than the expected value.

Supplemental Files:

All Supplemental materials are publicly available here:

https://github.com/pspealman/CNV_essentiality/tree/main/Supplemental_Files/

Supplemental_File_S1.zip - Modified reference genome for Euploid ancestor (DGY1657) and evolved strains
Modified reference FASTA and GFF files with the Gresham GFP Reporter added to the appropriate coordinates on ChrXI.

Supplemental_File_S2.zip - Collection of custom source code used in this study
Numerous custom scripts were generated for the analyses in this paper. This zip compressed file contains a Code File Description document that details the contents and the data and source code use in the generation of all Figures.