Gene-level heritability analysis explains the polygenic architecture of cancer

Viola Fanfani [*], Luca Citi [†], Adrian L. Harris, [‡]Francesco Pezzella, [§]
Giovanni Stracquadanio [¶]

April 5, 2019

Abstract

It is unknown how high-frequency single nucleotide polymorphisms (SNPs) affect the risk of developing cancer in the broader population. Understanding the contribution of these variants is crucial to use genetic information in risk assessment. More importantly, identifying genes harbouring SNPs explaining a large proportion of the heritable risk could provide new insights in the aetiology of this disease and opportunities for treatment of patients. We addressed these questions by developing a new method to estimate genome-wide and gene-level cancer heritability. Using data from the UK Biobank, we identified a new class of $1,430$ genes, which we called cancer heritability genes, whose SNPs account for a significant contribution to the heritable risk of 35 malignancies. Functional analysis and comparison with cancer genomics data showed that cancer heritability genes are mostly tumour suppressors and control cell proliferation, invasion and metastasis. Our study suggests that SNPs in cancer heritability genes could provide a selective advantage for tumour proliferation and metastasis, thus representing a potential target for personalised treatment.

Decades of research have shown that inherited genomic mutations affect the risk of individuals to develop cancer [28, 25]. In familial studies, inherited mutations in cancer susceptibility genes, such as *TP53* [18], *BRCA1* and *BRCA2* [21, 33], confer up to an 8-fold increase in cancer risk in first degree relatives [28]. However, this evidence is limited to rare, highly penetrant inherited mutations, which explain only a small fraction of the relative risk for all cancers [19].

It has been hypothesised that cancer heritability could be apportioned to high-frequency low-penetrant variants, such as single nucleotide polymorphism (SNPs), which are co-inherited and synergistically modulate the risk of cancer in the broader population. Genome-Wide Association Studies (GWAS) have been instrumental in discovering SNPs associated with an increased risk of many diseases, including many types of cancers, such as breast [8, 30, 11], prostate [29, 9], testicular[16, 32, 17], chronic lymphocytic leukaemia [7, 27, 15], acute lymphocytic leukaemia [31,

[*]Institute of Quantitative Biology, Biochemistry, and Biotechnology, SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, UK

[†]School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom

[‡]Molecular Oncology Laboratories, Department of Oncology, The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK.

[§]Nuffield Department of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford, UK

[¶]Institute of Quantitative Biology, Biochemistry, and Biotechnology, SynthSys, School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3BF, UK. Corresponding author.

23] and several lymphomas [6, 10]. However, less than thousand cancer susceptibility SNPs have been identified, each accounting for a limited increase in cancer risk (odds-ratio, $OR < 1.5$). In general, the vast majority of SNPs show only subtle effects and their contribution is usually cancelled by multiple hypotheses correction procedures applied in GWAS analysis [24].

Although most SNPs show only weak, non statistically significant effects, they still contribute to the heritable risk, suggesting that cancer heritability will likely be polygenic; thus, accurate estimates of cancer heritability should take into account the contribution of all common SNPs in the genome. Moreover, since the functional impact of SNPs depends on the region where they are located [26], quantifying the amount of heritability apportioned to SNPs in protein-coding genes and nearby regulatory regions could provide new insights into the aetiology of this disease and identify new intervention points for treatment. Here we developed a new method, called BAyesian Gene HERitability Analysis (BAGHERA), which implements a new hierarchical Bayesian extension of Linkage Disequilibrium (LD) score regression [3], to obtain estimates of the heritability explained by all genotyped SNPs (genome-wide heritability) and SNPs in or around protein coding genes (gene-level heritability) from summary statistics.

We used BAGHERA to analyse 35 self-reported cancers in the UK Biobank [4], a large-scale prospective study aiming at systematically screening and phenotyping more than $500,000$ individuals. In our study, we used available UK Biobank summary statistics for $N = 337,159$ patients [22]; each cancer has 798 cases on average and a relative prevalence ranging from $< 0.1\%$ to 2%, with higher estimates for common malignancies in European populations, such as breast and prostate cancer [2]. A custom pipeline was developed for pre-processing summary statistics to apply filters for SNP quality control, assign LD scores and annotate SNPs to genes, with overlapping genes merged into non-overlapping genic regions (see Online Methods).

We then estimated the genome-wide heritability of each cancer by computing the median of the posterior distribution of h^2_{SNP}, and transforming this value on to the liability scale, $h^2_{SNP_L}$, to obtain estimates independent from prevalence and comparable across malignancies. We observed heritability values ranging from 4% for ovarian cancer up to 44% for adnexal cancer (Supplementary Table 1). Overall, 22 out of 35 cancers show heritability higher than 10%, consistent with other recent estimates and suggesting a significant contribution of SNPs to the heritable risk of cancer [25]. It is important to note that BAGHERA provided accurate estimates even for low-heritability cancers, whereas LD score regression returns negative values (see Supplementary Materials).

We then identified heritability genes for each cancer, defined as those with per-SNP heritability higher than the genome-wide estimate (see Online Methods). We found 13 heritability genes per malignancy on average, ranging from 2 genes for Non-Hodgkin's lymphoma to 246 for breast cancer (Figure 1A and Supplementary Table 2). Gene-level heritability across the 35 cancers has a long-tail distribution (Figure 1B), with a median 22.6-fold change compared to the genome-wide estimate, ranging from 4.5-fold for *SDK1* in breast cancer to 374.1-fold for *KITLG* in testicular cancer. Interestingly, 95% of heritability genes show heritability 10-fold higher than the genome-wide estimate. Heritability genes represent less than 1% of all the genes in the genome across all cancers; this result is consistent with cancer heritability being polygenic. However, heritability genes account for up to 43% of all the heritable risk (breast cancer), suggesting that a small set of genes has

the largest contribution to cancer heritability (Figure 1A). Consistent with our hypotheses, when we looked at the contribution of SNPs outside our protein-coding regions, we did not observe any difference compared to the genome-wide estimate. We then tested whether higher heritability could be explained by the presence of genome-wide significant SNPs ($P < 5 \times 10^{-8}$) nearby protein-coding regions. For each cancer, we identified genes harbouring at least 1 genome-wide significant SNP, and denoted this set as minSNPs; significant SNPs falling outside genic regions are assigned to the non-coding nuisance gene, for consistency with our analysis. We found 67 (56 in coding regions) minSNPs in total, with at least 1 minSNP in only 8 of the 35 cancers (Supplementary Table 3). This is a striking difference compared to the 13 heritability genes found on average in our analysis (Figure 1C and Supplementary Table 3); interestingly, our method was able to recover 47 (70%) of the minSNPs (84% of the coding ones), suggesting that it can detect heritability genes regardless of the association strength of their SNPs. Taken together, our analysis found 976 non-overlapping genic regions, which identify $1,430$ genes in the genome, having a significant contribution to the heritable risk in at least 1 cancer. We denoted these $1,430$ genes as Cancer Heritability Genes (CHGs).

We then characterised the functional role of CHGs to understand whether they are involved in biological processes and molecular functions mediating cancer phenotypes. We performed gene ontology enrichment analysis, using the slim gene ontology for human, finding $1,324$ out of $1,430$ CHGs to be annotated with at least one term. We found a statistically significant enrichment for 10 terms (False Discovery Rate, FDR $< 10\%$, Figure 2A and Supplementary Table 4). CHGs are enriched in cell morphogenesis ($OR : 1.62$ $P : 8.76 \times 10^{-6}$), cell division ($OR : 1.59,\ P : 2.78 \times 10^{-4}$) and cell adhesion ($OR : 1.35,\ P : 1.23 \times 10^{-3}$), which are biological processes underpinning tumor proliferation, invasion and metastasis; these classes include heritability genes such as *EGFR*, *FGFR2*, *SMAD3* and *NF1*. We also observed a significant enrichment of genes associated with cytoskeleton organization ($OR : 1.33,\ P : 1.64 \times 10^{-3}$) and anatomical structure development ($OR : 1.19,\ P : 2.68 \times 10^{-3}$), which includes members of the SWI/SNF complex, such as *ARID1B*. We further validated this finding by testing whether cancer heritability genes are associated with any hallmark of cancer [12]. Interestingly, we found 30 CHGs associated with at least one hallmark ($OR : 1.98,\ P : 1.1 \times 10^{-3}$); in particular, invasion, metastasis, growth and cell death are controlled by cancer heritability genes, such as *APC*, *CDKN2A*, *EGFR*, *ESR1*, *SMAD2* and *SMAD3* (2C and Supplementary Table 5). Taken together, we demonstrated that CHGs are directly implicated in biological processes underpinning tumorigenesis.

We then studied whether CHGs mediate cancer processes by acting either as tumour suppressor genes (TSGs) or oncogenes. To do that, we used the Precision Oncology Knowledge Base (OncoKB, [5]), an expert-curated list of 519 genes, including 197 tumour suppressor genes (TSGs), 148 oncogenes and other cancer genes of unknown function. We found that 58 CHGs (4%) are also cancer genes (OR: 10.3, $P : 0.0012$). Specifically, 30 CHGs are tumour suppressors (OR: 1.83, $P : 0.024$), whereas only 14 are reported as oncogene (OR: 0.66, $P : 0.92$) (Figure 2C). Tumour suppressor CHGs include well-known cancer driver genes, such as *CDKN2A*, *CDKN2B* and *CDKN2C*, which regulate cell growth, and DNA repair genes, such as *MUTYH* and *FANCA*. Our results suggest a functional role for cancer heritability genes consistent with a two-hit model [14]; while inherited mutations associated with oncogene activation are likely to be under purifying selection, mutations in tumour suppressor genes can be observed at higher frequency because

deleterious effects are only observed upon complete loss of function. We hypothesised that SNPs impairing the function of tumour suppressor genes could provide a selective advantage for tumour development.

Since CHGs are predominantly involved in common cell development functions, they may be thought unlikely to be specific for a particular cancer. Thus, we estimated the amount of heritability explained by CHGs across all cancers to test for pleiotropic effects. We observed that CHGs explain 22% of heritability on average, ranging from 43% for breast cancer to 2% for leukaemia (Figure 3A). However, the contribution of the heritability genes specific to each cancer was consistently higher than the CHGs contribution, suggesting the lack of pleiotropic effects. This observation is consistent with results from tumour sequencing studies, which have shown that most genes are mutated in only one type of cancer [1] and pleiotropic effects are limited to few master regulators, such as *TP53*, *PIK3CA* or potent oncogenes, such as *KRAS*. Analogously, of the $1,430$ CHGs, $\approx 15.7\%$ show a significant heritability enrichment in at least 2 cancers, and only $\approx 2\%$ in 3 or more (Figure 3B-C and Supplementary Table 5). Our analysis suggests that while cancer heritability is controlled by mechanisms underpinning all cancers, the specific contribution of each gene is cancer specific; this could be explained by the fact that the same molecular function is accomplished by different genes in a tissue-specific way.

Our study provides new fundamental evidence demonstrating a strong contribution of high-frequency inherited mutations to the heritable risk of cancer. To do that, we developed a new method, called Bayesian Gene Heritability Regression Analysis (BAGHERA), which enabled the identification of $1,430$ genes with a significant contribution to the heritable risk of 35 malignancies. We showed that these loci are tumor suppressors controlling cell development, adhesion and motility, which are fundamental processes required for tumorigenesis. Ultimately, our results support a two-hit model, where inherited mutations in tumour suppressor genes could create a favourable genetic background for tumorigenesis. It is therefore conceivable that SNPs make normal cells more likely to evade the cell-cell contact inhibition of proliferation, to elude the anatomical constrains of their tissue and to achieve more easily independent motility in presence of other early oncogenic events. Preliminary support for this model has been provided by studies in hereditary diffuse gastric cancer (HDGC) [20]; specifically, a germline mutation causing loss of E-Cadherin function, a crucial gene for cell adhesion, has been found to be a causal factor in up to 50% of all HDGC patients [13].

We also recognise the limitations of our work; while our method provides accurate estimates of genome-wide heritability, extremely low heritability diseases could lead to negative gene-level heritability estimates; this was a trade-off to ensure reasonable computational efficiency, although a rigorous model is provided as part of our software. Our analysis does not incorporate functional information, such as gene expression, which limits our power of detecting tissue-specific contributions. On this point, as the genes may be expressed in different cellular compartments, they may contribute to the stromal niches in which cancers develop and their role in tissue specificity of mutations will be of interest to analyse experimentally.

Taken together, our study provides a new view of the genetic architecture of cancer with gene-level resolution. We anticipate that the availability of genome editing techniques will enable testing of the functional mechanisms mediated by cancer heritability genes. We also expect that integrating our results with tumour sequencing data will provide new venues for personalised treatment and patients stratification.

References

[1] Matthew H Bailey et al. "Comprehensive Characterization of Cancer Driver Genes and Mutations." In: *Cell* 173.2 (Apr. 2018), 371–385.e18. ISSN: 1097-4172. DOI: 10.1016/j.cell.2018.02.060. URL: http://www.ncbi.nlm.nih.gov/pubmed/29625053%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6029450.

[2] Freddie Bray et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." In: *CA: a cancer journal for clinicians* (Sept. 2018). ISSN: 1542-4863. DOI: 10.3322/caac.21492. URL: http://www.ncbi.nlm.nih.gov/pubmed/30207593.

[3] Brendan K Bulik-Sullivan et al. "LD Score regression distinguishes confounding from polygenicity in genome-wide association studies". In: *Nature genetics* 47.3 (2015), p. 291.

[4] Clare Bycroft et al. "The UK Biobank resource with deep phenotyping and genomic data". In: *Nature* 562.7726 (2018), pp. 203–209. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0579-z. URL: http://www.nature.com/articles/s41586-018-0579-z.

[5] Debyani Chakravarty et al. "OncoKB: a precision oncology knowledge base". In: *JCO precision oncology* 1 (2017), pp. 1–16.

[6] W. Cozen et al. "A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus". In: *Nature Communications* 5.1 (Dec. 2014), p. 3856. ISSN: 2041-1723. DOI: 10.1038/ncomms4856. URL: http://www.nature.com/articles/ncomms4856.

[7] Maria Chiara Di Bernardo et al. "A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia". In: *Nature Genetics* 40.10 (Oct. 2008), pp. 1204–1210. ISSN: 1061-4036. DOI: 10.1038/ng.219. URL: http://www.nature.com/articles/ng.219.

[8] Douglas F. Easton et al. "Genome-wide association study identifies novel breast cancer susceptibility loci". In: *Nature* 447.7148 (June 2007), pp. 1087–1093. ISSN: 0028-0836. DOI: 10.1038/nature05887. URL: http://www.nature.com/doifinder/10.1038/nature05887.

[9] Rosalind A Eeles et al. "Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array". In: *Nature Genetics* 45.4 (Apr. 2013), pp. 385–391. ISSN: 1061-4036. DOI: 10.1038/ng.2560. URL: http://www.nature.com/articles/ng.2560.

[10] Victor Enciso-Mora et al. "A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3)". In: *Nature Genetics* 42.12 (Dec. 2010), pp. 1126–1130. ISSN: 1061-4036. DOI: 10.1038/ng.696. URL: http://www.nature.com/articles/ng.696.

[11] Maya Ghoussaini et al. "Genome-wide association analysis identifies three new breast cancer susceptibility loci". In: *Nature Genetics* 44.3 (Mar. 2012), pp. 312–318. ISSN: 1061-4036. DOI: 10.1038/ng.1049. URL: http://www.nature.com/articles/ng.1049.

[12] Douglas Hanahan and Robert A Weinberg. "Hallmarks of cancer: the next generation". In: *cell* 144.5 (2011), pp. 646–674.

[13] Samantha Hansford et al. “Hereditary diffuse gastric cancer syndrome: CDH1 mutations and beyond”. In: *JAMA oncology* 1.1 (2015), pp. 23–32.

[14] Alfred G Knudson. “Mutation and cancer: statistical study of retinoblastoma”. In: *Proceedings of the National Academy of Sciences* 68.4 (1971), pp. 820–823.

[15] Philip J. Law et al. “Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia”. In: *Nature Communications* 8 (Feb. 2017), p. 14175. ISSN: 2041-1723. DOI: 10.1038/ncomms14175. URL: http://www.nature.com/doifinder/10.1038/ncomms14175.

[16] Kevin Litchfield et al. “Identification of 19 new risk loci and potential regulatory mechanisms influencing susceptibility to testicular germ cell tumor”. In: *Nature Genetics* 49.7 (2017), pp. 1133–1140. ISSN: 15461718. DOI: 10.1038/ng.3896.

[17] Kevin Litchfield et al. “The genomic landscape of testicular germ cell tumours: From susceptibility to treatment”. In: *Nature Reviews Urology* 13.7 (2016), pp. 409–419. ISSN: 17594820. DOI: 10.1038/nrurol.2016.107.

[18] D Malkin et al. “Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms.” In: *Science (New York, N.Y.)* 250.4985 (Nov. 1990), pp. 1233–8. ISSN: 0036-8075. URL: http://www.ncbi.nlm.nih.gov/pubmed/1978757.

[19] Teri A. Manolio et al. “Finding the missing heritability of complex diseases”. In: *Nature* 461.7265 (2009), pp. 747–753. ISSN: 00280836. DOI: 10.1038/nature08494. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1038/nature08494.

[20] Alisha M Mendonsa, Tae-Young Na, and Barry M Gumbiner. “E-cadherin in contact inhibition and cancer”. In: *Oncogene* (2018), p. 1.

[21] Y Miki et al. “A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1”. In: *Science* 266.5182 (Oct. 1994), pp. 66–71. ISSN: 0036-8075. DOI: 10.1126/science.7545954. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.7545954.

[22] Benjamin Neele. *UKBB Heritability*. URL: https://nealelab.github.io/UKBB_ldsc/.

[23] L Orsi et al. “Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE)”. In: *Leukemia* 26.12 (Dec. 2012), pp. 2561–2564. ISSN: 0887-6924. DOI: 10.1038/leu.2012.148. URL: http://www.nature.com/articles/leu2012148.

[24] Bogdan Pasaniuc and Alkes L. Price. “Dissecting the genetics of complex traits using summary association statistics”. In: *Nature Reviews Genetics* 18.2 (Nov. 2017), pp. 117–127. ISSN: 14710064. DOI: 10.1038/nrg.2016.142. arXiv: 15334406.

[25] Joshua N Sampson et al. “Analysis of Heritability and Shared Heritability Based on Genome-Wide Association Studies for Thirteen Cancer Types.” In: *Journal of the National Cancer Institute* 107.12 (Dec. 2015), djv279. ISSN: 1460-2105. DOI: 10.1093/jnci/djv279. URL: http://www.ncbi.nlm.nih.gov/pubmed/26464424%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4806328.

[26] Andrew J Schork et al. “All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs”. In: *PLoS genetics* 9.4 (2013), e1003449.

[27] Helen E Speedy et al. “A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia”. In: *Nature Genetics* 46.1 (Jan. 2014), pp. 56–60. ISSN: 1061-4036. DOI: 10.1038/ng.2843. URL: http://www.nature.com/articles/ng.2843.

[28] Amit Sud, Ben Kinnersley, and Richard S. Houlston. “Genome-wide association studies of cancer: current insights and future perspectives”. In: *Nature reviews. Cancer* 17.11 (Nov. 2017), pp. 692–704. ISSN: 14741768. DOI: 10.1038/nrc.2017.82. URL: https://www.nature.com/articles/nrc.2017.82.

[29] Gilles Thomas et al. “Multiple loci identified in a genome-wide association study of prostate cancer”. In: *Nature Genetics* 40.3 (Mar. 2008), pp. 310–315. ISSN: 1061-4036. DOI: 10.1038/ng.91. URL: http://www.nature.com/articles/ng.91.

[30] Clare Turnbull et al. “Genome-wide association study identifies five new breast cancer susceptibility loci”. In: *Nature Genetics* 42.6 (June 2010), pp. 504–507. ISSN: 1061-4036. DOI: 10.1038/ng.586. URL: http://www.nature.com/articles/ng.586.

[31] Jayaram Vijayakrishnan et al. “Genome-wide association study identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia”. In: *Nature Communications* 9.1 (Dec. 2018), p. 1340. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03178-z. URL: http://www.nature.com/articles/s41467-018-03178-z.

[32] Zhaoming Wang et al. “Meta-analysis of five genome-wide association studies identifies multiple new loci associated with testicular germ cell tumor”. In: *Nature Genetics* 49.7 (2017), pp. 1141–1147. ISSN: 1061-4036. DOI: 10.1038/ng.3879. URL: http://www.nature.com/doifinder/10.1038/ng.3879.

[33] R Wooster et al. “Identification of the breast cancer susceptibility gene BRCA2.” In: *Nature* 378.6559 (1995), pp. 789–92. ISSN: 0028-0836. DOI: 10.1038/378789a0. URL: http://www.ncbi.nlm.nih.gov/pubmed/8524414.

Figures

Figure 1: **Heritability genes across 35 cancers in the UK Biobank**. A) For each malignancy we report the observed heritability (h^2_{SNP}, blue boxes), the percentage of h^2_{SNP} explained by heritability genes (red barplot) and the number of heritability genes (blue barplot). B) Gene-level heritability distribution across heritability genes, expressed as fold-change with respect to the genome-wide estimate, including the top 5 genes observed and the one with the lowest fold-change. C) Comparison of heritability genes (HGs) and genes harbouring genome-wide significant SNPs (minSNPs). For each cancer, we report the number of cancer heritability genes (blue) and minSNPs (red), and those that are both HGs and minSNPs. For those cancers with at least one minSNP, we show the number of minSNPs captured in our analysis. With non-coding minSNP we report if the genome-wide significant SNPs are in the non coding areas of the genome.

Figure 2: **Cancer heritability genes are associated with the hallmarks of cancer**. A) Gene ontology enrichment analysis of cancer heritability genes (CHGs). For each GO term, we report the odds ratio and its false discovery rate (FDR). Terms significant at the 10% FDR are reported in red, whereas non-significant ones in blue. B) Cancer heritability genes are associated with the hallmarks of cancer. C) Number of CHGs reported as cancer genes in the Precision Oncology Knowledge Base (OncoKB). For each cancer, we report cancer heritability genes annotated as cancer genes in red and those with uncharacterized function in blue. The inset plot shows the number of CHGs reported as tumour suppressors (TSG) or oncogenes.

Figure 3: **A subset of cancer heritability genes are associated with multiple cancers**. A) Percentage of heritability explained by all CHGs, with the red line representing the expected level of heritability explained by the genes assuming equal per SNP heritability contribution. B) Number of cancer heritability genes associated with multiple cancers, with more than 10% of CHGs being common to multiple malignancies. C) Cancer heritability genes associated with at least 2 malignancies.

Figure 1

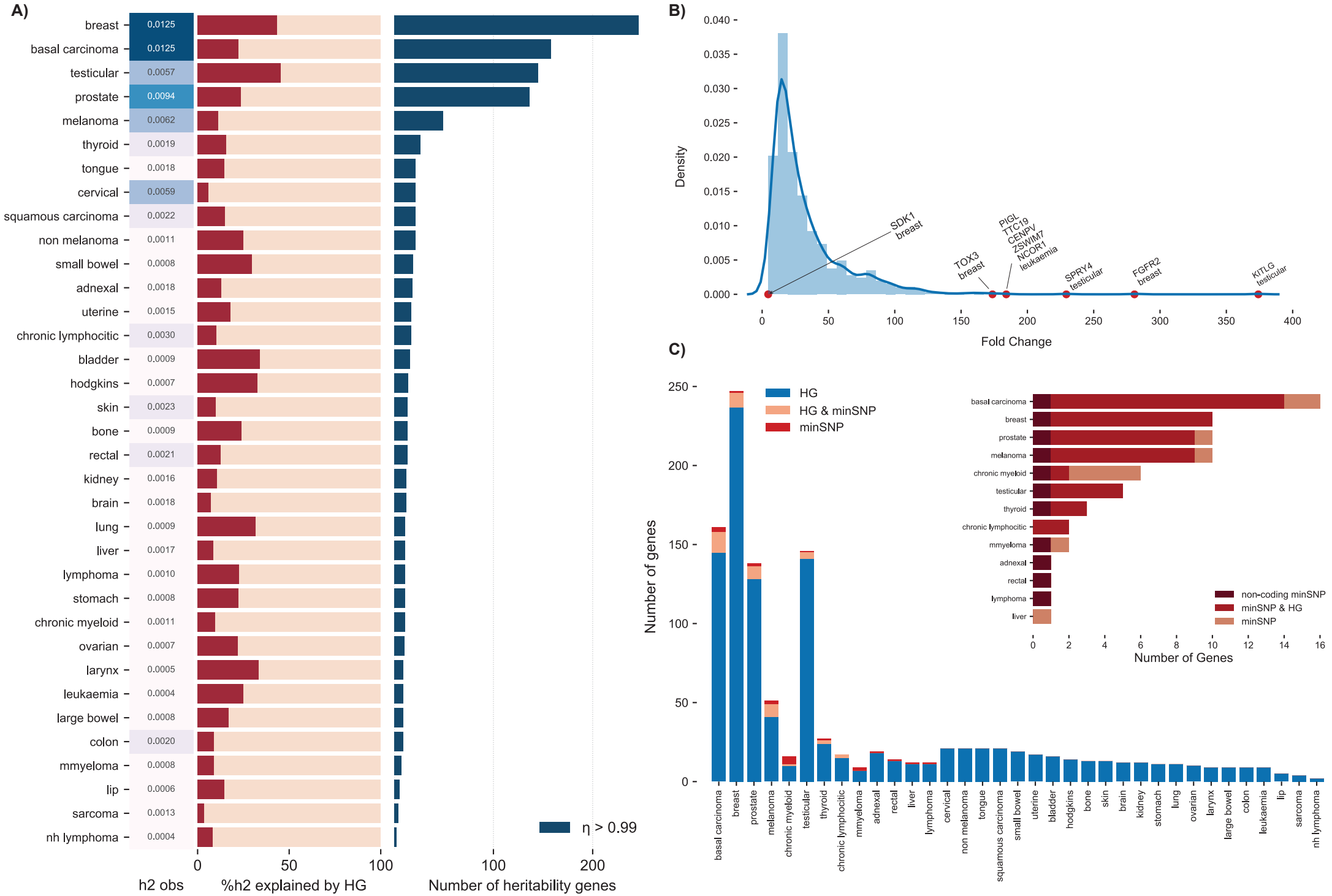

A)
breast 0.0125
basal carcinoma 0.0125
testicular 0.0057
prostate 0.0094
melanoma 0.0062
thyroid 0.0019
tongue 0.0018
cervical 0.0059
squamous carcinoma 0.0022
non melanoma 0.0011
small bowel 0.0008
adnexal 0.0018
uterine 0.0015
chronic lymphocitic 0.0030
bladder 0.0009
hodgkins 0.0007
skin 0.0023
bone 0.0009
rectal 0.0021
kidney 0.0016
brain 0.0018
lung 0.0009
liver 0.0017
lymphoma 0.0010
stomach 0.0008
chronic myeloid 0.0011
ovarian 0.0007
larynx 0.0005
leukaemia 0.0004
large bowel 0.0008
colon 0.0020
mmyeloma 0.0008
lip 0.0006
sarcoma 0.0013
nh lymphoma 0.0004
h2 obs
%h2 explained by HG
Number of heritability genes
η > 0.99
B)
Density
Fold Change
SDK1 breast
TOX3 breast
PIGL TTC19 CENPV ZSWIM7 NCOR1 leukaemia
SPRY4 testicular
FGFR2 breast
KITLG testicular
C)
HG
HG & minSNP
minSNP
Number of genes
non-coding minSNP
minSNP & HG
minSNP
Number of Genes

Figure 2

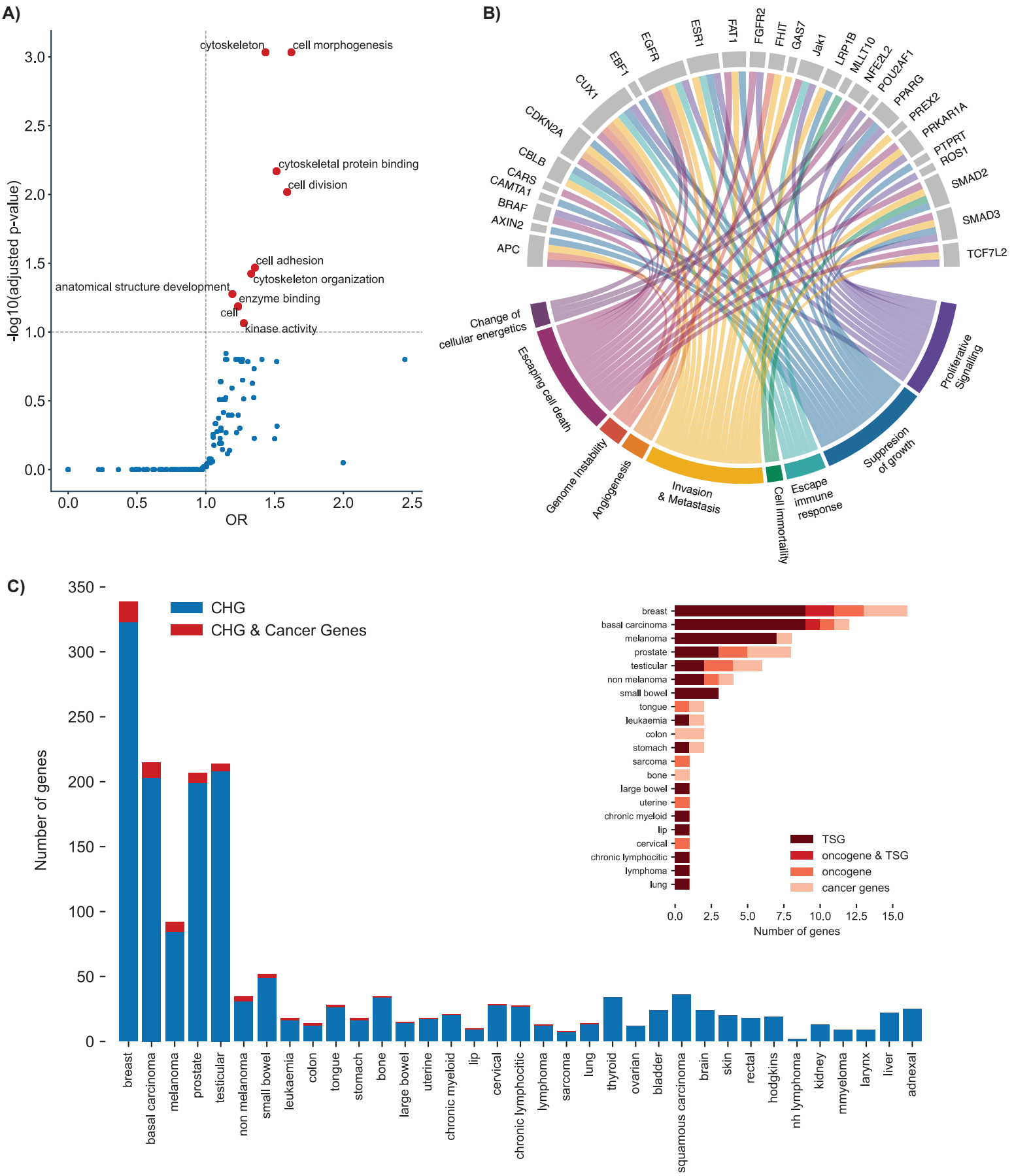
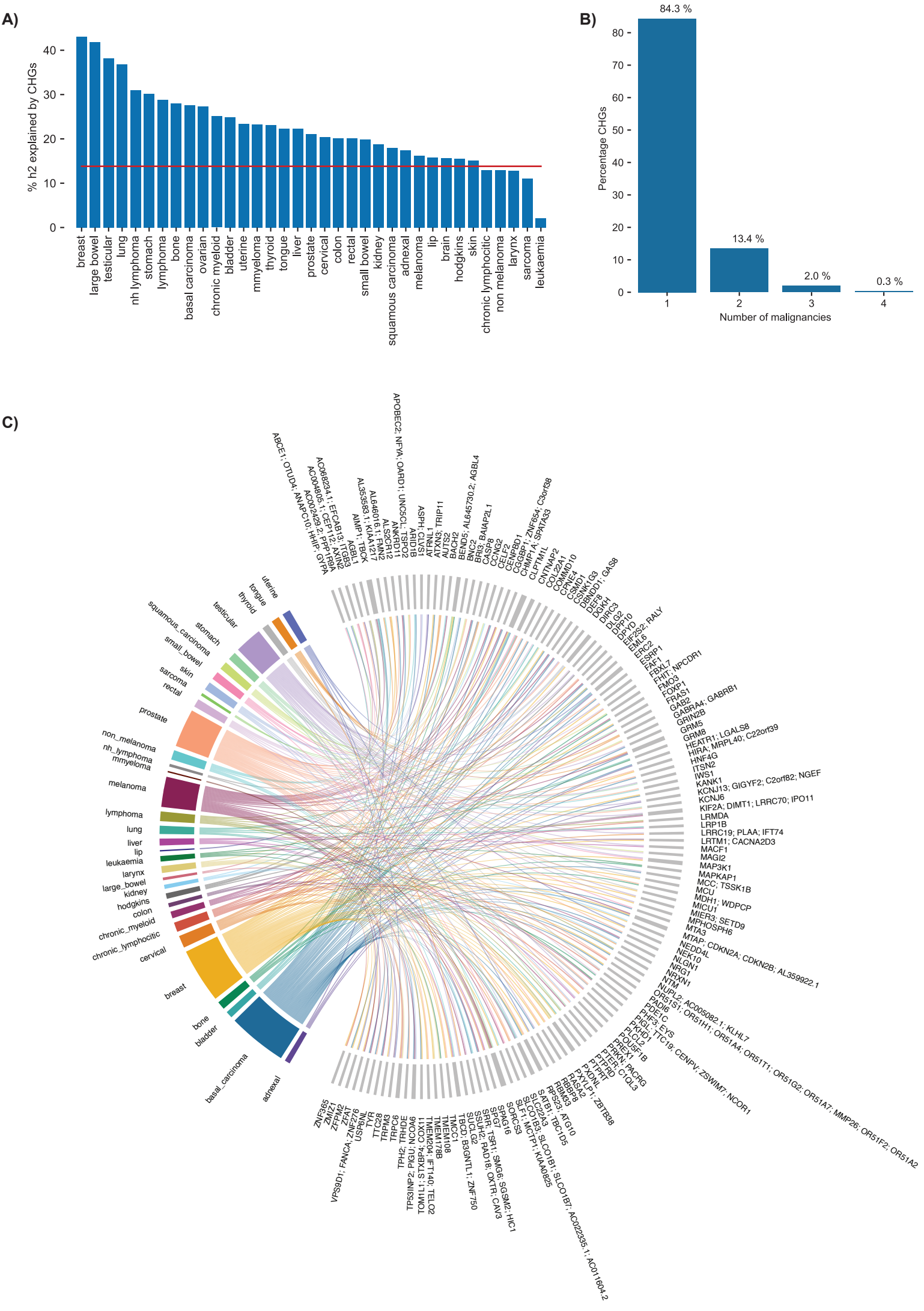

A)
cytoskeleton
cell morphogenesis
cytoskeletal protein binding
cell division
cell adhesion
cytoskeleton organization
anatomical structure development
enzyme binding
cell
kinase activity
-log10(adjusted p-value)
OR
B)
EGFR
ESR1
FAT1
FGFR2
FHIT
GAS7
Jak1
LRP1B
MLLT10
NFE2L2
POU2AF1
PPARG
PREX2
PRKAR1A
PTPRT
ROS1
SMAD2
SMAD3
TCF7L2
EBF1
CUX1
CDKN2A
CBLB
CARS
CAMTA1
BRAF
AXIN2
APC
Change of cellular energetics
Escaping cell death
Genome Instability
Angiogenesis
Invasion & Metastasis
Cell immortality
Escape immune response
Suppresion of growth
Proliferative Signalling
C)
CHG
CHG & Cancer Genes
Number of genes
breast
basal carcinoma
melanoma
prostate
testicular
non melanoma
small bowel
leukaemia
colon
tongue
stomach
bone
large bowel
uterine
chronic myeloid
lip
cervical
chronic lymphocitic
lymphoma
sarcoma
lung
thyroid
ovarian
bladder
squamous carcinoma
brain
skin
rectal
hodgkins
nh lymphoma
kidney
mmyeloma
larynx
liver
adnexal
TSG
oncogene & TSG
oncogene
cancer genes
Number of genes

Figure 3


A)
% h2 explained by CHGs
0
10
20
30
40
breast
large bowel
testicular
lung
nh lymphoma
stomach
lymphoma
bone
basal carcinoma
ovarian
chronic myeloid
bladder
uterine
mmyeloma
thyroid
tongue
liver
prostate
cervical
colon
rectal
small bowel
kidney
squamous carcinoma
adnexal
melanoma
lip
brain
hodgkins
skin
chronic lymphocitic
non melanoma
larynx
sarcoma
leukaemia
B)
Percentage CHGs
0
10
20
30
40
50
60
70
80
84.3 %
13.4 %
2.0 %
0.3 %
1
2
3
4
Number of malignancies
C)
uterine
tongue
thyroid
testicular
stomach
squamous_carcinoma
small_bowel
skin
sarcoma
rectal
prostate
non_melanoma
nh_lymphoma
mmyeloma
melanoma
lymphoma
lung
liver
lip
leukaemia
larynx
large_bowel
kidney
hodgkins
colon
chronic_myeloid
chronic_lymphocitic
cervical
breast
bone
bladder
basal_carcinoma
adnexal