

**Supplementary information**

---

**Sex disparities matter in cancer  
development and therapy**

---

In the format provided by the  
authors and unedited

Supplementary Table 1. **Kyoto Encyclopedia of Genes and Genomes Pathways and Immune signatures enriched among the X and Y chromosome genes as determined by over-representation analysis.**

Pathway or signature	adjusted_p_value	term_size	query_size	inter-section_size	intersections
X chromosome					
Metabolic pathways	1.20E-06	1482	732	46	DGKK, OCRL, MTM1, PRPS2, NDUFB11, PIGA, EBP, ALG13, PFKFB1, TKTL1, ATP6AP1, PCYT1B, RENBP, MTMR8, ACSL4, OTC, SAT1, NSDHL, GUCY2F, G6PD, NDUFA1, PGK1, IDH3G, CA5B, GYG2, GLUD2, MAOB, ASMT, PDHA1, RGN, HSD17B10, CTPS2, PGAM4, C1GALT1C1, HPRT1, ALAS2, PRPS1, MAOA, UPRT, MTMR1, SMS, TMLHE, GK, GLA, COX7B, SUV39H1
Carbon metabolism	7.57E-05	117	732	10	PRPS2, TKTL1, G6PD, PGK1, IDH3G, GLUD2, PDHA1, RGN, PGAM4, PRPS1
Pentose phosphate pathway	0.000999425	30	732	5	PRPS2, TKTL1, G6PD, RGN, PRPS1
Biosynthesis of amino acids	0.001032299	75	732	7	PRPS2, TKTL1, OTC, PGK1, IDH3G, PGAM4, PRPS1
Glycine, serine and threonine metabolism	0.017030845	40	732	4	MAOB, PGAM4, ALAS2, MAOA
Arginine and proline metabolism	0.03044943	49	732	4	SAT1, MAOB, MAOA ,SMS
mRNA surveillance pathway	1.45E-05	91	732	10	GSPT2, CSTF2, UPF3B, PABPC1L2B, PPP2R3B, NXF3, PABPC5, NXT2, NXF2B, NXF5
RNA transport	4.08E-05	159	732	12	EIF2S3, UPF3B, PABPC1L2B, NXF3, FMR1, PABPC5, EIF1AX, NXT2, NXF2B, NXF5, GEMIN8, THOC2
NF-kappa B signaling pathway	0.000999425	100	732	8	TAB3, CD40LG, BTK, IRAK1, EDA, EDA2R, IKBK, XIAP
Cytokine-cytokine receptor interaction	0.006256074	292	732	12	BMP15, IL9R, CRLF2, IL2RG, IL13RA1, IL13RA2, CSF2RA, CXCR3, CD40LG, IL3RA, EDA, EDA2R

JAK-STAT signaling pathway	0.01580168 2	162	732	8	IL9R, FHL1, CRLF2, IL2RG, IL13RA1, IL13RA2, CSF2RA, IL3RA
Influenza A	0.01583581	167	732	8	NXF3, TLR7, SLC25A5, NXT2, IKBKG, NXF2B, NXF5
Primary immunodeficiency	0.01583581	37	732	4	IL2RG, CD40LG, BTK, IKBKG
Herpes simplex virus 1 infection	0.01583581	487	732	15	HCFC1, ZNF630, ZNF81, ZNF674, ZNF41, ZNF182, ZNF275, NXF3, ZNF157, IRAK1, IKBKG, NXF2B, NXF5, ZFP92, CFP
Measles	0.01796710 6	138	732	7	RAB9B, MSN, IL2RG, RAB9A, TLR7, IRAK1, IKBKG
Salmonella infection	0.04300724 9	214	732	8	RAB9B, RAB9A, TAB3, IRAK1, IKBKG, FLNA, DYNLT3, PAK3
Ubiquitin mediated proteolysis	0.01703084 5	135	732	7	UBE2A, UBA1, CUL4B, HUWE1, KLHL13, MID1, XIAP
Regulation of actin cytoskeleton	0.01703084 5	212	732	9	DIAPH2, FGF16, MSN, ARHGEF6, FGD1, TMSB4X, LPAR4, PAK3, ARAF
MAPK signaling pathway	0.03446666 4	295	732	10	FGF16, CACNA1F, ELK1, IRAK1, IKBKG, FLNA, DUSP9, ARAF, RPS6KA6, RPS6KA3
Necroptosis	0.03599150 6	162	732	7	SLC25A6, AIFM1, CYBB, H2AFB1, GLUD2, SLC25A5, XIAP
Pathways in cancer	0.04909849 3	529	732	14	FGF16, PIM2, ELK1, AR, IL2RG, IL13RA1, CSF2RA, IL3RA, COL4A5, COL4A6, IKBKG, LPAR4, XIAP, ARAF
Neuroactive ligand-receptor interaction	5.47E-05	338	732	17	P2RY4, HTR2C, AVPR2, GLRA2, GABRE, AGTR2, BRS3, CYSLTR1, P2RY8, GPR50, GABRA3, LPAR4, GABRQ, APLN, GRIA3, P2RY10, GRPR
Ribosome biogenesis in eukaryotes	0.00119864 2	78	732	7	GNL3L, NXF3, NXT2, NXF2B, NXF5, DKC1, UTP14A
Calcium signaling pathway	0.01297104 4	192	732	9	SLC25A6, PHKA1, CACNA1F, ATP2B3, HTR2C, PHKA2, SLC25A5, CYSLTR1, GRPR
Mineral absorption	0.01379207 9	58	732	5	ATP2B3, ATP1B4, HEPH, S100G, ATP7A
Nicotine addiction	0.01703084 5	40	732	4	GABRE, GABRA3, GABRQ, GRIA3
Retrograde endocannabinoid signaling	0.02557764 7	148	732	7	NDUFB11, CACNA1F, GABRE, NDUFA1, GABRA3, GABRQ, GRIA3

Renin-angiotensin system	0.02831690 4	23	732	3	AGTR2, ATP6AP2, ACE2
GABAergic synapse	0.04275285 2	89	732	5	CACNA1F, SLC38A5, GABRE, GABRA3, GABRQ
Y chromosome					
M2 Macrophages	0.00040416	186	5	5	USP9Y, RPS4Y1, EIF1AY, KDM5D, DX3Y

Foot note:

X chromosome genes correspond to the Figure 3 **Kyoto Encyclopedia of Genes and Genomes** (KEGG) pathways enrichment visualisation as determined by Over-representation analyses (ORA) using the R package gprofiler2 (v0.1.8)<sup>1</sup>. P-values were false discovery rate (FDR) adjusted. Significant gene sets were inferred, using a 0.05 significance level.

Y chromosome genes subjected to ORA identified custom selected immune pathways. ORA was applied using the R package gprofiler2 (v0.1.8). P-values were false discovery rate (FDR) adjusted. Significant gene sets were inferred, using a 0.05 significance level.

## Reference

- 1 Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**, doi:10.12688/f1000research.24956.2 (2020).

Supplementary Table 2: **Among thirteen percent of non-reproductive cancers are associated with Infectious pathogens, the incidence in males is double that of females<sup>1</sup>.**

Cancer burden from infections	Infectious pathogen	Associated cancer risk	References
90%	<i>Helicobacter Pylori</i>	Gastric cancer, non-Hodgkin lymphoma	1
	Hepatitis B virus (HBV)	Hepatocellular carcinoma	1
	Hepatitis C virus (HCV)	Hepatocellular carcinoma, Non-Hodgkin lymphoma	1
	<b>Oncoviruses</b>		
	Epstein-Barr virus (EBV)	Nasopharngeal carcinoma, Non-Hodgkin lymphoma, Burkitt lymphoma	1,2
<10%	Human herpes virus 8 (HHV-8; also known as Kaposi sarcoma-associated herpes virus; KSHV)	Kaposi sarcoma	1,2
	Human T cell leukemia/lymphoma virus type 1 (HTLV-1)	Adult T cell Leukaemia/Lymphoma	1,2
	Merkel cell Polyomavirus (MCP $\gamma$ V)	Merkel cell carcinoma	1,2
	Human immunodeficiency virus (HIV-1)	EBV-associated cancers HHV-8 associated cancers	1,2

**Notes:**

- HTLV-1, EBV and MCP $\gamma$ V are regarded as a direct oncoviruses, as they encode either viral oncoproteins or active host oncoproteins; in contrast to indirect oncoviruses such as HCB and HCV that promote cancer for example by inducing chronic inflammation, or HIV-1 that imposes immune suppression<sup>2</sup>.
- High-risk Human Papilloma Virus (HPV), which accounts for 31.6% of all infection-associated cancers has been excluded from this analysis. HPV was omitted from this table as it contributes to very few non-reproductive cancers, but predominantly drives cancers of reproductive organs<sup>1</sup>.

**References**

- 1 de Martel, C., Georges, D., Bray, F., Ferlay, J. & Clifford, G. M. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob Health* **8**, e180-e190, doi:10.1016/S2214-109X(19)30488-7 (2020).
- 2 Guven-Maiorov, E., Tsai, C. J. & Nussinov, R. Oncoviruses Can Drive Cancer by Rewiring Signaling Pathways Through Interface Mimicry. *Front Oncol* **9**, 1236, doi:10.3389/fonc.2019.01236 (2019).

Supplementary Table 3. **Male and Female Identity of Cell lines in the Broad Institute Cancer Cell Line Encyclopedia<sup>1</sup>**

Cell Lines from females and males

	pleura	liver	upper aerodigestive tract	kidney	cns	lung	oesophagus	large intestine	haematopoietic & lymph	stomach	ganglia	skin	pancreas	urinary tract
Female	1	3	4	3	11	49	6	18	62	11	5	20	14	11
Male	9	22	28	11	35	119	12	32	107	16	7	27	20	13
Unknown	1	3	1	22	23	19	9	12	12	11	5	15	12	4
Ratio: M/F	9.0	7.3	7.0	3.7	3.2	2.4	2.0	1.8	1.7	1.5	1.4	1.4	1.4	1.2

Cell Lines with no females

	biliary tract	salivary gland	Small intestine
Female	0	0	0
Male	1	2	0
Unknown	7	0	1
Ratio: M/F	-	-	-

Source: Broad Institute Cancer Cell Line Encyclopedia: <https://portals.broadinstitute.org/ccle/about#contact>

Reference

1. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503-508, doi:10.1038/s41586-019-1186-3 (2019).