

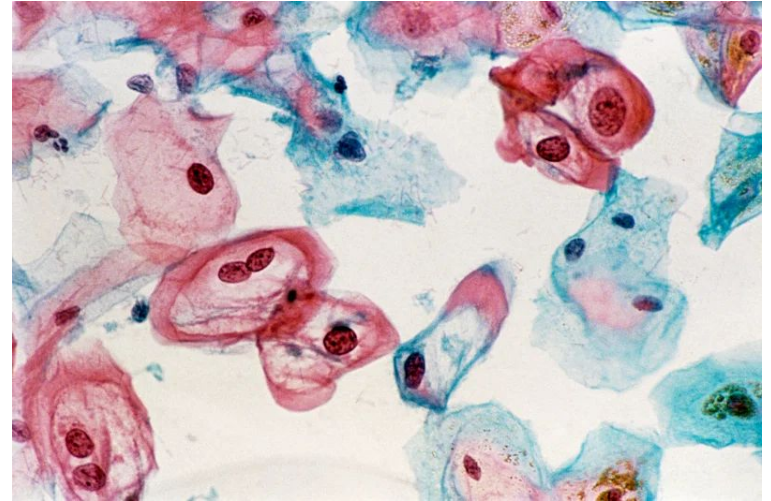
# Genomic Signal Analysis of Cervical Cancer:

---

A look at gene expression

# Cervical Cancer:

- In the US there are:
  - Nearly 14,000 cases diagnosed each year
  - Nearly 4300 deaths attributed to cervical cancer per year
- Pap test has significantly decreased death rate due to early detection
- Most frequently diagnosed between 35-40 w/average age of diagnosis 50
  - 20% of cases are diagnosed after 65
- Two main subtypes:
  - Squamous cell carcinoma - up to 90% of cases
  - Adenocarcinomas - most of the rest
  - Mixed carcinomas



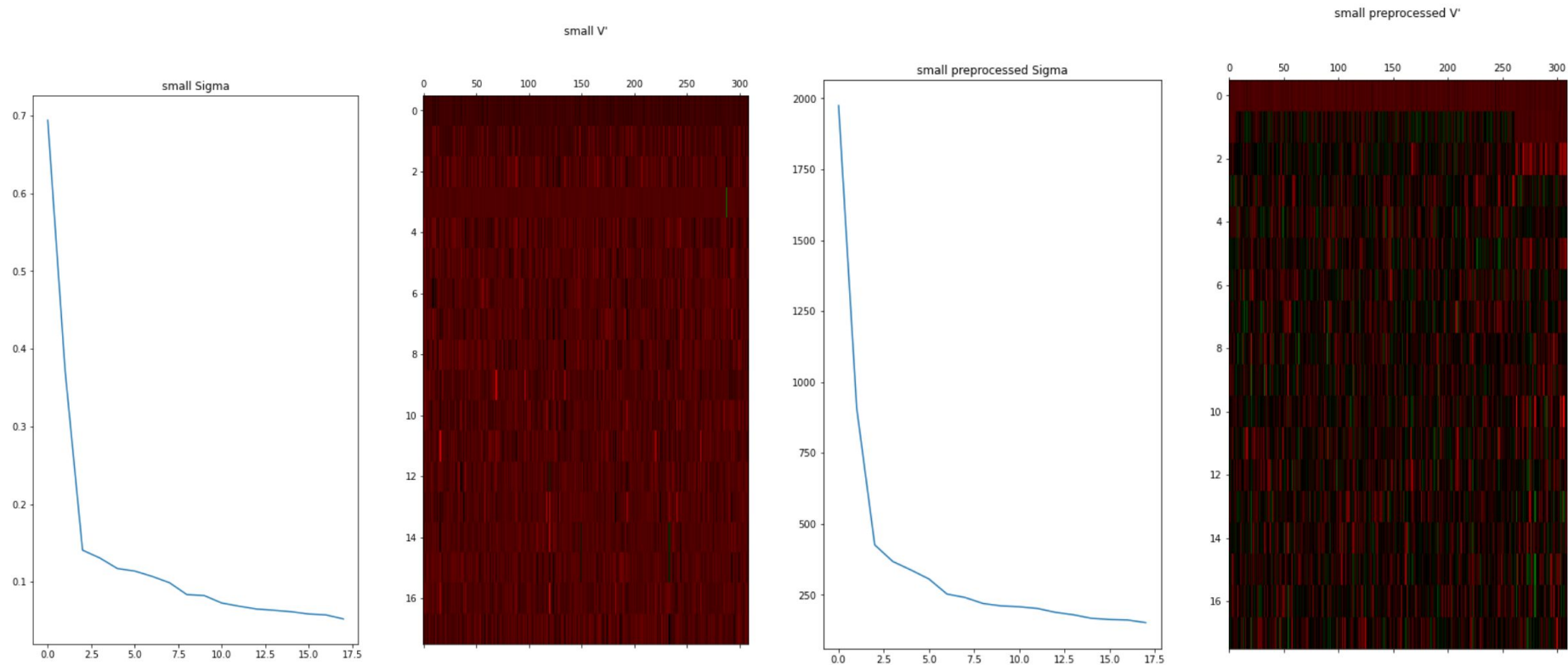
Cervical Cells - blue stain are HPV infected cells

# The Data: Gene expression from Illumina HiSeq

- 309 patients were included in this data set taken from TCGA legacy archive
- 255 patients had Squamous cell carcinoma - 82% of cases vs 90%
- Overall 13% of patients were over 65 vs the expected 20%
  - Looking at Squamous cell specifically increases to 18%

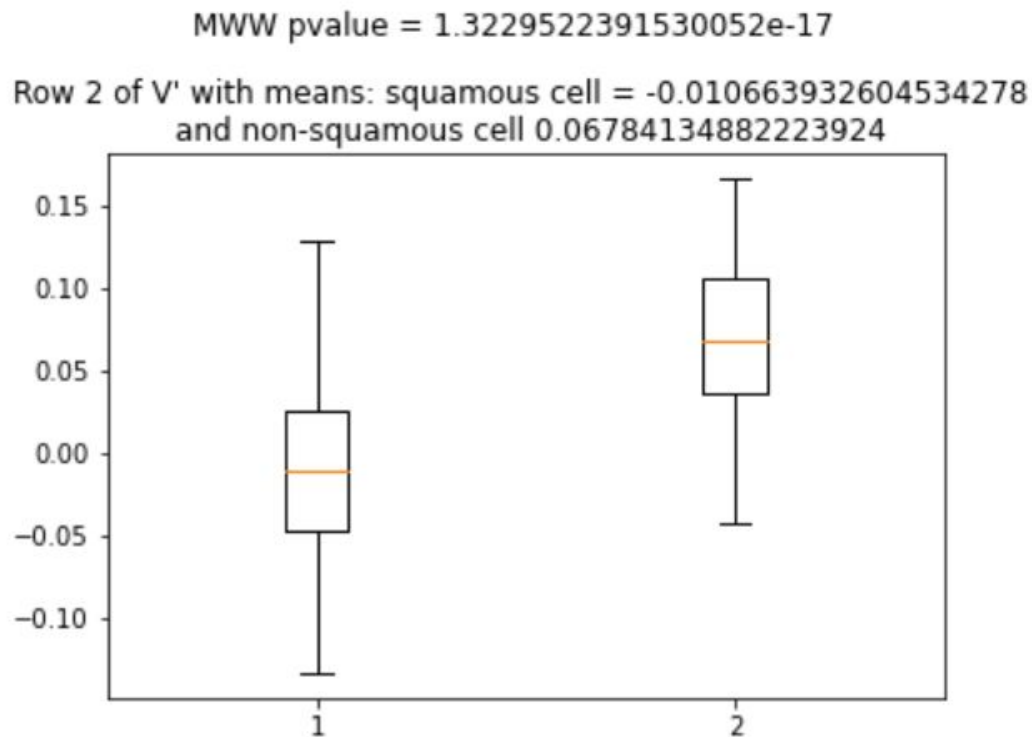


# Preprocessing and sorting by subtype



# Patients were grouped by subtype, Squamous cell vs non-squamous cell

- Most rows had significant p-values
- This row was chosen as it was the most significant prior to preprocessing and second most after

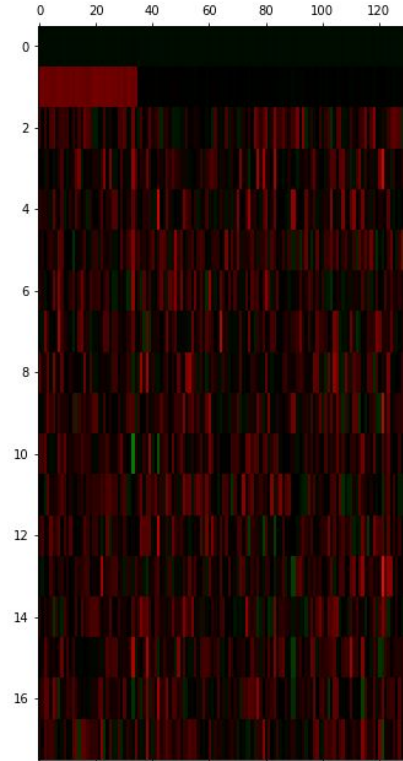


- Gorilla results for the third pattern
- Overexpressed in Squamous Cell Cervical cancer
- Highlighted processes seemed familiar

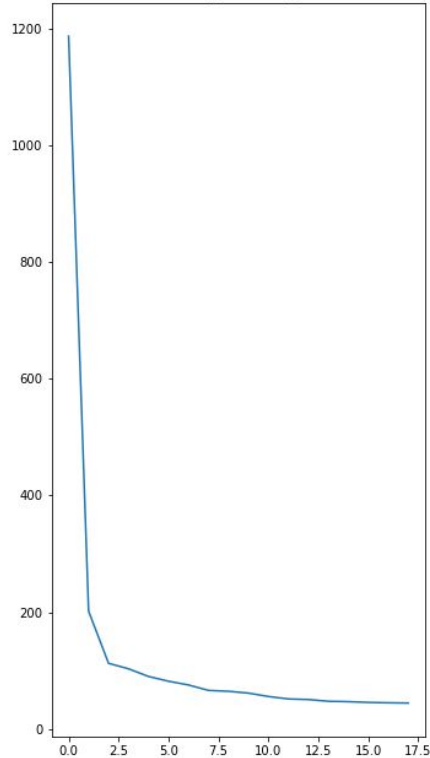
GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
<a href="#">GO:0006614</a>	SRP-dependent cotranslational protein targeting to membrane	3.46E-81	5.36E-77	51.33 (18697,89,221,54)
<a href="#">GO:0006613</a>	cotranslational protein targeting to membrane	2.91E-79	2.26E-75	48.60 (18697,94,221,54)
<a href="#">GO:0045047</a>	protein targeting to ER	2.7E-77	1.39E-73	43.90 (18697,106,221,55)
<a href="#">GO:0072599</a>	establishment of protein localization to endoplasmic reticulum	4.58E-76	1.77E-72	42.30 (18697,110,221,55)
<a href="#">GO:0000184</a>	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	8.38E-75	2.6E-71	38.75 (18697,118,229,56)
<a href="#">GO:0019083</a>	viral transcription	1.28E-72	3.31E-69	39.73 (18697,115,221,54)
<a href="#">GO:0070972</a>	protein localization to endoplasmic reticulum	1.79E-72	3.96E-69	37.83 (18697,123,221,55)
<a href="#">GO:0006413</a>	translational initiation	3.71E-72	7.19E-69	34.22 (18697,136,229,57)
<a href="#">GO:0006612</a>	protein targeting to membrane	1.52E-65	2.63E-62	38.93 (18697,158,152,50)
<a href="#">GO:0000956</a>	nuclear-transcribed mRNA catabolic process	7.47E-61	1.16E-57	34.08 (18697,192,140,49)
<a href="#">GO:0006402</a>	mRNA catabolic process	3.09E-60	4.36E-57	31.50 (18697,212,140,50)
<a href="#">GO:0006412</a>	translation	2.09E-58	2.7E-55	32.54 (18697,197,140,48)
<a href="#">GO:0006401</a>	RNA catabolic process	5.31E-57	6.33E-54	27.59 (18697,242,140,50)
<a href="#">GO:0043043</a>	peptide biosynthetic process	5.2E-56	5.76E-53	29.41 (18697,218,140,48)
<a href="#">GO:0090150</a>	establishment of protein localization to membrane	8.3E-54	8.58E-51	24.31 (18697,253,152,50)
<a href="#">GO:0034655</a>	nucleobase-containing compound catabolic process	6.93E-53	6.71E-50	17.23 (18697,407,152,57)
<a href="#">GO:0046700</a>	heterocycle catabolic process	5.48E-50	5E-47	15.41 (18697,455,152,57)
<a href="#">GO:0044270</a>	cellular nitrogen compound catabolic process	6.25E-50	5.38E-47	15.38 (18697,456,152,57)
<a href="#">GO:0019439</a>	aromatic compound catabolic process	3.76E-49	3.07E-46	14.92 (18697,470,152,57)
<a href="#">GO:0006518</a>	peptide metabolic process	3.28E-48	2.54E-45	13.48 (18697,341,244,60)
<a href="#">GO:0006605</a>	protein targeting	3.37E-48	2.49E-45	14.84 (18697,325,221,57)
<a href="#">GO:1901361</a>	organic cyclic compound catabolic process	1.84E-47	1.3E-44	13.97 (18697,502,152,57)
<a href="#">GO:0072657</a>	protein localization to membrane	4.52E-46	3.05E-43	14.24 (18697,475,152,55)
<a href="#">GO:0043604</a>	amide biosynthetic process	5.85E-46	3.78E-43	18.91 (18697,339,140,48)
<a href="#">GO:0072594</a>	establishment of protein localization to organelle	2.84E-43	1.76E-40	16.00 (18697,409,140,49)
<a href="#">GO:0016032</a>	viral process	1.01E-37	6.02E-35	9.77 (18697,705,152,56)
<a href="#">GO:0044403</a>	symbiont process	1.01E-37	5.8E-35	9.77 (18697,705,152,56)
<a href="#">GO:0044419</a>	interspecies interaction between organisms	3.32E-37	1.84E-34	8.27 (18697,890,155,61)

# Squamous cell lung cancer - Microarray Gene expression

First 18 rows V' of lung data

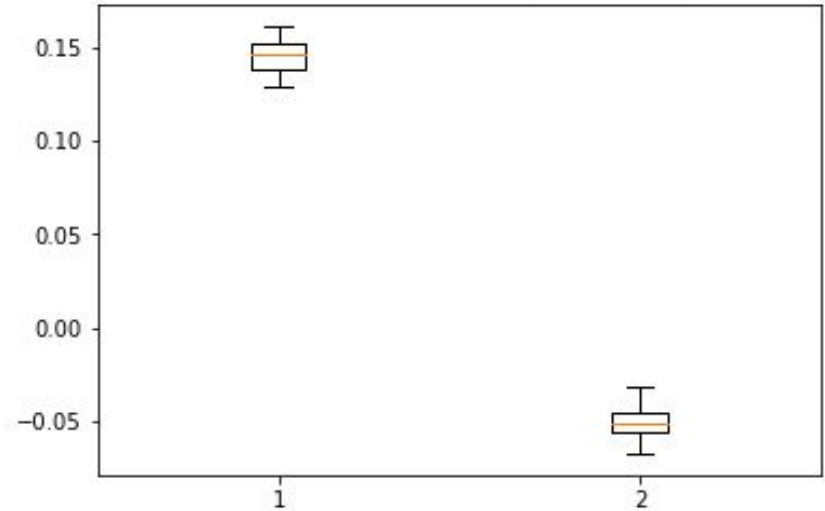


First 18 Sigma of lung data



MWW pvalue =  $2.4637188161967967e-18$

Row 1 of V' with means: Female = 0.1449718357716035  
and Male = -0.051021794280773816





# Processes overexpressed in female patients with Squamous cell lung cancer

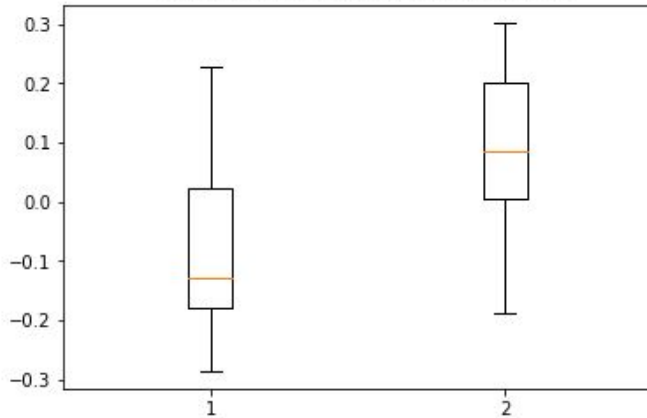
GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
<a href="#">GO:0006614</a>	SRP-dependent cotranslational protein targeting to membrane	9.86E-89	1.47E-84	77.54 (11702,77,98,50)
<a href="#">GO:0006613</a>	cotranslational protein targeting to membrane	1.12E-86	8.32E-83	32.81 (11702,82,274,63)
<a href="#">GO:0006413</a>	translational initiation	9.04E-85	4.49E-81	49.55 (11702,114,116,56)
<a href="#">GO:0000184</a>	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	7.75E-84	2.88E-80	59.66 (11702,102,100,52)
<a href="#">GO:0045047</a>	protein targeting to ER	3.34E-83	9.93E-80	65.61 (11702,91,98,50)
<a href="#">GO:0072599</a>	establishment of protein localization to endoplasmic reticulum	3.45E-82	8.55E-79	63.51 (11702,94,98,50)
<a href="#">GO:0019083</a>	viral transcription	2.68E-80	5.71E-77	59.70 (11702,100,98,50)
<a href="#">GO:0070972</a>	protein localization to endoplasmic reticulum	3.99E-79	7.43E-76	57.41 (11702,104,98,50)
<a href="#">GO:0006612</a>	protein targeting to membrane	5.36E-77	8.86E-74	50.33 (11702,121,98,51)
<a href="#">GO:0000956</a>	nuclear-transcribed mRNA catabolic process	4.89E-71	7.28E-68	38.51 (11702,158,100,52)
<a href="#">GO:0006412</a>	translation	2.89E-69	3.91E-66	40.34 (11702,148,98,50)
<a href="#">GO:0006402</a>	mRNA catabolic process	4.8E-69	5.95E-66	35.79 (11702,170,100,52)
<a href="#">GO:0043043</a>	peptide biosynthetic process	1.43E-66	1.64E-63	36.40 (11702,164,98,50)
<a href="#">GO:0090150</a>	establishment of protein localization to membrane	6.86E-66	7.3E-63	31.68 (11702,196,98,52)
<a href="#">GO:0006401</a>	RNA catabolic process	1.16E-65	1.15E-62	31.53 (11702,193,100,52)
<a href="#">GO:0006605</a>	protein targeting	4.79E-60	4.46E-57	23.97 (11702,264,98,53)
<a href="#">GO:0034655</a>	nucleobase-containing compound catabolic process	3.94E-57	3.45E-54	19.50 (11702,330,100,55)
<a href="#">GO:0006518</a>	peptide metabolic process	2.95E-56	2.44E-53	19.03 (11702,267,129,56)
<a href="#">GO:0044419</a>	interspecies interaction between organisms	2.58E-55	2.02E-52	3.68 (11702,775,735,179)
<a href="#">GO:0043604</a>	amide biosynthetic process	3.46E-55	2.57E-52	23.05 (11702,259,98,50)
<a href="#">GO:0016032</a>	viral process	6.63E-55	4.7E-52	9.66 (11702,622,148,76)
<a href="#">GO:0044403</a>	symbiont process	6.63E-55	4.48E-52	9.66 (11702,622,148,76)
<a href="#">GO:0044270</a>	cellular nitrogen compound catabolic process	6.96E-55	4.51E-52	17.88 (11702,360,100,55)
<a href="#">GO:0046700</a>	heterocycle catabolic process	9.67E-55	5.99E-52	17.78 (11702,362,100,55)
<a href="#">GO:0072594</a>	establishment of protein localization to organelle	4.15E-54	2.47E-51	18.95 (11702,334,98,53)
<a href="#">GO:0019439</a>	aromatic compound catabolic process	7.76E-54	4.44E-51	17.16 (11702,375,100,55)
<a href="#">GO:0072657</a>	protein localization to membrane	1.84E-53	1.01E-50	18.17 (11702,367,93,53)



# Female Lung Cancer Patients (left) vs Squamous Cell Cervical Cancer Patients (right) - grouped at 64 years

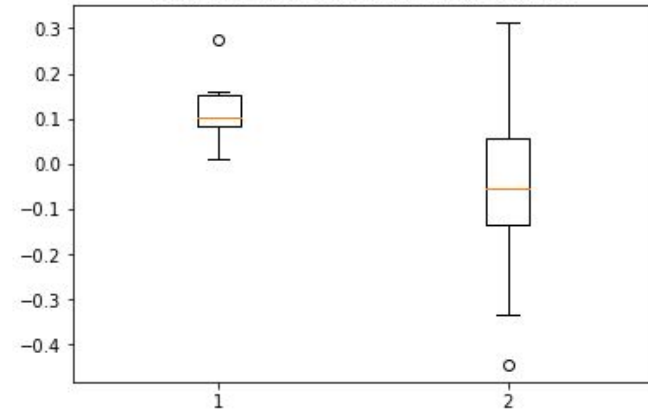
MWW pvalue = 0.0029229968574925884

Row 2 of Female Lung V' with means: younger than 64 = -0.07960838414742612  
and older than 0.08514580556374705



MWW pvalue = 0.002807106774182669

Row 2 of Cervical Subset V' with means: younger than 64 = 0.11849603964098926  
and older than -0.039642997660643886



# The processes enriched for in older patients of both cancer types were quite similar

## Lung Cancer

GO term	Description	P-value
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	7.2E-102
GO:0006613	cotranslational protein targeting to membrane	3.84E-99
GO:0045047	protein targeting to ER	6.08E-95
GO:0072599	establishment of protein localization to endoplasmic reticulum	1.07E-93
GO:0006413	translational initiation	2.12E-93
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	4.56E-93
GO:0019083	viral transcription	2.15E-91
GO:0070972	protein localization to endoplasmic reticulum	5.66E-90
GO:0006612	protein targeting to membrane	5.56E-87
GO:0006412	translation	2.5E-80
GO:0000956	nuclear-transcribed mRNA catabolic process	2.66E-78
GO:0043043	peptide biosynthetic process	3.67E-77
GO:0006402	mRNA catabolic process	4.49E-76
GO:0090150	establishment of protein localization to membrane	2.64E-73
GO:0006401	RNA catabolic process	2.65E-72
GO:0006605	protein targeting	1.38E-68
GO:0006518	peptide metabolic process	4.8E-65
GO:0034655	nucleobase-containing compound catabolic process	3.24E-64
GO:0043604	amide biosynthetic process	6.46E-64
GO:0044270	cellular nitrogen compound catabolic process	8E-62
GO:0046700	heterocycle catabolic process	1.13E-61
GO:0019439	aromatic compound catabolic process	1.04E-60
GO:0072594	establishment of protein localization to organelle	3.25E-60
GO:0072657	protein localization to membrane	2.16E-59
GO:0044419	interspecies interaction between organisms	1.37E-58
GO:1901361	organic cyclic compound catabolic process	1.48E-58
GO:0016032	viral process	3.57E-58
GO:0044403	symbiont process	3.57E-58

## Cervical Cancer

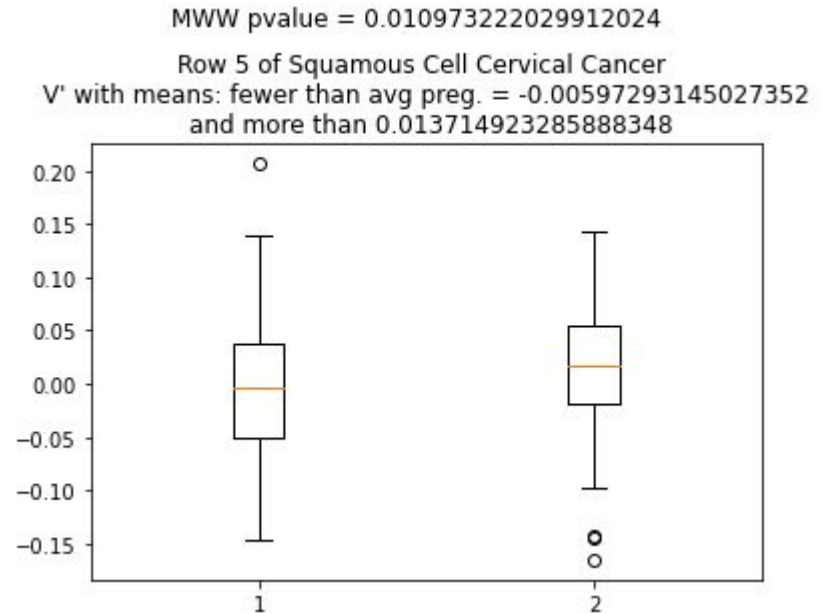
GO term	Description	P-value
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	5.29E-102
GO:0006613	cotranslational protein targeting to membrane	1.07E-99
GO:0044419	interspecies interaction between organisms	6.03E-96
GO:0045047	protein targeting to ER	2.83E-95
GO:0006413	translational initiation	3.13E-95
GO:0072599	establishment of protein localization to endoplasmic reticulum	5.7E-94
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	3.74E-93
GO:0019083	viral transcription	1.95E-92
GO:0016032	viral process	4.03E-91
GO:0044403	symbiont process	4.03E-91
GO:0070972	protein localization to endoplasmic reticulum	3.62E-90
GO:0006612	protein targeting to membrane	1.94E-84
GO:0006412	translation	3.09E-78
GO:0051704	multi-organism process	5.94E-77
GO:0000956	nuclear-transcribed mRNA catabolic process	1.21E-76
GO:0006402	mRNA catabolic process	2.74E-75
GO:0043043	peptide biosynthetic process	4.35E-75
GO:0006401	RNA catabolic process	3.71E-71
GO:0090150	establishment of protein localization to membrane	6E-71
GO:0006605	protein targeting	9.46E-68
GO:0034655	nucleobase-containing compound catabolic process	9.68E-65
GO:0006518	peptide metabolic process	1.17E-64
GO:0043604	amide biosynthetic process	4.6E-63
GO:0046700	heterocycle catabolic process	2.48E-61
GO:0044270	cellular nitrogen compound catabolic process	2.86E-61
GO:0009987	cellular process	4.37E-61
GO:0006810	transport	9.73E-61
GO:0019439	aromatic compound catabolic process	1.95E-60

25/28 top enriched processes were in both with 3 direct matches

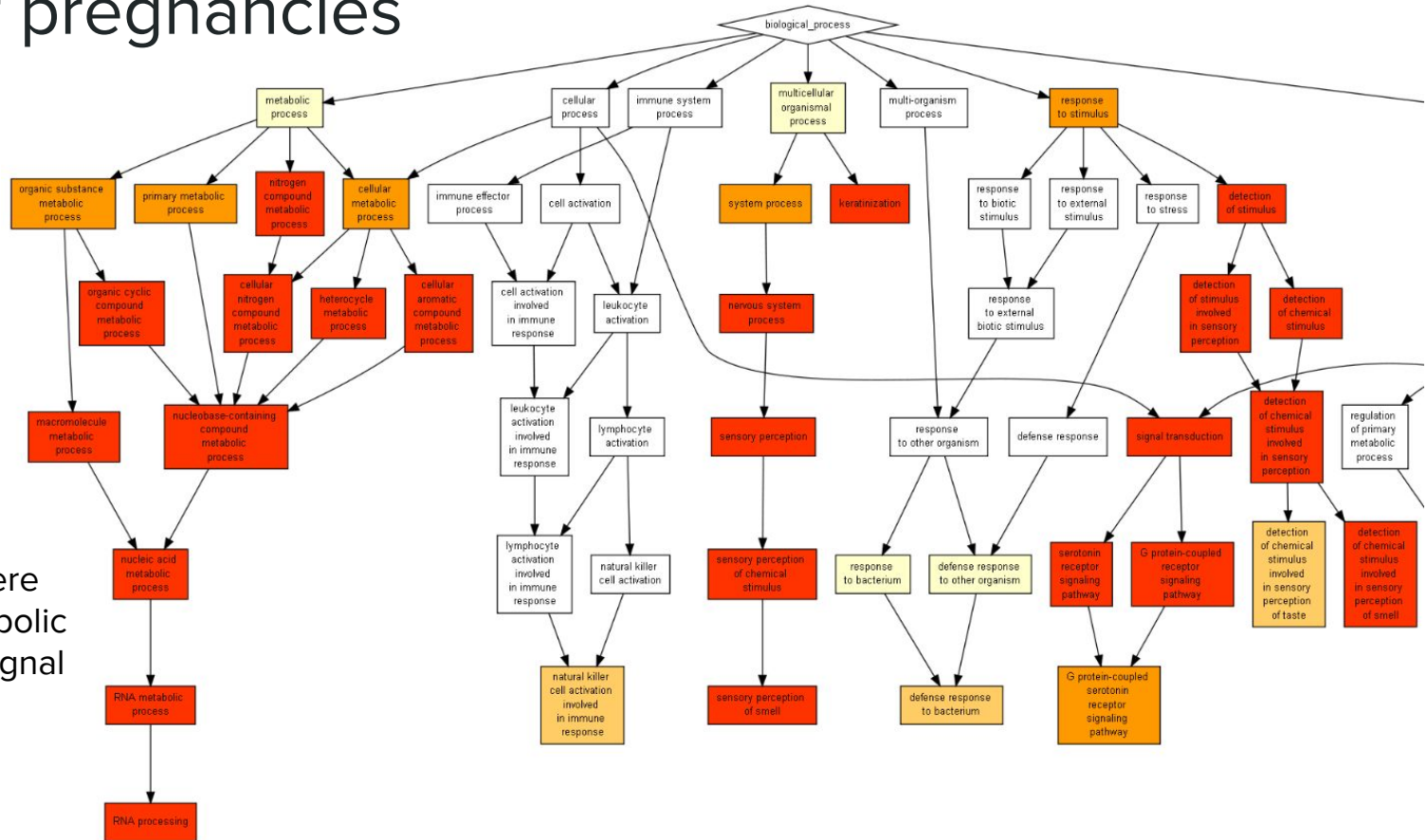
SRP targets proteins to the ER and directly interacts with actively translating ribosomes (Elvekrog, et al. *Curr Opin Chem Biol.* 2015)

# Other labels: number of pregnancies - Squamous Cell

- 37 patients had N/A and were removed leaving 218 patients
- Grouped by over/under the average number of pregnancies
  - 3.8 pregnancies was average
  - 147 under and 71 over



# GOrilla: processes overexpressed in above average number of pregnancies



- Enriched processes where primarily metabolic and sensory/signal based

# GOrilla: processes overexpressed in above average number of pregnancies

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)
<a href="#">GO:0050911</a>	detection of chemical stimulus involved in sensory perception of smell	0E0	0E0	10.39 (18697,370,1624,334)
<a href="#">GO:0050907</a>	detection of chemical stimulus involved in sensory perception	8.28E-305	6.42E-301	8.43 (18697,415,1939,363)
<a href="#">GO:0009593</a>	detection of chemical stimulus	8.81E-283	4.55E-279	9.22 (18697,447,1551,342)
<a href="#">GO:0050906</a>	detection of stimulus involved in sensory perception	7.97E-272	3.09E-268	8.76 (18697,475,1551,345)
<a href="#">GO:0051606</a>	detection of stimulus	9.33E-223	2.89E-219	6.35 (18697,624,1745,370)
<a href="#">GO:0007186</a>	G protein-coupled receptor signaling pathway	4.25E-168	1.1E-164	3.70 (18697,1229,2010,489)
<a href="#">GO:0006396</a>	RNA processing	3.79E-133	8.39E-130	7.40 (18697,1184,459,215)
<a href="#">GO:0016070</a>	RNA metabolic process	3.22E-96	6.24E-93	5.23 (18697,1872,384,201)
<a href="#">GO:0090304</a>	nucleic acid metabolic process	1.82E-73	3.14E-70	3.93 (18697,2501,384,202)
<a href="#">GO:0006139</a>	nucleobase-containing compound metabolic process	7.37E-59	1.14E-55	3.24 (18697,3048,384,203)
<a href="#">GO:0046483</a>	heterocycle metabolic process	3.57E-56	5.04E-53	3.10 (18697,3222,384,205)
<a href="#">GO:0006725</a>	cellular aromatic compound metabolic process	4.52E-55	5.84E-52	3.05 (18697,3269,384,205)
<a href="#">GO:0007608</a>	sensory perception of smell	1.56E-54	1.87E-51	12.58 (18697,94,980,62)

# Conclusion:

- It seems to me that the overexpression of SRP cotranslational targeting to the membrane and viral/symbiont processes may be a marker for the female immune response to Squamous Cell cancers
  - Overexpressed in Squamous Cell Cervical cancer
  - Overexpressed in female patients with Squamous Cell Lung cancer
    - Underexpressed in male patients with squamous Cell Lung cancer
- Compare these enriched processes with other those in other Squamous Cell Cancer Types



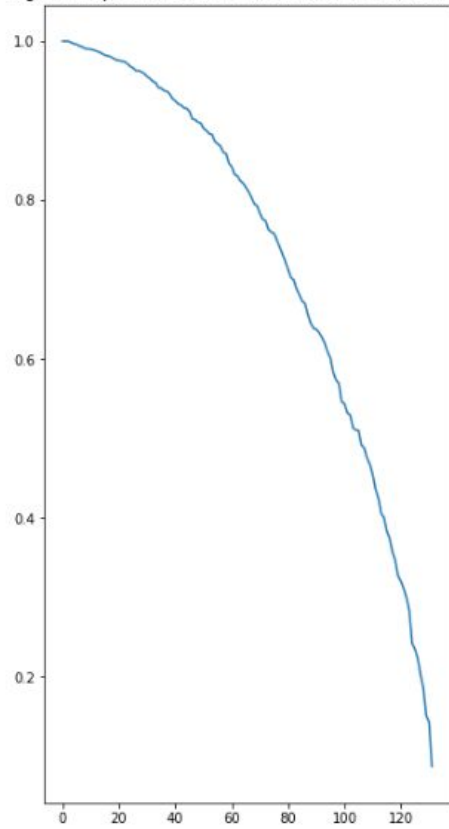
# HPV related overexpression in Cervical Cancer

- 123 genes were found to be consistently overexpressed in HPV associated Cervical Cancers<sup>1</sup>
- These genes were all related to cell cycle regulation or DNA regulation

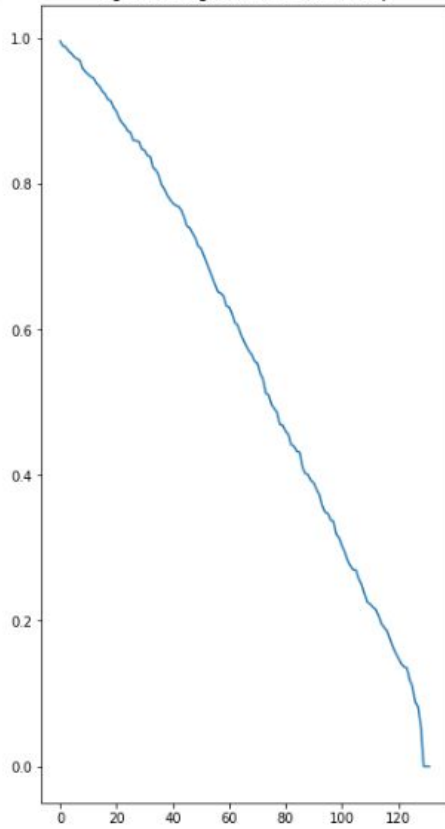
<sup>1</sup>Rosty, C. *et al.* Oncogene. 2005. "Identification of a proliferation gene cluster associated with HPV E6/E7 expression level and viral DNA load in invasive cervical carcinoma"

# Attempted GSVD of paired Lung and Cervical cancer

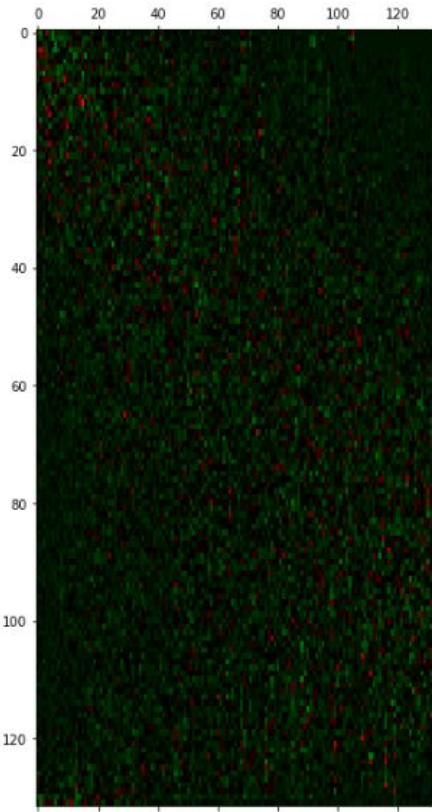
Sigma of Squamous Cervical subset data after QR decomp



Sigma of lung data after QR decomp



V' of Squamous Cervical subset data after QR decomp



V' of lung data after QR decomp

