

# Mapping putative regulatory regions in breast cancer using histone H3K4me1 in multiple breast cancer cell lines

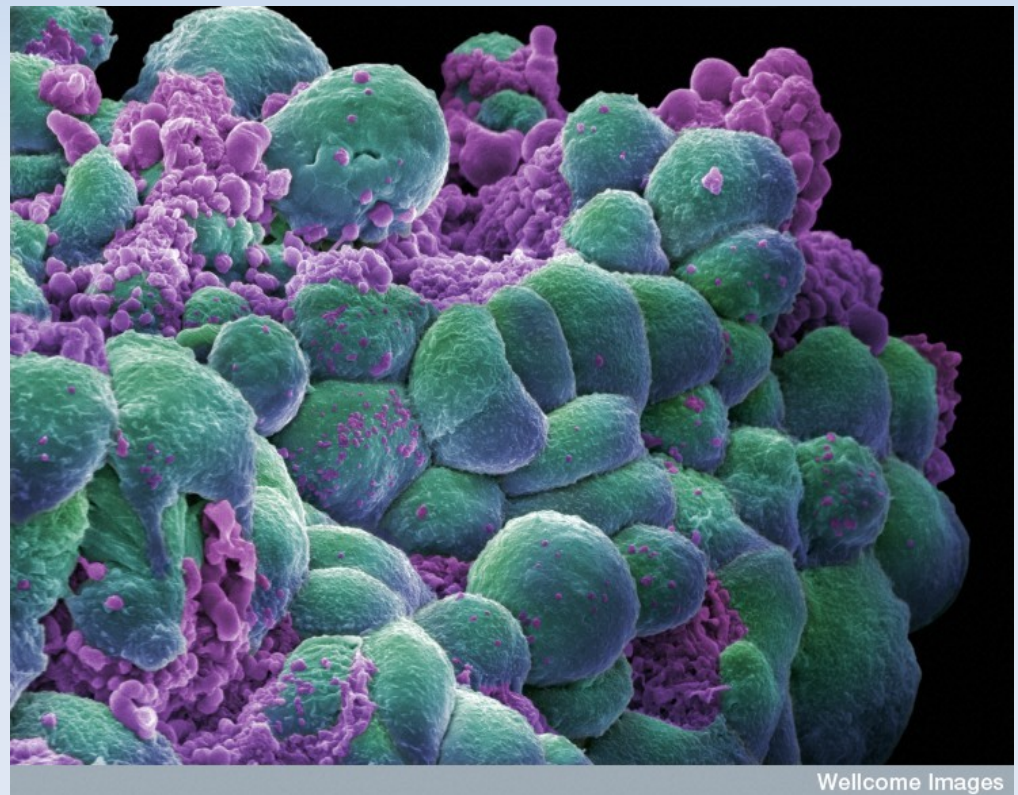
Denil Wickrama  
MSc. Candidate  
Dr. Steven Jones' lab

# Outline

- Background
- Method
- Results
- Conclusions

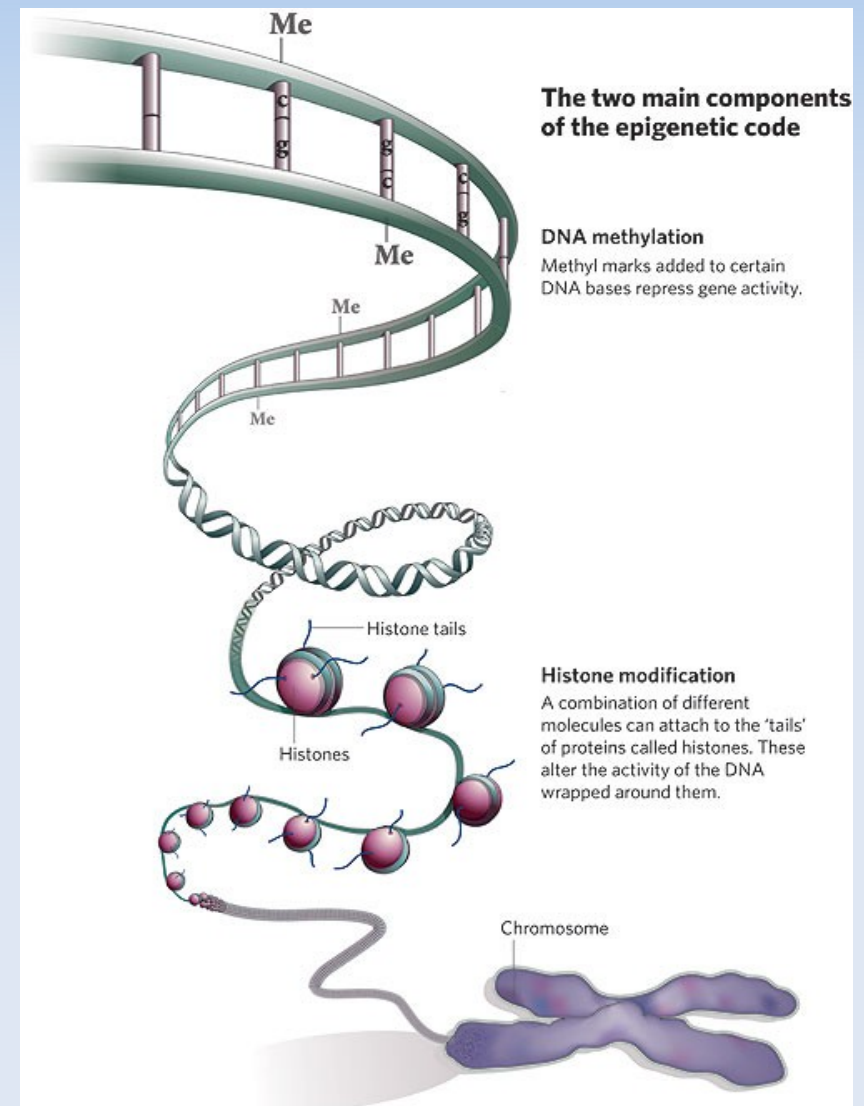
# Breast cancer

- 100 Canadian women will die of breast cancer every week
- Luminal
  - ER+ or PR+
  - Most prevalent
  - Best prognosis
- Basal
  - 15-20% of cancers
  - ER- and PR-, usually HER2-
  - poorer prognosis



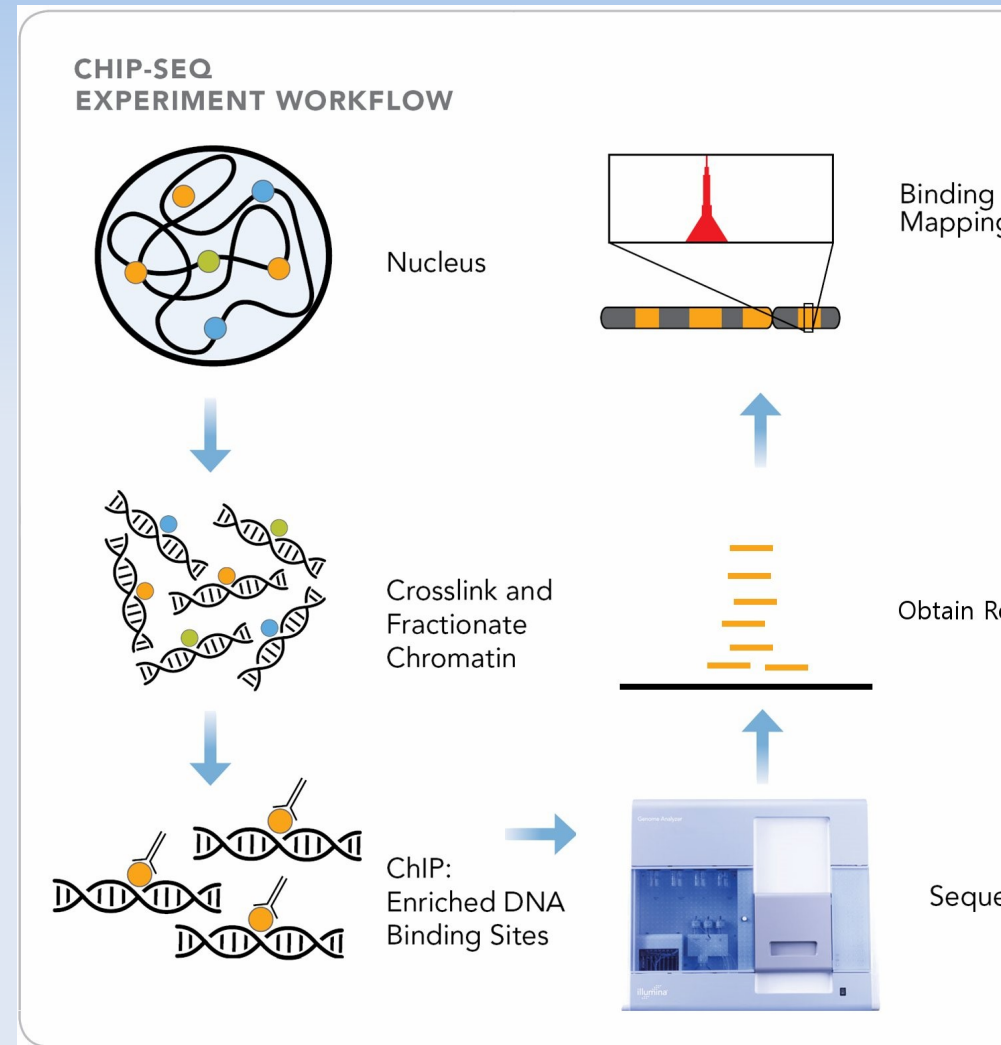
# H3K4me1

- Previous studies found this was a mark for potentially active distal regulatory regions
- Enriched relative to H3K4me3 at putative enhancers
- H3K4me1 in promoter regions diffuse
- Positive correlation exists between gene expression H3K4me1 in both promoter regions and gene bodies.



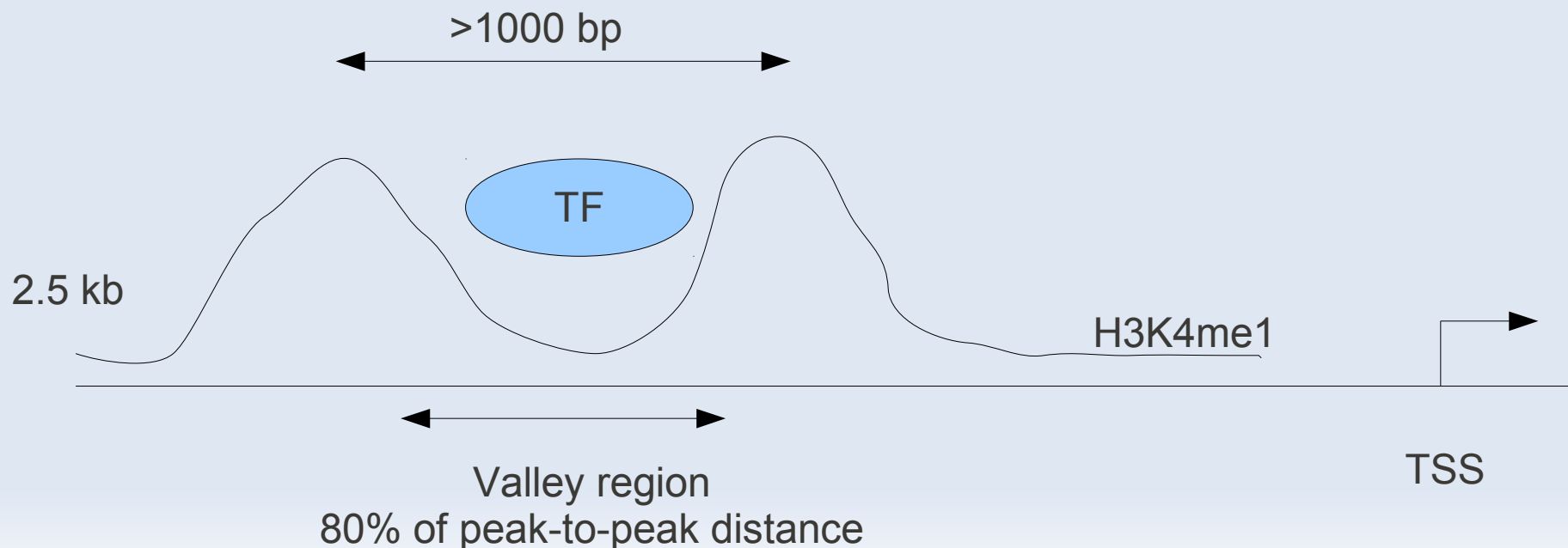
# ChIP-Seq

- ChIP-Seq, is used to analyze interactions of DNA with points of interest such as proteins
- Process
  - Bind POI with specific Ab
  - Chromatin immunoprecipitation
  - massively parallel DNA sequencing
  - Align reads to genome
  - Find overlapping reads
- \$1000 genome is approaching



# General Method

- Find peaks with the Vancouver Short Read Analysis Package
- Define valley regions that are have peaks  $>1000$  bp apart
- Find valley regions in the promoter regions of genes
- Correlate expression of genes found by RNA-seq with



# Motifs

- H3K4me1 marked sites are significantly ( $P < 10^{-3}$ ) enriched in motifs

		ORegAnno	Valleys	Percent	Shuffled Percent
MDA-MB-231	Basal	6	138	4.35	0.18
BT549	Basal	289	3825	7.56	0.35
HS578T	Basal	293	2543	11.52	0.37
MCF7	Luminal	293	3543	8.27	0.36
TD47D	Luminal	245	2417	10.14	0.38
HS578T	Cancer	297	2553	11.63	0.38
HS578BST	Control	197	2256	8.73	0.33

# Regulation of downstream gene

- Combined breast cancer cell lines
  - Enrichment of breast cancer genes downstream of H3K4me1 marked promoters
- HS578Bst and is the match control for HS578T
  - Taken from the same patient
  - Most marked valleys are unique to a cell line

Enrichment of breast cancer genes in dataset marked with H3K4me1

	Total	Marked with H3K4me1
Ensemble Gene IDs	63281	12466
Breast Cancer Genes	2180	1322
Percent	3.4%	10.6%

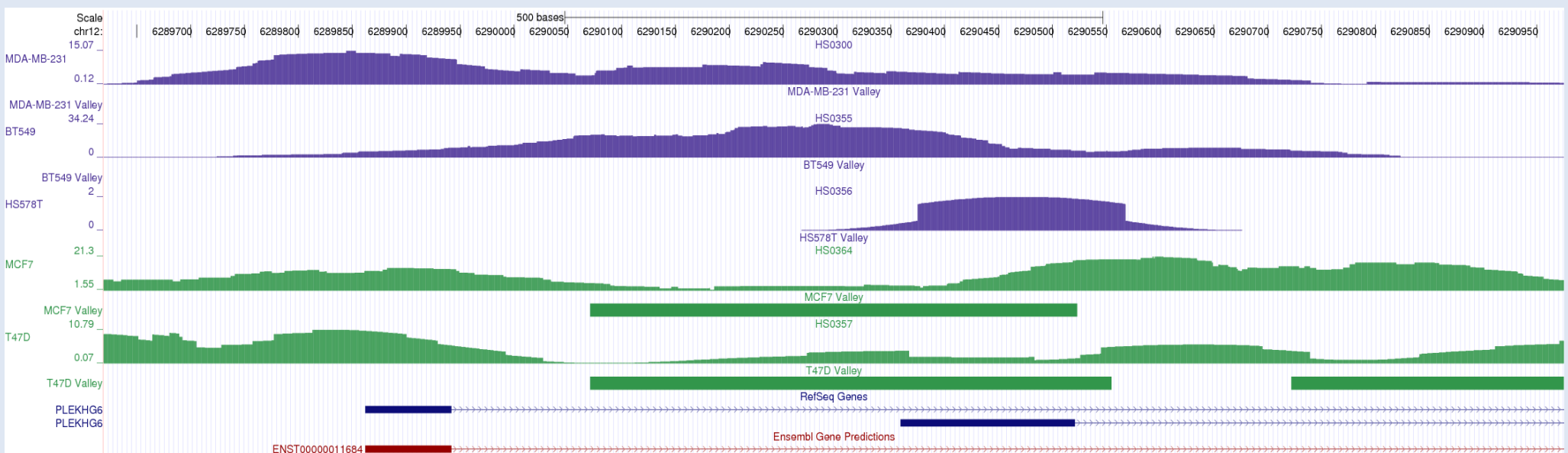
Promoter regions of genes marked with flanking H3K4me1

Shared	Unique to Cancer cell line	Unique to Control cell line
547	2006	4264

# Basal vs. Luminal

	Total	Shared
Total	6206	3996
Unique	2210	0
Shared	3996	3996

Luminal	Basal
MCF7	MDA-MB-231
T47D	BT549
	HS578T

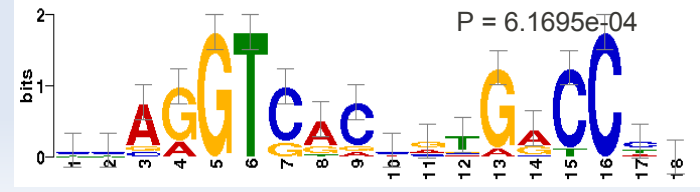


# ESR1

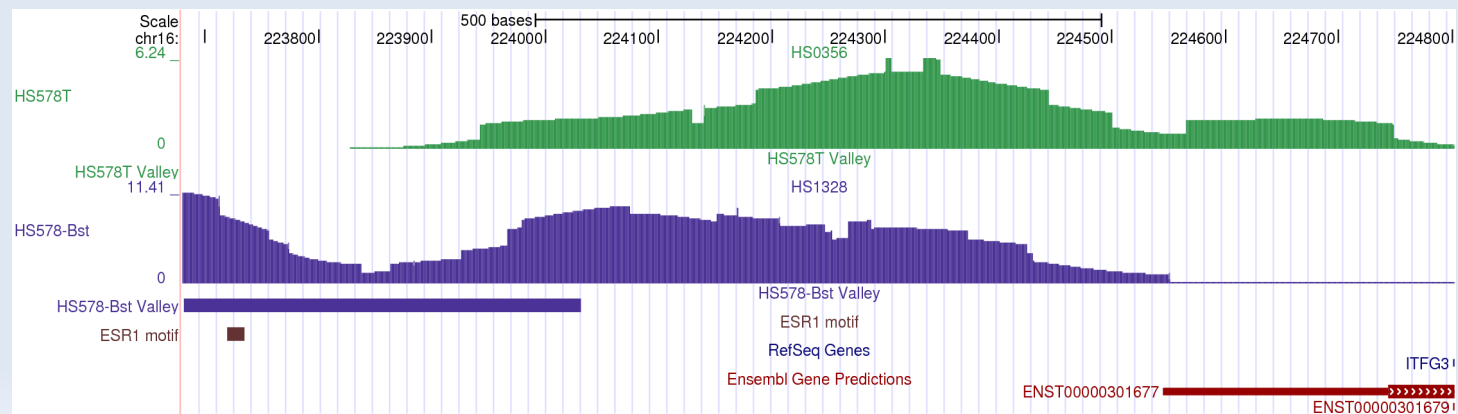
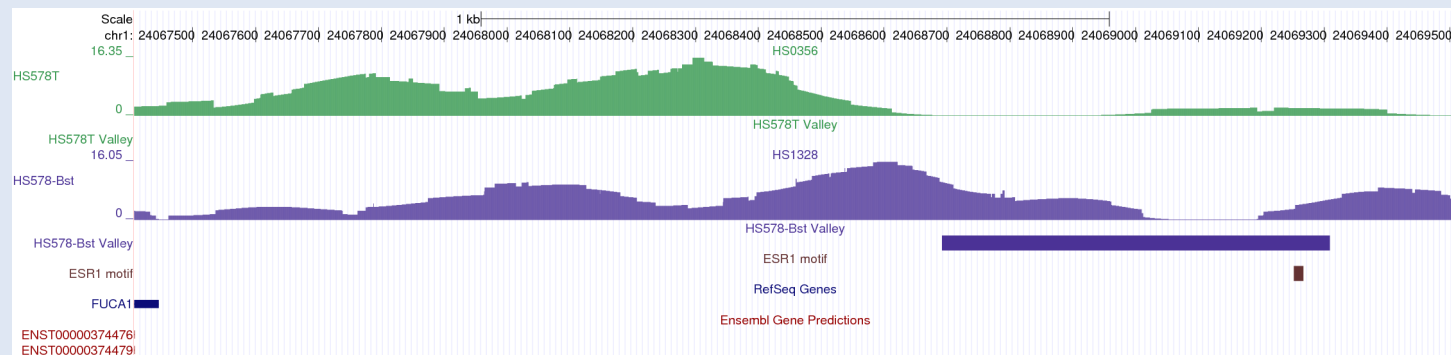
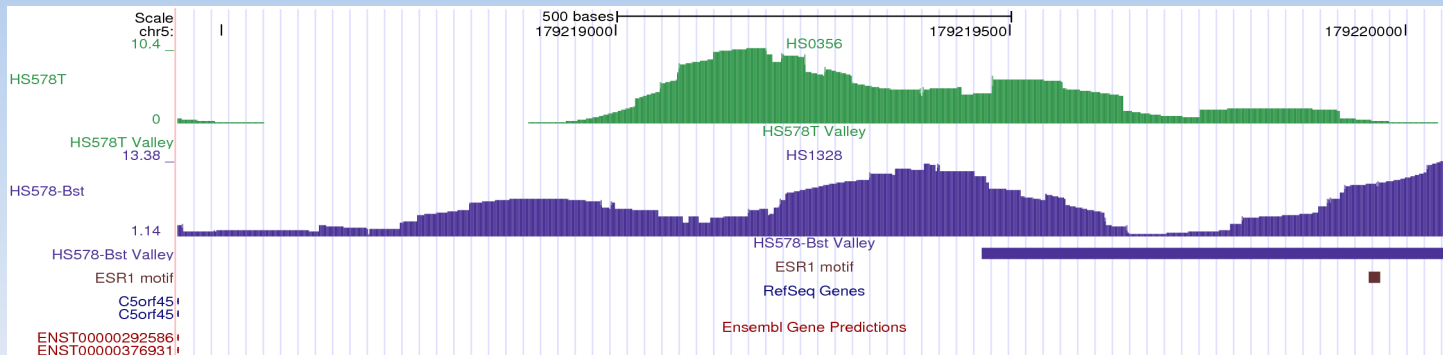
- Genes in matched control cell lines were divided into 4 categories based on:
  - H3K4me1 mark present uniquely in one cell line
  - 4x overexpression in cell line
- When searching the group of genes marked in the control cell line (HS578Bst) and also overexpressed in the control

- P4HTM
- CYP1B1
- AC092162.1
- PIK3IP1
- ITFG3
- PIK3IP1
- KCNK2
- C5ORF45
- S100A4
- IGFBP3

Name	Strand	Start	p-value	Sites ?
chr3:49000698-49001435~ENST00000383729	+	140	8.28e-10	CCGGGAGGCA <b>GAGGTTGCAGTGAGC</b> CGAGATGAAG
chr2:38158526-38158950~ENST00000407341	+	146	8.28e-10	GTGGGAGGCA <b>GAGGTTGCAGTGAGC</b> AGAGATTGCA
chr1:24068692-24069309~ENST00000374476	-	560	8.28e-10	CCAGGAGGTG <b>GAGGTTGCAGTGAGC</b> TGAGATTGCA
chr22:30019305-30020081~ENST00000402249	-	689	2.55e-09	CTGGGAGGTG <b>GAGGTTGCAGTAAGC</b> CGAGATGGTG
chr16:223681-224031~ENST00000301679	+	39	2.55e-09	CTGGGAAGCA <b>GAGGTTGCAGTAAGC</b> TGTGATCCCG
chr22:30018637-30019145~ENST00000402249	-	289	5.80e-09	AACCCGGCAG <b>GTGGTTGCAGTGAGC</b> CGAGATTGTG
chr1:213243007-213243247~ENST00000391895	-	225	1.02e-08	G <b>AAGGTTGCAGTGAGC</b> TGAGATTGTG
chr5:179219464-179220056~ENST00000376931	-	490	7.28e-08	GGTACTGCA <b>GAGGCTGCAGTCAGT</b> TTATTGTCAT
chr1:151786904-151787406~ENST00000368716	-	169	7.28e-08	TCAGAGCAGG <b>GAGGTTGAAGTCGGC</b> CCTCAGAAGG
chr7:45929217-45929896~ENST00000381086	-	515	1.31e-07	GAGGGGCAGG <b>GAGCTTGCAGTTATC</b> AAATAAACT



# Uniquely marked promoters with overexpression of downstream gene



# Conclusions

- Flanking H3K4me1 sites proximal to genes are enriched for regulatory regions
- The regions are associated with changes in transcription downstream
- Motif searching in these marked genes with concordant expression changes may yield insights into regulatory mechanisms and genes involved in cancer

# Acknowledgements

- People

- Dr. Steven Jones
- Anthony Fejes
- Gordan Robertson
- Mikhail Bilenky
- Yongjun Zhao
- Timothee Cezard

- Funding

- MSFHR/CIHR Bioinformatics Training Program

