Identification of novel SNPs in differentially expressed genes and its association with Horn Cancer of *Bos indicus* Bullocks by Next-Generation Sequencing

Thakor P.B.*, Koringa P.G.; Jakhesara S.J.; Rank D.N. and Joshi C.G.

Department of Animal Biotechnology

College of Veterinary Science and Animal Husbandry

Anand Agricultural University, Anand

Guajrat, India

Introduction





- Carcinoma of Horn is only reported in *Bos indicus.*
- It arise from pseudo stratified columnar epithelium of the horn core mucosa.
- It affects approximately 1% of the cattle population and account 83.34% of total tumor reported in the cattle population.

- The incidence of horn cancer is more frequent in Kankrej breed than other zebu cattle, crossbred or nondescript cattle .
- It is most commonly found in working bullocks (~95% cases), sometimes in cows (~5% cases) and rarely in bulls, buffaloes, sheep and goats.
- The main difficulties in early diagnosis include the absence of specific markers.
- Gene associated SNPs are important and discussed by many scientific groups with evidence suggesting that the various carcinogenic pathways harbor functional inherited SNPs that affect cell signalling, cell proliferation, cell migration and stability in tissue, resulting in differences in cancer risk and clinical outcome in humans .

Objectives

- To generate transcriptome profile using RNA-Seq and identify SNPs correlated with Horn Cancer in Kankrej bullock *(Bos indicus*) .

Experimental Design



Step 1 : Transcriptome Profiling



Sample Collection



Histopathological confirmation of collected tissue



RNA Extraction



Qualitative and quantitative evaluation of RNA



mRNA separation and fragmentation



cDNA synthesis



DNA Library Preparation for Genome Sequencer FLX



Emulsion PCR



Pyrosequencing in Pico Titer Plate



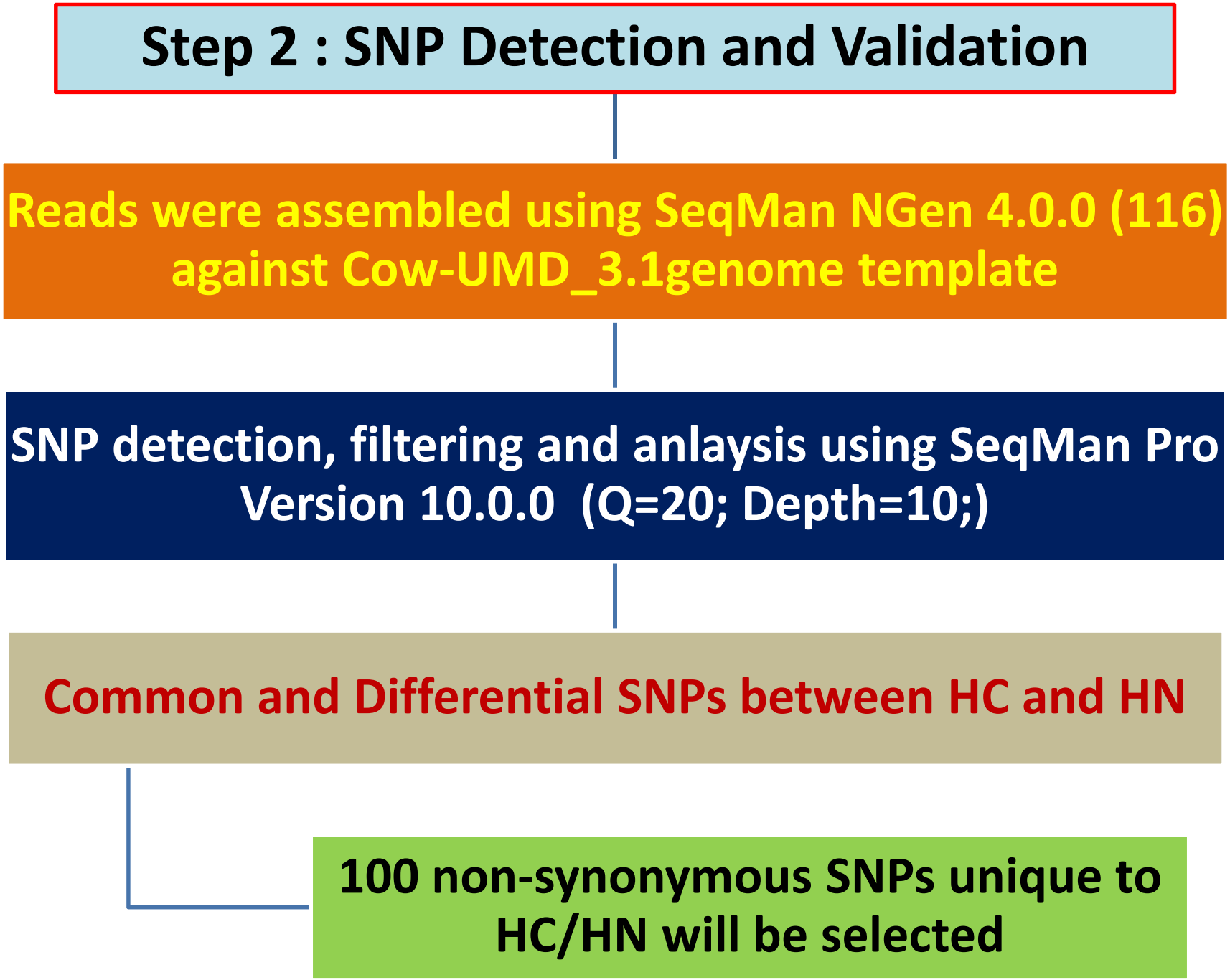
Image Processing and Data filtering



Bioinformatics analysis of data to identify differential gene expression and AS events



Validation of transcriptome data by RT qPCR


Step 2 : SNP Detection and Validation
Reads were assembled using SeqMan NGen 4.0.0 (116) against Cow-UMD_3.1genome template
SNP detection, filtering and anlaysis using SeqMan Pro Version 10.0.0 (Q=20; Depth=10;)
Common and Differential SNPs between HC and HN
100 non-synonymous SNPs unique to HC/HN will be selected

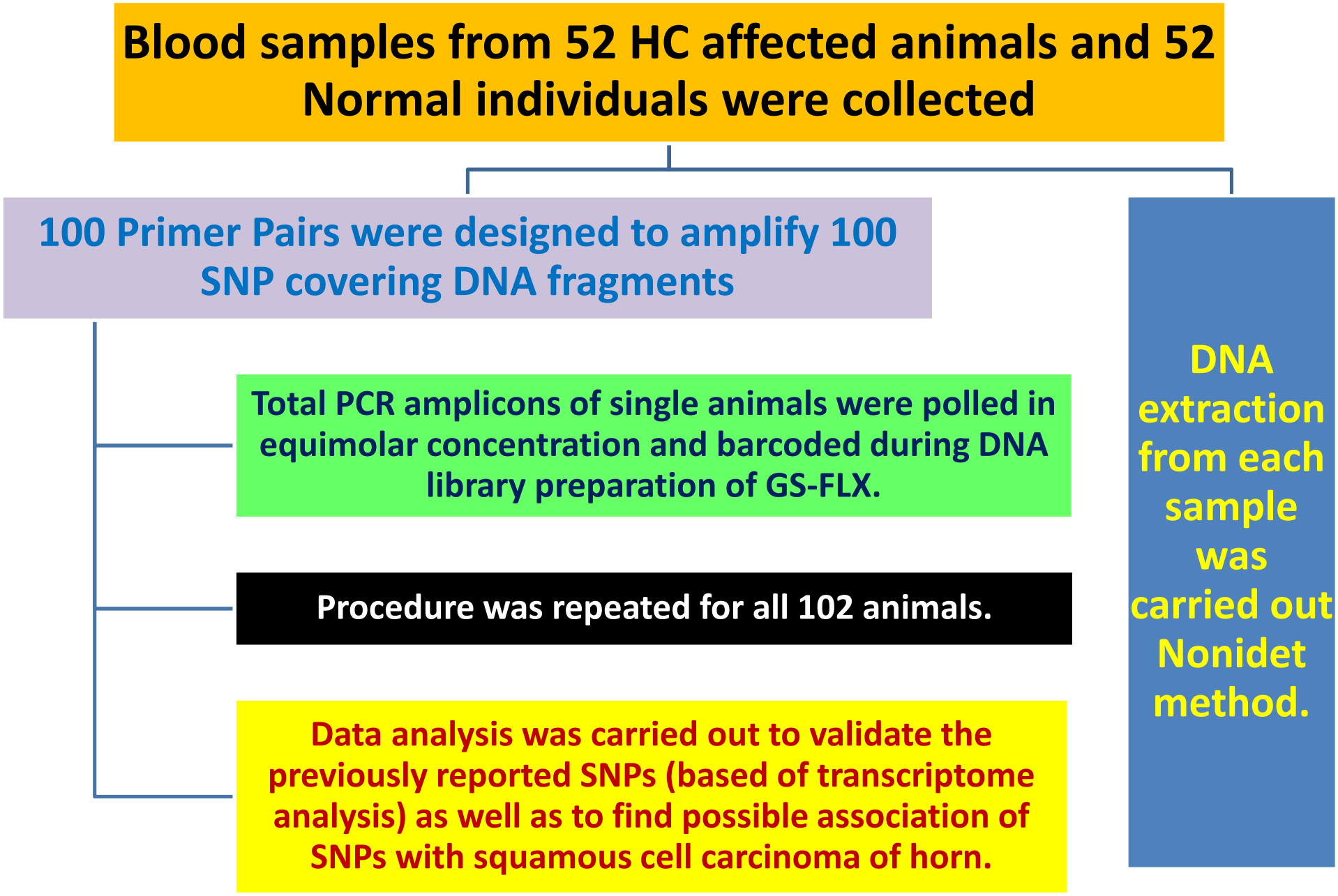

Blood samples from 52 HC affected animals and 52 Normal individuals were collected
100 Primer Pairs were designed to amplify 100 SNP covering DNA fragments
Total PCR amplicons of single animals were polled in equimolar concentration and barcoded during DNA library preparation of GS-FLX.
Procedure was repeated for all 102 animals.
Data analysis was carried out to validate the previously reported SNPs (based of transcriptome analysis) as well as to find possible association of SNPs with squamous cell carcinoma of horn.
DNA extraction from each sample was carried out Nonidet method.

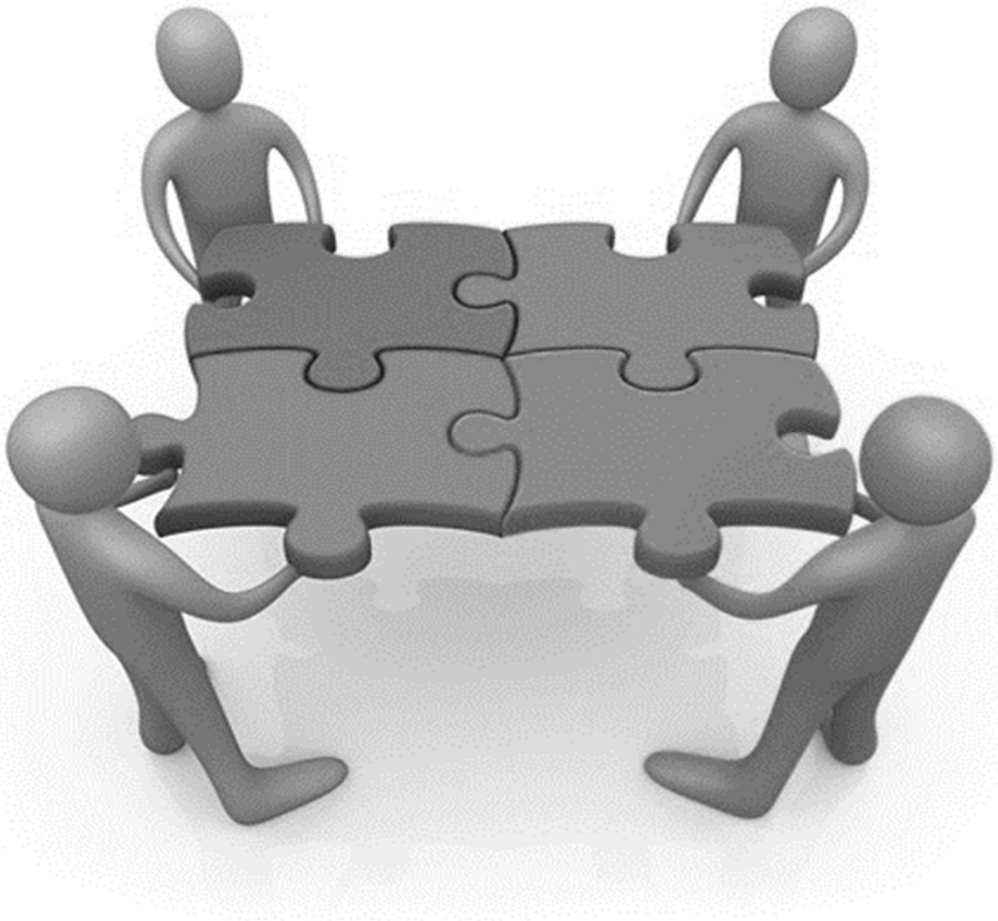


Figure-1: Kankrej bullock suffering from horn cancer [Animal's left horn tilted due to osteolysis and increased weight of horn as sinus filled with tumorous growth and debilitated body score]

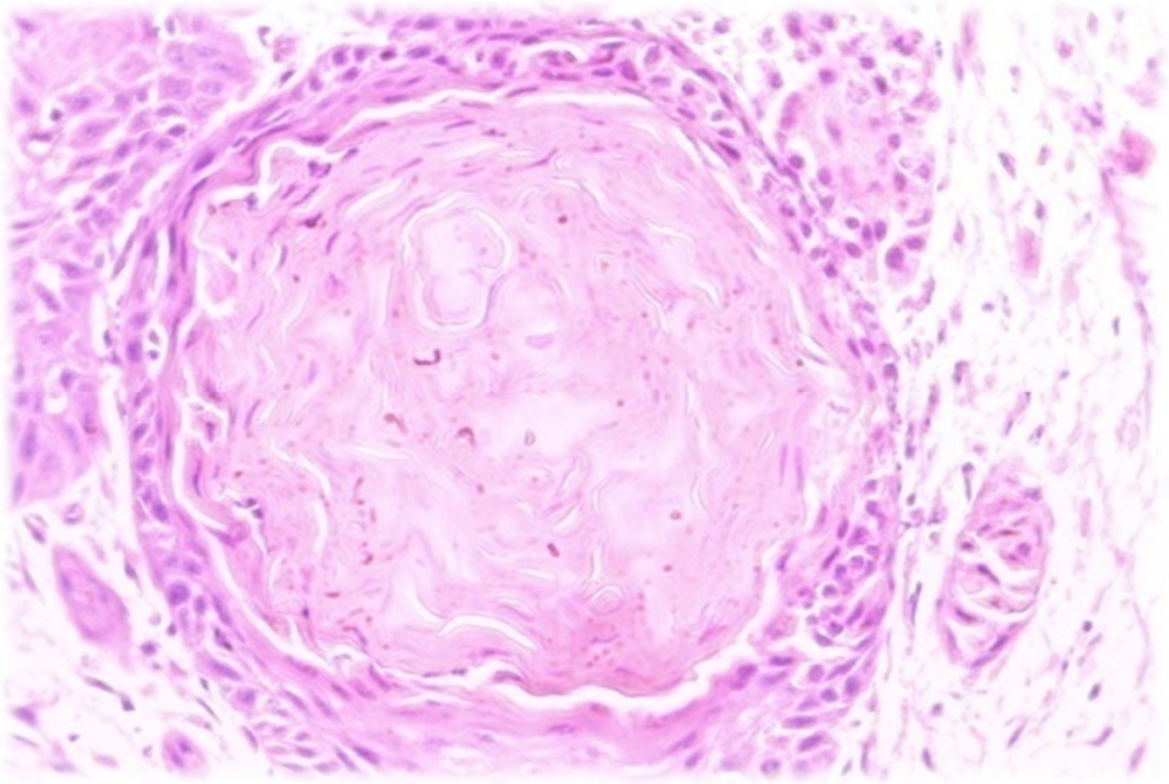


Figure-2: Mid part of the horn filled with tumorous mass & loss of bony sinusoidal structure [After surgical removal of horn from it's base]

Results & Discussion



Histopathological Examination



HC tissue shows proliferation of squamous cell in the horn corn mucosa with typical pattern of cell nest (epithelial pearl formation), circular region in the middle is well formed keratin mass. The squamous cell vesicular nucleous with few mitotic figures, whereas other areas of tumor tissue show myxomatous tissue with stylet shaped cells

Transciptome sequencing run-1				
Particulars	Region1(HC)	Region 2(HN)	Region 3(HN)	Region 4(HC)
No. of Reads	251,214	283,365	269,343	258,243
Total bases	104,964,935	117,483,469	110,446,245	107,310,595
Avg. Length	417.83	414.6	410.06	415.54
Transciptome sequencing run-2				
Particulars	HN		HC	
No. of Reads	26,613		323,871	
Total bases	10,131,001		128,612,747	
Avg. Length	380.68		397.11	
Transciptome sequencing run-3				
Particulars	HN		HC	
No. of Reads	5,129		87,122	
Total bases	1,838,485		31,608,392	
Avg. Length	358.45		362.81	
Summary of all GS-FLX Titanium runs				
Particulars	HN		HC	
No. of Reads	584,450		920,450	
Total bases	239,899,200		372,496,669	
Avg. Length	410.47		404.6897	

	Horn Normal	Horn Cancer
No. of Sequences	583491	909362
Total bases	239814801	372496669
Average Length (in bases)	411	405
No. of sequences aligned using Seq Man NGen 4.0.0*	585167	913139
No. of unaligned sequences	12324	20334
No. of unaligned bases	2941006	4432748

Summary of Alignment

DNA Star

- **SeqMan Pro**
  - Visualization and filtering of SNP in one sample using different parameters
- **ArrayStar**
  - Comparative gene expression, SNP, microRNA analysis of different samples
  - Detailed output can be exported

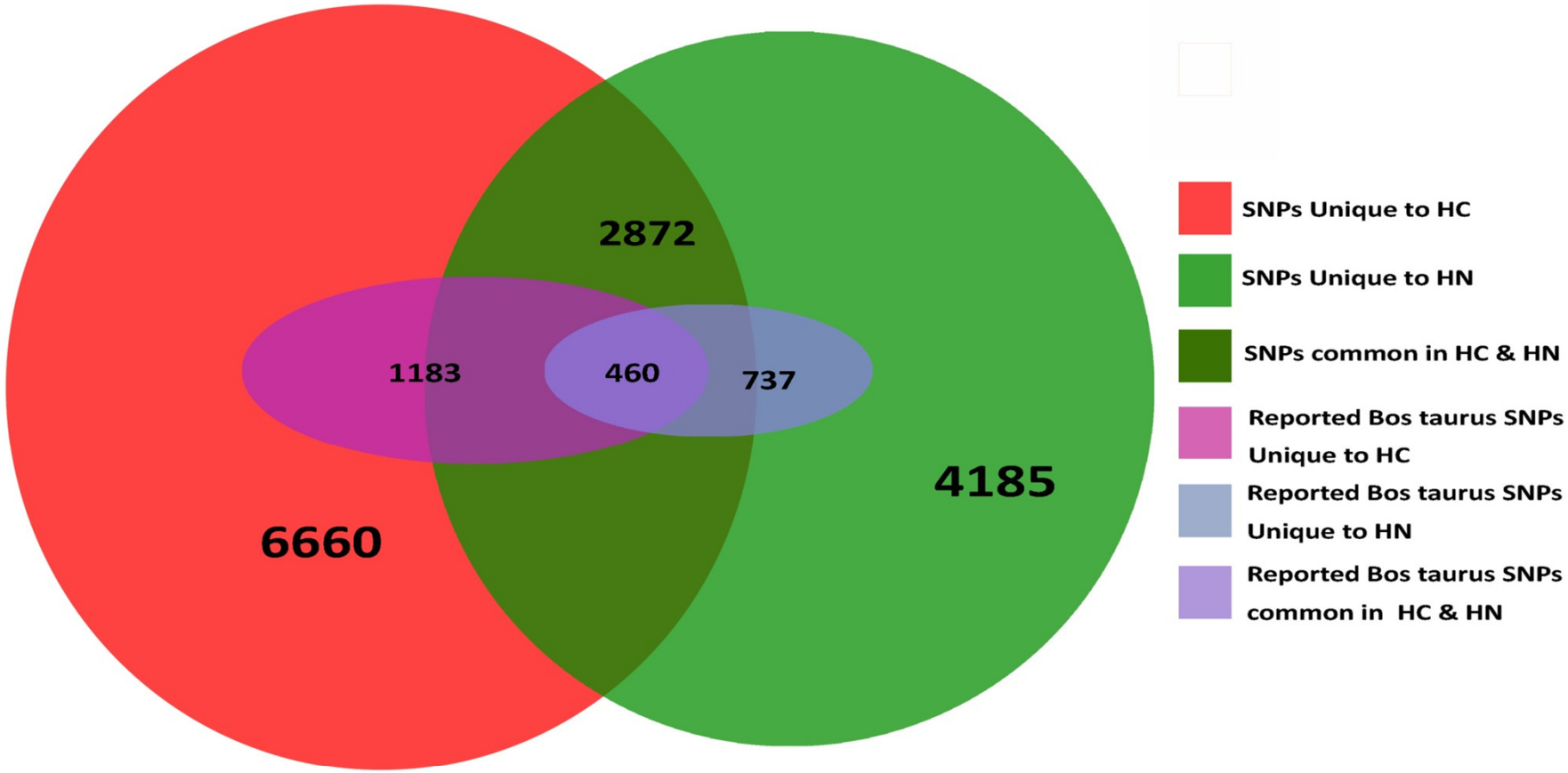


Figure 2 : Venn Diagram showing distribution of SNPs in HN and HC in *Bos taurus* as well as in *Bos indicus*

Gene ID	Name of gene	Function/ Pathways Involved	HC Only	HN Only
Genes Up regulated in HC				
KRT6A	Keratin 6A	Transcription regulator with positive and negative domains and activities	6 (4)	0
KRT6B	Keratin 6B	It is a histone acetyltransferase involved in both positive (N-terminus) and negative (C-terminus) regulation of transcription	14	0
CA1	Carbonic Anhydrase I	Carbonate dehydratase activity; Lyase activity	14 (5)	0
GJB2	Gap Junction Protein, beta-2	Cell Communication	4 (1)	0
TGM3	Transglutaminase 3	Keratinocyte differentiation; Hair follicle morphogenesis	2 (7)	0
TNS4	Tensin 4	Role in cancer cell metastasis through enhancing the migration as well as the invasion through tubulogenesis,	5 (1)	0
KRT14	Keratin 14	Cell Communication	15 (3)	0
COL17A1	Collagen, type XVII, alpha 1	Cell Communication	27 (3)	0
MKI67	Antigen Identified by monoclonal antibody Ki-67	Cell proliferation; Organ regeneration; Cellular response to heat	4	0
DSG1	Desmoglein 1	Cell Communication	13 (1)	0
HK2	Hexokinase 2	Transmembrane transport ; Gluconeogenesis	5 (1)	0
ODZ2	Teneurin Transmembrane Protein 2	Signal transduction; Integral to membrane	7	0
DSC3	Desmocollin 3	Cell Communication	12 (1)	0

Genes Up regulated in HN			HC Only	HN Only
BPIFB1	BPI fold containing Family B, member 1	Extracellular space; Lipid binding	0	2 (1)
FOSB	FBJ murine Osteosarcoma viral oncogene homolog B	Regulation of progression through cell cycle	0	11
NR4A2	Nuclear receptor subfamily 4, group A, member -1	Regulation of transcription; Signal transduction	0	0
CXCL17	Chemokine (C-X-C motif) ligand 17	Angiogenesis; Extracellular region; Chemotaxis; Cell differentiation	0	7
TNXB	Tenascin XB	Focal adhesion; ECM-receptor interaction; Cell Communication	0	0
FAM107A	Family with sequence similarity 107, member A	Regulation of cell growth	0	8 (2)
ATF3	Activating transcription factor 3	Cell Stability ; Tumor Suppressor	0	9
ALDH1A1	Aldehyde dehydrogenase 1 family, member A1	Retinal dehydrogenase activity	0	(4)
LOC531699	-	-	0	0
LRIG1	Leucine-rich repeats and immunoglobulin-like domains 1	Tumor suppressor through suppression of receptor tyrosine kinases,	0	4
ZBTB16	Zinc finger and BTB domain containing 16	Tumor Suppressor; Specific transcriptional repressor activity	0	0
FXYD3	FXYD domain containing ion transport regulator 3	Integral to plasma membrane; Chloride transport; Regulation of catalytic activity	0	0
C4A	Complement Component 4A	Complement and coagulation cascades	0	2
ATP6V0A4	ATPase, H+ transporting, lysosomal V0 subunit a4	Oxidative phosphorylation; ATP synthesis; Epithelial cell signaling	0	0
FHAD1	Forkhead-associated (FHA) phosphopeptide binding domain 1	Cell Communication	0	0

Amplicon sequencing using Ion Torrent PGM for validation of identified SNP in 50 HC and 50 HN samples

Sample No.	Data	Sample No.	Data
HN1	3.4 MB	HC1	68 MB
HN2	4.8 MB	HC2	79 MB
HN3	3.6 MB	HC3	76 MB
HN4	4.2 MB	HC4	95 MB
HN5	3.7 MB	HC5	60 MB
HN6	3.0 MB	HC6	68 MB
HN7	3.9 MB	HC7	70 MB
HN8	3.6 MB	HC8	64 MB
HN9	3.3 MB	HC9	55 MB
HN10	3.4 MB	HC10	34 MB
HN11	3.2 MB	HC11	78 MB
HN12	3.5 MB	HC12	99 MB
HN13	3.2 MB	HC13	77 MB
HN14	3.6 MB	HC14	46 MB
HN15	2.9 MB	HC15	89 MB
HN16	6.4 MB	HC16	84 MB
HN17	6.3 MB	HC17	8.3 MB
HN18	6.6 MB	HC18	7.8 MB
HN19	4.3 MB	HC19	7.3 MB
HN20	4.3 MB	HC20	6.1 MB
HN21	3.5 MB	HC21	5.9 MB
HN22	3.8 MB	HC22	6.6 MB
HN23	2.9 MB	HC23	5.7 MB
HN24	2.8 MB	HC24	7.7 MB
HN25	5.6 MB	HC25	5.3 MB

Sample No.	Data	Sample No.	Data
HN26	4.7 MB	HC26	5.9 MB
HN27	6.6 MB	HC27	62 MB
HN28	6.8 MB	HC28	67 MB
HN29	3.2 MB	HC29	69 MB
HN30	3.5 MB	HC30	48 MB
HN31	3.2 MB	HC31	77 MB
HN32	3.6 MB	HC32	73 MB
HN33	112 MB	HC33	56 MB
HN34	50 MB	HC34	138 MB
HN35	53 MB	HC35	62 MB
HN36	111 MB	HC36	124 MB
HN37	82 MB	HC37	54 MB
HN38	70 MB	HC38	44 MB
HN39	80 MB	HC39	80 MB
HN40	73 MB	HC40	73 MB
HN41	172 MB	HC41	85 MB
HN42	72 MB	HC42	68 MB
HN43	97 MB	HC43	86 MB
HN44	74 MB	HC44	80 MB
HN45	53 MB	HC45	82 MB
HN46	108 MB	HC46	80 MB
HN47	106 MB	HC47	84 MB
HN48	112 MB	HC48	84 MB
HN49	72 MB	HC49	50 MB
HN50	100 MB	HC50	84 MB

List of SNP showing different genotype between HC and HN

Ref Pos	dbSNP ID	Feature Name	Neg Log10 Prob Allele	Ref Seq	HC - Called Seq			HN - Called Seq			Protein Change
63251805	**136870681**	**BPIFA1**	**3.398**	**T**	**T|C**	**0.5**	**0.5**	**T|C**	**0.2**	**0.8**	**F189S**
87328340	109751733	ODAM	2.804	-	A|-	0.14	0.86	A|-	0.36	0.64	F54fs
74274533	137420399	PI3	2.148	G	G|A	0.37	0.63	G|A	0.2	0.8	R172Q
147024541	110653421	COL18A1	1.920	T	C		1	T|C	0.06	0.94	L750P
14700388		LMNA	1.710	A	T|A	0.36	0.64	T|A	0.21	0.79	D262E
112889632	109829723	SERPINE2	0.795	C	T|C	0.39	0.61	T|C	0.3	0.7	W107.
105366897	211290601	IGFBP2	0.712	T	T|C	0.77	0.23	T|C	0.67	0.33	
120090702	41608018	GYG1	0.541	T	T|C	0.35	0.65	T|C	0.28	0.72	S293G
14615220	108978307	WFDC18	0.451	C	T|C	0.73	0.27	T|C	0.67	0.33	A48V
25861437	382138331	BLA-DQB	0.253	T	T|A	0.69	0.31	T|A	0.71	0.29	E78V
147024648		COL18A1	0.022	G	G|A	0.48	0.53	G|A	0.47	0.53	A786T

- **The SNP identified in our study has already been reported as dBSNP ID rs136870681 in whole genome assembly of *B. taurus* cattle.**

- **BPI fold containing family A, member 1 (BPIFA1) gene is 7696bp long gene with six exons in it [NCBI Reference Sequence: AC_000170.1; region from base 63247184 to 63254779].**

- **A missense mutation at 63251805 position results in to allele change TTC ⇒ TCC at 189 position amino acid substitution resulted in to F (Phe) ⇒ S (Ser) residual change. (Zimin et al., 2009).**

- **Comparative analysis with SCC of human lungs with HC in bullocks, resulted in identification of numbers of commonly altered aberrant signaling pathways responsible for carcinogenesis.**

- **This is the first high throughput RNA-Seq based approach to identify SNPs and Indels in squamous cell carcinoma of *Bos indicus* animals.**

Acknowledgement








GUINNESS
WORLD RECORDS














શેઠ આણંદજી કલ્યાણજી
પેઢી





