Chapter 5

Comprehensive Discussion

5.1 JCPyV prevalence in the sub-Himalayan part of West Bengal:

JCPyV is widely spread in the human population (Padgett and Walker, 1976) and is the causative agent of PML in individuals with suppressed immune system. The present study was conducted within the sub-Himalayan part of West Bengal where no such study was ever conducted. Six hundred and thirteen (613) samples were collected from different regions of Jalpaiguri, Darjeeling, Coochbehar and Alipurduar districts of West Bengal. The study consisted of two groups: one group containing only immunocompromised subjects and the other one containing non-immunocompromised individuals. The immunocompromised group contained pregnant women in their 3rd trimesters, patients receiving steroids and cancer patients undergoing chemotherapy. The non-immunocompromised group consisted of individuals with healthy immune system. Some of the subjects in this group belong to the tribal group residing in this part of India.

The sub-Himalayan parts of West Bengal have long been inhabited by several tribal groups like Santhal, Oraon, Munda, Rabha, Chik-Baraik, Toto, Koch, Mech, Bhutia and others who belong to different linguistic groups, Austro-Asiatic, Dravidian and Tibeto-Chinese (Bhasin, 2006). Oraons are the second largest and the Mundas are the third largest tribal population after the Santhals, constituting about 14% and 7.8% respectively of total tribal population of West Bengal State of India (Census of India, 2001). Oraon and Mundas are immigrants in the tea gardens of West Bengal from the Chotanagpur plateau of the Jharkhand State. These tribes are divided into a number of small groups which practice exogamy (Bell *et al*., 1970). Rabha are an indigenous Assamese community mostly residing in Assam, Meghalaya and West Bengal. They mainly speak Assamese language, but in some areas, Rabha dialect is also used. In West Bengal, they are mostly found in Jalpaiguri and Coochbehar districts. Mech tribe (also known as the Bodo-kachari tribe) belongs to the kachari tribal group and mainly speaks Bodo language which is a Tibeto-Burman dialect. Mech tribes currently reside in parts of Assam and West Bengal.

JCPyV is ubiquitous and has been found in about 70% to 90% of the human population worldwide (Agostini *et al.*, 1997; Shackelton *et al.*, 2006). Out of the 613 samples, 50 samples tested positive for JCPyV with an overall incidence rate of 8.15% in this region. The incidence rate of JCPyV in the immunocompromised group was 10.60% i.e., 28 individuals among the 264 individuals tested positive for JCPyV. And JCPyV status within the non-immunocompromised group was 6.30%. Twenty-two (22) of 349 samples tested positive for JCPyV in this group. The rate of occurrence of JCPyV in the immunocompromised group of people was higher compared to the group consisting of non-immunocompromised individuals which was in agreement with the results reported earlier (Azzi *et al.*, 1996; Zanotta *et al.*, 2013; Boukoum *et al.*, 2016). The frequency of the prevalence of the virus varies between the genders and was found to be higher in males which were in accordance with the studies done on Native Americans (Agostini *et al.*, 1997) and tribals of Africa (Chima *et al.*, 1998).

A wide range of viral load was observed in the samples, ranging from 3.52×10^2 to 6.71×10^6 copies/ml of sample. The mean viral load in urine (2.19×10^6) was higher than that observed in the blood (7.77×10^5) samples. The mean viral load in the present study was comparatively higher than that reported in other studies done around the world. The mean and median viral load in the samples were comparatively higher than normal healthy subjects of Switzerland (Egli *et al.*, 2009), USA (Husseiny *et al.*, 2010), Portugal (Matos *et al.*, 2010), Kuwait (Chehadeh *et al.*, 2013) and Pakistan (Hussain *et al.*, 2017). It has long been suggested that nutritional deficiency might result in an immunodeficient condition in humans, which in turn enhances the susceptibility towards infection and disease (Harbige, 1996). Majority of the population in India cannot meet their dietary requirements due to poor socio-economic conditions and this could be the reason for presence of high viral load. However, the study of the implications of this higher viral load vis-à-vis underlying disease prognosis was beyond the scope of the present study. It also remains to be seen the cause of this higher JCPyV load.

5.2 NCCR Architecture among the endemic JCPyV strains:

The non-coding control region (NCCR) of JCPyV varies considerably among different natural isolates. Naturally occurring variant of the NCCR found in the CY strain of JCPyV is termed as 'Archetype' strain which lacks sequence repeats in the regulatory region and contains additional sequences, is rarely been associated with PML and is mostly found in kidney and urine.

Another variant of NCCR detected in brain, CSF and blood of PML patients is derived from the archetype virus strain during reactivation via rearrangement, insertion or deletion of sequences and is referred to as 'Prototype' (Sabath and Major, 2002). The prototype strain Mad-1 is characterized by the presence of a 98 bp tandem repeat that results in the duplication of TATA-Box and increase in the number of transcription factor binding sites. It has been postulated that the rearrangement in the NCCR may change the biological properties of the virus in due of a persistent infection.

In this study, DiAlign program of Genomatix software suite (Catharius *et al.*, 2005) was used to check similarity of Non coding control region (NCCR) of endemic JCPyV strains NB1, NB2, NB3, NB4, NB5, NB6, NB-PU1, NB-PU2 and NB-PU3 among themselves and also with other strains of the virus such as archetype CY, prototype Mad1, LH3 (Tibet), Tai3 (Taiwan) and IN8 (India). The NB1, NB2, NB3, NB4 and NB5 isolates were from Oraon and Munda tribes. The NB6 isolate was from the Rabha tribal group and NB-PU1, NB-PU2 and NB-PU3 were isolated from the pregnant women group. NB-PU3, NB3, NB4 and NB5 were identical. All the NCCR sequences i.e. NB1, NB2, NB3, NB4, NB5, NB-PU1, NB-PU2 and NB-PU3 except the NB6 NCCR sequence paired closely with each other (97-100% similar). Pairwise comparison with other strains revealed that the NCCR sequence of NB1, NB2, NB3, NB4, NB5, NB-PU1, NB-PU2, NB-PU3 were almost similar to the Tibetan LH3 (97-98% similar). The NCCR sequence of NB6 from the Rabha group is different from the Oraon/Munda group and largely similar to the north-east Asian JCPyV strain, the

archetypal CY. These NCCR sequences including that of NB6 were divergent from Tai3, IN8 and as expected from the Mad1 control regions.

Multiple sequence alignment of endemic JCPyV NCCRs with other previously reported strains revealed one 10 nucleotide (169-178) deletion in block B and one di-nucleotide (454-455) deletion in block F in NB-PU1, NB-PU2, NB-PU3, NB1, NB2, NB3, NB4 and NB5 isolates. Point mutations in different sites within the NCCR of the endemic strains have also been recorded. Point mutations were observed at 4, 13, 26, 27, 69, 226 and 452 nucleotide positions of the sequence alignment when compared with the archetype strain CY (**Fig.12**).

Several transcription factors have a role in the regulation of JCPyV gene expression which include NF-κB (Ranganathan and Khalili, 1993), NFAT4 (Wollebo *et al*., 2012), upstream Target or up-TAR (Chowdhury *et al*., 1993), Tst-1 (Wegner *et al*., 1993), Sp-1 (Henson *et al*., 1992), Spi-B (Marshall *et al*., 2010), GBP-i (Raj and Khalili, 1994), Y-box binding protein 1 (YB-1) and Purα (Chen and Khalili, 1995), Nuclear factor 1 or NF-1 (Amemiya *et al*. 1989), CREB/ATF-1 (Lonze and Guinty, 2002), Activator Protein 1 (AP-1) family members (Sadowska *et al*., 2003), p53 (Ariza *et al*., 1994), Early growth response-1 protein or Egr-1 (Romagnoli *et al*., 2008), Bcl-2-associated athano gene-1 or BAG-1 (Devireddy *et al*., 2000) and CAAT/enhancer binding protein beta or C/EBPβ (Romagnoli *et al*., 2009). In most cases, binding of host transcription factors activates viral gene expression, however, repression of viral gene expression has been reported for AP-1 (Kim *et al*., 2003; Ravichandran *et al*., 2006), C/EBPβ (Romagnoli *et al*., 2009), and NF-1A (Ravichandran and Major, 2008).

Two prominent features within the endemic JCPyV NCCR Box B are the absence of full-length Purα/YB-1 binding pentanucleotide (5'-AGGGAAGGGA-3') (Chen and Khalili, 1995) and the presence of Sp1 binding site (GA Box) (5'-AGG-GAGGAGC-3') (Henson *et al*., 1992) in the same region. These features are also presented by the archetypal CY, Tibetan LH3, Taiwanese Tai3 and the Central-North Indian IN-8 strains. Sp1 has been found to promote early gene transcription in both

glial and non-glial cells (Henson, 1994) and TAg-mediated transactivation of viral late genes. Therefore, the presence of Sp1 binding site in the endemic strain may enable replication of the virus in both glial and non-glial cells. Purα/YB-1 binding site is considered as a switch system to regulate early and late viral gene transcription (Chen and Khalili, 1995; Chen *et al.*, 1995). This site is absent in endemic strains of this region like that of archetype strain CY. The endemic strains also lack additional NF1-binding sites present in the 98 bp repeat of the NCCR of Mad-1 strain. This deletion of NF1 binding sites in the endemic strain may have effect on its ability to replicate in brain and lymphoid tissues.

Cellular tumour suppressor protein p53 is reported to bind to JCPyV large T antigen to repress viral replication (Staib *et al.*, 1996). We have found a di-nucleotide deletion within this p53 binding site of endemic JCPyV strains from Oraon/Munda group as well as NCCR sequences of pregnant women but the deletion was not present in NB6 NCCR which was isolated from the Rabha tribal group. p53 gene is a tumour suppressor gene. Wild-type p53 directly arrests growth of proliferating cells at the G1–S boundary (Diller *et al.*, 1990; Martinez *et al.*, 1991). Inactivation of this gene plays a critical role in malignant transformation. The deletion in the p53 binding site was observed in the strains of Oraon/Munda groups and pregnant women group but the not in the NCCR of Rabha tribal group. The binding site for p53 and SP1 have also been reported from the NCCRs of tumourigenic Merkel Cell Polyomavirus or MCPyV and other novel human PyV isolates (Moens et al., 2020), however, the exact role of these binding sites and the possible interactions of these sites with that of p53 have not been explored yet. However, this deletion or retention of the p53-binding sites within the NCCRs may have a role in viral replication, which needs to be further evaluated by *in vivo* studies.

Putative transcription factor binding sites (TFBSs) on the non-coding control region was also searched *in silico* using families of general core and vertebrate transcription factor matrix using the MatInspector program of the Genomatix software suite (Catharius *et al.*, 2005). POU domain factors, EGR/nerve growth factor, human and murine ETS1 factors, fork head domain factors, Krueppel-like C2H2 zinc finger

factors, homeodomain transcription factors, NeuroD, Nuclear factor 1, Nuclear factor κB, pleomorphic adenoma gene, Purα, SWI/SNF related nucleophosphoproteins, Spalt-like transcription factors, SP1 transcription factors were present in all the NCCR sequences of endemic JCPyV strains (**Table 24**). Few transcription factors such as Neuron-specific Olfactory factor or NOLF, TALE homeodomain class recognizing TG motifs or TALE, X-box binding factor or XBBF and Activator/repressor binding to transcription initiation site or YY1F matched exclusively to either positive (+) or negative (-) strands of NB6 NCCR owing to its sequence differences with that of other endemic NCCRs. Potential roles of some of the transcription factors listed in **Table 24** in viral replication and tissue tropism have been studied in the past. Further investigation is needed to validate the effect of all these transcription factors on JCPyV replication.

5.3 JC virus genotypes based on viral protein VP1 and T-Ag Sequences:

JCPyV genome classification by creating a phylogenetic tree based on variation in the 610 bp IG sequence has identified 12 genotypes- EU, Af1, Af2, Af3, B1-a, B1-b, B1-c, B1-d, B2, CY, MY and SC (Sugimoto *et al*., 1997; Guo *et al*., 1998). Some genotypes were further subdivided (Sugimoto *et al*., 2002). EU type is mainly spread over Europe and Mediterranean. Af2 is spread throughout Africa, West and South Asia, Af1 and Af3 are localized in West and Central Africa. B1-a, B1-b, B1-d, B2, CY, MY and SC are distributed throughout East Asia. Full-length sequencing of JCPyV have identified 8 JCPyV types that were numbered 1 through 8 each with multiple subtypes (Cubitt *et al*., 2001). JCPyV distribution has been used to study migration of human population. It was hypothesized that type 6 is the original JCPyV type (Pavesi, 2003). Type 1 and type 4 are prevalent in Europeans and European-Americans, while type 2A in Asians and Native American populations. Types 3 and 6 are associated with Africans and African-Americans. Types 2D and 7C are found in both Asians and South Asians (Yanagihara *et al*., 2002; Cui *et al*., 2004). Types 2E, 8A, and 8B are found in Western Pacific populations (Yanagihara *et al*.,

2002). Type 8A is detected only in Papua New Guinea populations (Jobes *et al*., 2001).

JCPyV mutates relatively rapidly like mtDNA compared to nuclear genes but slower than other viruses. This low rate of mutation rate and ubiquitous distribution of JCPyV in the human population makes it useful for the study of human migration.

In the present study, phylogenetic tree was created by maximum likelihood method based on both VP1 sequences and T-Ag sequences of endemic JCPyV strains. The analysis revealed that the isolates from Rabha tribes NB7, NB8, NB9, NB10, NB11, NB12, NB13 and NB14 form a distinct group and the other group was formed by the isolates from Oraon/Munda tribal group (NB2 and NB4). NB7, NB8, NB9, NB10, NB11, NB12, NB13 and NB14 clustered with the strains #123, #124, #402, G2 and Mad1 suggesting a close affinity with the EU/Type 1/Type 4 group. EU/Type 1/Type 4 is generally found in Europeans and European-Americans and is mainly spread over Europe and Mediterranean areas. Type 1/Type 4/EU strains have also been found in Japan and South Korea. Detection in population of Japan and South Korea can be the result of ancient colonization by Caucasians. These findings suggest that some distinct groups of Caucasian origin may have migrated to the East Asia during ancient times. Rabha are an indigenous Assamese community mostly residing in Assam, Meghalaya and West Bengal. Rabhas are mostly concentrated in two districts of West Bengal, Jalpaiguri and Coochbehar. Studies based on Y-chromosome haplogroup diversity, KIR gene frequencies showed that the Rabhas clustered with the North East Asians and have proximity towards mongoloid ethnicity (Debnath *et al*., 2011; Guha *et al*., 2015). Hence, the fact that JCPyV has been postulated to have been co-evolving wth the humans. The proximity of the Rabha population to the North East Asian population can be the reason for the presence of Type 1/Type 4/EU genotype in this population.NB2 and NB4 clustered with the strain #230 suggesting an affinity towards Type B1-b/Type 2D group. Types B1-b/Type 2D are found in both Asians and South Asians and are distributed throughout Asia.B1 could have originated in the near East and its extension eastward might have accompanied the migration of protoMongoloids.

Similar types of results were observed during pairwise sequence comparison of VP1 and T-antigen sequences of JCPyV isolates. It was done using DiAlign alignment software of Genomatix suite v2.5 GmbH. The isolates from Rabha tribes NB7, NB8, NB9, NB10, NB11, NB12, NB13 and NB14 were identical. These endemic isolates were most similar to the JCPyV Type 1B i.e. to the #123 strain of JCPyV. NB2 and NB4 VP1 sequences from Oraon/Munda tribes appeared almost identical to the Type 2D sequence.