Indian Heart Journal 73 (2021) 511–515

Contents lists available at ScienceDirect

Indian Heart Journal

journal homepage: www.elsevier.com/locate/ihj

Research Brief

Susceptibility of *CTLA-4* −1661A/G polymorphism towards severity of rheumatic heart disease

Ankit Bansal [a], Sana Tasnim [b], Mohit D. Gupta [a,*], Girish MP [a], Vishal Batra [a], Samantha Kohli [b], Sanjay Tyagi [a], M.A. Qadar Pasha [b]

[a] GB Pant Institute of Post Graduate Medical Education and Research, New Delhi, India
[b] CSIR-Institute of Genomics and Integrative Biology, Delhi, India

ARTICLE INFO

Article history:
Received 11 January 2021
Received in revised form
3 May 2021
Accepted 30 May 2021
Available online 8 June 2021

Keywords:
Acute rheumatic fever
Rheumatic heart disease
Valve lesions
Polymorphism
CTLA-4

ABSTRACT

Aim: Genetic contribution in acute rheumatic fever (ARF)/rheumatic heart disease (RHD) has been suggested but not according to severity of the valve involvement. This study attempts to identify the relevance of *CTLA-4* polymorphism with severity of the disease.
Methods: In a case-control design, 291 healthy controls and 83 patients were genotyped for association between RHD and single-nucleotide polymorphisms −1661A/G of *CTLA-4*.
Results: Segregation of patients on the basis of severity i.e., MVL (Mitral Valve Lesion) and CVL (Combined Valve Lesion) revealed that the frequency of *CTLA-4* −1661G allele depleted as the disease progressed to CVL ($p < 0.05$). Patients in the age group of 31–45 years were significantly more susceptible ($p < 0.046$). Whereas, female patients were more susceptible than the male patients.
Conclusion: Our study suggests the risk associated with decreased frequency of *CTLA-4* −1661G allele in the CVL group and in females.

1. Introduction

Acute rheumatic fever (ARF) is a post-infectious, non-suppurative, multisystemic inflammatory disease that contributes to typical clinical characteristics such as arthritis, chorea and carditis/valvulitis.[1–4] These characteristics lead to a most serious complication called Rheumatic heart disease (RHD) that occurs in about 30–45% of RF patients and that further leads to chronic valvular lesions.[5,6] The pathogenesis of RHD is quite intricate that involves both genetic and environmental factors.[6] ARF and RHD are post-infectious diseases that involve an inflammatory reaction in addition to T and B cells, hence several SNPs in genes that code for inflammatory molecules and contribute to predisposition and manifestation of the disease have been investigated.[1,7–11] However, interestingly, literature is scant on the role of these genes in the susceptibility of the valves and severity of the disease. Among the several genes that have been investigated in RHD, the Cytotoxic T lymphocyte associated antigen-4 (*CTLA-4*) has attracted much attention.

CTLA-4 gene is co-localized on q33 of human chromosome 2. It is a co-stimulatory molecule and is expressed on the surface of activated T-cells, and plays a pivotal role in the inhibition of T-cell activation and peripheral tolerance.[10,11] It is a negative regulator of T-cell activation and alteration of its expression may have a notable effect in immune-mediated diseases.[12] In a multifactorial immune checkpoint system, known as peripheral tolerance, CTLA-4 immune checkpoint pathway regulates the activation of T cells during an immune response to prevent autoimmunity. CD28 molecules on the T cells binds to B7 molecules on antigen presenting cells producing a stimulatory signal. Thus, the activation/inhibition of a T cell depends upon the binding between CD28: B7 and CTLA-4: B7. The upregulation of CTLA-4 on the cell surface is induced as a result of stimulatory signals from T-cell receptor and CD28: B7 binding. The increased binding of CTLA-4: B7 results in a negative signal limiting IL-2 production and proliferation and thus, limiting T cell survival.[13,14]

Two SNPs −318C/T and +49A/G of *CTLA-4* have been studied with respect to RHD[1–3,5]; whereas, a third SNP −1661A/G has been less characterized, although it has been associated with type 1 diabetes mellitus,[15] systemic sclerosis,[16] multiple sclerosis[17] and

* Corresponding author. GB Pant Institute of Post Graduate Medical Education & Research, New Delhi, 110002, India.
E-mail address: drmohitgupta@yahoo.com (M.D. Gupta).

https://doi.org/10.1016/j.ihj.2021.05.002

oral squamous cell carcinoma[18] thus indicating a correlation between this SNP and autoimmune diseases.

2. Materials and methods

2.1. Study participants and clinical evaluation

The human ethics committees of the institutes approved the study protocols and consent form. A total of 83 RHD patients (34 males and 49 females, mean age 36.8 ± 10.2 years), diagnosed on the basis of echocardiography, were included. The control group consisted of 291 healthy unrelated North-Indians (239 males and 52 females, mean age 31.4 ± 7.2 years) without any history of ARF or clinical or echocardiographic evidence of RHD. All the participants were aged between 20 and 60 years. Comprehensive history was recorded, followed by clinical examination and detailed 2D echocardiography. Five ml of blood was drawn in acid citrate dextrose (ACD) anticoagulant in supine position from the subjects after overnight fasting. Plasma was isolated and peripheral blood leucocytes were used for DNA extraction by modified salting out protocol.[19] Plasma was stored at −80 °C and DNA samples at −20 °C, if not used immediately.

2.2. Genotyping

SNPs were selected on the basis of their location, clinical and functional relevance with respect to the other SNPs. The specificity of all the primers was evaluated using nucleotide BLAST (Basic Local Alignment Search Tool). The SNP of *CTLA-4* −1661A/G was screened using SNaPshot ddNTP Primer Extension PCR method (Applied Biosystems, Foster City, USA) in the two groups.

2.3. Statistical analysis

SPSS version 16.0 and EPIINFO version 6.0 software were used for statistical analyses. Genotype and allelic distributions of *CTLA-4* was compared by multivariate logistic regression analysis; the covariates taken into consideration were age and gender. The HWE (Hardy–Weinberg equilibrium) was scrutinized using a $\chi 2$ goodness-of-fit test. The OR (Odds ratio) and 95% CI (Confidence interval) were calculated using multivariate logistic regression analysis and were used to measure the strength for the association of genotypes and their combinations with the disease. Correction for multiple comparisons was performed by false discovery rate (FDR) correction (BenjaminiHochberg.xlsx calculator). The power of the sample size to detect the association at $\alpha = 0.05$ was calculated using an online tool "OSSE-An Online Sample Size Estimator" (link: http://osse.bii.a-star.edu.sg/calculation2.php). A p value of ≤ 0.05 was considered statistically significant after FDR correction.

3. Results

3.1. Population characteristics

The clinical characteristics and echocardiographic findings of the patients are depicted in Table 1. Among the 83 patients, 45% were affected with mitral valve lesion (MVL), while 55% were affected with combined valve lesion (CVL). Also, 25% were diagnosed with tricuspid regurgitation while 23% and 17% were diagnosed with mitral and aortic regurgitation, respectively, in moderate to severe cases.

Table 1
Demographic and clinical characteristics of RHD patients and controls.

Parameter	Control ($n = 291$)	Patients ($n = 83$)
Age, years	31.4 ± 7.1	36.8 ± 10.2
Gender		
Male	239 (82%)	34 (41%)
Female	52 (18%)	49 (59%)
Valvular lesion		
Mitral stenosis		
Mild	–	5 (6%)
Moderate-Severe	–	78 (94%)
Mitral valve lesion	–	37 (45%)
Combined Valvular Lesion	–	46 (55%)
Mitral regurgitation		
Mild	–	51 (77%)
Moderate-Severe	–	15 (23%)
Aortic regurgitation		
Mild	–	38 (83%)
Moderate-Severe	–	8 (17%)
Tricuspid regurgitation		
Mild	–	61 (75%)
Moderate-Severe	–	20 (25%)

n, number of samples. The age is represented as mean ± standard deviation (SD). Each column under controls and patients represents number of samples (% distribution). p-value was calculated using the Epi Info™ software version 6. Significance was maintained at $p \leq 0.05$.

3.2. Single locus association analysis

We analyzed our genotype data for single locus association of *CTLA-4* −1661 A/G by applying 6 genetic models. Under a dominant model, carrying G allele would increase the risk of RHD, where AG and GG are pooled (AA versus AG and GG). While under a recessive model for G allele, AA and AG would be pooled (AA and AG versus GG). An over-dominant model would assume that the heterozygote AG would be of the strongest impact versus AA and GG. The additive model evaluated the impact of A and G alleles individually towards the disease. The co-dominant model evaluated the disease risk associated with AG individuals, the heterozygotes.[20,21]

Plausible association and the respective genotype and allele distributions of SNP −1661A/G of *CTLA-4* was investigated and is represented in Table 2. In our pursuit we first screened these genes for understanding the role in the disease and followed with the valvular association.

3.3. CTLA-4 −1661A/G SNP inclines toward disease susceptibility

The *CTLA-4* −1661A/G SNP revealed a statistically significant association for AG and GG genotype between the two groups ($p = 0.015$) with RHD. Also, on comparing the additive model, G allele revealed a significant association among the groups ($p = 0.012$). The heterozygotes of *CTLA-4* −1661A/G differed between the two groups ($p = 0.064$) as can be seen from Table 3. Importantly, however, the dominant model A/G significantly associated with the CVL subgroup ($p = 0.043$) but marginally with MVL subgroup ($p = 0.070$). The additive model revealed that the G allele distribution was significantly depleted in CVL ($p = 0.036$) and marginally depleted in MVL ($p = 0.059$) when compared with the distribution in controls (Table 3).

In a comparison of age and gender, the −1661 A/G heterozygotes were depleted in all the age group patients, however the middle age group i.e., 31–45 years as more affected ($p = 0.062$). The dominant model for −1661 A/G was significantly associated with RHD in the age group 31–45 as compared to the same age group controls ($p = 0.046$). The additive model with depleted distribution of G allele in the middle age group further strengthened the risk association ($p = 0.041$) (Table 4). The gender-based analysis was

Table 2
Genotype and allele distribution of the *CTLA-4* −1661 A/G gene polymorphisms in healthy controls and RHD patients.

Gene/SNPs	Genetic Model	Genotype/allele	Controls ($n = 291$)	Patients ($n = 83$)	χ2	*p*-value	OR (95%CI)
CTLA-4 −1661A/G (rs4553808)	Co-dominant	−1661AA	224 (77%)	73 (88%)	–	–	Reference
		−1661AG	63 (22%)	10 (12%)	4.95	**0.026**	0.39 (0.17–0.90)
		−1661GG	4 (1%)	0 (0%)	–	–	–
	Dominant	−1661AA	224 (77%)	73 (88%)	–	–	Reference
		−1661AG + GG	67 (23%)	10 (12%)	5.89	**0.015**	0.36 (0.16–0.82)
	Recessive	−1661AA + AG	287 (99%)	83 (100%)	–	–	Reference
		−1661GG	4 (1%)	0 (0%)	–	–	–
	Over dominant	−1661AA + GG	228 (78%)	73 (88%)	–	–	Reference
		−1661AG	63 (22%)	10 (12%)	4.69	**0.030**	0.40 (0.18–0.92)
	Additive	−1661A	511 (88%)	156 (94%)	–	–	Reference
		−1661G	71 (12%)	10 (6%)	6.38	**0.012**	0.40 (0.18–0.92)

n, number of samples (%distribution); SNP, Single nucleotide polymorphism; χ2, Chi-square; OR, odds ratio; CI, confidence interval; CTLA-4, Cytotoxic T-lymphocyte associated protein-4. *p*-value, χ2and odds ratio (OR) were calculated by *performing* multivariate logistic regression analysis after adjustment for age and gender by using SPSS version 16.0. Significance was maintained at $p < 0.05$ after FDR correction.
The p values written in bold are statistically significant values.

Table 3
Single locus association analysis of *CTLA-4* −1661 A/G SNP based on MVL and CVL subgroups.

SNPs	Genetic Model	Genotype/allele	Control ($n = 291$)	Patients ($n = 83$)		*p-value* OR (95%CI)		
				MVL ($n = 37$)	CVL ($n = 46$)	MVL vs Control	CVL vs Control	CVL vs MVL
CTLA-4 −1661A/G	Co-dominant	AA	224 (77%)	32 (86%)	41 (89%)	Reference	Reference	Reference
		AG	63 (22%)	5 (14%)	5 (11%)	0.101	0.064	0.844
						0.40 (0.14–1.20)	0.37 (0.13–1.06)	0.87 (0.22–3.44)
		GG	4 (1%)	0 (0%)	0 (0%)	–	–	–
	Dominant	AA	224 (77%)	32 (86%)	41 (89%)	Reference	Reference	Reference
		AG + GG	67 (23%)	5 (14%)	5 (11%)	0.07	**0.043**	0.844
						0.36 (0.12–1.08)	0.33 (0.12–0.97)	0.87 (0.22–3.44)
	Recessive	AA + AG	287 (98%)	37 (100%)	47 (100%)	Reference	Reference	Reference
		GG	4 (2%)	0 (0%)	0 (0%)	–	–	–
	Overdominant	AA + GG	228 (78%)	32 (86%)	41 (89%)	Reference	Reference	Reference
		AG	63 (22%)	5 (14%)	5 (11%)	0.114	0.071	0.844
						0.42 (0.14–1.23)	0.38 (0.13–1.09)	0.87 (0.22–3.44)
	Additive	A	511 (88%)	69 (93%)	87 (95%)	Reference	Reference	Reference
		G	71 (12%)	5 (7%)	5 (5%)	0.059	**0.036**	0.850
						0.37 (0.13–1.03)	0.34 (0.12–0.93)	0.88 (0.23–3.31)

n, number of samples(% distribution); MVL-Mitral valve lesion; CVL-Combined valve lesion. *p*-value and odds ratio (OR) were calculated by performing multivariate logistic regression analysis after adjustment for age and gender by using SPSS *version* 16.0. Significance was maintained at $p < 0.05$ after FDR correction.
The p values written in bold are statistically significant values.

Table 4
Age-based single locus association analysis of *CTLA-4*−1661 A/G SNP.

SNPs	Genetic Model	Genotype/allele	20–30 years			31–45 years			46–60 years		
			Control ($n = 160$)	Patients ($n = 29$)	*p-value*; OR (95%)	Control ($n = 121$)	Patients ($n = 39$)	*p-value*; OR (95%)	Control ($n = 10$)	Patients ($n = 15$)	*p-value*; OR (95%)
CTLA-4 −1661A/G	Co-dominant	AA	122 (76%)	25 (86%)	Reference	95 (78%)	35 (90%)	Reference	7 (70%)	13 (87%)	Reference
		AG	35 (22%)	4 (14%)	0.264	25 (21%)	4 (10%)	0.062	3 (30%)	2 (13%)	0.233
					0.50 (0.15−1.69)			0.32 (0.10−1.06)			0.21 (0.02−2.72)
		GG	3 (2%)	0 (0%)	–	1 (1%)	0 (0%)	–	0 (0%)	0 (0%)	–
	Dominant	AA	122 (76%)	25 (86%)	Reference	95 (78%)	35 (90%)	Reference	7 (70%)	13 (87%)	Reference
		AG + GG	38 (24%)	4 (14%)	0.237	27 (22%)	4 (10%)	**0.046**	3 (30%)	2 (13%)	0.233
					0.48 (0.15−1.61)			0.29 (0.09−0.98)			0.21 (0.02−2.72)
	Recessive	AA + AG	157 (98%)	29 (100%)	Reference	120 (99%)	39 (100%)	Reference	10 (100%)	15 (100%)	Reference
		GG	3 (2%)	0 (0%)	–	1 (1%)	0 (0%)	–	0 (0%)	0 (0%)	–
	Overdominant	AA + GG	125 (78%)	25 (86%)	Reference	96 (79%)	35 (90%)	Reference	7 (70%)	13 (87%)	Reference
		AG	35 (22%)	4 (14%)	0.272	26 (21%)	4 (10%)	0.07	3 (30%)	2 (13%)	0.233
					0.51 (0.15−1.70)			0.33 (0.10−1.10)			0.21 (0.02−2.72)
	Additive	A	279 (87%)	54 (93%)	Reference	216 (89%)	74 (95%)	Reference	17 (85%)	28 (93%)	Reference
		G	41 (13%)	4 (7%)	0.237	28 (11%)	4 (5%)	**0.041**	3 (15%)	2 (7%)	0.251
					0.50 (0.16−1.57)			0.31 (0.10−0.95)			0.26 (0.03−2.58)

n, number of samples (% distribution). *p*-value and odds ratio (OR) were calculated by performing multivariate logistic regression analysis after adjustment for gender by using SPSS version 16.0. Significance was maintained at $p < 0.05$ after FDR correction.
The p values written in bold are statistically significant values.

another revelation with the frequency of G allele being significantly higher in females compared to their counterparts. A significant difference was observed in the frequencies of the dominant genotype model i.e. AA versus AG + GG of *CTLA-4* −1661A/G ($p = 0.049$) as well as the additive model's G allele ($p = 0.040$) in the female subgroup; the same models although showed similar trend in males but it did not reach significance suggesting that females were at higher risk of RHD with the lower distribution of the G allele (Supplementary Table 1).

We next did extensive association study of this SNP −1661A/G with the clinical parameters that were pre-specified into various sub-groups. Our observations were no doubt interesting; however, we could note that the smaller sample size due to distribution into several groups depleted the power.

4. Discussion

Population based studies on prevalence of RHD in Indian population and also the role of inflammation in this disease have been conducted, in contrast no concerted efforts have been made to understand the genetic contribution; whereas, predisposing genetic markers of RHD are anticipated to be strongly effective in the detection of susceptible individuals. This study is an attempt in this direction and hence, screened −1661A/G SNP of *CTLA-4* for association with RHD in North-Indians.

Of interest, a detailed evaluation for association with respect to MVL and CVL, gender, age and clinical parameters of RHD patients revealed encouraging results for the *CTLA-4* −1661A/G SNP; the distribution of G allele in controls and the subgroups of patients made the difference. While the AG and GG genotype of *CTLA-4* −1661A/G showed a marginal association with MVL, the association with CVL subgroup of patients was statistically significant suggesting that absence of G allele increased the susceptibility toward the disease. It was equally interesting to note that the G allele followed a trend that was visible with frequency further decreasing in the upper age group i.e., 46–60 years; however, it could not be taken into consideration due to poor number of subjects in both controls and patients. It nevertheless suggested that the risk of disease increased with increasing age. Similarly, in the gender-based analysis the females were at higher risk of RHD with the lower distribution of the G allele. We attribute the risk of RHD to G allele, however, to counterfeit the very statement it may be said that the G allele, because of its greater distribution in healthy subjects or controls associates with lesser risk of RHD.

We do realize that the −1661A/G being a promoter SNP may play a relevant role in the gene regulation thereby in regulating the physiological outcome. Studies based on chromatin immunoprecipitation and electrophoretic mobility shift assays have reported that a transcription factor c/EBPβ binds specifically at position −1661 of the CTLA-4 gene and regulates its expression in the presence of the G allele rather than the A allele. Factor c/EBPβ participates in the CD152 (a CTLA-4 receptor) transcription as its binding activity is instigated after activation of T cell when the membrane form of CD152 is distinctly increased. In addition, an allelic variant may also contribute epigenetically by regulating the gene expression quantitatively or qualitatively by altering transcription factor binding sites or other controlling domains.[22,23] Hence, *CTLA-4* −1661A/G SNP may have a significant role in pathophysiology of RHD.

To conclude, to the best of our knowledge, this is the first study concerning an association of the *CTLA-4* −1661A/G especially with the various clinical parameters (MVL and CVL) of RHD associating with the minor allele. Our findings are interesting and may have long-lasting consequences. Overall, the present study emerged relevant with *CTLA-4* −1661A/G SNP as a potent candidate for RHD susceptibility in North Indian population.

The limitation of the present study that it is a single gene and single SNP investigation, while several genes could contribute towards RHD progression. The outcome also needed further investigations in diversified global populations with a larger sample size.

Declaration of competing interest

All the authors state that they have no conflict interest to declare with respect to the present article titled “Susceptibility of CTLA-4 −1661A/G polymorphism towards severity of Rheumatic Heart Disease”.

Acknowledgements

We appreciate the cooperation of all the volunteers who participated in the study and also that of CSIR-IGIB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ihj.2021.05.002.

References

1. Duzgun N, Duman T, Haydardedeoglu FE, Tutkak H. Cytotoxic T lymphocyte-associated antigen-4 polymorphism in patients with rheumatic heart disease. *Tissue Antigens*. 2009;74:539–542.
2. Makrexeni ZM, Pepeta L. Cardiovascular Topics Clinical presentation and outcomes of patients with acute rheumatic fever and rheumatic heart disease seen at a tertiary hospital setting in Port Elizabeth, South Africa. *Cardiovasc J Afr*. 2017;28:248–250.
3. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis*. 2005;5:685–694.
4. Guilherme L, Kohler KF, Kalil J. Rheumatic heart disease: genes, inflammation and autoimmunity. *Rheumatol Curr Res*. 2012. S4:001.
5. Carapetis JR, McDonald M, Wilson NJ. Acute rheumatic fever. *Lancet*. 2005;366:155–168.
6. Carapetis JR. Rheumatic heart disease in Asia. *Circulation*. 2008;118:2748–2753.
7. Faé KC, Oshiro SE, Toubert A, et al. How an autoimmune reaction triggered by molecular mimicry between streptococcal M protein and cardiac tissue proteins leads to heart lesions in rheumatic heart disease. *J Autoimmun*. 2005;24:101–109.
8. Shen YC, Yang Ting, Wan Chun, et al. Tumor necrosis factor-alpha polymorphisms and rheumatic heart disease risk: a meta-analysis. *Int J Cardiol*. 2013;168:2878–2880.
9. Liu J, Zhang H. −1722T/C polymorphism (rs733618) of CTLA-4 significantly associated with systemic lupus erythematosus (SLE): a comprehensive meta-analysis. *Hum Immunol*. 2013;74:341–347.
10. Lee YH, Choi SJ, Ji JD, Song GG. CTLA-4 and TNF-α promoter −308 A/G polymorphisms and ANCA-associated vasculitis susceptibility: a meta-analysis. *Mol Biol Rep*. 2012;39:319–326.
11. Gough SC, Walker LS, Sansom DM. CTLA4 gene polymorphism and autoimmunity. *Immunol Rev*. 2005 Apr;204:102–115.
12. Walker LS. Treg and CTLA-4: two intertwining pathways to immune tolerance. *J Autoimmun*. 2013 Sep;45(100):49–57.
13. Parry RV, Chemnitz JM, Frauwirth KA, et al. CTLA-4 and PD-1 receptors inhibit T-cell activation by distinct mechanisms. *Mol Cell Biol*. 2005;25:9543–9553.
14. Buchbinder EI, Desai A. CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition. *Am J Clin Oncol*. 2016;39:98–106.
15. Bouqbis L, Izaabel H, Akhayat O, et al. Association of the CTLA4 promoter region (−1661G allele) with type 1 diabetes in the South Moroccan population. *Gene Immun*. 2003;4:132–137.
16. Almasi S, Erfani N, Mojtahedi Z, Rajaee A, Ghaderi A. Association of CTLA-4 gene promoter polymorphisms with systemic sclerosis in Iranian population. *Gene Immun*. 2006;7:401–406.
17. Yousefipour G, Erfani N, Momtahan M, Moghaddasi H, Ghaderi A. *CTLA4* exon 1 and promoter polymorphisms in patients with multiple sclerosis. *Acta Neurol Scand*. 2009;120:424–429.
18. Kämmerer PW, Toyoshima Takeshi, Schöder Fabian, et al. Association of T-cell regulatory gene polymorphisms with oral squamous cell carcinoma. *Oral Oncol*. 2010;46:543–548.

19. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16, 1215.
20. Horita N, Kaneko T. Genetic model selection for a case-control study and a meta-analysis. *Meta Gene*. 2015;5:1–8.
21. Lewis CM. Genetic association studies: design, analysis and interpretation. *Brief Bioinf*. 2002;3:146–153.
22. Wang X, Pirskanen R, Giscombe R, Lefvert AK. Two SNPs in the promoter region of the CTLA-4 gene affect binding of transcription factors and are associated with human myasthenia gravis. *J Intern Med*. 2008;263:61–69.
23. Ling V, Wu PW, Finnerty HF, et al. Complete sequence determination of the mouse and human CTLA4 gene loci: cross-species DNA sequence similarity beyond exon borders. *Genomics*. 1999;60:341–355.