

# Identifying candidate genes for the regulation of the response to Trypanosoma congolense infection

# Introduction

Agaba M<sup>2</sup>

Noyes HA<sup>1</sup>

Ogugo M<sup>2</sup>

Brass A<sup>3</sup>

Anderson S<sup>4</sup>

Archibald A<sup>4</sup>

Hulme H<sup>3</sup>

Kemp SJ<sup>1</sup>

<sup>1</sup>School of Biological Sciences University of Liverpool **Crown Street** Liverpool L69 7ZB

<sup>2</sup>International Livestock Research Institute (ILRI)

African cattle breeds differ significantly in their ability to survive low to moderate levels of challenge with *Trypanosoma congolense*. Similarly the survival times of inbred mouse strains vary substantially after infection. We have previously identified three regions of the mouse genome that regulate survival after infection in two crosses (A/J x C57BL/6 and Balb/c x C57BL/6) (Kemp et al 1997 Nature Genetics). These have been designated *Tir1*, *Tir2* and *Tir3* for Trypanosoma infection response loci on mouse chromosomes 17, 5 and 1 respectively. We have now used two strategies to reduce the size of the regions that appear to be regulating survival. Firstly, congenic mice lines carrying defined regions of the C57BL/6 genome on an A/J background were developed to identify physical boundaries of the regions and confirm its effect. Secondly the response to infection has been mapped in an additional mouse strain (129J). The mapping data has been combined with haplotype maps to identify a 70kb high priority region containing just 5 strong candidates for the causative gene for resistance to trypanosomiasis in mice on chromosome 17.

### **Congenic Mice**

Congenic lines are created by crossing resistant C57BL/6 with the susceptible A/J mice. At each generation the offspring are genotyped to identify those animals that are carrying alleles from C57BL/6 in the target region of interest and these mice are then selected to be backcrossed to the recipient genome. After seven generations of backcrossing to A/J, heterozygous carriers of the C57BL/6 donor region of interest were intercrossed and homozygous carriers were selected as the congenic line and were designated TirnCC in this study. Homozygous carriers of the A/J alleles were selected as the control line and are designated TirnAA. The creation of these mouse lines makes it possible to study the effect of each locus in isolation from the other loci and most background effects.

The positions of the regions of the C57BL/6 genome that were introgressed into the A/J background were determined by genotyping the three lines with the Illumina 1536 SNP marker panel (Figure 1). The introgressed regions on chromosome 5 and 17 had a significant effect on survival (Figure 2) but the C57BL/6 region on chromosome 1 had no effect on survival indicating that the genes regulating the response to infection are elsewhere on this chromosome.

## Mapping loci controlling the response to infection in 129J mice

Identifying the regions controlling the response to infection in additional mouse strains make it possible refine the list of candidate genes that might regulate survival time.

F2 C57BL/6 x 129J were bred by crossing 129J and C57BL/6 mice to create and F1 generation and then by intercrossing their offspring to create an F2. 135 F2 C57BL/6 x 129J mice were genotyped with the Illumina 384 SNP mapping panel and the data was analysed with the JQTL package (Figure 3). The data confirmed the presence of QTL on chromosomes 1 and 17 but there was no evidence of a QTL on chromosome 5 suggesting that 129J might carry the C57BL/6 allele at this locus.

The locus on chromosome 17 was significantly associated with survival. The locus on distal chromosome 1 was not significant after correction for multiple testing, but since it was supported by multiple SNP markers and it coincides with previously identified loci in A/J and Balb/c mice, it may be real but the effect size of the alleles may be insufficient to show a significant effect in this relatively small panel of mice.



Figure 3. Plot of LOD scores for markers on chromosomes 1, 5 and 17. Showing evidence of QTL on chromosomes 1 and 17 but



<sup>4</sup>Roslin Institute **Roslin BioCentre** Midlothian **EH25 9PS** 

PO Box 30709

Nairobi 00100



Figure 1. Positions of QTL are shown by red arrows. Positions of regions of C57BL/6 genome introgressed into the A/J genome are shown by the coloured blocks



Figure 2. Survival times. The plots show that the Tir1CC mice carrying C57BL/6 DNA at the chromosome 17 locus survived longer than the Tir1AA mice carrying A/J DNA at this locus. Also the Tir2CC mice carrying C57BL/6 DNA at the chromosome 5 locus survived longer than the Tir2AA mice carrying A/J DNA at this locus. There was no difference in survival of the Tir3CC and Tir3AA mice (not shown) indicating that the genes carrying resistance alleles on chromosome 1 are outside this region.

# Haplotype analysis

The availability of mapping data from three pairs of Description mouse strains makes it possible to look for

correlation between haplotypes across the QTL regions and response to infection. The genome of inbred mouse strains is believed to be composed of mosaic of regions derived from three or four ancestral strains. Given that we have observed the some QTL in multiple pairs of mouse strains it is likely that the polymorphisms that make C57BL/6 more resistant are derived from one of those ancestral strains. If the ancestral strain from which gene is inherited can be identified then it should be trefoil factor 2 (spasmolytic protein 1) [Source:MarkerSymbol;Acc:MGI:104638] possible to produce a short list of candidate genes for assessment and a start a start a start and a start which C57BL/6 inherits its allele from one strain and dial spoke head 1 homolog (Chlamydomonas) [Source:MarkerSymbol;Acc:MGI:1926074] ash3a Solute carrier family 37 (glycerol-3-phosphate transporter), member 1 [Source:MarkerSymbol;Acc:MGI:194909] Rsph1 A/J, Balb/c and 129J inherit their alleles from a different strain.

The recent publication of 8 million SNP in the fiftee pox/knotted 1 homeobox [Source:MarkerSymbol;Acc:MGI:1201409] most widely used mouse strains makes it possible to U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) 1 [Source Marker Symbol: Acc: MGI:88515] identify the haplotype of origin of most regions of the mouse genome (Frazer et al. 2007 Nature). http://mouse.perlegen.com/mouse/mousehap.html We have used this data to assign each gene to an ancestral haplotype (Figure 4). Only five genes were RIKEN CDNA A430107D22 gene [Source:MarkerSymbol;Acc:MGI:2444128] on the same ancestral haplotype in 129J, A/J and Balb/c and on a different haplotype in C57BL/6. These genes are therefore strong candidates for the gene that regulates the difference in survival after infection, Quantative Trait Genes (QTG). An additional 19 genes were on different haplotypes in 129J, A/J and Balb/c but also all different from C57BL/6. These genes are also possible candidate QTG.

It is as important to exclude genes as include them and the addition of data from the  $129J \times C57BL/6$ cross makes it possible to exclude the class 1 and 2 MHC genes. Since the chromosome 17 QTL overla the the MHC region these were natural candidate genes. 129J and C57BL/6 share the same MHC haplotype (b) as can be seen in figure 4.

#### ID C57 AJ Balb 129 peptidase inhibitor 16 [Source:MarkerSymbol;Acc:MGI:1921366] Pi16 Mtch1 mitochondrial carrier homolog 1 (C. elegans) [Source:MarkerSymbol;Acc:MGI:1929261] FYVE, RhoGEF and PH domain containing 2 [Source:MarkerSymbol;Acc:MGI:1347084] Fgd2 proviral integration site 1 [Source:MarkerSymbol:Acc:MGI:97584] TBC1 domain family, member 22B [STbc1d22b Aring finger protein 8 [Source:MarkerSymbol;Acc:MGI:1929069] RIKEN cDNA 1110021J02 gene [Source:MarkerSymbol;Acc:MGI:1915847] 1300018I05R 1110021J02Rik Mdga1 SNORA70 Small nucleolar RNA SNORA70 [Source: RFAM 8.0] [Source: RFAM 8.0] zinc finger, AN1-type domain 3 [Source:MarkerSymbol;Acc:MGI:1096572] Zfand3 BTB (POZ) domain containing 9 [Source:MarkerSymbol;Acc:MGI:1916625] Btbd9 predicted gene, ENSMUSG00000052388 [Source:MarkerSymbol;Acc:MGI:3642006] ENSMUSG0000 glyoxalase 1 [Source:MarkerSymbol;Acc:MGI:95742] dynein, axonemal, heavy chain 8 [Source:MarkerSymbol;Acc:MGI:107714] U6 spliceosomal RNA [Source: RFAM 8.0] [Source: RFAM 8.0] Dnahc8 predicted gene, ENSMUSG0000054148 [Source:MarkerSyr ENSMUSG0000 glucagon-like peptide 1 receptor [So Glp1r predicted gene, ENSMUSG00000073427 [Source:MarkerSymbol;Acc:MGI:3642784 uromodulin-like 1 [Source:MarkerSymbol;Acc:MGI:1929785] ATP-binding cassette, sub-family G (WHITE), member 1 [Source:Mark Umodl1 ool;Acc:MGI:10 Abcg1 Tff2 Tff1 phosphodiesterase 9A [Source:MarkePde9a predicted gene, ENSMUSG000000532ENSMUSG WD repeat domain 4 [Source:MarkerSymbol:Acc:MGI:1889002] Wdr4 RIKEN cDNA 1500032D16 gene [Source:MarkerSymbol;Acc:MGI:1925580 1500032D16Ri RIKEN cDNA 4833413E03 gene [Source:MarkerSymbol;Acc:MGI:1925354] 4833413E03Ri Pknox1 athionine beta-synthase [Source:MarkerSymbol;Acc:MGI:88285] Cbs MG U2af1 crystallin, alpha A [Source:MarkerSymbol;Acc:MGI:88515 Cryaa Snf1lk SNF1-like kinase [Source:MarkerSymbol;Acc:MGI:104754] heat shock transcription factor 2 binding protein [Source:MarkerSymbol;Acc:MGI:1921627]Hsf2bp ribosomal RNA processing 1 homolog B (S. cerevisiae) [Source:MarkerSymbol;Acc:MGI:191 Rrp1b Notch gene homolog 3 (Drosophila) [Source:MarkerSymbol;Acc:MGI:99460 Notch3 abhydrolase domain containing 9 [Source:MarkerSymbol;Acc:MGI:1919182] Abhd9 bromodomain containing 4 [Source:MarkerSymbol;Acc:MGI:1888520] Brd4 A kinase (PRKA) anchor protein 8 [Source:MarkerSymbol;Acc:MGI:1928488] Akap8 A kinase (PRKA) anchor protein 8-like [Source:MarkerSymbol;Acc:MGI:1860606 Akap8 widely-interspaced zinc finger motifs [Source:MarkerSymbol;Acc:MGI:1332638] Wiz A430107D22R RIKEN cDNA A530088E08 gene [Source:MarkerSymbol;Acc:MGI:3603459] A530088E08R cytochrome P450, family 4, subfamily f, polypeptide 39 [Source:MarkerSymbol:Acc:MGI:244Cyp4f39 ome P450, family 4, subfamily f, polypeptide 16 [Source:MarkerSymbol;Acc:MGI:191Cyp4f1

cytochrome P450, family 4, subfamil Cyp4f16 Cvp4f4 cytochrome P450, family 4, subfamily f, polypeptide 15 [Source:MarkerSymbol;Acc:MGI:214Cyp4f15

RIKEN cDNA 9030612M13 gene [Source:MarkerSymbol;Acc:MGI:1921793] 9030612M13 zinc finger protein 811 [Source:MarkerSymbol;Acc:MGI:2682944] Zfp811 U1 spliceosomal RNA [Source: RFAM 8.0] [Source: RFAM 8.0] 6030490I01R<mark>i</mark> RIKEN cDNA 6030490I01 gene [Source:MarkerSymbol:Acc:MGI:2443934]

cDNA sequence BC066107 [Source:MarkerSymbol;Acc:MGI:3029586] BC066107 cytochrome P450, family 4, subfamily f, polypeptide 14 [Source:MarkerSymbol;Acc:MGI:192Cyp4f14 cytochrome P450, family 4, subfamily f, polypeptide 13 [Source:MarkerSymbol;Acc:MGI:215Cyp4f13 cytochrome P450, family 4, subfamily f, polypeptide 41 pseudogene [Source:MarkerS bo Cyp4f41-ps zinc finger protein 472 [Source:MarkerSymbol;Acc:MGI:2385049] Zfp472 RIKEN cDNA C920016K16 gene [Source:MarkerSymbol;Acc:MGI:2441928 C920016K16<mark>Ril</mark> RIKEN cDNA 1700065013 gene [Source:MarkerSymbol;Acc:MGI:1920701] 1700065013Ri PHD finger protein 8 [Source:RefSeq\_peptide;Acc:NP\_001009544] NP\_001009544 zinc finger protein 563 [Source:MarkerSymbol;Acc:MGI:2677168] Zfp563 microrchidia 2B [Source:MarkerSymbol;Acc:MGI:3045293] Morc2b predicted gene, EG383232 [Source:MarkerSymbol;Acc:MGI:3643505] EG383232 zinc finger protein 422, related sequence 1 [Source:MarkerSymbol;Acc:MGI:3028594] Zfp422-rs1

	zinc finger protein 81 [Source:MarkerSymbol;Acc:MGI:1890752]	Zfp81
	zinc finger protein 101 [Source:MarkerSymbol;Acc:MGI:107547]	Zfp101
	RIKEN cDNA 1700029I08 gene (1700029I08Rik), mRNA [Source:RefSeq_dna;Acc:NM_1832	NM_183282
	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 moti	Adamts10
)	myosin IF [Source:MarkerSymbol;Acc:MGI:107711]	Myo1f
1	zinc finger protein 414 [Source:MarkerSymbol;Acc:MGI:1915641]	Zfp414
	PML-RAR alpha-regulated adaptor molecule 1 [Source:MarkerSymbol;Acc:MGI:3576625]	Pram1
ır	Sembrane-associated ring finger (C3HC4) 2 [Source:MarkerSymbol;Acc:MGI:1925915]	Mar-0
T	RAB11B, member RAS oncogene family [Source:MarkerSymbol;Acc:MGI:99425]	Rab11b
	angiopoietin-like 4 [Source:MarkerSymbol;Acc:MGI:1888999]	Angptl4
	ankyrin repeat domain 47 [Source:MarkerSymbol;Acc:MGI:1098615]	Ankrd47
	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 7 (B14.5a) [Source:MarkerSymbo	Ndufa7
	CD320 antigen [Source:MarkerSymbol;Acc:MGI:1860083]	Cd320
	Fas death domain-associated protein [Source:MarkerSymbol;Acc:MGI:1197015]	Daxx
	zinc finger and BTB domain containing 22 [Source:MarkerSymbol;Acc:MGI:1931870]	Zbtb22
	TAD binding protoin [Courses MarkerCurshelsAcc.MCI.1201600]	Teehe

#### UNIVERSITY OF LIVERPOOL

27.7 Mb

31.0 Mb

ILRI

MANCHESTER 1824



nucleotide Single morphsims are not the only form of genetic variation. Regions of the mouse genome can be duplicated or deleted and are known as copy number variants (CNV). CNV between C57BL/6 and A/J mice were detected using Agilent 240k whole genome CGH arrays. An amplifcation of the Glo1 gene within the Tir1 QTL on chromosome 17. This gene was not identified by the haplotype analysis but expression analysis indicated that the gene was more highly expressed in A/J and Balb/c consistent with copy number.

Consequently it seems unlikely that the classical MHC genes are responsible for the large difference MHC genes are responsible for the large difference

		. Bogaiei	
6 I I I I I I I I I I I I I I I I I I I	ribosomal protein S18 [Source:MarkerSymbol;Acc:MGI:98146]	Rps18	
in survival time associated with this locus	vacuolar protein sorting 52 (yeast) [Source:MarkerSymbol;Acc:MGI:1330304]	Vps52	
In survival time associated with this locus.	predicted gene, EG630499 [Source:RefSeq_peptide;Acc:NP_001074484]	NP_0010744 <mark>84.1</mark>	
	expressed sequence AA388235 (AA388235), mRNA [Source:RefSeq_dna;Acc:NM_0010137	7 <b>9\GP</b> _0010138 <mark>15.1</mark>	
	histocompatibility 2, K1, K region [Source:MarkerSymbol;Acc:MGI:95904]	H2-K1	
	ring finger protein 1 [Source:MarkerSymbol;Acc:MGI:1101770]	Ring1	
	mmu-mir-219-1 [Source:miRBase;Acc:MI0000702]	mmu-mir-219 <mark>-1</mark>	
	H2-K region expressed gene 6 [Source:MarkerSymbol;Acc:MGI:95911]	H2-Ke6	
	solute carrier family 39 (zinc transporter), member 7 [Source:MarkerSymbol;Acc:MGI:9590	199jlc39a7	
	retinoid X receptor beta [Source:MarkerSymbol;Acc:MGI:98215]	Rxrb	
	<del>pr</del> φcollagen, type XI, alpha 2 [Source:MarkerSymbol;Acc:MGI:88447]	Col11a2	
Figure 4 Haplotypes of C57DI /6 AI Dalb/a and 120			
Figure 4. haplotypes of C5/bL/0, AJ, balo/c and 129	histocompatibility 2, O region alpha locus [Source:MarkerSymbol;Acc:MGI:95924]	H2-Oa	
	bromodomain containing 2 [Source:MarkerSymbol;Acc:MGI:99495]	Brd2	
across the firl OIL on chromosome 17. The ancestra	histocompatibility 2, class II, locus Mb1 [Source:MarkerSymbol;Acc:MGI:95922]	H2-DMb1	
······	histocompatibility 2, class II, locus Mb2 [Source:MarkerSymbol;Acc:MGI:95923]	H2-DMb2	
haplotype of each strain is indicated by the coloured protessme (prosome, macropain) subunit, beta type 9 (large multifunctional pepti			
haplotype of each strain is indicated by the coloured	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP) [Source:MarkerSymbol;Acc:M	(Tap1	
blocks to the might of the same names C57DI /6 is	proteasome (prosome, macropain) subunit, beta type 8 (large multifunctional peptidase 7	/ Psmb8	
blocks to the right of the gene names. C5/BL/01s	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP) [Source:MarkerSymbol;Acc:Mi	(Tap2	
	histocompatibility 2, O region beta locus [Source:MarkerSymbol;Acc:MGI:95925]	H2-Ob	
always shown in vellow. Where the other strains diffe	nigtocompatibility 2, class II antigen A, beta 1 [Source:MarkerSymbol;Acc:MGI:103070]	H2-Ab1	
	-ingtocompatibility 2, class II antigen A, aipna [Source:MarkerSymbol;Acc:MGI:95895]	HZ-Aa	
in ancestral hanlotung they are shown in red green or	histocompatibility 2, class II antigen E beta [Source:MarkerSymbol;Acc:MGI:95901]	HZ-ED1	
in ancesular naprocype mey are shown in red, green or	RIKEN CDNA AI 30038H09 gene [Source:MarkerSymbol;Acc:MGI:3588187]	A130038H09Rik	
11	nistocompatibility 2, class II antigen E alpha [Source:MarkerSymbol;Acc:MGI:95900]	HZ-Ea	
blue.	Duryrophilin-like 2 [Source:MarkerSymbol;Acc:MGI:1859549]	Btni2	
	Ducyropniin-like 1 [Source:MarkerSymbol;Acc:MGI:1932027]	BCUIT	

### **Conclusions**

The combination of congenic mice, additional mapping data and discovery of copy number variations within the QTL regions has made it possible to identify a short list of genes that might regulate the response to infection with *T. congolense*. This list is now sufficiently short that it is practicable to undertake detailed studies on the role of individual genes in the response to infection and hence determine whether they cause the difference in survival time after infection. The identification of these genes is expected to give an insight into the pathways that regulate the response to infection and may lead to new approaches to treatment.



Acknowledgements: We thank John Wambugu, Moses Ogugo and Leanne Wardlesworth for technical assistance. These studies were funded by the Wellcome Trust

poly-



H2-Ke2 Wdr46