

Lezione XXXIII-XXXIV

giovedì 15-XII-2011



D. Frezza corso di genomica

Laurea Magistrale Biotecnologia Industriale

aula 8

orario : Martedì ore 14.00 - 16.00

Giovedì ore 13.00 - 15.00

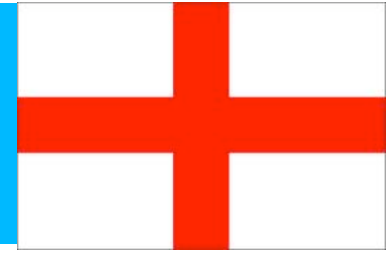
Esami: I appello Martedì 31 Gennaio

II appello Martedì 7 Febbraio

III appello Martedì 28 Febbraio,

A gennaio esercitazioni se arrivano i soldi per farle

somatic retrotransposition



Somatic retrotransposition alters the genetic landscape of the human brain

J. Kenneth Baillie^{1*}, Mark W. Barnett^{1*}, Kyle R. Upton^{1*}, Daniel J. Gerhardt², Todd A. Richmond², Fioravante De Sapio¹, Paul Brennan³, Patrizia Rizzu⁴, Sarah Smith¹, Mark Fell¹, Richard T. Talbot¹, Stefano Gustincich⁵, Thomas C. Freeman¹, John S. Mattick⁶, David A. Hume¹, Peter Heutink⁴, Piero Carninci⁷, Jeffrey A. Jeddloh² & Geoffrey J. Faulkner¹

Nature. 2011 Oct 30;479(7374):534-7. doi: 10.1038/nature10531.

danni genomici somatici



Invecchiamento e trasformazione di cellule maligne sono associate ad accumulo di mutazioni deleterie con perdita di funzione, morte cellulare o crescita cellulare incontrollata. La retrotrasposizione porta danni per mutagenesi. Nella popolazione umana è stimata la presenza di 400×10^6 varianti per retrotrasposizione e più di 70 patologie associate ereditabili o “de novo” da retrotrasposizioni. Per questo forse i retrotrasposoni competenti per la trasposizione sono fortemente metilati e trascrizionalmente repressi.

riattivazione di retrotra-sposoni

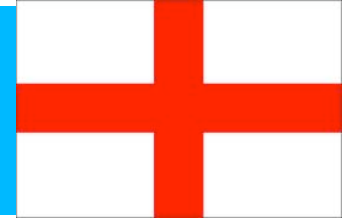


nelle linee cellulari neuronali nonostante la forte repressione sono state trovate retrotrasposizioni nuove.



data la complessa struttura ed organizzazione del cervello dei mammiferi: adattabilità, plasticità e capacità rigenerative, malattie degenerative e disordini mentali con eziologia non risolta,
potrebbe essere importante risolvere e capire i meccanismi di queste inserzioni somatiche.

una possibile causa della riattivazione

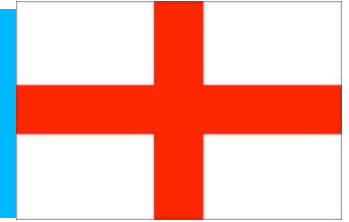


L'attività trasposizionale nel cervello può essere spiegata dal fatto che il promotore di L1 può essere temporaneamente distolto dalla soppressione epigenetica durante la neurogenesi (demetilato ecc.).

I trasposoni L1 capaci di trasposizione (completi) possono muoversi ripetutamente in luoghi diversi in singole cellule dando origine ad un mosaicismo somatico.

Varie evidenze convalidano questo modello con la trascrizione di L1: la variazione del numero di copie (CNV) nei tessuti di cervello umani di donatori di varie età, la mobilità di retrotrasposoni L1 modificati in vitro ed in roditori transgenici.

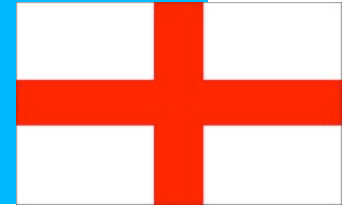
sconosciuti i bersagli



Non si sa dove avvengano le inserzioni nel genoma e dato che la cromatina aperta è suscettibile a questi eventi di integrazione di L1, potrebbe esserci una preferenzialità di integrazioni in loci codificanti proteine ed espressi nel cervello.

Mappare gli eventi di retrotrasposizione che collettivamente formano un mosaico somatico è complesso dato che gli eventi sono rari in una popolazione di cellule eterogenea. Per trovarli si dovrebbero clonare le cellule per avere gruppi isogenici omogenei di cellule.

Metodo di individuazione



Sviluppo di un protocollo globale ad alta risoluzione chiamato: cattura e sequenziamento di retrotrasposoni (RC-seq) “retrotransp. capture sequencing.”

- si frammenta il DNA genomico e si ibrida su array commerciale di cattura di sequenze con bersagli delle estremità 5' e 3' degli interi retrotrasposoni L1, Alu ed SVA (figura 1 e tavole suppl. 1 e 2).

Le lunghe terminal repeats immobili di ERVK ed ERVI servono da controllo negativo.

- i DNA catturati sequenziati ripetutamente, ~25 milioni di “reads” (sequenze risolte) con estremità appaiate (complete) di 101 basi per campione,
- le estremità sono mappate usando un algoritmo restrittivo per identificare i retrotrasposoni noti e nuovi (supl fig 1a-d), le nuove inserzioni con estremità uniche che mappano (da una parte sola).

leggenda fig 1 metodo

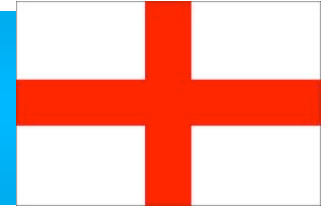


Figure 1 | Overall RC-seq methodology.

a, Retrotransposon capture: sheared genomic DNA is hybridized to custom tiling arrays probing full-length retrotransposons (nucleotides highlighted with light-blue background).

[DNA genomico tagliato è ibridato su array scelto che copra l'intero retrotrasposone]

b, Sequencing: after hybridization, DNA fragments are eluted and analysed with an Illumina sequencer, producing 2.5×10^7 paired-end reads per library that are subsequently aligned to the reference genome.

[dopo l'ibridazione i frammenti di DNA vengono eluiti e analizzati con il sequenziatore "Illumina" ottenendo 2.5×10^7 (letture) sequenze con estremità appaiate per library (da tipo di cellule) che vengono dopo allineate sulla sequenza genomica di riferimento.]

c, Reads mapping as a pair to a single locus indicate known retrotransposon insertions.

[sequenze che si appaiano con tutte e due le estremità ad un locus unico indicano la presenza di una inserzione già nota]

d, Unpaired reads where one end maps to a single locus and the other end maps to a distal retrotransposon indicate novel retrotransposition events.

[sequenze non appaiate che mappano con una estremità solo su un locus e con l'altro estremo su un altro retrotrasposone distante, indicano un nuovo evento di retrotrasposizione.]

supl. Fig 1



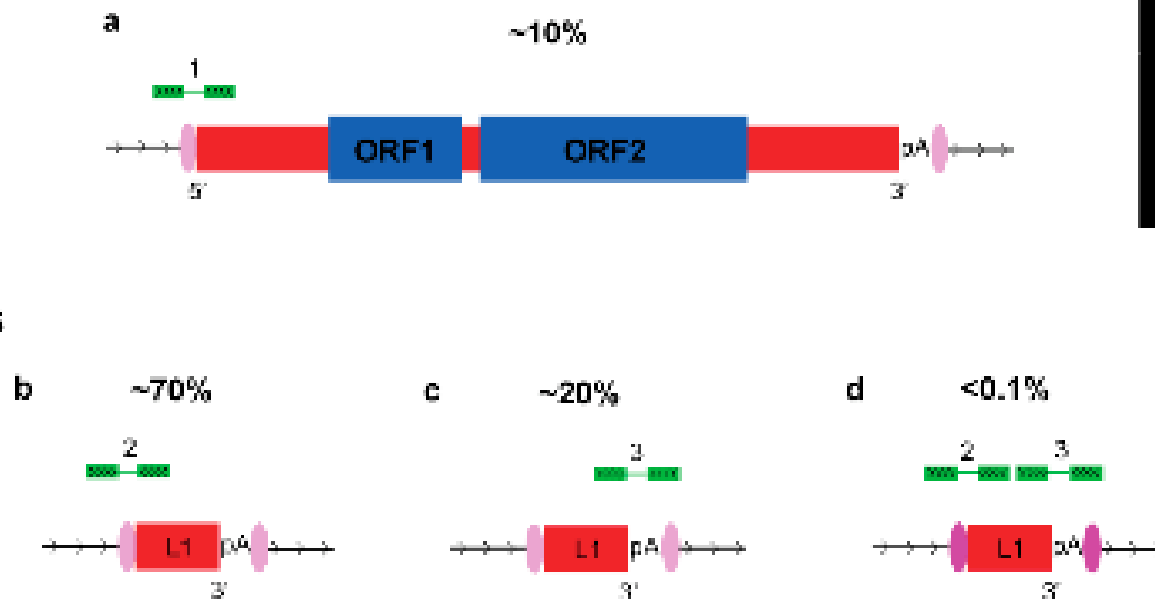
LEGEND



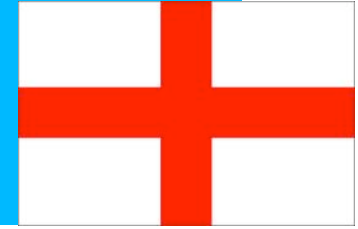
DESIGN



DETECTION OF NOVEL INSERTIONS

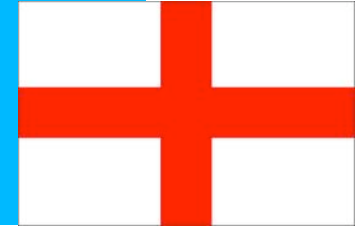


legend to suppl fig 1



Supplementary Figure S1: Four general types of somatic L1 insertion detected by RC-seq. Capture array probes targeting the termini of L1 sequences yield RC-seq reads that span insertion 5' and 3' junctions (illustrated as RC-seq reads 1-3). These are then mapped to the reference genome and correspond to (a) the 5' end of full-length or near full-length insertions (~10% of junctions detected) or (b) truncated insertions sequenced at their 5' end and captured by a probe at their 3' end (~70% of junctions detected) or (c) the 3' end of insertions of unknown length (~20% of junctions detected) or (d) insertions of known length detected at both ends (rare). As illustrated here, only (d) can conclusively determine whether target site duplications (TSDs) occur and therefore whether retrotransposition is driven by TPRT or an endonuclease independent mechanism. Also note that, as the average RC-seq insert size was ~250nt for most libraries, paired-end 101mer reads were usually separated by a gap of ~50nt (as depicted here). Longer 2x150mer sequencing performed on donor B samples, with an insert size of ~200nt, served to remove this gap.

Screening di cervelli indipendenti



Precedentemente era stata analizzata la CNV con la mobilitazione in vivo.

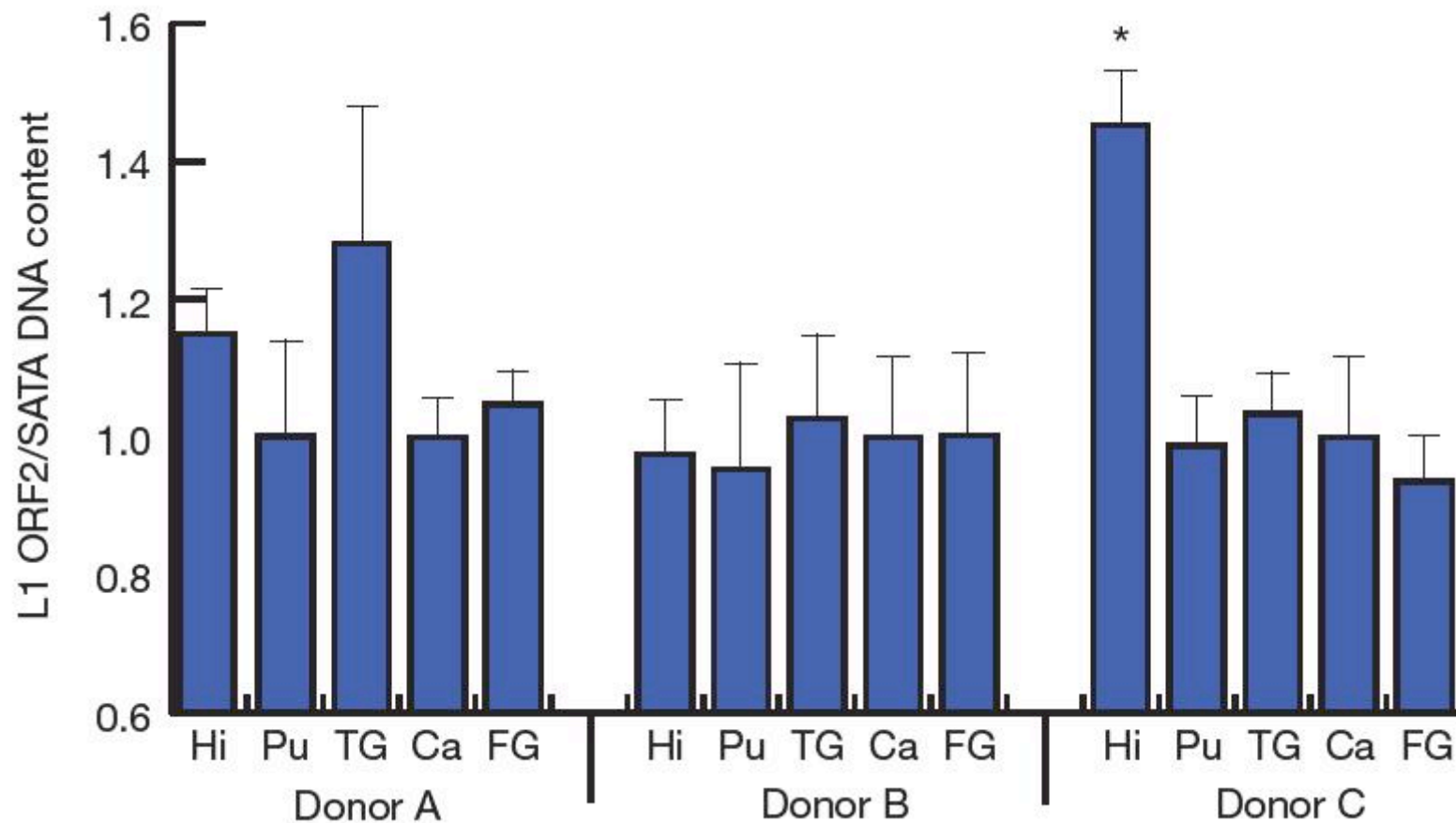
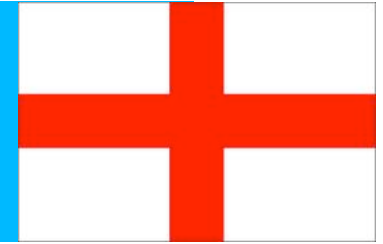
Analisi su tre cervelli (A;B;C) e cinque regioni viene contato il CNV di L1. Aumento di copie L1 (ORF 2) nell'ippocampo di C ($p < 0.001$) e minore in A (fig.2).

Applicata tecnica RC-seq all'ippocampo con CNV più alta e al nucleo caudato con CNV più bassa sui tre cervelli A;B;C

(incluso un controllo tecnico negativo di replica sul nucleo caudato.

Totale di $177,4 \times 10^6$ RC-seq con estremità appaiate da sei libraries (S table 3) alta copertura di sequenze (reads) di retrotrasposoni attivi (completi), alta riproducibilità e bassa cattura di seq bias (S results).

Figure 2 multiplex qPCR



Hippocampus; putamen; middle temporal gyrus; caudate nucleus; middle frontal gyrus.

Legend fig 2

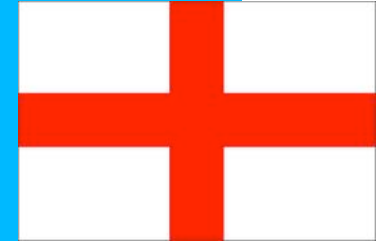
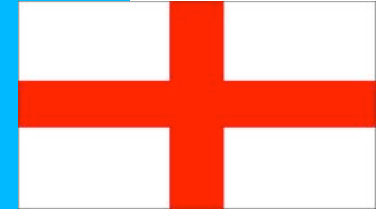


Figure 2 legend Multiplex quantitative PCR confirms L1 CNV in the human brain. The relative abundance of L1 open reading frame 2 (ORF2) with respect to a-satellite repeats (SATA) was quantified using an existing TaqMan-based approach¹⁰. Genomic DNA from five brain regions was assayed in three donors (A, B and C). Hi, hippocampus; Pu, putamen; TG, middle temporal gyrus; Ca, caudate nucleus; FG, middle frontal gyrus. Values are normalized to caudate nucleus for each donor. Error bars, s.e.m. *P , 0.001 for repeated-measures one-way analysis of variance within each donor, followed by pairwise least-significant-difference post hoc tests with Bonferroni correction.

Ricerca delle inserzioni nuove



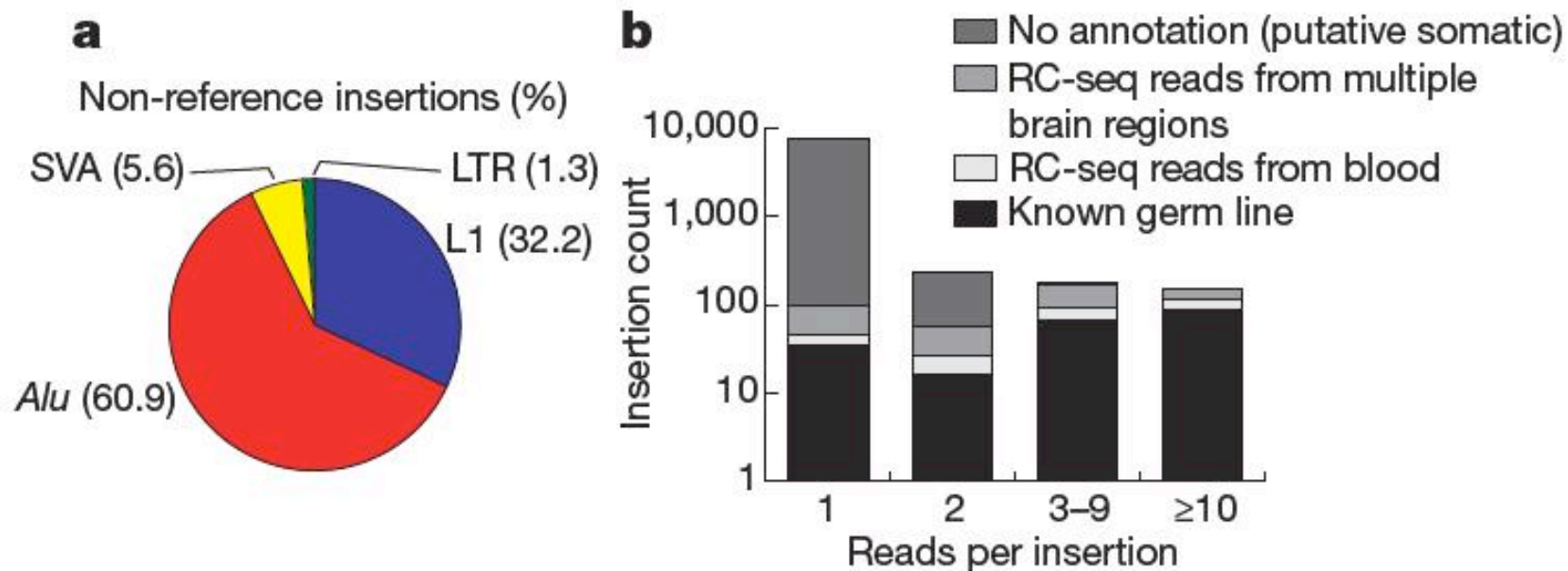
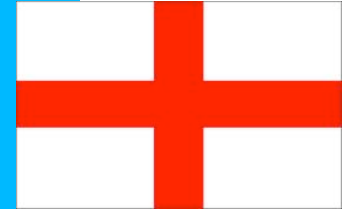
Le nuove inserzioni del retrotrasposone sono state raccolte (raggruppate) sulla base del sito di inserzione, orientamento relativo ed appartenenza alla famiglia di retrotrasposone.

Ottenuto un totale di 25'229 clusters.

Cluster prossimali posti su strand opposti indicano i due estremi di una inserzione e venivano appaiati generando un catalogo di 24'540 inserzioni nuove (S table 4).

Come atteso la maggioranza era L1 (32,2%) o Alu (60,9%) in fig.3a.

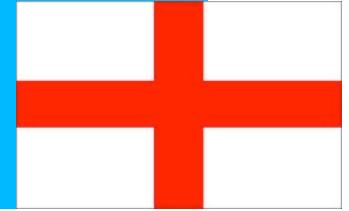
fig 3 non reference genome insertions



Characterization of non-reference genome insertions.

a, Proportions of novel insertions found for each family. b, Annotation of novel L1 insertions (note logarithmic scale) across all brain libraries. The great majority of insertions detected by fewer than three reads could not be annotated and were considered putative somatic events.

eventi somatici e germline



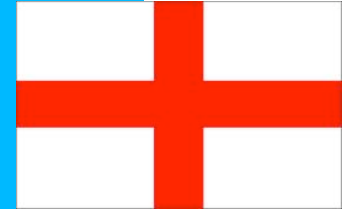
Confronto delle 3 più grandi banche dati dei polimorfismi di Alu ed L1 come riferimento,

Analisi RC-seq di Dna genomici uniti (pooled) estratti da sangue

Banca di 6'150 seq cluster (S table 5) che erano incrociate con i cluster già ottenuti con RC-seq di cervello. Ogni cluster che contenesse RC-seq da più di una regione o singole con una sovrapposizione su un altro cluster* RC-seq da sangue o coincideva con un polimorfismo noto venivano designate come inserzione germline (non somatica).

8,4% inserzioni Alu nel cervello erano germline contro 1.9% di L1. Quasi tutte le inserzioni L1 non marcate corrispondevano a meno di tre RC-seq reads e considerate nuove potenziali inserzioni. *regione di inserzione ricorrente in più reads

conferma delle inserzioni nuove



Validazione dei siti candidati di inserzione tramite PCR e sequenziamento con capillare.

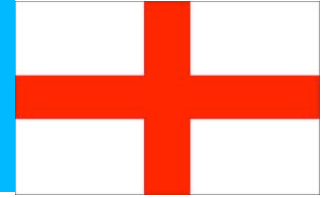
35 inserzioni germline L1, Alu, SVA e LTR sono state confermate per PCR diretta (S table 6).

Data la bassa abbondanza di molecole bersaglio e l'alta frequenza genomica dell'estremo 3' di L1 è stato fatta una PCR nested 5' di validazione per le inserzioni somatiche.

Da 850 e 2'601 inserzioni di L1 ed Alu complete (11% e 19% delle inserzioni possibili (putative) per ogni famiglia sono state validate 29 (14 L1 e 15 Alu).

Tutti i campioni scelti erano esoni o intronici e preferenzialmente sulla base della interruzione al 5', preferendo inserzioni più lunghe.

Ottimizzazione dell'esperimento



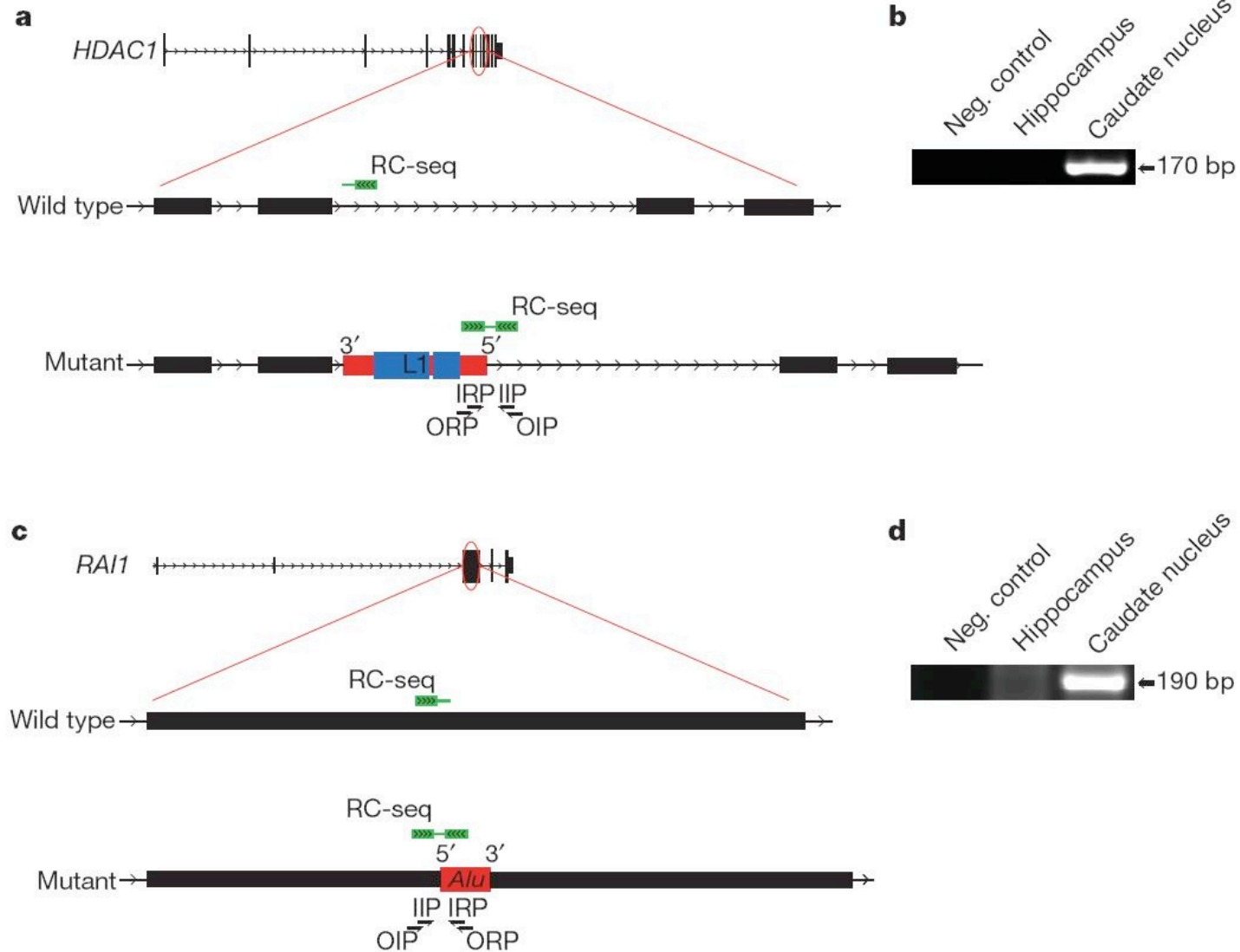
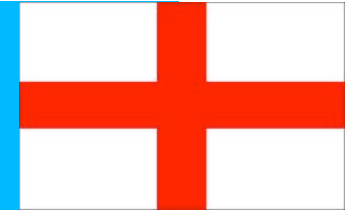
Con un nuovo campione di DNA (100 ng) quantità che un tempo si usava per un Southern

È stato ripetuto l'esperimento e ritrovato: a) tutte le inserzioni per L1 e 12 su 15 inserzioni per Alu, 2 per SVA.

In diversi casi la validazione per PCR è stata possibile solo al 5' per mancanza della sequenza al 3'.

Mobile elements drivers of genome evolution

fig 4 somatic insertions



legend to fig 4

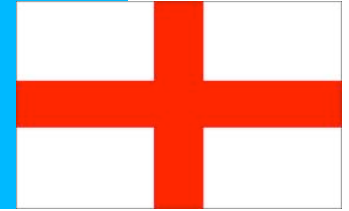


Figure 4 | Discovery of somatic insertions in HDAC1 and RAI1.

a, Alignment of an RC-seq read (green) from donor C caudate nucleus indicated an antisense L1 insertion in intron 9 of HDAC1. Nested PCR primers were designed to span the L1 5' terminus, with an initial reaction combining outside retrotransposon primers (ORPs) and insertion site primers (OIPs) and a second reaction using inside retrotransposon primers (IRP) and insertion site primers (IIP).

b, Amplification of the nested PCR target, confirmed for specificity by capillary sequencing, was achieved in caudate nucleus but not in hippocampus. Sequencing indicated that the L1 insertion mobilized from chromosome 9 and was accompanied by 5' transduction. bp, base pairs.

c, Alignment of an RC-seq read pair from donor A caudate nucleus indicated a sense Alu insertion in exon 3, and the coding sequence, of RAI1.

d, As for b, amplification of the nested PCR target was achieved in caudate nucleus but not in hippocampus. Sequencing indicated that the Alu insertion mobilized from chromosome 4. We note that L1 and Alu elements in a and c are not drawn to scale, that the L1 open reading frames in a are coloured in blue and that the untranslated regions of L1 and Alu are coloured in red.