Available online at www.sciencedirect.com

SCIENCE DIRECT®

Matrix Biology 23 (2004) 487–496

MATRIX BIOLOGY

www.elsevier.com/locate/matbio

Genomic organization and sequence variation of the human integrin subunit α8 gene (*ITGA8*)

Catherine Ekwa-Ekoka[a], George A. Diaz[a,b], Chris Carlson[c], Tomonobu Hasegawa[d], Ram Samudrala[e], Kim-chew Lim[f], Julie M. Yabu[g], Brynn Levy[a], Lynn M. Schnapp[g,*]

[a]*Department of Pediatrics, Mount Sinai School of Medicine, New York, NY, USA*
[b]*Department of Human Genetics, Mount Sinai School of Medicine, New York, NY, USA*
[c]*Department of Genome Sciences, University of Washington, Seattle WA, USA*
[d]*Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan*
[e]*Computational Genomics, Department of Microbiology, University of Washington, Seattle, WA, USA*
[f]*Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA*
[g]*Department of Medicine, Harborview Medical Center, University of Washington, Box 359640, 325 Ninth Ave, Seattle, WA 98104, USA*

Received 17 June 2004; received in revised form 31 July 2004; accepted 9 August 2004

Abstract

The integrin α8 is highly expressed during kidney and lung development. α8-deficient mice display abnormal renal development suggesting that α8 plays a critical role in organogenesis. Therefore, it would be of considerable interest to understand the genomic structure, localization and sequence variation of the α8 gene. Using FISH and genomic database analysis, we show that α8 gene maps to chromosome 10p13 and consists of >200 kbp organized into 30 exons. Examination of 47 individuals from two different ethnic groups (European and African descent) identified 286 varying sites. The diversity of α8 is comparable to that of other regions within the human genome. Eight of the varying sites were located in the coding regions: six resulted in nonsynonymous substitutions of which two lead to non-conservative changes in protein. None of the sites showed significant deviation from Hardy–Weinberg equilibrium. We mapped the coding region single nucleotide polymorphisms (SNPs) onto a model of the predicted α8 structure and found all the SNPs were located in the "calf" of the extracellular domain. In the European population, the linkage disequilibrium statistic D′ showed three blocks of relatively non-recombinant regions in the α8 gene while the African population showed more evidence of recombination. The observed patterns of the linkage disequilibrium statistic R^2 suggest that a large number of sites will need to be genotyped to ensure coverage of the entire gene for genetic association studies. Identification of the sequence variation will allow genetic association studies of α8 in kidney and lung disease.

Keywords: Integrin alpha chain; Polymorphism, single nucleotide

1. Introduction

The integrin family is a ubiquitously expressed group of cell surface receptors that mediate cell–cell and cell–extracellular matrix (ECM) interactions (Hynes, 2002). By providing an essential link between cell and ECM, integrins play a pivotal role in organogenesis and response to injury. The integrin subunit α8 associates exclusively with the β1 subunit to form an RGD-binding integrin. The integrin α8β1 expression is essential for normal kidney development and is critically important for the epithelial conversion of mesenchymal cells during nephrogenesis (Muller et al., 1997). Disruption of the α8 gene in mice is associated with severely impaired kidney development and early postnatal death due to kidney failure (Muller et al., 1997). A subset of α8−/− mice survive long-term and have defects in stereocilia development (Littlewood Evans and Muller, 2000).

The integrin α8β1 is also important in organ response to injury (Hartner et al., 1999; Levine et al., 2000). In the adult,

* Corresponding author. Tel.: +1 206 341 5389; fax: +1 206 341 5392.
E-mail address: lschnapp@u.washington.edu (L.M. Schnapp).

0945-053X/$ - see front matter © 2004 Elsevier B.V./International Society of Matrix Biology. All rights reserved.
doi:10.1016/j.matbio.2004.08.005

α8β1 is highly expressed in smooth muscle cells, lung myofibroblasts and kidney mesangial cells (Schnapp et al., 1995a). α8β1 is upregulated in experimental models of pulmonary fibrosis, liver fibrosis, cardiac fibrosis and glomerularnephritis (Bouzeghrane et al., 2004; Hartner et al., 1999; Levine et al., 2000). In a model of hypertension, α8-deficient mice had a significant increase in glomerular disruption (Hartner et al., 2002), suggesting that α8β1 is important to the stability of the glomerular capillary tuft.

Mutations or polymorphisms of the α8 gene in humans may contribute to the pathogenesis of kidney malformations or response to injury. Therefore, it would be of considerable interest to understand the genomic structure, localization and sequence variation of the α8 gene. We report the genomic organization of the α8 gene, including the signal peptide and exon/intron junctions. We identified single nucleotide polymorphisms (SNPs), determined allele frequencies in two different ethnic groups, and mapped the coding region SNPs onto a model of the predicted α8 structure. This information will provide insight into the regulation of α8 and allow the design of genetic studies of kidney disease based on marker association.

2. Results

2.1. Chromosomal location of the α8 gene

Query of genomic sequence databases with the 3 kb nucleotide sequence of α8 cDNA (GenBank accession L36531) identified several human BAC clones containing part of or all of the *ITGA8* gene. Clones RP11-39J5 (AL359645) and RP11-149I8 (AL139338) were assigned localization to chromosome 10; however, clone RP11-496B1 (AC018794) was assigned to chromosome 15. Sequence analysis confirmed identity between the three clones, establishing that the assignment of clone RP11-496B1 to chromosome 15 was erroneous.

Confirmation of this conclusion was obtained by using BAC clones and the α8 cDNA as probes for FISH. FISH analysis of metaphase chromosomes from a normal male showed hybridization of the α8 cDNA, the BAC-39J5 and the BAC-93B7 digoxigenin-labeled probes to one or both chromatids of chromosome 10, band p13 (Fig. 1). No specific signal was obtained in any other chromosomal location. From these data, we conclude that the α8 integrin subunit gene maps to chromosome 10p13.

2.2. Organization of the α8 integrin gene

Prior to the completion of the local genome sequence, we ordered the non-overlapping assemblies from the three identified BAC clones containing regions of homology to α8 and determined the *ITGA8* genomic organization, a finding that was confirmed upon completion of the regional sequence (Fig. 2A). The α8 integrin gene consists of 30 exons spanning >200 kb of genomic DNA. The exons were 44–206 bp in length, while the intron sizes ranged from 400 to 30,400 bp. The predicted post translational cleavage site, RKR, is contained in exon 26. The 25 residues of the transmembrane domain are encoded by exon 29. The last exon (exon 30) contains the cytoplasmic domain and the 3′ untranslated region. Direct reading of the sequences at the exon/intron boundaries allowed us to determine 29 splice junctions, all of which are in agreement with the consensus splice site sequences MAG/GTRAGT (M=A or C, R=A or G) and YAG (Mount et al., 1992).

The previously published α8 cDNA sequence (L36531) does not contain the translation start site or the signal peptide. To characterize the 5′ end of the gene, we used BLAST analysis to identify spliced ESTs that extended upstream of the known sequence. Three human (AI991802, AI865694 and AI674182) and two orthologous murine (BE624254 and BB643134) ESTs were identified. The human ESTs did not encode a potential start codon; therefore, 5′ RACE experiments were carried out which extended the transcription start site to 160 bp upstream of the start of the mature peptide ORF. The mature mouse and human proteins share 91% identity or similarity, suggesting that regions of homology represented bona fide translated sequence. ClustalW alignment of the nucleic acid sequences revealed sequence divergence upstream of a potential start

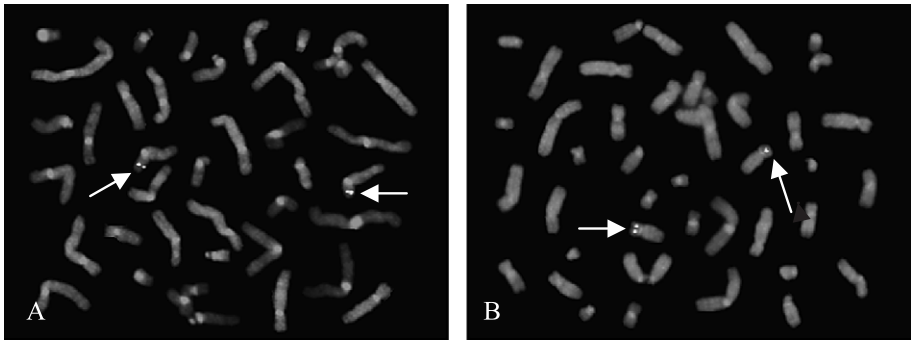

Fig. 1. Localization of α8 gene to chromosome 10p13. FISH analysis with the digoxigenin-labeled α8 cDNA probe (A) and digoxigenin-labeled BAC 39J5 probe (B) was performed on human metaphases. Arrow indicates specific sites of hybridization of the α8 probe to chromosome 10p13. Chromosomes were counterstained with DAPI.

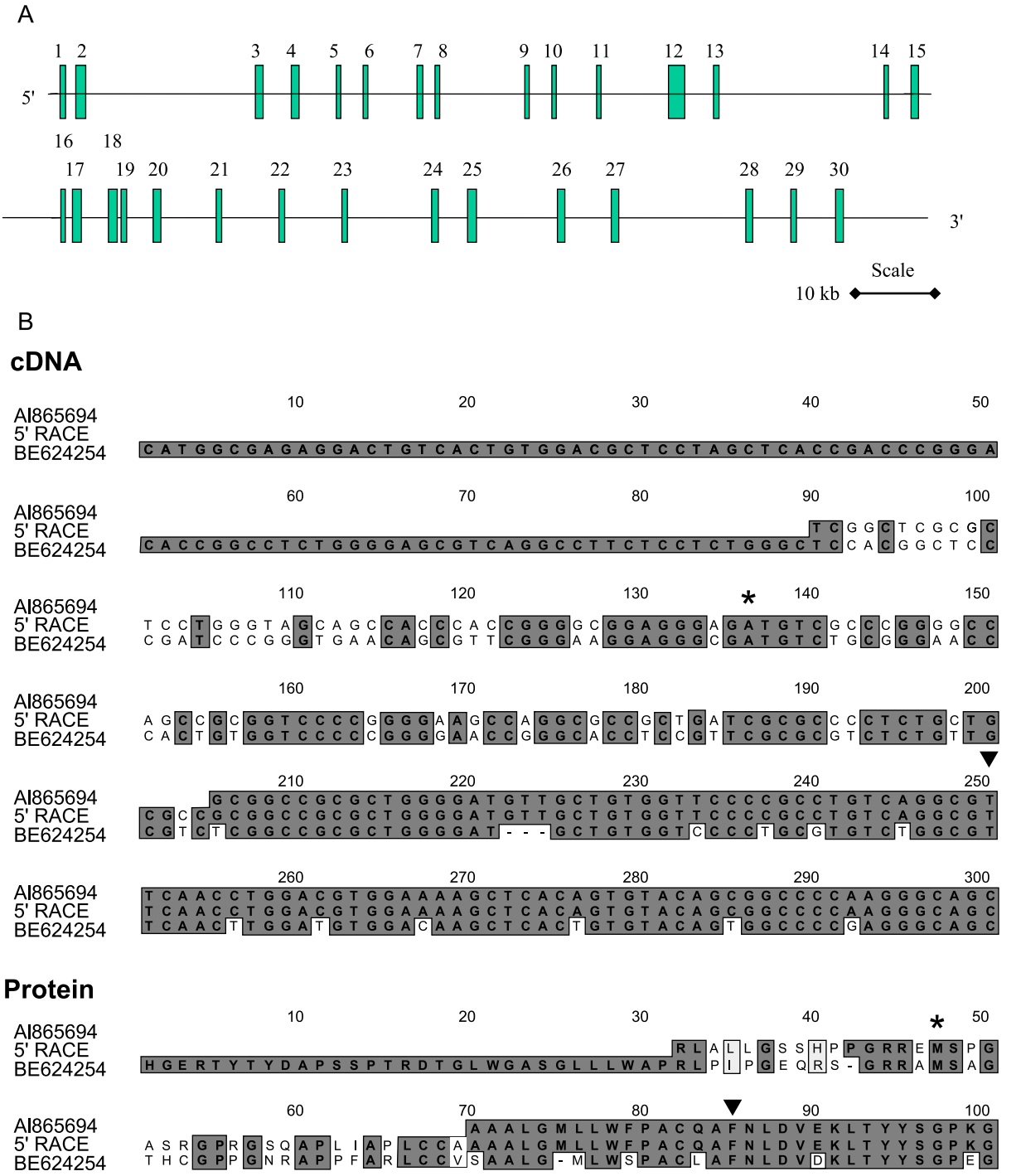

Fig. 2. (A) *ITGA8* genomic organization. *ITGA8* consists of 30 exons spaced over approximately 200 kb. (B) *ITGA8* 5′ sequence alignments. Nucleotide (top) and peptide (bottom) ClustalW sequence alignments are shown for the most upstream human (AI865694) and murine (BE624254) ESTs and a composite sequence derived from the longest 5′ RACE product obtained (5′ RACE). The putative ATG initiation codon and the encoded methionine are indicated with asterisks in the cDNA and peptide sequences. The start of the mature peptide (5′ end of sequence in GenBank) is indicated in the cDNA and peptide sequences by arrowheads.

codon, a pattern confirmed by alignment of the putative N terminal peptides (Fig. 2B). Analysis with the pSort II (http://www.psort.nibb.ac.jp) and SignalP (http://www.cbs.dtu.dk/services/SignalP) programs were consistent with the presence of a signal peptide sequence (Nakai and Kanehisa, 1992; Nielsen et al., 1997) and both programs predicted a signal peptide cleavage site immediately preceding the FNL sequence at the start of the mature peptide, suggesting that the putative initiation ATG identified by the 5′ RACE clones represented the actual translation start site.

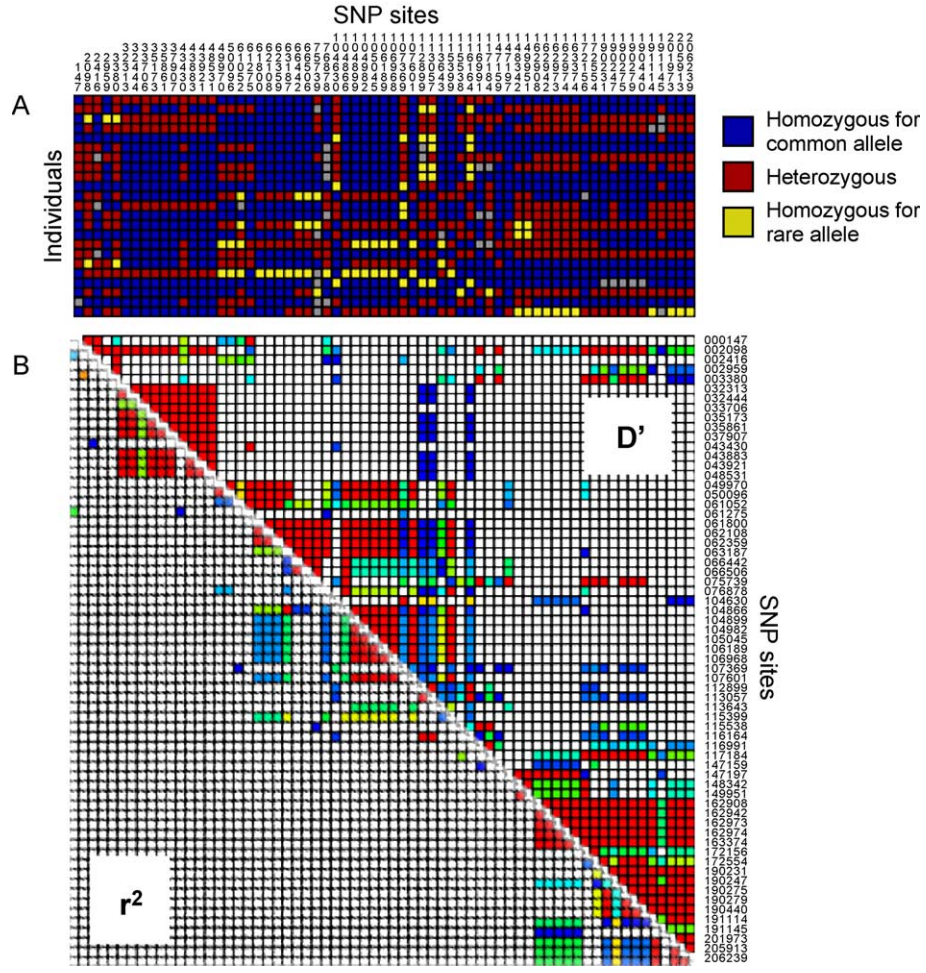

Fig. 3. (A) *ITGA8* genotypes of individuals of European descent at each of the variant sites with MAF>10%. Individual sample identifiers are shown on the left side of the array and variants are numbered consecutively across the top. At each site, individuals homozygous for the common allele are blue, heterozygotes are red, and homozygotes for the rare allele are yellow. Purple are undetermined. (B) Associations between *ITGA8* SNPs with MAF≥10%. The linkage disequilibrium statistic D′ is shown above the diagonal, and r^2 is shown below the diagonal. Each cell represents the observed linkage disequilibrium value for a pair of SNPs. Each cell is colored by value: Values below 0.5 are white, values between 0.5 and 0.75 are blue, values between 0.75 and 1 are green, and values of 1.0 are red.

2.3. Sequence diversity of ITGA8

Approximately 48 kbp of DNA containing the *ITGA8* gene were amplified and scanned for variation. In all, a sample of 47 individuals representing two populations were scanned across the target regions, and 286 varying sites were identified by comparing the amplified sequences using Poly Phred Program. Of these, 243 (85%) were diallelic single nucleotide substitutions. Of the remaining varying sites, 26 were insertions, 16 were deletions and 1 was trialleleic. As

Table 1
ITGA8 Coding Region Polymorphisms

Site	cDNA position	Description	Allele1	Codon 1	Amino1	AD freq[a]	ED freq	Allele2	Codon 2	Amino2	AD freq	ED freq
044010	00646	Non-syn	G	GTG	Val	0.91	1.00	C	CTG	Leu	0.09	0.00
113933	01730	Non-syn	C	TCC	Ser	0.96	0.98	T	TTC	Phe	0.04	0.02
113945	01742	Non-syn	A	CAG	Gln	1.00	0.93	C	CCG	Pro	0.00	0.07
117418	02100	Synon	A	GGA	Gly	0.95	1.00	G	GGG	Gly	0.05	0.00
129371	022243	Non-syn	G	CGT	Arg	0.98	1.00	A	CAT	His	0.02	0.00
135040	02352	Synon	A	GTA	Val	0.98	1.00	G	GTG	Val	0.02	0.00
190592	02977	Non-syn	A	ATA	Ile	0.95	1.00	G	GTA	Val	0.05	0.00
190596	02981	Non-syn	C	GCA	Ala	0.93	1.00	T	GTA	Val	0.07	0.00

AD: African Descent, ED: European Descent.
Non-syn: Non-synonymous (results in amino acid change).
Syn: Synonymous (no change in amino acid).
[a] Allele frequency at polymorphic site for each sample population.

Table 2
ITGA8 Non-synonymous polymorphisms

α8 Protein Position	cDNA position	Exon	α8 Allele 1	α8 Allele 2	α8 Mouse[a]	α8 Chicken	α•	α5	αIIb
529	00646	6	Val	Leu	Thr	Val	Ser	Thr	Leu
539	01730	17	Ser	**Phe**[b]	Ser	Lys	Leu	Arg	Ser
543	01742	17	Gln	Pro	Gln	Gln	Glu	Arg	His
710	02243	22	Arg	His	Arg	His	Gln	His	Asn
955	02977	28	Iso	Val	Thr	Iso	Tyr	Arg	Ala
956	02981	28	Ala	Val	Ala	Ala	Leu	Gln	Gln

[a] Amino acid at comparable positions in mouse and chicken α8 sequences and related integrin alpha subunits αv, α5 and αIIB sequences are presented.
[b] Predicted as intolerant amino acid change.

expected, nucleotide diversity was higher for the African American population than the European population (264 vs. 122 varying sites, respectively). For European descent, 64 SNP were detected with a MAF>10%; for African descent, 92 SNP with a MAF>10% were detected (Fig. 3A, and not shown). None of the SNPs showed significant deviation from Hardy–Weinberg equilibrium. Eight of the varying SNPs were located in the coding regions of *ITGA8* (Table 1). Of the eight coding SNPs, six lead to nonsynonymous substitutions in the proteins and two of these changes were non-conservative changes in protein (Ser to Phe, Gln to Pro), both located in exon 17 (Table 2). In order to gain insight into the potential consequences of the coding region SNPs, we modeled the predicted secondary and tertiary protein structure of α8 extracellular domain and mapped the coding SNPs. The model of α8 reveals a typical α subunit structure: β propeller domain (important in ligand binding), "thigh" domain and two "calf" domains (Fig. 4) (Humphries et al., 2003a). Coding region SNPs map to the "thigh" and "calf" domains of the subunit. We analyzed the coding region SNPs with SIFT (http://www.blocks.fhcrc.org/sift/SIFT.html), an algorithm for predicting deleterious SNPs using homologous

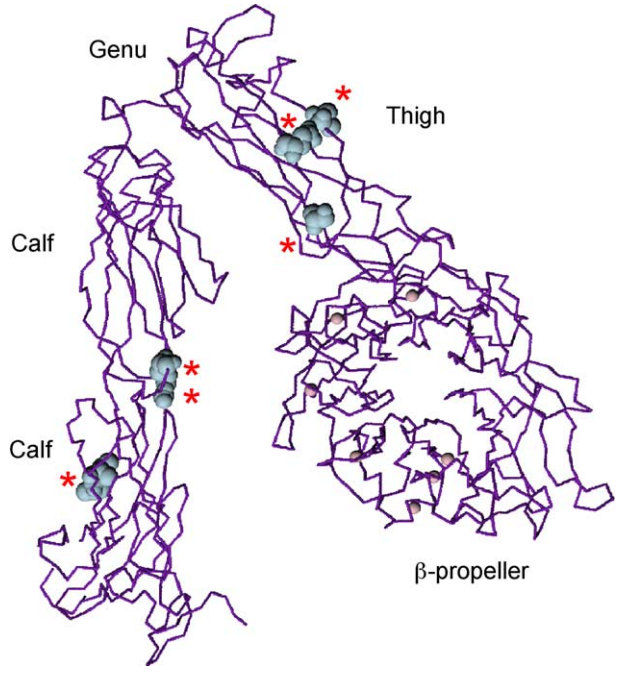

Fig. 4. Predicted structure of α8 extracellular domain and location of SNPs. The model of α8 reveals a typical α subunit structure: β-propeller domain, "thigh" domain and two "calf" domains. Coding region SNPs (*) map to the "thigh" and calf" domains of the subunit. Divalent cations are shown as spheres in β-propeller domain.

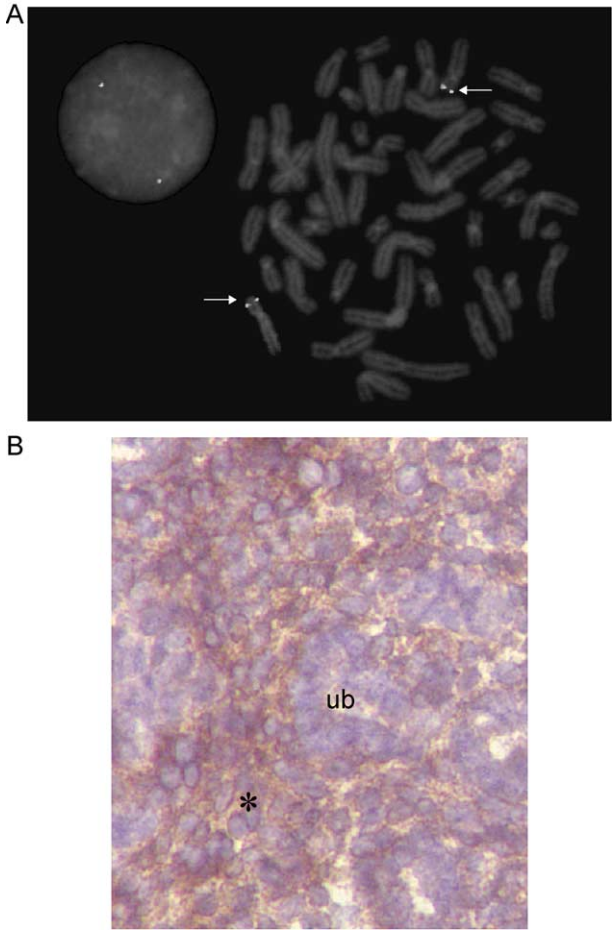

Fig. 5. (A) Localization of α8 in patient with HDR syndrome. FISH analysis with the digoxigenin-labeled α8 cDNA probe was performed on metaphase spread of HDR patient. FISH analysis shows hybridization of α8 cDNA probe to chromosome 10p13 (arrows), indicating that the α8 gene is not deleted in this patient. Note truncated chromosome 10, consistent with diagnosis of HDR. Chromosomes were counterstained with DAPI. (B) Immunohistochemistry for α8 in dopamine rescued GATA3−/− d13.5 embryos. Cryosections of mouse embryo were stained with anti-α8 antibody and counterstained with hematoxylin. Immunoreactivity of α8 is present in the mesenchyme (*), surrounding the ureteric bud (UB).

protein sequences (Ng and Henikoff, 2003). Threshold score of <0.05 was used to predict a non-tolerant substitution. Coding SNP at allele position 1139323 (Ser to Phe) was predicted to be intolerant while the other five nonsynonymous substitutions were predicted to be tolerated with threshold scores >0.05. We also calculated the all-atom scoring scores (measure of protein stability) for each nonsynonymous SNP. Gln to Pro and Iso to Val changes are predicted to be more stable than the major allele. Similar to SIFT, all-atom scoring predicted SNP 539 (Ser to Phe) likely to be deleterious. We compared the coding region SNPs to the homologous region in chimpanzee. Sequence information was available for 7/8 of the coding region SNPs. In all 7 cases, the major allele was present in the chimpanzee, suggesting that the major allele is the ancestral allele. Both populations displayed high haplotype diversity: the European descent clustered into 31 and the African group into 47 haplotypes, and only one haplotype in both groups was represented three or more times. Because of the high inferred haplotype diversity, the inferred haplotypes are less reliable and should be used with caution.

2.4. Linkage disequilibrium between common polymorphic sites in ITGA8

The linkage disequilibrium between polymorphic sites is shown graphically in Fig. 3B. The linkage disequilibrium statistic D′ (shown above the diagonal) is useful in identifying recombination events. If D′ is less than 1, either recombination has occurred between the two SNPs or a recurrent mutation or gene conversion has occurred at one of the SNPs. Likewise, D′ is expected to be 1 for pairwise combinations within a nonrecombinant area. For the European population, the *ITGA8* gene is characterized by three blocks of relatively non-recombinant regions (spanning approximately 3244–48531, 62108–107601, 162942–206239). In contrast, the African-American population shows much shorter non-recombinant regions.

The disequilibrium statistic r^2 (Pearson correlation coefficient for two biallelic SNPs) describes the allelic correlation between a pair of SNPs, which provides information for genetic association studies. Specifically, it describes how redundant genotype information is between a pair of sites. The observed patterns (Fig. 3B, below the diagonal) suggest that a large number of sites will need to be genotyped to ensure coverage of the entire gene for genetic association studies. In order to describe all genotype patterns above MAF>10% in European population, 28 SNPs would need to be analyzed.

2.5. α8 localization in HDR patient and GATA3-deficient embryo

Deletions of 10p13 are associated with the syndrome of HDR (renal dysplasia, deafness and hypoparathyroidism). Because α8 localized to the same chromosome as HDR, we asked whether α8 was deleted in HDR patients. We performed FISH analysis of a patient with HDR and renal dysplasia (Hasegawa et al., 1997). We detected signal from both chromatids at 10p13 (Fig. 5A), showing that α8 was not deleted in this patient. A recent report showed that many cases of HDR were due to haploinsufficiency of the transcription factor GATA3 (Van Esch et al., 2000). Since the expression pattern of GATA3 and α8 overlap in the developing kidney, we investigated whether GATA3 expression was necessary for α8 expression during development. Disruption of the GATA3 gene leads to early embryonic lethality prior to kidney development (Pandolfi et al., 1995). However, administration of dopamine to pregnant heterozygote dams partially rescues the GATA3 lethality and allows kidney development to begin (Lim et al., 2000). We performed immunohistochemistry for α8 on pharmacologically rescued E13.5 embryos (GATA3−/−). Immunoreactivity for α8 was present in the condensing mesenchyme surrounding the ureteric bud in the kidney, similar to normal embryonic kidney expression (Fig. 5B). In addition, α8 immunoreactivity was detected in the developing lung mesenchyme as expected (not shown).

3. Discussion

We describe for sequence variation, chromosomal location and genomic organization of the human integrin α8. Sequence homology analysis shows that α8 is most closely related to the αv, α5 and αIIb integrins. Like these α subunits, α8 is non-I-domain, RGD-binding and post-translationally cleaved. Compared to α8, the αv, α5 and αIIb integrins have similar genomic structures. αv has 30 exons ranging from 45 to 226 bp in size, and intron sizes ranging from 86 to over 20,000 bp. There is a 38-residue signal peptide encoded by exon 1 which also encodes for the first 32 residues of the mature peptide. α5 also has 30 exons with the 44-residue signal peptide and initial residues of the mature peptide encoded by the first exon (htgs database). αIIb spans approximately 17 kb. Its 30 exons range from 45 to 249 bp in size and the introns from 77 to 2800 bp (Heidenreich et al., 1990).

The genes encoding the I-domain α chains are clustered on chromosome 16 (αL, αM, αX and αD) and chromosome 5 (α1, α2). The high degree of homology of the integrin genes within these chromosomes suggests that they may represent gene duplications. The non-I-domain α genes are grouped on chromosomes 2 (α4, αV, α6), 12 (α5, α7), 17 (α3, αIIb) and now chromosome 10 (α8). Since αv, α5 and αIIb integrins display sequence homologies, as do α6, α7 and α3, it has been suggested that they form paralogous groups evolving in parallel with the homeobox genes which are also clustered on chromosomes 2, 12 and 17 (Wang et al., 1995). The localization of the α8 gene, however, is not associated with that of a homeobox

gene. Therefore, its evolution would have occurred separately from that of αv, α5, αIIb and the homeobox genes. Although there does not seem to be a relationship between the different α–β combinations and chromosomal location, it is interesting to note that α8, which combines exclusively with the β1 integrin, clusters with this subunit on chromosome 10.

Interestingly, during the original cloning of the human α8 cDNA, we identified one clone from the aorta cDNA library that lacked nucleotide sequence between positions 688 and 732 (Schnapp et al., 1995a). This interval corresponds to the entire exon 8 and may represent alternative splicing of the extracellular domain. Exon 8 is found within the predicted β propeller domain which is involved in ligand binding (Springer, 1997). Furthermore, our original description of α8 included identification of two distinct RNA species by Northern analysis (Schnapp et al., 1995a). Thus, the α8 subunit may exist in at least two different forms that may be developmentally regulated and dictate different ligand affinities and signaling capacities. Alternative splicing of both the extracellular (α6, α7 and αIIb) and cytoplasmic (β1, β3, β4, α3 and α6) domains of integrins have been described (de Melker and Sonnenberg, 1999).

Disruption of the α8 gene in mice is associated with severely impaired kidney development and early postnatal death due to renal failure (Muller et al., 1997). Interestingly, deletions of 10p13 are associated with the HDR (hypoparathyroidism, deafness and renal dysplasia) syndrome (Hasegawa et al., 1997). To test whether deletion of the α8 gene contributed to renal dysplasia in the HDR syndrome, we performed FISH analysis on metaphase spreads from a previously described patient with HDR whose clinical manifestations included renal anomalies (Hasegawa et al., 1997). We detected a signal for α8 using the BAC DNA and α8 cDNA on both chromatids of chromosome 10p13 (Fig. 5A), suggesting that a complete deletion of the α8 gene was not responsible for renal manifestations in this patient.

Recent work has shown that haploinsufficiency of the GATA3 gene leads to the clinical syndrome of HDR although two patients with HDR did not have mutations of the GATA3 gene (Van Esch et al., 2000). GATA3 is expressed during development of the human embryonic kidney, the inner ear and the parathyroids (Debacker et al., 1999; Karis et al., 2001; Labastie et al., 1995), and overlaps with α8 expression. Specifically, both α8 and GATA3 are expressed during mesenchymal condensation during kidney development (Labastie et al., 1995). Thus, GATA3 is a potential regulator of α8 expression. To test this hypothesis, we examined dopamine-rescued GATA3 knock-out mice embryos. Neuroendocrine rescue overcomes some of the early lethality defects and allows partial kidney development to occur (Lim et al., 2000). α8 signal is clearly identified in the mesenchyme in embryonic kidney and lung tissue at E13.5, suggesting that GATA3 is not an absolute requirement for α8 expression. Further investigations on the genes regulating α8 expression are needed to understand the factors controlling kidney development.

Our analysis detected eight coding SNPs, which is slightly higher than the expected frequency of one SNP every 1000 bases (Landegren et al., 1998). Notwithstanding, the diversity of *ITGA8* is comparable to that of other regions within the human genome. The coding region SNPs map to the "leg" domain; none are located near the ligand binding pocket formed by the β-propeller. Although the majority of function perturbing mutations are found in the β-propeller region, several natural and engineered mutations that perturb ligand binding are mapped to the "leg" in the α5 subunit (Humphries et al., 2003b). In addition, this region of α subunits can bind to integrin-associated proteins, such as tetraspanins (Yauch et al., 2000). In addition, based on energy calculations, the SNPs may affect the structural stability of the protein.

As expected, variants are not uniformly distributed among different populations. Analysis of two ethnic backgrounds identified different patterns of linkage disequilibrium relationships that exist across this gene in different human populations (Reich et al., 2001). Is the population size sufficient to detect significant SNPs? A screen of 47 individuals detects approximately 99% of the SNPs with minimal allele frequency of 5% (Kruglyak and Nickerson, 2001). It is still possible that we missed functionally significant SNPs by screening only 47 individuals. However, missed SNPs will tend to be rare in the population (<5%) and will not be as useful for association analysis, where the population sizes needed to test the effects of these rarer variants would be quite large. In order to detect >99% of SNPs with a MAF of >1%, we would need to screen 192 chromosomes or 96 individuals (Kruglyak and Nickerson, 2001). Nonetheless, by screening ~24 individuals from each population, we expect to have detected >85% of SNPs with MAF>1% in either population. Since the most effective SNPs for association studies are those with MAF>10%, this sample size is more than sufficient to capture the SNPs needed to design population-based association studies. The identification of *ITGA8* common SNPs provides the basis of future population studies to discover associations between α8 SNPs and the development of renal disease. An increased prevalence of a specific SNP or a sequence haplotype (specific combinations of SNPs across the sequence) in affected individuals suggests an association between sequence polymorphism and the disease (Collins et al., 1997). We have identified several coding SNPs with potential functional significance that will be tested in more detail.

In summary, α8 integrin expression is an essential component of the normal renal developmental pathway. Identification of the sequence variation of α8 will facilitate genetic association studies to determine the contribution of α8 in patients with clinical renal syndromes. Understanding the genomic organization and regulation of α8 will increase our understanding of the molecular basis of kidney development and injury response.

4. Experimental procedures

4.1. Preparation of DNA

Construction of the full-length cDNA of the human α8 integrin subunit was previously described (Schnapp et al., 1995b). The RPCI-11-39J5 and RPCI-11-93B7 BAC clones homologous with the α8 cDNA were obtained from the RPCI-11 Human BAC library created in the laboratory of Pieter J. de Jong in the Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY. BAC plasmid DNA was grown and purified following the Qiagen Large-Construct Purification kit (Valencia, CA).

4.2. Fluorescence in situ hybridization (FISH)

To generate probe, purified RPCI-11-39J5 BAC DNA, RPCI-11-93B7 BAC DNA and pBS-α8 cDNA were labeled with the Dig-Nick Translation Mix (Roche Molecular Biochemicals, Indianapolis, IN). Incubation time was optimized to obtain labeled DNA fragment sizes 300–800 bp. Metaphase spreads were prepared from phytohemagglutinin-stimulated human peripheral blood lymphocytes by standard techniques (Cannizzaro and Shi, 1997). Digoxigenin-labeled DNA probe was applied to the target DNA on the slides, which were incubated in a humid slide-chamber at 37°C overnight. The slides were washed to remove the unbound DNA. Hybridization signal was detected and amplified using FluorAmp kit (Oncor, Gaithersburg, MD) per manufacturer's instructions. Chromosome identification was facilitated by counterstaining with 0.5 μg/ml DAPI in Vectashield (Vector Laboratories). FISH images were captured with an ImagePoint cooled CCD video camera from Photometrics through a Labophot-2A fluorescence microscope from Nikon. Digital image analysis was performed using a Cytovision Probe system from Applied Imaging.

4.3. Determination of the α8 gene organization

Exon/intron boundaries were determined by direct comparison of the α8 cDNA with the genomic sequences on the clones. The lengths of the fully sequenced introns 2, 4–21, 25 and 27–30 were available from the genomic clones. To determine the size of the incompletely sequenced introns 3, 22–24 and 26–27, we designed primers (Primer3, http://www-genome.wi.mit.edu/genome_software/other/primer3.html) from the 5′ and 3′ ends of these introns and performed PCR amplification at annealing temperatures of 50–55 °C and at an extension time of 30 s. PCR amplicons were purified using the Qiagen PCR Purification and Gel extraction kits (Valencia, CA). DNA was used for PCR cycle sequencing on an ABI 377 automated sequencer. Sequence information and genomic organization was subsequently confirmed during the SNP analysis (see below).

4.4. Rapid amplification of cDNA ends (RACE) experiments

Total RNA was obtained from human H4 cells (ATCC). RNA was dephosphorylated, decapped and ligated to a 5′ oligomer using the GeneRacer kit (Invitrogen) per the manufacturer's recommendations. Reverse transcription was performed with a gene-specific downstream primer from exon 1 of the α8 cDNA. PCR was performed using the GeneRacer 5′ primer and the reverse gene specific 3′ primer. The amplicon was nested with the GeneRacer nested 5′ primer and a gene-specific nested 3′ primer. The nested amplicon was cloned into a TA vector for sequencing. Primer sequences are available upon request.

4.5. Sequence analysis and polymorphism detection

As part of the SeattleSNPs Program for Genomic Applications, 47 individuals from two populations were scanned for sequence polymorphism across the α8 gene: 23 of European descent (12 male/11 female) from the CEPH families and 24 of African descent (12 male/12 female) from the Coriell Cell Repositories (SeattleSNPs NHLBI Program for Genomic Applications, UW-FHCRC, Seattle, WA (URL: http://www.pga.gs.washington.edu) [accessed 9/2003]). Overlapping primer sets spanning the genomic sequence of *ITGA8* were designed using the program PCR Overlap (http://www.droog.mbt.washington.edu/PCR-Overlap.html). All primer sequences are available at http://www.pga.gs.washington.edu/data/itga8/ITGA8.primers.fasta. Universal forward or reverse sequence (M13) was added to the 5′ end of the primers to generate fragments compatible with dye–primer fluorescence-based sequencing. All samples were amplified from genomic DNA. PCR conditions are available at http://www.pga.gs.washington.edu/data/primer_descriptions.html. Sequence analysis was performed as previously described (Rieder et al., 1999). Briefly, sequences were base-called by Phred (which reports differences in the quality of the sequence read), assembled by Phrap and called for single-nucleotide substitutions by the PolyPhred program (Nickerson et al., 1997). PolyPhred is a program that integrates the output of several other programs to accurately identify heterozygous sites in sequences produced with fluorescence-based chemistries. This program not only identifies polymorphic sites but also genotypes the individuals at these locations. All variants have been deposited in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/).

4.6. Protein modeling of α8 extracellular domain

Modeling of the α8 protein was accomplished using the RAMP software suite of programs (http://www.software.compbio.washington.edu) that is encapsulated in the

PROTINFO structure prediction server (http://www.protinfo.compbio.washington.edu). Modeling was performed using the comparative modeling protocol, which has been shown to work well in the CASP prediction experiments (Hung and Samudrala, 2003; Samudrala and Levitt, 2002). The structure of extracellular domain of αvβ3 (PDB code: 1l5g-A) was used as a template to construct the initial models of α8 using a minimum perturbation approach that aims to preserve as much information as possible from the template X-ray structure. Variable side chains and main chains were constructed using a graph-theory clique-finding approach that explores a variety of possible conformations for the respective side chains and main chains and finds the optimal combination using an all-atom scoring function (Samudrala and Moult, 1998). These approaches are described in further detail in (Samudrala and Levitt, 2002).

4.7. Statistical analysis

Allele frequency for each variable site was estimated by gene counting for each population. Hardy–Weinberg test for genotype frequency distributions was performed on the observed genotype frequencies for each site and population with significance based on a standard observed–expected χ^2 with 1 df. Linkage disequilibrium across *ITGA8* in both populations was assessed. The linkage disequilibrium statistics D′ (useful to detect recombination events between two SNPs) and r^2 (which describes how similar patterns of genotypes are for any pair of SNPs) were calculated for each pair of SNPs with an observed minor allele frequency (MAF)>10%. Haplotypes were inferred for all sites with MAF≥10% using the program PHASE v2.0 (Stephens et al., 2001).

4.8. Immunohistochemistry

Pregnant females from heterozygote transgenic mice with GATA3 germline-targeted mutation were fed DOPS as previously described (Lim et al., 2000). E13.5 embryos were dissected and immersed at 4°C in 10% sucrose in PBS. Following sucrose preservation, tissues were placed in OCT embedding medium (Tissue-Tek, Sakura Finetek, Torrance, CA) and frozen in pre-chilled methylbutane. Generation of rabbit anti-α8 peptide polyclonal antiserum was previously described (Schnapp et al., 1995a). Sections were incubated overnight at 4°C with α8 antibody (1:16,000 to 1:32,000 dilution) followed by 1 h room temperature incubation with biotinylated goat anti-rabbit IgG antibody (1.5 μg/ml) (Vector). Immunoperoxidase complexes were formed using Vectastain Elite ABC Kit (Vector). Color development was performed with DAB (Sigma, St. Louis, MO) for 10 min followed by DAB enhancer (Zymed) for 1 min. Sections were counterstained with Hematoxylin (Vector), serially dehydrated and mounted with Permount (Sigma).

Acknowledgements

This work was supported by AHA Grant-in-Aid UW Royalty Research Fund (LMS), NSF DBI-0217241 and NIH GM068152 (RS) and SeattleSNPs NHLBI Program for Genomic Applications, to UW-FHCRC, Seattle, WA.

References

Bouzeghrane, F., Mercure, C., Reudelhuber, T.L., Thibault, G., 2004. Alpha8beta1 integrin is upregulated in myofibroblasts of fibrotic and scarring myocardium. J. Mol. Cell Cardiol. 36, 343–353.

Cannizzaro, L.A., Shi, G., 1997. Fluorescent in situ hybridization (FISH) for DNA probes in the interphase and metaphase stages of the cell cycle. Methods Mol. Biol. 75, 313–322.

Collins, F.S., Guyer, M.S., Charkravarti, A., 1997. Variations on a theme: cataloging human DNA sequence variation. Science 278, 1580–1581.

Debacker, C., Catala, M., Labastie, M.C., 1999. Embryonic expression of the human GATA-3 gene. Mech. Dev. 85, 183–187.

de Melker, A.A., Sonnenberg, A., 1999. Integrins: alternative splicing as a mechanism to regulate ligand binding and integrin signaling events. BioEssays 21, 499–509.

Hartner, A., Schocklmann, H., Prols, F., Muller, U., Sterzel, R.B., 1999. Alpha8 integrin in glomerular mesangial cells and in experimental glomerulonephritis. Kidney Int. 56, 1468–1480.

Hartner, A., Cordasic, N., Klanke, B., Muller, U., Sterzel, R.B., Hilgers, K.F., 2002. The alpha8 integrin chain affords mechanical stability to the glomerular capillary tuft in hypertensive glomerular disease. Am. J. Pathol. 160, 861–867.

Hasegawa, T., Hasegawa, Y., Aso, T., Koto, S., Nagai, T., Tsuchiya, Y., Kim, K.C., Ohashi, H., Wakui, K., Fukushima, Y., 1997. HDR syndrome (hypoparathyroidism, sensorineural deafness, renal dysplasia) associated with del(10)(p13). Am. J. Med. Genet. 73, 416–418.

Heidenreich, R., Eisman, R., Surrey, S., Delgrosso, K., Bennett, J.S., Schwartz, E., Poncz, M., 1990. Organization of the gene for platelet glycoprotein IIb. Biochemistry 29, 1232–1244.

Humphries, M.J., McEwan, P.A., Barton, S.J., Buckley, P.A., Bella, J., Paul Mould, A., 2003a. Integrin structure: heady advances in ligand binding, but activation still makes the knees wobble. Trends Biochem. Sci. 28, 313–320.

Humphries, M.J., Symonds, E.J., Mould, A.P., 2003b. Mapping functional residues onto integrin crystal structures. Curr. Opin. Struck. Biol. 13, 236–243.

Hung, L.H., Samudrala, R., 2003. PROTINFO: secondary and tertiary protein structure prediction. Nucleic Acids Res. 31, 3296–3299.

Hynes, R., 2002. Integrins. Bidirectional, allosteric signaling machines. Cell 110, 673.

Karis, A., Pata, I., van Doorninck, J.H., Grosveld, F., de Zeeuw, C.I., de Caprona, D., Fritzsch, B., 2001. Transcription factor GATA-3 alters pathway selection of olivocochlear neurons and affects morphogenesis of the ear. J. Comp. Neurol. 429, 615–630.

Kruglyak, L., Nickerson, D.A., 2001. Variation is the spice of life. Nat. Genet. 27, 234–236.

Labastie, M.C., Catala, M., Gregoire, J.M., Peault, B., 1995. The GATA-3 gene is expressed during human kidney embryogenesis. Kidney Int. 47, 1597–1603.

Landegren, U., Nilsson, M., Kwok, P.Y., 1998. Reading bits of genetic information: methods for single-nucleotide polymorphism analysis. Genome Res. 8, 769–776.

Levine, D., Rockey, D.C., Milner, T.A., Breuss, J.M., Fallon, J.T., Schnapp, L.M., 2000. Expression of the integrin alpha8beta1 during pulmonary and hepatic fibrosis. Am. J. Pathol. 156, 1927–1935.

Lim, K.C., Lakshmanan, G., Crawford, S.E., Gu, Y., Grosveld, F., Engel, J.D., 2000. Gata3 loss leads to embryonic lethality due to noradrena-

line deficiency of the sympathetic nervous system. Nat. Genet. 25, 209–212.

Littlewood Evans, A., Muller, U., 2000. Stereocilia defects in the sensory hair cells of the inner ear in mice deficient in integrin alpha8beta1. Nat. Genet. 24, 424–428.

Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O., Fields, C., 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res. 20, 4255–4262.

Muller, U., Wang, D., Denda, S., Menese, J.J., Pedersen, R.A., Reichardt, L.F., 1997. Integrin alpha8 beta 1 is critically important for epithelial–mesenchymal interactions during kidney morphogenesis. Cell 88, 603–613.

Nakai, K., Kanehisa, M., 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. Genomics 14, 897–911.

Ng, P.C., Henikoff, S., 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814.

Nickerson, D.A., Tobe, V.O., Taylor, S.L., 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. 25, 2745–2751.

Nielsen, H., Engelbrecht, J., Brunak, S., von Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6.

Pandolfi, P.P., Roth, M.E., Karis, A., Leonard, M.W., Dzierzak, E., Grosveld, F.G., Engel, J.D., Lindenbaum, M.H., 1995. Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. Nat. Genet. 11, 40–44.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., Lander, E.S., 2001. Linkage disequilibrium in the human genome. Nature 411, 199–204.

Rieder, M.J., Taylor, S.L., Clark, A.G., Nickerson, D.A., 1999. Sequence variation in the human angiotensin converting enzyme. Nat. Genet. 22, 59–62.

Samudrala, R., Moult, J., 1998. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J. Mol. Biol. 275, 895–916.

Samudrala, R., Levitt, M., 2002. A comprehensive analysis of 40 blind protein structure predictions. BMC Struct. Biol. 2, 3.

Schnapp, L.M., Breuss, J.M., Ramos, D.M., Sheppard, D., Pytela, R., 1995a. Sequence and tissue distribution of the human integrin alpha 8 subunit: a beta 1-associated alpha subunit expressed in smooth muscle cells. J. Cell. Sci. 537–544.

Schnapp, L.M., Hatch, N., Ramos, D.M., Klimanskaya, I.V., Sheppard, D., Pytela, R., 1995b. The human integrin alpha 8 beta 1 functions as a receptor for tenascin, fibronectin, and vitronectin. J. Biol. Chem. 270, 23196–23202.

Springer, T.A., 1997. Folding of the N-terminal, ligand-binding region of integrin alpha-subunits into a beta-propeller domain. Proc. Natl. Acad. Sci. U. S. A. 94, 65–72.

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989.

Van Esch, H., Groenen, P., Nesbit, M.A., Schuffenhauer, S., Lichtner, P., Vanderlinden, G., Harding, B., Beetz, R., Bilous, R.W., Holdaway, I., et al., 2000. GATA3 haplo-insufficiency causes human HDR syndrome. Nature 406, 419–422.

Wang, W., Wu, W., Desai, T., Ward, D.C., Kaufman, S.J., 1995. Localization of the alpha 7 integrin gene (ITGA7) on human chromosome 12q13: clustering of integrin and Hox genes implies parallel evolution of these gene families. Genomics 26, 568–570.

Yauch, R.L., Kazarov, A.R., Desai, B., Lee, R.T., Hemler, M.E., 2000. Direct extracellular contact between integrin alpha 3beta 1 and TM4SF protein CD151. J. Biol. Chem. 275, 9230–9238.