JCI The Journal of Clinical Investigation

Retroviral integration and human gene therapy

Frederic D. Bushman

J Clin Invest. 2007;117(8):2083-2086. https://doi.org/10.1172/JCI32949.

Commentary

Long-term correction of genetic diseases requires permanent integration of therapeutic genes into chromosomes of affected cells. Retroviral vectors are the most widely used delivery vehicles because of their efficiency and precision of integration. However, retroviral integration can take place at a variety of chromosomal sites, and examples have been reported of integration of therapeutic vectors activating oncogenes and causing cancer in patients. This issue of the *JCI* presents three articles that update successful human gene therapy trials and furthermore evaluate the sites of integration in cells from treated patients, including samples from individuals experiencing serious adverse events following therapy (see the related articles beginning on pages 2225, 2233, and 2241).

Retroviral integration and human gene therapy

Frederic D. Bushman

Department of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA.

Long-term correction of genetic diseases requires permanent integration of therapeutic genes into chromosomes of affected cells. Retroviral vectors are the most widely used delivery vehicles because of their efficiency and precision of integration. However, retroviral integration can take place at a variety of chromosomal sites, and examples have been reported of integration of therapeutic vectors activating oncogenes and causing cancer in patients. This issue of the *JCI* presents three articles that update successful human gene therapy trials and furthermore evaluate the sites of integration in cells from treated patients, including samples from individuals experiencing serious adverse events following therapy (see the related articles beginning on pages 2225, 2233, and 2241).

The first unambiguous success for human gene therapy was reported in 2000 by Marina Cavazzana-Calvo, Alain Fischer, and their coworkers (1). Two boys with X-linked SCID (SCID-X1) caused by a genetic defect in a cytokine receptor gene were successfully treated by restoring the missing coding region. Previous attempts to treat patients with adenosine deaminase–deficient SCID (ADA-SCID) had been widely reported, but how much the gene therapy treatments benefited the patients was never fully clarified (the patients remained on effective enzyme replacement therapy throughout) (2). The gene therapy treatments for SCID-X1, in contrast, were stunningly successful. Adding back the gene for the missing cytokine chain restored a proliferation signal, allowing the gene-modified cells to expand in number and repopulate the treated patients, resulting in clear-cut reconstitution of immune function. Ten patients were ultimately treated by the French team of Cavazzana-Calvo and Fischer (ref. 3 and references therein). Subsequently, another group treated a further ten SCID-X1 patients (ref. 4 and references therein), a third group successfully treated two patients with chronic granulomatous disease (CGD) (5), and a fourth group achieved unambiguous success treating five ADA-SCID patients (6). At last, gene therapy had arrived.

However, with these successes came the first serious adverse events in retrovirus-based gene therapy. Three of the SCID-X1 patients treated by the French team developed a leukemia-like lymphoproliferative disease (7, 8). In the two adverse events reported in detail to date, integration of the therapeutic vector, a gammaretrovirus derivative, activated transcription of the LIM domain only 2 (*LMO2*) protooncogene. Two additional factors appear to have contributed to transformation in addition to insertional activation of *LMO2*: an activation signal from the newly introduced cytokine chain and a third "hit" in the form of a chromosomal rearrangement. One patient died of the lymphoproliferative disease. The other two patients responded to chemotherapy and are doing well, and furthermore continue to benefit from the gene therapy.

In this issue of the *JCI*, we get an update on the two SCID-X1 trials and the successful ADA trials, together with a detailed look at the distribution of genomic sites targeted for integration by the retroviral vectors in patient cells. The analysis of integration site patterns is useful in several respects. From a basic science perspective, such data add to our understanding of the molecular mechanisms of retroviral integration. Clinically, following integration site distributions over time helps elucidate the effects of insertional activation on persistence of the selected surviving clones of modified cells. Potentially most important is the possibility, although not yet realized, that such information can help forecast incipient adverse events.

Target site selection during retroviral integration

The genomic features guiding retrovirus integration site selection have now been characterized in some detail (9). The gammaretroviruses, the type of retrovirus adapted for use as vectors in all three gene therapy trials, favor integration near the 5′ ends of transcription units and associated CpG islands (10). In contrast, the lentiviruses — the group that includes HIV — strongly favor integration within active transcription units and show no particular favoring of gene 5′ ends (11, 12). The alpharetroviruses, exemplified by the avian sarcoma–leukemia virus group, show fairly random integration, with only weak favoring of transcription units (12, 13). The integration site specificity of gammaretroviruses is determined primarily by the viral integrase protein, as demonstrated by studies of integration targeting in hybrid retroviruses with "transplanted" integrase coding sequences (14).

Studies of integration site distribution in gene therapy patients capture another dimension. Not only can the initial distribution of integration sites in patient cells be analyzed, but for cells recovered from patients at varying times after therapy, it is possible to ask how selective forces in vivo have altered the distribution of integration site populations. One important question is whether vector integration near genes involved in growth control could have promoted cellular proliferation, and whether this may foreshadow malignant transformation. In the CGD trial, cells recovered from patients were enriched in integration sites in several known oncogenes, which probably promoted cell growth and contributed to the success of the gene therapy (5). So far there are no reported adverse events in the two patients from this trial, but there is considerable danger that further genotoxic events in the gene-corrected cells could result in transformation.

Integration targeting in successful gene therapy trials

The articles in this issue of the *JCI* begin by updating us on the status of gene therapy patients from the SCID-X1 and ADA-SCID trials. Schwarzwaelder et al. (4) report that the patients treated in the British SCID-X1 trial continue to do well, and Deichmann et al. (3) update the first and longer-running SCID-X1 trial. At present there is no expla-

Nonstandard abbreviations used: ADA-SCID, adenosine deaminase–deficient SCID; CGD, chronic granulomatous disease; SCID-X1, X-linked SCID.

Conflict of interest: The author has declared that no conflict of interest exists.

Citation for this article: *J. Clin. Invest.* **117**:2083–2086 (2007). doi:10.1172/JCI32949.

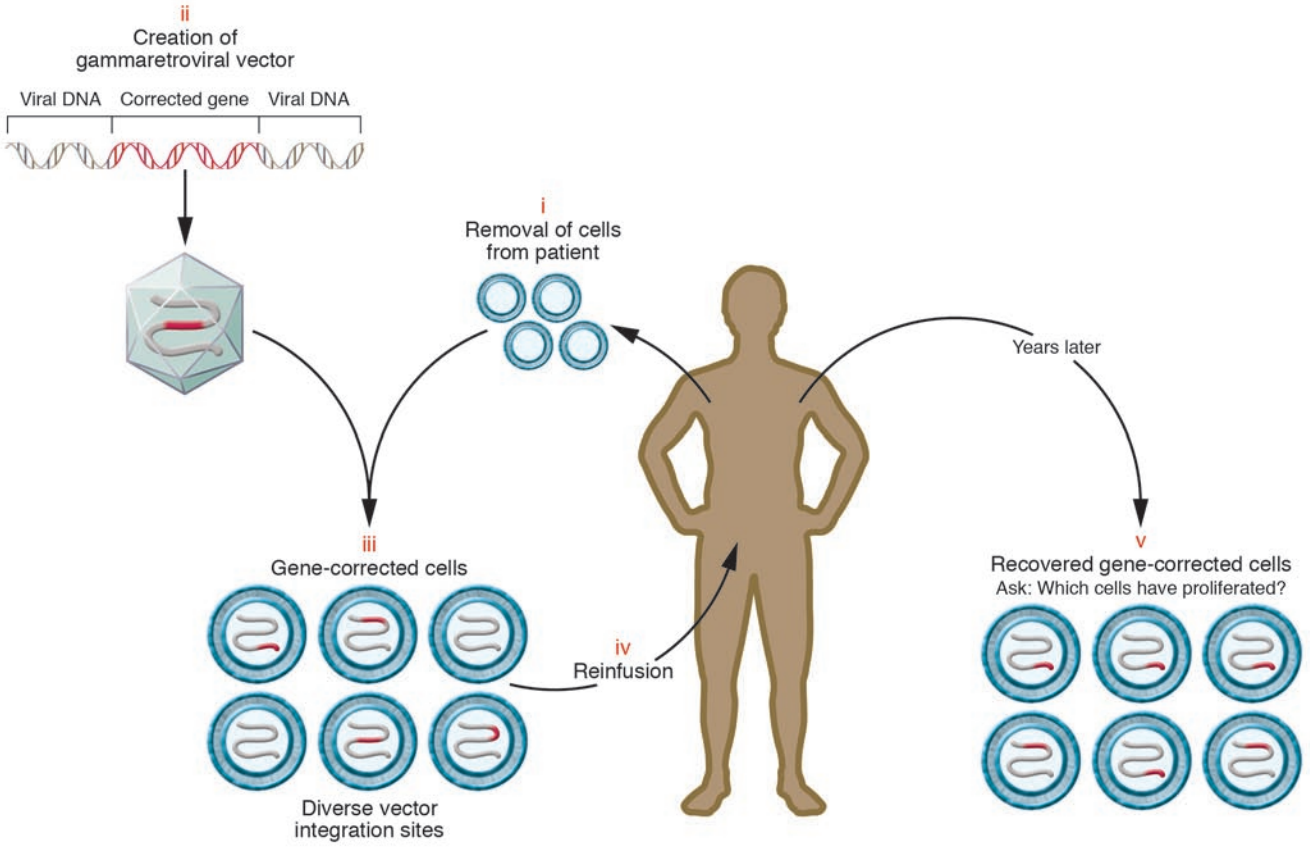

Figure 1
Gene correction using retroviral vectors, and monitoring the placement of integration sites after evolution in patients. A population of cells is removed from the patient (i) and transduced ex vivo with a retroviral vector carrying the corrected gene (ii–iii). The gene-corrected cells are subsequently reinfused into the patient (iv). The population of vector integration sites is diverse prior to transplantation, with each cell harboring integrated proviral copies at different locations in the genome. Examination years later of the population of gene-corrected cells reveals that the population of provirus has changed because of evolution of cells in vivo (v). In some cases this can include proliferation of cells where integrated proviruses activated oncogenes, leading to an adverse event.

nation for why adverse events occurred in one SCID-X1 trial and not the other. Less time has elapsed since cell infusion for the patients from the second trial, so differences are still hard to evaluate. The numbers of patients are still very low, so low, in fact, that the differences between the two trials are not statistically significant. Even if the Cavazzana-Calvo and Fischer et al. (1) trial experiences a fourth adverse event, so that 4 of 10 patients suffer adverse events versus 0 of 10 in the Thrasher trial (4), the difference still is not statistically significant. In the third related study reported in this issue of the *JCI*, Aiuti et al. (15) report that the five ADA-SCID patients continue to do well up to 47 months after cell infusion and are free of adverse events.

The three studies (3, 4, 15) then go on to investigate the locations of the integrated vectors in patient cells (Figure 1). All three studies used gammaretrovirus-based vectors, and the global integration patterns were as expected based on previous work using cell culture models (10, 14). Integration sites were enriched near gene 5′ ends and associated CpG islands compared with expectation based on random placement of sites. Analysis of the activity of genes near integration sites using transcriptional profiling data from the cell types transduced ($CD34^+$ progenitor cells) indicated that active genes were particularly favored for integration.

Integration of retroviral DNA marks the multipotent progenitor cells that then give rise to the mature cell types of the hematopoietic lineage. Thus it is possible to investigate what types of progenitors can contribute to each lineage by asking whether the same integration sites, presumably derived by infection of a single multipotent progenitor cell, appear in different lineages. Aiuti et al. (15) separated cells into myeloid (granulocytes) and lymphoid (B and T cells) lineages and found integration sites in common between the two lineages. Similar data has been previously reported for the SCID-X1 trial (16). A technical concern is that imperfect cell fractionation could have yielded the observed result, but the results were obtained for multiple integration sites in several samples. Thus the simplest interpretation is that the therapeutic vector, at least on occasion, integrated in a precursor cell capable of giving rise to multiple lineages.

Integration site monitoring in clinical management

The key question underlying the integration site studies, however, centers on understanding the role of retroviral integration and outgrowth of specific cellular clones, which may possibly aid in identifying impending adverse events. All three of

the present studies investigated this issue (3, 4, 15). One approach involves analysis of clustering of integration sites. To what degree are integration sites clustered in the patient samples, and does increased clustering — for example, near the promoters of oncogenes — indicate possible future malignant transformation?

Aiuti et al. (15) reported a significant enrichment of integration sites near transcription start sites in transduced cells recovered from patients compared with the cell population prior to infusion, indicating evolution in vivo. Similarly, Schwarzwaelder et al. (4) and Deichmann et al. (3) reported a significant increase in the proportion of integration sites clustered near promoters (referred to as common integration sites) in cells from patients. However, comparison of the cell populations from healthy patients with cells of patients that experienced adverse events during the SCID-X1 trial did not show a significant difference (3). Thus the data from the three articles supports the idea that there are consistent changes in the populations of integration sites resulting from selective pressures on cells after reinfusion in patients. As mentioned above, data from the CGD trial provides a particularly striking example of this (5). There is not yet any support for the idea that information about integration site distribution contains any data predictive of adverse events. An unresolved question is whether the behavior of a large population of gene-corrected cells contains information that could help predict the transformation events that take place in a single cell. Given that there are steps toward transformation that do not involve vector integration, such as the chromosomal rearrangements seen in the first two adverse events in the SCID-X1 trial, it seems that integration studies will, at best, be only partially predictive in nature. Still, work in this area is just beginning, and it is hard to predict how useful this information will ultimately be.

Some of the bioinformatic methods utilized here are in early stages of development and will likely evolve in future studies. One ongoing challenge is controlling the increased likelihood of false-positive calls as investigators ask more and more questions of a single data set. All three articles in this issue analyze the types of genes associated with integration sites, referred to as gene ontology. Many ontology classes have been delineated in the course of annotating genome sequences. Comparing integration site populations over the many ontology classes increases the chances of obtaining a significant association by chance. Aiuti et al. (15) took the precaution of applying the Bonferroni correction, which makes it incrementally harder to obtain significant *P* values as the number of comparisons increases. The other two studies did not apply such a correction (3, 4). Many classes of genes are classified as significantly enriched in these studies, but some of these probably will not survive a correction for multiple comparisons. The problem of controlling inflation of error pops up in various other statistical tests of genomic features affecting integration, providing an interesting challenge for future development of bioinformatic methods. Similarly, further methods for assessing the statistical significance of clustering will likely also be useful. A number of groups have published further bioinformatic methods for evaluating integration targeting (e.g., refs. 17, 18).

Evolution of integration strategies

Returning to the biology of retroviral integration, the results from the gene therapy trials provide food for thought on the evolution of integration targeting in the gammaretroviruses. There are many examples of integrating genomic parasites in which their targeting strategies have evolved to optimize their interactions with the host cell (19, 20). For HIV, there is evidence that integration in active transcription units optimizes subsequent proviral gene expression (21, 22). This strategy makes sense in the context of HIV infection, because HIV-infected cells persist only for a day or two, so it is important to integrate in an optimal location for rapid production of progeny virions in the short time available. It has been less clear why gammaretroviruses would integrate near gene 5′ ends — so far there is no evidence that this maximizes gammaretroviral gene expression, although this remains a possibility. Scientists studying insertional activation of oncogenes by gammaretroviruses in animal models have long wondered whether oncogene activation might confer some advantage to the virus. Data from the three gene therapy trials supports the idea that the gammaretroviral integration near gene 5′ ends can promote outgrowth of cells harboring proviruses at these sites. This fits with the idea that promoting proliferation of infected cells may be a mechanism of replicating proviral genomes in addition to de novo infection. Although the gene therapy vectors used are obviously incapable of multicycle replication, the integration-targeting biases of the original viruses is preserved in the vectors. Similarly, human T lymphotropic virus 1 is thought to replicate as much by promoting proliferation of infected cells as by infecting new cells (23). If gammaretroviruses have evolved to activate genes for growth control by integrating into the genes' promoters, then these would be the worst possible choice for gene therapy vectors — the increasingly popular lentiviral vectors may be a safer alternative (24, 25).

In conclusion, the three studies of integration sites reported in this issue of the *JCI* (3, 4, 15) provide rich data on the genetic consequences of manipulating human cells and provide a point of departure for investigating whether such information can be used to diminish the dangers of human gene therapy.

Address correspondence to: Frederic D. Bushman, Department of Microbiology, University of Pennsylvania School of Medicine, 3610 Hamilton Walk, Philadelphia, Pennsylvania 19104-6076, USA. Phone: (215) 573-8372; Fax: (215) 573-4856; E-mail: bushman@mail.med.upenn.edu.

1. Cavazzana-Calvo, M., et al. 2000. Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science.* **288**:669–672.
2. Blaese, R.M., et al. 1995. T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science.* **270**:475–480.
3. Deichmann, A., et al. 2007. Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J. Clin. Invest.* **117**:2225–2232. doi:10.1172/JCI31659.
4. Schwarzwaelder, K., et al. 2007. Gammaretrovirus-mediated correction of SCID-X1 is associated with skewed vector integration site distribution in vivo. *J. Clin. Invest.* **117**:2241–2249. doi:10.1172/JCI31661.
5. Ott, M.G., et al. 2006. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* **12**:401–409.
6. Aiuti, A., et al. 2002. Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning. *Science.* **296**:2410–2413.
7. Hacein-Bey-Abina, S., et al. 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N. Engl. J. Med.* **348**:255–256.
8. Hacein-Bey-Abina, S., et al. 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science.* **302**:415–419.
9. Bushman, F., et al. 2005. Genome-wide analysis of retroviral DNA integration. *Nat. Rev. Microbiol.* **3**:848–858.
10. Wu, X., Li, Y., Crise, B., and Burgess, S.M. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science.* **300**:1749–1751.
11. Schröder, A.R., et al. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* **110**:521–529.
12. Mitchell, R.S., et al. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site

preferences. *PLoS Biol.* **2**:e234.
13. Narezkina, A., et al. 2004. Genome-wide analyses of avain sarcoma virus integration sites. *J. Virol.* **78**:11656–11663.
14. Lewinski, M.K., et al. 2006. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* **2**:e60.
15. Aiuti, A., et al. 2007. Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with stem cell gene therapy. *J. Clin. Invest.* **117**:2233–2240. doi:10.1172/JCI31666.
16. Schmidt, M., et al. 2005. Clonal evidence for the transduction of CD34$^+$ cells with lymphomyeloid differentiation potential and self-renewal capacity in the SCID-X1 gene therapy trial. *Blood.* **105**:2699–2706.
17. Berry, C., Hannenhalli, S., Leipzig, J., and Bushman, F.D. 2006. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput. Biol.* **2**:e157.
18. Wu, X., and Burgess, S.M. 2004. Integration target site selection for retroviruses and transposable elements. *Cell. Mol. Life Sci.* **61**:2588–2596.
19. Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M. 2002. *Mobile DNA II.* ASM Press. Washington, DC, USA. 1204 pp.
20. Bushman, F.D. 2001. *Lateral DNA transfer: mechanisms and consequences.* Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York, USA. 448 pp.
21. Lewinski, M., et al. 2005. Genome-wide analysis of chromosomal features repressing HIV transcription. *J. Virol.* **79**:6610–6619.
22. Jordan, A., Bisgrove, D., and Verdin, E. 2003. HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* **22**:1868–1877.
23. Coffin, J.M., Hughes, S.H., and Varmus, H.E. 1997. *Retroviruses.* Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York, USA. 843 pp.
24. Levine, B.L., et al. 2006. Gene transfer in humans using a conditionally replicating lentiviral vector. *Proc. Natl. Acad. Sci. U. S. A.* **103**:17372–17377.
25. Naldini, L., et al. 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science.* **272**:263–267.

When EGF is offside, magnesium is wasted

Shmuel Muallem[1] and Orson W. Moe[1,2,3]

[1]Department of Physiology, [2]Department of Internal Medicine, and [3]Charles and Jane Pak Center for Mineral Metabolism, University of Texas Southwestern Medical Center, Dallas, Texas, USA.

Our understanding of magnesium (Mg^{2+}) regulation has recently been catapulted forward by the discovery of several disease loci for monogenic disorders of Mg^{2+} homeostasis. In this issue of the *JCI*, Groenestege et al. report that their study of a rare inherited Mg^{2+} wasting disorder in consanguineous kindred shows that EGF acts as an autocrine/paracrine magnesiotropic hormone (see the related article beginning on page 2260). EGF stimulates Mg^{2+} reabsorption in the renal distal convoluted tubule (DCT) via engagement of its receptor on the basolateral membrane of DCT cells and activation of the Mg^{2+} channel TRPM6 (transient receptor potential cation channel, subfamily M, member 6) in the apical membrane. These authors show that a point mutation in pro-EGF retains EGF secretion to the apical but not the basolateral membrane, disrupting this cascade and causing renal Mg^{2+} wasting. This work is another seminal example of the power of the study of monogenic disorders in the quest to understand human physiology.

Magnesium (Mg^{2+}) is a critical cofactor in many enzymatic reactions and as such participates in all cellular functions. It is the second most common intracellular ion and the fourth most abundant cation in the body, and plasma and cellular Mg^{2+} concentrations are both tightly controlled. The renal regulation of Mg^{2+} excretion can range from 100% reabsorption of the filtered load (0% excretion) to excretion of greater than 100% of the filtered load (renal secretion) under experimental conditions (1, 2). This extraordinary homeostatic feat performed by the kidney still evades our comprehension after several decades of investigation. The initial advances in our understanding of Mg^{2+} handling stemmed from clearance, micropuncture, microcatheterization, and microperfusion studies. While these experiments furnished the key foundations of understanding the regulation of Mg^{2+} balance, this process is surprisingly poorly understood at the cellular and molecular levels, largely due to a lack of good surrogate cell model systems and a slow rate of emergence and hence paucity of cDNAs and specific reagents for Mg^{2+} homeostatic proteins. Almost all of the seminal progress in enlightening our understanding of the molecular mechanisms of Mg^{2+} handling arose from identification of disease loci of rare human monogenic Mg^{2+} disorders (3–10). The discovery reported by Groenestege et al. in this issue of the *JCI* (11) appends a new page to this catalog of pedagogical disorders. These authors show that EGF is an autocrine/paracrine magnesiotropic hormone that regulates renal Mg^{2+} reabsorption by regulating the activity of the Mg^{2+}-permeable channel TRPM6 (transient receptor potential cation channel, subfamily M, member 6). They go on to demonstrate that a point mutation in pro-EGF that disrupts sorting of the protein to the basolateral membrane of distal convoluted tubule (DCT) cells in kidney nephrons and thus release of EGF to the basolateral space or inhibition of the EGFR by anti-EGFR antibodies led to suppressed activity of TRPM6 and renal Mg^{2+} wasting in humans.

Nonstandard abbreviations used: DCT, distal convoluted tubule; HSH, hypomagnesemia with secondary hypocalcemia; IRH, isolated recessive renal hypomagnesemia; NCC, NaCl cotransporter; TRPM6, transient receptor potential cation channel, subfamily M, member 6.

Conflict of interest: The authors have declared that no conflict of interest exists.

Citation for this article: *J. Clin. Invest.* **117**:2086–2089 (2007). doi:10.1172/JCI33004.

Mg^{2+} homeostasis

The systemic balance of Mg^{2+} and its intracellular concentration are determined by intestinal absorption and renal excretion. The main site of intestinal Mg^{2+} absorption is the small bowel, with some additional absorption in the large bowel. Renal handling commences with glomerular filtration of the non–protein bound plasma fraction (free and complex) followed by passive absorption through the paracellular pathway in the proximal tubule and the thick ascending loop of Henle and active transcellular absorption by the DCT (Figure 1) (12). The molecular mechanism of these processes remained elusive for many years until identification of disease genes underlying hereditary Mg^{2+} homeostatic disorders. Analysis of the mutations leading to familial hypomagnesemia with hypercalciuria and nephrocalcinosis (FHHNC) disclosed that passive Mg^{2+} absorption by the thick ascending