Introduction

'Saint-Adolf-Grand-Grand-God-Father' – Switzerland, 1915, Adolf Wölfli.

Adolf Wölfli (1864 – 1930), who at 31 was hospitalised with symptoms of schizophrenia, remained at the Waldau Asylum near Bern, until his death 35 years later. He is regarded as one of the foremost artists in the Art Brut movement. He produced a huge number of works, including a semi-autobiographical epic which eventually stretched to 45 volumes, containing a total of over 25000 pages and 1600 illustrations, which tells how he transformed himself from a child to 'Knight Adolf' to 'Emperor Adolf' and finally to 'St Adolf II'. Today his art is on display at the Museum of Fine Arts in Bern. His music compositions inspired many, including the Danish composer Per Nørgård, who wrote the opera 'The Divine Circus' on Wölfli's life.

A. Mental retardation

It has now been indisputably proven that genetics plays an important role in the aetiology of MR. However, genetic causes underlying MR remain largely elusive, as mental dysfunction represents a set of very heterogeneous conditions. Due to its societal, familial and financial burden, and its high prevalence, mental handicap is one of the major unsolved problems in medicine.

A.1. Definition of mental retardation

Onset during childhood, IQ < 70, diminished ability to adapt to a normal social environment.

People within +/- 2.0 SD of the mean IQ score (i.e. 70 – 130, encompassing 95.4% of the general population) are considered to be within the normal range of intelligence. Individuals with an IQ deviating more than + 2.0 SD from the mean (i.e. IQ > 130) are deemed 'highly gifted', whereas individuals deviating more than -2.0 SD. (i.e. IQ < 70) are said to be mentally retarded (Fig. I-1a).

Given that intelligence is a complex trait influenced by a multitude of different factors (see I.A.4) and that IQ measurements have their limitations and drawbacks as outlined in Appendix A, MR cannot be defined solely as an IQ measure. The WHO therefore defines MR as 'a global deficiency in cognitive function (IQ < 70) with an onset during childhood and associated with a diminished ability to adapt to the daily demands of a normal social environment'. Based on IQ, the WHO classifies MR in four subcategories. However, owing to limitations in determining IQ, most researchers only distinguish between mild (ranging from -2.0 to -3.3 SDs; that is, an IQ between 50 – 69) and severe (the range below -3.3 SDs; that is, an IQ < 50) forms of MR[6] (Fig. I-1a).

A.2. Costs of mental retardation

Costs of MR far exceed those relating to any other class of disease.

Due to its high incidence and associated life-long need for medical care, MR represents a considerable financial load on society. In the Netherlands, for example, health care expenditure for Mental Disorders, as defined by the ICD-9 classification of the WHO, increased to over 6.95 billion Euros in 1994, claiming 25.8% of the total health care budget[7], by far exceeding the costs relating to any other ICD-9 category. With a 9.0% share of the total health care budget, MR alone is almost as expensive as cancer (4.4%), stroke (3.6%) and coronary heart diseases (2.5%) – the main causes of death – together (Fig. I-1b)[7,8]. All economic evaluations are also likely underestimates, since indirect costs, such as lost production, are not taken into account.

More importantly, decreased quality of life and other intangible costs represent a huge burden for patients and their families.

A.3. Aetiology of mental retardation

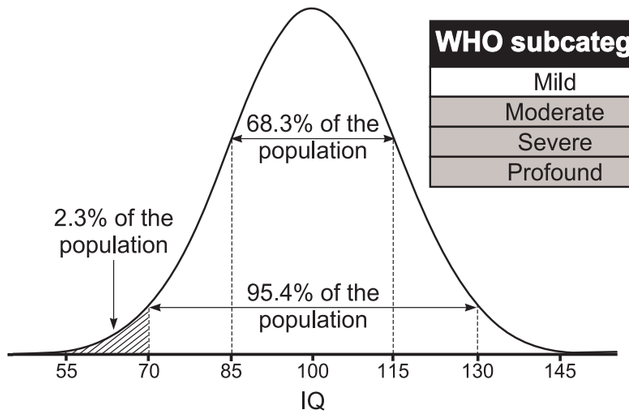
Likely owing to the complexity of the human brain, both in its structure and its function, the underlying causes of MR are manifold.

A.3.1. General considerations

MR can be caused by a multitude of genetic and environmental factors.

MR can be divided into forms with a genetic and a non-genetic or environmental basis. Genetic factors include chromosomal anomalies and monogenic diseases. Infectious diseases (e.g. postnatal meningitis), perinatal anoxia, maternal alcohol abuse during pregnancy or very premature birth are some examples of environmental factors acting prenatally or during early infancy that result in brain injury. Apart from the IQ argument (see I.A.1), it is well validated that mild and severe MR represent two different entities[9]. Mild MR is believed to represent one end of the intelligence range and is often thought to be multifactorial, resulting from a combination of environmental and/or genetic factors (hence, the difficulty in estimating its prevalence rate). It is usually familial and frequently affects people from socially disadvantaged backgrounds[10]. On the other hand, severe MR more commonly results from discrete events having a catastrophic effect on mental development

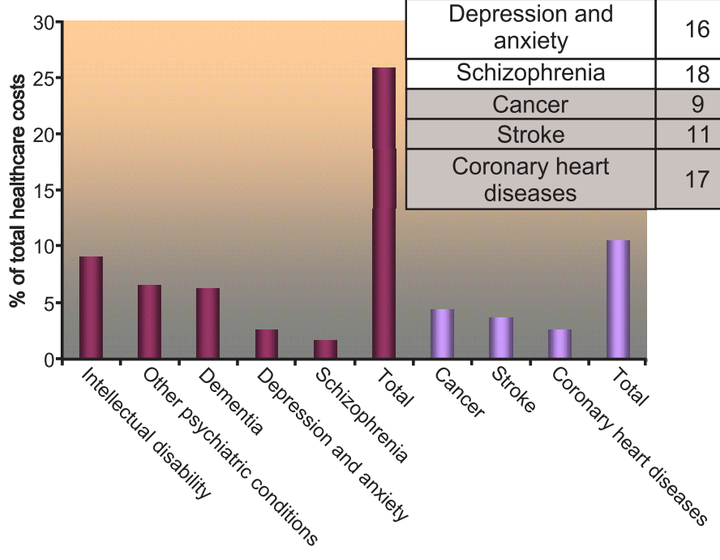
a



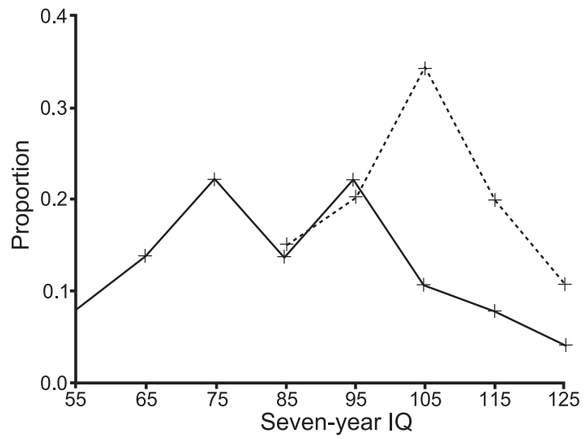
WHO subcategory	IQ	% of MR
Mild	50 - 69	85
Moderate	35 - 49	10
Severe	20 - 34	3.5
Profound	<20	1.5

b

Diagnostic group	Rank	Cost (10^9 €)	% of total costs
Intellectual disability	1	2.43	9.0
Other psychiatric conditions	4	1.75	6.5
Dementia	5	1.67	6.2
Depression and anxiety	16	0.67	2.5
Schizophrenia	18	0.43	1.6
Cancer	9	1.19	4.4
Stroke	11	0.97	3.6
Coronary heart diseases	17	0.67	2.5



c



d

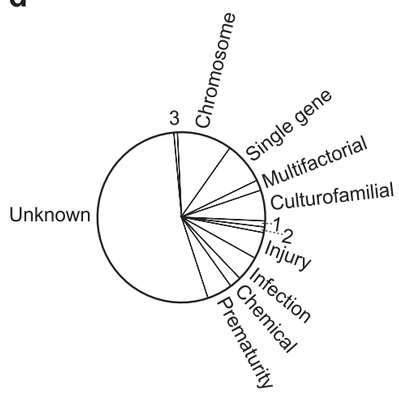


Fig. I-1 | **Cognitive impairment.**

a. Incidence of mental retardation (MR). Assuming a Gaussian distribution of intelligence quotient (IQ) with a mean of 100 and a standard deviation (SD) of 15, 68.3% and 95.4% of the general population have IQs deviating one and two SDs from the mean, respectively. By definition, 2.3% of the general population have an IQ below 70 (dashed area), which categorises them as mentally retarded. The tenth version of the international classification of diseases (ICD-10) of the World Health Organisation (WHO) subdivides MR in four categories[6]. For reasons detailed in the text, this study only distinguishes between mild and severe (shaded) MR.
b. Costs of MR. The bar graph plots the costs for several ICD-9 disease classes as a percentage of the total health care budget. Aubergine represents diseases affecting mental health and violet main causes of death. The table shows the ranking within ICD-9, and the absolute and relative costs of those diseases presented in the diagram. Shading accentuates details on the main causes of death.
c. Heritability of MR. Whereas siblings of mildly mentally retarded children (full line) tend to have lower than normal IQs, siblings of severely retarded individuals (dashed line) tend to have normal IQs, suggesting that mild MR is heritable, but severe MR is not.
d. Aetiology of MR. A pie chart shows 11 causation categories based on 10997 mentally retarded patients studied by Stevenson *et al.*[11]. These categories are: other environmental insults (< 1%), chromosomal aberrations (11%), single-gene defects (8%), multifactorial origins (2%), culturofamilial conditions (6%), syndromes with a presumed genetic deficit (1%), other genetic syndromes (1%), injuries (5%), infections (5%), chemical insults (2%) and prematurity (5%). For the majority of cases (56%), no cause for the MR could be established.

Data are converted from *8* and *7* (b), and graphs adapted from *12* (c) and *11* (d).

(e.g. trauma) and patients' social class distribution approximates that of the general population[13]. Consistent with these observations, siblings of severely retarded children tend to have normal IQs, whereas siblings of mildly retarded children tend to have lower than normal IQs, again suggesting that mild MR is familial and perhaps heritable, but severe MR is not (Fig. I-1c). Importantly, biomedical causes are identified far more frequently in the severely mentally impaired than in the mildly affected[14].

A positive correlation between MR and malformations has been observed in several studies[11,15,16], indicating that the structural and mental abnormalities originate during embryonic development. Because of the somatic manifestations characteristic in S-MR, this form is much more amenable to epidemiological studies focusing on aetiology than is NS-MR, although studies have investigated the aetiology of idiopathic MR[17]. It is likely that the over 50% of cases with unknown causation in a study of 10997 MR individuals by Stevenson *et al.* represents a significant number of non-syndromic cases[11]. In their study, the authors distinguish six genetic causation categories and eleven environmental causation categories. The study shows that 63% of cases in which a cause could be identified have an underlying genetic defect, with 28% of genetic causality resulting from single-gene defects. A disproportionate fraction (42%) of monogenic causes were X-linked, with 52% of these accounted for by Fragile X syndrome (OMIM 309550). With an overall incidence of 1.7%, this syndrome is the single most common cause of MR. The most common autosomal disorder, tuberous sclerosis (OMIM 191100), represented 1.3% of the genetic causes and had an overall prevalence of 0.4%. An overview of all these data is presented in Fig. I-1d.

A.3.2. X-linked mental retardation

Due to the obvious mode of inheritance of X-linked conditions, genetic causes underlying MR have primarily been studied in families in which mental deficit segregates with the X chromosome.

A.3.2.1. Biology and DNA sequence of the X chromosome

Sequencing of the X chromosome revealed low gene content, large numbers of interspersed repeats and a low heterozygosity level.

Whereas females inherit an X chromosome from each parent, males only receive a maternal X. Only 54 of the 1098 annotated X-linked genes have a counterpart on

the male-determining Y chromosome[18]. Therefore, males are hemizygous for the vast majority of X-chromosomal genes. This results in the exposure of recessive phenotypes and, together with the relative ease of recognizing an X-chromosomal mode of inheritance, explains the disproportionately large numbers of conditions and traits that have been associated with the X chromosome. To compensate for the sex-dependent difference in gene dosage, one female X chromosome is randomly inactivated. This process is described in detail in the next section.

The recent sequencing of the human X chromosome[18] has added an unprecedented level of detail to the understanding of its biology. A spatial overview of some key features along its length is given in Fig. I-2. The X chromosome is 155 Mb in length, covering 5.4% of the euchromatic genome[19]. With 7.1 genes/Mb, the gene density on the X chromosome is below the genome average. Only 1.7% of its sequence is exonic, accounting for 33% of transcribed chromosome, which is below the numbers calculated for other chromosomes that were annotated using the same procedures at similar dates (Table I-1).

Table I-1 | **Comparison of gene densities on human chromosomes X, 6, 9 and 10**

Chromosome[§]	**X**	**6**	**9**	**10**	**Whole genome**
Number of protein coding genes	1098	1557	1149	1357	22287
Protein coding gene density (Mb^{-1})	7.1	9.2	10.5	10.4	7.7
Protein coding gene coverage (%)	33	42	47	51	25.1
Protein coding exon coverage (%)	1.7	2.2	2.5	2.3	1.2

[§] Data for chromosomes X, 6, 9 and 10 from *18* and references therein; whole genome data from *19*, apart from protein coding gene coverage[20].

In line with its low gene content, the X chromosome contains only one predicted CpG island per 5.25 Mb, which is half of the estimated genome average[21]. Cross-species comparison of X chromosomes revealed 4493 evolutionarily conserved regions. Of these, 97.8% correspond to annotated exons and the 100 remaining regions are likely to be non-annotated exons or highly conserved regulatory regions[18].

Interspersed repeats account for 55.6% of the X chromosome's euchromatic sequence, which is remarkably high in comparison to the genome average of 44.7%. LINEs of the L1 family are especially enriched (Fig. I-2d, 28.9% on the X vs. 16.9% genome-wide)[18,21]. This overabundance of L1 LINEs is consistent with their proposed function as 'way stations' for XCI (see I.A.3.2.2.2).

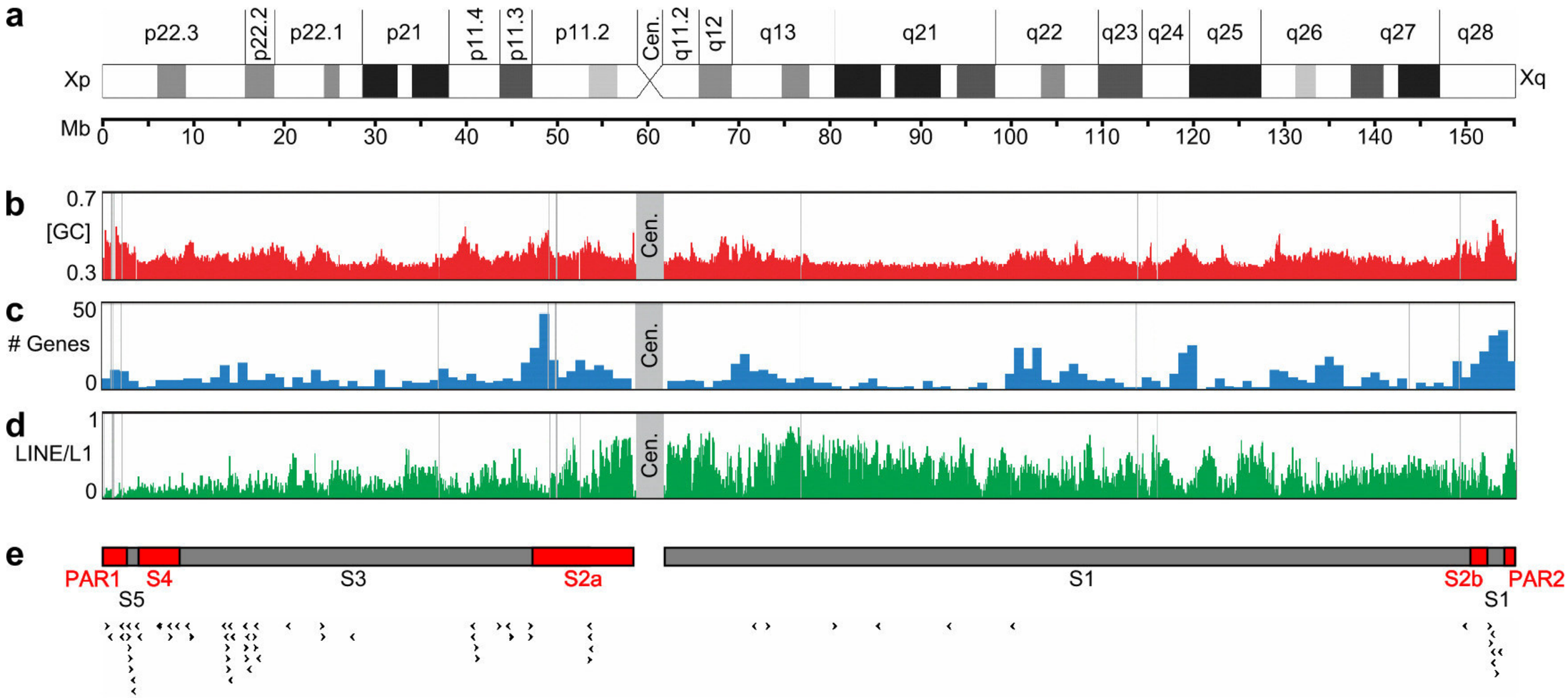
Throughout the X chromosome, 153146 SNPs have been identified. To date, 901 of these result in non-synonymous changes in protein coding sequence, forming the basis for functional protein variants. Using 62334 SNPs of flow-sorted and partially sequenced X chromosomes, Ross *et al.* calculated the heterozygosity level of the X chromosome to be 57% of that of chromosome 20[18], which is in line with the fact that the heterozygosity level of the X chromosome is well below that of the autosomes[22].

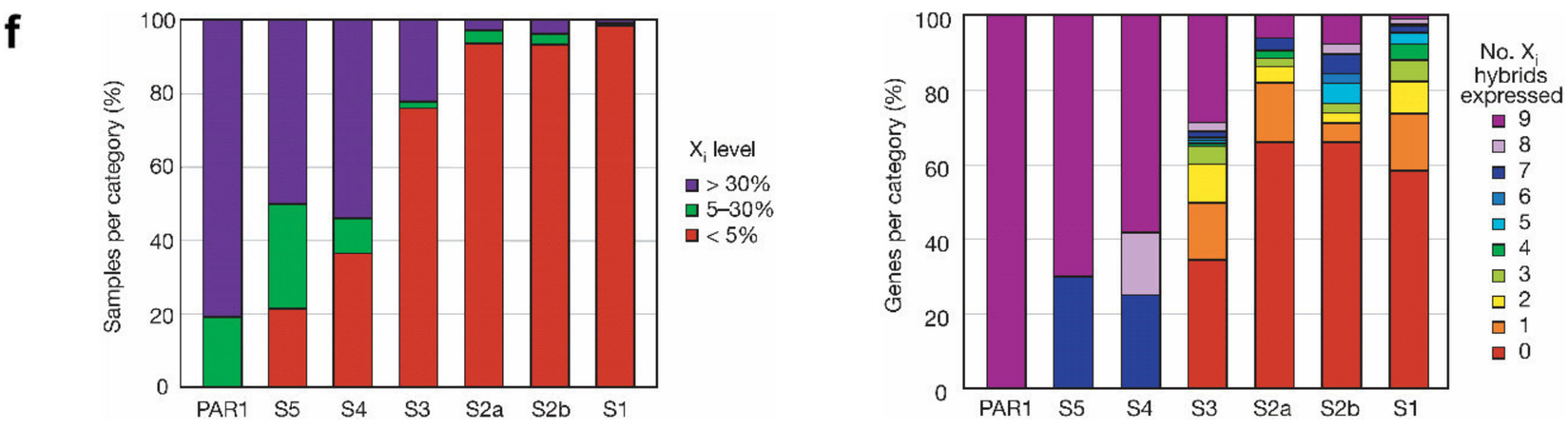
A large part of the low heterozygosity of the X chromosome ($\mu_{\Sigma X}$) may be explained by population genetics. Two factors are notably important in these considerations. First, as males are hemizygous for the X, the mutation rate of the X chromosome (μ_{HX}) is only three-quarters that of the autosomes (μ_{HA}). Second, the mutation rate in the male germ line (μ_{mgl}) is 1.7 to 2.1-fold higher than in the female germ line (μ_{fgl})[21,23]. Due to larger sequence size and longer evolutionary time span used in the analyses, the latter estimate is the more reliable one. The difference in mutation rate between the male and female germ lines is important as the X chromosome only undergoes male meiosis one-third of the time compared to half of the time for the autosomes.

Fig. I-2 | *Next page.* **Spatial overview of some key features along the length of the X chromosome**.

Grey vertical lines in panels b – d represent the centromere (Cen.) and gaps in the euchromatic sequence.
a. Ideogram of the X chromosome from Xp (left) to Xq (right). The sequence scale shows that the X chromosome is 155 Mb in size.
b. Plot of the GC content per 100 kb.
c. Plot of the number of genes, excluding pseudogenes, per Mb.
d. Plot of the fractional coverage of L1 long interspersed nuclear elements (LINE/L1) per 100 kb.
e. While the top panel shows a diagram of the evolutionary strata along the X chromosome (PAR1 & 2, pseudoautosomal regions; S1 – S5, strata 1 – 5), the bottom panel summarises the distribution of genes escaping X chromosome inactivation. Arrowheads indicate the direction of transcription.
f. A subdivision of transcripts according to expression levels from the inactive X chromosome (X_i, left panel) and a classification of genes according to the number of X_i mouse – human somatic cell hybrids expressing them (right panel), shows that the spatial distribution of genes expressed from X_i is non-random.

Adapted from *18* (a – e, bottom), and *24* (e, top and f).


a
p22.3
p22.2
p22.1
p21
p11.4
p11.3
p11.2
Cen.
q11.2
q12
q13
q21
q22
q23
q24
q25
q26
q27
q28
Xp
Xq
Mb
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
b
0.7
[GC]
0.3
Cen.
c
50
Genes
0
Cen.
d
1
LINE/L1
0
Cen.
e
PAR1
S4
S3
S2a
S5
S1
S2b
PAR2
S1


f
Samples per category (%)
100
80
60
40
20
0
PAR1
S5
S4
S3
S2a
S2b
S1
X_i level
> 30%
5–30%
< 5%
Genes per category (%)
No. X_i hybrids expressed
9
8
7
6
5
4
3
2
1
0

Assuming $\frac{\mu_{mgl}}{\mu_{fgl}} = 2.1$:

$$\mu_{Agl} = \frac{\mu_{mgl} + \mu_{fgl}}{2} = 1.55\mu_{fgl}$$

and

$$\mu_{Xgl} = \frac{\mu_{mgl} + 2\mu_{fgl}}{3} = 1.37\mu_{fgl}$$

and hence

$$\frac{\mu_{Xgl}}{\mu_{Agl}} = 0.88$$

Combination with $\mu_{HX} = 0.75\mu_{HA}$, based on the hemizygosity argument, yields

$$\mu_{\Sigma X} = (0.75 \times 0.88)\mu_{\Sigma A} = 0.66\mu_{\Sigma A}$$

These calculations are in agreement with the observed $\mu_{\Sigma X} = 0.61\mu_{\Sigma A}$[22]. Owing to the hemizygosity in males, it is likely that X-linked gene defects are subject to strong selection.

A.3.2.2. X chromosome inactivation

The random transcriptional inactivation of one of the two X chromosomes in every cell of the female mammal, termed XCI, is one of the most intriguing aspects of X chromosome biology. XCI is believed to counteract male hemizygosity for the majority of X-chromosomal genes. The molecular mechanism and pattern of XCI are specified in Appendix B.

A.3.2.2.1. The X chromosome inactivation hypothesis

Occurring early in female development, XCI randomly affects the paternal or maternal X in different cells of the same individual.

Ever since Morgan discovered the sex chromosomes in *Drosophila* (see Appendix C), researchers have been intrigued by the nature of the mechanism that would allow for dosage compensation of sex-linked genes. In 1949, Barr observed a particular female-specific intranuclear entity, the so-called Barr body[25]. In 1960, Ohno and Hauschka showed that these Barr bodies represent single X chromosomes in several murine tissues[26]. Shortly

before, it had been established in the mouse that the Y chromosome carries male-determining factors and that XO mice are fertile, healthy females. Considering all of these observations and combining them with the fact that female mice heterozygous for sex-linked genes affecting coat colour showed a mottled fur resembling that of somatic mosaics, Mary Lyon postulated her XCI hypothesis[27,28]:

- The Barr body is a genetically inactive X chromosome.
- In different cells of the same animal, X chromosomes of either paternal or maternal origin can be inactivated.
- The inactivation occurs early in development.

Several observations in humans were in concordance with her hypothesis. From the study of individuals with supernumeral X chromosomes, it became apparent that the number of Barr bodies was always one less than the number of X chromosomes[29-31] and that 49,XXXXY males were only slightly more affected than 47,XXY males. The hypothesis could also explain the variability seen in females affected with, for example, colour blindness and haemophilia A, both sex-linked traits.

A.3.2.2.2. X chromosome inactivation in X;autosome translocations

In unbalanced X;A translocations, the der(X) forms X_i. Transcriptional silencing in the autosomal material is incomplete and discontinuous but occurs at long distances. In balanced X;A translocations, the normal X is inactivated.

Shortly after Lyon's XCI hypothesis was proposed, classical mouse genetics on X;A translocation carriers indicated that inactivation did not extend in a simple linear fashion to the translocated autosome. Several mouse labs had already noticed (i) that the degree of variegation of an autosomal gene in X;A carriers depends on the distance between that locus and the BP, and (ii) that not all X;A translocations resulted in variegation at the autosomal locus, but that it was crucially dependent on the part of the X chromosome involved in the rearrangement. Russell proposed – in retrospect – a very accurate model[32]:

- The negative correlation between variegation at a particular locus and the distance between that locus and the BP suggests a gradient of inactivation.
- Non-variegation of an autosomal gene translocated to the X chromosome occurs because of:
    - ✓ Attachment to a non-inactivating portion of the X chromosome, or
    - ✓ Excessive distance from an inactivated portion of the X chromosome.

Recent molecular work on similar X;A rearrangements confirms Russell's model: *XIST* RNA coating of X_i, the first event observed in the initiation of XCI, and H4 acetylation, another marker of XCI (see Appendix B), clearly occur more efficiently on X chromatin than on adjacent autosomal material[33].

A similar situation as that described in the mouse has been observed in human t(X;A) translocations. In contrast to the mouse work, however, human geneticists mainly used examples of unbalanced translocations. In such cases, the translocated X chromosome usually forms the X_i, which tends to restore genetic balance and is therefore positively selected[34]. In a balanced translocation, on the other hand, inactivation of the derivative X chromosome leads to a lethal genetic imbalance, which is why in most cases female carriers of such a translocation are characterised by skewed XCI. Such skewed XCI is believed to assure normal transcription of the translocated autosome[35].

An overview of recent investigations into the inactivation of autosomal material in X;A translocations in humans is given in Table I-2. Although not completely coherent, these data are in broad agreement: transcriptional silencing in autosomal material fused to part of the X chromosome occurs at long distances. The incompleteness and discontinuity of this process is reflected in the variability of various XCI hallmarks. The consistent lack of *XIST* RNA associating with the autosomal material suggests that transcriptional silencing of autosomal genes can occur in the absence of coating by *XIST* RNA. The absence of *XIST* RNA could indicate that autosomal sequences are resistant to the XCI signal or that autosomal DNA lacks chromosome X-specific elements. Gartler and Riggs have hypothesised such elements with their 'way station' model, in which they propose particular sequences along the X chromosome that boost the condensation signal[36]. Lyon postulated that L1 LINEs could, indirectly, act as 'way stations'[37]. The completed sequence of the human X chromosome revealed an enrichment and distribution of L1 elements that is consistent with the 'way station' and LINE hypotheses[18]. While the *XIST* locus lies in a region that is virtually devoid of L1 repeats, L1 levels are extremely high in the adjacent regions. In evolutionary strata where genes consistently escape XCI[24], L1 repeat coverage is particularly low (Figs. I-2d – f). Such distribution is not observed for any of the other interspersed repeats[18].

Table I-2 | **Inactivation status of autosomal sequences in selected human X;A translocations**

Case	XCI features of autosomal sequence	Ref.
46,X,der(X)t(X;4)(q22;q26)	* Late replicating * Six out of 20 tested genes escape XCI * Discontinuous escape from XCI * XCI spreading over >100 Mb	38
46,X,der(X),t(X;6)(p11;21.1) 46,X,ins(X;16)(q26;p12-13.1)	* Early replicating * No H4 hypoacetylation * Infrequent spreading of *XIST* RNA	39
47,Y,t(X;14)(Xpter>Xq13::14q32 >14qter;14pter>14q32::Xq13>Xqter)$_{mat}$ 46,X,-X,+der(9) t(X;9)(9pter>9q34::Xq13>Xqter)$_{mat}$	* Variable replication timing * Chr. 14: gradient in transcriptional activity * Chr. 9: consistent H4 hypoacetylation (some acetylated H4 far from the BP) * Deficient interaction with *XIST* RNA	40
46,X,der(X)t(X;11)(q26.3;p12)$_{pat}$ 46,X,der(X)t(X;7)(q27.3;q22.3)$_{mat}$ 46,X,der(X)t(X;6)(p11.2;p21.1)$_{mat}$ 46,X,der(X)t(X;6)(q28;p12)$_{pat}$ 46,X,der(X)t(X;10)(q26.3;q23.3)$_{mat}$	* Long-range (~ 45 Mb) transcriptional silencing, but incomplete or discontinuous * Three out of five cases show a sharp late-replication/histone hypoacetylation/H3-K4 dimethylation boundary at the BP * XCI gradient effects in three cases	41

A.3.2.3. Evidence for X-linked mental retardation from classical genetics

Male excess in MR, families with sex-linked MR, males showing a greater variability in intelligence, female-specific decrease in IQ in inbred populations, maternal IQ predicting a child's IQ better than paternal IQ, tendency for male MR to run in families and X-linked MR genes are all evidence for XLMR.

Since 1890, many studies have reported a male excess among the mentally handicapped[42]. One of the early proponents of the concept that MR was biologically determined was Lionel Penrose[43]. From his elaborate Colchester study, Penrose noticed that significantly more males than females were affected, indicating an involvement of a sex-linked factor. In the sample of 1280 individuals that Penrose studied, 710 were males and 570 were females, a male excess of 24.5%[44]. Soon afterwards, classical human genetics provided evidence that one such sex-linked factor could be genes on the X chromosome[45]. Later, Lehrke formulated his hypothesis that major genetic loci related to intellectual function were located on the X chromosome[46]. Support for Lehrke's hypothesis came from the following observations:

- Males show a greater variability in intelligence:
  - ✓ Male excess in institutions caring for the mentally disabled.
  - ✓ In his study on genius, Terman noticed 37% more 'clever' boys (n = 1405)[47].

  - ✓ WAIS minimises gender differences, but still the variance on ten of eleven scales is greater for males[48].
  - ✓ Similarly, the variance of test scores on 44 out of 56 academic tests was greater for boys than for girls, although the average scores did not significantly differ between both sexes[49].
- Studies in a Japanese inbred population showed a female-specific, significant decrease in IQ test scores as compared to a control population[50].
- Maternal IQ predicts a child's IQ better than paternal IQ or the educational level of either parent:
  - ✓ Reed and Reed found that there are more than five times as many children with severe MR born to retarded mothers than to retarded fathers[51].
  - ✓ High-IQ fathers have a higher percentage of high-IQ offspring than high-IQ mothers, but the high-IQ fathers are married to brighter women[51].
- There is a tendency for male MR to run in families. Studies of families with more than one retarded child indicated that there are almost twice as many MR boys as girls (n = 179)[52].

Formal proof for Lehrke's hypothesis, and for the notion of X-linked genes involved in MR in general, came with the cloning of *FMR1*, the gene mutated in Fragile X syndrome[53] and the first XLMR gene to be identified at the molecular level. Interestingly, the Martin – Bell family, reported in 1943[45] with a pedigree indicative of sex linkage of MR, turned out to carry a mutation at the *FMR1* locus[54]. In the Renpenning pedigree[55], another early reported family with apparent sex-linked MR[56], it has been shown that the MR segregates with a mutation in *PQBP1*[57], a previously identified XLMR gene[58]. These and other molecular studies, beautifully explaining observations of significant X-linked inheritance of MR from the field of classical genetics, firmly established the concept of XLMR.

Finally, epidemiological studies have reported a disproportional fraction of X-linked forms of MR[11].

A.3.2.4. Clinical variability of X-linked mental retardation

S-XLMR: *specific* phenotypic abnormalities associated with MR; NS-XLMR: no *consistent* somatic features associated with MR. This clinical division cannot be sustained at the molecular level.

MR linked to the X chromosome can be subdivided into two categories: syndromic forms which are characterised by *specific* phenotypic abnormalities associated with

MR, and non-syndromic forms, which have *no consistent* somatic features associated with the mental disability[59]. However, this clinically driven division cannot be sustained at the molecular level[60]; several S-XLMR genes were found to be mutated in patients with NS-XLMR, and *vice versa* (see Tables I-3 and I-6), leading to splitting and lumping in the nosology of XLMR[61]. Therefore, the distinction between S-XLMR and NS-XLMR is mainly of pragmatic use. The recently cloned *PQBP1* and *ARX* genes serve as two of the most convincing examples for the clinical variability of XLMR. Identical mutations in *PQBP1* result in widely different disorders, ranging from moderate NS-XLMR to very severe syndromic forms[58]. In the ARX protein, the same aberrant poly-alanine expansion was shown to cause different syndromic forms of XLMR, as well as NS-XLMR[62,63]. To date, eleven genes have been shown to be involved in both S- and NS-XLMR. Moreover, within the syndromic forms, clinical variability is frequently observed. One of the most compelling examples is that of mutations in *L1CAM* underlying extensive *intrafamilial* clinical variability[64].

Finally, it should be noted that allelism is not an (XL)MR-specific feature, but that other, apparently unrelated, disorders also turned out to be allelic. For example, mutations in the prion protein gene *PRNP* are associated with Gerstmann-Straussler disease (OMIM 137440)[65], Creutzfeldt-Jakob disease (OMIM 123400)[66] and familial fatal insomnia (OMIM 600072)[67].

A.3.2.5. Heterogeneity of X-linked mental retardation

~165 and ~100 gene defects are believed to be involved in S- and NS-XLMR, respectively.

MR is a characteristic finding in a multitude of human pathologies; searching OMIM for 'Mental Retardation' yields > 1300 entries. In September 2005, the online catalogue of XLMR conditions[a] comprised ~270 entries[68], a compilation of which is given in Table I-3.

Even despite the clinical variability described in the previous section, S-XLMR is more amenable to positional cloning than NS-XLMR, as linkage data from several families sharing similar clinical manifestations can be pooled in order to narrow down the candidate region (see I.B.1.3.1). This experimental bias contributes, at least in part, to the relative excess of characterised S-XLMR genes. It is estimated that ~165 genes may be involved in S-XLMR[69], one third of which have been identified (Table I-4). Among the S-XLMR genes, special note

[a] http://xlmr.interfree.it/home.htm

Table I-3 | **Overview of genes involved in XLMR**[§]

	Total # of entries[†]	Subset of entries with known chromosomal locations[‡]	Subset of entries with known gene defects	Total # of different genes involved in XLMR
S-XLMR[¿]	171	135	79	52
NS-XLMR	103	96	44	31
Σ	274	231	123	72[¶]

[§] Data based on the September 2005 update of genes involved in XLMR (http://xlmr.interfree.it/home.htm) and on the recent literature[70-74].
[†] Due to splitting and lumping in the nosology of XLMR[60,61,75], these numbers may change rapidly.
[‡] Numbers include entries with known gene defects.
[¿] Syndromic forms of XLMR include metabolic and neuromuscular disorders.
[¶] Corrected for the 11 genes implicated in both S- and NS-XLMR. See Table I-6.

should be made of mutations at the *FMR1* locus, causing Fragile X syndrome. Affecting 1:4000 males and 1:7000 females in the general population[76], mutations in this gene are the single most common cause of XLMR. With an incidence of 2 – 3% in MR males and 1% in MR females, who are generally less affected[77,78], routine diagnostic testing for Fragile X syndrome is conducted on all male patients with NS-MR.

Table I-4 | **Compilation of data on genes involved in S-XLMR**[§]

Mutated gene	Gene product Function	Pathology[†] (OMIM)[‡]
ABCD1	Peroxisomal ATP-binding cassette transporter Long-chain fatty acid metabolism	Adrenoleucodystrophy (300100)
ATP6AP2	Renin receptor Activation of ERK1 and 2	XMRE (300423)
ATP7A	ATPase Copper transport	Menkes disease (309400)
BCOR	Transcriptional co-repressor Modulation of histone acetylation and chromatin remodelling	Lenz microphthalmia (300412) & Oculofaciocardiodental spectrum (300166)
CDKL5	Cyclin-dependent Ser/Thr kinase Chromatin remodelling	X-linked infantile spasms (308350) & atypical Rett syndrome (312750)
DCX	Microtubule-associated protein Neuronal migration	X-SCLH/LIS (300067)
DKC1	RNA-associating protein rRNA processing and telomerase function	Dyskeratosis congenita (305000) & Hoyeraal-Hreidarsson syndrome (300240)
DMD	Actin-binding, brain-expressed Dp71 isoform Neurite outgrowth	Duchenne (310200) & Becker (300376) muscular dystrophy

Table I-4 | **Compilation of data on genes involved in S-XLMR**[§]

Mutated gene	**Gene product Function**	**Pathology**[†] **(OMIM)**[‡]
FLNA	Actin-binding protein Neurite outgrowth & dendritic spine formation	Periventricular heterotopia (300049), frontometaphyseal dysplasia (305620), Melnick-Needles syndrome (309350), otopalatodigital syndrome I (311300) & II (304120),
FMR1	mRNA-binding protein mRNA transport and regulation of translation	Fragile X syndrome (309550)
GK	Glycerol kinase Nuclear translocation of the glucocorticoid receptor complex	Hyperglycerolemia (307030)
GPC3	Cell-surface proteoglycan Glypican 3 Inhibition of cell proliferation and regulation of cell survival	Simpson-Golabi-Behmel syndrome (312870)
HADH2	MHBD Isoleucine catabolism	MHBD deficiency (300256)
HPRT1	Hypoxanthine guanine phosphoribosyltransferase Purine metabolism	Lesch-Nyhan syndrome (300322)
IDS	Iduronate 2-sulfatase Glycosaminoglycan metabolism	Hunter syndrome (309900)
IGBP1	Immunoglobulin binding protein Regulatory subunit of phosphatase PP2A	FGS1 (305450)
IKBKG	Regulatory subunit of IκB kinase Modulation of NF-κB	Incontinentia pigmenti (308300)
KIAA2022	1516 AA protein Unknown	Neonatal hypotonia, progressive quadriparesia, gastroeosophageal reflux, infantile autism, stereotypic hand movement (300524)
L1CAM	Neural cell glycoprotein Neurite outgrowth, axon fasciculation, remodelling of the cytoskeleton	HSAS (307000), spastic paraplegia I (312900) & MASA syndrome (303350)
LAMP2	Lysosomal membrane protein Autophagy, protection of the lysosomal membrane	Danon disease (300257)
MAOA	MAO Serotonin metabolism	MAOA deficiency, disturbed regulation of impulsive aggression (309850)
MID1	Microtubule-associated Ubiquitin ligase Protein degradation	Opitz syndrome (300000)
MTM1	Tyrosine phosphatase Muscle cell differentiation	Myotubular myopathy (310400)
NDP	Ligand of Frizzled Vascular development in the eye and ear	Norrie disease (310600)

Table I-4 | **Compilation of data on genes involved in S-XLMR**[§]

Mutated gene	**Gene product Function**	**Pathology**[†] **(OMIM)**[‡]
NHS	1630 AA protein Unknown	Nance-Horan syndrome (302350)
NLGN3	TM cell adhesion molecule Formation and remodelling of pre-synaptic synapses	Autism-spectrum disorder (300425 & 300494)
NXF5	Nuclear RNA export factor Possible role in (i) synaptic plasticity and (ii) polarised cytoplasmic transport and localisation of mRNAs in neurons	Short stature, pectus excavatum, general muscle wasting, facial dysmorphism (300319)
OCRL	PIP_2 5-phosphatase Translocation to membrane ruffles upon Rac GTPase activation, interaction with Clathrin and regulation of protein trafficking	Lowe syndrome (309000) & Dent disease (300009)
OFD1	116 kDa protein containing a large number of coiled-coil domains Formation of primary cilia and specification of the left – right axis	Oro-facial-digital syndrome I (311200)
OTC	Ornithine transcarbamylase Detoxification of ammonia	Hyperammonemia (311250)
PDHA1	Subunit of Pyruvate decarboxylase Part of the Pyruvate dehydrogenase complex, role in energy metabolism	Lactic acidosis (312170)
PGK1	Phosphoglycerate kinase Major enzyme in glycolysis, generation of ATP	Myoglobinuria, epilepsy, hemolytic anemia (311800)
PHF6	PHD-containing ZFP Likely transcriptional regulator	Börjeson-Forssman-Lehmann syndrome (301900)
PHF8	PHD- and jmjC-containing ZFP Possibly modification and regulation of DNA structure	Siderius-Hamel cleft lip/palate syndrome (300263)
PLP	Myelin proteolipid protein Major component of myelin	Pelizaeus-Merzbacher disease (312080) & spastic paraplegia II (312920)
PRPS1	Phosphoribosylpyrophosphate synthase Role in purine, pyrimidine, and pyridine biosynthesis	Hyperuricaemia (311850)
SLC16A2	Thyroid hormone transporter protein Iodothyronine transport across the cell membrane	Allan-Herndon-Dudley syndrome (300523)
SMS	Spermine synthase Poly-amine metabolism	Snyder-Robinson syndrome (309583)
SOX3	HMG-containing TF Formation of the hypothalamo-pituitary axis	Isolated growth hormone deficiency (300123)

Table I-4 | **Compilation of data on genes involved in S-XLMR**[§]

Mutated gene	Gene product Function	Pathology[†] (OMIM)[‡]
SYN1	SV-associated protein Modulation of neurotransmitter release and regulation of synaptogenesis	Epilepsy, learning difficulties, macrocephaly, aggressive behaviour (300491)
TIMM8A	MIM transport protein Import and insertion of membrane proteins into the MIM	Mohr-Tranebjaerg (304700) & Jensen (311150) syndromes

[§] Data based on the September 2005 update of genes involved in S-XLMR (http://xlmr.interfree.it/home.htm) and on the recent literature[70,71,74].
[†] Disorders include MR as a consistent or an inconsistent feature.
[‡] For appropriate references, please consult OMIM.

Initially, it was assumed that the genetic component of NS-XLMR would originate from defects in a handful of genes[79,80]. However, only five years later and based on elaborate linkage studies[81], it was estimated that > 100 X-chromosomal genes could contribute to NS-XLMR[82]. More recent estimates settle the number of NS-XLMR genes somewhat below a hundred[83], which seems an acceptable figure in light of the observation that mutations in NS-XLMR genes are typically found in ~1% of the patient cohort, with notable exceptions such as *ARX*, mutations in which are found in 8.1% of MRX families[84]. Although mutational hot spots on the X chromosome have been identified[83], it is the extensive heterogeneity underlying NS-XLMR – precluding straightforward pooling of linkage data from different families – that makes NS-XLMR gene hunting an arduous task. The identification of 16 genes implicated in NS-XLMR – that is, over 50% of the total known so far – during the last five years (April 2001 – April 2006, Fig. I-3) is indicative of the rapid progress that is being made towards the elucidation of genetic causes of NS-XLMR, but many of the genes causing NS-XLMR are still unknown[83]. Finally, it should be noted that, due to the specific methodology employed, NS-MR research is strongly biased towards the identification of X-linked genes. To date, only four genes[85-88] and two loci[89-91] have been implicated in autosomal NS-MR.

A.4. Molecular mechanisms underlying cognitive (dys)function

Due to the recent rapid advances in the search for XLMR gene defects, molecular mechanisms putatively causing mental handicap start to emerge. These mechanisms also allow first

insights into the development and function of brain structures involved in learning and memory, two aspects coupled to human intelligence.

A.4.1. A note of caution

Due to genetic MR research focussing on XLMR and because of the 'pure' MR in NS-MR, NS-XLMR genes are overrepresented in the following sections.

Well over 10000 transcripts are thought to be expressed in the human brain[a]. Most of them will likely be relevant to the brain's development, structure and/or function, and it is conceivable that mutations in many of the genes encoding them will lead to mental dysfunction.

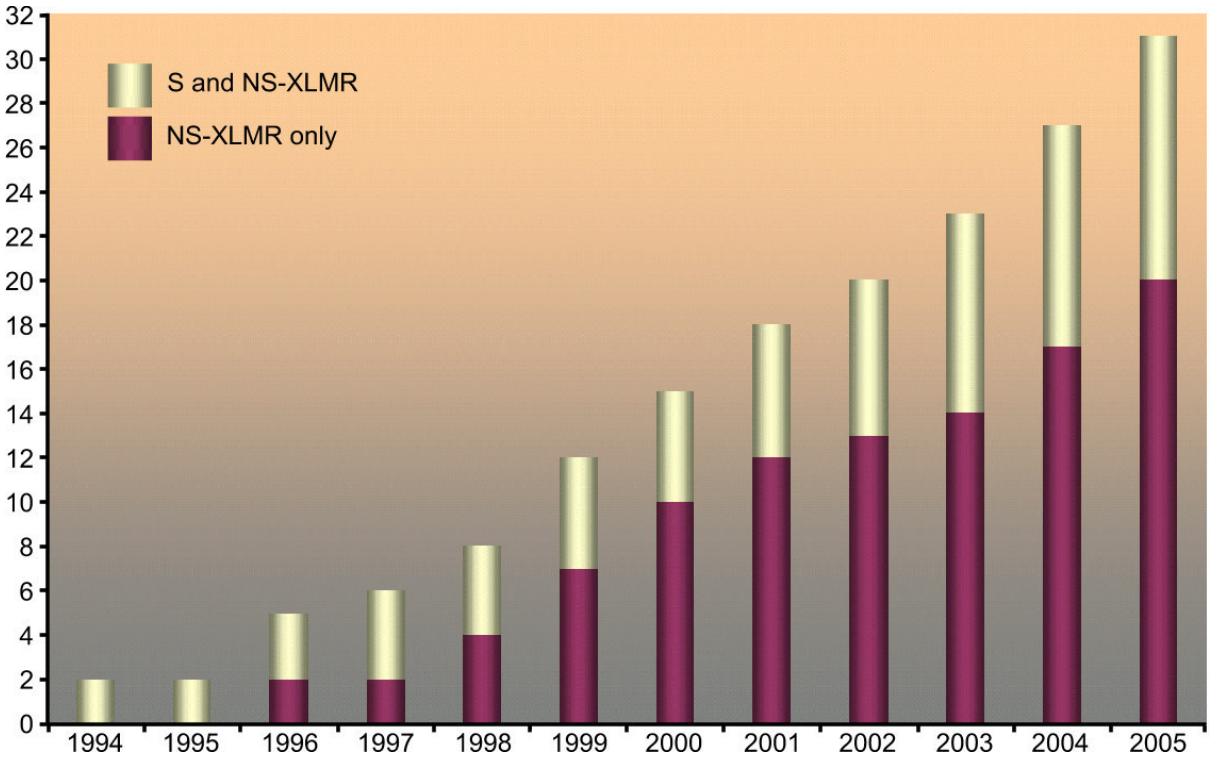


Fig. I-3 | **Identification rate of genes implicated in X-linked mental retardation.**

Plotting the cumulative frequency of gene discovery in the X-linked mental retardation (XLMR) field shows a steady increase in the identification of novel candidate disease genes. Genes are subdivided into those involved in both syndromic (S) and non-syndromic (NS) forms of XLMR (yellow) and those believed to underlie only non-syndromic phenotypes (aubergine). Entries were binned per year from April 1994 through April 2006.

Some of these genes have already been shown to result in MR with or without associated features when mutated. Unfortunately, syndromic forms of MR are less suited for the study of underlying mechanisms of brain function, as it is less than straightforward to determine

[a] http://www.biomart.org/index.html

whether the mental impairment is a direct consequence of the gene defect responsible for the syndrome or whether the MR is a secondary effect. Therefore, sections I.A.4.3 – I.A.4.6, describing some of the molecular mechanisms believed to play a role in mental deficit, will discriminate between insights gained from the study of S- or NS-MR. To this extent, genes – and their proteins – involved in syndromic (Table I-4), non-syndromic (Table I-5) or both (Table I-6) forms of MR are labelled with $^{'S'}$, $^{'NS'}$ or $^{'S+NS'}$, respectively (e.g. ^{S}*FMR1*, ^{NS}ZNF41, $^{S+NS}$*MECP2*). The bias towards identification of X-linked genes in MR combined with the above arguments on the usefulness of S- vs. NS-MR genes in the delineation of normal brain function led to the artificial overrepresentation of genes involved in NS-XLMR in the following sections.

Table I-5 | **Compilation of data on genes involved in NS-XLMR**

Mutated gene	**Gene product Function**	**Refs.**
ACSL4	Long-chain fatty acid-CoA ligase Membrane synthesis and/or recycling	92,93
AGTR2[§]	Angiotensin receptor Regulation of blood pressure and body temperature; role in behaviour (exploration, fear)	94,95
ARHGEF6	Cdc42/Rac1-specific GEF Stimulation of neurite outgrowth, integrin-mediated cell spreading	96-98
DLG3	Post-synaptic scaffolding protein Clustering of NMDA receptors in the PSD, NMDA receptor-mediated synaptic activity and plasticity	99,100
ELK1	MAPK docking site-containing TF Chromatin remodelling	101-104
FMR2	Transcriptional activator Possible role in long-term memory and enhanced LTP	105-109
FMR3	Likely a non-coding RNA Unknown	72,110
FTSJ1	RNA methyltransferase Possible role in post-transcriptional modification of tRNA, translation	111-113
GDI1	GDI of Rab GTPases RAB3A/3C-dependent regulation of SV fusion, short-term synaptic plasticity, formation and retention of temporal associations, social behaviour, astrocytic morphology	114-122
IL1RAPL1	Homologous to IL-1 receptor accessory protein NCS1-dependent regulation of Ca^{2+}-dependent exocytosis at nerve terminals	123-125

Table I-5 | **Compilation of data on genes involved in NS-XLMR**

Mutated gene	Gene product Function	Refs.
PAK3	Ser/Thr protein kinase Regulation of Actin cytoskeleton, dendritic spine and synapse formation; contribution to hippocampal late-phase LTP, learning and memory	126-130
RSK4	Ser/Thr protein kinase Specification of receptor tyrosine kinase signals, regulation of p53-induced growth arrest	131-133
TM4SF2	Four-fold TM tetraspanin Possible facilitation of synapse formation	134-140
VCX-3A[§]	186 AA protein Putative regulation of ribosome assembly during spermatogenesis	141-144
ZDHHC15	DHHC domain-containing ZFP Possible Palmitoyl transferase	145,146
ZNF261	MYM motif-containing ZFP TF with possible role in telomere capping	147-149
ZNF41	KRAB-containing ZFP KAP1-dependent transcriptional repression, role in chromatin remodelling	150-152
ZNF674	KRAB-containing ZFP TF	73
ZNF741	CACCC-box-binding ZFP CtBP-dependent transcriptional repression	153-155
ZNF81	KRAB-containing ZFP KAP1-dependent transcriptional repression, role in chromatin remodelling	150,151,156

[§] Involvement of *AGTR2*[157-159] and *VCX-3A*[160] mutations in the aetiology of XLMR is controversial.

Table I-6 | **Compilation of data on genes involved in both S- and NS-XLMR**

Mutated gene	Gene product Function	Refs.[§]	Pathology[†] (OMIM)
ARX	Homeodomain TF Transcriptional activator and repressor; proliferation, migration and differentiation of GABA-ergic neurons; role in olfactory system and forebrain development	62,78,161-173	XLAG (300215), myoclonic epilepsy (300432), Partington (309510), Proud (300004) & West (308350) syndromes
ATRX	PHD-containing ATPase/DNA helicase Regulation of transcription through ATP, HP1 and EZH2-dependent chromatin remodelling; alignment of chromosomes and organisation of the meiotic spindle; mediator of neuronal survival during corticogenesis	174-188	ATR-X (301040), Juberg-Marsidi, Chudley-Lowry, Smith-Fineman-Myers, Carpenter-Waziri & Holmes-Gang syndromes (309580)

Table I-6 | **Compilation of data on genes involved in both S- and NS-XLMR**

Mutated gene	**Gene product Function**	**Refs.**[§]	**Pathology**[†] **(OMIM)**
FGD1	SH3-BD-containing Rho GEF Promotion of Actin assembly at the cell cortex	189-193	Aarskog-Scott syndrome (305400)
JARID1C	JmjC & ARID DNA-BD-containing protein Transcriptional regulation and chromatin remodelling	194-198	Microcephaly, spasticity, epilepsy, short stature, facial anomalies (300534)
KIAA1202	PDZ- and ASD2-containing protein Unknown	199,200	Stocco dos Santos syndrome (300434)
MECP2	Methyl CpG-binding protein Silencing of neuronal gene expression through chromatin remodelling; neuronal survival and maturation, hippocampal synaptic plasticity; regulation of excitatory presynaptic function, involvement in learning and memory	201-225	Rett syndrome (312750), PPM-X (300055), progressive spasticity (300279), male fatal neonatal encephalopathy
NLGN4	Post-synaptic TM cell adhesion molecule in glutamatergic synapses Formation and remodelling of pre-synaptic synapses	226-230	Autism-spectrum disorder (300425 & 300494)
OPHN1	Rho-GAP Involvement in dendritic spine morphogenesis	231-234	Cerebellar ataxia 2 (302500)
PQBP1	Poly-glutamine-binding protein Regulation of transcription & neuronal cell death; involvement in cell cycle, neuronal proliferation and maturation; putative role in long-term memory and courtship	57,58,235-243	Renpenning & Sutherland-Haan (309500) syndrome, Golabi-Ito-Hall [244], Porteous & Hamel cerebro-palatocardiac [245] syndromes
RSK2	Ser/Thr protein kinase Chromatin remodelling, regulation of transcription; role in G-protein-coupled receptor signalling, the formation of long-term memory, osteoblast differentiation and in regulation of the Actin cytoskeleton, excitatory synaptic transmission and the cell-cycle	246-255	CLS (303600)
SLC6A8	Creatine transporter Creatine metabolism	256-258	Creatine deficiency syndrome (300352)

[§] Apart for gene function, references are given for the involvement in NS-XLMR. For appropriate S-XLMR references, please consult OMIM.
[†] Disorders include MR as a consist or an inconsistent feature.

A.4.2. The 'structural' and the 'functioning' brain

Cognitive deficits can arise from defects upsetting structural features of the CNS and/or affecting its function.

Basically, the brain, seat of human intelligence, can be considered from two different, but not necessarily mutually exclusive, angles.

On the one hand, there is the 'structural brain' with its specific anatomy, its plethora of neuronal cells and its innumerable synapses. An example of a structural CNS feature affected in human disease is the myelin sheathing of neurons, which electrically shields them from their environment, allowing for faster and less dissipative nerve impulses. Mutations in ^{S}*PLP*, the gene encoding PLP and DM-20, two major myelin proteins[259], give rise to Pelizaeus-Merzbacher disease (OMIM 312080)[260] and spastic paraplegia II (OMIM 312920)[261], both dysmyelinating CNS-specific disorders.

On the other hand, there is the relentlessly 'functioning brain', with its firing synaptic connections, its neurotransmitter household and its incessantly adapting neurons. In sections I.A.4.3 – I.A.4.6, four mechanisms in which XLMR genes have been shown to play important roles and which are relevant to the 'functioning brain' will be discussed. It should be noted, however, that the distinction between 'structural' and 'functioning' brain is becoming increasingly fuzzy, especially with more recent studies, and that many pathways are situated on the interface.

Clearly, mutations upsetting either one of these 'brains', or both, will affect intelligence and could lead to mental handicap. Moreover, since the brain is so complex and requires such extraordinary precision, it is conceivable that mutations affecting more basic cellular or metabolic processes, which may go unnoticed in less 'demanding' tissues, could have a profound effect on the brain, again resulting in MR. A few examples of XLMR genes affecting basic cellular functions are the ^{NS}FTSJ1 protein[113], involved in translation; ^{S}DKC1, implicated in rRNA processing[262] and telomerase function[263]; and ^{S}LAMP2, thought to protect the lysosomal membrane and to aid in importing proteins into the lysosome[264].

Finally, it is worth mentioning that biochemical pathways and global molecular mechanisms associated with human cognition slowly start to emerge, but that none of them have been established in enough detail to unravel mental deficit nor, for that matter, human intelligence.

A.4.3. Remodelling of the Actin cytoskeleton and dendrite outgrowth in mental retardation

A first molecular mechanism possibly underlying mental function in humans is the remodelling of the Actin cytoskeleton. To date, several XLMR genes have been shown to regulate GTPases, a family of signal transducers exerting an effect on cytoskeletal rearrangements.

A.4.3.1. General overview

Connectivity in the brain relies on neurite guidance. Aberrant dendritic morphology, guidance and connectivity have been associated with MR. The cellular basis underlying these processes is the rapid remodelling of the Actin cytoskeleton in response to a variety of extracellular cues. These signals are relayed to the Actin remodelling machinery via a pathway involving Rho GTPases.

In order to relay electrical signals from other neurons, each neuron itself is covered with dendrites radiating from its cell body. Like axons, conducting signals away from the cell body, dendrites extend by means of a growth cone at their tip, a broad expansion with many finger-like extrusions, the filopodia[265]. These motile filopodia can form and retract with great speed due to (de)polymerisation of Actin, one of the three major types of cytoskeletal proteins[266], at their plasma membrane[267]. These mechanisms of extension and retraction lie at the basis of neurite (that is, dendritic and axonal) guidance, a process that is of paramount importance in establishing connectivity in the brain[268] and that has been shown to be responding to a number of extracellular cues, such as physical character of the substratum (stereotropism)[269], gradients of adhesivity (haptotaxis)[270], differential adhesive specificities[271] or a concentration gradient of soluble molecules (chemotaxis)[272]. Taken together, this means that connectivity in the brain is dependent on rapid remodelling of the Actin cytoskeleton in axons and dendritic spines steered by environmental cues (Figs. I-4a – c').

Over thirty years ago, abnormalities in dendrites and dendritic spines were associated with MR. One study showed a reduction in the number of dendritic spines and a predominance of thin, long spines in the brains of mentally retarded children, contrasting with the abundance of stubby and mushroom-like spines found in the brains of normal children (Fig. I-4c'')[273]. Another study associated reduced dendritic branching with MR[274]. Over the years, further evidence consolidated the link between dendritic spine morphology and intellectual capacity:

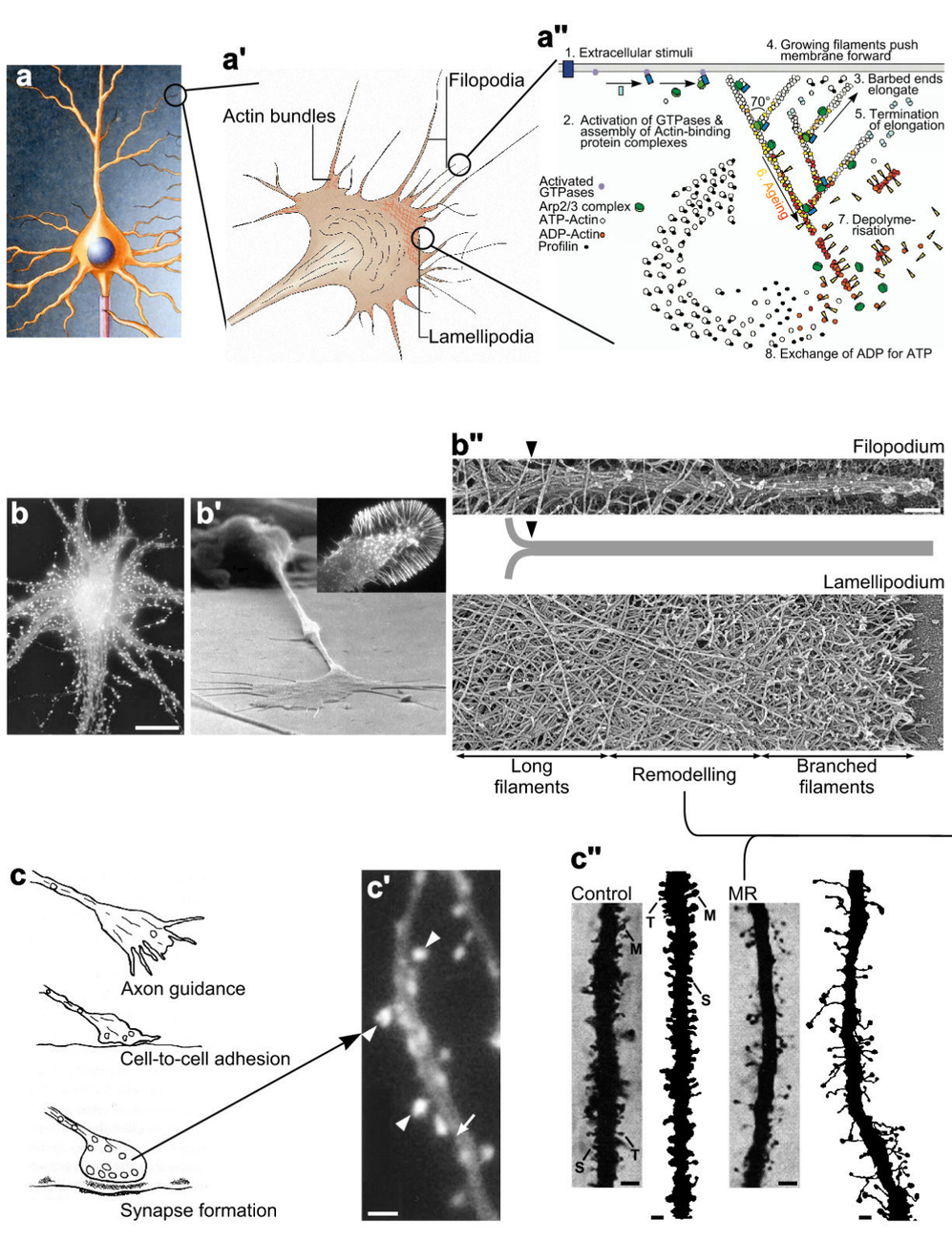
- Dendritic anomalies were found in adults with untreated phenylketonuria[275] and in infants with malnutrition[276], two conditions known to lead to MR.

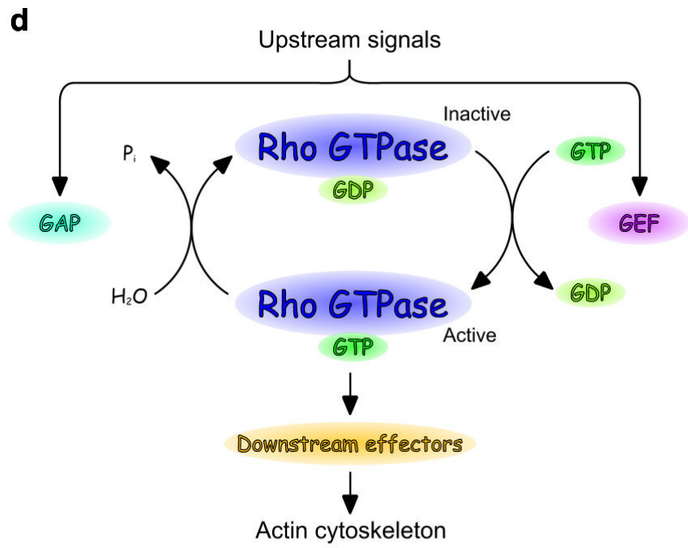
- Similar anomalies were observed in the pyramidal neurons of the hippocampus and cerebral cortex of patients with Down's, Rett's or Fragile X syndrome[277].
- ^{S}*FMR1* KO mice, mimicking the human Fragile X syndrome, showed learning deficits[278] and abnormal dendritic spines[279-281].
- Other animal studies recapitulating MR by prenatal alcohol exposure[282], protein deprivation[283], foetal hypoxemia[284] or hypothyroidism[285] all observed reduced dendritic arborisation and/or spine deficits.

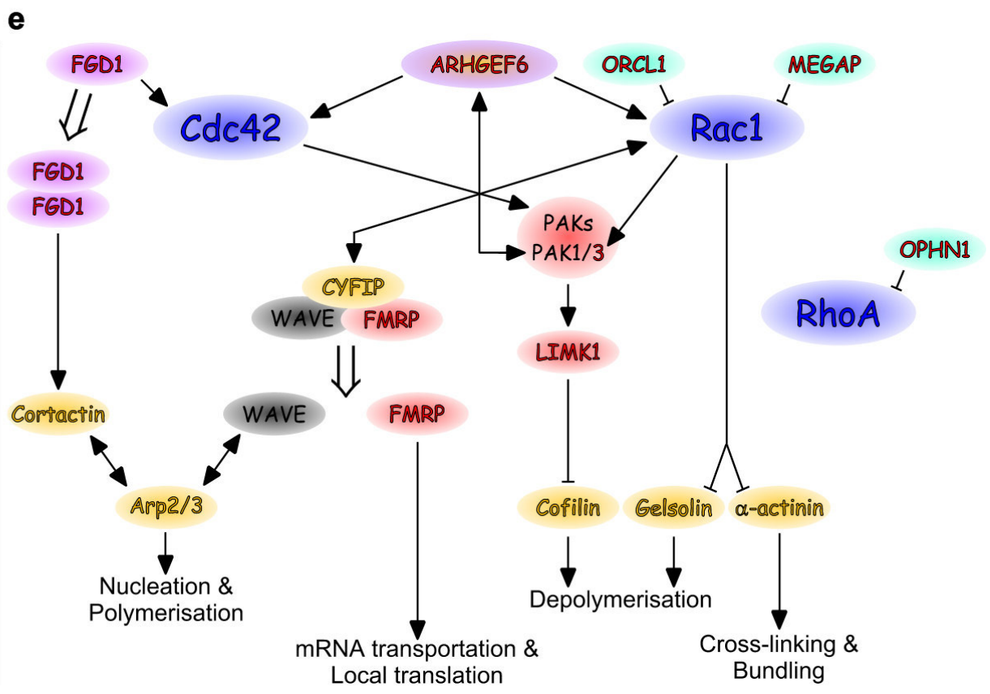
Fig. I-4 | *Next page.* **Mental retardation and the Actin cytoskeleton.**

a, a', a''. Schematic representation of the role of Actin remodelling in neurite outgrowth. **a.** Pathfinding of neuronal processes depends on the terminal growth cone. **a'.** Specific structures, such as filopodia and lamellipodia, characterise the growth cone. **a''.** Filopodia and lamellipodia are Actin-rich structures that sense guidance cues from the environment, which translate into cytoskeletal rearrangements. Arp2/3, Actin-related proteins 2 and 3.
b, b', b''. Microscopic studies showing some aspects of the Actin remodelling pathway in neurite outgrowth as schematised in panel series a. **b.** The neuron in this image expresses Actin tagged with green fluorescent protein. Dendritic spine heads appear as bright spots on the dendrites (see also panel c'). Scale bar, 10 μm. **b'.** This scanning electron microscopic (EM) recording of a growth cone clearly shows filopodia and lamellipodia (compare with a'). The inset is an epifluorescence micrograph of a growth cone from a neuron injected with labelled monomeric Actin that incorporated into Actin filaments. **b''.** The top panel shows a platinum replica EM recording of a filopodium. It contains a tight bundle of F-actin that splays apart at its root (arrowhead) and becomes an integral part of the surrounding Actin network (only partially shown). Scale bar, 0.2 μm. The line diagram (middle) represents the filopodium depicted in the top panel. The bottom panel is a transmission EM micrograph from a lamellipodium at the leading edge of a motile keratocyte. Three zones of Actin organisation are labelled. Molecular pathways involved in the remodelling zone are shown in panel e.
c, c', c''. Outline of the link between axon guidance, dendritic spines and mental retardation (MR). **c.** Synaptogenesis involves three phases: through axon guidance (top; specified in panel series a and b), the growth cone contacts its target cell (middle), where it has the capacity to form a synaptic bouton (bottom). **c'.** Dendritic spines (arrowheads) are highly enriched in Actin compared to the dendritic shaft (arrow). Dendritic spine heads are locations at which synaptic boutons with different morphologies form. Scale bar, 2 μm. **c''.** Photomicrograph and *camera lucida* representation of Golgi-stained motor cortex neurons show a dendritic segment of a layer V pyramidal neuron from a neurologically normal six month-old infant (left panel, Control). Three types of dendritic spines can be recognised: thin (T), stubby (S), and mushroom-shaped (M). The right panel shows a similar dendritic segment from a ten month-old retarded child (MR). Note the overall reduction in spine density and the predominance of abnormally long, thin spines which sometimes appear to be entangled. Panel e shows an overview of pathways that link MR to Actin remodelling. Such remodelling may explain (part of) the morphological aberrations seen in the dendritic spines of mentally retarded patients. Scale bars, 2 μm.
d. Internalised extracellular signals regulate the activities of guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs), which switch the Rho GTPases between their active and inactive states. In their active state, they bind and effect a battery of downstream regulators of Actin cytoskeletal organisation.
e. Red represents MR proteins that have been implicated in the regulation of Actin remodelling, yellow Actin-binding proteins. *ARHGEF6*, a GEF, encodes an isoform implicated in MR and one binding Actin. Colour coding for Rho GTPases, GEFs and GAPs is as in panel d. → activation, ↔ activating direct interaction and —| inhibition. Please note that several MR proteins that interact with Actin are not included in the diagram. See text for details.

Images adapted from http://www.mb.jhu.edu/tins.asp (a); *286* (a'); *287* (a''); *288* (b); *289* (b'', top); *287* (b'', bottom); *290* (c); *288* (c'); *273* (c'') or courtesy of Drs. Balazovich, University of Michigan, USA (b') and Cohan, University at Buffalo, USA (b', inset).


a
a'
a"
Actin bundles
Filopodia
Lamellipodia
1. Extracellular stimuli
2. Activation of GTPases & assembly of Actin-binding protein complexes
3. Barbed ends elongate
4. Growing filaments push membrane forward
5. Termination of elongation
6. Aging
7. Depolymerisation
8. Exchange of ADP for ATP
70°
Activated GTPases
Arp2/3 complex
ATP-Actin
ADP-Actin
Profilin
b
b'
b"
Filopodium
Lamellipodium
Long filaments
Remodelling
Branched filaments
c
Axon guidance
Cell-to-cell adhesion
Synapse formation
c'
c"
Control
MR
T
M
S


d
Upstream signals
Inactive
Rho GTPase
GDP
GTP
GAP
GEF
P_i
H_2O
Rho GTPase
GTP
GDP
Active
Downstream effectors
Actin cytoskeleton


e
FGD1
Cdc42
ARHGEF6
ORCL1
MEGAP
Rac1
FGD1
FGD1
PAKs
PAK1/3
OPHN1
RhoA
CYFIP
WAVE
FMRP
LIMK1
Cortactin
WAVE
FMRP
Arp2/3
Cofilin
Gelsolin
α-actinin
Nucleation & Polymerisation
mRNA transportation & Local translation
Depolymerisation
Cross-linking & Bundling

The environmental cues directing dendritic outgrowth are transmitted to the Actin cytoskeleton via a signal transduction cascade involving G-proteins, belonging to the GTPase superfamily. G-proteins themselves are subdivided into large, heterotrimeric, membrane-associated proteins and small, monomeric, cytosolic proteins. They act as molecular switches, cycling between an active, GTP-bound and an inactive, GDP-bound state. While the GEFs mediate the transition from the inactive to the active state by facilitating the exchange of GDP for GTP, the GAPs stimulate the opposite reaction by catalysing the hydrolysis of GTP. A cytosolic pool of GDP-bound, and thus inactive, GTPases is maintained via binding to the GDIs. Activated G-proteins interact with their downstream effectors, a battery of molecules including kinases and scaffold proteins involved in Actin dynamics and other processes (Fig. I-4d). A large body of literature shows that Rho GTPases (e.g. RhoA, Rac and Cdc42), a subfamily of the small G-proteins, regulate neuronal morphogenesis and connectivity[286].

A.4.3.2. Mental retardation gene products regulate remodelling of the Actin cytoskeleton

Eight genes linked to Actin remodelling, seven of them encoding regulators or effectors of the Rho GTPases and six genes encoding proteins affecting neurite outgrowth or cell motility, have been implicated in MR.

Seven recently identified MR genes, $^{S+NS}$*FGD1*, $^{S+NS}$*OPHN1*, ^{NS}*PAK3*, ^{NS}*ARHGEF6*, ^{S}*LIMK1*, ^{S}*MEGAP* and ^{S}*OCRL*, encode regulators or effectors of the Rho GTPases, suggesting an important role for Rho signalling and remodelling of the cytoskeleton in cognitive function (Fig. I-4e).

The $^{S+NS}$*FGD1* gene, first found to be mutated in patients with Aarskog-Scott syndrome (OMIM 305400)[189,191,192], encodes a protein with Cdc42-specific GEF activity[291] that contains two SH3-binding domains. Later, Lebel *et al.* described a family with NS-MR, in which a P312L mutation may upset the positioning of one of these SH3 BDs in affected males[190]. Dimerisation of FGD1's SH3 BD has been shown to increase the activity of Cortactin[193], an F-actin binding protein associated with sites of dynamic Actin assembly[292] that promotes F-actin network formation[293]. Moreover, microinjection of FGD1 into fibroblasts activated Cdc42 and induced Actin polymerisation[294].

Oligophrenin 1, encoded by the $^{S+NS}$*OPHN1* gene, acts as a Rho-GAP[231,234] and was shown to be required for dendritic spine morphogenesis, possibly through a direct interaction with Homer[232], a key adaptor protein that organises glutamate receptor signalling complexes

at the post-synaptic membrane[295], and that has been shown to be involved in dendritic spine morphogenesis and synaptic transmission[296].

^{NS}PAK3[126], the Ser/Thr protein kinase encoded by the *p21-activated kinase 3* gene[127], mediates, like other PAKs, effects on the Actin cytoskeleton downstream of the Rho GTPases Rac1 and Cdc42[297]. Using anti-sense and siRNA technologies, Boda *et al.* showed an involvement of ^{NS}PAK3 in the formation of dendritic spines and synapses, and in hippocampal synapse plasticity and LTP[128]. Apart from confirming these observations, an *mPak3* KO also showed learning and memory deficits[129].

αPIX, encoded by the ^{NS}*ARHGEF6* gene[298] implicated in NS-XLMR[96], is a protein with homology to Rho-GEFs[299]. It exhibits Cdc42/Rac1-specific GEF activity[300], interacts with PAK1[301], and a direct interaction between βPIX, an alternative ^{NS}*ARHGEF6* isoform, and F-actin has also been reported[302]. These observations link αPIX to cytoskeletal remodelling. Cell adhesion to the extracellular matrix is mediated through integrins, a family of TM receptors, and results in reorganisation of the Actin cytoskeleton through regulation of Rho GTPase activity[303]. Binding of αPIX to Affixin, an integrin-linked kinase-binding protein[304], and co-localisation of both proteins at the tips of lamellipodia in motile cells suggests that αPIX links integrin-dependent cell spreading to Cdc42/Rac1 activation[97]. Rosenberger *et al.* dissected the role of αPIX in cell spreading: αPIX' GEF activity contributes to enhanced formation of cellular protrusions and its GEF-independent Calpain 4 association results in cell spreading[98].

The protein kinase LIM kinase 1 (^{S}LIMK1), which accumulates at mature synapses[305], is a direct target of PAKs[306] and through inactivating the Actin depolymerisation factor Cofilin[307] it is a potent regulator of Actin dynamics, both *in vitro*[308,309] and *in vivo*[310,311]. Characterising the Williams-Beuren contiguous gene syndrome (OMIM 194050) critical region on 7q11.23 allowed Frangiskakis *et al.* to identify ^{S}*LIMK1* hemizygosity as a cause for impaired visuospatial constructive cognition, one of the manifestations of WBS[312]. However, this finding has been questioned[313,314]. Through regulation of Golgi dynamics and Actin reorganisation, ^{S}LIMK1 has been suggested to play a role in axon formation[315]. In line with this suggestion, several recent studies have indicated an involvement of ^{S}LIMK1 in relaying extracellular cues to the Actin remodelling machinery. First, ^{S}LIMK1 is critical for chemokine-induced polarised lamellipodium formation[316]; second, knock down of ^{S}LIMK1 prevents Thrombin-induced F-actin formation[317]; third, ^{S}LIMK1 co-localises and interacts with the BMP receptor BMPRII in the tips of neurites, an interaction that is required for

BMP-dependent induction of the dendritic arbor in cortical neurons[318] and fourth, translation of rat *Limk1* RNA is regulated by miR-134, a brain-specific microRNA localised to the synapto-dendritic compartment of hippocampal neurons. MicroRNAs are small, non-coding RNAs that control the translation of target messenger RNAs. Exposure of neurons to extracellular stimuli relieves miR-134 inhibition of Limk1 translation[319]. Taken together, these studies provide insights into how extracellular signals modulate dendritogenesis in a LIMK1-dependent way.

The mental disorder-associated GAP protein ^{S}MEGAP, with gene locus on 3p25, has been reported to be the cause of severe MR and hypotonia, two of the manifestations of the 3p^{-} contiguous gene syndrome[320]. An alternative name for ^{S}MEGAP is Slit – Robo GAP3, indicating its role in the Slit – Robo signal transduction pathway[321], which is involved in axonal guidance[322,323].

Mutations in ^{S}*OCRL* give rise to the oculocerebrorenal syndrome of Lowe (OMIM 309000)[324,325] and Dent disease (OMIM 300009)[74]. The ^{S}OCRL protein was originally identified to be a PIP_2 5-phosphatase located in the trans-Golgi network[326,327]. Suchy and Nussbaum noticed that Lowe syndrome fibroblasts are characterised by a decrease in long Actin stress fibers, an enhanced sensitivity to Actin depolymerising agents, an increase in punctate F-actin, and an abnormal distribution of Gelsolin and α-actinin[328]. Gelsolin and α-actinin, both Actin-binding proteins, are involved in cell motility[329] and spine morphology[330], respectively. The characteristics of the Lowe syndrome fibroblasts were attributed to a deficiency in PIP_2 5-phosphatase activity[328]. This finding was further delineated by the observation that ^{S}OCRL contains an active RhoGAP C-terminal that specifically interacts with Rac GTPase[331] and that, upon induction of Rac, a fraction of ^{S}OCRL translocates to the plasma membrane, where it concentrates in membrane ruffles. PIP_2 accumulation in Lowe fibroblast ruffles, suggests that ^{S}OCRL is active in Rac-induced membrane ruffles[332].

An eighth gene that has been connected to the Rho-GTPase pathway is ^{S}*FMR1*, encoding ^{S}FMRP and mutated in Fragile X syndrome[53]. ^{S}FMRP is an RNA-binding protein which forms a messenger – ribonucleoprotein complex that associates with translating polyribosomes[333]. It has been proposed that ^{S}FMRP is involved in synaptic plasticity through the regulation of mRNA transportation and translation[334]. However, in recent years, ^{S}FMRP has also been linked to Rho-GTPase signalling. CYFIP1, which interacts with Rac1 and F-actin[335], and CYFIP2, which is part of a complex including WAVE1, a downstream effector of Rac1[336], were both found to interact with ^{S}FMRP[337]. Studies in *Drosophila* showed that

[S]FMRP was positively regulated by Rac1[338]. Moreover, expression profiling identified a Rho GEF that was differentially expressed in *[S]FMR1* KO mice[339], which showed aberrant dendritic spines[281], suggesting a role for FMRP in Rac1-induced spine morphogenesis.

Apart from the eight proteins which are linked to Actin remodelling, seven of them through their function as regulators or effectors of the Rho GTPases, another six genes encoding proteins affecting neurite outgrowth or cell motility were shown to cause MR when mutated.

Mutations in *[S]FLNA*, the gene encoding [S]Filamin A[340], have been reported to underlie several forms of S-XLMR[341]. In periventricular heterotopia, mutations in *[S]FLNA* abolish the migration of cerebral cortical neurons[341]. [S]Filamin A is an Actin-binding[342] and very potent Actin filament-crosslinking protein that causes the gelation of Actin[343]. Gelation results from the crosslinking of Actin filaments into orthogonal networks. [S]Filamin A transduces ligand-receptor interactions into Actin reorganisation[344], including the induction of perpendicular branching of Actin filaments[345].

[S]*L1CAM*, the gene encoding the neural Cell adhesion molecule L1[346], was found to be mutated in hydrocephalus (OMIM 307000)[347-349], spastic paraplegia type I (OMIM 312900) and MASA syndrome (OMIM 303350)[64,350]. [S]L1CAM is a cell surface glycoprotein primarily expressed in the nervous system that has been implicated in a host of different neuronal mechanisms[351], including synapse formation[352] and neurite outgrowth[353]. Careful investigation of thalamocortical and corticothalamic axons in the brains of *L1cam* KO mice, revealed pathfinding errors, growth cone abnormalities and hyperfasciculation of these axons[354]. Homophilic [S]L1CAM adhesion mobilises cytoskeletal proteins to sites of cell contact, thereby remodelling the cytoskeleton[355]. This mechanism involves the acquisition of ankyrins, multi-domain linker proteins that bind cytoskeletal components[356-358], via the cytoplasmic domain of [S]L1CAM[359].

[S]DCX, a component of the microtubule cytoskeleton[360], is, together with DCX-like kinase[361,362], required for neuronal migration, since mutations in the human gene lead to the X-SCLH/LIS syndrome (OMIM 300067)[363,364], which is characterised by a disruption of the six-layered neocortex due to defective cortical neuronal migration[365]. Video microscopy analyses of migrating interneurons in the developing brain of *Dcx* KO mice highlighted a role for [S]DCX in branching and nucleokinesis[366]. Delineation of the latter revealed that [S]DCX mediates, along with Dynein, coupling of the nucleus to the centrosome during neuronal migration[367]. [S]DCX stabilises microtubules and preferentially assembles 13 protofilament microtubules[368]. The assembly of microtubules is one of the mechanisms underlying neuronal

migration[369]. Recent studies showed that SDCX is also associated with F-actin, raising the possibility that SDCX acts as a molecular link between microtubule- and Actin-cytoskeletal filaments[370].

S*OFD1*[371] has been found to be mutated in patients with orofaciodigital syndrome, type I (OMIM 311200)[372,373], an X-linked dominant disorder with lethality in males[374]. Recently, a family was identified in which frameshift mutations in S*OFD1* segregated with severe MR, recurrent respiratory tract infections and macrocephaly in males. Carrier females were clinically inconspicuous[375]. Study of *Ofd1* KO mice revealed that SOFD1 is required for ciliogenesis, left-right axis specification, patterning of the neural tube and correct expression of homeobox genes in the limb buds[376]. To date, SOFD1's cellular function remains elusive, but the novel sequence motif LisH, suggested to be implicated in protein dimerisation, protein half-life, subcellular localisation, microtubule dynamics and cell migration, has been identified in SOFD1[377,378].

S*DMD*, the gene mutated in Duchenne (OMIM 310200) and Becker (OMIM 300376) muscular dystrophies[379], has been shown to be expressed from tissue-specific promoters[380], leading to different isoforms of the protein[381-384]. The brain-expressed Dp71 isoform[385] has been shown to be required for neurite outgrowth *in vitro*[386], possibly mediated through a direct Dp71-specific binding to Actin[387].

A last protein, microtuble-associated SMID1, which causes Opitz syndrome (OMIM 300000) when mutated[388], is described in more detail under IV.E.2.

A.4.4. Chromatin remodelling and regulation of transcription in mental retardation

Regulation of transcription is achieved in manifold ways, including regulation of the transcriptional machinery or the condensation state of the chromatin (Fig. I-5a). Although regulation of transcription is omnipresent and may therefore play a role in a multitude of human pathologies by default, the remarkably large proportion of XLMR genes involved in transcription and chromatin remodelling indicates that these molecular mechanisms may indeed be important in human cognitive ability (Fig. I-5e).

A.4.4.1. General overview

Nuclear DNA is compacted by wrapping it around histone octamers. Such densely packed heterochromatin is mostly transcriptionally inactive. The condensation state of chromatin, and hence its transcriptional activity, is effected through changes in histone modifications.

Histones, first identified in the 1880s[389], are now subdivided into nucleosomal and H1 histones. The four nucleosomal histones[390], designated H2A, H2B, H3 and H4, are among the most highly conserved[391] of all proteins and 20 (H2A) to 25% (H4) of their composition are made up of the positively charged amino acids lysine and argenine, allowing for ionic interaction with the negatively charged phosphate groups of the DNA backbone. Two copies of each of these histones form a histone octamer[392], around which the DNA double helix is wound one and three-quarter times[393], resulting in the nucleosome (Figs. I-5b – b'')[394]. Beginning with the identification of the nucleosome as a basis for chromatin structure[395], the extraordinary packing of nuclear DNA has been studied at the molecular level since the 1970s. The nucleosomes are linked by 10 – 60 bp of DNA[396] and are packed together by H1 into a higher-order structure[397]. Apart from the four canonical nucleosomal histones, several different histone variants have been characterised[398].

Apart from compacting, it has long been realised that the condensation of nuclear DNA into chromatin is a way of regulating gene transcription; less compacted euchromatin is transcriptionally active, whereas densely packed heterochromatin is mostly transcriptionally inactive[399]. Still, it is only recently that the mechanisms governing transcriptional activity have been understood in some detail. The histone tails that protrude from the nucleosome are subject to reversible post-translational modifications (Fig. I-5c), such as acetylation and methylation of lysine and arginine, and phosphorylation of serine and threonine[400]. Moreover, di- and tri-methylation of lysine and addition of two methyl groups to arginine have been reported[401]. Histones are modified by a large number of enzymes, which have been grouped into histone acetyltransferase[402], HDAC[403], histone methyltransferase[404] and histone kinase[405] families. These enzymes are specific to a particular residue within a certain histone tail[406]. Also, ubiquitinylation[407], SUMOylation[408], glycosylation and ADP-ribosylation[409] of histones have been observed.

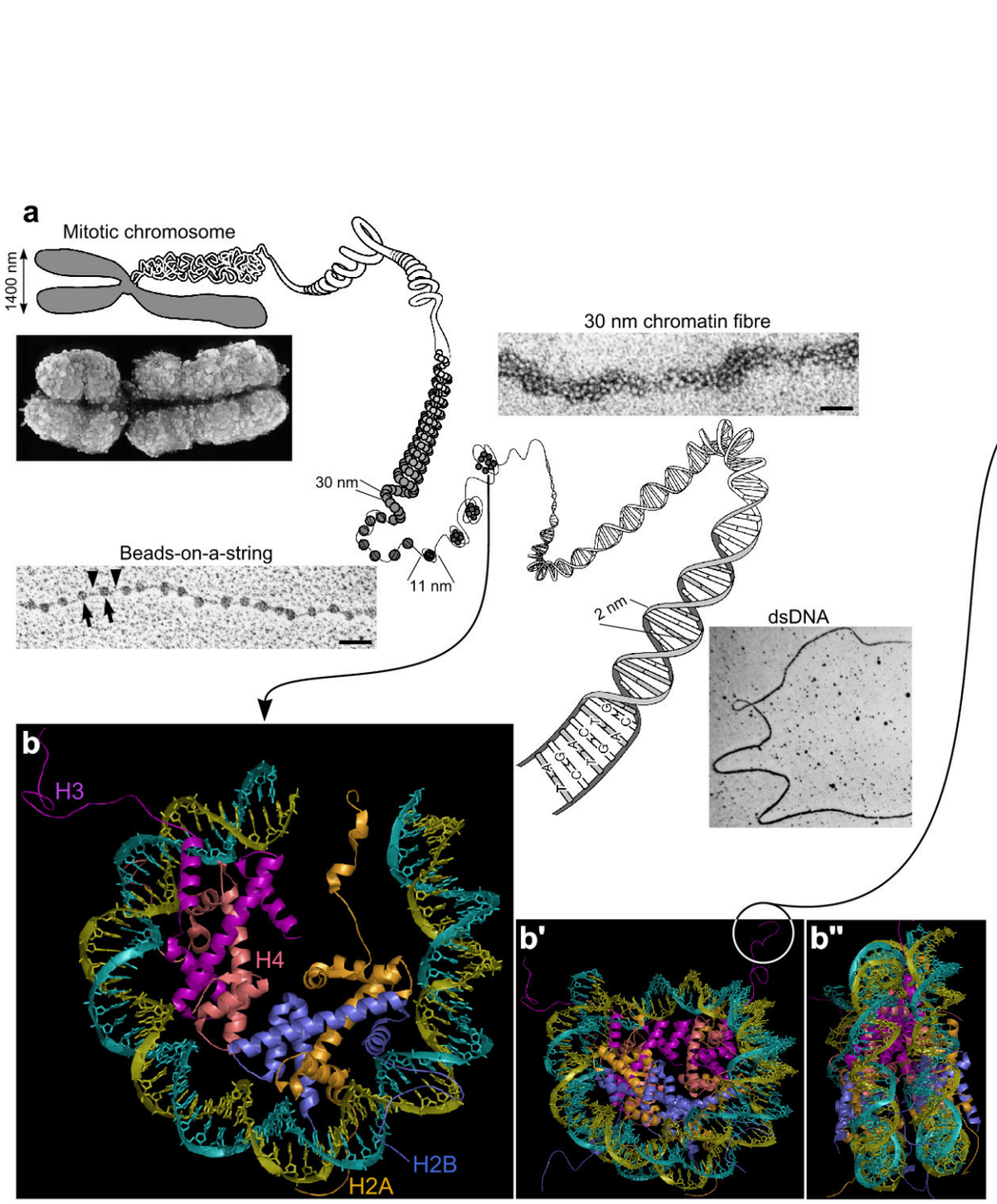
In the case of transcriptional control, the histone modifying enzymes interact with DNA sequence-specific regulators of transcription in order to target particular histones for modification[410]. Enzyme activity is further influenced by modifications that are already present; activating[411] and inhibitory[412] cross-talk among different histone marks results in the final set of modifications at each locus (Fig. I-5c). Such site-specific combinations of modifications led

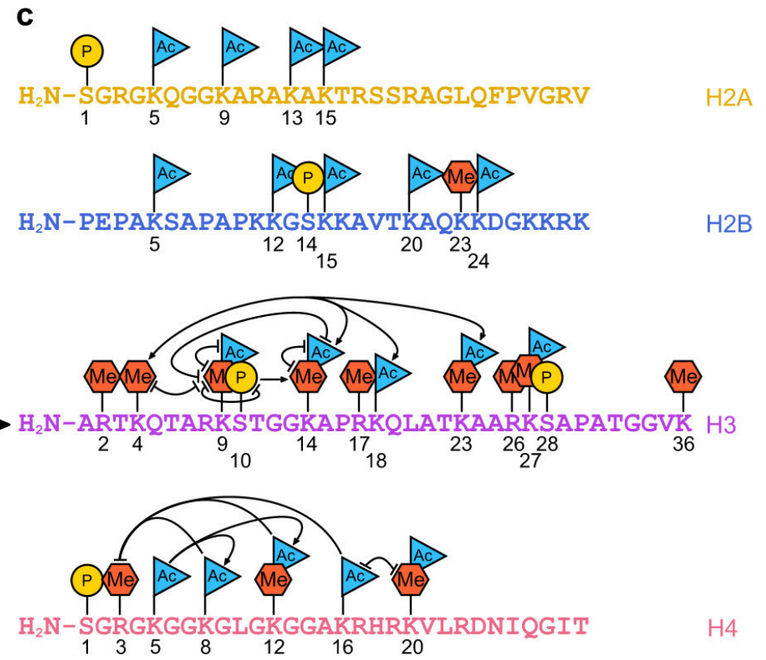
to the postulation of a 'histone code' extending the information potential of the genome[413]. At first, it was thought that histone modifications that resulted in a decreased positive charge altered the histone – DNA interaction and thus affected chromatin density[414]. However, it is now believed that histone modifications recruit non-histone proteins to the DNA. For example, chromodomains specifically bind methylated lysines[415], whereas bromodomains interact with acetylated lysine residues[416]. Histone tails are therefore considered to function as scaffolds that assemble all factors necessary for a particular biological function, such as the activation (generally mediated by hyperacetylation[417]) or silencing (generally characterised by hypoacetylation[418] and hypermethylation[419]) of transcription (Fig. I-5d).

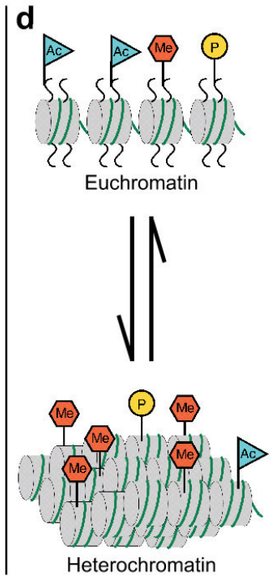
Fig. I-5 | *Next page.* **Mental retardation, chromatin condensation and regulation of transcription.**

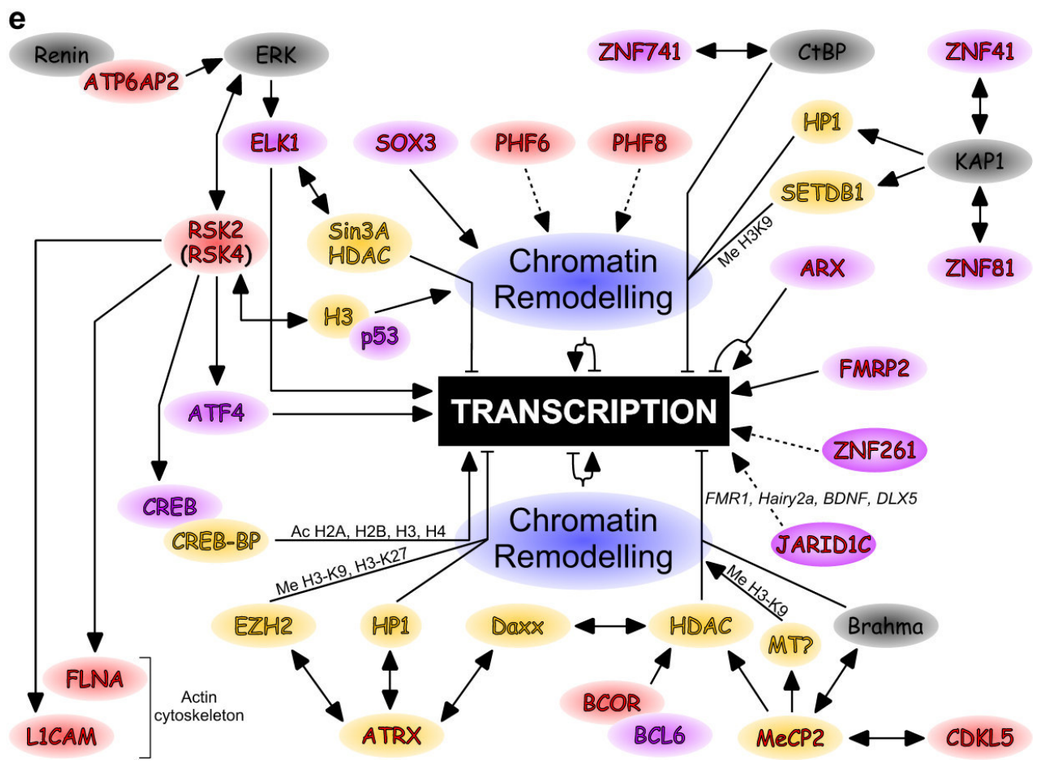
a. A mitotic chromosome consists of DNA that is densely packaged in highly organised structures. The scanning electron microscopic (EM) picture shows a human X chromosome. Chromatin isolated directly from an interphase nucleus appears as the so-called 30 nm fibre, in which the DNA is arranged in nucleosomes, which in turn are organised in higher order structures. After decondensation, individual nucleosomes (arrows) linked by spacer DNA (arrowheads) can be observed under the EM. Scale bars, 50 nm. Higher magnifications allow the visualisation of the non-compacted DNA, which is some ~50000 × longer than the mitotic chromosome in which it is packaged.
b, b', b''. Crystal structure of the nucleosome. **b.** The crystal structure of one half of the nucleosome core particle at 1.9 Å resolution illustrates the organisation of histones (H2A, bright orange; H2B, slate; H3, purple and H4, deep salmon) and DNA (deep olive and deep teal). Note that the representation of the protruding H2A, H2B and H3 tails has been smoothened for illustrative purposes and does not reflect the exact X-ray coordinates. **b'.** The entire structure shows how the histone octamer is assembled and how 147 base pairs of DNA are wrapped into a superhelix around it. Orientation and colour coding as in panel b. **b''.** Side view of the nucleosome, rotated 90° along its vertical axis with regard to panel b'. Note that the histone tails are only depicted partially. Colour coding as in panel b.
c. This diagram represents phosphorylation (P), acetylation (Ac) and methylation (Me) of residues in the tails of the four nucleosomal histones, which are colour-coded as in panel b. Arrows indicate cross-talk between some of these post-translational modifications (→ compatible and ⊣ incompatible modifications). The scheme only shows modifications for the histone tails but modifications for the central domains have also been reported. Moreover, other modifiers, such as Ubiquitin and Small Ubiquitin-like modifier, exist.
d. Through differential modification of histone tails, chromatin structure changes from transcriptionally active euchromatin (top panel) to transcriptionally inactive heterochromatin (bottom panel). This transition is thought to involve the recruitment of non-histone proteins (not shown).
e. Several proteins involved in mental retardation (red) regulate transcription either directly or via chromatin remodelling pathways (blue). Violet represents transcription factors (TFs, the colour gradient is reversed for predicted TFs), yellow chromatin-associated proteins. Dashed arrows indicate putative transcriptional regulation. Where appropriate, methylation (Me) of lysines 9 and 27 of Histone 3 (H3-K9 and H3-K27, respectively) and acetylation (Ac) of histone tails is shown. Four target genes of the MeCP2 pathway are presented, see text for details. → activation, ↔ direct interaction, ⊣ inhibition, and 'MT?', uncharacterised methyltransferase.

Images adapted from Darryl Leja, Access Excellence @ The National Health Museum, Washington, USA (a, main diagram); Andrew Syred, Photo Researchers Inc., New York, USA (a, SEM X chromosome); *420* (a, 30 nm chromatin fibre and beads-on-a-string); http://en.wikipedia.org (a, dsDNA) and *413* (d). Original X-ray coordinates 1KX5 from *421* were rendered using PyMOL (b, b' and b''). Diagram based on *422* (c).


a
Mitotic chromosome
1400 nm
30 nm chromatin fibre
30 nm
11 nm
Beads-on-a-string
2 nm
dsDNA
b
H3
H4
H2A
H2B
b'
b''


c
H2N-SGRGKQGGKARAKAKTRSSRAGLQFPVGRV
H2A
H2N-PEPAKSAPAPKKGSKKAVTKAQKKDGKKRK
H2B
H2N-ARTKQTARKSTGGKAPRKQLATKAARKSAPATGGVK
H3
H2N-SGRGKGGKGLGKGGAKRHRKVLRDNIQGIT
H4


d
Euchromatin
Heterochromatin


e
Renin
ATP6AP2
ERK
ELK1
SOX3
PHF6
PHF8
ZNF741
CtBP
ZNF41
HP1
KAP1
SETDB1
ZNF81
RSK2 (RSK4)
Sin3A HDAC
Chromatin Remodelling
Me H3K9
ARX
H3
p53
FMRP2
ATF4
TRANSCRIPTION
ZNF261
CREB
CREB-BP
Ac H2A, H2B, H3, H4
Chromatin Remodelling
FMR1, Hairy2a, BDNF, DLX5
JARID1C
Me H3-K9, H3-K27
Me H3-K9
EZH2
HP1
Daxx
HDAC
MT?
Brahma
FLNA
L1CAM
Actin cytoskeleton
ATRX
BCOR
BCL6
MeCP2
CDKL5

Finally, it should be noted that histone modification is not a transcription-specific mechanism. For example, DNA repair[423] and the formation of highly condensed centromeric heterochromatin[424] both use distinct mechanisms to generate histone modification patterns.

A.4.4.2. Mental retardation gene products regulate chromatin remodelling and transcription

Mutations in 18 genes encoding proteins involved in chromatin remodelling and/or regulation of transcription cause MR.

Eighteen established MR genes, $^{S+NS}$*ATRX*, ^{S}*PHF6*, ^{S}*PHF8*, ^{NS}*ZNF41*, ^{NS}*ZNF81*, ^{NS}*ZNF261*, ^{NS}*ZNF741*, $^{S+NS}$*RSK2*, ^{NS}*RSK4*, ^{S}*ELK1*, ^{S}*ATP6AP2*, $^{S+NS}$*MECP2*, ^{S}*CDKL5*, ^{NS}*FMR2*, ^{NS}*FMR3*, ^{S}*BCOR*, $^{S+NS}$*JARID1C* and ^{S}*SOX3*, encode proteins with functions in chromatin remodelling and/or transcriptional regulation (Fig. I-5e).

Mutations in $^{S+NS}$*ATRX*, which encodes the $^{S+NS}$ATRX protein [425], have been shown to cause the X-linked ATR-X syndrome (OMIM 301040)[175,186,426], characterised by severe MR and α-thalassaemia, but were also found in forms of S-XLMR without α-thalassaemia[178,179] and in NS-XLMR[176,177]. Using a conditional gene targeting approach, Bérubé and colleagues inactivated *Atrx* in the forebrain of mice and found that Atrx is required for neuronal survival during corticogenesis[182]. $^{S+NS}$ATRX associates with heterochromatin[181] and contains all seven motifs typical for the SNF2/SWI2 family of DNA-dependent ATPases/DNA helicases[180], which are involved in chromatin remodelling[427]. It interacts with the heterochromatin protein HP1 in human[181] and in mouse[428], binds EZH2[183], the homologue of *Drosophila*'s Enhancer of zeste, which is involved in chromatin remodelling[429], and forms an ATP-dependent chromatin remodelling complex with Daxx as a targeting subunit[184]. Daxx itself associates with HDAC chromatin remodelling complexes[430] and modulates apoptosis[431] and transcription[432]. Another line of evidence for the involvement of $^{S+NS}$ATRX in transcriptional regulation is the finding that mutations in $^{S+NS}$*ATRX* cause changes in the pattern of DNA methylation[433]. Moreover, $^{S+NS}$ATRX contains a PHD finger[174], a type of zinc-finger motif that chelates two Zn^{2+} ions per domain. The PHD finger domain is found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation[434].

Interestingly, two other PHD-encoding genes have been implicated in S- XLMR: ^{S}*PHF6* (OMIM 301900)[435-439] and ^{S}*PHF8* (OMIM 300263)[440]. The prominent nucleolar localisation and presence of two zinc-finger motifs in ^{S}PHF6 suggest a role for it as a transcriptional regulator[437]. Bioinformatic approaches suggest putative nuclear localisation signals, a PHD

finger domain and a jmjC domain[198] in ^{S}PHF8[440]. The jmjC domain is believed to be involved in chromatin organisation by modulating heterochromatisation[441], which, in combination with the other domains, made Laumonnier *et al.* propose a role for ^{S}PHF8 in modifying and regulating the structure of DNA[440]. Apart from these two, another four ZFPs, ^{NS}*ZNF41*, ^{NS}*ZNF81*, ^{NS}*ZNF261* and ^{NS}*ZNF741*, have been implicated in MR. ^{NS}*ZNF41*[152] and ^{NS}*ZNF81*[156], both encoding transcriptional repressors, are Krüppel-type Cys_2-His_2 ZFPs containing a KRAB domain. While the zinc-finger motif allows sequence-specific DNA binding[442], the KRAB domain represses transcription[150] via interaction with the KAP1 co-repressor[151]. KAP1 acts as a molecular scaffold targeted to specific loci by the KRAB ZFPs, where it then organises chromatin remodelling. This is achieved by histone methylation through binding with the H3-K9-specific methyltransferase SETDB1 and by depositing HP1[443], two mechanisms that effectively silence gene expression[444,445]. ^{NS}ZNF261[147] belongs to the novel MYM family of ZFPs[148], and *Drosophila* Woc, which shares homology with ^{NS}ZNF261, acts as a TF with a telomere capping function[149]. ^{NS}ZNF741[153] is a CACCC-box-binding Krüppel-like TF that associates with the co-repressor CtBP[154] and represses transcription[155], possibly through histone modifications[446].

Mutations in $^{S+NS}$*RSK2*, which encodes a member of the pp90^{rsk} kinase family[447], were found in patients with CLS (OMIM 303600)[246,252,448], as well as in affected members of a NS-XLMR family[247]. $^{S+NS}$RSK2 targets are manifold. It is required for phosphorylation of H3[248] and the tumour suppressor p53, whereby the $^{S+NS}$RSK2-p53-H3 complex is thought to contribute to chromatin remodelling[254]. $^{S+NS}$RSK2 also phosphorylates the CREB TF[449]. CREB interacts with the transcriptional co-activator CREB-binding protein[450], which has been shown to acetylate all four nucleosomal histones[451] and is required in the formation of long-term memory[249], a result supported by strong $^{S+NS}$*RSK2* expression in the hippocampus[452]. ATF4, a TF regulating osteoblast differentiation and function, is another $^{S+NS}$RSK2 substrate. Lack of ATF4 phosphorylation by $^{S+NS}$RSK2 may contribute to the skeletal phenotype of CLS[250]. Interestingly, ^{S}L1CAM[453] and ^{S}Filamin A[251], two MR proteins described earlier, are also $^{S+NS}$RSK2 targets.

Deletion of ^{NS}*RSK4*, a gene encoding another member of the pp90^{rsk} kinase family[454], has been shown to cause NS-XLMR[132]. Moreover, $^{S+NS}$RSK2 and ^{NS}RSK4 each contain a docking site for the MAPK ERK[454,455]. ERK phosphorylates ^{S}ELK1[456], leading to the recruitment of the Sin3A-HDAC I co-repressor complex[102], and thus links the ^{S}ELK1 TF with chromatin remodelling. ^{S}*ELK1* encodes the ETS-like protein 1, a TF containing a MAPK docking site[103] and a MAPK inducible transcription AD[457]. A p.T108A exchange in ^{S}ELK1 was found to co-

segregate with MR in a small three-generation family, implicating a role for ^{S}*ELK1* in the aetiology of MR[101]. ^{S}*Elk1* KO mice show mildly impaired neuronal gene activation, possibly because of functional redundancy[104]. Quite unexpectedly, mutations in ^{S}*ATP6AP2*[458], which codes for the human Renin receptor[459], were found to cause XLMR associated with epilepsy (OMIM 300423)[460]. Renin is an aspartyl protease essential for the control of blood pressure[461]. A possible link between the MR phenotype and the ^{S}*ATP6AP2* mutations could be the finding that binding of Renin on ^{S}ATP6AP2, which is highly expressed in the brain, activates ERK[459]. Indeed, functional analysis demonstrated that, although the mutated receptor could still bind Renin, ERK1/2 activation was modestly impaired[460].

$^{S+NS}$*MECP2*, mutations in which cause the female-specific Rett syndrome (OMIM 312750)[201], and different forms of S- and NS-XLMR in males[202-210,218,223,224], encodes $^{S+NS}$MeCP2[211]. $^{S+NS}$MeCP2 silences gene expression[212] by recruiting HDAC activity, resulting in chromatin modifications, thereby linking two global mechanisms of gene regulation, DNA methylation and histone deacetylation[462,463]. The finding of an *in vivo* interaction between $^{S+NS}$MeCP2 and an unidentified methyltransferase that specifically methylates lysine 9 of H3 establishes $^{S+NS}$MeCP2 as a link between DNA and histone methylation, thus reinforcing its role in chromatin remodelling[213]. Moreover, $^{S+NS}$MeCP2 has been suggested as a molecular connection between nuclear genome topology and epigenetic gene regulation maintaining cellular differentiation[464]. Recently, Harikrishnan *et al.* have shown that Brahma, a catalytic component of the SWI/SNF-related chromatin remodelling unit[465], binds $^{S+NS}$MeCP2 to form a co-repressor complex. Interestingly, their data suggest that the recruitment of this complex is a crucial feature of ^{S}*FMR1* repression[466]. Other genes silenced by $^{S+NS}$MeCP2 include *Hairy2a*[467], *BDNF*[468] and *DLX5*[221]. Hairy2a plays a role in neurogenesis[217] and BDNF is involved in neuronal survival[215,216]. It was noted that Bdnf levels were reduced in *Mecp2* KO mice and that the phenotype was alleviated upon over-expression of *Bdnf*, demonstrating a functional interaction between *Bdnf* and *Mecp2*[469]. Recent investigations of these mice showed that Mecp2-mediated chromatin looping at the *Dlx5 – Dlx6* locus is upset, resulting in a loss of maternal imprinting of the *Dlx5* gene, divulging a novel mechanism underlying gene regulation by Mecp2[221]. The finding that $^{S+NS}$MeCP2 deficiency causes epigenetic aberrations at the Prader-Willi syndrome imprinting centre on chromosome 15 supports these findings[470]. Further analyses of Mecp2-null mice, recapitulating Rett syndrome, indicated that $^{S+NS}$MeCP2 is important for neuronal survival[214] and revealed alterations in excitatory synaptic plasticity, possibly due to altered expression of NMDA receptor subunits[220,222], and highlighted a role for Mecp2 in learning and memory[219].

When screening a cohort of patients with a Rett-like phenotype, but without $^{S+NS}$*MECP2* mutations, several mutations were identified in ^{S}*CDKL5*[471,472], which encodes a cyclin-dependent Ser/Thr kinase[473]. Earlier, this gene had been implicated in severe X-linked infantile spasms (OMIM 308350) associated with MR[474]. Based on the extensive phenotypic overlap of patients with $^{S+NS}$*MECP2* and ^{S}*CDKL5* mutations, and the observation that activation of $^{S+NS}$MeCP2 target genes is associated with its increased phosphorylation[468], it has been proposed that ^{S}CDKL5 might be involved in $^{S+NS}$MeCP2 phosphorylation[471,472]. Indeed, it has been shown that $^{S+NS}$MeCP2 and ^{S}CDKL5 interact *in vitro* as well as *in vivo*, and that ^{S}CDKL5 mediates $^{S+NS}$MeCP2 phosphorylation[475], although the latter finding has been challenged recently[476].

$^{S+NS}$*ARX*, the gene encoding the orthologue of the *Drosophila* Aristaless[164] homeodomain protein[163], was found to be mutated in several forms of S-MR[62,161,166], as well as in NS-MR[78,162,165,167-169]. The homeodomain is a DNA-binding motif[477,478] and, from *Drosophila* classical and molecular genetics, it has long been known that homeodomain TFs control developmental switches in anterior-posterior patterning[479,480]. Therefore, it was thought that Aristaless-related TFs may play important roles in vertebrate embryogenesis, especially in head development[481]. Indeed, several studies now support this idea. *Xenopus* Arx acts as a bifunctional transcriptional regulator in forebrain development[173], *Arx* KO mice exhibit abnormal development of the forebrain because of decreased neural precursor proliferation[161], and mArx is essential for the establishment of functional olfactory neural circuitry[172].

Like FRAXA, described earlier, FRAXE is a folate-sensitive fragile site on the X chromosome[482]. FRAXE associates with mild MR, which was found to result from transcriptional silencing of the nearby ^{NS}*FMR2* gene[107,108] by hypermethylation of a CCG trinucleotide repeat[483] located in its first exon. Subsequent molecular studies on the ^{NS}FMR2 protein indicated that it is expressed in the nuclei of the cells making up the granular cell layer of the hippocampus[484], consistent with its proposed function as a potent transcriptional activator[109]. Animal studies showed that *Fmr2* KO mice exhibit impaired conditioned fear and enhanced LTP, suggesting a physiological role in regulation of synaptic plasticity[106]. It should be noted that ^{NS}*FMR3*, a gene transcribed from the opposite strand to ^{NS}*FMR2*, is also absent in FRAXE MR patients[110]. A boy with NS-MR has recently been reported in whom ^{NS}*FMR3* expression was abolished but ^{NS}*FMR2* expression was intact, suggesting a causal relationship between mutated ^{NS}*FMR3* and MR[72].

Another gene affecting mental ability[485,486] that is directly involved in transcription is the BCL-6 co-repressor ^{S}BCOR[487]. BCL-6 itself is a TF that was originally identified in B-cell

lymphoma by characterising the BP of a recurrent chromosome 3q27 translocation[488]. ^{S}BCOR has been shown to interact with HDACs[487], and knock down of its zebrafish orthologue recapitulated the human phenotype and confirmed the idea that ^{S}BCOR is a key transcriptional regulator during early embryogenesis[485].

$^{S+NS}$*JARID1C* was only recently identified as one of the more frequently mutated genes underlying XLMR[195,196,489]. It encodes one of the four known JARID1 proteins, all of which contain an ARID DNA-BD[194] and a jmjC domain.

A 33 bp in-frame insertion in ^{S}*SOX3*[490], leading to an additional 11 alanines in a poly-alanine tract of the ^{S}SOX3 protein, has been shown to co-segregate with XLMR associated with isolated growth hormone deficiency[491] (OMIM 300123). ^{S}SOX3 is required for the formation of specific CNS midline structures in mouse[492]. The ^{S}SOX3 protein contains an HMG domain [490], which bends DNA by binding it in its minor groove[493]. HMG-containing proteins are thought to affect chromatin architecture[494] and DNA characteristics[495], eventually resulting in the regulation of transcription[496].

A.4.5. Synaptic function in mental retardation

Neuronal trajectories are interrupted by chemical (and electrical) synapses, allowing for adaptable and versatile signal processing. Not surprisingly, several XLMR genes have been linked to synaptic function.

A.4.5.1. General overview

Regulation of signalling across chemical synapses depends on (i) which neurotransmitters are employed, (ii) their quantity and mode of release, and (iii) their rate and mode of removal.

By applying new staining techniques introduced by the Italian physician Camillo Golgi (1843 – 1926), Santiago Ramón y Cajal (1852 – 1934), a Spanish histologist and arguably the father of neuroscience, was the first one to show that neurons do not form a continuous network, but instead are physically separated from each other[497,498]. We now call this physical separation the synaptic cleft and, based on its dimensions, we distinguish between electrical and chemical synapses. In electrical synapses, specialised for rapid signal transmission, the synaptic cleft is only 20 Å wide and protein channels that allow ions to flow from one cell to the next interconnect the cytoplasm of adjacent neurons. These synapses are abundant in embryonic nervous tissues when neuronal connections are established, but most of them are gradually replaced with chemical synapses as development of the neural system

continues. In chemical synapses, the synaptic cleft is > 200 Å, requiring an indirect mode of signal transmission (Fig. I-6). Essentially, the information transfer across a chemical synapse involves five successive events (Fig. I-6c):

- Depolarisation at the pre-synaptic nerve terminal opens voltage-gated Ca^{2+} channels.
- The Ca^{2+} influx triggers exocytosis of neurotransmitter-loaded SVs, releasing their content into the synaptic cleft.
- The neurotransmitter diffuses across the synaptic cleft and activates transmitter-gated ion channels on the post-synaptic membrane.
- Influx of ions leads to localised changes in the membrane potential, which in turn result in the opening of voltage-gated ion channels.
- Depending on the type of neurotransmitter and its downstream effects, the post-synaptic neuron may be excited or inhibited.

Although chemical synapses have the disadvantage of resulting in a synaptic delay that is the rate-limiting step in neural transmission, they are very versatile and adaptable. For example, regulation of chemical synapses has been shown to occur via the nature of neurotransmitters employed[499], their quantity and mode of release[500], and their rate of removal through reuptake[501] or enzymatic degradation[502,503].

The mammalian nervous system is thought to employ more than 130 different neurotransmitters[504]. Some of the best characterised are acetylcholine[505], glutamate[506] and serotonin[507] (all with excitatory effects), and GABA[508,509] and glycine[510,511] (both inhibitory neurotransmitters). It is widely accepted that neurotransmitter metabolism plays an important role in mood, behaviour and neurodegenerative disease. Therefore, it should not be surprising that irregularities in the neurotransmitter household have been implicated in multifactorial psychological conditions such as schizophrenia[512] or mood and anxiety disorders[513], and in pathologies such as PD (OMIM 168600)[514].

A.4.5.2. X-linked mental retardation gene products regulate synaptic function

In eight instances, MR can be attributed to mutations in genes important in synaptic functionality.

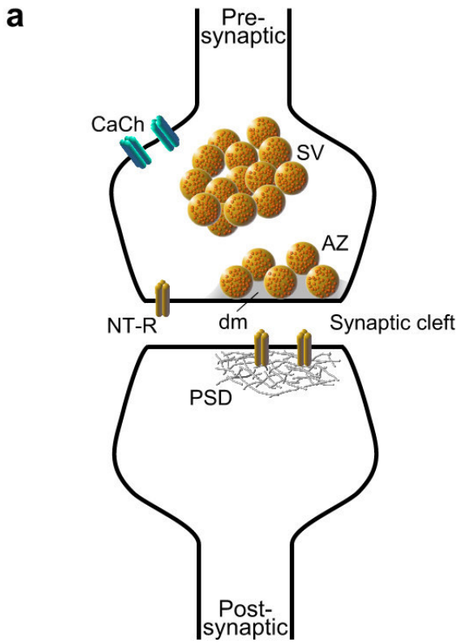
Eight genes that have been implicated in cognitive function, ^{NS}*GDI1*, ^{S}*SYN1*, $^{S+NS}$*NLGN4*, ^{S}*NLGN3*, ^{NS}*IL1RAPL1*, ^{NS}*DLG3*, ^{NS}TM4SF2 and ^{S}*MAOA*, were shown to play roles in synaptic functioning (Fig. I-6b).

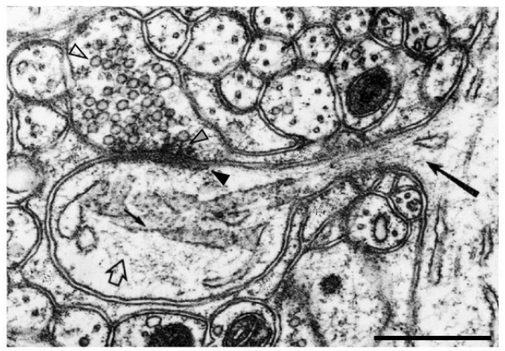
^{NS}*GDI1* was among the first genes to be implicated in NS-XLMR[114]. The gene encodes the GDI ^{NS}αGDI[116]. GDIs (see I.A.4.3.1) play an important role in the regulation of Rab GTPases[118], a subfamily of the small G-proteins, which are involved in membrane

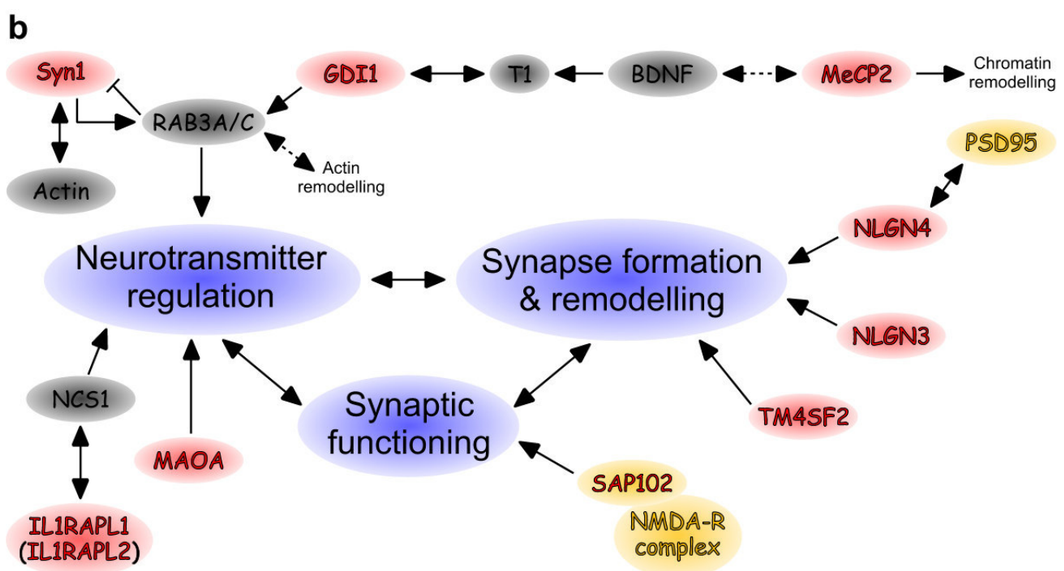
Fig. I-6 | *Next page.* **Mental retardation and synaptic function.**

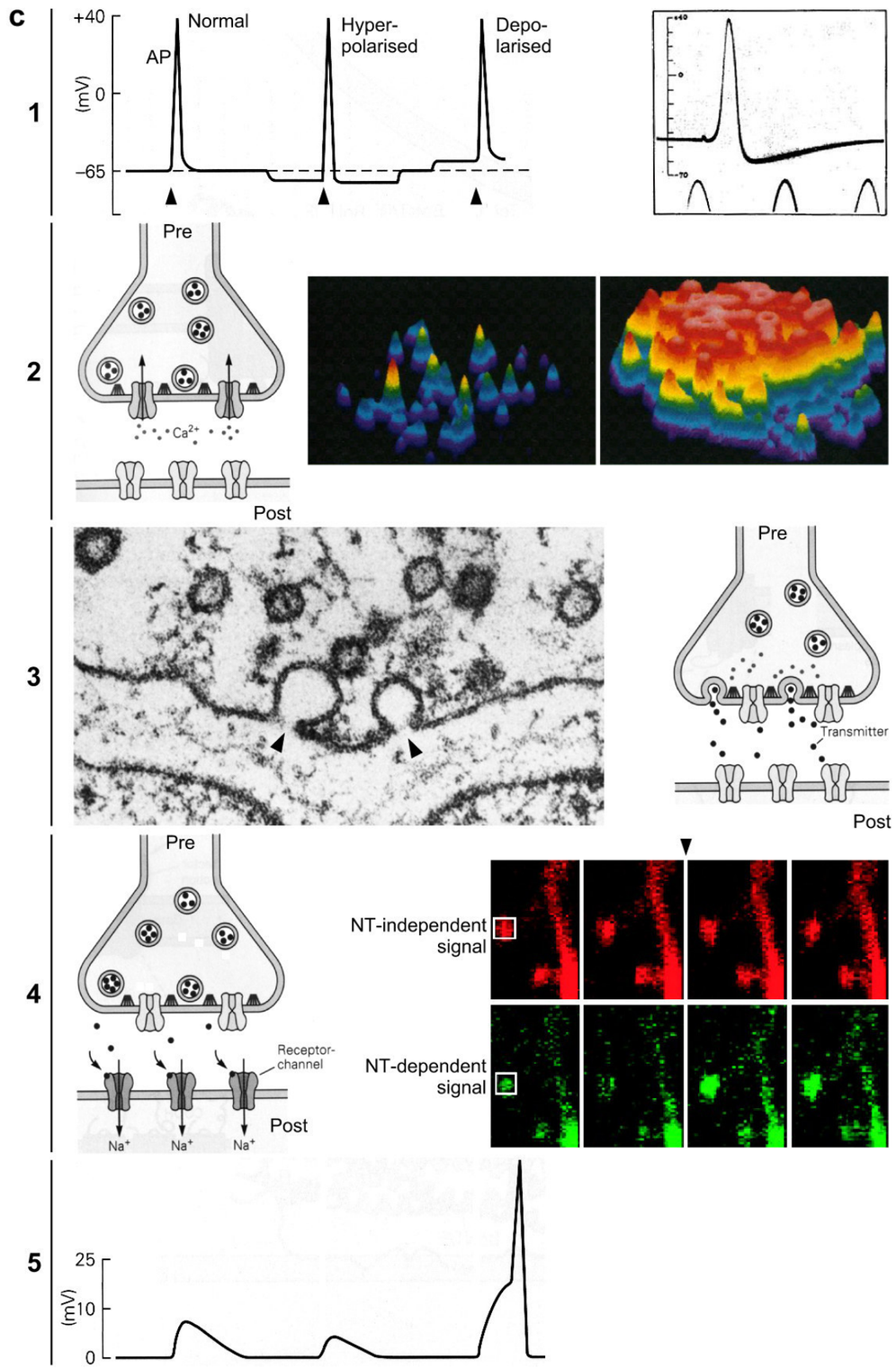
a. The left panel shows a schematic representation of a synapse. Calcium channels (CaCh, cyan rods), synaptic vesicles (SV, yellow spheres) loaded with neurotransmitters (red spheres), and neurotransmitter receptors (NT-R, yellow rods) are indicated. A dense matrix (dm) surrounds docked vesicles in the active zone (AZ) at the pre-synaptic plasma membrane. Across the synaptic cleft, the proteinaceous post-synaptic density (PSD) is present in precise register with the AZ at the pre-synaptic plasma membrane. Note that the post-synaptic contact site is not necessarily a synaptic bouton, but can also be a dendrite or a muscle fibre, for example. The right panel, an electron microscopic (EM) recording, illustrates the morphology of a dendritic spine. The spine receives a synapse from a parallel fibre axon. Actin forms a lattice in the spine head (open arrow) and parallel microfilaments in the spine neck (big arrow). The small arrow points to dense spots likely representing Ca^{2+} channels. Arrowheads indicate the PSD (black), the dense matrix at the AZ (grey) and SVs (open). Scale bar, 0.5 μm.
b. A number of proteins involved in mental retardation (red) affect synapse function (blue), either directly or through regulation of synaptic development or neurotransmitter release. See text for details. Yellow symbolises PSD-associated proteins, dashed double arrows indirect links to the Actin and chromatin remodelling pathways. → activation, ↔ direct interactions, ⊣ inhibition and NMDA-R, N-methyl-D-aspartate receptor.
c. Simplified illustration of the five events occurring during information transfer across a chemical synapse. **c1.** *Left* Stimuli (arrowheads) result in action potentials (APs) at the pre-synaptic terminal (see also c5). *Right* Historic tracing of the first published intracellular recording of an AP. It was measured in the squid giant axon, using glass capillary electrodes filled with sea water. The vertical scale indicates the differential potential in mV between the internal electrode and the external sea water. Time marker (bottom) is 500 Hz. **c2.** *Left* An AP arriving at the terminal of a pre-synaptic axon causes voltage-gated Ca^{2+} channels at the AZ to open. *Middle and right* 3D projection after image integration of quantum emission domains at a pre-synaptic site in the giant squid synapse before (middle) and during (right) stimulation shows a massive influx of calcium (blue, low [Ca^{2+}]; green, intermediate [Ca^{2+}]; red, high [Ca^{2+}]). The non-homogeneous signal distribution reveals the existence of calcium microdomains. **c3.** *Left* An EM recording (magnification 145000 ×) of a freeze-substituted neuromuscular junction in the frog displays SVs caught in the act of exocytosis (arrowheads). *Right* A high [Ca^{2+}] near the AZ causes SVs to release their neurotransmitter (NT) content into the synaptic cleft through exocytosis. **c4.** *Left* Released NT molecules diffuse across the synaptic cleft and bind to specific receptors on the post-synaptic membrane, causing ion channels to open (or close). Not all molecules reach the post-synaptic site; transmitters also re-enter the pre-synaptic bouton, are enzymatically degraded, diffuse away or are taken up by adjacent cells, such as glia. *Right* Two-photon imaging of NMDA-R activation reveals a clear increase in NT-dependent signal (bottom panel) in single dendritic spines (boxed area) after stimulation (arrowhead). Such an increase is not observed in the NT-independent channel (top panel). **c5.** The altered state of the ion channels after NT binding causes a change in the membrane conductance and potential of the post-synaptic cell. In conclusion, changes in membrane potential of the pre-synaptic terminal (see c1, left panel) affect the intracellular concentration of Ca^{2+} and, thus, the amount of NT released. When the pre-synaptic membrane is at its normal resting potential, an AP (c1, left panel, left trace) produces a post-synaptic potential of a given size (c5, left trace). Hyperpolarising the pre-synaptic terminal prior to an AP (c1, left panel, middle trace) decreases the steady-state Ca^{2+} influx, so that the same-size AP produces a smaller post-synaptic potential (c5, middle trace). Depolarising the pre-synaptic neuron increases the steady-state Ca^{2+} influx, so that the same-size AP (c1, left panel, right trace) produces a post-synaptic potential large enough to trigger an AP in the post-synaptic cell (c5, right trace).

Images adapted from *515* (a, right); *516* (c1, left; c2, left; c3, right; c4, left and c5); *517* (c1, right); *518* (c2, middle and right); *519* (c3, left) and *520* (c4, right).


a
Pre-
synaptic
CaCh
SV
AZ
NT-R
dm
Synaptic cleft
PSD
Post-
synaptic




b
Syn1
GDI1
T1
BDNF
MeCP2
Chromatin
remodelling
RAB3A/C
Actin
remodelling
Actin
PSD95
NLGN4
NLGN3
Neurotransmitter
regulation
Synapse formation
& remodelling
NCS1
Synaptic
functioning
TM4SF2
MAOA
SAP102
NMDA-R
complex
IL1RAPL1
(IL1RAPL2)


c
1
+40
0
–65
(mV)
AP
Normal
Hyper-
polarised
Depo-
larised
2
Pre
Ca2+
Post
3
Pre
Transmitter
Post
4
Pre
Receptor-
channel
Na+
Na+
Na+
Post
NT-independent
signal
NT-dependent
signal
5
25
10
0
(mV)

trafficking[521]. [NS]αGDI regulates RAB3A and RAB3C[119], both participating in SV fusion[120,121], thus linking MR to synaptic dysfunction. Interestingly, [NS]αGDI binds T1[117], a neurotrophin receptor predominantly expressed in the adult CNS[522]. Through T1, BDNF exerts its influence on many biological processes, such as neuronal survival[523]. As mentioned earlier, a functional interaction between *mBdnf* and *mMecp2* exists[469]. Physiological studies on the anatomically normal *mGdi1* KO revealed involvement of αGdi in short-term synaptic plasticity[122], formation and retention of temporal associations, and in social behaviour[115], thereby recapitulating some of the characteristics of NS-MR. Surprisingly, *Rab3a* KO mice demonstrate none of these behavioural changes; they show impaired spatial reversal learning and increased explorative activity[524].

In line with the link between synaptic dysfunction and MR, [S]*SYN1*, which encodes Synapsin I[525], a member of the SV-specific synapsins[526] regulating neurotransmitter release[527,528], was recently reported to cause epilepsy with or without associated features, such as learning difficulties and aggressive behaviour, when mutated[529]. [S]Synapsin I is a RAB3A interaction partner[530]. The presence of [S]Synapsin I (i) stimulates GTP binding and GTPase activity of RAB3A, and (ii) prevents GDI-induced RAB3A dissociation from SVs. Conversely, RAB3A inhibits (i) [S]Synapsin I's interactions with Actin and (ii) [S]Synapsin I-induced vesicle clustering. Taken together, the [S]Synapsin I – RAB3A interaction may play an important role in the modulation of neurotransmitter release[531].

Surprisingly, flies[532] and mice[533] lacking all synapsin isoforms are healthy and show no obvious defect in brain morphology. However, they exhibit a number of behavioural defects[532,533], and KO flies were shown to be impaired in learning and memory[532]. Careful investigation of cultured neurons from the hippocampus of KO mice indicated that, while synapsins maintain the reserve pool of glutamatergic vesicles, they also regulate the size of the readily releasable pool of GABAergic vesicles, showing different pre-synaptic roles at excitatory and inhibitory synapses[533].

Mutations in [S+NS]*NLGN4* were shown to be involved in MR with[227] or without[226] autism. [S+NS]*NLGN4* encodes Neuroligin 4, a member of the neuroligin protein family that binds to the PSD protein PSD-95[534]. Neuroligins are TM cell adhesion molecules, particularly abundant in the post-synaptic membrane of glutamatergic synapses[228], that trigger pre-synaptic development[229], suggesting an important role in the formation or remodelling of synapses. In addition, a mutation in [S]*NLGN3* has been reported in a family with autism-spectrum

disorders[227]. The disorder-associated mutant $^{S+NS}$NLGN4 and ^{S}NLGN3 proteins have been shown to remain in the endoplasmic reticulum, and hippocampal neurons over-expressing them did not stimulate the formation of pre-synaptic terminals as compared to the WT proteins[230]. These experiments further strengthen the connection between a loss of synaptic function and autism or NS-MR.

Mutations in ^{NS}*IL1RAPL1* were found in patients with NS-XLMR[123,535]. Through interaction with NCS1[125], ^{NS}IL1RAPL1 may regulate Ca^{2+}-dependent exocytosis at nerve terminals[124]. The *Drosophila* NCS1 homologue Frequenin has been shown to affect synaptic efficacy[536], and NCS1 itself has been implicated in facilitating pre-synaptic Ca^{2+} currents[537]. Intriguingly, *IL1RAPL2*, a homologue of ^{NS}*IL1RAPL1* also located on the X chromosome, is specifically expressed in the CNS, leading to the speculatation that mutations in *IL1RAPL2* may be responsible for disorders affecting cognitive delay, including NS-XLMR[538].

SAP102 is concentrated in the PSD and is expressed in neurons during early brain development[539]. It interacts with NMDA receptor complexes and is thought to take part in clustering them in the PSD[100]. It is widely accepted that NMDA receptors, ion channels that only gate when glutamate is bound *and* the membrane is depolarised[540], play an important role in LTP[541], the mechanism underlying learning[542], and in developmental adjustments of synaptic connections[268]. Four different mutations in ^{NS}*DLG3*, the gene encoding SAP102, all truncate SAP102 and are predicted to impair affinity for NMDA receptors. These mutations lead to NS-XLMR[543].

^{NS}TM4SF2[134-137,544] is a member of the four-fold TM tetraspanin protein family[138], highly expressed in the cerebral cortex and the hippocampus[134]. The *Drosophila* neural tetraspanin Lbm has been reported to facilitate synapse formation[139], and an analysis of the fly's genome identified 34 additional tetraspanin family members, including two that enhance the *lbm* phenotype when deleted[140].

MAOs[545], localised at the MOM[546], catalyse the oxidative degradation of biogenic amines, including the neurotransmitters adrenaline[547], dopamine[548] and serotonin[549]. It is worthy of note that nonsense mutations in ^{S}*MAOA*[550] leading to ^{S}MAOA deficiency have been implicated in borderline MR associated with disturbed regulation of impulsive aggression[551]. The generation of ^{S}*Maoa* KO mice supports the link between ^{S}MAOA deficiency and aggressive behaviour[552], and another study has found an association between low ^{S}MAOA levels and lowered IQs associated with autism[553]. Still, there are also reports that failed to replicate such associations[554]. Such a discrepancy is not uncommon in MAOA research; associations

between MAOA levels and bipolar disorder[555-560], schizophrenia[561,562] or PD[563,564] have all been observed as well as questioned.

A.4.6. The Ubiquitin – proteasome pathway in neurodegenerative disease

As a multitude of biological processes depends on cells progressing through the cell cycle, (regulated) protein degradation via the UPP is of paramount importance in biology. Faulty protein degradation, cellular inability to cope with misfolded proteins and inefficient degradation machinery all have been implicated in human disease.

A.4.6.1. General aspects of the Ubiquitin – proteasome pathway

The UPP is a tightly controlled and temporally regulated degradation machinery specific for intracellular 'self' proteins.

During proteolysis, proteases hydrolyse the peptide bonds that link the amino acids, releasing the free amino acids and energy that was previously invested to create the peptide bond. With regard to degradation, proteins can be divided into three main categories:

- The 'foreign' proteins that are mostly taken up as nutrients and degraded in the gastrointestinal tract into short non-antigenic peptides.
- The extracellular 'self' proteins, such as immunoglobulins and peptide hormones, which are taken up via pinocytosis or receptor-mediated endocytosis to be non-specifically degraded in the primary lysosomes.
- The intracellular 'self' proteins responsible for structure and function of the cell.

Several indications point towards the existence of a highly specific, tightly controlled and temporally regulated degradation machinery specific for the third group of proteins. First, proteins have a varying half-life, ranging from minutes (e.g. the enzyme Ornithine decarboxylase[565]) to days (e.g. the muscle protein Myosin[566]) or even years (e.g. the eye lens component Crystalline[567]). Second, inhibitors of lysosomal degradation do not have any effect on the destruction of intracellular proteins. And third, proteolytic enzymes and their substrates reside in the same subcellular compartments.

The discovery of the Ubiquitin proteasome proteolytic pathway by A. Hershko, A. Ciechanover, I. Rose and A. Varshavsky[568,569] proved the idea of a specific and tightly controlled degradation pathway to be correct. Since then, Ubiquitin-dependent proteolysis has been shown to be involved in cell cycle progression[570-572], differentiation and development[573], signal transduction[574] and transcriptional regulation[575,576], among other cellular processes. Not

surprisingly, aberrations in the UPP underlie a multitude of pathogeneses, including multiple neurodegenerative disorders[577].

A.4.6.2. Overview of the Ubiquitin – proteasome pathway

Ubiquitinylation involves three enzymatic activities: E1 activates Ubiquitin, E2 conjugates it and E3 ligates it to the target protein or to a previously attached Ubiquitin moiety. Poly-ubiquitinylated proteins are degraded by the proteasome.

Ubiquitin is a highly conserved 76 AA protein[578] that is covalently attached to substrate proteins, tagging them for degradation by the 26S proteasome. The UPP involves the activity of three enzymes or protein complexes: the Ubiquitin-activating enzyme (E1), a Ubiquitin-conjugating enzyme (E2) and a Ubiquitin ligase complex (E3). A single E1 activates Ubiquitin for all conjugation reactions and transfers it to all known E2s. Most E2s interact with several E3s, which, in turn, bind a large variety of substrates. A single proteasome then degrades all poly-ubiquitinylated proteins. E2 – E3 – substrate combinatorics, multiple post-translational modifications of the substrate, and subcellular localisation of E3s and substrates add a high complexity level to Ubiquitin-dependent proteolysis. DUBs, cysteine proteases that hydrolise the amide bond immediately after the C-terminal residue, recycle the Ubiquitin monomers[579].

In a two-step intramolecular, ATP-dependent reaction, E1 catalyses the formation of a high-energy thiolester bond between the C-terminus of Ubiquitin and a conserved E1 active site cysteine residue[580,581].

Any of the > 20 E2s to which the Ubiquitin moiety is transferred, forming a thiolester bond with a conserved E2 active site cysteine residue, binds to distinct E3s[582] or interacts directly with the substrate[583]. It should be noted, however, that the latter is the less frequent process.

E3s facilitate the formation of an isopeptide linkage between Ubiquitin and the ε-NH_2 group of an internal lysine residue of the target protein[584] or lysine 48 or 63 on the previously conjugated Ubiquitin molecule[585]. Repetition of this ligation reaction assembles a poly-Ubiquitin chain on the target protein that is thereby recognised by the 26S proteasome[586], a multi-subunit[587], ATP-dependent protease complex[588]. At least four Ubiquitin residues are necessary for recognition and subsequent degradation by the proteasome (Fig. I-7a)[589]. Occasionally, a Ubiquitin chain is added on the NH_2-terminus of the target protein, the so-called 'N-end rule' pathway in which the N-terminal amino acid of the substrate dictates recognition[590].

A.4.6.3. The E3 Ubiquitin ligase complexes

E3s can be divided into the HECT and RING domain-containing superfamilies. The latter includes the single protein E3s E3α and E3β, as well as the multi-subunit complexes APC, VBC and SCF.

The E3 ligases serve as the recognition factors in Ubiquitin-mediated protein degradation through their substrate specificity. Based on shared structural motifs, it is estimated that the human genome codes for ~1000 different E3s[591]. The combination of their specificity and abundance explains part of the versatility of the UPP.

With the emergence of several protein complexes that all showed Ubiquitin ligase activity and similar molecular architectures, it became apparent that there exist two superfamilies of E3 ligases, the HECT and the RING domain-containing superfamilies (Fig. I-7b)[584,592-594]. Whereas HECT domain E3s form a thioester intermediate with Ubiquitin before transferring Ubiquitin to the ligase-bound substrate[595], RING domain E3s catalyse the direct transfer of Ubiquitin to the poly-Ubiquitin chain or to the E3-bound substrate[596]. RING domain E3s can be single polypeptides or multi-subunit complexes. While single protein RING E3s contain both the RING domain and the substrate binding site[597], complexes in the RING superfamily contain an adaptor protein that binds the substrate to the core complex. This core contains four proteins: Skp1 or the Elongin C homologue of Skp1 that bridges the adaptor protein to a cullin, a cullin or cullin-like scaffold protein, and a RING protein and specific E2 that bind to the cullin (Fig. I-7b)[584,592,594]. Some complexes may contain additional proteins.

Among the RING E3 ligases, the following belong to the best-characterised to date:

- The single protein E3 Ubiquitin ligases E3α[598] and E3β[599] target proteins by the 'N-end rule' pathway.
- The APC, a multi-subunit E3 complex, targets mitotic cyclins and other regulators of mitosis for destruction, thereby controlling the metaphase to anaphase transition[600,601].
- Through the VHL protein, the VBC complex[602] induces oxygen-dependent ubiquitinylation and degradation of HIF-1α[603]. HIF-1α forms a heterodimer with HIF-1β to function as a TF central to cellular responses to hypoxia[604].
- The SCF complex is described in detail in the next section.

It should be noted that RING-like domain containing E3s, such as PHD[605] and U-box[606] E3s, have also been identified. The PHD[607] and U-box[608] domains represent structural variants of the RING domain.

A.4.6.4. The SCF E3 Ubiquitin ligase complex

Cullin 1 is the scaffold in the complex, binding Roc1, E2 and Skp1. Roc1 is a RING finger protein, which binds E2 and an FBP. E2's core enzymatic domain conjugates activated Ubiquitin. Skp1 bridges between Cullin 1 and an FBP. FBPs bind Skp1 via their F-boxes and confer substrate specificity to the complex via their divergent protein – protein interaction motifs.

The SCF E3 Ubiquitin ligase complex mediates the degradation of cyclins and CDKIs, as well as many other proteins[609,610]. SCF complexes contain five proteins: Cullin 1, Roc1, E2 (Cdc34), Skp1 and an FBP (Fig. I-7c). For degradation to take place, substrates must be phosphorylated.

scCdc53 was the first member of the Cullin protein family to be characterised[611]. So far, seven cullins are reported in mammals: Cul1[612], Cul2[613], Cul3[614], Cul4A[615], Cul4B[616], Cul5[617] and Cul7[618]. In addition, the human genome codes for at least two cullin-like proteins, Parc[619] and the APC-specific APC2[620]. The cullins organise the E3 complex[621] and recruit the E2[611]. scCdc53 consists of three domains. The first one is a 219 AA N-terminal domain that directly binds scSkp1[622]. The second domain is an internal CH domain[620], binding the scCdc34 E2 and the RING finger-containing Roc1 protein[623]. The third domain is a short C-terminal domain that mediates auto-rubynilation (i.e. the attachment of Rub1, the yeast homologue of Nedd8)[624,625]. Nedd8 is an 81 AA protein that is 60% identical and 80% homologous to Ubiquitin[626].

Roc1, the RING finger component of the SCF complex, interacts with Cullin 1[627], E2[628] and FBPs such as Grr1 and Cdc4[609,627,629]. Ubiquitin ligase activity *in vitro* is abrogated when conserved cysteine and histidine residues in the RING-H2 domain are mutated, but binding with Cdc53 is not affected, suggesting dysfunction through loss of the interaction with E2[627].

SCFs interact with the Ube2d1 family of E2s[630] and with scCdc34 E2[609]. E2s consist of a core enzymatic domain and a unique C-terminal tail[631]. Domain-swapping experiments showed that the scCdc34 tail is able to confer cell cycle regulatory function upon heterologous proteins[632]. Mutation analyses indicated that the proximal portion of the tail domain mediates scCdc34 assembly with the SCF complex[633,634].

So far, only one *Skp1* gene has been discovered in man[635]. Skp1, a CyclinA/CDK2-associated protein[635], binds the N-terminus of Cullin 1[622], and deletion mutants show that the first 91 AA of hSKP1 contain cullin and F-box BSs[636], thus acting as a bridge between cullin and an FBP.

First identified in Cyclin F as a protein – protein interaction motif, the F-box is a conserved motif of approximately 45 AA[610], characterised by a consensus submotif near the N-terminus (LPxEILxxILsyLL; x: any AA; capitals: conserved residues) and by a consensus of

the submotif in the middle (VCKxW)[637]. The most highly conserved is the LP dipeptide that is, for example, required for binding of FBP Cdc4 with scSkp1[610]. However, interaction of hFBXL1 with hSkp1 does not require the conserved proline[638], and some F-box family members lack the dipeptide altogether[637], although it is unknown whether these FBPs interact with Skp1.

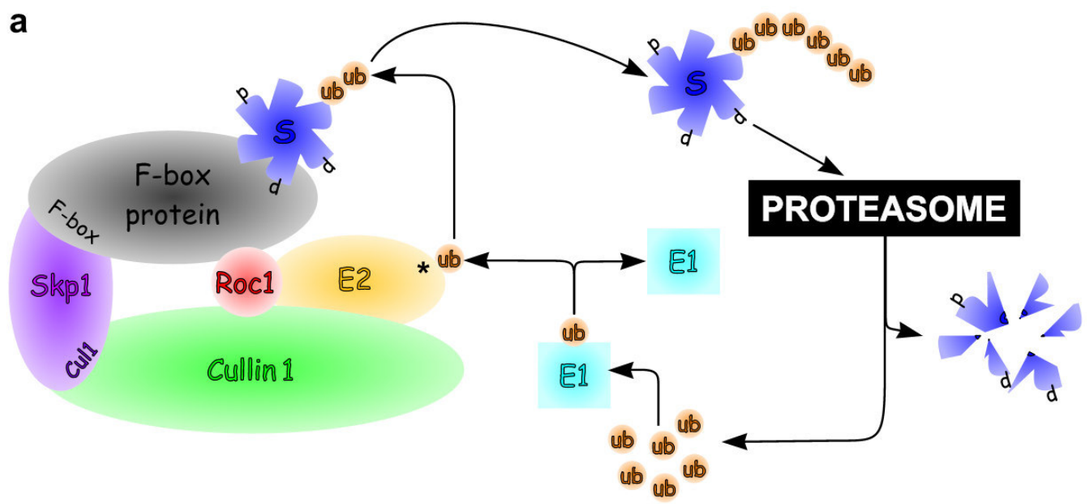
Several lines of evidence suggest that the F-box alone is not sufficient for binding to Skp1:

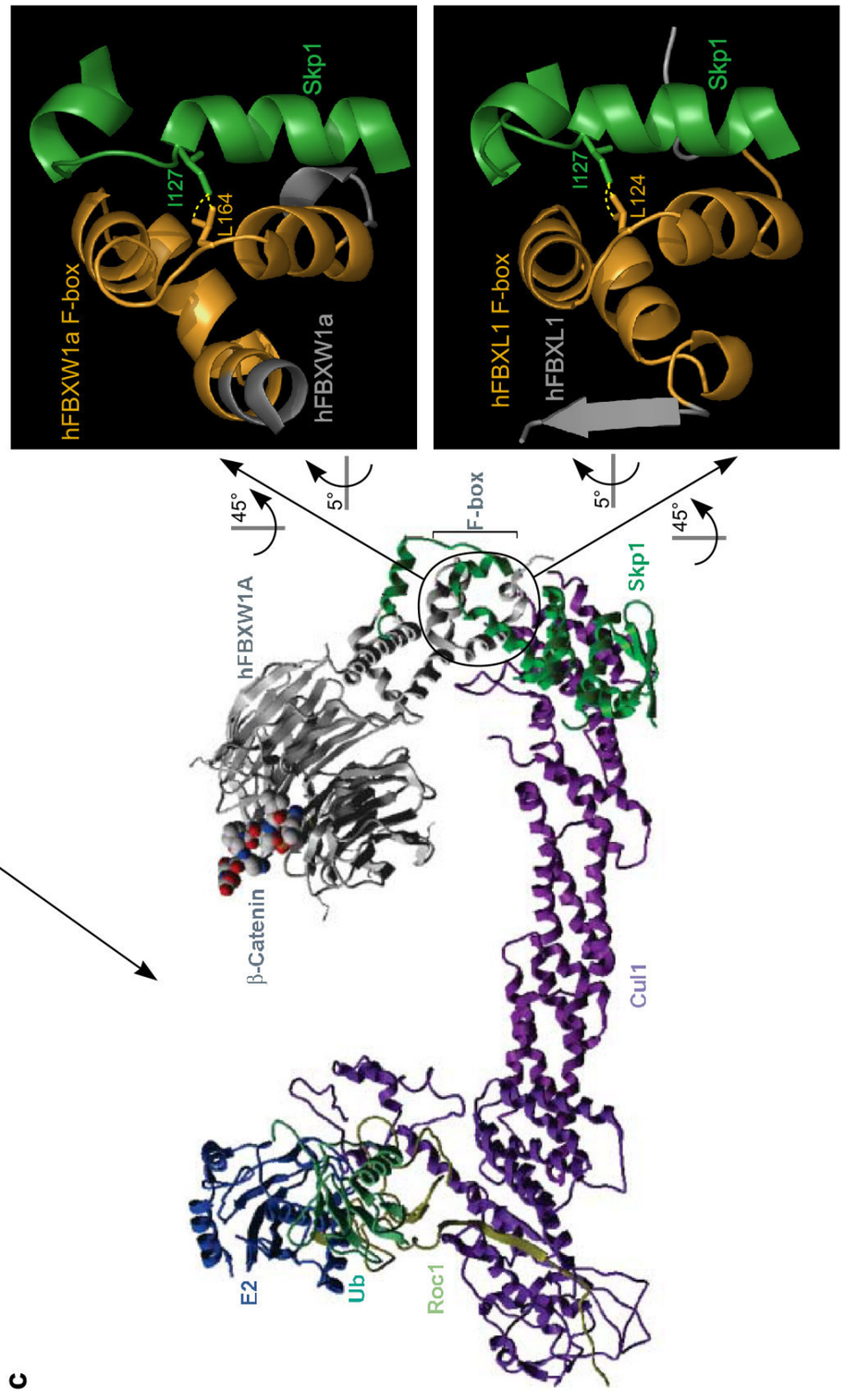
- The Grr1 F-box alone does not bind scSkp1[639]. Also, an N-terminal fragment of FBP scCdc4 lacking its WD-40 repeats but containing the F-box co-immunoprecipitates poorly with scSkp1[640].
- Mutation of two residues in hFBXL1 far downstream of the F-box abolishes binding to hCyclin A and hSkp1[612].

Fig. I-7 | *Next page.* **The Ubiquitin – proteasome pathway and E3 Ubiquitin ligases.**

a. Roc1 binds the E2 Ubiquitin-conjugating enzyme, which acquires Ubiquitin (ub) from the E1 Ubiquitin-activating enzyme. The asterisk indicates the site of ubiquitinylation. E2 itself is recruited to the E3 complex by Cullin 1. Skp1 links Cullin 1, via its cullin-binding domain (Cul1), to the F-box domain in the F-box protein (FBP). The latter confers substrate (S) specificity to the complex. Only phosphorylated (p) substrates are recognised. The proteasome degrades the poly-ubiquitinylated substrate. Ubiquitin is recycled.
b. The same colour depicts equivalent proteins in each of the E3 Ubiquitin ligase complexes; grey symbolises the proteins linking the substrate (blue) to the complex, yellow the E2s, red the RING finger-containing proteins, purple Skp1 and its equivalents, green Cullin 1 and its equivalents, and cyan accessory proteins. See text for details. The asterisks indicate the site of ubiquitinylation. The HECT E3s contain a HECT domain protein instead of a RING finger protein. In RING E3s, the RING finger protein confers substrate specificity to the complex. A representative substrate is given for each complex. EGFR, epidermal growth factor receptor and HIF-1α, hypoxia-inducible factor 1α.
c. *Left* Structural model for Skp1/Cul1/FBP hFBXW1a ($SCF^{hFBXW1a}$) composed by superimposing the Skp1 – F-box portions of the Skp1 – hFBXW1a – β-catenin [641], Cul1 – Roc1 – Skp1 – hFBXL1 F-box [642] and Skp1 – hFBXL1 [643] crystal structures. Each protein is named in the same colour as its structure. Space-filling depiction shows a short β-catenin substrate peptide that gets ubiquitinylated by the complex *in vitro*. The illustrated model represents only one possibility for an SCF complex, as several biochemical observations cannot yet be explained at the structural level [644]. For example, in this structural model, Roc1 does not contact the hFBXW1a FBP, which contradicts biochemical data showing interaction between Roc1 and FBPs (panels a and b). In addition, not all proteins in the model were crystallised in their entirety. *Right* Close-up of the region encircled in the left panel for the hFBXW1a F-box – Skp1 (top) and hFBXL1 F-box – Skp1 (bottom) interfaces. Leucine residues corresponding to a position in the Fbxo25 protein that has been investigated in this study contact Skp1 I127. Distances between the δ1- and δ2-methyl groups of the leucine residues and the δ-methyl group of Skp1 I127 (yellow dashed lines) are 2.25 and 2.46 Å for hFBXW1a and 2.25, and 2.65 Å for hFBXL1. Small differences in the Skp1 representation likely derive from differences in the resolution at which the crystal structures were determined.

Image adapted from *644* (c, left). Original X-ray coordinates 1P22 from *641* (c, right, top) and 1FQV from *643* (c, right, bottom) were rendered using PyMOL.


a
F-box protein
F-box
Skp1
Cul1
Cullin 1
Roc1
E2
E1
ub
PROTEASOME
b
HECT E3
e.g. p53
HECT
RING E3
e.g. EGFR
RING
VBC E3
e.g. HIF-1α
SOCS-box protein
Elongin B
Elongin C
Cullin2
SCF E3
e.g. β-Catenin
APC E3
e.g. Cyclin B
Cdc20
Apc11
Apc1
Apc2
Apc4
Apc5
Apc7
Apc3
Apc8
Apc10
Apc6


hFBXW1a F-box
hFBXW1a
Skp1
I127
L164
hFBXL1 F-box
hFBXL1
L124
45°
5°
F-box
hFBXW1A
β-Catenin
Cul1
E2
Ub
Roc1
c

- The crystal structure of an hSkp1 – hFBXL1 complex[643] sheds more light on the hSkp1 – hFBXL1 interaction: AA 109 – 151 of hFBXL1, comprising the F-box, are engaged with the C-terminal region of Skp1 in a superhelical arrangement. Mutations of the amino acids that are involved in this interaction disrupt the binding, indicating that this core interface is required for all Skp1 – F-box interactions.

Whereas the FBP – Skp1 interaction involves the F-box, binding of the substrate with the FBP is mediated by the divergent protein – protein interaction motifs present in FBPs, allowing the bipartite FBPs to confer substrate specificity to the SCF complex[637]. The occurrence of several other domains apart from the F-box explains the overall dissimilarity between F-box family members. FBPs are divided into three classes according to their additional motifs[645]:

- FBXWs contain WD-40 repeats[646]
- FBXLs contain leucine-rich repeats[647]
- FBXOs contain other or no additional motifs

Numerous studies indicate that target phosphorylation is a prerequisite for recognition by the SCF complex[648], which is remarkable, as the substrate binding domains of the different FBPs are very variable; for example, Cdc4 and hFBXW1a are FBXWs[609,630], whereas Grr1 and Fbxl1 are FBXLs[610,639]. Some of the phosphorylated targets tagged by SCF complexes for proteasomal degradation include the CDKIs p27[649], Sic1[650] and Far1[651], the replication activator Cdc6[652], the TF Gcn4[653], the G1 cyclins Cln1 and Cln2[609], and IκB[630] and β-catenin[654]. However, phosphorylation has also been shown to protect proteins from degradation by the Ubiquitin system[655].

A.4.6.5. The Ubiquitin – proteasome pathway in human (neurological) disease

Several neurodegenerative diseases and cancers are associated with mutations in members of the Ubiquitin degradation pathway.

Loss of function mutations in the UPP lead to the stabilisation of proteins which may, for example, contribute to the existence of cellular inclusions. A host of pathological inclusions, such as the neurofibrillary tangles in AD, Lewy bodies in idiopathic PD, Bunina bodies in amyotrophic lateral sclerosis and nuclear inclusions in CAG repeat expansion disorders such as HD, the spinocerebellar ataxias and spinobulbar muscular atrophy, consist of disease-specific proteins, and are associated with accumulated Ubiquitin and its conjugates[656]. The role of the UPP in these neurodegenerative diseases and its more general role in synaptic function are specified under IV.E.1 – E.2.

Apart from neurodegenerative disorders, members of the Ubiquitin degradation pathway have also been reported to be involved in other human pathologies, such as cancer (e.g. uterine cervical carcinoma[657], colorectal, prostate and breast tumours[658]) and pseudoaldosteronism (Liddle syndrome, OMIM 177200)[659].

B. Identification and confirmation of candidate disease genes

Different approaches allow the initial identification of genes putatively involved in human disease. As a second step, such candidate disease genes need to be confirmed.

B.1. Identification methodology for Mendelian disease genes

Since Ingram's work in the 1950s, which marked the era of modern human molecular genetics by establishing the molecular cause of sickle cell anemia (see Appendix C)[660], elucidation of congenital disorders has been undertaken in several different ways. The most common approaches include functional cloning, (positional) candidate gene approaches and positional cloning. Although these methodologies are inapproriate to elucidate multifactorial forms of MR, they have proven extremely valuable in uncovering gene defects underlying monogenic MR.

B.1.1. Functional cloning

Gene identification based on function.

Essentially, functional cloning is an approach in which knowledge of the gene product or gene function is employed to identify the gene. Successful methodologies have included the use of gene-specific oligonucleotides [e.g. haemophilia A (OMIM 306700)[661]], specific antibodies [e.g. phenylketonuria (OMIM 261600)[662]] or functional complementation assays [e.g. Fanconi's anaemia (OMIM 227650)[663]].

B.1.2. Candidate gene approaches

Gene identification through educated guesses based on function and chromosomal location.

Candidate gene approaches are very useful in the identification of disease genes, especially when the candidate gene and the disease gene map to the same subchromosomal location (so-called positional candidate gene approach). As the human genome is now com-

pletely sequenced[19] and annotated with a high degree of reliability, it can be expected that the positional candidate gene approach will be the method of choice to establish new genotype – phenotype correlations (Fig. I-8a). Genes involved in achondroplasia (OMIM 100800)[664] and Marfan syndrome (OMIM 154700)[665] are only two examples of the successful application of positional candidate approaches.

B.1.3. Positional cloning

Positional cloning, gene identification based on its chromosomal location, has frequently been approached through linkage analysis, LOH mapping or study of chromosomal abnormalities (Figs. I-8b – d).

B.1.3.1. Linkage analysis

Analyses of segregation between markers and disease loci.

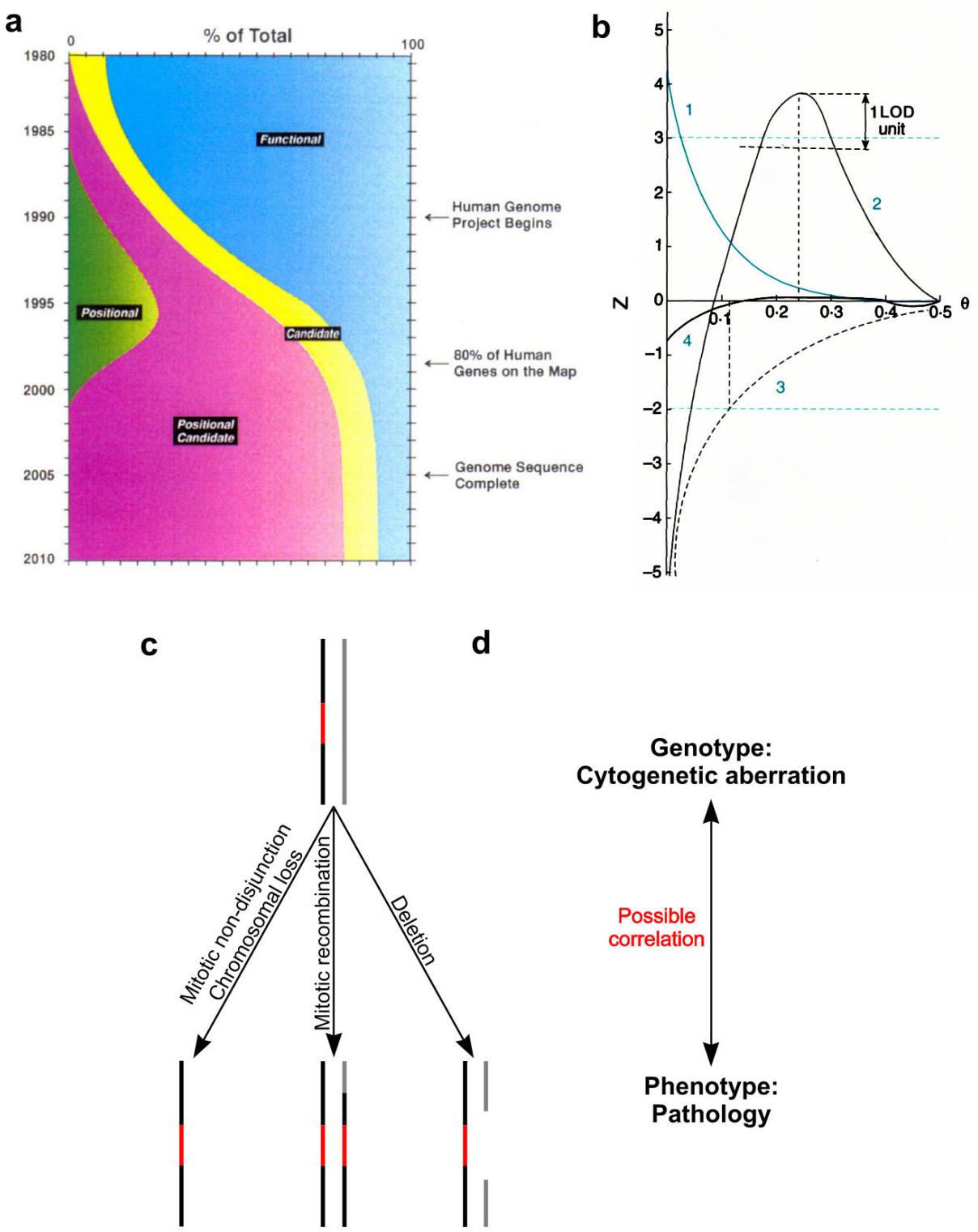
In cases of Mendelian inheritance, linkage analysis is the most widely applied method to localise a disease gene. Essentially, the aim is to identify a fragment of DNA that segregates with the disease. To this end, recombination events between genetic markers, polymorphic sites distributed throughout the genome, are recorded in family studies.

Recombination events, generating genetic diversity, result from crossing-over during meiosis. The recombination frequency θ depends on the distance between markers. There is little chance for recombination to take place if markers are close together and, hence, θ is low

Fig. I-8 | *Next page.* **Overview of some positional cloning approaches.**

a. Predictions made in 1995 suggested that an increase in the importance of positional candidate cloning as a means of identifying human disease genes would roughly be proportional to the amount of available genomic data. Exact quantitation is not intended.
b. Hypothetical logarithm of odds scores (Z) are plotted against the recombination fraction (θ). Plot 1, evidence of linkage without recombinants; Plot 2, evidence supporting linkage in the 0.17 – 0.32 θ interval with the most likely recombination fraction being 0.23; Plot 3, exclusion of linkage for $\theta < 0.12$ and inconclusive for other recombination fractions; Plot 4, inconclusive at all values of θ.
c. Black and grey lines represent part of a homologous pair of chromosomes. Several mechanisms (arrows), including mitotic non-disjunction and subsequent loss of an entire chromosome, mitotic recombination, and deletion (interrupted grey line), lead to loss of heterozygosity (LOH), thereby uncovering a hitherto hidden recessive mutation (red). Note that alternatives exist which also result in LOH, and that other mutational mechanisms, such as point mutations at the wild-type locus, may reveal recessive mutations.
d. Characterisation of the possible correlation between aberrant genotypes and phenotypes is the central idea in positional cloning. In practice, the study of chromosomal rearrangements in patients provides a powerful starting point in the identification of disease genes.

Diagram from *666* (a) and graph adapted from *667* (b).



and the markers are said to be linked. On the other hand, if markers are far apart there are always an equal number of recombinants and non-recombinants observed. This is because (i) there is an equal chance for an odd (recombinants) or even (non-recombinants) number of cross-overs and (ii) the three types of multiple cross-overs, involving two, three or four chromatids, occur with the same likelihood, leading to an overall equal distribution of recom-

binants vs. non-recombinants. Therefore, θ_{max} is 0.5 and the markers are said to be unlinked. θ is used as a measure of genetic distance: two markers are 1 cM apart if the recombination frequency between those two markers is 1%. It should be noted, however, that genetic distance does not equal physical distance: averaging across the entire genome, 1 cM equals 1.1 Mb. Such a general conversion is nevertheless not very useful, as it has been observed that recombination is non-random[668], and sequence-[669] and sex-specific[670]. The male human genome is estimated at 2545 cM[671] and the female genome at 3515 cM[672], although both males and females have approximately the same amount of DNA. Owing to the obligate cross-over in the pseudoautosomal region during male meiosis, the largest discrepancy between male and female genetic distances is seen for this region; 1 Mb of male DNA equals 19 cM and, in contrast, 1 Mb of female DNA equals just 2.7 cM.

The LOD score Z calculates the likelihood of linkage between markers at a certain recombination frequency compared to the null hypothesis of no linkage ($\theta = 0.5$)[673]. Generally, Z can be calculated as

$$Z(\theta) = \log \frac{L(\theta)}{L(\theta = 0.5)} \qquad (1)$$

L, likelihood.

In its simplest form, (1) becomes

$$Z(\theta) = \log \frac{\theta^R (1-\theta)^{NR}}{0.5^{R+NR}} \qquad (2)$$

R, number of recombinants; **NR**, number of non-recombinants.

However, the function describing the likelihood quickly becomes more complex, as it depends on the number of markers under consideration (two-point vs. multipoint mapping) and on the phase. When the phase for a particular individual is unknown – that is, if it is impossible to distinguish between paternal or maternal inheritance of a certain marker – it is also impossible to count the exact number of recombinants and non-recombinants. In such cases, the likelihood function would have to contain a term reflecting the situation for paternal inheritance and a term reflecting maternal inheritance. It is clear that a computer is needed when several individuals are phase-unknown and many markers are involved.

The threshold for accepting linkage, with a 5% error margin, is $Z = 3$. Linkage is excluded when $Z < -2$ (Fig. I-8b). These numbers take into account the prior probability of approximately 1 in 50 for autosomal linkage between two markers[673].

The situation is slightly different for X-linked disorders, as linkage to the X chromosome can often be deduced from the specific pattern of inheritance, implicating a higher prior probability of linkage. Since this prior probability is estimated at approximately 1 in 10, a LOD score of 2.3 suggests linkage with a 5% error margin.

The resolution of linkage analysis, i.e. the final minimal stretch of DNA that should be considered the disease locus, depends on the number of (i) available markers and (ii) recombinants. The latter depends on the number of meioses or, in other words, on family size. With the completion of the Human Genome Project[19], framework maps with over 10000 highly polymorphic microsatellite markers and 1.42 million diallelic SNPs became available[22], and marker density is therefore no longer a limiting factor in the majority of cases. Instead, it is the limited size of human pedigrees that determines the resolution. Fortunately, LOD scores can be added across families, allowing combination of linkage data from several families. However, locus heterogeneity precludes the pooling of linkage data; when mutations in several genes result in similar clinical features, combining linkage data from different families with similar phenotypes may lead to excessively large candidate regions. Conversely, genes with a low rate of mutation may be missed in a genetically heterogeneous group of families due to a bias towards a mutation-prone locus.

Apart from this intrinsic limitation, computational power and the need to specify a detailed genetic model, including the mode of inheritance, allele frequencies and estimates of penetrance, are important constraints that need to be taken into account. Errors introduced through sample mix-up or non-paternity are artefacts that should also be considered.

In 1982, *DMD* was the first gene to be cloned using linkage analysis[674]. Mutations in *DMD* cause Duchenne muscular dystrophy (OMIM 310200).

B.1.3.2. Loss of heterozygosity mapping

Genomic regions characterised by LOH are useful starting points in the identification of tumour suppressor genes.

Retinoblastoma (OMIM 180200) is a rare tumour of the retina with onset during childhood. While in familial retinoblastoma bilateral tumours are common, sporadic tumours are unilateral. This observation led Knudson to postulate his two-hit mechanism for cellular transformation[675]. Essentially, this hypothesis states that two successive allelic and recessive loss-of-function mutations are needed to initiate tumour growth (Fig. I-8c). The class of genes affected according to the two-hit mechanism are known as tumour suppressor genes. LOH at genomic loci has been used extensively to localise such genes by screening

paired blood and tumour samples with genome-wide arrays of genetic markers. Apart from the retinoblastoma gene[676], this approach, among others, has aided in the identification of several tumour suppressor genes, such as the one responsible for colon cancer (OMIM 175100)[677].

With steady advances in increased resolution, array comparative genome hybridisation[678] presumably will supersede LOH studies in the near future.

B.1.3.3. Characterisation of disease-associated chromosomal rearrangements

Disease-associated chromosomal rearrangements, such as deletions, insertions, inversions and reciprocal translocations, are a very informative class of mutations (Fig. I-8d). Patients carrying a cytogenetic abnormality have often been instrumental in positional cloning of disease genes[666].

B.1.3.3.1. Deletions and insertions

Unbalanced rearrangements, which may cause disease through gene disruption or dosage effects.

While deletions are characterised by a loss of genetic material and often lead to loss-of-function phenotypes (e.g. contiguous gene syndromes, Table XI-1), insertions are characterised by a gain of DNA and are less frequently associated with disease. When they are, the phenotypic aberration is mostly caused by gene disruption at the site of insertion (e.g. DNA transposition, Table XI-1) or is dependent on a dosage effect (e.g. gene duplication, Table XI-1). Because of the variation in the amount of genetic material, both deletions and insertions are said to be unbalanced.

B.1.3.3.2. Inversions

Balanced rearrangements, which may cause disease through gene disruption or position effects.

Inversions are generally balanced: there is no gain or loss of genetic material. However, they are sometimes associated with deletions, which may go unnoticed if they are smaller than the resolution of the applied method. Therefore, many inversions may actually be unbalanced at the molecular level.

There is a high frequency of correlation between disease-associated inversions and patient phenotype, although this may at least in part be accounted for by ascertainment bias[679]. When there is a causal link between an inversion and the observed phenotype, the BPs in this class of cytogenetic abnormalities almost always act as pointers to the exact physical location of

the disease gene[680]. Still, cases are described in which BPs alter the expression of genes that are located several kilobases away by affecting the overall chromatin structure or by disrupting distant regulatory sequences[681] (Table XI-1). In case of paracentric and pericentric inversions, two BPs are involved in the rearrangement. When no other positional information is available, both BPs may give rise to the disease phenotype and have to be considered as such.

B.1.3.3.3. Translocations

Translocations can arise during leptotene stage of meiosis. Even though Robertsonian translocations, arising through exchange of the p-arms of non-homologous acrocentric chromosomes, do not cause disease, their carriers produce unbalanced gametes. Exchange of two fragments of non-acrocentric chromosomes, leads to reciprocal translocations. X;A balanced reciprocal translocations disrupting X-chromosomal genes represent human 'KOs', in males and in females.

One source of translocations is the meiotic process, which leads to haploid gametes from a diploid germ cell. During the leptotene stage of meiosis, physiological double strand breaks occur[682]. Their repair is tightly coupled to the homologue pairing process. If the homologue search machinery fails to find the right partner for the broken chromosome, intermingling of non-homologous chromosomes may lead to repair synthesis between them, which, in turn, would result in a translocation event. One proposed stage during meiosis where such an event could take place is the bouquet stage, when all telomeres cluster in a limited region of the nuclear membrane[683]. This stage fulfills the requirements for generation of translocation chromosomes, as has been suggested earlier for mammalian karyotype evolution[684].

The human genome contains five acrocentric chromosomes (chromosomes 13, 14, 15, 21 and 22), which only contain repeated rDNA genes on their p-arms. When parts of the p-arms of two non-homologous acrocentric chromosomes are exchanged, this results in a so-called Robertsonian translocation[685]. While the fused p-arms are lost, the fused q-arms form a stable dicentric chromosome. Although there is a loss of DNA, there is no loss of genetic information, as arrays of rDNA genes are present on the other acrocentric chromosomes. This explains why there are no pathological consequences in carriers of Robertsonian translocations. However, this does not hold true for their offspring; carriers produce unbalanced gametes containing the dicentric chromosome, resulting in monosomic or trisomic zygotes. Although most of these zygotes are embryonic lethal, trisomy 21 zygotes are viable, resulting in Down syndrome (OMIM 190685)[686,687] (Table XI-1), the most frequent genetic cause of MR.

Reciprocal translocations occur when two fragments of non-acrocentric chromosomes are exchanged. Only the reciprocal exchange of two acentric fragments between non-homologous chromosomes leads to a stable translocation; the acentric and dicentric chromosomes that can also form are unstable. Fertilisation of a gamete carrying the translocation with a normal gamete can lead to a partial monosomy or partial trisomy. Whether such a zygote survives to term depends on the nature of the monosomy or trisomy.

Balanced reciprocal translocations are by far the most studied chromosomal abnormalities in the characterisation of human disease genes[666]. Translocation BPs typically lead to loss-of-function mutations that disrupt a gene or its regulatory sequences, leading to congenital disorders[474], whereas in tumourigenesis, they often lead to gain-of-function alterations by creating fusions between genes and unrelated regulatory sequences[688] (Table XI-1). Molecular characterisation of BPs in X;A reciprocal translocations is particularly useful in the identification of genes underlying X-linked disorders. Due to male hemizygosity for all X-chromosomal genes, apart from those in the pseudoautosomal region, and to skewed XCI in female X;A translocation carriers (see I.A.3.2.2.2), disruption of X-chromosomal genes often represents a KO.

B.2. Confirmation of candidate disease genes

Once the cloning strategies described above have uncovered a candidate disease locus, the causal relation between mutation at this locus and disease needs to be confirmed. Although several strategies are possible, proof of causation is usually obtained by screening panels of patients for pathogenic changes at the candidate locus.

B.2.1. Types of pathogenic changes in human disorders

Many different genomic changes, both local and global, can lead to human disease.

The following mutation classes have been implied in human pathogenic conditions:

- Local mutations
  - ✓ Single-base substitutions

    Silent, missense and nonsense mutations
  - ✓ Insertions and deletions
  - ✓ Expansions of triplet repeats
  - ✓ Mutations affecting mRNA and transcription

Mutations (i) of the poly-adenylation signal, (ii) affecting RNA splicing, (iii) in the UTRs and (iv) in regulatory sequences

- ✓ Epigenetic mutations

  Faulty (i) imprinting and (ii) methylation at the promoter

- ✓ Mitochondrial mutations

- Global mutations
  - ✓ DNA transposition
  - ✓ Altered chromosomal environment
  - ✓ Unequal cross-over and unequal sister chromatid exchange
  - ✓ Contiguous gene syndrome

    X-chromosomal and autosomal
  - ✓ Chromosomal aberrations
  - ✓ Numerical chromosomal abnormalities

    Uniparental disomy, aneuploidy and polyploidy

A more comprehensive explanation of these various categories of mutations is given in Appendix D.

B.2.2. Mutation screening

Comparing DNA sequence alterations of patient cohorts with those found in control panels may lead to the identification of pathogenic mutations.

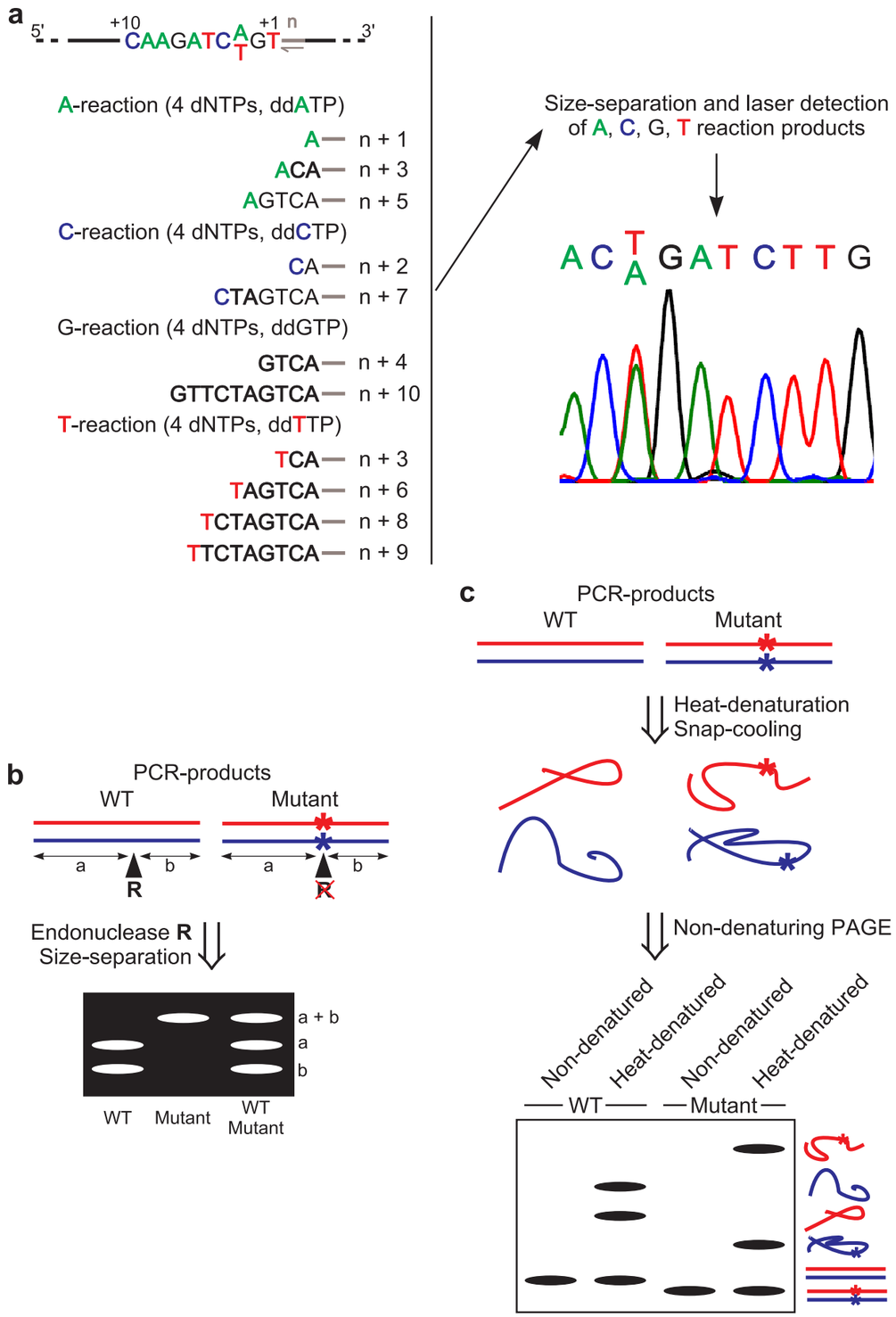
As implied by the name, candidate gene approaches will result in the identification of one or more candidate disease genes. In order to conclusively link a particular disease gene with a human disorder, additional evidence is required. With the invention of PCR[689] and subsequent development of a vast number of variations on the original technique, allowing rapid amplification of stretches of DNA under variable conditions, PCR-based methods dominate current mutation screening. Mutation screening implies the recovery of sequence alterations in the candidate disease gene(s) in a cohort of phenotypically similar patients, which, ideally, are linked to the disease locus.

Throughout evolution, the human genome has aquired sequence variations at a neutral substitution rate of around 2×10^{-9} per site per year, with a 2.8-fold lower rate in coding sequences[690], leading to an average heterozygosity of about 1 out of 1250 bases[691]. Due to this natural variation, it is of paramount importance to verify the absence of the putative pathogenic allele in a panel of healthy, ethnically-matched individuals. When DNA from family

members of the index patient in which the presumptive mutation was found are available, the next step is to show that this allele segregates with the disease in the pedigree. Further functional characterisation of recovered mutations can then conclude the genotype – phenotype correlation study.

Fig. I-9 | *Next page.* **Some methods used in mutation analysis.**

a. Dideoxynucleotide DNA sequencing. A black line depicts a DNA template from 5' to 3'. Single letters highlight positions +1 to +10 with regard to the primer binding site (grey line). The DNA is heterozygous at position +3, which may indicate the presence of a mutation. The primer (grey arrow) has a length of n base pairs. Four separate DNA synthesis reactions, which each use the same primer but randomly incorporate a specific fluorescently-labelled dideoxy-nucleotide that acts as a chain terminator generate products with sizes between n + 1 and n + 10 nucleotides. Electrophoresis on large gels or gel capillaries separates the reaction products according to size. A laser detects their fluorescence sequentially, which results in a string of nucleotides corresponding to positions n + 1, n + 2, n + 3,... of the original template DNA. A sequence chromatogram represents the fluorescence pattern. Note the double signal corresponding to template position +3. Homozygous exchanges only become apparent when compared to a reference sequence.
b. Restriction site polymorphism polymerase chain reaction (PCR). Red (direct strand) and blue (complementary strand) lines depict double-stranded PCR products amplified from a wild-type (WT) and a mutant (asterisk) allele. The mutation abolishes an endonuclease restriction site (**R**). Digestion with an endonuclease, which recognises the **R**-site and subsequent conventional agarose gel electrophoresis reveals the presence or absence of a homozygous or heterozygous sequence alteration.
c. Single-strand conformation polymorphism. Upon heat denaturation, PCR products (shown as in panel b) become single-stranded and, after snap-cooling, each of these strands assumes a complex structure, which is stabilised by weak intramolecular bonds and dictated by its sequence. A fraction of denatured amplicons renatures, which gives rise to the original double-stranded conformation. Different single- and double-stranded DNA species migrate at different speeds in non-denaturing polyacrylamide gel electrophoresis (PAGE). Note that the disparities in migration speeds are exaggerated for illustrative purposes and that these are much smaller in reality, necessitating careful comparison of patient and control samples.
d. Denaturing high-performance liquid chromatography (DHPLC). Upon heat denaturation, PCR products (shown as in panel b) become single-stranded and, after slowly cooling down, they renature, which forms homo- and hetero-duplexes. Partially denaturing conditions on an HPLC column separates such mixtures of duplexes (see text for experimental details). Typically, the retention time on the column is plotted against the absorbance, which results in four distinct peaks when homo- and hetero-duplexes are present in the mixture (yellow curve). As a control, a reference DNA is added to each run (magenta curve).
e. Southern hybridisation. Curved black lines symbolise genomic DNA, **R**s represent endonuclease recognition sites and the grey sections indicate hybridisation sites for a specific probe. This probe consists of a radioactively-labelled (RA), heat-denatured PCR product (grey lines). Note that other probes, such as complementary DNA, can also be employed. The mutant DNA differs from the WT by an insertion of length b (green section). Restriction digestion of genomic DNA with an endonuclease recognising **R**, size-separation and transfer yields a membrane with bound and fragmented DNA, including pieces with sizes of lengths a, b, and a + b that encompass the probe binding site. The pattern of DNA fragments that results from hybridisation of the membrane with the probe and subsequent exposure to X-ray film reveals the presence of WT, homozygous and heterozygous mutant DNA. Southern hybridisation can be employed for detection of all sequence alterations that change the expected size between two restriction sites, when separation of restriction fragments is amenable to conventional agarose gel electrophoresis.


a
5'
+10
CAAGATCAGT
+1
n
3'
A-reaction (4 dNTPs, ddATP)
A — n + 1
ACA — n + 3
AGTCA — n + 5
C-reaction (4 dNTPs, ddCTP)
CA — n + 2
CTAGTCA — n + 7
G-reaction (4 dNTPs, ddGTP)
GTCA — n + 4
GTTCTAGTCA — n + 10
T-reaction (4 dNTPs, ddTTP)
TCA — n + 3
TAGTCA — n + 6
TCTAGTCA — n + 8
TTCTAGTCA — n + 9
Size-separation and laser detection
of A, C, G, T reaction products
A C T G A T C T T G
b
PCR-products
WT
Mutant
a
b
R
Endonuclease R
Size-separation
a + b
a
b
WT
Mutant
WT
Mutant
c
PCR-products
WT
Mutant
Heat-denaturation
Snap-cooling
Non-denaturing PAGE
Non-denatured
Heat-denatured
Non-denatured
Heat-denatured
WT
Mutant
d
PCR-products
WT
Mutant
Heat-denaturation
Slow cooling
DHPLC
Absorbance (mV)
0
1
2
3
4
5
0.0
0.5
1.0
1.5
2.0
Retention time (minutes)
e
WT
R
R
a
Mutant
R
R
b
Endonuclease R
Size-separation
& Blotting
Heat
denaturation
RA-labelled
PCR-product
Hybridisation
Film exposure
a + b
a
WT
Mutant
WT
Mutant

B.2.3. Types of mutation screening

Due to the wide spectrum of methods available for mutation screening, only the techniques employed in this work will be discussed (Fig. I-9).

B.2.3.1. Single-strand conformation polymorphism analysis

Inconclusive method for the identification of localised changes based on conformational alteration of single-stranded DNA.

When double-stranded PCR products amplified from genomic DNA are denatured, they yield single-strand DNA molecules. These molecules adopt a folded conformation dictated by their sequence and stabilised by intrastrand interactions (Fig. I-9c). SSCP analysis is based on this principle[692]. When submitting several samples in parallel to non-denaturing poly-acrylamide gel electrophoresis, a mobility shift in one of the samples indicates a difference in conformation and hence in primary sequence. Direct sequencing then reveals the nature of the sequence alteration in this particular sample.

B.2.3.2. Denaturing high-performance liquid chromatography

Inconclusive method for the identification of localised changes based on thermal denaturation and chemical destabilisation of hetero-duplexes.

In the case of sequence variation within a pooled sample of up to four corresponding PCR products amplified from genomic DNA, a process of denaturing and subsequent reannealing results in the formation of a mixture of homo- and hetero-duplexes. Partial heat denaturation can be used to separate these duplexes; at a particular temperature, hetero-duplexes will already be partially denatured, whereas homo-duplexes are still completely double-stranded. This is the principle used in DHPLC[693,694]. DNA duplexes are adsorbed via their negatively charged phosphate groups to the hydrophobic stationary phase of an ion-pair reversed-phase liquid chromatography column through the ion-pairing reagent TEAA. Because of a reduced negative charge in single-stranded DNA, partially denatured hetero-duplexes elute earlier from the column than their double-stranded counterparts (Fig. I-9d). An additional layer of sensitivity is obtained by applying a step gradient of acetonitril across the separating column, which gradually destabilises the TEAA bridging. The combination of heat denaturation and destabilisation of TEAA bridging results in an earlier elution of hetero-duplexes. Subsequent sequencing of each of the pooled PCR products reveals the nature and the location of the sequence alteration.

B.2.3.3. Restriction site polymorphism polymerase chain reaction

Method based on DNA amplification and subsequent restriction digestion, allowing for quick allele typing.

Amplification across restriction site polymorphisms is a straightforward method to easily type different alleles[695]. For example, PCR products amplified from genomic DNA can be subjected to endonuclease digestion, allowing the different alleles to be inferred from the fragment pattern after size separation (Fig. I-9b).

B.2.3.4. Dideoxynucleotide DNA sequencing

Conclusive method for the identification of localised changes.

Dideoxynucleotide sequencing[696] of PCR products amplified from genomic DNA is the most straightforward and conclusive method of screening for localised mutations (Fig. I-9a). Due to its price tag, DNA sequencing is often only employed to screen a subset of samples that has been pre-selected by less costly methods, such as those described before.

B.2.3.5. Southern hybridisation

Method for investigation of global genomic changes based on hybridising digested, size-separated genomic DNA with labelled DNA probes.

In 1975, Southern developed a method that can be used to screen for large-scale mutations such as repeat expansions or chromosomal rearrangements. So-called Southern hybridisation involves the hybridisation of digested, size-separated genomic DNA with complementary, labelled DNA probes[697]. Variant alleles are recognised as aberrant banding patterns (Fig. I-9e).

C. Objectives

Characterisation of two novel genes disrupted by the BPs of a translocation in a mentally retarded patient may aid in elucidation of a pathway involved in human cognition.

The advent and subsequent success of positional cloning and the recent availability of several near-complete genome sequences have acted as a major impetus on the emergence of molecular pathology. Molecular pathology is characterised by an emphasis on genotype – phenotype correlations, and its goals are the mapping of all (human) genes and the

subsequent molecular characterisation of their products. Ultimately, this will result in an array of pathways connecting all distinct molecular entities in the cell, which, in turn, will lead to the understanding of normal cellular function and human physiology. This knowledge will ultimately aid in the prevention, diagnosis and treatment of human disease.

This thesis deals with the molecular analysis of the BPs of an apparently balanced *de novo* 46,X,t(X;8)(p11.2;p22.3) translocation in a mentally retarded patient. We found that the translocation disrupts two uncharacterised genes. We postulate that the patient's phenotype is correlated with the disruption of one or both of these genes, and test this hypothesis by employing several different approaches to identify and characterise them, their transcripts and the corresponding proteins. These studies fit into the general framework of molecular pathology outlined above and, in addition, may eventually better our understanding of human cognition.