Mutation Research 750 (2012) 52–59

Contents lists available at SciVerse ScienceDirect

Mutation Research/Reviews in Mutation Research

journal homepage: www.elsevier.com/locate/reviewsmr
Community address: www.elsevier.com/locate/mutres

Review

Transient hypermutability, chromothripsis and replication-based mechanisms in the generation of concurrent clustered mutations

Jian-Min Chen [a,b,c,d,*], Claude Férec [a,b,c,d], David N. Cooper [e]

[a] Institut National de la Santé et de la Recherche Médicale (INSERM), U613, Brest, France
[b] Etablissement Français du Sang (EFS) – Bretagne, Brest, France
[c] Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, France
[d] Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Universitaire (CHU), Hôpital Morvan, Brest, France
[e] Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK

ARTICLE INFO

Article history:
Received 5 May 2011
Received in revised form 25 October 2011
Accepted 27 October 2011
Available online 9 November 2011

Keywords:
Chromothripsis
Complex genomic rearrangement
Multiple mutation
Mutational mechanisms
Non-homologous end joining
Transient hypermutability

ABSTRACT

Clustered mutations may be broadly defined as the presence of two or more mutations within a spatially localized genomic region on a single chromosome. Known instances vary in terms of both the number and type of the component mutations, ranging from two closely spaced point mutations to tens or even hundreds of genomic rearrangements. Although clustered mutations can represent the observable net result of independent lesions sequentially acquired over multiple cell cycles, they can also be generated in a simultaneous or quasi-simultaneous manner within a single cell cycle. This review focuses on those mechanisms known to underlie the latter type. Both gene conversion and transient hypermutability are capable of generating closely spaced multiple mutations. However, a recently described phenomenon in human cancer cells, known as 'chromothripsis', has provided convincing evidence that tens to hundreds of genomic rearrangements can sometimes be generated simultaneously via a single catastrophic event. The distinctive genomic features observed in the derivative chromosomes, together with the highly characteristic junction sequences, point to non-homologous end joining (NHEJ) as being the likely underlying mutational mechanism. By contrast, replication-based mechanisms such as microhomology-mediated break-induced replication (MMBIR) which involves serial replication slippage or serial template switching probably account for those complex genomic rearrangements that comprise multiple duplications and/or triplications.

Contents

1. Introduction . . . 52
2. Transient hypermutability . . . 53
3. Chromothripsis . . . 55
4. Replication-based mechanisms . . . 56
5. Conclusions and perspective . . . 58
References . . . 58

1. Introduction

The term clustered mutations refers to the presence of two or more mutations in a spatially localized genomic region within a single chromosome, the precise meaning of the term 'spatially localized' being necessarily context-dependent (see below). Clustered mutations may of course constitute the observable net result of multiple independent lesions sequentially acquired over different cell cycles (Fig. 1a). Alternatively, they can be generated

Abbreviations: CSMMs, closely spaced multiple mutations; FoSTeS, fork stalling and template switching; MMBIR, microhomology-mediated break-induced replication; NHEJ, non-homologous end joining; SRS, serial replication slippage.

* Corresponding author at: INSERM, U613 and EFS – Bretagne, 46 rue Félix Le Dantec, 29218 Brest, France. Tel.: +33 2 9801 8102; fax: +33 2 9843 0555.
E-mail address: jian-min.chen@univ-brest.fr (J.-M. Chen).

1383-5742/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrrev.2011.10.002

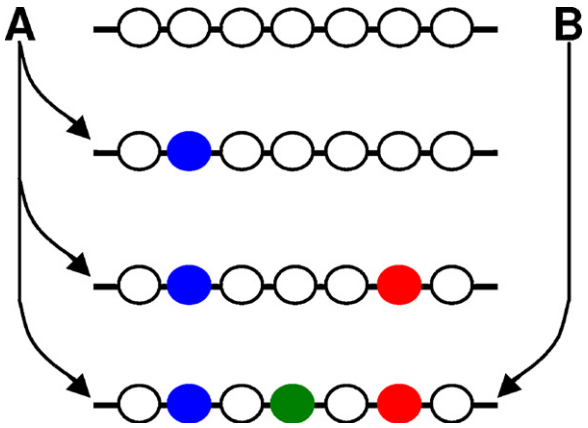

Fig. 1. Alternative models for the generation of clustered mutations. Clustered mutations can be acquired independently over different cell cycles (a) or simultaneously during the same cell cycle (b). Filled color circles indicate individual component mutations.

in a simultaneous or quasi-simultaneous manner within a single cell cycle (Fig. 1b), through mechanisms whose discussion will be the central theme of this article.

Tandem-base mutations or consecutive nucleotide changes (see Fig. 1 in [1]) represent a unique subtype of clustered mutation but will not be discussed here since their underlying mechanism(s) remains largely undefined [1–4]. Gene conversion, a process through which a DNA sequence tract (usually <1 kb in the context of human genes) is replaced by the copy of a paralogous sequence with high sequence similarity (usually >92%), is a well known means by which clustered mutations can be generated, but this process has been extensively reviewed elsewhere [5]. In this article, we therefore focus on transient hypermutability, the newly described phenomenon of 'chromothripsis' and several replication-based mechanisms, all of which are capable of giving rise to contemporaneous spatially clustered mutations in human somatic and germline cells.

2. Transient hypermutability

Data from diverse organisms including viruses, prokaryotes and yeast, as well as cell lines and tissues from higher eukaryotes, have firmly established that the observed frequency of multiple mutation (a category which comprises mainly single point mutations) is significantly greater than that which would be predicted simply from the mutation frequency and a random distribution of mutations [3,6–12]. Of particular note are the results of a study that sequenced the 1.4 kb lacI transgene in thousands of mutants from normal tissue as well as from spontaneous tumours of the Big Blue® transgenic mice; the distribution of the spacing between component mutations in doublets (i.e., two *cis*-linked mutations separated by at least one nucleotide) was highly non-random, with half the mutation doublets being separated by <120 bp [8].

Multiple mutations which have originated cumulatively over a number of different cell cycles exhibit an essentially random intercomponent spacing distribution, as would be expected for mutations of independent origin [3,13]. By contrast, those multiple mutations that exhibit non-random proximal spacing in higher eukaryotes [8] – termed 'closely spaced multiple mutations (CSMMs)' [1] – are most compatible with a model which posits that they have been generated simultaneously (and non-independently as a single event). Multiple simultaneous or synchronous mutations have been postulated to arise as a consequence of transient hypermutability [3,6,7,11].

We recently attempted to extend the concept of transient hypermutability from the soma to the germline, using multiple mutations causing human inherited disease as a model system [1]. Adopting stringent criteria for data inclusion (i.e., the selected examples of multiple mutations *in cis* were neither explicable by gene conversion nor known to result from sequentially acquired events, whilst the component mutations had to be both individually rare and not previously reported in a control population), we retrospectively collated 141 pathogenic multiple mutations bearing at least one single nucleotide substitution: 110 involved only single nucleotide substitution mutations and 31 comprised single nucleotide substitution mutation(s) plus another type of mutation such as small insertions, deletions or indels. 9.2% (13/141) of the multiple mutations comprised three or more components. Notably, the three or more mutational components in up to eight different examples were invariably located within a sequence tract of <100 bp, thereby providing the first convincing evidence that the human germline also experiences transient hypermutability [1].

We then sought to obtain additional evidence to support the postulate that the CSMMs causing human inherited disease arose through transient hypermutability. As illustrated in Fig. 2a–e, all the currently proposed mechanisms underlying transient hypermutability imply new DNA synthesis. Note here that the term 'new DNA synthesis' refers only to the synthesis of the first strand containing the mutation in question. In other words, replication against the first strand, which is necessary to convert the mutation from a unstable potentially transient state (prone to DNA mismatch repair) to a stable state (Watson–Crick base pairing), is not considered. One mutational mechanism that does not involve new DNA synthesis is methylation-mediated deamination of 5-methylcytosine, which gives rise to C > T transition (Fig. 2f). Since 5-methylcytosine in the human genome is largely confined to the CpG dinucleotide, this mechanism accounts for the CpG dinucleotide being a mutation hotspot [14]. We therefore surmised that the proportion of CpG mutations, manifested by the component mutations from a given set of multiple mutations, could be used as a crude indicator of the relative likelihood of transient hypermutability. The logic behind this postulate is that the lower the proportion of CpG substitution, the higher the likelihood that the multiple mutations would have arisen via transient hypermutability. For reasons of simplicity, we focused upon the 102 double mutations that comprised exclusively single nucleotide substitution mutations. Using a cut-off value of ≤100 bp (to denote the separation between the component mutations) to define CSMMs in the human context, we demonstrated that the 58 CSMMs so defined manifested a CpG substitution rate of 10%, significantly lower than that the 45% observed for the remaining 44 double single nucleotide substitution events whose component mutations were separated by >100 bp (χ^2 test, $P < 10^{-7}$) [1]. We also revisited the double single nucleotide substitution mutations detected in the Big Blue® transgenic mice [8,10]. CpG substitution rates were 13% for the 19 events of ≤100 bp and 32% for the 19 events of >100 bp ($P = 0.054$) [1]. The trend appears clear despite the marginal significance which may be related to the small size of the available dataset.

The notion that the two groups of double single nucleotide substitution mutations (i.e., ≤100 bp and >100 bp) had arisen via qualitatively quite different mutational mechanisms received further support from the analysis of the highly informative homocoordinate mutations (multiple mutations in the same gene involving the same mutation type but occurring at different sites in *cis* [8]). Of the 102 double SNS mutations, 17 were found to be homocoordinate mutations; only one of the 6 homocoordinate events in the ≤100 bp group was a CpG mutation whereas 10 of the 11 homocoordinate events in the >100 bp group were CpG mutations [1].