Genome Biology

CrossMark

New insights into the generation and role of de novo mutations in health and disease

Rocio Acuna-Hidalgo[1], Joris A. Veltman[2,3*] and Alexander Hoischen[2]

Abstract

Aside from inheriting half of the genome of each of our parents, we are born with a small number of novel mutations that occurred during gametogenesis and postzygotically. Recent genome and exome sequencing studies of parent–offspring trios have provided the first insights into the number and distribution of these de novo mutations in health and disease, pointing to risk factors that increase their number in the offspring. De novo mutations have been shown to be a major cause of severe early-onset genetic disorders such as intellectual disability, autism spectrum disorder, and other developmental diseases. In fact, the occurrence of novel mutations in each generation explains why these reproductively lethal disorders continue to occur in our population. Recent studies have also shown that de novo mutations are predominantly of paternal origin and that their number increases with advanced paternal age. Here, we review the recent literature on de novo mutations, covering their detection, biological characterization, and medical impact.

Introduction

Upon fertilization, a human zygote inherits half of its genome from the mother via the oocyte and the other half from the father through the sperm. In addition to the genetic information passed on from generation to generation, each of us is born with a small number of novel genetic changes—de novo mutations—that occurred

either during the formation of the gametes or postzygotically [1, 2]. Additionally, novel mutations continue arising throughout post-natal and adult life in both somatic and germ cells. Only mutations present in the germ cells can be transmitted to the next generation [3].

There is a long-standing interest in the study of the frequency and characteristics of de novo mutations in humans, as these are crucial to the evolution of our species and play an important role in disease. A typical human genome varies at 4.1 to 5.0 million positions compared with the human reference genome [4]. The vast majority of genetic variation observed in a typical human genome is common and shared by more than 0.5% of the population as a result of having been recombined, selected, and passed on for many generations [4]. By contrast, a typical human genome contains 40,000 to 200,000 rare variants that are observed in less than 0.5% of the population [4]. All of this genetic variation must have occurred as a de novo germline mutation in an individual at least once in human evolution [5]. Historically, the germline mutation rate in humans has been calculated by analyzing the incidence of genetic disorders; in 1935, Haldane estimated the mutation rate per locus per generation based on the prevalence of hemophilia in the population [6, 7]. More recently, in 2002, Kondrashov accurately calculated the de novo mutation rate in humans by examining the mutation rate at known disease-causing loci [8]. Nowadays, next-generation sequencing (NGS) approaches in parent–offspring trios can be used to directly study the occurrence of all types of de novo mutations throughout the genome, from single-nucleotide variants (SNVs) to small insertions–deletions (indels) and larger structural variations (Box 1). Genome-wide NGS studies place the germline de novo mutation rate for SNVs in humans at 1.0 to 1.8×10^{-8} per nucleotide per generation [1, 9–13], with substantial variation among families [11, 13, 14]. This number translates into 44 to 82 de novo single-nucleotide mutations in the genome of the average individual, with one to two affecting the coding sequence [9, 10, 12, 13, 15]. These state-of-the art genomic

* Correspondence: joris.veltman@radboudumc.nl
[2]Department of Human Genetics, Donders Institute of Neuroscience, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands
[3]Department of Clinical Genetics, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands
Full list of author information is available at the end of the article

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 2 of 19

Box 1 Sequencing technology and de novo mutations
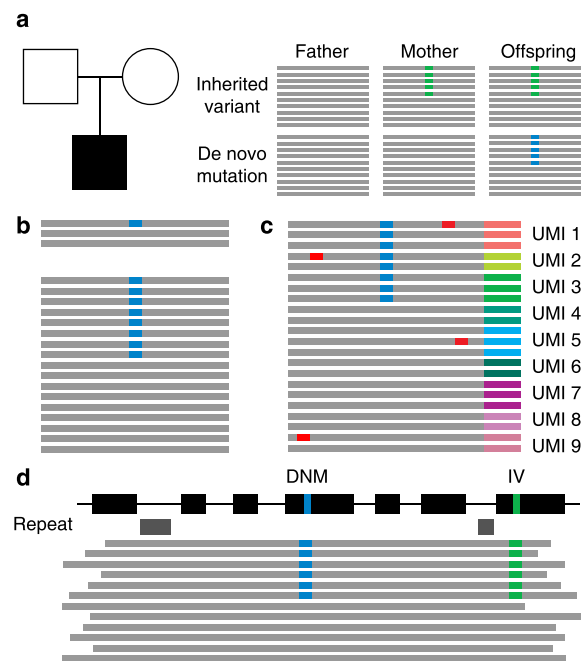
Box Fig. 1 Technical improvements to the detection of de novo mutations (*DNMs*). **a** Trio-based sequencing allows the identification of de novo mutations in an individual. **b** Increased sequencing coverage benefits the detection of de novo mutations (in *blue*). Low coverage (*upper*) reduces the probability that a de novo mutation will be sequenced and called, compared with high sequencing coverage (*lower*). **c** Using random tags or unique molecular identifiers (UMIs) decreases the number of false positives (in *red*) by making consensus calls from all reads with the same UMI. Furthermore, UMIs can be used to remove PCR-derived duplicate reads to determine accurately the allelic ratio. **d** Long sequencing reads improve mappability, even across difficult genomic regions such as those containing repeats (*gray boxes*). Additionally, long reads can be used to phase mutations (shown in *blue* and in *green*) and generate haplotypes, to help identify the parent of origin of a mutation. *IV* inherited variant.

Whole-exome sequencing (WES) and whole-genome sequencing (WGS) provide the possibility to perform an untargeted exome- or genome-wide analysis of an individual's DNA and, in theory, detect all the genetic variation present in an individual. By applying these approaches in parent–offspring trios, one can determine which variants are inherited and which have occurred as de novo mutations in the offspring. The trio design shown in Box Fig. 1a allows investigators to focus directly on the 44 to 82 de novo mutations arising in the human genome per generation. Most current technologies rely on re-sequencing, which is short-read sequencing followed by mapping and comparison to the human reference genome [199], relying on the raw sequencing quality [200] as well as the mapping quality of the NGS reads [201].

The detection of de novo mutations requires high-quality and high-coverage sequencing (Box Fig. 1b) of three samples; de novo mutations need to be detected in the offspring, and the corresponding base pair needs to be called as wild type in both parental samples in a reliable manner. Poor quality and/or low sequencing coverage of any one of the three analyzed samples severely compromises the reliable detection of de novo mutations. Postzygotic de novo mutations represent an additional challenge as mutations are only present in a small percentage of cells and, upon sequencing, can resemble false-positive sequencing artifacts. For this type of mutation, high sequence coverage is even more crucial. Independent validation by Sanger sequencing (only for validation of germline de novo mutations owing to the low sensitivity of the method) and/or targeted NGS approaches remains essential in the case of uncertainty, especially if a de novo mutation might be of diagnostic relevance. Improvements in raw sequencing quality and higher-coverage sequencing will greatly improve the detection of de novo mutations and allow the consistent identification of postzygotic de novo mutations present in small subsets of cells.

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 3 of 19

(See box on previous page.)

A recent improvement for targeted re-sequencing is single-molecule tracing (Box Fig. 1c), which is based on the incorporation of random tags or unique molecular identifiers (UMIs) to each DNA molecule during capture. UMIs can be used to remove duplicates from the sequencing reads, but they can also allow consensus calling of PCR-derived duplicates of the same DNA molecule [202]. This is of particular interest for the study of mosaicism, in which a mutation is present in only a fraction of all analyzed cells, such as postzygotic de novo mutations [88].

In addition, the affordable and widespread use of long sequencing-read technology (Box Fig. 1d) [203] in the coming years is expected to improve greatly the detection and characterization (including precise breakpoint mapping, length measurement, and exact copy-number state) of small indels, repeat-expansions, as well as CNVs and more-complex structural genomic variation such as inversions and chromosomal rearrangements by improving mappability and even allowing for de novo genome assembly [204–206]. For instance, long-read sequencing technologies identify 85% of novel indels and CNVs with an average size close to 500 bp that were missed by other methods [205]. The application of this technology in parent–offspring trios will provide a better insight into the frequency and role of different types of de novo mutations in health and disease. The use of longer sequencing reads is also particularly useful to determine the parental origin of a de novo mutation that requires mutation phasing (that is, deriving haplotypes) by making use of inherited SNVs on the mutant allele. With currently available short-read sequencing technology, phasing can identify the parental origin for ~20% of de novo mutations [13].

approaches allow us to determine additional characteristics of de novo mutations, such as the parental origin and whether they occurred in the germline or postzygotically. We now know that the majority of germline de novo mutations have a paternal origin and that a higher paternal age at conception results in an increase in the number of de novo mutations in the offspring [15–18]. Furthermore, the study of large cohorts of parent–offspring trios provides insight into the distribution of mutations throughout the genome, the genomic context in which they arise, and possible underlying mechanisms [11–13] (see Fig. 1 for an overview of different mechanisms resulting in de novo mutations).

Mutations conferring a phenotypic advantage propagate rapidly through a population [19–21], whereas neutral mutations can disseminate merely as a result of genetic drift [22]. However, damaging mutations resulting in deleterious traits before or during the reproductive phase undergo purifying selection, and their spread through the population is averted [23]. This entails that de novo mutations are genetically distinct from inherited variants, as they represent the result of the mutagenic processes taking place between one generation and the next, before undergoing selection (Table 1). Loss or acquisition of traits at the population level drives evolution of a species, whereas, at the level of an individual, loss or acquisition of traits can result in disease.

Germline de novo genetic alterations have been implicated in human disease for decades. Virtually all disease-causing aneuploidies arise as de novo events. The best known example for this is trisomy 21, identified in 1959 as the cause of Down syndrome [24]. In the beginning of this millennium, genomic microarray technology provided insight into the role of de novo copy-number variations (CNVs) in disease [25]. Even though large CNVs occur at a very low rate, arising at a frequency of only 0.01 to 0.02 events per generation [25–27], they contribute significantly to severe and early-onset neurodevelopmental disorders and congenital malformations owing to their disruptive effect on many genes [28]. The magnitude of the contribution of de novo genetic alterations to human disease, however, has only recently become fully apparent now that NGS approaches allow the reliable and affordable detection of all types of de novo mutations [25]. Damaging de novo point mutations and indels affecting important genes in development have been established as a prominent cause of both rare and common genetic disorders [29–35].

In this review, we first touch on the biological aspects of de novo mutations in humans, such as their origin, distribution throughout the genome, and factors related to their occurrence and timing. Later, we discuss the increasingly recognized role of de novo mutations in human disease and other translational aspects. Throughout, we will focus mostly on de novo SNVs; readers should refer to Box 2 and previous work from others for more information on the role of de novo CNVs and other structural genomic variation in human disease [36, 37].

Causes of de novo mutations

Mistakes during DNA replication can give rise to de novo mutations as a result of the erroneous incorporation of nucleotides by DNA polymerases [38]. DNA polymerases ε and δ catalyze replication predominantly in the leading and lagging strand, respectively. Both polymerases integrate nucleotides during polymerization in a highly selective way, with an average of one mismatch per 10^4–10^5 bp in vitro [39, 40]. A proofreading subunit present in both
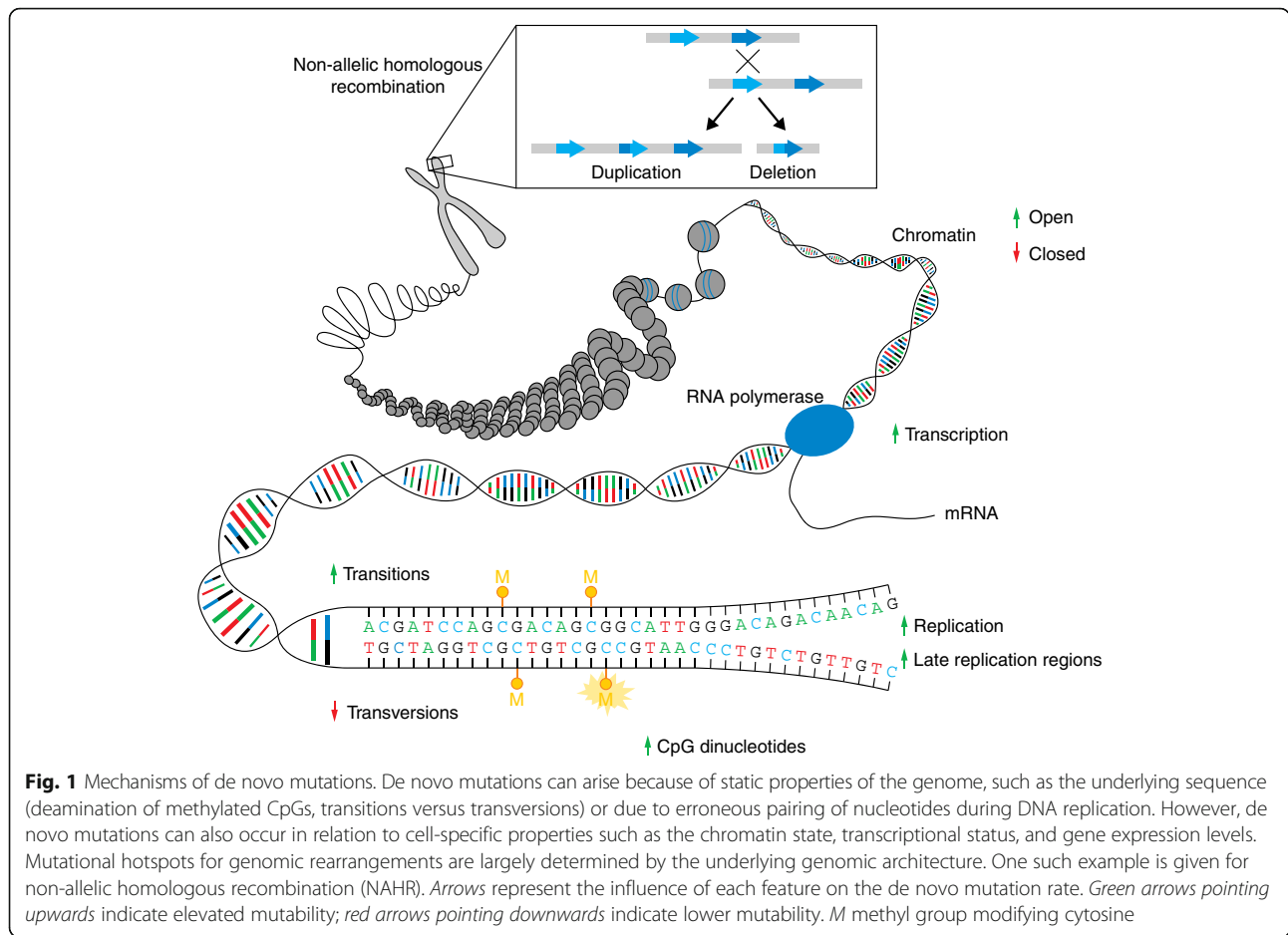
Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 4 of 19

Fig. 1 Mechanisms of de novo mutations. De novo mutations can arise because of static properties of the genome, such as the underlying sequence (deamination of methylated CpGs, transitions versus transversions) or due to erroneous pairing of nucleotides during DNA replication. However, de novo mutations can also occur in relation to cell-specific properties such as the chromatin state, transcriptional status, and gene expression levels. Mutational hotspots for genomic rearrangements are largely determined by the underlying genomic architecture. One such example is given for non-allelic homologous recombination (NAHR). *Arrows* represent the influence of each feature on the de novo mutation rate. *Green arrows pointing upwards* indicate elevated mutability; *red arrows pointing downwards* indicate lower mutability. *M* methyl group modifying cytosine

Table 1 Comparison of inherited and de novo variants

	Inherited variants	De novo mutations
Single-nucleotide variants (SNVs)	3.5 to 4.4 million [4]	44 to 82 [9, 10, 12, 13, 15]
Number of coding SNVs	22,186 [10]	1–2 [25]
Insertions and deletions (indels <50 bp)	~550,000 [4]	2.9–9 [26, 91]
Large indels (50–5000 bp)[a]	~1000 [4]	0.16 [26]
Copy-number variations (CNVs)	~160 [4]	0.0154 [26][b]
Selection pressure in previous generation(s)	High	None
Damaging capacity of variants	Majority with small effect	High
Differences in population	Yes	None
Parental/paternal age effect	None	Strong
Detection of variants	Imputable	Not imputable
Amenable to positional cloning[c]	Yes	No

[a]Owing to technical limitations, the number and mutation rate for large indels ranging between 50 and 5000 bp remain uncertain. Novel sequencing approaches will likely provide more-accurate estimates (see Chaisson et al. [205])
[b]Per generation for CNVs larger than 100 kb
[c]Positional cloning by linkage analysis or homozygosity mapping

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 5 of 19

Box 2 De novo copy number variations and other structural variations

Copy-number variations (CNVs) are defined as deletions or duplications affecting 1000 nucleotides or more [207]. Because of their size, CNVs often have a negative effect on fitness and therefore undergo purifying selection. As a result, there are relatively few inherited CNVs per genome (approximately 160) [4], and de novo germline CNVs are a well-known cause of severe congenital malformations and neurodevelopmental disorders [208–210].

Genomic microarrays have been instrumental for the initial detection and characterization of CNVs with a size below light-microscope resolution [25]. However, NGS has recently shown superior resolution for the detection of CNVs using both short and long sequencing-read approaches and data analysis focusing on depth of sequence reads, split reads, and paired-end reads [211, 212]. The rate at which large de novo CNVs (over 100,000 bp) arise in the human genome is estimated to lie between 0.01 and 0.02 events per generation [25–27]. The mutation rate for indels and CNVs, ranging in size between 10 and 10,000 bp, remains uncertain owing to technical limitations in the detection of these events with short-read sequencing technology.

The CNV mutation rate varies several orders of magnitude depending on the genomic region and parent-of-origin as a result of differences in the mechanism by which the de novo CNV arises [213]. Similar to what has been observed for SNVs, non-recurrent de novo CNVs also show a strong paternal bias and age effect [15, 214]. This correlation highlights a possible mitotic origin for these mutations, resulting from fork stalling and template switching during DNA replication [215]. By contrast, the occurrence of recurrent de novo CNVs, many of which cause well-known developmental syndromes [216], is strongly dependent on the underlying genomic architecture [36]. For instance, the distribution and orientation of segmental duplications (also termed "low-copy repeats") is known to create "hot spots for structural variation" mediating recurrent CNVs by non-allelic homologous recombination during meiosis (NAHR; Fig. 1) [36, 216]. A strong maternal bias for these types of CNVs has been observed at specific loci [217], which might be explained by a higher local maternal recombination rate. Additionally, for a number of recurrent de novo CNVs, it has been shown that the parental allele carries an inversion that places the duplicated flanking regions in tandem. Some of these inversions have reached high frequencies in specific populations; for instance, the inversion leading to the so-called H2 haplotype on 17q21 is present in 20% of Europeans [218] and predisposes to the occurrence of 17q21 microdeletion syndrome [219].

---

polymerases subsequently verifies the geometry of the paired nucleotides to ensure that the incorporated base is correct [38].

Single or multiple base-pair mismatches can cause alterations in the structure of the replicating DNA and can be restored by the mismatch repair (MMR) pathway [41]. The MMR pathway is highly efficient, which explains why the amount of mutations generated during DNA replication is much lower than the polymerase error rate. The frequency at which specific base-pair substitutions arise can be different from the speed at which they are repaired, which defines the mutation rates for specific base-pair substitutions [41]. Incomplete repair can lead to single or multiple base-pair substitutions or indels. Additionally, damaged nucleotides can be incorporated during replication, leading to mispairings and base substitutions [42].

DNA lesions can also appear spontaneously as a consequence of exogenous or endogenous mutagens—UV or ionizing radiation and DNA-reactive chemicals are an example of the former, whereas reactive oxygen species belong to the latter [38]. Before replication, these spontaneous lesions are repaired mainly by the nucleotide excision repair system and base excision repair pathways [43]. However, inefficient repair of pre-mutations before a new round of DNA replication can lead to the mutation becoming permanently fixed in either one or both daughter cells [44]. If mutation repair fails, DNA replication might also be completely arrested and ultimately lead to cell death [44].

The difference between the rate at which pre-mutagenic damage appears in DNA and the rate at which it is repaired defines the rate at which de novo mutations arise. It is often assumed that germline de novo mutations originate from errors in DNA replication during gametogenesis, particularly in sperm cells and their precursors (see section below on parental origin of de novo mutations). However, inefficient repair of spontaneous DNA lesions can also give rise to de novo mutations during spermatogenesis, as continuous proliferation and short periods between cell divisions can translate into there being less time to repair these lesions [44, 45]. Furthermore, in oogenesis, spontaneous DNA mutations coupled to inefficient repair mechanisms might play a more prominent role [44]. Therefore, while the de novo mutation rate is a reflection of the replication error rate and the number of mitoses a cell has undergone, this number is also influenced by the amount of time between mitoses and the efficiency of the DNA repair [44].

Distribution of de novo mutations in the genome

While the typical human mutation rate is $1-1.8 \times 10^{-8}$ per nucleotide per generation [1, 9–13], mutagenesis does

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 6 of 19

not occur completely at random across the genome [9]. Variation in mutability across different areas of the genome can be explained by intrinsic characteristics of the genomic region itself, related to its sequence composition and functional context [46]. Certain factors playing a role in the mutability of the genomic region are predicted to be shared by all cell types in the human organism. These include the local base-pair context, recombination rate, and replication timing [9, 13, 47]. Replication timing refers to the order in which different areas of the genome are replicated during the S-phase of the cell cycle. Genomic regions that are replicated late display more genetic variation than regions that are replicated early [47]. It has been suggested that this could be due to a higher mutability that is secondary to depletion of dNTPs at the end of replication, although other changes such as alterations in polymerase activity and decreased MMR repair activity have also been implicated [38, 48, 49].

Other factors influencing mutability can vary from cell to cell, depending on the transcriptional activity and chromatin state [50–52]. In addition, recent whole-genome sequencing (WGS) studies have revealed the presence of so-called "mutational clusters" and "mutational hotspots". Mutational clusters correspond to the observation of multiple de novo mutations in very close vicinity in a single individual, whereas multiple de novo mutations occurring at the same location in several individuals are an indication of the existence of mutational hotspots [53].

Nucleotide differences: transitions, transversions, and CpGs

The molecular events underlying transitions occur more frequently than those leading to transversions, resulting in a two-fold greater rate of transitions over transversions across the genome [27, 38]. Transitions arise predominantly as a result of C > T mutations, which is at least partially explained by the mutability of CpG dinucleotides [54]. The cytosine in a CpG dinucleotide often undergoes methylation at the fifth position of the six-atom ring, leading to 5-methylcytosine (5-mC). In humans, methylated CpG dinucleotides are known to be chemically unstable and highly mutable due to deamination of 5-mC at CpG dinucleotides, resulting in G:T mismatches [12]. Indeed, the mutability of CpG dinucleotides is approximately ten to eighteen times higher than that of other dinucleotides [27], and, as a result, CpG dinucleotides are found at only a fraction of their expected frequency in the human genome [54]. The high de novo mutation rate at CpG sites is also illustrated by the recent work of the Exome Aggregation Consortium (ExAC). Through the work of this consortium, exome data from more than 60,000 individuals without severe pediatric disease are currently available (Box 3). Analysis of the data in ExAC shows that the discovery of new mutations at CpG dinucleotides reaches

saturation at 20,000 exomes [55, 56]. This emphasizes that identical CpG mutations do not necessarily reflect an ancestral event but are likely the result of independent de novo mutations.

Remarkably, the mutability of CpG dinucleotides is lower in genomic regions enriched for CpG and with higher GC content than in the rest of the genome [44]. In fact, the mutation rate for CpGs in the GC-richest regions of the genome are two to threefold lower than in the rest of the genome [44, 48]. This could be the result of lower methylation levels, the effect of selection because the regions play a role in gene regulation, or secondary to stronger binding between DNA strands impeding separation and spontaneous deamination [38, 44, 57].

Mutational signatures underlying specific mutational processes

While errors in DNA replication, exposure to mutagens, or failure to repair DNA damage can all result in mutations, there are differences in the pattern of mutations arising from each of these processes. A "mutational signature" has been defined as a pattern of mutations that is specific to a mutational process occurring in a cell, tissue, or organism [58]. A recent study based on the analysis of 4.9 million somatic mutations in more than 12,000 cancer genomes defined 21 mutational signatures associated with mutational processes active in somatic cells (termed signature 1 to 21) [58]. Detailed descriptions of each signature are available at http://cancer.sanger.ac.uk/cosmic/signatures. Each of these millions of mutations is placed into one of 96 possible mutation types based on six possible base pair substitutions (C > A, C > G, C > T, T > A, T > C, and T > G) and one of four possible base pairs adjacent to the mutation both at the 5′ and at the 3′ position of the mutation. Concisely, each mutation type is a trinucleotide in which the middle base pair is mutated to a specific nucleotide and each mutational signature is defined by the frequency of each mutation type observed [59].

A recent study showed that the mutational spectrum of germline de novo mutations correlated best with two of these previously described mutational signatures, currently known as signatures 1 and 5 [11, 13]. This suggests that the mutational processes associated with these signatures in somatic cells might also be active in germ cells, although the mechanisms underlying the processes remain elusive. Mutational signature 1 represents close to 25% of de novo germline mutations and is characterized by a high proportion of C > T transitions at CpG dinucleotides, which is associated with deamination of methylated cytosine [11, 58]. Mutational signature 5, which corresponds to the remaining 75% of de novo mutations, is characterized mainly by

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 7 of 19

Box 3 Establishing causality for a de novo mutation

Although identifying de novo mutations is becoming increasingly easy, interpreting them (i.e., linking them to a phenotype) often remains challenging [220]. Clinical interpretation of de novo mutations requires evaluation at the level of the affected locus or gene, as well as at the variant level [221].

For the interpretation of candidate disease-causing de novo mutations, it is important to verify that the phenotype of the patient with the identified mutation matches that of patients described in the literature possessing similar mutations. Next, the de novo mutation can be evaluated by the same methods used to interpret inherited variations, such as in silico prediction programs such as SIFT, PolyPhen, MutationTaster, and CADD [221–224]. Traditionally, evidence linking a gene or a mutation to a phenotype has been established experimentally [221, 223], although functional validation is laborious and the necessary assays can differ per gene and per mutation. Many recent developments can support the interpretation of de novo mutations in human disease. For instance, to study the consequences of a mutation, induced pluripotent stem cells from patient-derived samples can be differentiated into cell types relevant for the respective disease [225]. Furthermore, as a robust method for in vitro and in vivo genetic manipulation, the "clustered regularly interspaced short palindromic repeats" CRISPR–Cas9 system can be used to establish cell and animal models for functional studies [226, 227]. Other CRISPR/Cas9-based methods, such as "saturation genome editing", hold promise for the evaluation of hundreds of mutations in a single assay [228], allowing the interpretation of de novo mutations to keep pace with their discovery in the genomics era.

Replication is essential to establish the link between de novo mutations in a novel disease gene and a phenotype [189]. This initially involves the identification of de novo mutations in the same gene in two or more individuals sharing a similar phenotype. However, large-scale parent–offspring sequencing studies have made apparent that this, by itself, is not sufficient to establish causality for a disease [221]. The number of de novo events identified in a specific gene in individuals with the same phenotype must exceed the expected background rate of de novo mutations, which depends on specific features of each gene, such as its size, sequence, and constraint [126]. This approach has been used successfully to identify new disease genes for autism spectrum disorders [229], epileptic encephalopathies [31], and intellectual disability [128]. A novel way to find more patients with de novo mutations in the same gene is emerging from genetic matchmaking platforms such as Matchmaker exchange (http://www.matchmakerexchange.org/) [230] or GeneMatcher (https://www.gene-matcher.org/) [231], which enable easy data sharing. Establishing unequivocally a link between a genotype and a phenotype requires the same meticulousness in patient phenotyping as in their genotyping, and objective criteria are needed to be able to compare clinical features in patients. For large heterogeneous patient cohorts, systematic phenotyping, including an assessment in Human Phenotype Ontology (HPO) terms, can prove beneficial and increase the diagnostic yield in exome sequencing [127, 232–234].

Large-scale databases of genetic variation can be used to see whether a gene or gene region shows constraint against variation in controls, as the frequency of a mutation in the population is often a good indirect estimation of its pathogenicity [223]. To this end, RVIS and selective constraint scores have become routine in the interpretation of de novo variants, both in research and in the clinic [126, 188]. Population databases, such as the Exome Aggregation Consortium (ExAC) [55], are expected to be depleted of de novo disease-causing mutations for severe and early-onset disorders. Given that de novo mutations are the rarest type of variation, the absence of a mutation from the ExAC database is not in itself sufficient evidence for its pathogenicity. By contrast, the presence of a mutation in ExAC does not automatically entail that the mutation is not disease causing. Pathogenic mutations involved in dominant disease are present in ExAC [55], which might be explained by variable penetrance for these variants [235], the presence of false-positive variants in the control database [236], or undiagnosed disease in control individuals. Possible other explanations for these observations could be the presence of these mutations as somatic events in control individuals [106–108] or resilience to disease in a few selected individuals [237].

A > G transitions [11]. While the mechanism underlying this signature remains unclear, the mutations observed as part of this signature might be secondary to spontaneous deamination of adenine to hypoxanthine, which is then read as guanine [60]. This mutational signature is associated with transcriptional strand bias, suggesting that some of these mutations arise from adducts subject to transcription-coupled repair [60].

Mutational clusters and hotspots

De novo mutations occur throughout the human genome, but occasionally several mutations can arise at a closer distance than expected by random distribution [9]. The term "mutational clusters" refers to the occurrence of de novo mutations in an individual at a closer distance than expected, with multiple de novo mutations within regions ranging from 10 to 100 kb [9, 12, 13, 53]. Mutational

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 8 of 19

clusters display a unique mutational spectrum, with a lower rate of transitions and a large proportion of C > G transversions [13]. This phenomenon has been described to arise in somatic cells in the context of cancer, where it is known as "kataegis", and is linked to the family of enzymes known as APOBEC (for "apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like") [53, 58]. It has been suggested that clusters involving C > G transversions could be related to the formation of single-stranded DNA in diverse cellular processes, such as double-strand breaks and dysfunctional replication forks [61]. Single-stranded DNA might be mistaken for retroelements and attacked by APOBEC enzymes, which convert cytosine to uracil [53]. The mutations are then repaired through base-excision repair and subsequent translesional DNA synthesis with error-prone polymerases [38]. Indeed, mutational clusters have been described to be reminiscent of APOBEC-mediated mutations, albeit with a different sequence context [12, 13]. The occurrence of mutational clusters has been found to correlate with increased parental age [13].

Another origin for some of these clusters could be chromosomal rearrangements. It has been shown that the mutation rate for SNVs is elevated and SNVs can cluster in proximity to the breakpoints of de novo CNVs [62, 63]. This is likely the result of the replicative CNV mechanism in which a low-fidelity, error-prone DNA polymerase is used during repair of DNA. Indeed, work performed in yeast supports the observation that double-strand-break-induced replication is a source of mutation clusters [61].
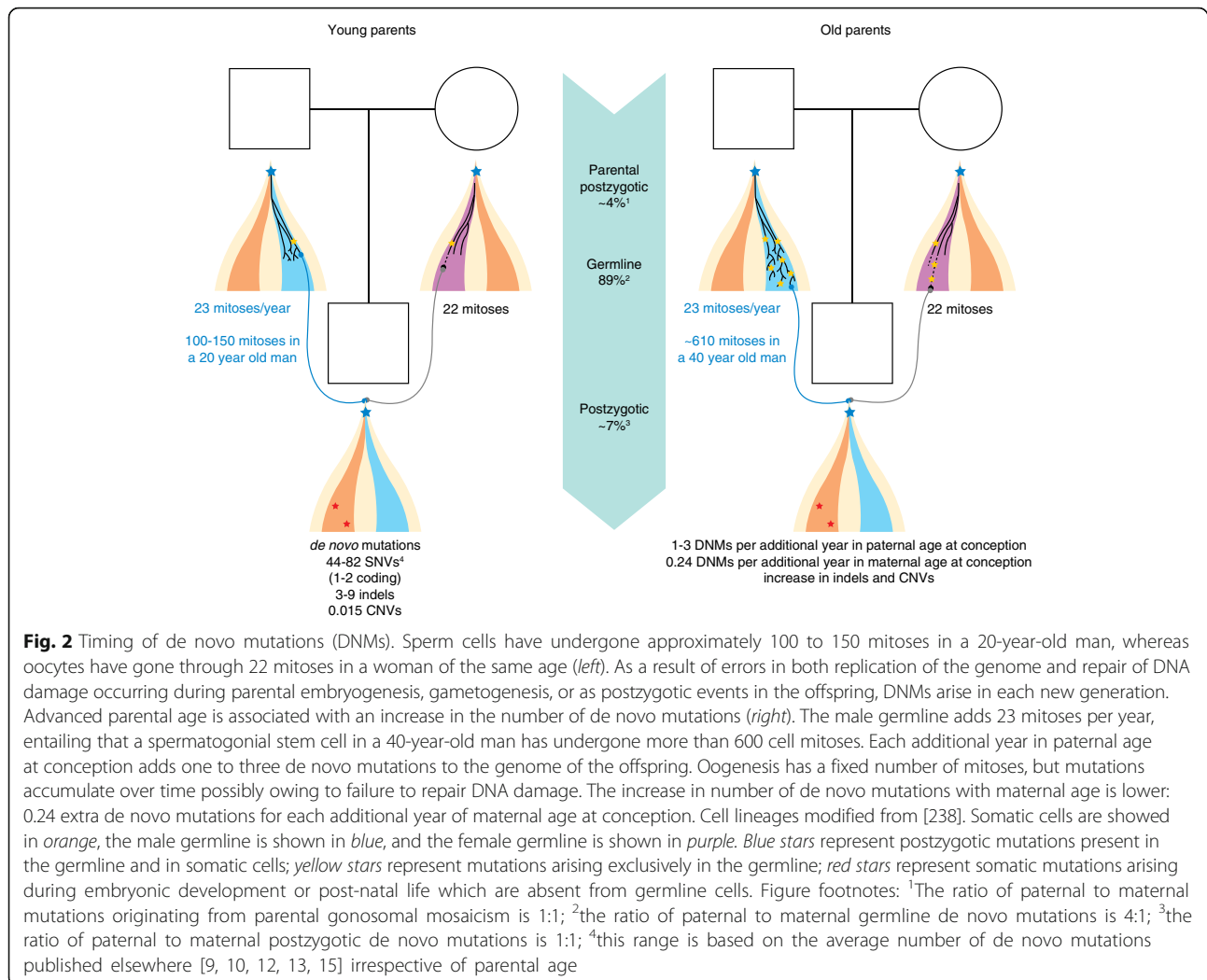
In contrast to the mutation clusters that occur within one individual, mutational hotspots are considered overlapping loci that are found to be mutated more frequently than expected in different individuals. Recent research based on WGS datasets and modeling has identified such hotspots in coding sequences [9]. Furthermore, the existence of these mutational hotspots has been recently confirmed in a larger study that showed specific bins of 1 Mb within the human genome with elevated mutation rates [13]. Interestingly, in this study, two bins including genes *CSMD1* and *WWOX* were shown to have a higher maternal than paternal mutation rate. The mechanism for this is still largely unknown, but the latter is a well-known fragile site within the human genome [64]. Other sites of the human genome that are especially prone to de novo mutations include ribosomal DNA (rDNA) gene clusters [65], segmental duplications [66], and microsatellites [67], with mutation rates three to four orders of magnitude higher than average [68].

Parental origin of de novo germline mutations

In human embryos, the primordial germ cells (PGCs) emerge from the epiblast, eight to fourteen cell divisions after fertilization [69]. In these first cell divisions, the mutation rate appears to be similar in male and female embryos (approximately 0.2–0.6 mutations per haploid genome per cell division, according to models estimating the mutation rate during gametogenesis) [11]. After their specification, PGCs expand to form the pool of spermatogonial stem cells and the complete population of primary oocytes in male and female embryos, respectively [11, 69]. Despite differences in the expansion of PGCs to oogonia or spermatogonia, the mutation rate during this step is similar in both sexes, with approximately 0.5 to 0.7 mutations per haploid genome per cell division, according to computational modeling [11]. However, after puberty, the processes involved in spermatogenesis and oogenesis diverge further. Spermatogonial stem cells divide by mitosis approximately every 16 days, maintaining the spermatogonial stem cell pool while generating differentiated spermatogonial cells which produce sperm cells through an additional round of mitosis followed by meiosis [70]. By contrast, each menstrual cycle, a few oocytes escape from meiotic arrest and complete the first meiotic division. After ovulation, the oocyte becomes arrested once more until fertilization, when it completes the second meiotic division. Thus, after PGC expansion in embryogenesis, oocytes only undergo one additional round of DNA replication in their evolution to a mature ovum. In contrast, spermatogonial cells can undergo hundreds of rounds of DNA replication and cell division before their maturation to sperm cells.

Approximately 80% of all de novo germline point mutations arise on the paternal allele, and advanced paternal age at conception has been established as the major factor linked to the increase in the number of de novo mutations in the offspring, both at the population level and within the same family (Fig. 2) [11, 13, 15]. Spermatogonial cells continue to divide throughout life, which is likely to allow the progressive accumulation of mutations due to errors during DNA replication but also as a result of failure to repair non-replicative DNA damage between cell divisions [44]. Furthermore, the efficiency of endogenous defense systems against radical oxygen species and of DNA repair mechanisms might also decline with age [71, 72]. De novo mutations in children of young fathers show a different signature and localize to later-replicating regions of the genome compared with those of children of old fathers, suggesting that additional factors contribute to de novo mutations with age [12, 13]. It has been calculated that one to three de novo mutations are added to the germline mutational load of the offspring for each paternal year at conception, but this effect varies considerably between families [11, 13]. This variability has been suggested to be due to individual differences in the rate of mutagenesis, in the frequency of spermatogonial stem cell division and even to genetic variation in DNA mismatch repair genes

Fig. 2 Timing of de novo mutations (DNMs). Sperm cells have undergone approximately 100 to 150 mitoses in a 20-year-old man, whereas oocytes have gone through 22 mitoses in a woman of the same age (*left*). As a result of errors in both replication of the genome and repair of DNA damage occurring during parental embryogenesis, gametogenesis, or as postzygotic events in the offspring, DNMs arise in each new generation. Advanced parental age is associated with an increase in the number of de novo mutations (*right*). The male germline adds 23 mitoses per year, entailing that a spermatogonial stem cell in a 40-year-old man has undergone more than 600 cell mitoses. Each additional year in paternal age at conception adds one to three de novo mutations to the genome of the offspring. Oogenesis has a fixed number of mitoses, but mutations accumulate over time possibly owing to failure to repair DNA damage. The increase in number of de novo mutations with maternal age is lower: 0.24 extra de novo mutations for each additional year of maternal age at conception. Cell lineages modified from [238]. Somatic cells are showed in *orange*, the male germline is shown in *blue*, and the female germline is shown in *purple*. *Blue stars* represent postzygotic mutations present in the germline and in somatic cells; *yellow stars* represent mutations arising exclusively in the germline; *red stars* represent somatic mutations arising during embryonic development or post-natal life which are absent from germline cells. Figure footnotes: [1]The ratio of paternal to maternal mutations originating from parental gonosomal mosaicism is 1:1; [2]the ratio of paternal to maternal germline de novo mutations is 4:1; [3]the ratio of paternal to maternal postzygotic de novo mutations is 1:1; [4]this range is based on the average number of de novo mutations published elsewhere [9, 10, 12, 13, 15] irrespective of parental age

[11]. Indeed, one could speculate that deleterious variation in genes involved in replication and repair could predispose to elevated de novo mutation rates not only in somatic cells but also in the germline, as has been observed in mouse models lacking exonuclease activity in DNA polymerase δ [73].

The effect of increased maternal age is well established for errors leading to chromosomal nondisjunction involved in aneuploidies [74, 75], but less so for de novo point mutations. The fixed number of mitoses required for oogenesis would entail that maternal age would not be linked to an increase in DNA-replication-associated mutations. However, an effect of maternal age on the number of de novo mutations has been reported recently [13, 76], likely reflecting an excess of non-replicative DNA damage that is not properly repaired [44]. This maternal age effect was initially reported in a study analyzing de novo mutations in WGS data from a large cohort of parent–offspring trios, in which maternal age correlated with the total number of de novo mutations after correcting for paternal age [76]. A

more detailed analysis of the same cohort confirmed a subtle but significant increase in the number of maternal de novo mutations with advancing maternal age, comprising 0.24 additional de novo mutations per extra year of maternal age at conception [13]. Previous studies had failed to identify a maternal age effect on the number of de novo mutations [12, 15]. This might be explained by differences in the parental age distribution between cohorts or due to a lack of statistical power to detect this subtle effect for which paternal age is a confounder [76]. The increase of de novo mutations with advanced paternal and maternal age support the possibility that the accuracy of DNA repair mechanisms in germ cells decreases with age [72].

Selective advantage of de novo mutations in the testes
A striking increase with paternal age has been observed for a small subset of de novo mutations that are highly recurrent and localize to specific nucleotides in the genome. These de novo mutations are thought to grant spermatogonial stem cells a growth advantage, leading to clonal

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 10 of 19

expansion of mutated cells in the testis [77]. For instance, gain-of-function mutations in genes in the RAS–MAPK pathway have been shown to cause clonal expansion of mutant spermatogonial stem cells owing to proliferative selective advantage [77, 78]. Computational modeling suggests that this would result from a slightly increased ratio of symmetric versus asymmetric divisions in mutant spermatogonial stem cells, favoring the production of two mutated spermatogonial stem cells compared with a single mutated stem cell and one differentiated spermatogonial stem cell harboring the mutation [79, 80]. Therefore, over time, spermatogonial stem cells carrying these mutations undergo positive selection owing to higher self-renewal than surrounding wild-type cells and expand clonally in the testis [81]. The occurrence and enrichment of mutations in spermatogonial stem cells is thought to take place in all men and would entail that the testes of older men contain a higher number of clones of mutant spermatogonial stem cells [77, 78].

Interestingly, the first mutations implicated in clonal expansion in spermatogonial stem cells were initially shown to cause developmental disorders such as Noonan and Costello syndrome (caused by *PTPN11* and *HRAS* mutations, respectively) [78, 81, 82], Apert, Crouzon, and Pfeiffer syndromes (*FGFR2*) [81, 83], achondroplasia, Muenke syndrome and thanatophoric dysplasia (*FGFR3*) [81, 82], and multiple endocrine neoplasia (*RET*) [84]. Mutations that are positively selected at the spermatogonial stem cell level but are detrimental at the organism level have been termed to behave selfishly and are therefore referred to as "selfish mutations" [82]. Owing to the expansion of mutant cells over time, the incidence of these developmental disorders shows an exponential increase with paternal age at conception, well beyond the increase observed for other disorders caused by de novo mutations [85]. Appropriately, these disorders are known as "recurrent, autosomal dominant, male-biased, and paternal" (RAMP) age effect disorders or, simply, paternal age effect (PAE) disorders [45, 78]. Because of the selfish selection of mutant spermatogonial cells, PAE disorders have an incidence up to 1000-fold higher than expected based on the mutational target size and the average mutation rate [45, 85]. It has been hypothesized that "selfish mutations" with a weaker effect on spermatogonial stem cell behavior could be involved in more-common phenotypes, such as intellectual disability, autism, or epilepsy [86]. Furthermore, "selfish" behavior is a characteristic of certain mutations driving cancer as they lead to positive cellular selection despite being harmful for the organism. Predictably, several mutations behaving selfishly in spermatogonial stem cells have also been identified as somatic events driving clonal growth in tumorigenesis [82].

Following the identification of genomic regions enriched for maternal de novo mutations [13], the possibility of selfish mutations in the maternal germ line has also been put forward [72]. It appears that these genomic regions harbor genes with a role in tumor suppression, and some de novo mutations could, it is speculated, provide mutant oocytes in aging women with a survival advantage over wild-type ones [72].

Timing of de novo mutations

De novo mutations have traditionally been considered to occur as germline events, but the advent of NGS allowed scientists to demonstrate that de novo mutations occur as non-germline events more often than previously estimated [3, 87–89]. Mosaicism, which is the existence of two or more genetically distinct cell populations in an individual developing from a single fertilized egg [90], is the norm rather than the exception. Postzygotic mutations, that is, mutations arising in the first few cell divisions after fertilization, can lead to high-level mosaicism and be present in many different tissues of an organism. Mutations that arise later in development or post-natal life, by contrast, can remain restricted to a single tissue or even to a small number of somatic cells (Fig. 2).

Approximately 7% of seemingly de novo mutations are present in blood as high-level mosaic mutations, having likely occurred as early postzygotic events [88, 89, 91]. This, together with the observation that chromosomal instability and structural rearrangements are common in cleavage-stage human embryos, has led to the suggestion that early embryogenesis might be a period of high mutability [92, 93]. Before the initiation of transcription and translation in the zygote, human embryos rely on maternal proteins contributed by the oocyte [94], which could lead to a shortage of proteins involved in DNA replication and repair, resulting in genomic instability [3]. Depending on the timing at which a de novo mutation arises during embryonic development, it could be present at different levels in multiple tissues or be organ specific [95]. A recent study examined multiple samples from the same individual and showed the widespread presence of postzygotic de novo mutations in tissues of different embryonic origin, including somatic and germ cells [96]. Furthermore, mutations can arise in the germ cell lineage after the specification of PGCs during early embryonic development, remaining isolated from somatic cells [3]. Although these mutations are undetectable in sampled tissues such as blood or buccal swabs, they can be transmitted to the offspring as germline events.

Somatic cells are predicted to accumulate hundreds of different mutations throughout post-natal and adult life [97]. Large chromosomal abnormalities have been observed in many tissues in the human body [98], such as the blood, where the presence of these lesions increases with age [99–101]. For instance, loss of the Y chromosome in blood cells has been described as a

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 11 of 19

frequent event in aging males, affecting over 15% of men aged 70 years or older [102, 103]. Somatic mutations resulting in low-level mosaicism are prevalent in healthy tissues [104], including the brain [105], blood [106–108], and skin, where the somatic mutation rate has been calculated at two to six SNVs per megabase of coding sequence per cell [109]. As a result of the accumulation of somatic mutations, the genome sequence is certain to vary among different cells of an individual, a level of genetic diversity that is best observed with single-cell sequencing technologies [110]. Studies in mouse models have shown that the mutation frequency is higher in somatic cells than in germ cells [111, 112]. The comparison of the somatic and germline mutation rate in humans supports this finding, which might stem from differences in the efficiency of DNA replication and repair mechanisms in germ and somatic cells, in addition to differences in exposure to mutagens [72].

De novo mutations in human disease

The medical relevance of de novo mutations has only recently been fully appreciated, mainly because advances in sequencing technology have allowed a comprehensive analysis of these mutations [25]. The field of human genetics had previously focused primarily on inherited diseases, leaving sporadic disorders largely untouched. This was because traditional disease gene identification methods relied mainly on positional mapping of disease loci in large pedigrees with multiple affected members, followed by Sanger sequencing to identify disease-causing mutations in candidate genes. By contrast, NGS techniques such as whole-exome sequencing (WES) or WGS now provide the possibility to detect most, if not all, genetic variation present in a patient. To this end, trio-based WES or WGS has been instrumental in detecting and characterizing de novo mutations in patients with a wide variety of diseases (Box 1) [25, 35].

De novo mutations in pediatric disease

De novo mutations are now well known to play an important role in severe early-onset diseases, which for the most part arise sporadically because of their impact on fitness; owing to the severity of the phenotype in which they often result, an individual with a deleterious de novo mutation will not produce offspring and the phenotype therefore only arises through de novo mutations.

In the first 5 years of widespread availability of WES, more than 500 novel disease–gene associations have been identified, with the strongest increase in sporadic diseases caused by de novo mutations [35, 113, 114]. Recent studies applying exome sequencing in the clinic have shown that of all sporadic cases that received a molecular diagnosis through clinical exome sequencing,

between 60 and 75% could be explained by de novo mutations [115, 116]. De novo mutations affecting the coding region have also been established as an important cause of common neurodevelopmental disorders, such as autism [29, 30], epilepsy [31], and intellectual disability [33, 34], which affect over 1% of the population [117, 118]. Clearly, these common genetic disorders are not explained by de novo mutations affecting the same locus in every patient. Instead, an extreme genetic heterogeneity is observed, and patients with common genetic disorders carry de novo mutations in many different genes. The population frequency of a disorder caused by de novo mutations is determined in large part by the number of genes or genetic loci that can result in this disorder when mutated, which we have referred to previously as the "mutational target" [25]. Rare disorders are most often caused by mutations in a single gene or a small number of genes, while common genetic disorders usually have a large mutational target, often comprising hundreds to thousands of genes or genetic loci. [25]. As an example, more than 700 genes have now been identified to cause autosomal dominant intellectual disability when mutated [117], and this number is rapidly increasing since the widespread application of NGS technology. Based on these sequencing studies, it appears that the majority of the most severe neurodevelopmental phenotypes, such as severe intellectual disability with an IQ below 50, are the consequence of damaging de novo germline mutations in the coding region [10]. An enrichment for damaging de novo mutations has also been observed in individuals with milder phenotypes such as autism spectrum disorder without cognitive deficits [16, 18, 29, 30, 119]. For these milder phenotypes that have less impact on fitness, the exact contribution of de novo mutations to the disease burden is not yet firmly established, and inherited variation is likely to be at least as important in the expression of the phenotype [120–122]. Next to neurodevelopmental disorders, de novo mutations also play a prominent role in pediatric diseases such as congenital heart defects (CHDs) [123–125]. In agreement with the observation made in neurodevelopmental disorders, recent studies found the highest contribution of de novo mutations to disease in individuals with the most severe and syndromic forms of CHD [123, 125]. Finally, it is essential in large-scale sequencing studies to test formally whether the recurrence of de novo mutations in a gene exceeds the number of observations expected by chance (Box 3) [126].

The vast majority of pathogenic de novo mutations are involved in dominant genetic disorders. This appears logical, as a single damaging de novo mutation can be sufficient to cause these kinds of disorders. However, there are examples of recessive disorders that can be caused by the combination of an inherited mutation on

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 12 of 19

one allele and the occurrence of a de novo mutation on the other [33]. In a cohort of 100 trios with severe ID, we identified one case of autosomal recessive ID that was due to the inheritance of one pathogenic allele and the occurrence of a de novo hit in the other [33], and similar observations in the context of late-onset disease are described below. Furthermore, there are reports of cases with a merged phenotype comprising two clinically distinct disorders of which either one or both are caused by a pathogenic de novo mutation [115]. Phenotype-based and classic genetic approaches are insufficient to diagnose individuals with this kind of combined disease, illustrating the power of an unbiased genotype-first approach. Additionally, this approach reduces the need for clinical homogeneity for disease–gene identification studies, as was required for phenotype-first approaches [127, 128].

De novo mutations in late-onset disorders

Few studies until now have addressed the role of de novo mutations in late-onset diseases. The role of de novo mutations is likely to be smaller in late-onset disorders than in pediatric disorders given the effect of de novo mutations on reproductive fitness. Nevertheless, genes involved in adult-onset disorders are just as likely to be affected by de novo mutations as genes involved in pediatric disorders. A complicating factor in these late-onset disorders, however, is the collection of parental samples for the study of de novo mutations [129]. Despite this obstacle, recent publications have suggested a link between de novo mutations and late-onset neurological and psychiatric disorders: Parkinson's disease, amyotrophic lateral sclerosis, schizophrenia, and bipolar disorder have been associated with de novo SNVs and CNVs [130–137]. For example, one study found that 10% of individuals with sporadic schizophrenia have a rare de novo CNV compared with 1.26% for controls [132]. Exome sequencing of a cohort of 623 schizophrenia trios identified an enrichment for de novo point mutations in genes encoding synaptic proteins in cases compared with controls [130]. A large meta-analysis recently identified both an excess of loss-of-function mutations in the histone methyltransferase *SETD1A* and an excess of de novo occurrence of these mutations in individuals with schizophrenia compared with controls [138]. Recent studies have exposed a genetic overlap between neurodevelopmental disorders and schizophrenia, with de novo mutations in the same gene being involved in both early and late-onset disorders [138–140]. While de novo mutations have been firmly linked to neurodevelopmental disorders, their involvement in late-onset psychiatric phenotypes is more controversial. This could be the result of a more complex underlying genetic architecture [141], together with a more prominent role for environmental factors in the expression of the phenotype [142].

Cancer, particularly in relatively young individuals without relevant family history, has been associated with de novo mutations in genes involved in cancer-predisposition syndromes. For example, at least 7% of germline mutations in *TP53* (encoding cellular tumor antigen p53) in individuals with Li-Fraumeni syndrome occurred de novo [143], and a similar proportion has been identified for mutations in *APC* involved in familial adenomatous polyposis [144]. Nevertheless, the rate of de novo mutations in genes involved in other cancer-predisposition syndromes, such as *BRCA1* and *BRCA2* [145], or in DNA mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) [146] has been reported to be much lower.

Interestingly, de novo mutations have also been identified as causative mutations in genetic disorders that are typically inherited, such as hereditary blindness. For instance, the rate of causative de novo mutations among sporadic cases within a cohort of patients with retinitis pigmentosa was close to 10% [147], a result that was later confirmed by an independent study [148]. Although for the majority of this group the de novo mutation represented a single dominant hit causative of the phenotype, in one case the de novo mutation was in fact the second hit in an autosomal recessive form of retinitis pigmentosa. Similarly, in a cohort suffering from mild-to-moderate sensorineural hearing loss, de novo mutations were identified in two out of eleven sporadic cases [149], also suggesting a role for de novo mutations in this heterogeneous disorder.

As de novo mutations are known to play an important role in disorders that affect fitness, it might also be very relevant to investigate their role in disorders linked to fertility, such as male infertility. Both de novo chromosome Y deletions as well as de novo point mutations in a few genes have been found to cause this disorder [150, 151], but a systematic screen is lacking so far.

Postzygotic de novo mutations in disease

The timing of a pathogenic de novo mutation can have an important influence on the expression of the phenotype. Postzygotic mutations are currently receiving more and more attention as technological improvements allow the detection of (low level) mosaic mutations for the first time at a genome-wide scale (Box 1). Postzygotic de novo mutations have been identified as the cause of several human diseases, ranging from developmental disorders [152–154] to cancer [155–157]. While de novo mutations arising later in development and leading to gonadal or gonosomal mosaicism might be clinically silent in that individual,

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 13 of 19

there is an increased likelihood that the mutation is transmitted to the offspring as a germline event, resulting in a clinical disorder [158].

Regardless of whether they occur in the germline or postzygotically, some de novo mutations lead to a single Mendelian phenotype in which the mosaic and constitutive form are part of the same clinical spectrum [159]. For example, pathogenic mutations in genes involved in epileptic encephalopathies [160] and cerebral cortical malformations [161] have been shown to cause similar phenotypes when they arise either in the germline or as postzygotic de novo mutations leading to mosaicism in the brain. However, in some of these cases, mosaicism might cause a clinical phenotype milder than a constitutive mutation [162, 163].

De novo mutations can also result in different phenotypes when they are present in the germline or arise postzygotically [164]. Some de novo mutations lead to developmental disorders only if the de novo mutation occurs postzygotically, as the constitutive presence of the mutation is suspected to be lethal [165, 166]. Examples of this include Proteus syndrome (caused by *AKT1* mutations) [152], Sturge-Weber syndrome (*GNAQ*) [153], and CLOVES syndrome (*PIK3CA*) [167]. A common feature to these disorders is that they are caused by mutations known to lead to activation of cellular proliferation pathways and overgrowth. The mutations with the strongest effect generally result in more-severe developmental alterations [168], suggesting that the type of de novo mutation influences the expression of the phenotype. Remarkably, the mutations with the strongest effect on activation have also been observed as somatic events in cancer [168], for which constitutive activation of cellular proliferation pathways is a major hallmark [169]. This finding supports the view that not only the type of pathogenic mutation but also the time at which the mutation occurs is crucial in defining its consequences.

The timing of a postzygotic mutation determines the percentage of affected cells in the organism and the type of tissues involved [90, 153]. For instance, the same genetic alteration in genes in the RAS–MAPK pathway can result in very diverse phenotypes, depending on the timing at which they arise [164, 170, 171]. Mutations in *HRAS* mutating codon G12 of the HRAS protein have been identified in Costello syndrome when present in the germline [172], but postzygotic and embryonic occurrences of mutations in this residue have been observed in Schimmelpenning syndrome [164], sebaceous nevus [164], keratinocytic epidermal nevi [173], and early-onset bladder cancer [157, 174]. Furthermore, identical mutations in the phosphoinositide-3-kinase *PIK3CA* can cause different phenotypes, ranging from different overgrowth syndromes [154] to lymphatic [175] and venous malformations [176],

depending on the tissue distribution. Therefore, the timing of a pathogenic de novo mutation is likely instrumental in defining its phenotypic consequences as it determines the burden placed by the mutation upon the organism, including the type of tissues affected and the percentage of cells in which the mutation is present [90, 153].

Finally, an important characteristic of postzygotic mutations is that they generate genetically distinct populations of cells that coevolve within a single organism. This can lead to competition between populations of cells [177] or generate interference in signal transduction between cells [178, 179]. For example, craniofrontonasal syndrome is an X-linked disorder in which women with germline mutations and men with postzygotic mutations have a more severe phenotype than men with germline mutations, owing to interference in cell signaling between different cell populations [179].

Postzygotic de novo mutations have been implicated in early-onset cancer [155, 157] and could well represent an early mutational event in the development of cancer in the general population [156]. Additionally, the high degree of mosaicism observed in a normal human brain has led to the suggestion that pathogenic postzygotic and somatic mutations could be at the source of psychiatric disorders [180, 181]. The role of mosaic de novo mutations is not yet fully appreciated, and it is to be expected that our understanding of this class of mutations will increase rapidly in the coming years because of further technological improvements as well as access to DNA from other (affected) tissues or even cell-free DNA (cfDNA) as a source of DNA from multiple tissues [182–184].

De novo mutations in clinical practice

The recent recognition of the importance of de novo mutations in human disease has many implications for routine genetic testing and clinical practice. De novo mutations are now established as the cause of disease in a large fraction of patients with severe early-onset disorders, ranging from rare congenital malformation syndromes [185, 186] to more-common neurodevelopmental disorders, such as severe forms of intellectual disability [33], epilepsy [31], and autism [29]. Together, these disorders represent a substantial proportion of all patients seen at neuropediatric and clinical genetics departments around the world.

Pinpointing the genetic cause of a disorder caused by a de novo mutation in an individual can be challenging from the clinical point of view because of pleiotropy as well as genetic heterogeneity underlying a single phenotype. For instance, intellectual disability can be caused by de novo point mutations, indels, or CNVs in any of hundreds of genes [117]. This obstacle to providing a clinical diagnosis strongly argues for a reliable and

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 14 of 19

affordable genomics approach that can be used to detect these de novo mutations in large groups of patients. Exome and genome sequencing (which additionally offers the possibility of accurate detection of structural variation) of patient–parent trios is ideal for this and will soon become the first-tier diagnostic approach for these disorders. A key advantage of this trio-based sequencing approach is that it helps prioritize candidates by de novo occurrence, allowing clinical laboratories to focus on the most likely candidate mutations for follow-up and interpretation (Box 3) [187]. The interpretation of candidate de novo mutations can be guided by the use of different scores, such as the "residual variation intolerance score" (RVIS), based on the comparison of rare versus common missense human variation per gene [188]. Alternatively, "selective constraint scores" can be used, based on the observed versus expected rare functional variation per gene within humans [126].

The identification of a de novo mutation as the cause of disease in a patient has several implications for the patient and his or her family. First, the detection of the genetic defect underlying the phenotype establishes a genetic diagnosis that can be used to provide a prognosis based on data from other patients with similar mutations [189] and information about current treatment options [190] and, in the future, for the development and application of personalized therapeutic interventions [191]. Furthermore, the identification of a de novo mutation offers the parents of the affected patient an explanation as to why the disorder occurred and might help deal with feelings of guilt [192, 193]. In terms of family planning, the identification of a de novo mutation as the cause of disease in a child can be positive news with regard to recurrence risk, as it is much lower than for recessive or dominant inherited disorders (slightly above 1% versus 25 and 50%, respectively) [11, 158]. However, the recurrence risk is strongly dependent on the timing of the mutation as parental mosaicism for the mutation increases the risk of recurrence [158]. Approximately 4% of seemingly de novo mutations originate from parental mosaicism detectable in blood [11], and recent work suggests that transmission of parental mosaicism could explain up to 10% of de novo mutations in autism spectrum disorder [194]. This entails that a fraction of de novo mutations have an estimated recurrence risk above 5% [158]. Furthermore, close to 7% of seemingly de novo mutations arise as postzygotic events in the offspring [88, 89, 91]. Parents of an individual with a postzygotic mutation have a low risk for recurrence of the mutation in an additional child, estimated as being the same as the population risk [90]. Targeted deep sequencing of a disease-causing mutation can be performed to test for its presence in parental blood and detect mosaicism in the offspring. Although it is not yet offered on a routine basis, this kind of testing can provide a personalized and stratified estimate of the recurrence risk based on the presence or absence of mosaicism in the parents or in the offspring.

Finally, it is impossible to prevent de novo mutations from arising in the germline of each new generation, but attention must be brought to the factors that increase the number of de novo mutations in the offspring. The single most important risk factor is advanced paternal age at conception [15], which is of great importance from an epidemiological perspective since most couples in Western countries are having children at later ages. In fact, this increase in de novo mutations with paternal age at conception might explain epidemiological studies that link increased paternal age to increased risk of neurodevelopmental disorders in offspring [195]. A recent population-genetic modeling study, however, indicated that de novo mutations might not explain much of the increased risk of psychiatric disorders in children born to older fathers [122]. While this might be the case for relatively mild and later-onset phenotypes such as schizophrenia, de novo mutations are responsible for the majority of the most severe pediatric disorders arising in outbred populations [10, 196]. At present, most attention, advice, and guidelines are focused on advanced maternal age as a public health issue. It is evident from current work on de novo mutations that advising the public, including policy makers, on potential risks of advanced paternal age and the burden it might bring on society is crucial. An extreme "solution" if reproduction is to be postponed might be to promote cryopreservation of oocytes and sperm [197], a measure under much debate that has been termed "social freezing".

Conclusions and future directions

Advances in sequencing technologies have provided us with the ability to identify systematically most if not all de novo mutations in a genome. This has boosted fundamental research into the evolution of our genome by providing insight into the mechanisms that play a role in mutagenesis, the origins of these mutations, and their distribution throughout the genome. While most of this research has been focused on germline mutations, we now see a shift towards the detection and study of somatic de novo mutations also for non-cancer phenotypes, greatly facilitated by more accurate and deeper-coverage sequencing technologies. Next-generation sequencing has also boosted research and diagnostics on sporadic diseases. The routine detection of de novo mutations by trio-based sequencing of patients and their unaffected parents in research as well as in diagnostics will soon allow the identification of most disease-causing genes involved in sporadic monogenic disorders. This will allow for the classification of different developmental and neurodevelopmental disorders based on the underlying genotype

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 15 of 19

rather than solely on the phenotype. In turn, this offers the possibility of targeted medical consultations and interventions, engagement in gene-specific patient groups, and, in some cases, treatment. The study of de novo mutations will shift more and more towards the detection and characterization of non-coding de novo mutations in disease. Although a phenomenal challenge that will require large-study cohorts and detailed functional validation, the limited number of de novo mutations per genome reduces the search space for pathogenic non-coding mutations, as was shown recently for non-coding de novo CNVs [198].

Abbreviations
CHD: Congenital heart defect; CNV: Copy number variation; DNM: De novo mutation; ExAC: Exome Aggregation Consortium; Indel: Insertion–deletion; MMR: Mismatch repair; NAHR: Non-allelic homologous recombination; NGS: Next-generation sequencing; PAE: Paternal age effect; PGC: Primordial germ cell; rDNA: Ribosomal DNA; RVIS: Residual variation intolerance score; SNV: Single-nucleotide variant; UMI: Unique molecular identifier; WES: Whole-exome sequencing; WGS: Whole-genome sequencing

Authors' contributions
All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands. [2]Department of Human Genetics, Donders Institute of Neuroscience, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, The Netherlands. [3]Department of Clinical Genetics, GROW – School for Oncology and Developmental Biology, Maastricht University Medical Centre, Universiteitssingel 50, 6229 ER Maastricht, The Netherlands.

References
1. Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science. 2010;328:636–9.
2. Lynch M. Rate, molecular spectrum, and consequences of human mutation. Proc Natl Acad Sci U S A. 2010;107:961–8.
3. Campbell IM, Shaw CA, Stankiewicz P, Lupski JR. Somatic mosaicism: implications for disease and transmission genetics. Trends Genet. 2015;31:382–92.
4. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.
5. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011;147:32–43.
6. Haldane JBS. The rate of spontaneous mutation of a human gene. J Genet. 1935;31:317–26.
7. Nachman MW. Haldane and the first estimates of the human mutation rate. J Genet. 2008;87:317.
8. Kondrashov AS. Direct estimates of human per nucleotide mutation rates at 20 loci causing mendelian diseases. Hum Mutat. 2003;21:12–27.
9. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell. 2012;151:1431–42.
10. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511:344–7.
11. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. Nat Genet. 2015;48:126–33.
12. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, et al. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet. 2015;47:822–6.
13. Goldmann JM, Wong WSW, Pinelli M, Farrah T, Bodian D, Stittrich AB, et al. Parent-of-origin-specific signatures of de novo mutations. Nat Genet. 2016;48:935–9.
14. Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang Y, Casals F, et al. Variation in genome-wide mutation rates within and between human families. Nat Genet. 2011;43:712–4.
15. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488:471–5.
16. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP L, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. 2012;485:246–50.
17. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485:237–41.
18. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. 2012;485:242–5.
19. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jorgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science. 2015;349:1343–7.
20. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 2010;329:75–8.
21. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet. 2004;74:1111–20.
22. Fu W, Akey JM. Selection and adaptation in the human genome. Annu Rev Genomics Hum Genet. 2013;14:467–89.
23. Hurst LD. Fundamental concepts in genetics: genetics and the understanding of selection. Nat Rev Genet. 2009;10:83–93.
24. Lejeune J, Gautier M, Turpin R. Etude des chromosomes somatiques de neuf enfants mongoliens. [Study of somatic chromosomes from 9 mongoloid children]. Comptes rendus Hebd des séances l'Académie des Sci. 1959;248:1721–2.
25. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet. 2012;13:565–75.
26. Kloosterman WP, Francioli LC, Hormozdiari F, Marschall T, Hehir-Kwa JY, Abdellaoui A, et al. Characteristics of de novo structural changes in the human genome. Genome Res. 2015;25:792–801.
27. Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. Trends Genet. 2013;29:575–84.
28. Weischenfeldt J, Symmons O, Spitz F, Korbel JO. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet. 2013;14:125–38.
29. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515:216–21.
30. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011;43:585–9.
31. Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, et al. De novo mutations in epileptic encephalopathies. Nature. 2013;501:217–21.
32. Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. A de novo paradigm for mental retardation. Nat Genet. 2010;42:1109–12.
33. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med. 2012;367:1921–9.
34. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012;380:1674–82.
35. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. Am J Hum Genet. 2015;97:199–215.

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 16 of 19

36. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet. 2016;17:224–38.

37. Girirajan S, Campbell CD, Eichler EE. Human copy number variation and complex genetic disease. Annu Rev Genet. 2011;45:203–26.

38. Ségurel L, Wyman MJ, Przeworski M. Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet. 2014;15:47–70.

39. Korona DA, LeCompte KG, Pursell ZF. The high fidelity and unique error signature of human DNA polymerase. Nucleic Acids Res. 2011;39:1763–73.

40. Schmitt MW, Matsumoto Y, Loeb LA. High fidelity and lesion bypass capability of human DNA polymerase δ. Biochimie. 2009;91:1163–72.

41. Kunkel TA, Erie DA. Eukaryotic mismatch repair in relation to DNA replication. Annu Rev Genet. 2015;49:291–313.

42. Maki H. Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. Annu Rev Genet. 2002;36:279–303.

43. Lindahl T. Quality control by DNA repair. Science. 1999;286:1897–905.

44. Gao Z, Wyman MJ, Sella G, Przeworski M. Interpreting the dependence of mutation rates on age and time. PLoS Biol. 2016;14:e1002355.

45. Goriely A, Wilkie AOM. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. Am J Hum Genet. 2012;90:175–200.

46. Shendure J, Akey JM. The origins, determinants, and consequences of human mutations. Science. 2015;349:1478–83.

47. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nat Genet. 2009;41:393–5.

48. Chen CL, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res. 2010;20:447–57.

49. Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. Am J Hum Genet. 2012;91:1033–40.

50. Green P, Ewing B, Miller W, Thomas PJ, Green ED. Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet. 2003;33:514–7.

51. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. Cell. 2016;164:538–49.

52. Schuster-Böckler B, Lehner B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature. 2012;488:504–7.

53. Chan K, Gordenin DA. Clusters of multiple mutations: incidence and molecular mechanisms. Annu Rev Genet. 2015;49:243–67.

54. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. Nat Rev Genet. 2011;12:756–66.

55. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

56. Shendure J. Human genomics. A deep dive into genetic variation. Nature. 2016;536:277–8.

57. Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. Nat Rev Genet. 2015;16:213–23.

58. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415–21.

59. Petljak M, Alexandrov LB. Understanding mutagenesis through delineation of mutational signatures in human cancer. Carcinogenesis. 2016;37(6):531–40.

60. Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, et al. Clock-like mutational processes in human somatic cells. Nat Genet. 2015;47:1402–7.

61. Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, et al. Break-induced replication is a source of mutation clusters underlying kataegis. Cell Rep. 2014;7:1640–8.

62. Carvalho CMB, Pehlivan D, Ramocki MB, Fang P, Alleva B, Franco LM, et al. Replicative mechanisms for CNV formation are error prone. Nat Genet. 2013;45:1319–26.

63. Neumann R, Lawson VE, Jeffreys AJ. Dynamics and processes of copy number instability in human globin genes. Proc Natl Acad Sci U S A. 2010;107:8304–9.

64. Smith DI, Zhu Y, McAvoy S, Kuhn R. Common fragile sites, extremely large genes, neural development and cancer. Cancer Lett. 2006;232:48–57.

65. Stults DM, Killen MW, Pierce HH, Pierce AJ. Genomic architecture and inheritance of human ribosomal RNA gene clusters. Genome Res. 2007;18:13–8.

66. Bailey JA. Recent segmental duplications in the human genome. Science. 2002;297:1003–7.

67. Weber JL, Wong C. Mutation of human short tandem repeats. Hum Mol Genet. 1993;2:1123–8.

68. Sun JX, Helgason A, Masson G, Ebenesersdóttir SS, Li H, Mallick S, Gnerre S, et al. A direct characterization of human mutation based on microsatellites. Nat Genet. 2012;44:1161–5.

69. Campbell IM, Yuan B, Robberecht C, Pfundt R, Szafranski P, McEntagart ME, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. Am J Hum Genet. 2014;95:173–82.

70. Crow JF. The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet. 2000;1:40–7.

71. Paul C, Robaire B. Ageing of the male germ line. Nat Rev Urol. 2013;10:227–34.

72. Goriely A. Decoding germline de novo point mutations. Nat Genet. 2016;48:823–4.

73. Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, et al. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. Genome Res. 2015;25:1125–34.

74. Sherman SL, Petersen MB, Freeman SB, Hersey J, Pettay D, Taft L, et al. Non-disjunction of chromosome 21 in maternal meiosis I: evidence for a maternal age-dependent mechanism involving reduced recombination. Hum Mol Genet. 1994;3:1529–35.

75. Robinson W. Maternal meiosis I non-disjunction of chromosome 15: dependence of the maternal age effect on level of recombination. Hum Mol Genet. 1998;7:1011–9.

76. Wong WSW, Solomon BD, Bodian DL, Kothiyal P, Eley G, Huddleston KC B, et al. New observations on maternal age effect on germline de novo mutations. Nat Commun. 2016;7:10486.

77. Goriely A, Wilkie AOM. Missing heritability: paternal age effect mutations and selfish spermatogonia. Nat Rev Genet. 2010;11:589.

78. Yoon S-R, Choi S-K, Eboreime J, Gelb BD, Calabrese P, Arnheim N. Age-dependent germline mosaicism of the most common noonan syndrome mutation shows the signature of germline selection. Am J Hum Genet. 2013;92:917–26.

79. Giannoulatou E, McVean G, Taylor IB, McGowan SJ, Maher GJ, Iqbal Z, et al. Contributions of intrinsic mutation rate and selfish selection to levels of de novo HRAS mutations in the paternal germline. Proc Natl Acad Sci U S A. 2013;110:20152–7.

80. Arnheim N, Calabrese P. Germline stem cell competition, mutation hot spots, genetic disorders, and older fathers. Annu Rev Genomics Hum Genet. 2016;17:219–43.

81. Maher GJ, McGowan SJ, Giannoulatou E, Verrill C, Goriely A, Wilkie AOM. Visualizing the origins of selfish de novo mutations in individual seminiferous tubules of human testes. Proc Natl Acad Sci U S A. 2016;113:2454–9.

82. Goriely A, Hansen RMS, Taylor IB, Olesen IA, Jacobsen GK, McGowan SJ, et al. Activating mutations in FGFR3 and HRAS reveal a shared genetic origin for congenital disorders and testicular tumors. Nat Genet. 2009;41:1247–52.

83. Goriely A, McVean GAT, Röjmyr M, Ingemarsson B, Wilkie AOM. Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. Science. 2003;301:643–6.

84. Choi S-K, Yoon S-R, Calabrese P, Arnheim N. Positive selection for new disease mutations in the human germline: evidence from the heritable cancer syndrome multiple endocrine neoplasia type 2B. PLoS Genet. 2012;8:e1002420.

85. Arnheim N, Calabrese P. Understanding what determines the frequency and pattern of human germline mutations. Nat Rev Genet. 2009;10:478–88.

86. Goriely A, McGrath JJ, Hultman CM, Wilkie AOM, Malaspina D. "Selfish Spermatogonial Selection": a novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. Am J Psychiatry. 2013;170:599–608.

87. Lupski JR. New mutations and intellectual function. Nat Genet. 2010;42:1036–8.

88. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, et al. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. Am J Hum Genet. 2015;97:67–74.

89. Dal GM, Ergüner B, Sağıroğlu MS, Yüksel B, Onat OE, Alkan C, et al. Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. J Med Genet. 2014;51:455–9.

90. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet. 2013;14:307–20.

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 17 of 19

91. Besenbacher S, Liu S, Izarzugaza JMG, Grove J, Belling K, Bork-Jensen J, et al. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. Nat Commun. 2015;6:5969.

92. Voet T, Vanneste E, Vermeesch JR. The human cleavage stage embryo is a cradle of chromosomal rearrangements. Cytogenet Genome Res. 2011;133:160–8.

93. Vanneste E, Voet T, Le Caignec C, Ampe M, Konings P, Melotte C, et al. Chromosome instability is common in human cleavage-stage embryos. Nat Med. 2009;15:577–83.

94. Lee MT, Bonneau AR, Giraldez AJ. Zygotic genome activation during the maternal-to-zygotic transition. Annu Rev Cell Dev Biol. 2014;30:581–613.

95. Youssoufian H, Pyeritz RE. Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet. 2002;3:748–58.

96. Huang AY, Xu X, Ye AY, Wu Q, Yan L, Zhao B, et al. Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. Cell Res. 2014;24:1311–27.

97. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. Proc Natl Acad Sci U S A. 2013;110:1999–2004.

98. O'Huallachain M, Karczewski KJ, Weissman SM, Urban AE, Snyder MP. Extensive genetic variation in somatic human tissues. Proc Natl Acad Sci U S A. 2012;109:18018–23.

99. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012;44:642–50.

100. Jacobs KB, Yeager M, Zhou W, Wacholder S, Wang Z, Rodriguez-Santiago B, et al. Detectable clonal mosaicism and its relationship to aging and cancer. Nat Genet. 2012;44:651–8.

101. Forsberg LA, Rasi C, Razzaghian HR, Pakalapati G, Waite L, Thilbeault KS, et al. Age-related somatic structural changes in the nuclear genome of human blood cells. Am J Hum Genet. 2012;90:217–28.

102. Stone JF, Sandberg AA. Sex chromosome aneuploidy and aging. Mutat Res. 1995;338:107–13.

103. Dumanski JP, Rasi C, Lonn M, Davies H, Ingelsson M, Giedraitis V, et al. Smoking is associated with mosaic loss of chromosome Y. Science. 2015;347:81–3.

104. Yadav VK, DeGregori J, De S. The landscape of somatic mutations in protein coding genes in apparently benign human tissues carries signatures of relaxed purifying selection. Nucleic Acids Res. 2016;44:2075–84.

105. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015;350:94–8.

106. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371:2477–87.

107. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371:2488–98.

108. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20:1472–8.

109. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015;348:880–6.

110. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 2016;17:175–88.

111. Walter CA, Intano GW, McCarrey JR, McMahan CA, Walter RB. Mutation frequency declines during spermatogenesis in young mice but increases in old mice. Proc Natl Acad Sci U S A. 1998;95:10015–9.

112. Kohler SW, Provost GS, Fieck A, Kretz PL, Bullock WO, Sorge JA, et al. Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice. Proc Natl Acad Sci U S A. 1991;88:7958–62.

113. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012;20:490–7.

114. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013;14:681–91.

115. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014;312:1870.

116. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016;18:678–85.

117. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2015;17:9–18.

118. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. Psychol Med. 2015;45:601–13.

119. Hoischen A, Krumm N, Eichler EE. Prioritization of neurodevelopmental disease genes by discovery of new mutations. Nat Neurosci. 2014;17:764–72.

120. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee Y-H, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. 2015;112:E5600–7.

121. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016;22:345–61.

122. Gratten J, Wray NR, Peyrot WJ, McGrath JJ, Visscher PM, Goddard ME. Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. Nat Genet. 2016;48:718–24.

123. Homsy J, Zaidi S, Shen Y, Ware JS, Samocha KE, Karczewski KJ, et al. De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science. 2015;350:1262–6.

124. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature. 2013;498:220–3.

125. Sifrim A, Hitz M-P, Wilsdon A, Breckpot J, Turki SH A, Thienpont B, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet. 2016;48:1060–5.

126. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014;46:944–50.

127. Fitzgerald TW, Gerety SS, Jones WD, van Kogelenberg M, King DA, McRae J, et al. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2014;519:223–8.

128. Lelieveld SH, Reijnders MRF, Pfundt R, Yntema HG, Kamsteeg E, de Vries P, et al. Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. Nat Neurosci. 2016;19:1194–6.

129. Pamphlett R, Morahan JM, Yu B. Using case-parent trios to look for rare de novo genetic variants in adult-onset neurodegenerative diseases. J Neurosci Methods. 2011;197:297–301.

130. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature. 2014;506:179–84.

131. Gauthier J, Champagne N, Lafrenière RG, Xiong L, Spiegelman D, Brustein E, et al. De novo mutations in the gene encoding the synaptic scaffolding protein SHANK3 in patients ascertained for schizophrenia. Proc Natl Acad Sci U S A. 2010;107:7863–8.

132. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet. 2008;40:880–5.

133. Kun-Rodrigues C, Ganos C, Guerreiro R, Schneider SA, Schulte C, Lesage S, et al. A systematic screening to identify de novo mutations causing sporadic early-onset Parkinson's disease. Hum Mol Genet. 2015;24:6711–20.

134. Chesi A, Staahl BT, Jovičić A, Couthouis J, Fasolino M, Raphael AR, et al. Exome sequencing to identify de novo mutations in sporadic ALS trios. Nat Neurosci. 2013;16:851–5.

135. Steinberg KM, Yu B, Koboldt DC, Mardis ER, Pamphlett R. Exome sequencing of case-unaffected-parents trios reveals recessive and de novo genetic variants in sporadic ALS. Sci Rep. 2015;5:9124.

136. Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. Science. 2015;349:1489–94.

137. Georgieva L, Rees E, Moran JL, Chambert KD, Milanova V, Craddock N, et al. De novo CNVs in bipolar affective disorder and schizophrenia. Hum Mol Genet. 2014;23:6677–83.

138. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016;19:571–7.

139. Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet. 2012;44:1365–9.

140. Zhu X, Need AC, Petrovski S, Goldstein DB. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. Nat Neurosci. 2014;17:773–81.

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 18 of 19

141. Gratten J, Wray NR, Keller MC, Visscher PM. Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat Neurosci. 2014;17:782–90.

142. van Os J, Kenis G, Rutten BPF. The environment and schizophrenia. Nature. 2010;468:203–12.

143. Gonzalez KD, Buzin CH, Noltner KA, Gu D, Li W, Malkin D, et al. High frequency of de novo mutations in Li-Fraumeni syndrome. J Med Genet. 2009;46:689–93.

144. Aretz S, Uhlhaas S, Caspari R, Mangold E, Pagenstecher C, Propping P, et al. Frequency and parental origin of de novo APC mutations in familial adenomatous polyposis. Eur J Hum Genet. 2004;12:52–8.

145. Golmard L, Delnatte C, Laugé A, Moncoutier V, Lefol C, Abidallah K, et al. Breast and ovarian cancer predisposition due to de novo BRCA1 and BRCA2 mutations. Oncogene. 2016;35:1324–7.

146. Win AK, Jenkins MA, Buchanan DD, Clendenning M, Young JP, Giles GG, et al. Determining the frequency of de novo germline mutations in DNA mismatch repair genes. J Med Genet. 2011;48:530–4.

147. Neveling K, Collin RWJ, Gilissen C, van Huet RAC, Visser L, Kwint MP, et al. Next-generation genetic testing for retinitis pigmentosa. Hum Mutat. 2012;33:963–72.

148. Glöckle N, Kohl S, Mohr J, Scheurenbrand T, Sprecher A, Weisschuh N, et al. Panel-based next generation sequencing as a reliable and efficient technique to detect mutations in unselected patients with retinal dystrophies. Eur J Hum Genet. 2014;22:99–104.

149. Kim NKD, Kim AR, Park KT, Kim SY, Kim MY, Nam J, et al. Whole-exome sequencing reveals diverse modes of inheritance in sporadic mild to moderate sensorineural hearing loss in a pediatric population. Genet Med. 2015;17:901–11.

150. Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, et al. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat Genet. 1999;23:429–32.

151. Moro E. Male infertility caused by a de novo partial deletion of the DAZ cluster on the Y chromosome. J Clin Endocrinol Metab. 2000;85:4069–73.

152. Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, et al. A mosaic activating mutation in AKT1 associated with the proteus syndrome. N Engl J Med. 2011;365:611–9.

153. Shirley MD, Tang H, Gallione CJ, Baugher JD, Frelin LP, Cohen B, et al. Sturge–Weber syndrome and port-wine stains caused by somatic mutation in GNAQ. N Engl J Med. 2013;368:1971–9.

154. Rivière J-B, Mirzaa GM, O'Roak BJ, Beddaoui M, Alcantara D, Conway RL, et al. De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. Nat Genet. 2012;44:934–40.

155. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline mutations in predisposition genes in pediatric cancer. N Engl J Med. 2015;373:2336–46.

156. Fernández LC, Torres M, Real FX. Somatic mosaicism: on the road to cancer. Nat Rev Cancer. 2015;16:43–55.

157. Hafner C, Toll A, Real FX. HRAS mutation mosaicism causing urothelial cancer and epidermal nevus. N Engl J Med. 2011;365:1940–2.

158. Campbell IM, Stewart JR, James RA, Lupski JR, Stankiewicz P, Olofsson P, et al. Parent of origin, mosaicism, and recurrence risk: probabilistic modeling explains the broken symmetry of transmission genetics. Am J Hum Genet. 2014;95:345–59.

159. Huisman SA, Redeker EJW, Maas SM, Mannens MM, Hennekam RCM. High rate of mosaicism in individuals with Cornelia de Lange syndrome. J Med Genet. 2013;50:339–44.

160. Halvorsen M, Petrovski S, Shellhaas R, Tang Y, Crandall L, Goldstein D, et al. Mosaic mutations in early-onset genetic diseases. Genet Med. 2016;18:746–9.

161. Jamuar SS, Lam A-TN, Kircher M, D'Gama AM, Wang J, Barry BJ, et al. Somatic mutations in cerebral cortical malformations. N Engl J Med. 2014;371:733–43.

162. Okajima K, Warman ML, Byrne LC, Kerr DS. Somatic mosaicism in a male with an exon skipping mutation in PDHA1 of the pyruvate dehydrogenase complex results in a milder phenotype. Mol Genet Metab. 2006;87:162–8.

163. Plant KE, Boye E, Green PM, Vetrie D, Flinter FA. Somatic mosaicism associated with a mild Alport syndrome phenotype. J Med Genet. 2000;37:238–9.

164. Groesser L, Herschberger E, Ruetten A, Ruivenkamp C, Lopriore E, Zutt M, et al. Postzygotic HRAS and KRAS mutations cause nevus sebaceous and Schimmelpenning syndrome. Nat Genet. 2012;44:783–7.

165. Happle R. Lethal genes surviving by mosaicism: a possible explanation for sporadic birth defects involving the skin. J Am Acad Dermatol. 1987;16:899–906.

166. Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating mutations of the stimulatory G protein in the McCune–Albright syndrome. N Engl J Med. 1991;325:1688–95.

167. Kurek KC, Luks VL, Ayturk UM, Alomari AI, Fishman SJ, Spencer SA, et al. Somatic mosaic activating mutations in PIK3CA cause CLOVES syndrome. Am J Hum Genet. 2012;90:1108–15.

168. Mirzaa G, Timms AE, Conti V, Boyle EA, Girisha KM, Martin B, et al. PIK3CA-associated developmental disorders exhibit distinct classes of mutations with variable expression and tissue distribution. JCI Insight. 2016;1:1–18.

169. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.

170. Hafner C, Groesser L. Mosaic RASopathies. Cell Cycle. 2013;12:43–50.

171. Pollock PM, Harper UL, Hansen KS, Yudt LM, Stark M, Robbins CM, et al. High frequency of BRAF mutations in nevi. Nat Genet. 2002;33:19–20.

172. Aoki Y, Niihori T, Kawame H, Kurosawa K, Ohashi H, Tanaka Y, et al. Germline mutations in HRAS proto-oncogene cause Costello syndrome. Nat Genet. 2005;37:1038–40.

173. Levinsohn JL, Teng J, Craiglow BG, Loring EC, Burrow TA, Mane SS, et al. Somatic HRAS p.G12S mutation causes woolly hair and epidermal nevi. J Invest Dermatol. 2014;134:1149–52.

174. Beukers W, Hercegovac A, Zwarthoff EC. HRAS mutations in bladder cancer at an early age and the possible association with the Costello Syndrome. Eur J Hum Genet. 2014;22:837–9.

175. Luks VL, Kamitaki N, Vivero MP, Uller W, Rab R, Bovée JVMG R, et al. Lymphatic and other vascular malformative/overgrowth disorders are caused by somatic mutations in PIK3CA. J Pediatr. 2015;166:1048–54.e1-5.

176. Limaye N, Kangas J, Mendola A, Godfraind C, Schlögel MJ, Helaers R, et al. Somatic activating PIK3CA mutations cause venous malformation. Am J Hum Genet. 2015;97:914–21.

177. Clavería C, Giovinazzo G, Sierra R, Torres M. Myc-driven endogenous cell competition in the early mammalian embryo. Nature. 2013;500:39–44.

178. Wieacker P, Wieland I. Clinical and genetic aspects of craniofrontonasal syndrome: towards resolving a genetic paradox. Mol Genet Metab. 2005;86:110–6.

179. Twigg SRF, Babbs C, van den Elzen MEP, Goriely A, Taylor S, McGowan SJ, et al. Cellular interference in craniofrontonasal syndrome: males mosaic for mutations in the X-linked EFNB1 gene are more severely affected than true hemizygotes. Hum Mol Genet. 2013;22:1654–62.

180. Poduri A, Evrony GD, Cai X, Walsh CA. Somatic mutation, genomic variation, and neurological disease. Science. 2013;341:1237758.

181. Insel TR. Brain somatic mutations: the dark matter of psychiatric genetics? Mol Psychiatry. 2014;19:156–8.

182. Sun K, Jiang P, Chan KCA, Wong J, Cheng YKY, Liang RHS, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A. 2015;112:E5503–12.

183. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell. 2016;164:57–68.

184. Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A. 2016;113:E1826–34.

185. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010;42:790–3.

186. Hoischen A, van Bon BWM, Rodríguez-Santiago B, Gilissen C, Vissers LELM, de Vries P, et al. De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. Nat Genet. 2011;43:729–31.

187. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011;12:745–55.

188. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet. 2013;9:e1003709.

189. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. Cell. 2014;156:872–7.

190. Klepper J, Scheffer H, Leiendecker B, Gertsen E, Binder S, Leferink M, et al. Seizure control and acceptance of the ketogenic diet in GLUT1 deficiency syndrome: a 2- to 5-year follow-up of 15 children enrolled prospectively. Neuropediatrics. 2005;36:302–8.

Acuna-Hidalgo *et al. Genome Biology* (2016) 17:241

Page 19 of 19

191. Brandler WM, Sebat J. From de novo mutations to personalized therapeutic interventions in autism. Annu Rev Med. 2015;66:487–507.

192. James CA, Hadley DW, Holtzman NA, Winkelstein JA. How does the mode of inheritance of a genetic condition influence families? A study of guilt, blame, stigma, and understanding of inheritance and reproductive risks in families with X-linked and autosomal recessive diseases. Genet Med. 2006;8:234–42.

193. McAllister M, Davies L, Payne K, Nicholls S, Donnai D, MacLeod R. The emotional effects of genetic diseases: implications for clinical genetics. Am J Med Genet Part A. 2007;143A:2651–61.

194. Krupp DR, Barnard RA, Duffourd Y, Evans S, Bernier R, Rivière J-B, et al. Exonic somatic mutations contribute risk for autism spectrum disorder. bioRxiv. 2016. http://dx.doi.org/10.1101/083428.

195. D'Onofrio BM, Rickert ME, Frans E, Kuja-Halkola R, Almqvist C, Sjölander A, et al. Paternal age at childbearing and offspring psychiatric and academic morbidity. JAMA Psychiatry. 2014;71:432.

196. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. N Engl J Med. 2013;369:1502–11.

197. Cobo A, García-Velasco JA, Coello A, Domingo J, Pellicer A, Remohí J. Oocyte vitrification as an efficient option for elective fertility preservation. Fertil Steril. 2016;105:755–764.e8.

198. Flottmann R, Wagner J, Kobus K, Curry CJ, Savarirayan R, Nishimura G, et al. Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. J Med Genet. 2015;52:1–8.

199. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet. 2010;11:31–46.

200. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30:434–9.

201. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 2008;18:1851–8.

202. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013;23:843–54.

203. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17:333–51.

204. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733–5.

205. Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, et al. Resolving the complexity of the human genome using single-molecule sequencing. Nature. 2014;517:608–11.

206. Ritz A, Bashir A, Sindi S, Hsu D, Hajirasouliha I, Raphael BJ. Characterization of structural variants with single molecule and hybrid sequencing approaches. Bioinformatics. 2014;30:3458–66.

207. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444:444–54.

208. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46:1063–71.

209. Sanna-Cherchi S, Kiryluk K, Burgess KE, Bodria M, Sampson MG, Hadley D, et al. Copy-number disorders are a common cause of congenital kidney malformations. Am J Hum Genet. 2012;91:987–97.

210. Glessner JT, Bick AG, Ito K, Homsy JG, Rodriguez-Murillo L, Fromer M, et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circ Res. 2014;115:884–96.

211. Chiang DY, Getz G, Jaffe DB, O'Kelly MJT, Zhao X, Carter SL, et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods. 2009;6:99–103.

212. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011;12:363–76.

213. Lupski JR. Genomic rearrangements and sporadic disease. Nat Genet. 2007;39:S43–7.

214. Hehir-Kwa JY, Rodríguez-Santiago B, Vissers LE, de Leeuw N, Pfundt R, Buitelaar JK, et al. De novo copy number variants associated with intellectual disability have a paternal origin and age bias. J Med Genet. 2011;48:776–8.

215. Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell. 2007;131:1235–47.

216. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet. 1998;14:417–22.

217. Duyzend MH, Nuttle X, Coe BP, Baker C, Nickerson DA, Bernier R, et al. Maternal modifiers and parent-of-origin bias of the autism-associated 16p11.2 CNV. Am J Hum Genet. 2016;98:45–57.

218. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, et al. A common inversion under selection in Europeans. Nat Genet. 2005;37:129–37.

219. Koolen DA, Sharp AJ, Hurst JA, Firth HV, Knight SJL, Goldenberg A, et al. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. J Med Genet. 2008;45:710–20.

220. Vermeesch JR, Balikova I, Schrander-Stumpel C, Fryns J-P, Devriendt K. The causality of de novo copy number variants is overestimated. Eur J Hum Genet. 2011;19:1112–3.

221. MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. Nature. 2014;508:469–76.

222. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet. 2011;12:628–40.

223. Sunyaev SR. Inferring causality and functional significance of human coding DNA variants. Hum Mol Genet. 2012;21:R10–7.

224. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.

225. Higurashi N, Uchida T, Lossin C, Misumi Y, Okada Y, Akamatsu W, et al. A human Dravet syndrome model from patient induced pluripotent stem cells. Mol Brain. 2013;6:19.

226. Kuechler A, Zink AM, Wieland T, Lüdecke H-J, Cremer K, Salviati L, et al. Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. Eur J Hum Genet. 2015;23:753–60.

227. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of Topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161:1012–25.

228. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. Nature. 2014;513:120–3.

229. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012;338:1619–22.

230. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat. 2015;36:915–21.

231. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat. 2015;36:928–30.

232. Vissers LELM, Veltman JA. Standardized phenotyping enhances Mendelian disease gene identification. Nat Genet. 2015;47:1222–4.

233. Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Baynam G, Zemojtel T, et al. The Human phenotype ontology: semantic unification of common and rare disease. Am J Hum Genet. 2015;97:111–24.

234. Bone WP, Washington NL, Buske OJ, Adams DR, Davis J, Draper D, et al. Computational evaluation of exome sequence data using human and model organism phenotypes improves diagnostic efficiency. Genet Med. 2016;18:608–17.

235. Minikel EV, Vallabh SM, Lek M, Estrada K, Samocha KE, Sathirapongsasuti JF, et al. Quantifying prion disease penetrance using large population control cohorts. Sci Transl Med. 2016;8:322ra9.

236. Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, et al. Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. Genet Med. 2016. doi:10.1038/gim.2016.90

237. Chen R, Shi L, Hakenberg J, Naughton B, Sklar P, Zhang J, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. Nat Biotechnol. 2016;34:531–8.

238. Scally A. Mutation rates and the evolution of germline structure. Philos Trans R Soc B Biol Sci. 2016;371:20150137.