CellPress
REVIEWS

Review

Somatic Variants: New Kids on the Block in Human Immunogenetics

L. Van Horebeek,[1,2] B. Dubois,[1,2,3] and A. Goris[1,2,*]

Somatic variants are not inherited but acquired during an individual's lifetime, and individuals are increasingly considered as complex mosaics of genetically distinct cells. Whereas this concept is long-recognized in cancer, this review focuses on the growing role of somatic variants in immune cells in nonmalignant immune-related disorders, such as primary immunodeficiency and autoimmune diseases. Older case reports described somatic variants early in development, leading to large numbers of affected cells and severe phenotypes. Thanks to technological evolution, it is now feasible to detect somatic variants occurring later in life and affecting fewer cells. Hence, only recently is the scale at which somatic variants contribute to monogenic diseases being uncovered and is their contribution to complex diseases being explored systematically.

Somatic Variants: An Older Concept in a New Context

Somatic **variants** (see Glossary) arise as postzygotic **mutations** in a single progenitor cell and can through transmission to its daughter cells accumulate in cells of the human **soma** (Figure 1). The presence of somatic variants results in **mosaicism.** Hence, we can consider individuals as complex mosaics of genetically distinct cells. Variants restricted to somatic cells (**somatic mosaicism**) are typically not inherited over generations, with the exception of somatic variants co-existing in somatic and germ cells (**gonosomal mosaicism**).

Linkage studies and subsequently genome-wide association studies (GWASs) have made much progress in understanding genetic determinants inherited through the germline for monogenic and complex diseases. Somatic variation has a physiological role in generating immune receptor diversity in the immune system and has long been recognized in cancer. Early on, somatic mosaicism has also been put forward as an important and sometimes dramatic cause for phenotypic variation in the expression of genetic traits [1]. Subsequently, a multistep pathogenesis for immune diseases has been suggested, in which multiple variants, both inherited and somatic, contribute to emergence of disease [2]. Case reports where important genes are mutated early in development, leading to large numbers of affected cells and severe immune disease, appeared around 15 years ago [3,4]. Thanks to technological evolution, it is now becoming feasible to detect somatic variants occurring later in life and affecting fewer cells [5]. This has revealed that somatic variants arise relatively frequently (Box 1). Moreover, only recently is the scale at which somatic variants contribute to monogenic diseases being uncovered and the contribution of somatic variants to complex diseases being explored systematically (Figure 2, Key Figure).

In this review, we focus on somatic and gonosomal mosaicism due to **single nucleotide variants (SNVs)** affecting cells of the immune system and on the increasing evidence that they may contribute to immune-related diseases beyond cancer. Somatic mosaicism in other cell types, such as brain or skin, has been reviewed elsewhere [6,7].

An Under-Recognized Cause of Monogenic Diseases

Recognizing the role of somatic variants in immune-related diseases started from phenotypes overlapping between proliferative disorder and autoimmunity. Autoimmune lymphoproliferative syndrome (ALPS) is characterized by nonmalignant lymphoproliferation and signs of autoimmunity. Both dominant and recessive germline mutations in different genes of the Fas pathway have been described. In 2004, six patients were reported who carried somatic *FAS* (*TNFRSF6*) mutations for which either the variant itself or a variant with the same effect was previously described as a germline mutation [3]. Later estimates indicated that somatic mutations account for up to 20% of ALPS cases [8,9],

Highlights

Somatic variants arise frequently, at least one order of magnitude more often than germline variants. They can occur early in development or later in life, and lead to mosaicism in a wide range or a specific subset of cell lineages.

Somatic mosaicism is an under-recognized cause of monogenic immune disorders. Moreover, gonosomal mosaicism in mostly unaffected or sometimes mildly affected parents can lead to vertical transmission of a presumed *de novo* germline variant.

Somatic variants may act as stochastic factors increasing risk of complex diseases together with inherited germline variants and environmental risk factors.

Somatic variants tagging cells involved in immune pathogenesis may act as biomarkers guiding treatment.

Improving technologies will facilitate detection of somatic variants affecting smaller cell subsets. Integration of germline and somatic genomes may provide novel insights into disease pathogenesis.

[1]KU Leuven Department of Neurosciences, Laboratory for Neuroimmunology, 3000 Leuven, Belgium

[2]Leuven Brain Institute, 3000 Leuven, Belgium

[3]University Hospitals Leuven, Department of Neurology, 3000 Leuven, Belgium

*Correspondence: an.goris@kuleuven.be

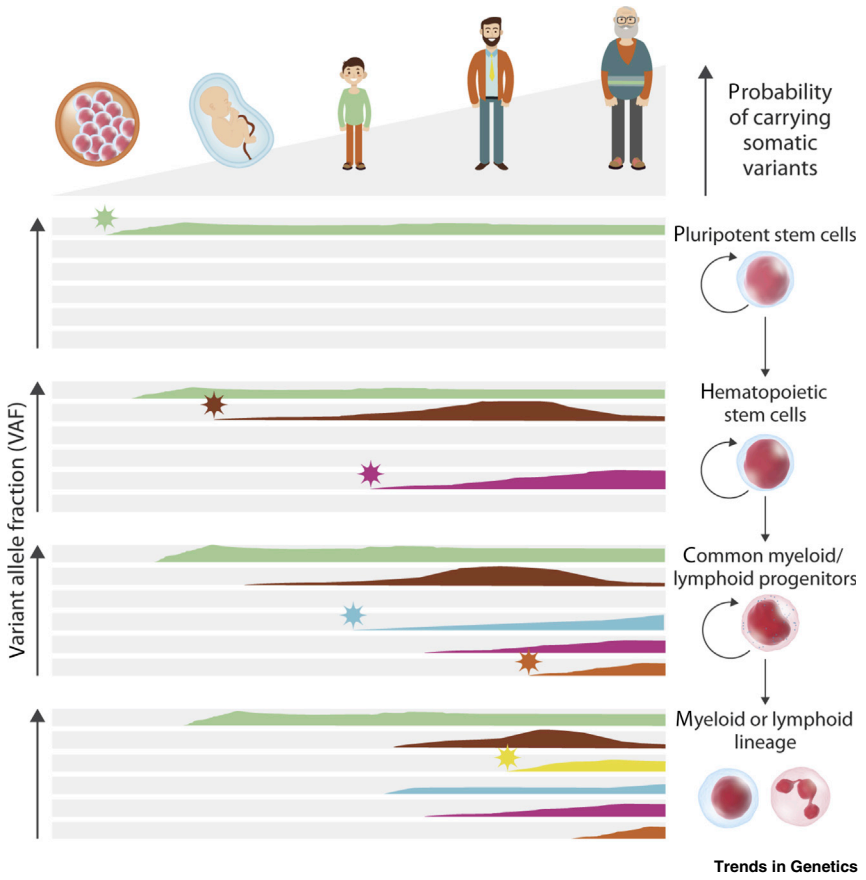

Figure 1. Somatic Variants May Occur Early in Development or Later in Life, Lead to Mosaicism in a Wide Range or a Specific Subset of Cell Lineages, and Are Relatively Stable over Time.

Somatic variants (color-coded) can arise at any time point and at any step of the differentiation into immune cells. Differentiation of affected cells may result in the detection of variants in the derived cells. Examples reflect possible scenarios for a selection of variants reported in the literature. Green: *FAS* variant originating in pluripotent stem cells during embryonic development resulting in autoimmune lymphoproliferative syndrome. Brown, blue, pink, and orange: *NLRP3* variants originating in hematopoietic stem cells or common myeloid/lymphoid progenitors during fetal development, childhood or adulthood, resulting in (late-onset) cryopyrin-associated periodic syndrome (CAPS). Yellow: *STAT3* variants originating in the lymphoid lineage during adulthood, increasing the risk for rheumatoid arthritis. Early occurrence or a survival or proliferation advantage may lead to an increasing number of cells carrying the somatic variant. Available data suggest that the variant allele fraction (VAF) remains quite stable over time for the majority of the variants. However, factors such as disease progression or treatment may alter abundance.

and that disease-causing germline and somatic mutations concentrate in a few small regions, probably representing hot spots [9]. Moreover, the incomplete penetrance of heterozygous germline *FAS* mutations seen in seven families can be explained by somatic changes in the second **allele** [10]. This provides a survival advantage to cells that are normally constrained by Fas-dependent apoptosis, in particular $CD3^{+}CD4^{-}CD8^{-}$ double negative T lymphocytes, leading to uncontrolled proliferation and high circulating levels of these cells, a hallmark of ALPS.

From 2005 onwards, somatic mutations were described in additional nonproliferative immune disorders, either **primary immunodeficiency diseases (PIDs)** or **autoinflammatory diseases (AIDs)** [4]. The first and still most-recognized example is cryopyrin-associated periodic syndrome (CAPS), comprising the three

Glossary

Allele: one of the possible nucleotide variations.
Autoimmune diseases: a heterogeneous group of diseases characterized by the specific adaptive immune system reacting towards self-antigens.
Autoinflammatory disease: a heterogeneous group of diseases characterized by uncontrolled activation of the nonspecific innate immune system.
Bulk DNA: DNA from a large number of cells.
Clonal hematopoiesis of indeterminate potential (CHIP): the presence of an expanded somatic blood-cell clone in persons without other hematologic abnormalities.
Driver mutations: mutations that provide the cells carrying them a survival or proliferation advantage.
Gonosomal mosaicism: mosaicism with postzygotic mutations affecting both the soma and the reproductive cells.
***HPRT* assay**: an assay to test for the presence of inactivating mutations in the *HPRT* gene, which encodes for hypoxanthine guanine phosphoribosyl transferase, based on whether the cultured cells survive in medium containing thioguanine.
Mosaicism: the state of being composed of two or more genetically distinct cell populations
Mutation: used for both the mutation process giving rise to a new variant and a variant with a known pathogenic effect on the organism in the context of monogenic disorders.
Primary immunodeficiency diseases (PIDs): a heterogeneous group of inherited diseases characterized by absent or dysregulated immune responses.
Revertant mosaicism: mosaicism due to spontaneous correction of a pathogenic mutation in a somatic cell.
Second-site revertant mutation: a somatic variant at a second location that counteracts the effect of a germline variant at another location.
Single-cell sequencing: examination of the sequence information of individual cells.

Box 1. Somatic Mutation Rates during Embryonic Development and Lifetime

Germline mutation rates are estimated at 0.6×10^{-10} mutations per base pair per cell division [67] or 0.4–1.4 mutations per diploid genome per cell division [23]. *De novo* germline mutation rates are estimated at 1.28×10^{-8} mutations per base pair per generation, with an average of 64 *de novo* SNVs in the nonrepetitive genome of an individual, a number that varies with parental age [23].

During early embryonic development, 2.8 (95% confidence interval 2.4–3.3) mutations occur per cell per cell-doubling event (which due to cell loss can entail >1 cell division), comparable with or slightly higher than for germline [21]. Approximately 14 early embryonic substitutions are predicted in the blood of any individual, and most are also detectable in other tissues. However, current methods find only few (0–4) from the personal whole-genome sequence of an individual [21].

Somatic mutation rates later in life are $\sim6.4\times10^{-10}$–7.8×10^{-10} mutations per base pair per cell division across cell types [68,69], and can reach an order of magnitude higher than for germline because they only need to be compatible with normal cells and not organism function [67]. Mutation rates per cell division seem to be comparable across cell types, and rates per year hence depend on cell renewal rates [57]. Rates for lymphocytes are estimated at 7.67×10^{-9} mutations per base pair per year [68], and may occur more often within hotspots in the genome [52]. Exome sequencing in three hematopoietic stem/progenitor cells (HSPCs) from each of seven healthy individuals (0–80 years) [69] and whole-genome sequencing in blood of a 115-year-old woman [70] provided estimates of 0.13 exonic SNVs per year or four or five SNVs in the nonrepetitive genome per year [69,70]. In total, the female centenarian carried an extreme number of 424 SNVs, largely confined to the blood compartment, with characteristics very similar to germline variants in public databases (dbSNP) and very few nonsense/missense or damaging variants [70].

Whole-genome data from blood of 11 262 individuals across a wide age range (10–100 years) more clearly demonstrate the effect of age [32]. Younger subjects carry a median of three somatic variants, whereas the number of people with a high count of somatic variants climbed rapidly above 35–45 years [32]. Extremes as defined by >20 somatic variants occurred in 12.5% of the overall population, but in only 0.5% of those <35 years and increasing to over half of those >85 years.

Single nucleotide variant (SNV): variation in a single nucleotide at a specific position in the genome.
Soma: all of an organism except the reproductive cells.
Somatic mosaicism: mosaicism with postzygotic mutations restricted to the soma.
Variant: an alteration from the reference sequence.
Variant allele fraction (VAF): the proportion of the alleles in a specific cell population that carry the variant allele at a particular position.

entities of chronic infantile neurological, cutaneous articular (CINCA) syndrome, Muckle–Wells syndrome (MWS), and familial cold autoinflammatory syndrome (FCAS). Germline mutations are found in the *NLRP3* gene and lead to enhanced activation of the NLRP3-inflammasome and overproduction of interleukin (IL)-1β. Up to 40% of patients do not carry germline *NLRP3* mutations, and may be explained by somatic *NLRP3* mutations [11,12]. By now, somatic variants in immune cells may explain up to 25% of PID/AID patients [11–13], and are observed in a range of monogenic immune-related phenotypes (Table 1). Many patients carrying mutations with low abundance; that is, characterized by a low **variant allele fraction (VAF)**, may still be missed (Box 2). As an example, the range of currently observed VAFs for *NLRP3* (median 11%, range 3–36%) in peripheral blood overlaps the typical detection limit (>5%) [4,11–19].

Depending on the stage during embryonic development or later in life when somatic mutations occur, mosaicism is either widely present across cell types and tissues or restricted to particular cell types (Figure 1). *NLRP3* and *FAS* mutations may arise early, before pluripotent stem cells commit to hematopoietic progenitor stem cells or ectoderm-derived nonhematopoietic stem cells, leading to the mutation occurring across cell types and lineages [3,4,10,11,15,17]. In other patients, the mutation has occurred later in embryonic development or even later in life, and may lead to myeloid-restricted (*NLRP3*) [17–19] and myeloid- and lymphoid- or lymphoid-restricted mosaicism (*FAS*) [3,10]. Patients present with typical disease, although onset may vary and later onset, even during adulthood, is not uncommon [10,17–20].

Besides the cell type affected, also the VAF and whether or not the variant is known as germline may affect phenotype (Figure 2). Among the group of patients carrying somatic *NLRP3* mutations also known as germline mutations, those with intermediate VAFs (10–30%) may present with milder symptoms, especially less mental retardation and neurological involvement [11], than those with VAFs >30%, which appear clinically indistinguishable from patients with germline mutations. By contrast,

Key Figure

Spectrum of Somatic Variants Based on Their Abundance and Effect on Disease Phenotypes

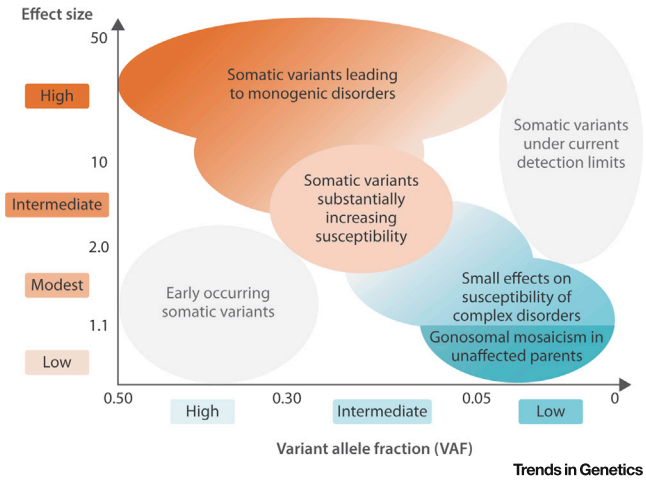

Figure 2. Somatic variants may have different effect sizes, depending on the abundance (variant allele fraction; VAF) but also the identity of the variant, the affected cell types, and the moment of arising. A simplified depiction based on current emerging, but still limited, knowledge is given. The scale in which somatic variants contribute to monogenic immune disorders is only recently becoming clear. Such variants, for example, *NLRP3* variants resulting in autoinflammatory disease, have been frequently reported with high and intermediate abundance. The identification of variants with relatively high effect size but low abundance may still be hampered by detection limits in most current studies in this context starting from total blood. Identification may be improved through deep sequencing in total blood and/or investigating different immune cell subsets. Lowly abundant somatic variants are also seen as part of gonosomal mosaicism in unaffected parents, which can explain the unexpected recurrence of disease within families. First examples of somatic variants with intermediate abundance in immune cells substantially increasing susceptibility for complex diseases include *STAT3* variants increasing risk of autoimmune disease and clonal hematopoiesis of indeterminate potential (CHIP) increasing risk of cardiovascular diseases. Thanks to technological evolution, it is now clear that somatic variants occurring later in life and affecting fewer cells or limited to specific immune cell subsets are a common source of genetic variation. When sufficiently large study populations are possible, small to modest effect sizes on the risk of complex diseases such as autoimmune diseases may be detected for these variants. Somatic variants that occur early in development may reach high abundance even without providing the cell with a strong advantage. However, current methods detect too few of them to draw conclusions on their possible effects on human traits (Box 1). This figure was made in analogy to the classical figure depicting the spectrum of germline variants [74].

those patients harboring novel mutations not previously reported in germline status may be more prone to severe symptoms independent of VAF [4,13]. An explanation could be that these variants are too damaging as germline variants to be compatible with normal organism function. Somatic variants, however, only need to be compatible with normal cell function. Examples of gonosomal variants described in the next section indicate that low VAFs may also remain asymptomatic or lead to milder disease.

De Novo Variants: Somatic Variants in Disguise?

De novo variants occur in a germ cell or zygote, whereas somatic variants are defined as postzygotic. Somatic variants that have arisen during early embryogenic development before the separation of

Gene	Phenotype	Variant type	Known[a] germline/ novel	VAF	Origin or affected cell types	Refs
FAS	ALPS[c]	D-S	Known/same effect[b]/ novel	Leukocyte subsets: <1–10% DN-T: 50%	Pluripotent origin, HSC origin	[3,8,9,20]
	ALPS	R-S/AR-G LOH-S/AR-G	Novel	Leukocyte subsets: 7–35% DN-T: 50%/100% (if LOH)	Pluripotent origin, DN-T restricted	[10]
	Evans syndrome	D-S	Known somatic in ALPS	NA	NA	[75]
NLRP3	CAPS (CINCA, MWS)	D-S	Known/novel	Peripheral blood/PBMCs: 4.2–35.8%	HSC origin or earlier	[4,11–16]
	CAPS adult onset	D-S	Known/novel	Myeloid subsets: 6.6–44.8%	HSC origin or earlier, myeloid progenitor origin	[13,17–19]
NLRC4	CAPS (CINCA)	D-S	Novel	PBMCs: 30.3%	Pluripotent origin	[54]
TNFRSF1A	TRAPS	D-GS	Novel	Leukocyte subsets: 18–30%	Pluripotent origin	[66]
NOD2	Blau syndrome	D-S D-GS	Known	Peripheral blood: 7.7–12.9% Leukocyte subsets: 7.3–11%	Pluripotent origin	[13,25,76]
TMEM173	SAVI	D-S	Known	NA	Pluripotent origin	[77]
CYBB	CGD	X-S	Novel/known	Myeloid subsets: ~50%	Granulocytes + monocytes affected	[78,79]
STAT3	HIES	D-GS	Novel	NA	NA	[80]
STAT5B	Autoimmunity and allergy	D-S	Novel	Leukocyte subsets: 10.4–46.1%	HSC origin	[22]
KRAS	Multiple autoimmunity	D-S	Known somatic in RALD and cancer	NA	T + B + NK affected	[75,81]
	RALD	D-S	Known somatic in cancer	NA	HSC origin	[82–84]
NRAS	RALD	D-S	Known somatic in cancer	Peripheral blood + bone marrow: ~50%	Pluripotent origin	[85]

Table 1. Overview of Known Genes Affected by Somatic Mutations in Monogenic Immune Disorders

[a]Indicates whether the same variants have been observed as germline variants leading to the same disease phenotype, except when indicated differently.

[b]Different nucleotide change resulting in the same amino acid change or with the same effect on splicing has been observed previously as a germline variant leading to the same phenotype.

[c]Abbreviations: AD, autosomal dominant; ALPS, autoimmune lymphoproliferative syndrome; AR, autosomal recessive; B, B cell; CAPS, cryopyrin-associated autoinflammatory syndrome; CGD, chronic granulomatous disease; CINCA syndrome, chronic infantile neurological, cutaneous, and articular syndrome; D, dominant; DN-T, double-negative ($CD4^{-}CD8^{-}$) T cell; G, germline; GS, gonosomal; HIES, hyper IgE syndrome; HSC, hematopoietic stem cell; LOH, loss of heterozygosity; MWS, Muckle–Wells syndrome; NA, not available; NK, natural killer; PBMCs, peripheral blood mononuclear cells; R, recessive; RALD, RAS-associated ALPS-like disease; S, somatic; SAVI, STING-associated vasculopathy with onset in infancy; T, T cell; TRAPS, tumor-necrosis-factor-receptor-associated periodic syndrome; X, X-linked.

germline and soma lead to gonosomal mosaicism [21], and have a high probability of being transmitted to offspring as germline variants.

Early occurrence and/or strong selection may lead to somatic variants with VAFs around 50%, and render these indistinguishable from germline *de novo* variants, with 50% VAF expectations, unless

Box 2. Challenges in Identifying Somatic Variants in Immune Cells

In the 1990s, the occurrence of somatic mutations in immune cells was determined with clonal assays in which disruption of genes such as *HPRT* by a somatic mutation confers a measurable phenotype to a cell [42]. Sanger sequencing has anecdotally identified somatic mutations but generally lacks the resolution now offered by deep second-generation sequencing. Both high and low ends of the VAF spectrum indeed pose technical challenges in an immune context, and may have important clinical consequences in genetic counseling. The required sequencing depth is determined by VAF to be identified (e.g., lower VAFs) and accuracy with which VAF needs to be determined (e.g., higher VAFs to be distinguished from 50% germline expectations). Hence, a balance needs to be found between sequencing depth and extent of sequence covered. Exome and whole-genome screening approaches are able to identify somatic variants in total blood, but are typically limited to high VAFs (>10%). In order to identify variants with lower VAF, as low as 1%, targeted sequencing remains most cost-efficient.

Approaches to determine somatic variants in total blood or those of early embryonic origin have relied on deviations from the expected VAF of 50% for heterozygous germline variants [29]. Identification of somatic variants in a specific cell type, in contrast, can be based on comparison with a well-selected different cell type or lineage assumed not to be affected by the same somatic variants [47], replacing the pairwise tumor–normal comparison often applied in cancer.

Mutation callers originally developed for cancer have good sensitivity to call somatic variants, but positive predictive values remain unacceptably low for the lower VAFs typically seen in a noncancer context [71]. Artefacts, which may be the result of DNA contamination, DNA damage, sequencing errors, and read–mapping problems, are typically in the same low VAF range as true somatic variants [5]. Moreover, the somatic mutation rate in normal samples is lower than in cancer, thereby causing artefacts to outnumber somatic variants. Together, these two challenges mean that studies of somatic variants in normal immune cells are substantially less tolerant of false positives, and appropriate study design and validation is essential. Bioinformatics approaches combining different somatic variant callers and adding technical filtering criteria increase true positive or replication rates to >50% [47]. Advanced sequencing and analytical approaches are increasingly becoming available [5]. For a more extensive discussion on the causes underlying the challenge of distinguishing somatic variants with low VAF in normal and nontumor disease tissues from artefacts and on appropriate study designs, we would like to refer the reader to a separate review in this journal [5].

deep sequencing is applied for sufficient resolution or multiple cell subsets are investigated. As an example, a *NOD2* PID mutation presumed as *de novo* based on Sanger sequencing in a PID patient turned out to be a postzygotic somatic variant with high VAF (40.5%) in blood upon deep amplicon-based sequencing [13]. Two patients with autoimmune and allergy symptoms shared a somatic gain-of-function *STAT5B* mutation reaching VAFs of 46% in T cells, with lower VAFs in B cells and dendritic cells and absence in other tissues suggesting a strong T cell survival advantage [22].

At least 4% of germline presumed *de novo* variants may be mosaic in >1% of parental blood [23] (Figure 2). The first report of vertical transmission of a presumed *de novo* variant described a MWS patient inheriting an *NLRP3* mutation from an asymptomatic mother, for whom gonosomal mosaicism with low VAF in blood (2–3%) was demonstrated [24]. In a family with Blau syndrome, the father with late-onset mild disease had gonosomal mosaicism for a *NOD2* mutation (VAF in blood 13%), which was inherited into the germline of his two daughters, leading to early-onset and more-severe disease [25]. Recently, a systematic reinvestigation of 92 PID families with presumed *de novo* mutations found evidence for gonosomal mosaicism in the parents in six families (7%), most of them unaffected and some mildly affected [13]. This vertical transmission explains the unexpected recurrence of disease in families.

Not Always the Bad Guys: Revertant Mosaicism

As somatic variants arise frequently, occasionally they occur in a gene in which a germline disease mutation is present and entirely or partially abolish its consequences in affected cells (**revertant mosaicism**). Multiple revertant T cells are often seen in patients with inherited disease mutations in genes like *ADA*, *IL2RG*, and CD3-ζ (*CD247*) [2]. Revertant mosaicism is estimated at around 10% in Wiskott–Aldrich syndrome (WAS) patients [26]. Many revertants arise independently, as multiple independent

second-site revertant mutations are observed [27]. Revertant mutations offer a substantial survival and/or proliferating advantage to the affected cells [28], and the time point at which a revertant mutation arises is expected to play a role in whether or not the disease phenotype is alleviated or altered.

Adding to the Complexity of Multifactorial Diseases

The occurrence of somatic variants increases with age, and hematopoietic stem cells (HSCs) accumulate somatic variants during their life history, most with no apparent effect on cellular phenotype [29–31]. However, in around 12.5% of the overall healthy population, and especially in older age groups, an extreme number of recurrent somatic variants originate in long-lived HSCs that expand (Box 1) [32]. As mutated stem cells maintain the ability to differentiate in circulating granulocytes, monocytes, and lymphocytes, clones affect both the myeloid and lymphoid lineage [33] and are detectable in blood with deep sequencing methods. This phenomenon is called **clonal hematopoiesis of indeterminate potential (CHIP)**. CHIP **driver mutations** in known immunoproliferative cancer genes, such as *DNMT3A, TET2,* and *ASXL1*, are detected in a subset of individuals [29–36], but CHIP can also occur without known driver mutations [29,32]. Individuals with CHIP, but absence of other hematological abnormalities, have a 2–13 times higher probability of developing hematological cancers [29,30,32]. Of more interest for this review, these individuals are also at increased risk [with odds ratio (OR) = 1.2–2] for cardiovascular diseases and possibly psychiatric diseases, and of overall increased mortality [30,32,35] (Figure 2). This increased susceptibility has been attributed to inflammatory aspects [32,37,38].

A link between somatic variants and increased risk of autoimmunity was first suggested in immunoproliferative disorders (Figure 2). As a first example, large granular lymphocytic leukemia (LGL) patients with somatic *STAT3* mutations in $CD8^+$ T cells develop rheumatoid arthritis (RA) four times more frequently than patients without these mutations [39,40]. Secondly, the extent and abundance of *BRAF* V600E somatic mutations explains part of the heterogeneity in Langerhans cell histiocytosis (LCH). The 5% of LCH patients who additionally develop neurodegenerative disease have a significantly higher frequency of *BRAF* $V600E^+$ peripheral blood mononuclear cells compared with those who did not (OR = 8) [41]. In postmortem tissue of these patients, *BRAF* $V600E^+$ myeloid/monocytic cells are associated with white matter injury, gliosis, and demyelination, and a leukoencephalopathic pattern of neurodegeneration along with infiltrating lymphocytes [41].

Studies from the 1990s based on the ***HPRT* assay** [42] demonstrate that T cells from patients with **autoimmune diseases** have an increased somatic mutation rate compared with T cells from control individuals without autoimmune disease [43,44]. Whereas *HPRT* cell-based assays are a measure of overall somatic mutation rates, the technology to detect and identify specific somatic variants *ex vivo* only emerged recently, with the advent of second-generation sequencing. First studies applying targeted sequencing for the systematic screening of autoimmune disease patients have now been reported. Somatic variants in clonally expanded cytotoxic $CD8^+$ T lymphocytes are observed in around 20% of newly diagnosed and untreated RA patients [45]. In multiple sclerosis (MS), no large clonal expansions are observed, but two studies agree that approximately 60% of patients carry somatic variants with a median VAF of around 1% in subpopulations of immune cells (T/$CD8^+$/$CD4^+$/B cells) [46,47]. Variants arise developmentally late, and occur, for example, in $CD8^+$ but not in $CD4^+$ T cells, whereas VAF in blood would be largely undetectable at <0.5% [46,47]. Systematic comparisons of the occurrence, VAF, and other characteristics in patients versus healthy controls are currently limited [45], hampering conclusions on the contribution of somatic variants to disease pathogenesis.

Both in MS and RA, somatic variants occur most frequently in $CD8^+$ T cells, which are known to harbor more clonal expansions [45–47]. Positive cells are typically characterized by a memory phenotype [45,47] and a survival- and proliferation-gene expression signature [45], which is particularly biologically relevant from an autoimmune perspective [48,49]. Although sample sizes have been too small to perform systematic pathway analyses, a subset of variants occurs in genes expressed in immune cells, and in known regulators of (auto-)immunity and cell proliferation [45–47].

Somatic variants may not only be relevant in the context of disease but may influence quantitative traits within the population. Examples of noncoding somatic variants acting as expression quantitative trait loci in cancer have indeed been described [50,51], and could extend to somatic variants in normal immune cells.

In summary, complex diseases have often been attributed to genetic, environmental, and so-called stochastic factors. Somatic variants can be seen as stochastic effects of DNA replication and/or as potential mechanisms of action for environmental factors, and there is now growing evidence that such variants may contribute in a substantial way to a range of traits.

From Variants to Function

The relevance of somatic mosaicism to the onset of both monogenic and complex diseases is becoming increasingly clear through sequence-based approaches. Somatic variants accumulate at a significantly slower pace with increasing age in protein-coding genes compared with the overall genome, suggesting that many random somatic variants are damaging to cellular function [52]. A range of gene-level and variant-level bioinformatics tools is available for prioritizing disease-relevant germline variants. Limited data suggest these tools may be applicable for somatic variants as well, but more research is required (Box 3). However, direct evidence that the small subsets of cells carrying a somatic variant have disease-specific characteristics, and can thereby cause or contribute to disease, is lacking in these experiments. Several experimental functional approaches are currently being undertaken and focus mostly on coding variants. Beyond the coding region, somatic variants in lymphocytes are increased fivefold in active open chromatin regions and may affect expression [52].

In vitro studies demonstrate similar or even more pronounced effects for somatic compared with germline PID/AID mutations [11,14]. Genetic heterogeneity with mutant cells showing altered protein expression or phosphorylation can be detected by *ex vivo* intracellular staining [45]. *In vivo* animal models allow investigating effects restricted to specific cell types. When occurring in HSCs, BRAF

Box 3. Prioritization Strategies for Biological Impact

For prioritizing potentially immune disease-relevant germline variants, gene-level strategies based on cumulative mutational damage to a particular human gene in the general population are typically applied [72]. They are used to remove genes that are highly mutated in the general population and, therefore, less likely to cause severe disease. It has been noted that most genes in which mosaicism is known to cause disease have a moderate-to-high intolerance to functional genetic variation [13]. The fact that *NLRP3* is among the 10% of most intolerant genes has been suggested as an explanation for this gene having the highest number of detected postzygotic variants amongst PID patients. The 2% of genes in the genome with highest cumulative mutational damage (gene damage index; GDI) contribute a large proportion of the rare germline variants seen in the general population [73]. For somatic variants detected in complex autoimmune diseases, it was similarly observed that 16–28% are located in genes with a high GDI, whereas such genes only correspond to 3–5% of genes sequenced [47]. This suggests such variants may receive lower prioritization for biological importance.

For germline variants, the pathogenicity is furthermore often supported by variant-level strategies, such as prediction to be damaging by a high combined annotation dependent depletion (CADD) score adjusted by mutation significance cutoff (MSC) or prediction to be conserved by a high genomic evolutionary rate profiling (GERP++) score [72]. The pathogenicity of somatic variants in monogenic immune disorders has similarly been supported by high CADD scores [13]. Somatic variants observed in immune cells of MS patients were significantly more novel or rare and trend towards being more damaging and located at more conserved positions [47].

In summary, for germline variants both gene- and variant-level approaches are used in synergy to create a phenotypic impact gradient: benign variants of highly damaged genes have the lowest predicted impact, and putatively damaging variants of genes with low levels of damage are predicted to have the highest impact [73]. Further investigations are needed to establish whether these prediction tools can be extended to somatic variants.

V600E leads to a highly penetrant immune-proliferative phenotype in the bone marrow, spleen, and lungs in mice [53]. Mosaicism for the same mutation restricted to brain-resident macrophages, in contrast, results in clonal expansion of activated microglia, causing lethal neurodegenerative disease in adult mice [53], corroborated by human data described above [41]. In mice, *TET2* deficiency recapitulates some effects of CHIP, such as accelerating atherosclerosis development, and appears to do so through its effect on histone deacetylation in the macrophages leading to increased IL-1β production, increased P-selectin expression, and more recruitment of myeloid cells into the vascular wall [35,37]. However, the precise mechanisms underlying increased risk of disease for CHIP and CHIP-driver mutations in humans remain as yet unknown [38].

Newer approaches allow investigating genotype–phenotype correlations in the individual at the single cell level. iPSC lines can be directly reprogrammed from somatic cells into a range of cell types relevant to disease pathogenesis. As each iPSC clone originates from a single somatic cell, mutant and wild-type clones can be compared against an otherwise similar genetic background from a single mosaic patient. As an example, the cytokine hallmark of autoinflammation was observed only in *NLRC4*- or *NLRP3*-mutant but not wild-type macrophages derived from iPSC-derived immortalized myeloid cell lines [54,55]. This methodology could subsequently be used for drug screening [55]. Whereas second-generation sequencing has mainly been applied on **bulk DNA, single-cell sequencing** can link genetic and transcriptomic information in single cells. This could reveal whether transcriptional heterogeneity *ex vivo* is explained by mutational heterogeneity, and if so may identify expression signatures associated with somatic variants and cell surface markers that can be used to isolate specific cells for downstream studies [56].

Towards Future Integration of the Germline and Somatic Genomes

The majority of somatic variants appear to be due to 'bad luck'; that is, random mutations arising during DNA replication and most strongly being correlated with ageing and the number of divisions of normal self-renewing cells [32,57]. Environmental factors such as smoking [32], UV radiation [58], and Epstein–Barr virus infection [59] are also associated with the number of somatic variants. Surprisingly, somatic variants may also recur within families more often than expected, pointing to an inherited predisposition. As an example, sibling recurrence risks (λ_s) of up to 2.7 are seen for somatic variants in the *TET2* gene associated with CHIP [36]. A GWAS in 11 262 Icelanders indeed identified germline variants, for example, in the *TERT* gene, associated with an increased risk (OR 1.2) of CHIP [32]. This indicates that the emerging impact of germline variants on somatic events in tumors, as reviewed elsewhere in this journal [60], may also extend to somatic variants in healthy individuals and noncancer diseases.

Evidence is increasing that both germline and somatic variants are contributing to a range of traits, and even interactions between germline and somatic variants on disease outcome have been reported in cancer [61]. Over the past decade, GWASs using genotyping arrays have been highly successful in identifying germline variants associated with complex traits. Most somatic variants, even when recurrent across individuals, will likely be rare within the population. Hence, as for rare germline variants, there is a need for sequencing and for aggregating variants in a systematic approach; for example, within genomic windows or in gene burden tests. Whereas most common disease-associated variants are noncoding, the search for somatic variants has so far focused mostly on coding variants. The importance of noncoding somatic variants is increasingly being realized [62], but would require deep sequencing across large parts of or the entire genome becoming affordable (Box 2). The list of both germline and somatic variants emerging from individual or aggregate association tests may be combined for computationally based functional prioritization and interpretation, such as pathway-based analyses and fine-mapping approaches [62].

Although currently available data preclude a systematic analysis, germline and somatic variants in immune cells may indeed affect the same genes or pathways in both monogenic and complex diseases. One example is *STAT3*, where both germline [63–65] and somatic variants [39,40] are associated with risk of autoimmune disease, and somatic variants, some leading to increased STAT3 phosphorylation and activation, have been observed in $CD8^+$ T cells of autoimmune disease patients [45,46].

In summary, germline and somatic variants have so far been studied separately. In future, integrating germline and somatic genome data may provide novel insights into disease pathogenesis and offer clinical implications.

Clinical Applications in Diagnosis and Genetic Counseling

Data reviewed so far indicate that somatic mosaicism is a relevant and under-recognized mechanism underlying monogenic immune disorders, and may explain up to 25% of cases [13]. Receiving a correct genetic diagnosis has important implications for these patients and their families. A correct diagnosis may enable initiation and reimbursement of appropriate therapies. Examples of patients with somatic mutations demonstrate that the same therapies used for patients with corresponding germline mutations, such as anti-IL-1, lead to marked and sustained improvement [12,18,19,66]. For other somatic variants, such as *BRAF* or *STAT3*, it may be possible to develop new therapies inhibiting the affected pathway [45,53].

The distinction between a germline variant presumed to have occurred *de novo* and either a somatic variant with high VAF or vertical transmission from a parent with gonadal or gonosomal mosaicism is not always easy or even possible to make [23], but is particularly important for genetic counseling. The overall probability for a presumed *de novo* variant identified in a proband to be shared with a sibling is estimated at around 1%. However, this increases substantially to 24% for variants that turn out to be mosaic in >1% of parental blood cells and may be up to 50% for those mosaic in >6% of parental blood cells [23]. Deep sequencing of parental blood or other tissues for pathogenic somatic mutations seen in children may enable meaningful counseling on disease risks in future pregnancies.

Potential as Biomarkers Guiding Treatment

Limited data are available on the stability of somatic variants over time. In the subset of healthy controls with CHIP or patients with monogenic or complex immune diseases for whom up to 10-year longitudinal follow-up is available, variants overall remain present with strong stability of VAFs [3,12,13,19,30,46,47]. In a few CAPS patients, VAFs either rose by ninefold over 12 years in parallel with progressively more severe disease and increasing treatment dosages of anti-IL-1 required [19], or VAFs of >20% dropped by half, possibly related to the effectiveness of anti-IL-1 therapy [13]. In an MS patient, a mutated $CD8^+$ clone increased sevenfold from 2% to 16% upon change of treatment to fingolimod [46]. This suggests that treatment, retaining T lymphocytes in the lymph nodes, targeted other $CD8^+$ populations more efficiently than the mutant clone.

The examples above illustrate a general application, most evident in monogenic disorders but possibly extending into complex immune diseases. If somatic variants – whether drivers or passengers – tag cells involved in the disease pathogenesis, they offer the opportunity of follow-up over time. For this purpose, technologies such as droplet digital PCR allow accurate quantifications of VAFs as low as 0.1% [47]. Changes in VAF may be used to detect whether treatment effectively depletes cells carrying somatic variants and when a subsequent treatment administration, increased dosage or treatment change is required.

Concluding Remarks and Future Perspectives

Individuals are increasingly recognized as complex mosaics of genetically distinct cells resulting from the occurrence of somatic variants during embryonal development or throughout life, and future studies will need to take this concept into account (see Outstanding Questions). Relatively abundant somatic variants in immune cells have recently been shown as an important cause of monogenic immune disorders, with important clinical implications. They have also been associated with 1.2–10-fold increased risk of a range of complex diseases with inflammatory aspects. Emerging technological advances increasingly enable the detection of less abundant somatic variants. Single-cell approaches are required to establish whether these variants confer an altered cellular phenotype to the small subsets of cells affected. Examples suggest somatic variants may affect an individual's phenotype. However, the overall occurrence and consequences of somatic variants in healthy individuals and their

contribution to disease is not currently understood. Large-scale germline genome data have over the past 10 years substantially advanced our understanding of genetic determinants for monogenic and complex diseases. Integrating somatic genome data may in future provide novel insights into disease pathogenesis and offer clinical implications.

Outstanding Questions

Will somatic variants resolve more unexplained cases with monogenic disorders?

How can technological and analytical improvements overcome remaining challenges to study individuals as complex mosaics of genetically distinct cells?

Can large-scale case–control comparisons for the occurrence and the characteristics of somatic variants provide more evidence for their contribution to susceptibility for or heterogeneity in complex diseases?

Can single-cell approaches provide the resolution to capture clonal evolution and demonstrate an altered, for example, more autoreactive or proinflammatory, phenotype of affected cells?

How do the germline and the somatic genomes within an individual interact?

Can somatic variants provide a mechanism for environmental factors to contribute to disease?

Acknowledgments

L.V.H. is a PhD Fellow of the Belgian Charcot Foundation, and B.D. is a Clinical Investigator of the Research Foundation Flanders (FWO-Vlaanderen). The authors have been supported for their work on this topic by the Research Fund KU Leuven (CREA14/023 and C24/16/045), the Research Foundation Flanders (G0A73415), the Belgian Charcot Foundation, and Roche. We thank Dr Stephanie Humblet-Baron as well as members of our research group, in particular Marijne Vandebergh and Dr Emanuela Oldoni, for critical reading and discussion of the draft manuscript.

References

1. Hall, J.G. (1988) Review and hypotheses: somatic mosaicism: observations related to clinical genetics. *Am. J. Hum. Genet.* 43, 355–363
2. Goodnow, C.C. (2007) Multistep pathogenesis of autoimmune disease. *Cell* 130, 25–35
3. Holzelova, E. *et al.* (2004) Autoimmune lymphoproliferative syndrome with somatic Fas mutations. *N. Engl. J. Med.* 351, 1409–1418
4. Saito, M. *et al.* (2005) Somatic mosaicism of CIAS1 in a patient with chronic infantile neurologic, cutaneous, articular syndrome. *Arthritis Rheum.* 52, 3579–3585
5. Dou, Y. *et al.* (2018) Detecting somatic mutations in normal cells. *Trends Genet.* 34, 545–557
6. Lim, Y.H. *et al.* (2017) Mosaicism in cutaneous disorders. *Annu. Rev. Genet.* 51, 123–141
7. D'Gama, A.M. and Walsh, C.A. (2018) Somatic mosaicism and neurodevelopmental disease. *Nat. Neurosci.* 21, 1504–1514
8. Dowdell, K.C. *et al.* (2010) Somatic FAS mutations are common in patients with genetically undefined autoimmune lymphoproliferative syndrome. *Blood* 115, 5164–5169
9. Neven, B. *et al.* (2011) A survey of 90 patients with autoimmune lymphoproliferative syndrome related to TNFRSF6 mutation. *Blood* 118, 4798–4807
10. Magerus-Chatinet, A. *et al.* (2011) Onset of autoimmune lymphoproliferative syndrome (ALPS) in humans as a consequence of genetic defect accumulation. *J. Clin. Invest.* 121, 106–112
11. Tanaka, N. *et al.* (2011) High incidence of NLRP3 somatic mosaicism in patients with chronic infantile neurologic, cutaneous, articular syndrome: results of an International Multicenter Collaborative Study. *Arthritis Rheum.* 63, 3625–3632
12. Nakagawa, K. *et al.* (2015) Somatic NLRP3 mosaicism in Muckle-Wells syndrome. A genetic mechanism shared by different phenotypes of cryopyrin-associated periodic syndromes. *Ann. Rheum. Dis.* 74, 603–610
13. Mensa-Vilaro, A. *et al.* (2019) Unexpected relevant role of gene mosaicism in patients with primary immunodeficiency diseases. *J. Allergy Clin. Immunol.* 143, 359–368
14. Saito, M. *et al.* (2008) Disease-associated CIAS1 mutations induce monocyte death, revealing low-level mosaicism in mutation-negative cryopyrin-associated periodic syndrome patients. *Blood* 111, 2132–2141
15. Arostegui, J.I. *et al.* (2010) A somatic NLRP3 mutation as a cause of a sporadic case of chronic infantile neurologic, cutaneous, articular syndrome/neonatal-onset multisystem inflammatory disease: novel evidence of the role of low-level mosaicism as the pathophysiologic mechanism underlying mendelian inherited diseases. *Arthritis Rheumatol.* 62, 1158–1166
16. Omoyinmi, E. *et al.* (2014) Brief report: whole-exome sequencing revealing somatic NLRP3 mosaicism in a patient with chronic infantile neurologic, cutaneous, articular syndrome. *Arthritis Rheumatol.* 66, 197–202
17. Zhou, Q. *et al.* (2015) Brief report: cryopyrin-associated periodic syndrome caused by a myeloid-restricted somatic NLRP3 mutation. *Arthritis Rheumatol.* 67, 2482–2486
18. Mensa-Vilaro, A. *et al.* (2016) Brief report: late-onset cryopyrin-associated periodic syndrome due to myeloid-restricted somatic NLRP3 mosaicism. *Arthritis Rheumatol.* 68, 3035–3041
19. Rowczenio, D.M. *et al.* (2017) Late-onset cryopyrin-associated periodic syndromes caused by somatic NLRP3 mosaicism-UK single center experience. *Front. Immunol.* 8, 1410
20. Martinez-Feito, A. *et al.* (2016) Autoimmune lymphoproliferative syndrome due to somatic FAS mutation (ALPS-sFAS) combined with a germline caspase-10 (CASP10) variation. *Immunobiology* 221, 40–47
21. Ju, Y.S. *et al.* (2017) Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* 543, 714–718
22. Ma, C.A. *et al.* (2017) Somatic STAT5b gain-of-function mutations in early onset nonclonal eosinophilia, urticaria, dermatitis, and diarrhea. *Blood* 129, 650–653
23. Rahbari, R. *et al.* (2016) Timing, rates and spectra of human germline mutation. *Nat. Genet.* 48, 126–133
24. Jimenez-Trevino, S. *et al.* (2013) First report of vertical transmission of a somatic NLRP3 mutation in cryopyrin-associated periodic syndromes. *Ann. Rheum. Dis.* 72, 1109–1110
25. Mensa-Vilaro, A. *et al.* (2016) Brief report: first identification of intrafamilial recurrence of Blau syndrome due to gonosomal NOD2 mosaicism. *Arthritis Rheumatol.* 68, 1039–1044
26. Stewart, D.M. *et al.* (2007) X-linked hypogammaglobulinemia and isolated growth hormone deficiency: an update. *Immunol. Res.* 38, 391–399
27. Boztug, K. *et al.* (2008) Multiple independent second-site mutations in two siblings with somatic mosaicism for Wiskott-Aldrich syndrome. *Clin. Genet.* 74, 68–74
28. Humblet-Baron, S. *et al.* (2007) Wiskott-Aldrich syndrome protein is required for regulatory T cell homeostasis. *J. Clin. Invest.* 117, 407–418
29. Genovese, G. *et al.* (2014) Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* 371, 2477–2487

30. Jaiswal, S. *et al.* (2014) Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* 371, 2488–2498
31. Xie, M. *et al.* (2014) Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* 20, 1472–1478
32. Zink, F. *et al.* (2017) Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* 130, 742–752
33. Young, A.L. *et al.* (2016) Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* 7, 12484
34. Busque, L. *et al.* (2012) Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* 44, 1179–1181
35. Jaiswal, S. *et al.* (2017) Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* 377, 111–121
36. Buscarlet, M. *et al.* (2017) DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* 130, 753–762
37. Fuster, J.J. *et al.* (2017) Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* 355, 842–847
38. Sano, S. *et al.* (2018) Clonal hematopoiesis and its impact on cardiovascular disease. *Circ. J.* 83, 2–11
39. Koskela, H.L. *et al.* (2012) Somatic STAT3 mutations in large granular lymphocytic leukemia. *N. Engl. J. Med.* 366, 1905–1913
40. Rajala, H.L. *et al.* (2015) The analysis of clonal diversity and therapy responses using STAT3 mutations as a molecular marker in large granular lymphocytic leukemia. *Haematologica* 100, 91–99
41. McClain, K.L. *et al.* (2018) CNS Langerhans cell histiocytosis: common hematopoietic origin for LCH-associated neurodegeneration and mass lesions. *Cancer* 124, 2607–2620
42. Albertini, R.J. *et al.* (1982) T-cell cloning to detect the mutant 6-thioguanine-resistant lymphocytes present in human peripheral blood. *Proc. Natl. Acad. Sci. U S A* 79, 6617–6621
43. Allegretta, M. *et al.* (1990) T cells responsive to myelin basic protein in patients with multiple sclerosis. *Science* 247, 718–721
44. Cannons, J.L. *et al.* (1998) HPRT- mutant T cells in the peripheral blood and synovial tissue of patients with rheumatoid arthritis. *Arthritis Rheumatol.* 41, 1772–1782
45. Savola, P. *et al.* (2017) Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat. Commun.* 8, 15869
46. Valori, M. *et al.* (2017) A novel class of somatic mutations in blood detected preferentially in CD8+ cells. *Clin. Immunol.* 175, 75–81
47. Van Horebeek, L. *et al.* (2018) A robust pipeline with high replication rate for detection of somatic variants in the adaptive immune system as a source of common genetic variation in autoimmune disease. *Hum. Mol. Genet.* 28, 1369–1380
48. Dooley, J. *et al.* (2016) Immunological profiles of multiple sclerosis treatments reveal shared early B cell alterations. *Neurol. Neuroimmunol. Neuroinflamm.* 3, e240
49. Battistini, L. *et al.* (2003) CD8+ T cells from patients with acute multiple sclerosis display selective increase of adhesiveness in brain venules: a critical role for P-selectin glycoprotein ligand-1. *Blood* 101, 4775–4782
50. Zhang, W. *et al.* (2018) A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. *Nat. Genet.* 50, 613–620
51. Huang, F.W. *et al.* (2013) Highly recurrent TERT promoter mutations in human melanoma. *Science* 339, 957–959
52. Zhang, L. *et al.* (2019) Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc. Natl. Acad. Sci. U S A* 116, 9014–9019
53. Mass, E. *et al.* (2017) A somatic mutation in erythro-myeloid progenitors causes neurodegenerative disease. *Nature* 549, 389–393
54. Kawasaki, Y. *et al.* (2017) Identification of a high-frequency somatic NLRC4 mutation as a cause of autoinflammation by pluripotent cell-based phenotype dissection. *Arthritis Rheumatol.* 69, 447–459
55. Tanaka, T. *et al.* (2012) Induced pluripotent stem cells from CINCA syndrome patients as a model for dissecting somatic mosaicism and drug discovery. *Blood* 120, 1299–1308
56. Nam, A.S. *et al.* (2019) Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* 571, 355–360
57. Tomasetti, C. and Vogelstein, B. (2015) Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 347, 78–81
58. Helleday, T. *et al.* (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* 15, 585–598
59. Kim, J.H. *et al.* (2013) Epstein-Barr virus latent membrane protein 1 increases genomic instability through Egr-1-mediated up-regulation of activation-induced cytidine deaminase in B-cell lymphoma. *Leuk. Lymphoma* 54, 2035–2040
60. Ramroop, J.R. *et al.* (2019) Germline variants impact somatic events during tumorigenesis. *Trends Genet.* 35, 515–526
61. Rachakonda, P.S. *et al.* (2013) TERT promoter mutations in bladder cancer affect patient survival and disease recurrence through modification by a common polymorphism. *Proc. Natl. Acad. Sci. U S A* 110, 17426–17431
62. Khurana, E. *et al.* (2016) Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* 17, 93–108
63. Jakkula, E. *et al.* (2010) Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am. J. Hum. Genet.* 86, 285–291
64. Barrett, J.C. *et al.* (2008) Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* 40, 955–962
65. Forbes, L.R. *et al.* (2016) Signal transducer and activator of transcription 3: a year in review. *Curr. Opin. Hematol.* 23, 23–27
66. Rowczenio, D.M. *et al.* (2016) Brief report: association of tumor necrosis factor receptor-associated periodic syndrome with gonosomal mosaicism of a novel 24-nucleotide TNFRSF1A deletion. *Arthritis Rheumatol.* 68, 2044–2049
67. Lynch, M. (2010) Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U S A* 107, 961–968
68. Tomasetti, C. *et al.* (2013) 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* 110, 1999–2004
69. Welch, J.S. *et al.* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150, 264–278
70. Holstege, H. *et al.* (2014) Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. *Genome Res.* 24, 733–742

71. Sandmann, S. *et al.* (2017) Evaluating variant calling tools for non-matched next-generation sequencing data. *Sci. Rep.* 7, 43169
72. Bucciol, G. *et al.* (2017) Whole exome sequencing in inborn errors of immunity: use the power but mind the limits. *Curr. Opin. Allergy Clin. Immunol.* 17, 421–430
73. Itan, Y. *et al.* (2015) The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc. Natl. Acad. Sci. U S A* 112, 13615–13620
74. McCarthy, M.I. *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369
75. Hadjadj, J. *et al.* (2019) Pediatric Evans syndrome is associated with a high frequency of potentially damaging variants in immune genes. *Blood* 134, 9–21
76. de Inocencio, J. *et al.* (2015) Somatic NOD2 mosaicism in Blau syndrome. *J. Allergy Clin. Immunol.* 136, 484–487 e2
77. Liu, Y. *et al.* (2014) Activated STING in a vascular and pulmonary syndrome. *N. Engl. J. Med.* 371, 507–518
78. Wolach, B. *et al.* (2005) Unusual late presentation of X-linked chronic granulomatous disease in an adult female with a somatic mosaic for a novel mutation in CYBB. *Blood* 105, 61–66
79. Yamada, M. *et al.* (2012) Somatic mosaicism in two unrelated patients with X-linked chronic granulomatous disease characterized by the presence of a small population of normal cells. *Gene* 497, 110–115
80. Holland, S.M. *et al.* (2007) STAT3 mutations in the hyper-IgE syndrome. *N. Engl. J. Med.* 357, 1608–1619
81. Moritake, H. *et al.* (2016) Autoimmunity including intestinal Behcet disease bearing the KRAS mutation in lymphocytes: a case report. *Pediatrics* 137, e20152891
82. Niemela, J.E. *et al.* (2011) Somatic KRAS mutations associated with a human nonmalignant syndrome of autoimmunity and abnormal leukocyte homeostasis. *Blood* 117, 2883–2886
83. Takagi, M. *et al.* (2011) Autoimmune lymphoproliferative syndrome-like disease with somatic KRAS mutation. *Blood* 117, 2887–2890
84. Lanzarotti, N. *et al.* (2014) RAS-associated lymphoproliferative disease evolves into severe juvenile myelo-monocytic leukemia. *Blood* 123, 1960–1963
85. Shiota, M. *et al.* (2015) Somatic mosaicism for a NRAS mutation associates with disparate clinical features in RAS-associated leukoproliferative disease: a report of two cases. *J. Clin. Immunol.* 35, 454–458