

SCIENTIFIC REPORTS



OPEN

The fetal thymus has a unique genomic copy number profile resulting from physiological T cell receptor gene rearrangement

Received: 06 October 2015

Accepted: 09 March 2016

Published: 24 March 2016

Anders Valind¹, C. Haikal¹, M.E.K. Klasson¹, M. C. Johansson², J. Gullander¹, M. Soller³, B. Baldetorp² & David Gisselsson^{1,4}

Somatic mosaicism, the presence of genetically distinct cells within an organism, has been increasingly associated with human morbidity, ranging from being a cause of rare syndromes to a risk factor for common disorders such as malignancy and cardiovascular disease. Previous studies interrogating the normal prevalence of somatic mosaicism have focused on adults. We here present an estimate of the baseline frequency of somatic mosaic copy number variation (CNV) at the time around birth, by sampling eight different organs from a total of five fetuses and newborns. Overall we find a significantly lower frequency of organ specific (i.e. mosaic) CNVs as compared to adults ($p = 0.003$; Mann-Whitney U-test). The rate of somatic CNV in adults has been estimated to around 2.2 CNV per organ assayed. In contrast, after stringent filtering, we found no organ-private CNVs in fetuses or newborns with exception of the thymus. This organ exhibited a specific genome profile in the form of deletions resulting from polyclonal T-cell receptor rearrangements. This implies that somatic non-immune related CNVs, if present at birth, are typically confined to very small cell populations within organs.

Somatic mosaicism is increasingly acknowledged as an important factor behind human disease^{1–4}. This is exemplified by recent publications connecting somatic copy number variants to an increased risk of cancer^{3,5} and to shorter overall survival⁶. To date, almost all published studies of somatic copy number variants (CNV) have focused on adults. It is therefore not known whether somatic variation at the copy number level is present at detectable levels already in utero, whether it is generated *de novo* after birth, or present in utero as low frequency clones that expand during the life of an individual. Another fundamental question is whether the baseline level of somatic CNV at birth differs between different parts of the human body. For example, it has been suggested that the human liver exhibits a high degree of somatic CNV in the form of aneuploidy even in small children⁷, but these data have not been independently validated. As there is a strong connection between age and the presence of somatic copy number variants in peripheral blood⁸, it is important to investigate whether this also holds for solid organs.

To address these issues, we set out to derive a first estimate of the level of somatic copy number mosaicism present around birth. To detect organ-private CNVs indicative of somatic variation^{9,10}, we analyzed tissue obtained at perinatal autopsy from eight different organs from four fetuses and one newborn using high resolution whole genome genotyping arrays (Table 1). We also mapped the levels of intra-organ aneuploidy using well-established *in situ* methods.

Results

The study was approved by the regional ethics review board as a specific project within a larger study on the role of CNVs in perinatal disease. Anonymized tissue samples were obtained at clinical autopsies of fetal and perinatal deaths, where cases were selected for studies of baseline CNV on the criteria that they should neither exhibit

¹Department of Clinical Genetics, Lund University BMCC13, Lund University, SE 221 84, Lund, Sweden. ²Department of Oncology, Lund University University Hospital, SE 221 85, Lund, Sweden. ³Department of Clinical Genetics, Skåne Regional and University Laboratories University Hospital, SE 221 85, Lund, Sweden. ⁴Department of Pathology, Skåne Regional and University Laboratories University Hospital, SE 221 85, Lund, Sweden. Correspondence and requests for materials should be addressed to A.V. (email: anders.valind@med.lu.se)

ID	Age (GW)	Sex	Cause of death	Clinical context	Organs analyzed after quality filtering								
					Lung	Muscle	Kidney	Heart	Adrenal	Spleen	Thymus	Liver	
1	24-25	Male	Hyaline membrane disease	Extreme prematurity									
2	24-28	Male	Placental abruption	Sudden intrauterine death									
3	28-31	Male	Heart failure	Down syndrome									
4	30-34	Male	Hydrancephaly	Disruption of cerebral vessels									
5	38-39	Male	Insufficient placental maturation	Sudden intrauterine death									

Table 1. Patient characteristics based on autopsy findings with age given in gestational weeks (GW), together with organs assayed for each patient. Blue cells in the table indicate organs that passed all quality controls and were analyzed by genomic array.

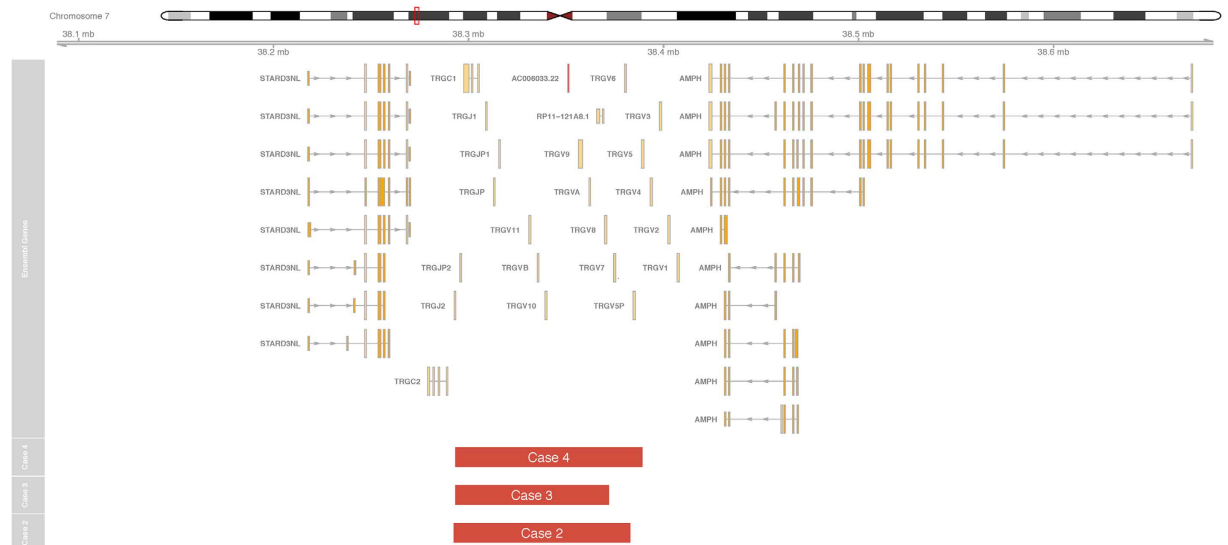


Figure 1. Recurrent *TCRG* deletions. RAG-mediated deletions at the T cell receptor gamma locus in 7p detected in three out of five individuals. The red bars denote the extension of the deletions in each of the cases using the stringent 50 probe threshold and the gene annotation track shows the location of *TCRG* genes.

developmental anatomic abnormalities, nor maceration threatening DNA quality. After strict filtering with respect to DNA quality and genotyping performance (See Methods for details) 3–8 tissue types (heart, lung, liver, spleen, thymus, kidney, adrenal gland, and muscle) per patient remained for analysis. In order to minimize the risk of false positives, we used a stringent calling method. In short, we ran each sample on two different high resolution genomic arrays (Affymetrix Cytoscan HD with a total of 2.67 million makers and Illumina Human Omni Quad 5 M with a total of 4.3 million markers) and retained only aberrations that spanned at least 50 consecutive probes and required a 50% reciprocal overlap between the segments from the two arrays to call a CNV. The latter criterion has previously been adopted by the 1000 Genomes project¹¹ and also used as a concordance measure for array-platform comparison¹². Using these calling criteria followed by manual visual inspection of the raw data, we detected a total of only eight true organ-private CNVs, all deletions, present in three of the patients. These eight variants were all found in the thymus and were situated at T cell receptor (*TCR*) loci, thus representing physiological RAG-mediated somatic rearrangements during T cell ontogeny. As a second step, we then re-analyzed the data by lowering the threshold for inclusion to ten consecutive probes. This resulted in more putative variants for cross validation and manual visual inspection (23 putative variants for visual inspection using 50 probes as a threshold compared 85 using 10 probes). However, after these validation steps only one additional CNV was retained. This consisted of a small 5 kb deletion located in the *TCRD* locus in one of the individuals (case 4).

The three CNV-positive thymus samples all carried deletions in the *TCRG* and *TCRD* loci (Figs 1 and 2). One sample also had a deletion at the *TCRB* locus. The deletions in the *TCRD* loci were similar in breakpoints to rearrangements found in lymphoid neoplasms¹³. However, all thymus samples were polyclonal at TCR clonality analysis, inferring that the detected deletions resulted from a variety of different somatic rearrangements leading to loss of overlapping genomic segments. A validation set of five additional thymus samples from a second set of fetuses from the third trimester all exhibited similar deletions in *TCRG* and *TCRD* (Table S1). Our findings of a preferential involvement of the *TCRG* and *TCRD* genes are in line with earlier work showing that most T cells have genomically reordered gamma/delta receptors even if they have an alpha/beta-phenotype¹⁴. It is also consistent with the sequential ordering of rearrangements of the T cell receptor loci, where *TCRD* and *TCRG* rearrange before *TCRB* and *TCRA*¹⁵. Even though our findings of thymus-specific TCR deletions are not entirely surprising considering that TCR rearrangements are well known to occur in utero, our data for the first time show a human somatic tissue type to have a recurrent unique copy number profile.

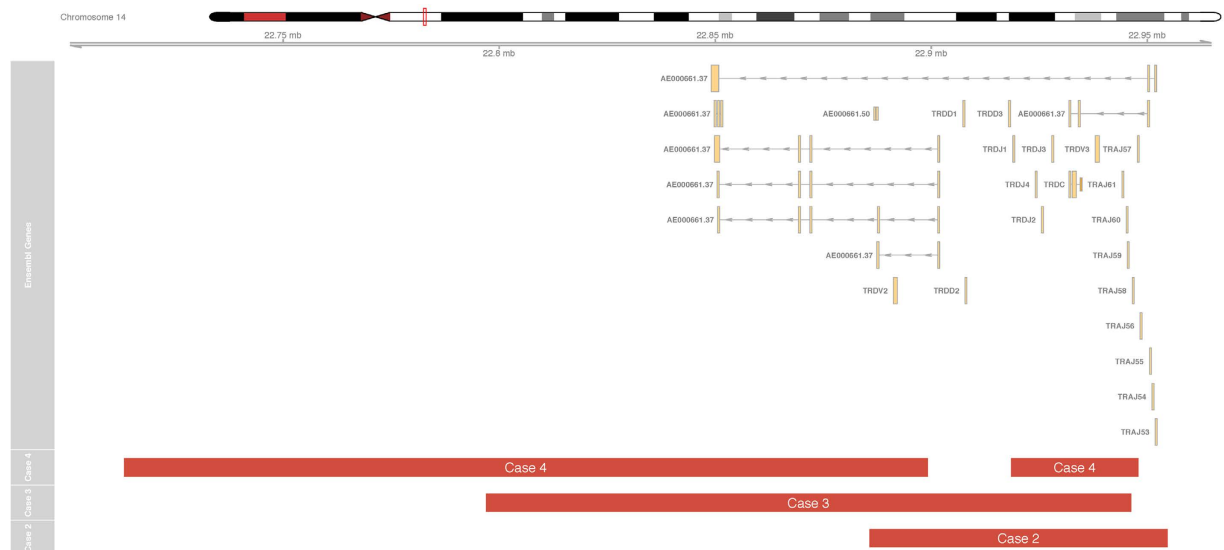


Figure 2. Recurrent *TCRD* deletions. RAG-mediated deletions at the T cell receptor delta locus in 14q detected in three out of five individuals using the stringent 50 probe threshold. Annotations are as in Fig. 1.

In addition to studying inter-organ variation, we also mapped intra-organ variation at the whole chromosome level in the three cases where cell nuclei were best preserved post mortem (cases 1, 3 and 4). The liver has been previously reported to harbor physiological chromosome copy number variation even in young individuals⁷. Using standard single color FISH we indeed found an elevated rate of aneuploidy in hepatocytes compared to cells in other organs (Fig. 3A–C). However, using a stringent, dual color FISH-method, previously shown to eliminate false positive noise signals¹⁶, liver cells failed to show an elevated prevalence of chromosome copy number changes. Instead false positive noise levels were considerably higher in the liver than in thymus samples used as a comparison (Fig. 3D). These findings are in concordance with a recent single cell sequencing study failing to show a higher frequency of aneuploidy in liver cells compared to keratinocytes¹⁷. In line with its role in post-natal hepatocyte differentiation and in contrast to adult livers¹⁸, there was no evidence of polyploidy in our fetal/neonatal liver samples, using either FISH or quantitative flow cytometry (DNA-FCM) (Fig. S1). We conclude that liver cells are no more aneuploid than most other somatic cells in fetuses and newborns. For the other organs assayed, the level of whole chromosome aneuploidy was at a level comparable with the noise inherent in single probe FISH (Fig. 3C).

Discussion

One important limitation of our study is that we were unable to obtain tissue samples from all major organ systems. In particular, there was a lack of samples from the central nervous system (CNS) as perinatal autopsy guidelines stipulate that the entire brain should be subjected to neuropathological analysis. Including tissue from the CNS was therefore not considered ethical within the frames of the present study protocol, as it could interfere with the diagnostic process. The lack of data from the CNS is particularly regrettable as several studies have shown a wide variety of somatic variation confined to the brain^{19–23}. Another weakness of the present study is its small sample size; this is both due to the difficulty of obtaining post mortem fetal tissue samples of sufficient quality for genome-wide analysis as well as the cost of that analysis. Nevertheless, our study contains a similar number of patients as the three previous studies of interorgan variation post mortem in adults^{9,10,24}. Finally it must be noticed that our study can not differentiate between a complete absence of somatic variation within an organ and its presence in a small fraction (<10%) of cells within that organ.

With these limitations in mind, our data still reinforce the notion that detectable somatic mosaicism at the copy number level is an age-related phenomenon and establishes that this applies also outside the hematopoietic system. We also conclude that earlier reports of aneuploidy in the human liver seem to be biased by method specific noise and that liver cell polyploidy appears to be a post-natal phenomenon. This is in line with earlier reports focusing on somatic copy number variation within a single tissue type, i.e. peripheral blood, showing that the presence of mosaic events increase linearly with age^{3,5,25}, although these studies only included samples from adult patients. There are very few studies focusing on somatic mosaicism within solid organs, possible due to the paucity of material^{9,10,24}. The only comparable study in terms of array resolution found 2.2 CNV/organ assayed⁹, a significantly higher frequency than we detected in tissues outside the thymus ($p = 0.003$; Mann-Whitney U-test). It is possible that our stringent cross-platform validation criteria led to failure to ascertain some CNVs due to their small size and/or low prevalence in tissue. However, the fact that we detected rearrangements in TCR loci down to 5 kb in the thymus, events that are subclonal within this organ, argues against this. As high resolution whole genome genotyping arrays detect subclonal variation present in as low as 10% of cells in the sample assayed²⁶, larger subpopulations carrying unique CNVs would most probably have been detected if present.

In the context of somatic mosaicism, the best-studied organ system besides human blood is probably the skin, assumedly due to the ease of getting access to material. Here, recent work has shown that smaller copy

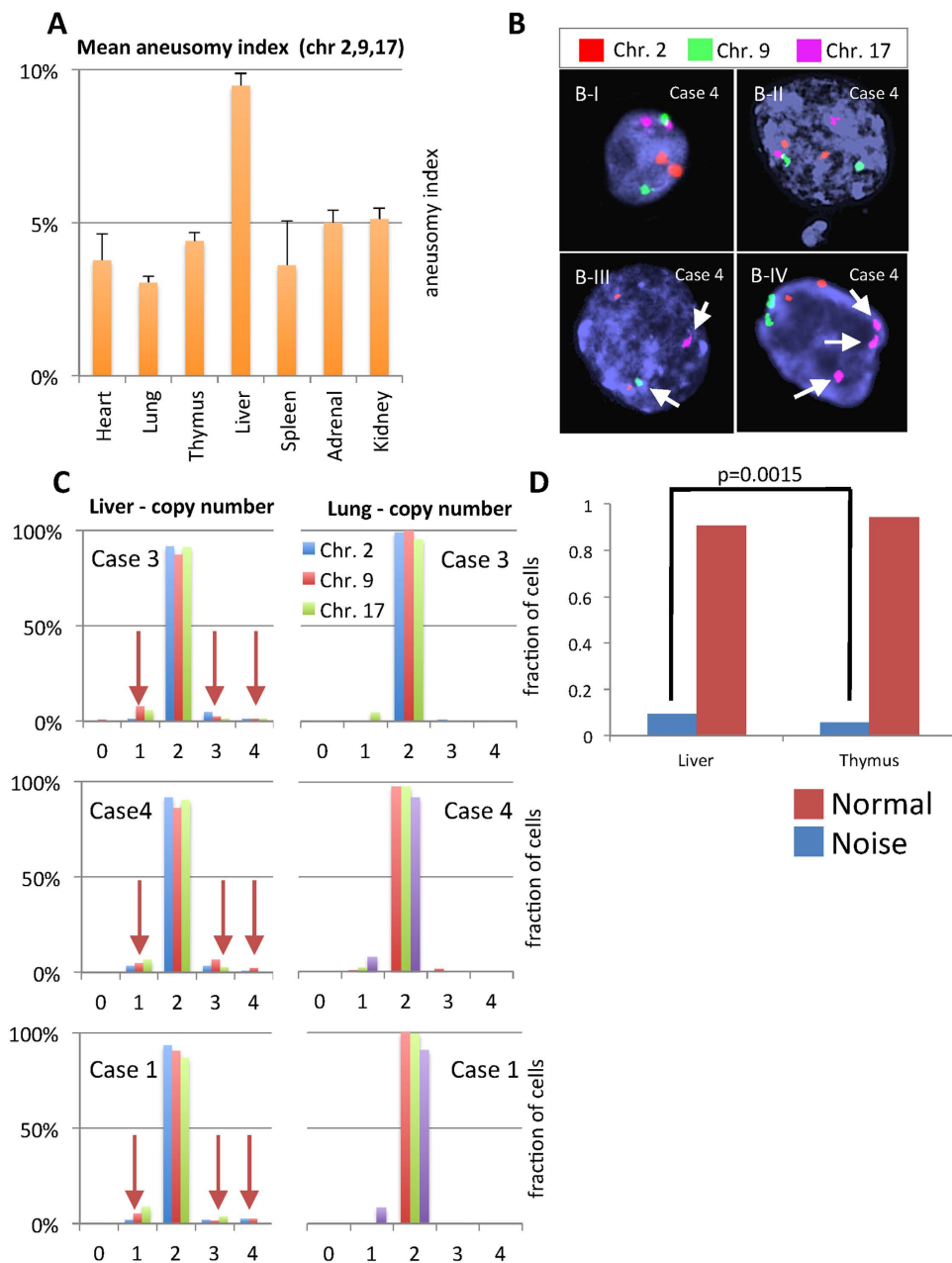


Figure 3. Intraorgan variation mapping using FISH. (A) Mean aneusomy index (AI) for all tissue types tested in cases 1, 3 and 4. For each tissue sample and chromosome, >200 nuclei were scored. The graph clearly shows a higher level of aneusomy in the liver. (B) Examples of FISH signal configurations that were scored as normal (B-I and B-II) as well as monosomic (B-III) and trisomic (B-IV) using the single probe method. (C) A comparison of chromosome number distributions for the liver (highest mean AI) and lung (lowest mean AI) for each of these cases, using single probe FISH. (D) The dual probe-based estimation of noise levels in the liver and the thymus for two samples (Case 1 and 4), showing that there is a significant increase (p-value from Fisher's exact test) of false positive signals in the liver compared to the thymus samples. No true positive (aneusomic cells were detected in either tissue) and the number of true negative (disomic cells) were similar.

number variants can be present at very low allele frequencies²⁷ and that the frequency of aneuploidy in human keratinocytes is at most a few percent¹⁷. The skin is presumably exposed to exogenous mutagens more often than other organs, resulting in a higher mutation rate and this may effect clonal dynamics²⁸, possibly causing a larger age-dependent accumulation effect of somatic mosaicism than in other organs. Interestingly, another known mutagen, cigarette smoke, has been associated with an increase in mosaic loss of chromosome Y in peripheral blood in men²⁹. Regarding somatic copy number variation at low frequencies at birth, recent studies have shown that low grade parental mosaicism can sometimes be transmitted, i.e. present in both the soma and the germline³⁰. This shows that—in situations associated with genetic disease – somatic copy number variants can form during the

very first stages of embryogenesis, followed by age dependent clonal expansion, possibly modulated by various environmental cues.

To summarize, we showed that there was no detectable somatic copy number mosaicism beyond physiological immunological rearrangements, in a total of 28 fetal/neonatal organs that passed stringent DNA quality criteria. Our data provide the first experimental evidence that detectable somatic CNVs are typically postnatal phenomena in individuals without congenital genetic disease.

Methods

Organ Sampling and DNA extraction. Autopsy, biobanking and genetic analyses were performed per clinical routine after informed consent from next of kin. All experiments were performed according to the ethical guidelines for research at the Lund University Faculty of Medicine. The study protocol was approved by the Regional Ethics Committee (No. L2012/405). Per this protocol, 5 mm tissue samples from multiple organs were immediately frozen under a serial number, which allowed safe transmission of basic clinical data (Table 1) for comparison to genomic results while keeping patient identity confidential. No case was included for research purposes only. Patients with malformation syndromes of potential genetic origin as well as severely macerated fetuses were excluded from the study. DNA was extracted from frozen samples using the DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany).

Quality Control on the Discovery Platform. The Affymetrix Cytoscan HD uses three different metrics to evaluate the quality of the generated data: SNPQC, the median absolute value pairwise difference (MAPD), and Waviness-SD. SNPQC is a metric measuring the separation of the genotype cluster that gives information on the quality of the genotyping for the subset of probes on the Cytoscan Array that are genotypeable. MAPD measures local variation in the log₂-ratio, while the Waviness-SD measures variation in log₂-ratio over larger genomic space, as described in the user manual for the Affymetrix Chromosome Analysis Suite software³¹. As the Waviness-SD metric is sensitive to the presence of true copy number variants we used only the MAPD and SNPQC while evaluating the quality of the generated data. All samples that were below the thresholds for MAPD and SNPQC as set by Affymetrix were excluded from downstream analysis. The MAPD and SNPQC values for all samples assayed are presented in Supplemental Table 2.

Calling process for organ specific CNVs. The Affymetrix Cytoscan HD data and the Illumina 5 M Omni Quad data were loaded into two platform specific Nexus Copy Number 7.5 projects, and then normalized and filtered according to the default setting for the corresponding array type. They were subsequently segmented using the FASTT2-SNP segmentation method using the standard parameter for calling thresholds. FASTT2-SNP is a Hidden Markov Model-based segmentation algorithm that incorporates both the Log₂-ratio information as well as the alternate allele frequency when calling copy number variants. All segments were then exported and, using custom python scripts, segments that spanned less than 50 (first analysis) and 10 (second analysis) consecutive markers were removed together with non-copy number variants (i.e. allelic imbalances and LOH-events). The filtered files were then converted to .bed files for downstream processing with the Bedtools-suite³², using the following steps:

- (1) Merging of directly adjacent segments of the same type.
- (2) Per-patient pairwise within-organ, between platform-intersection, with a minimum criterion of 50% reciprocal overlap.
- (3) Per-patient all-organ intersection (using bedtools multiinter) to remove variants present in all organs (i.e. germline variants)

This was followed by visual inspection of the remaining segments and the raw probe data in Nexus Copy Number 7.5.

Detection of aneuploidy. DNA-FCM analysis was performed by routine methods, as described³³, on formalin-fixed paraffin embedded tissue with >90% intact nuclei by microscopic inspection. Single and dual color interphase FISH was performed as described previously^{16,34}. In short, touch preparations were made by gently pressing a small frozen tissue sample onto a microscopic slide using a surgical forceps. This slide was subsequently washed and pepsinated and treated with 1% formamide at room temperature. After this the slides were dehydrated using a 75–80–100% ethanol series and air-dried followed by application of the probe mixture and hybridization over night. We defined true aneusomy as a non-disomic signal pattern consisting of the same number of signals for both probes located to the same chromosome. Patterns consisting of an uneven number of dual color signals were defined as noise and would have represented false positives using single color FISH.

TCR clonality analysis. This was performed as described in³⁵, using four validated multiplexed single tube PCR on DNA from the same extract as was genotyped on the array platforms. The first PCR reaction targeted the *TCRG* locus (as well as IGH, t(14;18) and t(11;14)). The second multiplex reaction targeted *TCRD*-rearrangements and the last two targeted the *TCRB* locus. For detailed information about primers used we refer to Dictor *et al.*³⁵.

References

1. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* 1–11 (2014), doi: 10.1056/NEJMoa1408617.
2. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* 141126135947008 (2014), doi: 10.1056/NEJMoa1409405.

3. Jacobs, K. B. *et al.* Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* **44**, 651–8 (2012).
4. Jamuar, S. S. *et al.* Somatic mutations in cerebral cortical malformations. *N. Engl. J. Med.* **371**, 733–43 (2014).
5. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44**, 642–50 (2012).
6. Forsberg, L. a. *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **46**, 624–8 (2014).
7. Duncan, A. W. *et al.* Frequent aneuploidy among normal human hepatocytes. *Gastroenterology* **142**, 25–8 (2012).
8. Machiela, M. J. *et al.* Characterization of Large Structural Genetic Mosaicism in Human Autosomes. *Am. J. Hum. Genet.* **96**, 487–497 (2015).
9. Huallachain, M. O., Karczewski, K. J., Weissman, S. M. & Ekehart, A. Extensive genetic variation in somatic human tissues. *Proc. Natl. Acad. Sci.* **109**, 18018–18023 (2012).
10. Piotrowski, A. *et al.* Somatic mosaicism for copy number variation in differentiated human tissues. *Hum. Mutat.* **29**, 1118–24 (2008).
11. Mills, R. E. *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).
12. Haraksingh, R. R., Abyzov, A., Gerstein, M., Urban, A. E. & Snyder, M. Genome-wide mapping of copy number variation in humans: comparative analysis of high resolution array platforms. *PLoS One* **6**, e27859 (2011).
13. van der Velden, V. H. J. *et al.* Age-related patterns of immunoglobulin and T-cell receptor gene rearrangements in precursor-B-ALL: implications for detection of minimal residual disease. *Leuk. Off. J. Leuk. Soc. Am. Leuk. Res. Fund, UK* **17**, 1834–1844 (2003).
14. Sherwood, A. M. *et al.* Deep sequencing of the human TCR γ and TCR β repertoires suggests that TCR β rearranges after $\alpha\beta$ and $\gamma\delta$ T cell commitment. *Sci. Transl. Med.* **3**, 90ra61 (2011).
15. Dik, W. a. *et al.* New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* **201**, 1715–23 (2005).
16. Valind, A., Jin, Y. & Gisselsson, D. Elevated tolerance to aneuploidy in cancer cells: estimating the fitness effects of chromosome number alterations by in silico modelling of somatic genome evolution. *PLoS One* **8**, e70445 (2013).
17. Knouse, K. a., Wu, J., Whittaker, C. a. & Amon, A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc. Natl. Acad. Sci.* **111**, 13409–13414 (2014).
18. Celton-Morizur, S. & Desdouets, C. Polyploidization of liver cells. *Adv. Exp. Med. Biol.* **676**, 123–135 (2010).
19. Hua, Y. & Crino, P. B. Single cell lineage analysis in human focal cortical dysplasia. *Cereb. Cortex* **13**, 693–699 (2003).
20. McConnell, M. J. *et al.* Mosaic Copy Number Variation in Human Neurons. *Science (80-)* **342**, 632–637 (2013).
21. Sakai, M. *et al.* Assessment of copy number variations in the brain genome of schizophrenia patients. *Mol. Cytogenet.* **8**, 46 (2015).
22. Knouse, K. A., Wu, J. & Amon, A. Assessment of megabase-scale somatic copy number variation using single cell sequencing. *Genome Res.* (2016), doi: 10.1101/gr.198937.115.
23. Pamphlett, R., Morahan, J. M., Luquin, N. & Yu, B. Looking for differences in copy number between blood and brain in sporadic amyotrophic lateral sclerosis. *Muscle Nerve* **44**, 492–8 (2011).
24. Žilina, O. *et al.* Somatic mosaicism for copy-neutral loss of heterozygosity and DNA copy number variations in the human genome. *BMC Genomics* **16**, 703 (2015).
25. Forsberg, L. a. *et al.* Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* **90**, 217–28 (2012).
26. Freed, D., Stevens, E. & Pevsner, J. Somatic Mosaicism in the Human Genome. *Genes (Basel)* **5**, 1064–1094 (2014).
27. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–42 (2012).
28. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-)* **348**, 880–887 (2015).
29. Dumanski, J. P. *et al.* Smoking is associated with mosaic loss of chromosome Y. *Science (80-)* **217**, 15–18 (2013).
30. Campbell, I. M. *et al.* Parental Somatic Mosaicism Is Underrecognized and Influences Recurrence Risk of Genomic Disorders. *Am. J. Hum. Genet.* 173–182 (2014), doi: 10.1016/j.ajhg.2014.07.003.
31. Affymetrix. *Chromosome Analysis Suite User Manual.* (2013) Available at www.affymetrix.com (Accessed: 27th August 2015).
32. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
33. Valind, A., Jin, Y., Baldetorp, B. & Gisselsson, D. Whole chromosome gain does not in itself confer cancer-like chromosomal instability. *Proc. Natl. Acad. Sci. USA* 1–5 (2013), doi: 10.1073/pnas.1311163110.
34. Gisselsson, D. Refined characterisation of chromosome aberrations in tumours by multicolour banding and electronic mapping resources. *Methods Cell Sci.* **23**, 23–8 (2001).
35. Dictor, M., Warenholt, J. & Isinger, A. Resolving T-cell receptor clonality in two and genotype in four multiplex polymerase chain reactions Michael. *Haematologica* **90**, 1524–1532 (2005).

Acknowledgements

The study was supported by grants from the Swedish Research Council, the Swedish Cancer Society, the Swedish Childhood Cancer Society, the Gunnar Nilsson Cancer Foundation, the Crafoord Foundation, and the Strategic Cancer Research Program BioCARE.

Author Contributions

A.V. and D.G. Designed and performed research, analyzed data and wrote the manuscript. M.E.K.K., C.H., M.C.J., J.G., M.S. and B.B. performed research and analyzed data. All authors approved of the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Valind, A. *et al.* The fetal thymus has a unique genomic copy number profile resulting from physiological T cell receptor gene rearrangement. *Sci. Rep.* **6**, 23500; doi: 10.1038/srep23500 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>