

Edinburgh/Amsterdam/Munich/Toulouse, 7 April 2014.

Dear Colleagues,

We wish to draw your attention to an issue of scientific concern that involves the publication of a manuscript in PNAS (Hu et al., *Proc Natl Acad Sci USA*. 2008 Dec 9;105(49):19199-19204), the validity of the data presented in that paper, and events subsequent to that publication. We believe that the events surrounding this story are damaging to the fields of nuclear organization and steroid receptor biology, detrimental to the integrity of the scientific process and scientific publishing and a disservice to scientists who have genuine concerns about published data.

We apologize to you all for the length of this communication, but by following the discreet and official procedures for the examination of scientific integrity for the last four years, and six years after the publication of the papers concerned, we have hit a wall of silence. Therefore, we have decided to share our concerns in full to a group of respected colleagues world-wide so as to render the papers concerned open for scientific reassessment.

On **3 September 2010**, we sent a formal Letter-of-Concern to UCSD, HHMI, and the NIH Office of Research Integrity (ORI), with copies to the editors of *Cell* and *Proc. Natl. Acad. Sci. USA*. In that letter we expressed our concern that a team of UCSD/HHMI investigators, led by Dr. Xiang-Dong Fu and Dr. Michael Geoff Rosenfeld, may have presented inappropriately manipulated data in two scientific publications:

(1) *Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules*. Nunez E, Kwon YS, Hutt KR, Hu Q, Cardamone MD, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG, Fu XD. *Cell*. 2008 Mar 21;132(6):996-1010.

(2) *Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules*. Hu Q, Kwon YS, Nunez E, Cardamone MD, Hutt KR, Ohgi KA, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG, Fu XD. *Proc Natl Acad Sci USA*. 2008 Dec 9;105(49):19199-19204.

The *Cell* paper was retracted in June 2008. The reason for the retraction, according to the authors, was *a series of data duplications and transversions in the original dataset*. The authors also stated in the Retraction that *they continue to believe the central conclusions of the paper*. One of us (BvS) expressed several serious concerns about massive data duplications and alterations to the authors around May 2008, which caused or contributed to the retraction. The Retraction text did not refer to the mistakes as *data fabrication*.

Surprisingly, essentially the same story was published again in PNAS in December 2008 by the same 11 authors, under almost the same title. The paper was communicated to PNAS by Dr Rosenfeld – a National Academy member - within 10 days between submission and acceptance).

In our Letter-of-Concern we provided multiple lines of evidence that suggest a substantial amount of data reported in these two publications may not be authentic. We requested that an independent committee should investigate this matter. Our extensively documented Letter-of-Concern (54 pages) can be downloaded [here](#).

Officials at UCSD responded that they would investigate this matter. However, it took ~16 months, until **26th January 2012**, before we were interviewed via telephone by an Investigation Committee convened by UCSD. On **29th November 2012** that Committee indicated to us that it had concluded its investigation. Their report was submitted to the UCSD Office of Research Affairs on **26th December 2012**. The final report from UCSD was submitted to the NIH Office of Research Integrity (ORI) on **19th June 2013**, i.e., with another ~6 months delay. Since then, UCSD and ORI have declined to comment on the case, despite requests from our end.

We therefore decided to re-engage with PNAS, the journal that had published the second paper and we sent them an expression of concern on **28th January 2014**. To our surprise, PNAS informed us that they had recently printed two ‘corrections’ to the Hu et al. paper. These corrections included relabeling of figure legends to indicate that the data presented were from a completely different cell line than that indicated in the original manuscript – note our concerns about MCF7 cells below. The other correction involved the replacement of micrographs for which we had provided evidence in 2010 that the images could not be correct.

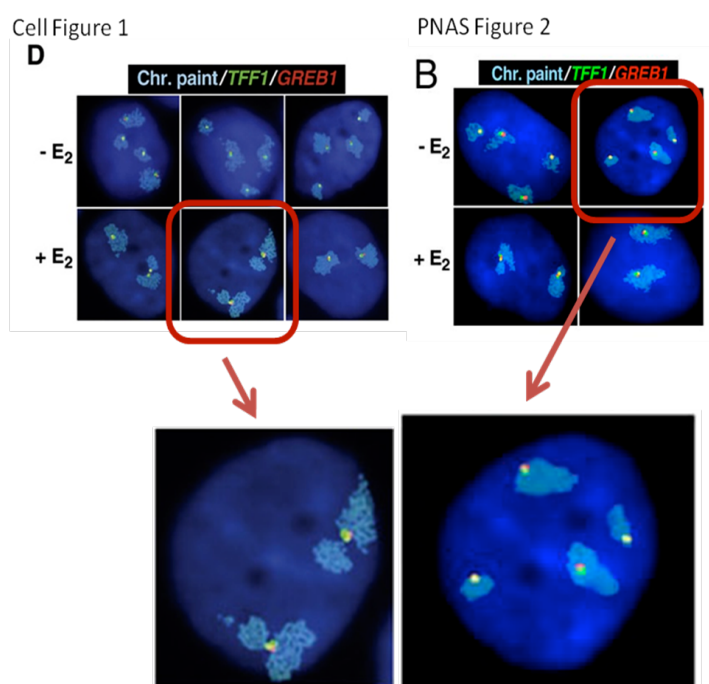
We are disturbed that PNAS should consider it acceptable to post such substantial data corrections to a manuscript that is under ORI investigation, and linked to a retracted Cell paper. Moreover, in 2010 the authors indicated to PNAS that the original microscopy images are no longer available. In our considered opinion, the numerous irregularities, duplications and mistakes associated with the publication either represent an unacceptable level of inadvertent errors, or suggest repeated data manipulation.

We are deeply concerned that – more than three years after the start of a formal investigation – the scientific community still has not been informed of the outcome of that investigation. We call upon UCSD and ORI to make the main findings of their investigation available without further delay, and we urge PNAS to publish an Expression of Concern, or retract the paper.

Specific concerns:

1. We found evidence that in at least two separate instances a microscopy image was duplicated and altered so as to represent different cells and/or different experimental conditions (Cell & PNAS). This raises concerns about the authenticity of these images.

a) The blue DAPI staining pattern of the nucleus (shape, position of heterochromatic foci and nucleoli) in the top right panel of Figure 2B of the PNAS paper appears strikingly similar to that of the nucleus shown in the middle of the bottom row in Figure 1D of the retracted Cell paper. It seemed likely to us that these two nuclei are one and the same, yet the original legend indicates that the image in the PNAS paper is from a cell cultured in the absence of estrogen (-E₂), whilst in the retracted Cell paper this is labeled as a cell in the presence (+) of E₂. Moreover, the probe hybridization signals in the two figures are completely different from each other.



b) Similarly, the DAPI (blue) images of two nuclei in the bottom row of Supplementary Figure S2A of the PNAS paper appeared to be identical, yet both the green and red FISH signals are different in each image.

An image correction has recently been posted to the PNAS paper (below). This is despite Dr Fu originally indicating to one of us (BvS) in March 2009 that raw image data were no longer available for these experiments, but had been lost due to a crashed hard-drive. From our correspondence with Dr. Randy Schekman (then Editor in Chief at PNAS) in 2010 it also became apparent that the authors could not provide the raw microscopy images; at that time the authors however stated that they had deleted the raw images from their hard drive because of disk space constraints.

The text of the correction pertaining to the PNAS Supplementary Information reads: "We regret that errors were made in two figure panels in the manuscript. The supplementary Figs. S2A and S5B each harbored an error in which the wrong DAPI-stained nucleus was incorrectly merged with the correct FISH image. The correctly merged images are now shown, with the affected panels: Fig. S2A, TFF1e, CASP7p+E2, and Fig. S5B, +E2+JP/SC35. We deeply regret carrying over these errors from a previous version of the figures and our failure to detect them at the time of publication. We apologize for any confusion that these errors may have caused."

2. We found strong statistical evidence that a large number of quantitative datasets of distance measurements between FISH probe pairs – presented as dotplots in the now retracted Cell paper, were duplicated and modified to represent measurements under different experimental conditions. BvS' two letters together directly identify more than 15 instances of indisputable duplications. The main figures of the Cell paper together displayed 62 dotplots that represent non-control samples. Statistical analysis of the corresponding datasets indicated that about 2/3 of these datasets are nearly identical (Pearson correlation coefficient >0.995) to at least one other dataset that contains supposedly independent measurements. Furthermore, a different constant number has been added to many duplicated datasets, effectively shifting the dotplots along the y-axis.

The scale of the data duplications, the fact that the data are nearly but not exactly identical, and the inexplicable additions of different constant values, together make it highly unlikely that the incorrect data were the result of simple human errors such as inadvertent copying or mis-labeling of columns in an Excel spreadsheet.

In **May and June of 2008**, one of us (BvS) extensively communicated with Dr. Fu and Dr. Rosenfeld, on this issue. In his email correspondence, Dr. Fu suggested that the y-axes differ due to the use of different microscopes, which require different pixel-to-micrometer conversion factors. This explanation however does not make sense: conversion factors must be *multiplied*, not *added*. Thus it remained unclear why various constant values were added to clearly duplicated measurements. In their response of **4th May 2008**, the authors also stated that *'all primary data used to reach the conclusions were "rock-solid" and all critical experiments were independently confirmed by two additional investigators'*. They went on to add that *'we want to unambiguously state that there is no data manipulation associated with any of the experiments reported'*.

On **10th June, 2008**, the authors wrote notifying us (BvS) of their intention to retract the Cell paper.

3. As extensively documented in *Kocanova et al, PLoS Genet 2010; 6(4):e1000922* and further substantiated by unpublished data from the Cremer lab, none of the key results can be reproduced by three independent labs. Nor are the data on TFF1-GREB interaction supported by the published ChIA-PET analysis for the estrogen-receptor in MCF7 cells (Fullwood et al., Nature, 2009).

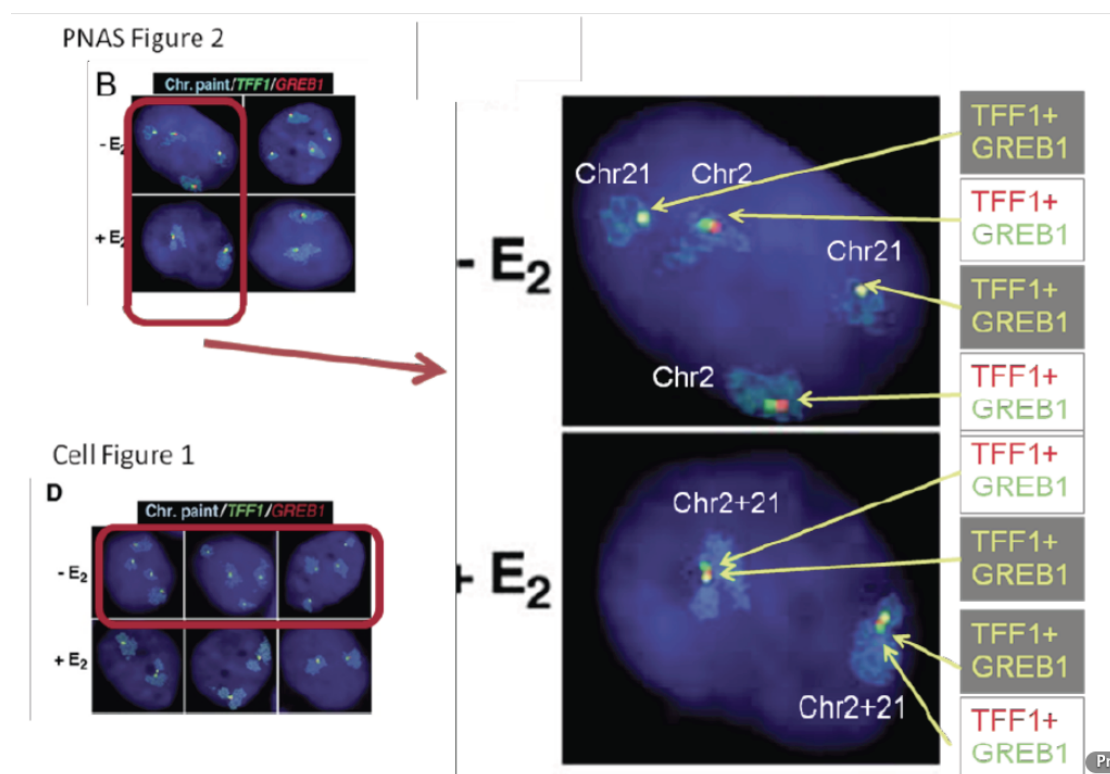
4. The expression level of the estrogen receptor is usually below detection levels in HMEC cells (Bowie et al., 2004 Oncogene; Sengupta et al 2004, Mol Cell Biochem; Koconova et al., 2010 Plos Genet), raising the question how estrogen-induced effects could have been observed. Estrogen receptor expression in the HMECs used is not presented in the PNAS paper. In response to an email

sent on **23rd October 2009**, from one of us (WB) to Dr Rosenfeld, informing him of problems reproducing data in the paper, he indicated that the HMEC primary cells ‘need to be replaced every few passages from Lonza, and thus batch-to-batch variance is a big problem, stemming in part from the ratio of stroma to epithelial cells, hence the population of ER-containing cells is variable, and therefore, their response to estrogen tends to differ dramatically’. We note that this serious reproducibility problem is not mentioned anywhere in the PNAS paper.

5. In some Fluorescence In Situ Hybridization (FISH) images purported to be from MCF7 breast cancer cells in the retracted Cell paper and the original version of the PNAS paper, the numbers of painted chromosomes and genes, as well as their relative positions, suggest a normal complement for these chromosomes in MCF7 breast cancer cells. This is not consistent with the fact that the MCF7 cells are aneuploid and rearranged for chromosomes 2 and 21, as reported in many previous publications (including Koconova et al., 2010 PLoS Genetics (Fig 2C); Jones et al., 2000 Cancer Genet. Cytogenet; Kytola et al., 2000 Genes Chromosomes Cancer; Neve et al., 2006 Cancer Cell; Osborne et al., 1987 Breast Cancer Res Treat.; Shadeo and Lam 2006, Breast Cancer Res.). Moreover, rearrangements on these chromosomes are confirmed by the sequence of the MCF7 genome (Hampton et al., Genome Res 2009).

The authors have recently (**January 2014**) posted a correction to the PNAS paper stating that images in Fig. 2 originally stated as being from MCF7 cells (aneuploid and rearranged for the chromosomes concerned), were all in fact from diploid HMECs. Most of the other micrographs presented now fail to state the cell line used. However, the legend to Figure S5 suggests that micrographs in Fig. S5B and C are from MCF7s, yet the presented images still suggest a normal diploid complement for *TFF1* and *GREB1* loci.

6. There are internal inconsistencies in the assignment of *GREB1* and *TFF1* spots to chromosome 2 and 21 territories in the control (-E₂) condition in Figure 2B of the PNAS paper – originally indicated to be from MCF7 cells, but now stated as from HMECs in the corrected version. On the larger pair of chromosome territories (therefore assumed to be Chromosome 2) there are both green and red superimposed FISH signals of the two studied genes – *TFF1* and *GREB1*, even though these two genes are supposed to be on different chromosomes (21 and 2 respectively).



7. Close inspection of (FISH) experiments suggest that several of the published images may not be genuine. In some FISH images in the Cell paper we noted nearly identical spatial and intensity patterns for the grey values of pixels allegedly recorded in separate color channels and representing co-localizing GREB1 and TFF1 genes. One of us (TC) reported this inconsistency to the Cell editor Marcus Emilie in an email on **August 2, 2008**.

8. The interphase chromosome paints in several images have a highly unusual rope-like appearance suggestive of inappropriate image processing or image manipulation (PNAS & Cell). Two of us (TC and WB) have decades of experience of chromosome painting using FISH, and have never seen chromosome territory images that look like this.

For detailed illustrations of these concerns, see our original [Letter-of-Concern](#).

Sincerely,

Bas van Steensel

Division of Gene Regulation
Netherlands Cancer Institute
Amsterdam, the Netherlands
b.v.steensel@nki.nl

Thomas Cremer

Department of Biology II
LMU Biocenter
Grosshadernerstr. 2
82145 Planegg-Martinsried, Germany
Thomas.Cremer@lrz.uni-muenchen.de

Wendy Bickmore

MRC Human Genetics Unit
Institute of Genetics and Molecular Medicine
University of Edinburgh
Edinburgh EH42XU, UK
Wendy.Bickmore@igmm.ed.ac.uk

Kerstin Bystricky

LBME – CNRS UMR5099 University of Toulouse
118 route de Narbonne
31062 Toulouse cedex, France
kerstin@biotoul.fr

