



OPEN

## Induced pluripotent stem cells from subjects with Lesch-Nyhan disease

Diane J. Sutcliffe<sup>1,12</sup>, Ashok R. Dinasarapu<sup>2,12</sup>, Jasper E. Visser<sup>3,4</sup>, Joery den Hoed<sup>1</sup>, Fatemeh Seifar<sup>1,5</sup>, Piyush Joshi<sup>1</sup>, Irene Ceballos-Picot<sup>6</sup>, Tejas Sardar<sup>1</sup>, Ellen J. Hess<sup>1,5,7</sup>, Yan V. Sun<sup>8</sup>, Zhexiong Wen<sup>1,9,10</sup>, Michael E. Zwick<sup>2,11</sup> & H. A. Jinnah<sup>1,2,5,11</sup>✉

Lesch-Nyhan disease (LND) is an inherited disorder caused by pathogenic variants in the *HPRT1* gene, which encodes the purine recycling enzyme hypoxanthine–guanine phosphoribosyltransferase (HGprt). We generated 6 induced pluripotent stem cell (iPSC) lines from 3 individuals with LND, along with 6 control lines from 3 normal individuals. All 12 lines had the characteristics of pluripotent stem cells, as assessed by immunostaining for pluripotency markers, expression of pluripotency genes, and differentiation into the 3 primary germ cell layers. Gene expression profiling with RNAseq demonstrated significant heterogeneity among the lines. Despite this heterogeneity, several anticipated abnormalities were readily detectable across all LND lines, including reduced *HPRT1* mRNA. Several unexpected abnormalities were also consistently detectable across the LND lines, including decreases in *FAR2P1* and increases in *RNF39*. Shotgun proteomics also demonstrated several expected abnormalities in the LND lines, such as absence of HGprt protein. The proteomics study also revealed several unexpected abnormalities across the LND lines, including increases in *GNAO1* and decreases in *NSE4A*. There was a good but partial correlation between abnormalities revealed by the RNAseq and proteomics methods. Finally, functional studies demonstrated LND lines had no HGprt enzyme activity and resistance to the toxic pro-drug 6-thioguanine. Intracellular purines in the LND lines were normal, but they did not recycle hypoxanthine. These cells provide a novel resource to reveal insights into the relevance of heterogeneity among iPSC lines and applications for modeling LND.

Lesch-Nyhan disease (LND) is an inherited disorder caused by pathogenic variants in the *HPRT1* gene<sup>1</sup>. The variants result in deficiency of a purine recycling enzyme, hypoxanthine–guanine phosphoribosyltransferase (HGprt)<sup>2,3</sup>. This enzyme is ubiquitously expressed in all tissues of the body at all stages of development<sup>4</sup>. HGprt deficiency produces pleiotropic clinical effects including overproduction of uric acid leading to kidney stones and gout<sup>5</sup>, brain abnormalities<sup>6–8</sup>, and macrocytic anemia<sup>9</sup>.

Many experimental models have been developed for LND. Animal models have been produced by knocking out the *Hprt1* gene in mice<sup>10,11</sup>, rats<sup>12</sup> and rabbits<sup>13</sup>. Several studies have described spontaneous or induced mutations in immortalized rodent cell lines<sup>14–20</sup>. However, rodents do not exhibit the clinical features seen in affected humans<sup>21,22</sup>. One of the reasons for the lack of overt abnormalities in rodents may be that purine metabolism differs across species. For example, purine turnover is at least tenfold higher in rodents compared to primates.

<sup>1</sup>Department of Neurology, Emory University School of Medicine, 101 Woodruff Circle, 6305 Woodruff Memorial Building, Atlanta, GA 30322, USA. <sup>2</sup>Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA. <sup>3</sup>Department of Neurology, Cognition and Behavior, Donders Institute for Brain, Radboud University Medical Center, Nijmegen, The Netherlands. <sup>4</sup>Department of Neurology, Amphia Hospital, Breda, The Netherlands. <sup>5</sup>Neurosciences Graduate Program, Graduate Division of Biological and Biomedical Sciences, Laney Graduate School, Emory University, Atlanta 30322, USA. <sup>6</sup>Laboratoire de Biochimie Métabolomique Et Protéomique, Hôpital Universitaire Necker, Paris, France. <sup>7</sup>Department of Pharmacology and Chemical Biology, Emory University School of Medicine, Atlanta, GA 30322, USA. <sup>8</sup>Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA. 30322, USA. <sup>9</sup>Department of Psychiatry & Behavioral Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA. <sup>10</sup>Department of Cell Biology, Emory University School of Medicine, Atlanta, GA 30322, USA. <sup>11</sup>Department of Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA. <sup>12</sup>These authors contributed equally: Diane J. Sutcliffe and Ashok R. Dinasarapu. ✉email: hjinnah@emory.edu

	LND Cases			Healthy controls		
	A	D	T	A	B	C
Sex	Male	Male	Male	Male	Male	Male
Age <sup>a</sup>	15	10	46	21	43	22
Race	Black	Caucasian	Caucasian	Caucasian	Black	Caucasian
<i>HPRT1</i> gene variant	c.151C>T	c.508C>T	c.371insTT	none	none	none
HGprt consequence	Premature stop	Premature stop	Frame shift, premature stop	Normal	Normal	Normal
Clinical features <sup>b</sup>	Classic	Classic	Classic	Normal	Normal	Normal
iPSC lines	L3.29 L3.30	L2.09 L2.12	L1.02 L1.04	C1.02 C1.03	C2.04 C2.06	C3.05 C3.08

**Table 1.** Subject characteristics. <sup>a</sup>Age relates to subject age at skin biopsy. <sup>b</sup>Clinical features of classic LND include motor dysfunction resembling cerebral palsy, cognitive impairment, self-injurious behavior, and overproduction of uric acid.

Rodents also express certain enzymes that are lacking in humans, such as uricase, which protects rodents from excess uric acid<sup>23–25</sup>. There are also substantial differences in organ development, particularly for the brain. Thus, species differences in purine metabolism or neurodevelopmental mechanisms have led to uncertainty regarding the translation of findings from rodent models to humans.

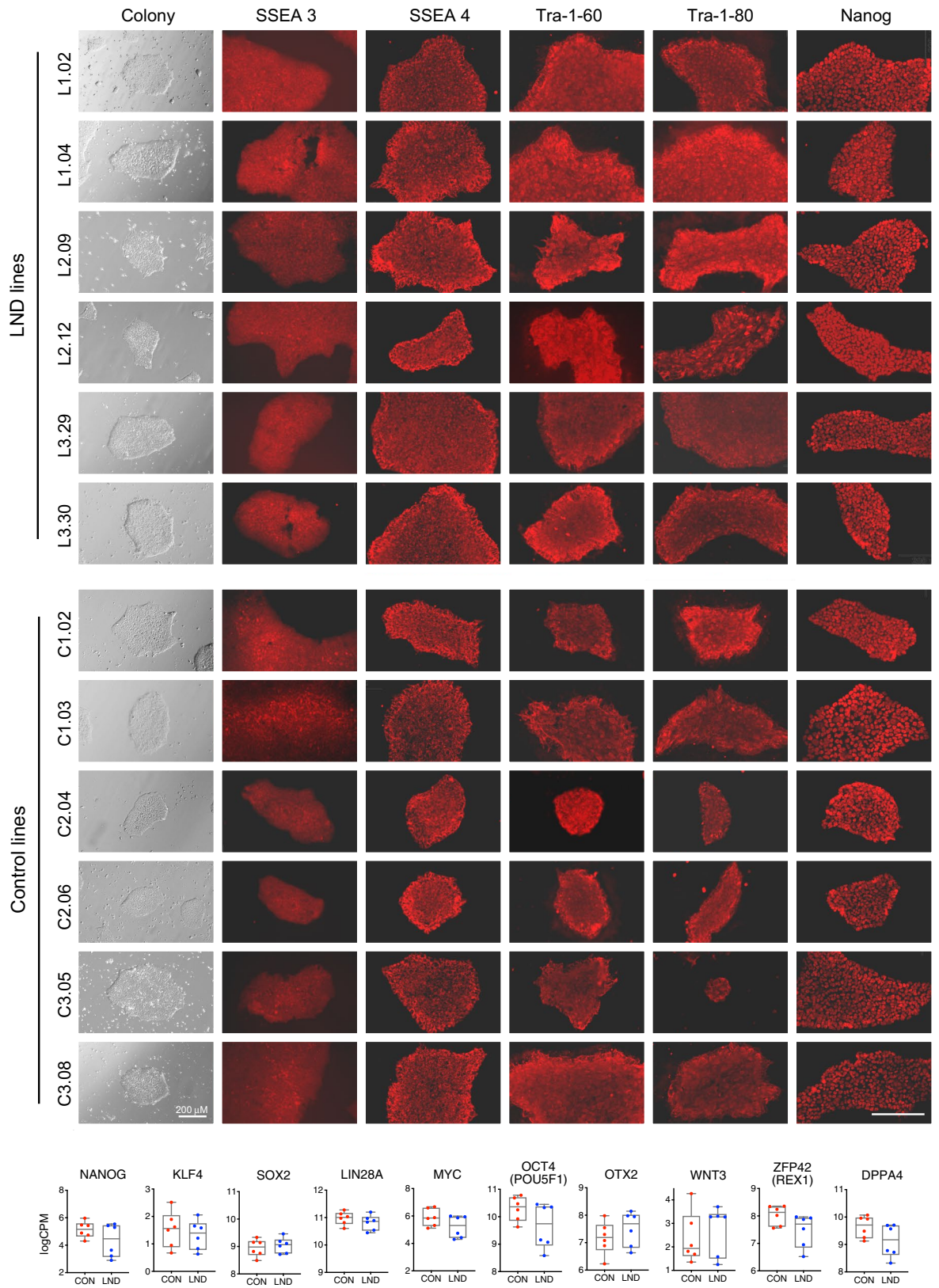
Other experimental models have been derived from human materials, such as blood cells<sup>26–32</sup> or fibroblasts<sup>3,33–38</sup> collected from individuals with LND. Some cell models have relied on spontaneous or induced mutations in immortalized human cell lines<sup>39–42</sup>. One study described human neural stem cells from an aborted LND fetus<sup>43</sup>, while another focused on partial knockdown of *HPRT1* expression in a human induced pluripotent stem cell (iPSC) line and a human embryonic stem cell line<sup>44</sup>. These studies of human cells have revealed that many of the consequences of HGprt deficiency are cell-specific<sup>42,45</sup>. Three reasons account for these cell type-specific differences. First, HGprt is expressed ubiquitously; but its level varies across tissues, suggesting that some cells are more dependent on HGprt than others<sup>46</sup>. Other enzymes of purine metabolism also are expressed at different levels in different tissues<sup>47</sup>, so downstream and compensatory consequences of HGprt deficiency may vary as well. Finally, purines are utilized in different ways by different tissues. For example, rapidly dividing cells use purine nucleotides predominantly for RNA synthesis and DNA replication<sup>48</sup>. In contrast, post-mitotic cells such as neurons do not need large amounts of purine nucleotides for DNA replication, but use nucleotides and nucleosides for signaling both within and between cells<sup>49</sup>.

Because the consequences of HGprt deficiency vary among different types of cells, the pathogenesis of the different clinical aspects of HGprt deficiency may need to be studied in different cell types relevant to the different clinical features of LND. Induced pluripotent stem cells (iPSCs) derived from patients carrying pathogenic variants in the *HPRT1* gene provide a novel strategy to study the consequences of HGprt deficiency in different lineages. The main goal of the current study was to describe the development and characterization of 6 iPSC lines derived from 3 unrelated LND cases as a resource for future studies. The first part of the manuscript describes the generation and characterization of the novel iPSC lines. The second part describes gene and protein expression abnormalities consistently found among the LND lines. The final part presents a functional characterization of the LND lines in relation to purine metabolism.

## Results

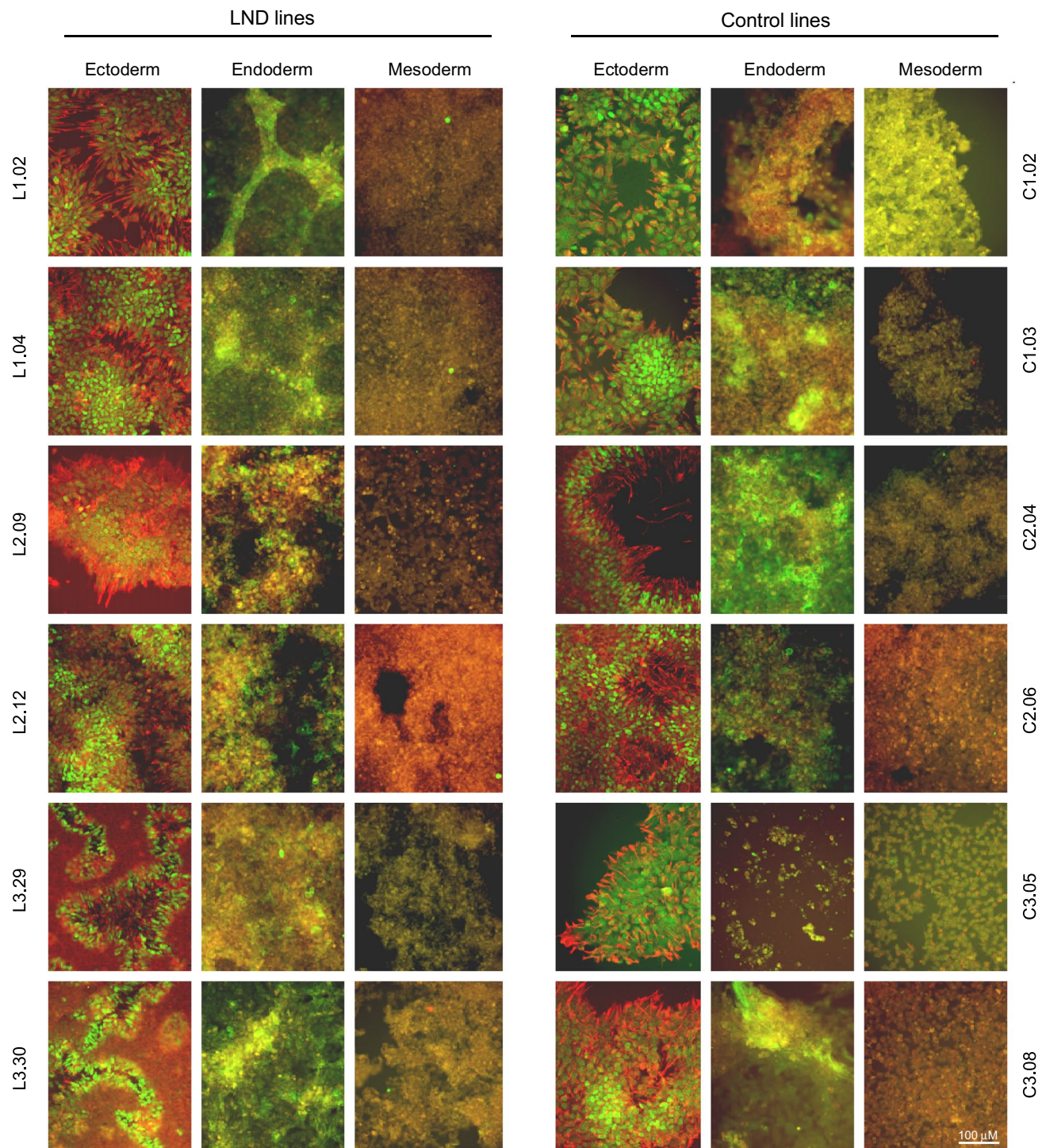
**Selection of fibroblasts for reprogramming.** From a collection of fibroblast cultures of 47 cases of HGprt deficiency, several criteria were used to select candidates for reprogramming. First, the mutation should be stable with a known consequence for HGprt enzyme function. Second, the mutation should be associated with multiple unrelated LND patients to ensure a consistent relationship with the clinical phenotype. Third, cultures from relatively young patients should be available, because age affects reprogramming efficiency and frequency of acquired genetic variants<sup>50</sup>. With these criteria in mind, three unrelated cases were selected, all with the classic LND phenotype that included overproduction of uric acid, motor and cognitive disability, and self-injurious behavior. Two involved hot-spot variants leading to different stop codons (c.151C>T and c.508C>T) and one was frame-shifting mutation (c.371insTT). All of these variants lead to low mRNA levels due to nonsense-mediated decay, as well as null enzyme function. Control fibroblast cultures were selected from a bank of 20 normal individuals free of any known medical or neurological problems. Because LND is inherited in an X-linked recessive manner, all cases and controls were males (Table 1).

**Characterization of iPSC lines.** At the morphological level, all 6 LND and all 6 control lines grew in rounded colonies with defined borders typical of other iPSC lines (Fig. 1). All showed immunostaining characteristics typical of pluripotency including expression of SSEA3, SSEA4, TRA-1-60, TRA-1-80, and Nanog (Fig. 1). Complete gene expression profiles were determined by RNAseq, with the percentage of mapped reads ranging between 88.8 and 91.7%. A total of 23,748 genes were detected. After filtering genes expressed at very low levels, 17,113 genes remained for quantification. All of the lines expressed genes typically associated with pluripotency, with no apparent differences between LND and control lines (Fig. 1). All lines were capable of differentiating into each of the three main germ cell layers as shown by immunostains (Fig. 2) for ectoderm (PAX6,



**Figure 1.** Pluripotency characteristics. This figure shows a phase-contrast photomicrograph of typical colonies for all 6 LND and all 6 control lines, along with immunostains for pluripotency markers (SSEA3, SSEA4, TRA-1-60, TRA-1-80, Nanog). All cultures were stained simultaneously, and all photomicrographs were taken using the same microscope settings. Scale bar is shown at bottom right and depicts 200  $\mu$ m. Gene expression profiling shown as box plots at the bottom of the figure reveal that the control and LND lines expressed similar levels of genes associated with pluripotency.





**Figure 2.** Trilineage differentiation. This figure shows the differentiation of all 6 LND and all 6 control lines into the 3 primary germ cell layers via merged immunostains for ectoderm (PAX6 = green, nestin = red), mesoderm (Brachyury = green, NCAM = red), and endoderm (SOX17 = green, FOXA2 = red). All cultures were stained simultaneously, and all photomicrographs were taken using the same microscope settings. Distinct red or green colors are evident for ectoderm, where the markers were often expressed in different cells. Shades of orange and yellow result from varied proportions of overlapping red and green markers. Scale bar is shown at bottom right and depicts 100  $\mu$ M.

NESTIN), mesoderm (Brachyury, NCAM), or endoderm (SOX17, FOXA2). Over several different staining runs, with or without a nuclear counterstain, there were no obvious differences between the LND and control lines.

RNAseq data for all iPSC lines were queried to assess for genetic variants in *HPRT1*. Each of the LND lines harbored genetic variants expected from their starting fibroblast lines (Table 1, Supplementary Fig. S1). The nonsense variant c.151C>T was confirmed for lines LND29 and LND30. The nonsense variant c.508C>T was confirmed for lines LND9 and LND12. The frame-shifting variant c.371insTT was confirmed for LND2 and LND4. All of these variants were independently confirmed by sequencing RT-PCR products. No other *HPRT1* sequence variants were detected in the LND iPSC lines and no variants were detected in any of the 6 control iPSC lines. Karyotypes were normal for all LND and control lines, except for line LND30 where 11/20 metaphases demonstrated an unbalanced translocation affecting the long arm of chromosome 1, with a karyotypic profile of 46,XY<sup>51</sup>/46,XY,t(1:10)(p36.3;q11.1)<sup>51</sup> (Supplementary Fig. S2). Although karyotypically abnormal lines may yield idiosyncratic results, this line was not discarded, but results were censored so that it could be compared against the other karyotypically normal lines.

**Differential gene expression in LND and control iPSCs.** The RNAseq data were next evaluated to identify differences in gene expression between the LND (n=6) and control (n=6) lines. With a minimum threshold of 1.5× change and correcting for multiple comparisons with FDR method, differential gene expression using DESeq2 method revealed significant differences for 3555 genes at FDR<0.10, 1797 transcripts at FDR<0.05, and 413 transcripts at FDR<0.01 (Fig. 3A). Differential gene expression using the edgeR method revealed significant differences for 1108 transcripts at FDR<0.10, 407 transcripts at FDR<0.05, and 69 transcripts at FDR<0.01 (Fig. 3A). These results demonstrate that these commonly used methods for differential gene expression yield different outcomes. However, at FDR<0.05, almost all transcripts (403 transcripts, 99%) were common to both the DESeq2 and edgeR methods (Fig. 3B). As shown in volcano plots (Fig. 3C) and heat map (Fig. 3D), the most prominently increased transcript was *RNF39* (ring finger protein 39). The most prominently reduced transcripts were *HPRT1* and *FAR2P1* (fatty acyl-coA reductase pseudogene 1). All three of these prominently altered transcripts in RNAseq were confirmed via qPCR using independently collected samples from each line (Supplementary Fig. S3). The significant decrement in *HPRT1* transcripts provides a useful positive control, because such a reduction is consistent with the fact that the genetic variants included here are all subject to nonsense-mediated mRNA decay.

Genes that were differentially expressed via either the edgeR and DESeq2 methods were next subject to Ingenuity Pathway Analysis. Although these methods revealed somewhat different lists of differentially expressed transcripts, they produced similar results for affected biological pathways (Fig. 3E). Of the top 10 pathways selected based on p-value and z-scores, two were significantly reduced (tRNA charging and NRF2-mediated oxidative stress) and 8 were significantly increased (Role of NFAT in cardiac hypertrophy, GPCR-mediated nutrient sensing in enteroendocrine cells, androgen signaling, synaptic long-term depression, endocannabinoid neuronal synapse pathway, G beta gamma signaling, calcium signaling, and relaxin signaling).

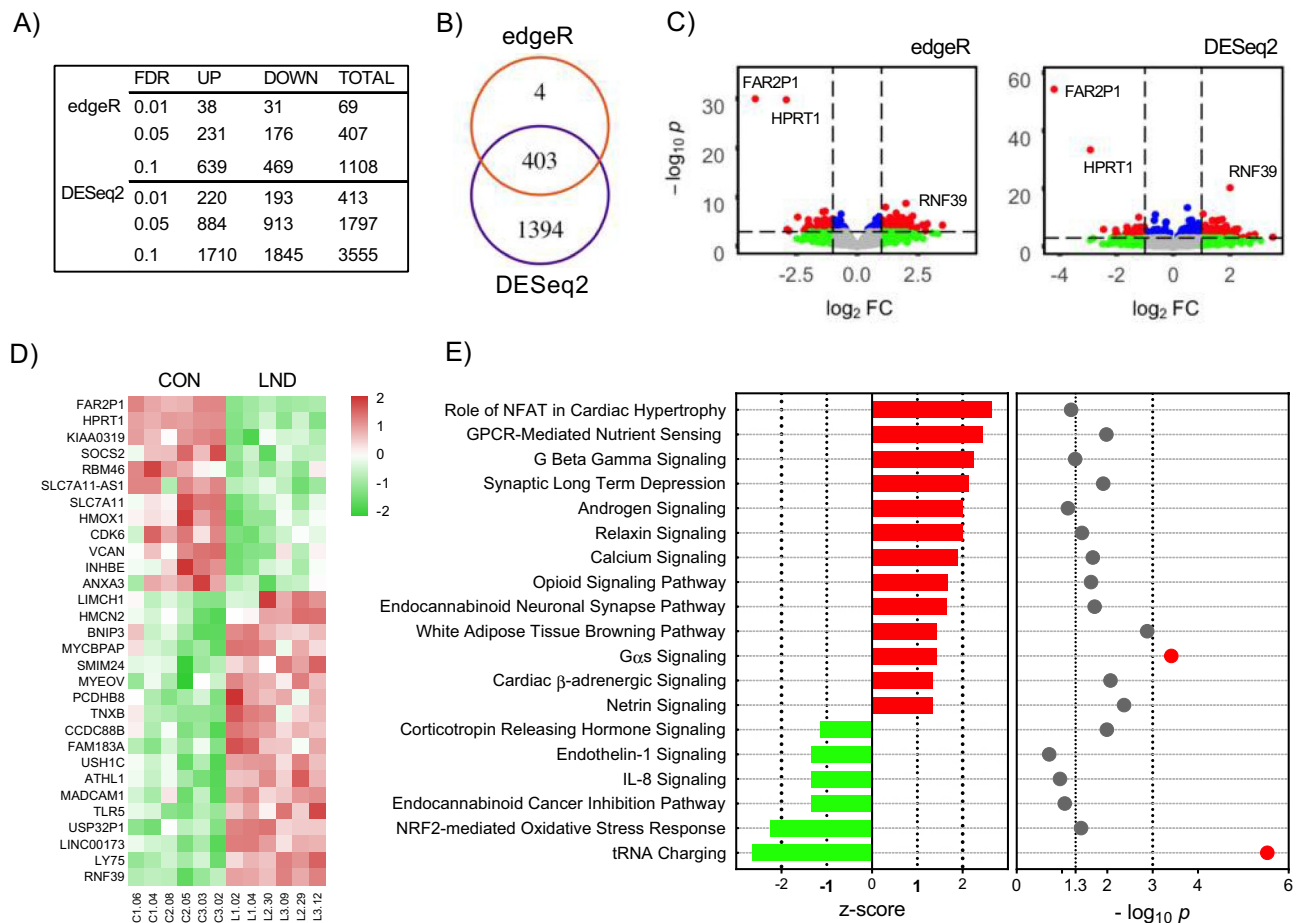
**Shotgun proteomics in LND and control iPSCs.** An unbiased shotgun approach was used to compare the proteomes of the LND (n=6) and control (n=6) iPSC lines. After filtering peptides with low expression, a total of 33,648 peptides mapping to 2,493 proteins were quantified. The proteomics dataset provided quantitative data for 12 proteins associated with pluripotency (ACACA, ACLY, CDH1, DNMT1, DNMT3A, FASN, RBM3, SOX2, SLC25A1, VIM, YAP1, ZIC2). None of these were significantly different between LND and control iPSCs (not shown).

The proteomics data were next evaluated to identify differences in protein expression between the LND and control lines. There were 7 significant protein changes at FDR<0.10 (Fig. 4B). Five proteins were decreased (HGprt, NSE4A/EID3, MPU1, TOIP1, PDIA2, GNAO, PDIA3), and 2 proteins were increased (GNAO, ATPK) (Fig. 4C). Two of these protein changes remained significant at FDR<0.01 (Fig. 4D). One was HGprt, as expected for the types of mutations involved. In fact, comparison of the raw vs imputed proteomics data implied a complete absence of HGprt protein in all of the LND samples (Fig. 4E). The other peptide significantly reduced at FDR<0.01 mapped to both NSE4A (non-structural maintenance of chromosomes element 4, homolog A) and EID3 (EP300 interacting inhibitor of differentiation 3, a paralog of NSE4A). Both the raw and imputed data for NSE4A/EID3 implied a complete absence in all LND iPSCs. Reductions in HGprt and NSE4A were confirmed using Western blotting in independently collected samples from all the lines (Supplementary Fig. 4).

**Comparison of RNAseq and proteomics results.** Among 2493 proteins identified, 2444 mapped to genes simultaneously quantified by RNAseq. Overall, there was a good quantitative correlation between the results from proteomics and RNAseq (Fig. 4A) for both the control (R=0.58) and LND lines (R=0.55). These correlations fall in the range of other studies that compared RNAseq and proteomic data generated from the same samples<sup>51</sup>. The lack of strict correlation between these methods is known to result from post-transcriptional, post-translational, and other mechanisms not detected by either method.

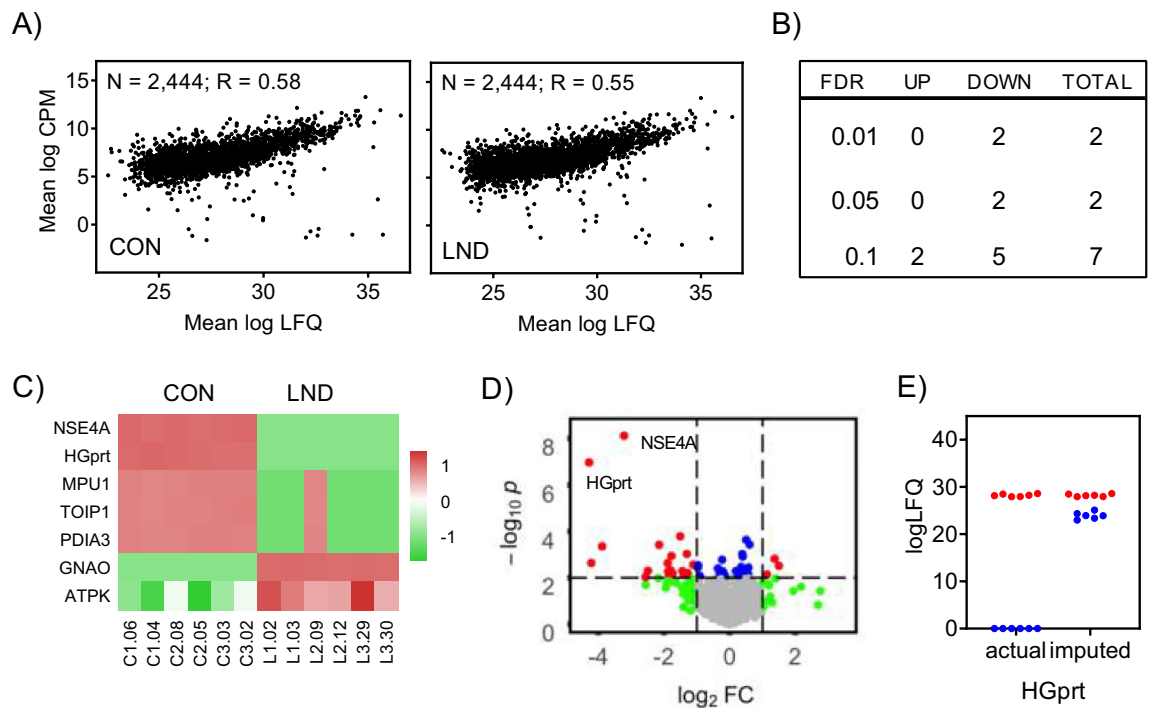
Among the analytes that were most significantly different between control and LND lines, RNAseq and proteomics revealed significantly reduced *HPRT1* transcripts and HGprt protein. However some of the transcripts significantly different between LND and controls in the RNAseq dataset were not significantly different in the proteomics study. For example *FAR2P1* is a pseudogene, so no protein was detected by proteomics. *RNF39* was significantly increased in the RNAseq dataset, but was not detectable in the proteomics dataset. Thus some of the apparent discrepancies in the RNAseq and proteomics datasets reflect differences in the biology of mRNA and protein expression.





**Figure 3.** Transcriptome analysis. Gene expression profiles were evaluated by RNAseq for all LND ( $n = 6$ ) and control ( $n = 6$ ) lines. Differential gene expression differed according to both analytical method (edgeR, DESeq2) and statistical threshold (A). However, 403 genes were differentially expressed using both methods at FDR < 0.05 (B). For both edgeR and DESeq2 methods, volcano plots (C) consistently demonstrate 2 genes to be expressed at markedly lower levels in the LND lines (*HPRT1* and *FAR2P1*), and 1 gene to be expressed at higher levels in the LND lines (*RNF39*). The red color shows genes with FDR < 0.05 and fold change > 2. The blue color shows genes with FDR < 0.05 and fold change < 2. Green and gray show gene changes with FDR > 0.05. Individual variation among the lines is shown as a heatmap of gene expression (D). Red depicts higher gene expression and green depicts lower expression. Considering only the 403 genes differentially expressed using both edgeR and DESeq2 at FDR < 0.05, 19 pathways appeared to be significantly different when comparing the LND and control lines (E). Red bars depict positive z-scores while green depict negative z-scores. Red dots represent  $p$ -values < 0.001 (i.e.  $-\log_{10}(0.001) > 3$ ).

Conversely, some statistically significant differences in the proteomics dataset between control and LND iPSCs were not significant in the RNAseq dataset. For example, a peptide mapping to both EID3 and NSMCE4 was detected in control iPSCs, but absent in LND iPSCs. However, the RNAseq revealed no significant change in either transcript (*NSMCE4*,  $p = 0.45$ ; *EID3*,  $p = 0.23$ ). Two other proteins significantly altered in the proteomics study also were not significantly altered by RNAseq. These proteins included ATPK (gene symbol *ATP5JK* or *ATP5MF*,  $p = 0.11$ ) and GNAO (gene symbol, *GNAO1*,  $p = 0.24$ ). ATPK is involved in synthesis of the purine ATP and GNAO is regulated by the purine GTP, so changes at the protein level may reflect post-transcriptional alterations caused by changes in intracellular purine metabolism in HGprt-deficient cells. Three other proteins significantly altered in the proteomics study were not significantly altered in the RNAseq study when statistically correcting for multiple comparisons, but they showed clearly significant trends in the same direction as proteomics when uncorrected. These proteins included PDIA3 (gene symbol *PDIA3*,  $p = 0.010$ ), MPU1 (gene symbol *MPU1*,  $p = 0.011$ ), TOIP1 (gene symbol, *TOR1AIP1*,  $p = 0.025$ ). Thus some of the apparent discrepancies between the RNAseq and proteomics datasets reflect the limits of statistical methods in RNAseq and proteomics, and some have argued that corrections for multiple comparisons using the FDR method are inappropriately stringent for proteomics<sup>52</sup>. These comparisons highlight the many different reasons that RNAseq and proteomics datasets from the same samples typically show imperfect correlations ranging from  $0.40 < R < 0.60$ <sup>51</sup>.



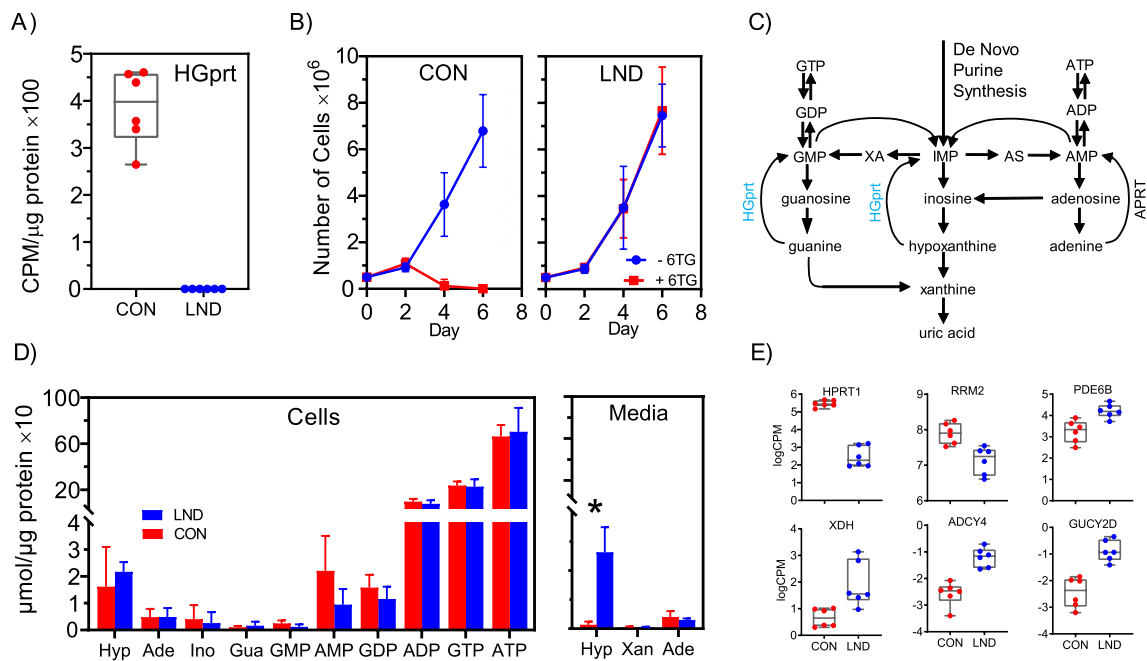
**Figure 4.** Proteomics analysis. Protein expression profiles were evaluated by shotgun proteomics for all LND (n=6) and control (n=6) lines. There was a good quantitative correlation between the gene and protein expression results for both the LND and control lines (A). Only 7 proteins reached statistical thresholds for differential expression at FDR<0.10 (B). Individual variation for among the lines for the 7 proteins meeting the FDR<0.10 criterion is shown as a heatmap (C). Red depicts a higher gene expression and green depicts lower gene expression. Only 2 proteins remained significantly different between LND and control at FDR<0.01, and a volcano plot revealed marked reductions in the levels of both HGprt and NSE4A (D). Red depicts genes with  $p < 0.001$  and fold change  $> 2$ . Green and gray depict FDR>0.05. The marked reductions in all LND lines (blue) compared to all control lines (red) were confirmed by evaluation of the raw and imputed values for HGprt (E).

**Functional assessment of LND iPSCs.** To assess the functional effect of the *HPRT1* genetic variants, all 6 LND and all 6 control lines were evaluated using an assay for HGprt enzyme activity. HGprt enzyme activity was readily detectable in all control iPSCs (Fig. 5A). Functional HGprt enzyme activity was absent from all LND lines, consistent with the transcriptome and proteome results described above. To further assess the impact of the *HPRT1* genetic variants on cell growth, cells were grown with or without 6TG, an inert synthetic purine base that is phosphorylated into a toxic purine nucleotide, only in cells which have functional HGprt. Overall, the average growth rates for the LND and control iPSCs were comparable in the standard culture medium (Fig. 5B). All of the control lines showed a clear toxic effect when 6TG was added, while all LND were resistant to it (Fig. 5B). These results demonstrate absence of functional HGprt enzyme in all of the LND iPSC lines.

Because HGprt plays a known role in purine salvage (Fig. 5C), purines were measured in all 6 LND and all 6 control iPSCs by HPLC with photodiode array UV detection. In keeping with other types of cells, the majority of intracellular purines were adenine nucleotides (ATP, ADP, AMP), with smaller amounts of guanine nucleotides (GTP, GDP, GMP), and trace amounts of several other purines. Multivariate ANOVA revealed no significant differences for any of the intracellular purines when comparing the LND vs control lines (Fig. 5D).

Purines accumulated in the tissue culture medium over a 24-h period were also quantified because some purines are released from cells to the extracellular environment. In keeping with other types of cells, the tissue culture medium contained purine bases and nucleosides, but no detectable nucleotides. Multivariate ANOVA revealed significant differences between the LND and control lines for hypoxanthine (Fig. 5D). The LND iPSCs released an average of 20 fold more hypoxanthine into the medium than control iPSCs. This result is consistent with the known role of HGprt in recycling hypoxanthine into purine nucleotides. Uric acid levels were undetectable in both LND and control iPSCs, implying that the oxidase form of xanthine oxidoreductase (XOR) is not active in these cells (Fig. 5D).

Although HGprt deficiency is the primary biochemical defect in LND, it is known to produce numerous downstream secondary changes in purine metabolism, such as increases in purine synthesis. We therefore interrogated the RNAseq data for changes the expression of other genes encoding proteins known to be involved in purine metabolism. Among the 109 purine-related genes in the KEGG pathway for purine metabolism, only 6 were significantly different in the LND vs control iPSC lines at FDR<0.05 (Fig. 5E). This result implies that any secondary changes in purine metabolism that occur in LND, such as increases in purine synthesis, may result from post-transcriptional mechanisms. This suggestion is consistent with prior studies have shown that many enzymes of purine metabolism not regulated by their levels of expression, but by other factors such as substrate



**Figure 5.** Functional assessment of HGprt deficiency. All control iPSC lines (red,  $n=6$ ) demonstrated readily detectable activity in the HGprt enzyme assay, while all LND iPSC lines (blue,  $n=6$ ) had no detectable activity (A). Results are shown as box plots with results for each line shown separately. There was no obvious difference in the growth of the control and LND lines in the absence of 6TG (blue lines, B). However, the control lines showed a clear toxic effect of 6TG, while the LND lines appeared to be resistant to it (red lines, B). In these growth curves, the error bars show the standard deviation. HGprt plays an important role in the metabolism of hypoxanthine (C), leading to marked hypoxanthine accumulation that is released into the tissue culture medium; with no apparent impact on intracellular purines (D). In this panel, controls are shown in red and LND in blue, with error bars reflecting standard deviations. Among 109 genes in the KEGG pathway for purine metabolism, only 6 were significantly different between control (red) and LND (blue) iPSC lines at  $FDR < 0.05$  (E).

availability, release from endogenous feedback inhibitors, post-translational mechanisms, or compartmentalization into purinosomes<sup>48,53</sup>.

## Discussion

The current studies describe the development and characterization of 6 iPSC lines from subjects with classic LND, along with 6 iPSC lines from normal healthy individuals for comparisons. All 12 lines have the typical characteristics of pluripotent stem cells, as assessed using multiple methods including immunostaining for pluripotency markers, expression of mRNA transcripts for pluripotency genes, and directed differentiation into the 3 primary germ cell layers (endoderm, ectoderm, and mesoderm). These cells provide novel tools for studying the pathogenesis of LND, as well as novel insights into the biology of iPSCs and their use for disease modeling.

**Relevance for iPSC biology.** Many prior studies have shown an important interplay between pluripotency and metabolism, and especially purine nucleotides such as ATP<sup>54</sup>. In addition to driving many energy-dependent reactions, ATP is required as a building block for DNA and RNA in dividing cells. Mammalian cells derive their purine nucleotides from two major sources, de novo purine synthesis and purine salvage<sup>48</sup>. However, all types of cells do not rely equally on de novo and salvage pathways to maintain purine levels. For example, actively dividing stem or neoplastic cells typically rely heavily on de novo synthesis, while terminally differentiated cells tend to depend more on salvage<sup>48,55–57</sup>.

Although *HPRT1* is listed among the many genes involved in cell cycling<sup>58</sup>, our studies imply that iPSCs do not require HGprt, because iPSCs with null variants had pluripotency characteristics that were similar to those of iPSCs derived from normal individuals (Figs. 1, 2), they grew at rates similar to iPSCs from normal individuals (Fig. 5B), and they demonstrated no significant changes in intracellular purines (Fig. 5D). However, all 6 LND iPSCs consistently wasted large amounts of purines, as judged by the 20-fold increases in hypoxanthine released from cells into the tissue culture medium (Fig. 5D). Because intracellular purine nucleotides were normal in the LND iPSCs, this result implies that the de novo pathway for synthesis of new purines is sufficient to compensate for any purines potentially lost because of the absence of HGprt-mediated purine salvage in iPSC cells.

**Relevance for iPSC disease modeling.** Many prior studies of iPSCs derived from subjects with various diseases have relied on relatively small numbers of iPSC lines, especially for rare diseases<sup>59</sup>. For LND, prior



studies have described only 1–2 iPSC lines, often from a single individual<sup>60,61</sup>. Others have described *HPRT1* gene-edited iPSCs<sup>62</sup>. However, few characteristics regarding these iPSCs were provided. Many other studies have emphasized phenotypic variation among iPSC lines, even those derived from the same subject<sup>63–70</sup>. There are many causes for this variability including sample source (disease characteristics, type of mutations, background genome, sex, age, race, etc....), cell handling methods (reprogramming method, growth conditions, passage level, etc....), and experimental measures examined (gene expression, protein expression, enzyme activity, purines measures, etc....). Our studies with 6 LND iPSC lines (2 independent lines from each of 3 unrelated subjects) provide some insights into the biological relevance of this variability.

On the one hand, our studies confirm significant variation between iPSC lines. For example, the L3.30 line acquired a genomic defect involving an unbalanced translocation of chromosome 1. Such chromosomal defects are commonly acquired during iPSC production and growth<sup>71,72</sup>. Significant variability was also evident across iPSC lines in the gene expression studies, with no obvious influence of within-subject or between-subject relationships (Fig. 3D). In the proteomic studies, line L2.09 demonstrated idiosyncratic results compared to other LND lines (Fig. 4C). This variability evident across the 6 LND iPSC lines confirms prior calls for caution when relying on a small number of iPSC lines for disease modeling, and calls into question results from prior smaller studies of LND<sup>60,61</sup>.

On the other hand, despite this variation, some very consistent abnormalities could be detected across the LND iPSC lines. For example, it is often claimed that karyotypical defects such as those found in LND30 should not be used, because they provide idiosyncratic results<sup>71,72</sup>. However, our results show that this line behaved similarly to the other 5 LND iPSC lines in all respects studied. Thus this karyotypical defect did not seem relevant to the molecular and biochemical consequences of *HGprt* deficiency. Despite the variation in gene expression, consistent abnormalities were clearly detectable for many transcripts. Most notably, consistent abnormalities could be detected in the LND iPSC lines for expression of *HPRT1* mRNA and protein, enzyme activity, growth characteristics with and without 6TG, and certain unexpected findings such as reductions in *FAR2P1* mRNA (Fig. 3) and NSEA4/EID3 protein (Fig. 4). Thus, biological variation certainly exists among iPSCs, but the significance of this variation depends on what is being studied. Our results demonstrate that the most robust disease-related abnormalities can be detected despite the variability inherent across iPSC lines.

**Relevance for LND.** The results described here also demonstrate the value of both hypothesis-driven and unbiased approaches to understanding the biology of disease. LND is caused by defects in a known gene encoding a known protein with a known function. This knowledge makes it possible to hypothesize and then verify findings from studies using iPSCs, such as reductions in *HPRT1* mRNA in gene expression studies, loss of *HGprt* protein in proteomic studies, absence of *HGprt* enzymic activity, resistance to 6-TG, and specific changes in purine metabolism.

The unbiased studies further revealed some novel and unexpected findings relating to the loss of *HGprt*, such as marked and consistent reductions in *FAR2P1* mRNA in the LND iPSCs (Fig. 3). The reasons for changes in *FAR2P1* are not immediately obvious. It is thought to be a pseudogene with no known function. Further, there is no obvious connection between *FAR2P1* and *HPRT1*, the known biology of *HGprt*, or purine metabolism. Similarly, there was a significant reduction in a peptide mapping to both NSEA4 and/or EID3 in proteomic studies of the LND iPSCs (Fig. 4). Once again, the reasons for these changes are not clear. NSEA4 is thought to play a role in maintaining chromosome structure, and EID3 plays a role in suppressing gene transcription from nuclear receptors. Neither protein bears any obvious relationship to the biology of *HGprt*. These findings and their relationship to the biology of LND will require further studies.

*HGprt* deficiency appears to have minimal impact on intracellular purines in iPSCs, possibly because they depend more on de novo purine synthesis, rather than purine recycling, for maintaining purines. There are significant variations in dependence of different cells on synthesis versus salvage pathways to maintain purines. Tissue differences are most obvious in bone marrow cells and particularly mature erythrocytes, which lack a functional de novo synthetic pathway because of the absence of the first enzyme in the pathway<sup>45,46,73</sup>. Indeed, the dependence of erythrocytes on *HGprt*-mediated salvage of hypoxanthine supplied into the circulation by the liver<sup>74</sup> may explain why macrocytic anemia is so common in LND<sup>9,45</sup>. Similarly, de novo purine synthesis seems to contribute to cell division and neurogenesis in the developing brain<sup>56,75–77</sup>, but the mature brain depends more on purine salvage<sup>55,56,78</sup>. Within the brain, neurons seem to express much more *HGprt* than glia<sup>79</sup>. In keeping with the concept that the ratio of de novo synthesis to salvage synthesis of purines is low in the adult brain, the Human Protein Atlas indicates high levels of *HGprt* across almost all brain regions, with variable or low levels of the 6 enzymes in the de novo pathway ([www.proteinatlas.org](http://www.proteinatlas.org)). However, the distribution and levels of each enzyme of purine synthesis in various types of neurons, and whether they can aggregate into purinosomes needed to drive purine synthesis, has never been methodically mapped out<sup>47,80,81</sup>. Although the exact mechanisms remain to be disclosed, the dependence of specific neurons on *HGprt*-mediated purine salvage to maintain purines may explain the prominent neurobehavioral abnormalities of LND<sup>82,83</sup>. Thus, certain cell types may experience significant purine deficiency in LND, even when the undifferentiated LND iPSCs do not. Hypotheses relating to cell-specific changes in purines can now be tested by differentiating the LND iPSCs into disease-relevant cell types.

## Methods

**Fibroblast cultures.** All procedures involving human subjects followed appropriate institutional and national guidelines and were approved by the Emory University Institutional Review Board. For all control subjects who provided data or materials for these studies, informed written consent was obtained. For LND subjects who have substantial cognitive impairment and usually cannot hold a pen, informed written consent was obtained from their legal guardians; verbal assent also was obtained from subjects who could speak. The diagnosis

of LND was made according to published criteria and was based on the presence of all elements of the classical clinical phenotype including self-injurious behavior, motor impairments, intellectual disability, and evidence for overproduction of uric acid<sup>6</sup>. The diagnosis was verified by biochemical testing showing reduced HGprt enzyme activity in fibroblasts or blood cells and/or an associated genetic variant predicting null HGprt enzyme activity. Cultures of human skin fibroblasts were prepared as previously described<sup>3</sup>.

**Reprogramming methods to generate induced pluripotent stem cells.** Fibroblasts were reprogrammed to iPSCs, using a microRNA/mRNA Reprogramming method (Stemgent, Cambridge MA). Two independent lines for each of 3 unrelated LND cases were prepared, for a total of 6 lines. Two independent lines for each of 3 unrelated normal controls (n = 6) were prepared in parallel. For each fibroblast culture, one well of a 6-well plate was seeded with  $5.0 \times 10^3$  cells. Twelve days of transfections with in vitro transcribed mRNAs encoding the human transcription factors Oct4, Sox2, Klf4, c-Myc, Lin28 and nGFP mRNA followed. In addition, transfections at day 1 and day 5 included a proprietary miRNA cocktail. Colonies began to emerge at day 13–15. Colonies with typical iPSC morphology were manually passaged approximately 4 times, before being enzymatically passaged using Accutase (Stemcell Technologies). Cultures were typically maintained on Matrigel-coated 6-well plates (BD Biosciences) in an atmosphere of 5% CO<sub>2</sub> at 37 °C in mTeSR1 medium (Stemcell Technologies) supplemented with 0.5% penicillin–streptomycin. The majority of studies were conducted after 10 passages.

**Genetic evaluation of iPSCs.** For each iPSC line, one Matrigel-coated T25 flask of cells was grown to 80% confluency, harvested, and trypsin G-banded. A minimum of 20 metaphase cells were evaluated for karyotypic abnormalities (Wicell Research Institute Cytogenetics Lab, Madison WI). At least 4 metaphases were photographed with a minimum of 400 band resolution.

The original *HPRT1* genetic variant was confirmed in the reprogrammed iPSCs by RT-PCR. RNA was extracted using TriZol Reagent (Life Technologies, Carlsbad CA) and the concentration of RNA was determined using a NanoDrop spectrophotometer 1000 (Thermo Fisher Scientific, Rockford IL). RT-PCR was performed using Superscript One-Step for Long Template Kit (Life Technologies, Carlsbad CA) with 1 µg RNA (forward primer GCGAACCTCTGGGCTTTC, reverse primer AAGCTCTACTAAGCAGATGGCCACAGAACTAGA). After assessing the quality of the product by gel electrophoresis, cDNA was purified using the QIAquick PCR purification Kit (Qiagen, Venlo, The Netherlands) and samples, premixed with the primers, were sequenced (Genewiz, South Plainfield NJ). The sequencing result was analyzed by aligning it to the sequence of *HPRT1* (NM\_000194.2) using BLAST aligning tool.

**Assessment of pluripotency.** Several methods were used to assess pluripotency potential including immunostaining for pluripotency markers (SSEA3, SSEA4, TRA1-60, TRA1-81, NANOG), assessing expression of pluripotency genes by gene expression profiling, and differentiating the lines towards each of the 3 major germ cell layers. Primary antibodies for staining included anti-Tra1-81 (mouse, R&D #MAB8495), Anti-Tra1-60 (mouse, R&D #MAB4770), Anti-NANOG (rabbit, R&D # AF1997), Anti-SSEA-3 (mouse, R&D #MAB1434) or Anti-SSEA-4 (rat, Abcam #ab16287). Secondary antibodies included anti-mouse Alexa Fluor 488 and Alexa Fluor 594, anti-rabbit Alexa Fluor 594, anti-rat Alexa Fluor 488 or anti-goat Alexa Fluor 488 (Life Technologies, Carlsbad CA) diluted 1:1000 in 5% BSA-PBS. Cells were visualized using an Olympus DP72 Microscope (Olympus, Shinjuku, Tokyo, Japan). All antibodies used had their staining characteristics verified using the previously validated KOLF-2 iPSC line<sup>84</sup>. All control and LND lines were subject to at least 2 separate staining runs for all markers, with or without a nuclear counterstain (DAPI) to identify unstained cells.

The STEMdiff Trilineage Differentiation Kit (Stemcell Technologies, Vancouver BC) was used to assess competence for differentiation into the 3 major germ cell layers (endoderm, ectoderm and mesoderm). For ectoderm, antibodies targeted PAX6 (rabbit, Abcam #ab195045) and Nestin (mouse, Stemcell Technologies #60,091). For mesoderm, antibodies targeted Brachyury (goat, R&D #AF2085) and NCAM (mouse, Stemcell Technologies #60,021). For endoderm, antibodies targeted SOX17 (goat, R&D Systems #AF1924) and FOXA2 (mouse, BD Pharmingen #561,580). All lines were subject to at least 2 separate staining runs for all markers, with or without a nuclear counterstain (DAPI).

**Gene expression profiling.** Prior studies of the RNAseq method have suggested ideal comparisons should involve at least 6 biological replicates<sup>85</sup>. As a result, each of the 6 control and 6 LND lines were grown to approximately 80% confluency, harvested by enzymatic release and centrifugation. At least 1 million cells per tube were flash-frozen in a mixture of ethanol and dry ice, and stored at -80 °C. Once samples from all lines were ready for simultaneous processing, RNA was isolated with the Qiagen miRNEasy kit with on-column DNase treatment. Variance from biological replicates is far greater than variance from technical replicates, so the value of technical replicates is limited<sup>86</sup>. As a result, RNAseq measures were conducted only once for each cell line. RNA concentrations were measured with UV absorbance on a Tecan Infinite m200 pro. RNA profiles were assessed with the Agilent 2100 Bioanalyzer and Agilent RNA 6000 Nano Chip assay. RNA sequencing (RNA-seq) was conducted at HudsonAlpha (Birmingham, Alabama). A total of 50 million paired-end, 100 bp reads were generated per library (100 million total reads per sample). Base calling and FASTQ generation were processed using Illumina's HiSeq Control Software version 1.5.15.1 (RTA v1.13.18 and bclfastq v 1.8.3).

Raw Illumina sequencing reads were first checked by FastQC (0.11.4) for quality and then aligned to the human indexed reference genome (UCSC RefGene; hg38 build, 26,485 genes) using STAR (Spliced Transcripts Alignment to a Reference)<sup>87</sup>. Quantification of gene expression was done using HTSeq-count<sup>88</sup>. To improve the fidelity of quantitative results, transcripts with very low counts across all libraries were filtered out<sup>89</sup>. This filtering strategy applied using filterByExpr function (default settings) from the edgeR (3.28.0) package retained

17,113 genes for differential expression analyses. Differential expression was examined using both edgeR and DESeq2 (1.26.0)<sup>86</sup>. Raw count data from HTSeq-count were imported into R (version 3.6.1; R Core Team, 2019, “R: A language and environment for statistical computing”, R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>) where they were filtered, normalized and identified. For edgeR, counts per million (CPM) data were generated with the trimmed method of M-values (TMM), as implemented in the Bioconductor package edgeR<sup>90</sup>. CPM is defined as read counts scaled by the number of sequenced fragments multiplied by 1 million. Correction for multiple comparisons employed the Benjamini–Hochberg method for the False Discovery Rate (FDR), and data are expressed at multiple thresholds. DESeq2 involves normalization of raw read counts using median ratio normalization where the geometric mean is calculated for each transcript across all samples, and the count of a transcript in each sample is then divided by this mean. DESeq2 uses the Wald test for significance testing.

Ingenuity Pathway Analysis (Qiagen Bioinformatics, Redwood City CA) was used to group significantly affected genes into biological pathways. This analysis was conducted on the 403 transcripts differentially expressed between LND and control samples at FDR < 0.05 for both edgeR and DESeq2 methods. Output included two statistical measures. The first (p-value) represents the probability that the correlation between the genes entered did not occur by chance. The second (activation z-score) represents the degree to which observed changes correlate with the expected changes associated with pathway activation or inhibition. Significant pathway activation is reflected by a positive z-score of > 2, (at p < 0.05), while pathway inhibition is indicated by a negative z-score.

**Proteomic analyses.** At least 1 million cells from each line were harvested as described above for gene expression profiles and stored at – 80 °C. Once samples from all lines were ready for simultaneous processing, they were subjected to trypsin digestion followed by LC–MS/MS of resulting peptides on an Orbitrap Fusion mass spectrometer.

All raw data files were processed together in a single run by MaxQuant (version 1.6.7) with the following settings: Group-specific parameters [Variable modifications: Oxidation(M), Acetyl (Protein N-term) and Deamidation (NQ); Digestion: trypsin]; Global parameters [Min. peptide length: 6 and Max. peptide mass [Da]: 6000, Modifications used for protein quantification: Oxidation(M), Acetyl (Protein N-term) and Deamidation (NQ)]; FTMS MS/MS match tolerance: 0.05 Da; Match between runs: true and a maximum of two missed cleavage sites were allowed. Database searches were conducted with the Andromeda search engine with the UniProt-SwissProt human canonical database as a reference and the database of common laboratory contaminants. MaxQuant reports summed intensity for each protein as well as its iBAQ value. Proteins that shared all identified peptides were combined into a single protein group. Peptides that matched multiple proteins were assigned to the protein group with the most unique peptides. MaxQuant employs the MaxLFQ algorithm for label-free quantification. Quantification was performed using unique and multiple-mapping peptides, including those modified by acetylation (N-terminal), oxidation (methionine), and deamidation (NQ). PTQX was used for general quality control<sup>91</sup>.

LFQ intensities of proteins generated from MaxQuant were imported into Perseus (version 1.6.7) to identify differentially expressed proteins (DEPs). Briefly, proteins identified as: only identified by site, reverse and contaminants were removed and protein group LFQ intensities were log<sub>2</sub>-transformed to reduce the effect of outliers and filtered to have 6 valid values (50% of samples). Missing LFQ values were imputed before performing Student’s t-tests. The imputation was achieved by using the “Replace missing values from normal distribution” function. Because prior studies have suggested that routine correction for multiple comparisons using the FDR method may be too conservative for discovery proteomics, differences are presented with both uncorrected p-values and FDR-corrected values. Graphic representations of the data were prepared using either R or GraphPad Prism 8.

**Functional assessment of LND lines.** The functional consequences of the *HPRT1* gene variants were assessed using three approaches. The first was a direct enzyme assay for HGprt from tissue extracts, modified for cultured cells, and adapted for use in a microplate platform<sup>92</sup>. It is based on conversion of radiolabeled hypoxanthine into inosine monophosphate (IMP). Individual iPSC cultures were grown in 6 well dishes. Cells were detached with Accutase (Stemcell Technologies) into single cell suspensions, pelleted by centrifugation at 300×g for 4 min, and mixed with 50 µL of assay buffer consisting of 50 mM HEPES and 15 mM MgCl<sub>2</sub>. Samples were then frozen at – 80 °C for storage. For assay, samples were thawed, sonicated briefly, and centrifuged at 10,000×g for 10 min. A 10 µL aliquot of each was then incubated in a reaction mixture that contained 100 µM [8-<sup>14</sup>C]-hypoxanthine (50 mCi/mmol, Moravek Biochemicals Inc, CA), assay buffer and 10 mM PRPP (Sigma). The assay was allowed to proceed for 30 min at 37 °C. The reaction was then terminated by the addition of 5 µL of 2 M PCA. Tubes were then incubated on ice for 15 min to precipitate potassium-perchlorate. Precipitate was centrifuged and a 5 µL sample of the supernatant was spotted onto microplates with diethylaminoethyl anion exchange filters (Millipore, Billerica MA). Samples were allowed to adsorb for 60 min and free radiolabeled substrate was washed from IMP bound to the filters with 200 µL of H<sub>2</sub>O followed by 200 µL of 50% methanol (3 times) by vacuum filtration using a MultiScreen HTS Vacuum Manifold (Millipore). Then 100 µL of ULTIMA GOLD MV scintillation liquid (Perkin Elmer, Waltham, MA) was added and plates were stored at room temperature for 24 h before counting with a 2450 MicroBeta Microplate Scintillation Counter (Perkin Elmer, Finland). To accommodate slight changes in cell density at harvest, HGprt enzyme activity was normalized to total protein concentrations using the Pierce BCA kit (Thermo Fisher Scientific, Rockford IL).

The second approach to functional assessment of the *HPRT1* gene variants involved growth in medium supplemented with 6-thioguanine (6TG), an inert prodrug that is phosphoribosylated into a toxic nucleotide by HGprt<sup>93</sup>. Normal cells expressing HGprt die after 1–2 days of exposure, while HGprt-deficient cells grow normally. For each iPSC line, 50,000 cells were plated in 24-well plates in mTeSR1 media (Stemcell Technologies),



with or without 30  $\mu\text{M}$  6TG (Sigma A-4660). Medium was replaced every other day. Cells survival was assessed on days 2, 4, and 6.

The third approach for functional assessment involved measuring intracellular purine levels and purine waste products released into the tissue culture medium. Cultures were grown to 80% confluency in triplicate samples (at least 1 million cells per sample). To measure purines released into culture medium, fresh medium was replaced 24 h before collecting a 450  $\mu\text{L}$  aliquot of the culture medium, and mixed with 50  $\mu\text{L}$  of 1 M perchloric acid and stored at  $-80^\circ\text{C}$ . To measure intracellular purines, the cells were removed with Accutase and pelleted at  $500\times g$  for 5 min. Purines were measured by high performance liquid chromatography with photodiode array ultraviolet detection (HPLC–UV) as previously described<sup>3,19,42</sup>. This method resolves and quantifies the most abundant biologically relevant purines including ATP, ADP, AMP, adenosine, adenine, GTP, GDP, GMP, guanosine, guanine, hypoxanthine, inosine, xanthine and uric acid<sup>17,42</sup>. Purine analytes were normalized to total protein with the Pierce BCA kit (Thermo Fisher Scientific, Rockford, IL).

## Data availability

Data or materials described in this report may be requested from the authors.

Received: 31 July 2020; Accepted: 24 March 2021

Published online: 19 April 2021

## References

1. Fu, R. *et al.* Genotype-phenotype correlations in neurogenetics: Lesch-Nyhan disease as a model disorder. *Brain* **137**, 1282–1303 (2014).
2. Fu, R. & Jinnah, H. A. Genotype-phenotype correlations in Lesch-Nyhan disease: Moving beyond the gene. *J. Biol. Chem.* **287**, 2997–3008 (2012).
3. Fu, R. *et al.* Clinical severity in Lesch-Nyhan disease: The role of residual enzyme and compensatory pathways. *Mol. Genet. Metab.* **114**, 55–61 (2015).
4. Uhlen, M. *et al.* Proteomics Tissue-based map of the human proteome. *Science* **347**, 1260419. <https://doi.org/10.1126/science.1260419> (2015).
5. Puig, J. G. *et al.* The spectrum of hypoxanthine-guanine phosphoribosyltransferase deficiency: Clinical experience based on 22 patients from 18 Spanish families. *Medicine* **80**, 102–112 (2001).
6. Jinnah, H. A. *et al.* Delineation of the motor disorder of Lesch-Nyhan disease. *Brain* **129**, 1201–1217 (2006).
7. Schretlen, D. S. *et al.* Behavioral aspects of Lesch-Nyhan disease and its variants. *Dev. Med. Child Neurol.* **47**, 673–677 (2005).
8. Schretlen, D. S., Harris, J. C., Park, K. S., Jinnah, H. A. & Ojeda del Pozo, N. Neurocognitive functioning in Lesch-Nyhan disease and partial hypoxanthine-guanine phosphoribosyltransferase deficiency. *J. Int. Neuropsychol. Soc.* **7**, 805–812 (2001).
9. Cakmakli, H. *et al.* Macrocytic anemia in Lesch-Nyhan disease and its variants. *Genet. Med.* **21**, 353–360 (2019).
10. Kuehn, M. R., Bradley, A., Robertson, E. J. & Evans, M. J. A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* **326**, 295–298 (1987).
11. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 (1987).
12. Meek, S. *et al.* Reduced levels of dopamine and altered metabolism in brains of HPRT knock-out rats: A new rodent model of Lesch-Nyhan Disease. *Sci. Rep.* **6**, 25592 (2016).
13. Yin, M. *et al.* Generation of hypoxanthine phosphoribosyltransferase gene knockout rabbits by homologous recombination and gene trapping through somatic cell nuclear transfer. *Sci. Rep.* **5**, 16023 (2015).
14. Egami, K. *et al.* Basal ganglia dopamine loss due to defect in purine recycling. *Neurobiol. Dis.* **26**, 396–407 (2007).
15. Bitler, C. M. & Howard, B. D. Dopamine metabolism in hypoxanthine-guanine phosphoribosyltransferase-deficient variants of PC12 cells. *J. Neurochem.* **47**, 107–112 (1986).
16. Yeh, J., Zheng, S. & Howard, B. D. Impaired differentiation of HPRT-deficient dopaminergic neurons: a possible mechanism underlying neuronal dysfunction in Lesch-Nyhan syndrome. *J. Neurosci. Res.* **53**, 78–85 (1998).
17. Lewers, J. C. *et al.* Consequences of impaired purine recycling in dopaminergic neurons. *Neuroscience* **152**, 761–772 (2008).
18. Ceballos-Picot, I. *et al.* Hypoxanthine-guanine phosphoribosyl transferase regulates early developmental programming of dopamine neurons: implications for Lesch-Nyhan disease pathogenesis. *Hum. Mol. Genet.* **18**, 2317–2327 (2009).
19. Goettle, M., Burhenne, H., Sutcliffe, D. & Jinnah, H. A. Purine metabolism during neuronal differentiation: The relevance of purine synthesis and recycling. *J. Neurochem.* **127**, 805–818 (2013).
20. Zoref-Shani, E., Bromberg, Y., Brosh, S., Sidi, Y. & Sperling, O. Characterization of alterations in purine nucleotide metabolism in hypoxanthine-guanine phosphoribosyltransferase-deficient rat neuroma cell line. *J. Neurochem.* **61**, 457–463 (1993).
21. Jinnah, H. A., Langlais, P. J. & Friedmann, T. Functional analysis of brain dopamine systems in a genetic mouse model of Lesch-Nyhan syndrome. *J. Pharmacol. Exp. Ther.* **263**, 596–607 (1992).
22. Jinnah, H. A. *et al.* Dopamine deficiency in a genetic mouse model of Lesch-Nyhan disease. *J. Neurosci.* **14**, 1164–1175 (1994).
23. Keebaugh, A. C. *et al.* PRTFDC1 is a genetic modifier of HPRT-deficiency in the mouse. *PLoS ONE* **6**, e22381 (2011).
24. Wu, X. *et al.* Hyperuricemia and urate nephropathy in urate oxidase-deficient mice. *Proc. Natl. Acad. Sci. (USA)* **91**, 742–746 (1994).
25. Engle, S. J. *et al.* HPRT-APRT deficient mice are not a model for Lesch-Nyhan syndrome. *Hum. Mol. Genet.* **5**, 1607–1610 (1996).
26. Pesi, R. *et al.* Cytosolic 5'-nucleotidase hyperactivity in erythrocytes of Lesch-Nyhan syndrome patients. *NeuroReport* **11**, 1827–1831 (2000).
27. Micheli, V. *et al.* Hypoxanthine-guanine phosphoribosyltransferase deficiency and erythrocyte synthesis of pyridine coenzymes. *Life Sci.* **64**, 2479–2487 (1999).
28. Simmonds, H. A., Reiter, S., Davies, P. M. & Cameron, J. S. Orotidine accumulation in human erythrocytes during allopurinol therapy: Association with high urinary oxypurinol-7-riboside concentrations in renal failure and in the Lesch-Nyhan syndrome. *Clin. Sci.* **80**, 191–197 (1991).
29. Micheli, V., Simmonds, H. A. & Ricci, C. Regulation of nicotinamide-adenine dinucleotide synthesis in erythrocytes of patients with hypoxanthine-guanine phosphoribosyltransferase deficiency and a patient with phosphoribosylpyrophosphate synthetase superactivity. *Clin. Sci.* **78**, 239–245 (1990).
30. Simmonds, H. A. *et al.* Central nervous system dysfunction and erythrocyte guanosine triphosphate depletion in purine nucleoside phosphorylase deficiency. *Arch. Dis. Child.* **62**, 385–391 (1987).
31. Fairbanks, L. D., Simmonds, H. A. & Webster, D. R. Use of intact erythrocytes in the diagnosis of inherited purine and pyrimidine disorders. *J. Inher. Metab. Dis.* **10**, 174–186 (1987).

32. Sidi, Y. & Mitchell, B. S. Z-nucleotide accumulation in erythrocytes from Lesch-Nyhan patients. *J. Clin. Invest.* **76**, 2416–2419 (1985).
33. Fairbanks, L. D., Jacomelli, G., Micheli, V., Slade, T. & Simmonds, H. A. Severe pyridine nucleotide depletion in fibroblasts from Lesch-Nyhan patients. *Biochem. J.* **366**, 265–272 (2002).
34. Hussain, S. P., Duley, J. A. & Connolly, G. P. Purine and pyrimidine levels in cultured skin fibroblasts from control and Lesch-Nyhan fibroblasts. *Cell. Mol. Biol. Lett.* **4**, 377 (1999).
35. Kelley, W. N. & Meade, J. C. Studies on hypoxanthine-guanine phosphoribosyltransferase in fibroblasts from patients with the Lesch-Nyhan syndrome. Evidence for genetic heterogeneity. *J. Biol. Chem.* **246**, 2953–2958 (1971).
36. Schretlen, D. J. *et al.* Do clinical features of Lesch-Nyhan disease correlate more closely with hypoxanthine or guanine recycling?. *J. Inherit. Metab. Dis.* **39**, 85–91 (2015).
37. Skaper, S. D. & Schafer, I. A. Monoamino oxidase activity reduced in cultured human fetal cells deficient in hypoxanthine-guanine phosphoribosyltransferase activity. *Biochem. Genet.* **16**, 1135–1138 (1978).
38. Becker, M. A., Losman, M. J. & Kim, M. Mechanisms of accelerated purine nucleotide synthesis in human fibroblasts with superactive phosphoribosylpyrophosphate synthetases. *J. Biol. Chem.* **262**, 5596–5602 (1987).
39. Kang, T. H., Guibinga, G. H., Jinnah, H. A. & Friedmann, T. HPRT deficiency coordinately dysregulates canonical Wnt and pre-senilin-1 signaling: A neuro-developmental regulatory role for a housekeeping gene?. *PLoS ONE* **6**, e16572 (2011).
40. Guibinga, G. H., Hrustanovic, G., Bouic, K., Jinnah, H. A. & Friedmann, T. MicroRNA-mediated dysregulation of neural developmental genes in HPRT deficiency: Clues for Lesch-Nyhan disease?. *Hum. Mol. Genet.* **21**, 609–622 (2012).
41. Guibinga, G. H., Hsu, S. & Friedmann, T. Deficiency of the housekeeping gene hypoxanthine-guanine phosphoribosyltransferase (HPRT) dysregulates neurogenesis. *Mol. Ther.* **15**, 54–62 (2010).
42. Shirley, T. L. *et al.* A human neuronal tissue culture model for Lesch-Nyhan disease. *J. Neurochem.* **101**, 841–853 (2007).
43. Cristini, S. *et al.* Human neural stem cells: a model system for the study of Lesch-Nyhan disease neurological aspects. *Hum. Mol. Genet.* **19**, 1939–1950 (2010).
44. Mastrangelo, L., Kim, J. E., Miyanohara, A., Kang, T. H. & Friedmann, T. Purinergic signaling in human pluripotent stem cells is regulated by the housekeeping gene encoding hypoxanthine guanine phosphoribosyltransferase. *Proc. Natl. Acad. Sci. USA* **109**, 3377–3382 (2012).
45. McKeran, R. O., Howell, A., Andrews, T. M., Watts, R. W. E. & Arlett, C. F. Observations on the growth in vitro of myeloid progenitor cells and fibroblasts from hemizygotes and heterozygotes for complete and partial hypoxanthine-guanine phosphoribosyltransferase (HGPRT) deficiency, and their relevance to the pathogenesis of brain damage in the Lesch-Nyhan syndrome. *J. Neurol. Sci.* **22**, 183–195 (1974).
46. Murray, A. W. The biological significance of purine salvage. *Ann. Rev. Biochem.* **40**, 811–826 (1971).
47. Moriwaki, Y., Yamamoto, T. & Higashino, K. Enzymes involved in purine metabolism: A review of histochemical localization and functional implications. *Histol. Histopathol.* **14**, 1321–1340 (1999).
48. Lane, A. N. & Fan, T. W. Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Res.* **43**, 2466–2485 (2015).
49. Burnstock, G., Krugel, U., Abbraccio, M. P. & Illes, P. Purinergic signalling: from normal behaviour to pathological brain function. *Prog. Neurobiol.* **95**, 229–274 (2011).
50. Lo Sardo, V. *et al.* Influence of donor age on induced pluripotent stem cells. *Nat. Biotechnol.* **35**, 69–74 (2017).
51. Jiang, L. *et al.* A quantitative proteome map of the human body. *Cell* **183**, 269–283. <https://doi.org/10.1016/j.cell.2020.08.036> (2020).
52. Pascovici, D., Handler, D. C., Wu, J. X. & Haynes, P. A. Multiple testing corrections in quantitative proteomics: A useful but blunt tool. *Proteomics* **16**, 2448–2453 (2016).
53. Pedley, A. M. & Benkovic, S. J. A new view into the regulation of purine metabolism: The purinosome. *Trends Biochem. Sci.* **42**, 141–154 (2017).
54. Harvey, A., Caretti, G., Moresi, V., Renzini, A. & Adamo, S. Interplay between metabolites and the epigenome in regulating embryonic and adult stem cell potency and maintenance. *Stem Cell Rep.* **13**, 573–589. <https://doi.org/10.1016/j.stemcr.2019.09.003> (2019).
55. Natsumeda, Y., Prajda, N., Donohue, J. P., Glover, J. L. & Weber, G. Enzymic capacities of purine de Novo and salvage pathways for nucleotide synthesis in normal and neoplastic tissues. *Cancer Res.* **44**, 2475–2479 (1984).
56. Allsop, J. & Watts, R. W. E. Purine synthesis de novo and salvage in hypoxanthine phosphoribosyltransferase-deficient mice. *Biochem. Soc. Trans.* **17**, 578–579 (1989).
57. Wang, X. *et al.* Purine synthesis promotes maintenance of brain tumor initiating cells in glioma. *Nat. Neurosci.* **20**, 661–673 (2017).
58. Giotti, B. *et al.* Assembly of a parts list of the human mitotic cell cycle machinery. *J. Mol. Cell Biol.* **11**, 703–718 (2019).
59. Freel, B. A., Sheets, J. N. & Francis, K. R. iPSC modeling of rare pediatric disorders. *J. Neurosci. Methods* **332**, 108533. <https://doi.org/10.1016/j.jneumeth.2019.108533> (2020).
60. Mekhoubad, S. *et al.* Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell* **10**, 595–609 (2012).
61. Ruillier, V. *et al.* Rescuing compounds for Lesch-Nyhan disease identified using stem cell-based phenotypic screening. *JCI Insight* <https://doi.org/10.1172/jci.insight.132094> (2020).
62. Frank, S., Skryabin, B. V. & Greber, B. A modified TALEN-based system for robust generation of knock-out human pluripotent stem cell lines and disease models. *BMC Genom.* **14**, 773 (2013).
63. Koyanagi-Aoi, M. *et al.* Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. *Proc. Natl. Acad. Sci. USA* **110**, 20569–20574 (2013).
64. Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67 (2011).
65. Bock, C. *et al.* Reference maps of human ES and iPSC cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452 (2011).
66. Vitale, A. M. *et al.* Variability in the generation of induced pluripotent stem cells: Importance for disease modeling. *Stem Cells Transl. Med.* **1**, 641–650 (2012).
67. Nazor, K. L. *et al.* Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–634 (2012).
68. Narsinh, K. H. *et al.* Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. *J. Clin. Invest.* **121**, 1217–1221 (2011).
69. Volpato, V. & Webber, C. Addressing variability in iPSC-derived models of human disease: Guidelines to promote reproducibility. *Dis. Model Mech.* <https://doi.org/10.1242/dmm.042317> (2020).
70. Steichen, C., Hannoun, Z., Luce, E., Hauet, T. & Dubart-Kupperschmitt, A. Genomic integrity of human induced pluripotent stem cells: Reprogramming, differentiation and applications. *World J. Stem Cells* **11**, 729–747. <https://doi.org/10.4252/wjsc.v11.i10.729> (2019).
71. Weissbein, U., Benvenisty, N. & Ben-David, U. Quality control: Genome maintenance in pluripotent stem cells. *J. Cell. Biol.* **204**, 153–163 (2014).
72. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521–531. <https://doi.org/10.1016/j.stem.2010.07.017> (2010).
73. Fontenelle, L. J. & Henderson, J. F. An enzymatic basis for the inability of erythrocytes to synthesize purine ribonucleotides de novo. *Biochim. Biophys. Acta* **177**, 175–176. [https://doi.org/10.1016/0304-4165\(69\)90085-3](https://doi.org/10.1016/0304-4165(69)90085-3) (1969).

74. Lajtha, L. G. & Vane, J. R. Dependence of bone marrow cells on the liver for purine supply. *Nature* **182**, 191–192 (1958).
75. Lo, Y. V. & Palmour, M. Developmental expression of murine HPRT. I. Activities, heat stabilities, and electrophoretic mobilities in adult tissues. *Biochem. J.* **17**, 737–746 (1979).
76. Planet, G. & Willemot, J. Changes in purine phosphoribosyltransferase activities in mouse brain, liver and muscle with age. *Biochem. Biophys. Acta* **364**, 236–242 (1974).
77. Yamada, S., Sato, A. & Sakakibara, S. I. Nwd1 regulates neuronal differentiation and migration through purinosome formation in the developing cerebral cortex. *iScience* **23**, 101058. <https://doi.org/10.1016/j.isci.2020.101058> (2020).
78. Howard, W. J., Kerson, L. A. & Appel, S. H. Synthesis de novo of purines in slices of rat brain and liver. *J. Neurochem.* **17**, 121–123 (1970).
79. Jinnah, H. A., Hess, E. J., Wilson, M. C., Gage, F. H. & Friedmann, T. Localization of hypoxanthine-guanine phosphoribosyltransferase mRNA in the mouse brain by in situ hybridization. *Mol. Cell. Neurosci.* **3**, 64–78 (1992).
80. Williamson, J., Petralia, R. S., Wang, Y. X., Mattson, M. P. & Yao, P. J. Purine biosynthesis enzymes in hippocampal neurons. *Neuromol. Med.* <https://doi.org/10.1007/s12017-017-8466-6> (2017).
81. Mangold, C. A. *et al.* Expression of the purine biosynthetic enzyme phosphoribosyl formylglycinamide synthase in neurons. *J. Neurochem.* **144**, 723–735. <https://doi.org/10.1111/jnc.14304> (2018).
82. McKeran, R. O. Factors in the pathogenesis of the brain damage and anaemia in the Lesch-Nyhan syndrome. *CIBA Found. Symp.* **48**, 83–96 (1977).
83. Visser, J. E., Baer, P. R. & Jinnah, H. A. Lesch-Nyhan syndrome and the basal ganglia. *Brain Res. Rev.* **32**, 449–475 (2000).
84. Skarnes, W. C., Pellegrino, E. & McDonough, J. A. Improving homology-directed repair efficiency in human stem cells. *Methods* **164–165**, 18–28. <https://doi.org/10.1016/j.ymeth.2019.06.016> (2019).
85. Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol.* **17**, 13 (2016).
86. Schurch, N. J. *et al.* How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. *RNA* **22**, 839–851 (2016).
87. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
88. Anders, S., Pyl, P. T. & Huber, W. HTSeq: A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
89. Sha, Y., Phan, J. H. & Wang, M. D. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* <https://doi.org/10.1109/EMBC.2015.7319872> (2015).
90. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).
91. Bielow, C., Mastrobuoni, G. & Kempa, S. Proteomics quality control: Quality control software for MaxQuant results. *J. Proteome Res.* **15**, 777–787. <https://doi.org/10.1021/acs.jproteome.5b00780> (2016).
92. Jinnah, H. A., Page, T. & Friedmann, T. Brain purines in a genetic mouse model of Lesch-Nyhan disease. *J. Neurochem.* **60**, 2036–2045 (1993).
93. Nelson, J. A., Carpenter, J. W., Rose, L. M. & Adamson, D. J. Mechanisms of action of 6-thioguanine, 6-mercaptopurine, and 8-azaguanine. *Cancer Res.* **35**, 2872–2878 (1975).

## Acknowledgements

This work was supported by the Lesch-Nyhan Syndrome Children's Research Foundation, the National Institute for Neurological Diseases and Stroke (NINDS) at the National Institutes of Health (NIH) through grants R56 NS102980 and R01 NS109242, a Netherlands Organisation for Scientific Research (NWO) ZonMW/Veni (916.12.167) grant, and a Hersenstichting Fellowship (F2014(1)-16). Additional support was provided in part by grants to the Dystonia Coalition (U54 TR001456 and U54 NS116025), a consortium of the Rare Diseases Clinical Research Network (RDCRN) that is supported by the Office of Rare Diseases Research (ORDR) at the National Center for Advancing Clinical and Translational Studies (NCATS) in collaboration with the NINDS. Additional support came in part from the Emory Integrated Genomics Core (EIGC), which is subsidized by the Emory University School of Medicine and is one of the Emory Integrated Core Facilities, as well as the Georgia Clinical & Translational Science Alliance of the National Institutes of Health under Award Number UL1TR002378. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health.

## Author contributions

D.J.S.: Conception and Design, Collection & assembly of data, Data analysis & interpretation, Manuscript writing, Final approval of manuscript. A.R.D.: Collection & assembly of data, Data analysis & interpretation, Manuscript writing, Final approval of manuscript. J.E.V.: Conception and Design, Collection & assembly of data, Financial support, Final approval of manuscript. J.D.H.: Collection & assembly of data, Data analysis & interpretation, Final approval of manuscript. E.S.: Collection & assembly of data, Data analysis & interpretation, Final approval of manuscript. P.J.: Collection & assembly of data, Data analysis & interpretation, Final approval of manuscript. I.C.P.: Conception and Design, Final approval of manuscript. T.S.: Collection & assembly of data, Final approval of manuscript. E.J.H.: Data analysis & interpretation, Final approval of manuscript. Y.V.S.: Conception and Design, Data analysis & interpretation, Final approval of manuscript. Z.W.: Conception and Design, Data analysis & interpretation, Final approval of manuscript. M.E.Z.: Conception and Design, Data analysis & interpretation, Final approval of manuscript. H.A.J.: Conception and Design, Provision of study materials, Collection & assembly of data, Data analysis & interpretation, Financial Support, Manuscript writing, Final approval of manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87955-9>.

**Correspondence** and requests for materials should be addressed to H.A.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021