bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A Higher Proportion of Craniosynostosis Genes Are Cancer Driver Genes

Suchir Misra[1], Andrew Shih[1], Xiao-Jie Yan[2], Wentian Li[1]

1. The Robert S. Boas Center for Genomics and Human Genetics

2. The Karches Center for Chronic Lymphocytic Leukemia Research

The Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA

December 10, 2019

ABSTRACT

Craniosynostosis (CRS) is a congenital abnormality deformity with a heterogenous genetic contribution. Previously, there are two attempts to collect genes that are genetically associated with craniosynostosis and some related syndromes with 57 (Twigg and Wilkie, 2015) and 39 (Goos and Mathijssen, 2019) genes identified, respectively. We expanded this list of craniosynostosis genes by adding another 17 genes with an updated literature search. These genes are shown to be more likely to be intolerant to functional mutations. Of these 113 craniosynostosis genes, 21 (19% vs. 1.5% baseline frequency) are cancer driver genes, a 14-fold enrichment. The cancer-craniosynostosis connection is further validated by an over-representation analysis of craniosynostosis genes in KEGG cancer pathway and several cancer related gene-sets. Many cancer-craniosynostosis overlapping genes participate in intracellular signaling pathways, which play a role in both development and cancer. This connection can be viewed from the oncogenesis recapitulates ontogenesis framework. Nineteen craniosynostosis genes are transcription factor genes (16.8% vs. 8.2% baseline), and craniosynostosis genes are also enriched in targets of certain transcription factors or micro RNAs.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

keywords: craniosynostosis; cancer driver genes; gene-set; enrichment analysis; signaling pathways; oncogenesis recapitulates ontogenesis.

abbreviations:

CADD: combined annotation dependent depletion;

CML: chronic myelogenous leukemia;

CNV: copy number variation;

CRS: craniosynostosis;

FGF: fibroblast growth factor;

GDI: gene damage index;

GEO: gene expression omnibus database;

GO: gene ontology;

GS: gene-set;

GWAS: genome-wide association studies;

HUGO: Human Genome Organisation;

IPA: Ingenuity Pathway Analysis;

KEGG: Kyoto Encyclopedia of Genes and Genomes;

LOEUF: loss-of-function observed-over-expected upper bound fraction;

MeSH: medical subject headings;

MSigDB: Molecular Signature Database;

OR: odds-ratio;

PID: pathway interaction database;

pv: p-value of a test;

RVIS: residual variation intolerance score;

SNP: single-nucleotide polymorphism;

TDT: transmission disequilibrium test;

TGF: transforming growth factor;

TF: transcription factor;

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Craniosynostosis is a congenital condition when the suture of a baby's skull does not close in a timely fashion, causing abnormal skull shape (Kabbani and Raghuveer, 2004). This premature fusion of sutures may also lead to other abnormal traits, a situation called a craniosynostosis syndrome (Muenke and Wilkie, 2014; Ko, 2016). There are 100-200 syndromes associated with craniosynostosis, including Apert, Pfeiffer, Crouzon, Antley-Bixler, Saethre-Chotzen, and Muenke syndromes (Wilkie et al., 1995; Carinci et al., 2005); at least 15 of the syndromes are listed in Online Mendelian Inheritance in Man (*www.omim.org*), and close to 40 are listed in National Organization for Rare Disorders (*rarediseases.org*). Craniosynostosis is highly heterogeneous in its potential causes, etiology of individual cases is often unknown (Wilkie, 1997), and has long been known to be a multifactorial disease (Hunter and Rudd, 1976). Through genetic studies, tens of genes are shown to be associated with craniosynostosis (Twigg and Wilie, 2015), providing a foundation for further elucidation of the mechanism of the trait. Like any phenotypes, there should also be environmental contributions to craniosynostosis (Shashi and Hart, 2002; Zeiger et al., 2002; Durham et al., 2017), illustrated by an observation of identical twins with discordant craniosynostosis (Magge et al., 2017).

If a specific gene is part of a known biological pathway, the genes function and its role in a disease might be hypothesized. However, due to the extreme heterogeneity of craniosynostosis, no single pathway stands out as the major pathway for all cases (Twigg and Wilie, 2015). We adopt a different strategy to treat the collection of genes that are genetically associated with craniosynostosis as a single unit (a gene-set, a molecular signature, etc.) (Jürgens, 1985; Subramanian et al., 2005; Li et al., 2015b). One gene-set can be checked with any other gene-sets for overlapping of common genes. If the two gene-sets share more common genes than expected by chance, an over-representation or enrichment is claimed, which provides evidence that the two gene-sets are biologically related (Khatri et al., 2012).

Naturally, the major focus on the molecular understanding of craniosynostosis is on early embryonic developmental pathways, such as the fibroblast growth factor (FGF) signaling pathway (Ornitz and Marie, 2002) that is associated with several downstream transduction pathways of RAS/MAPK (Takenouchi et al., 2014), PI3K/AKT (Dufour et al., 2008), PLCγ (Moen-

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ning et al., 2008), STAT (Lokau and Garbers, 2019), etc. The link between craniosynostosis and cancer was first suggested by the observation that a craniosynostosis-associated gene, FGFR2, experiences a higher than expected somatic mutation rate in endometrial cancer (Pollock et al., 2007).

In this paper, we aim at extending the craniosynostosis-cancer connection from a single gene (i.e., FGFR2) to the level of pathways and gene-sets. Towards this goal, we would like to have a more up-to-date list of genes that are genetically associated with craniosynostosis (i.e., genes that contain, uniquely or preferably, germline mutations in craniosynostosis patients). Previously, there have been two attempts to collect craniosynostosis genes: one in 2015 with 57 genes (Twigg and Wilie, 2015) and a more recent one in 2019 with 39 genes (Goos and Mathijssen, 2019).

Most of these genes are identified through a combination of family studies, candidate gene investigation, mouse model validation, etc. Recently, there are new strategies, such as GWAS (genome-wide association study) on common genetic variants (Justice et al., 2012), and exome sequencing followed by summing all rare variants (Timberlake et al., 2017). Without distinguishing different gene mapping methods, we proceed by a systematic search of all PUBMED titles/abstracts that contain the word of "craniosynostosis and "genetic, finding all gene names, then followed by a manual check of the original literature. An extra 17 genes have been added to the craniosynostosis genes list.

On the other hand, a list of cancer driver genes (i.e., somatic mutations in these genes are positive selected in cancer cell clonal expansion) is readily available with 125 genes in a 2013 study (Vogelstein et al., 2013), and 299 genes in a more recent analysis (Tokheim et al., 2016). We compare our craniosynostosis gene list and cancer driver gene lists to examine if the overlap between them is more than expected by chance.

The result section below is organized by a description of our compilation of craniosynostosis genes; the observation that these genes are more intolerant to functional mutations; these genes are enriched in cancer driver gene lists; an enrichment analysis of these genes in all gene-set from a gene-set database; an enrichment analysis of these genes in the Ingenuity pathway database; and their enrichment in the transcription factor list. The result section is followed by the Method/Data section and Discussion/Conclusion section.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Results

Compilation of a more complete list of craniosynostosis genes

In 2015, a set of 57 genes are summarized as being genetically associated with craniosynostosis (CRS) (Twigg and Wilie, 2015), i.e., variants/mutations are observed in these genes in CRS patients that are less likely to be observed in people without CRS. A more recent paper adds 39 more CRS genes published from 2015-2017 (Goos and Mathijssen, 2019), resulting in a total of 96 genes. The reason for so many genes being linked to the disease can be understood by the perspective of heterogeneity (McClellan and King, 2010).

When some public PUBMED-search programs such as PolySearch (*polysearch.ca*) (Liu et al., 2015) or HuGE Literature Finder (*phgkb.cdc.gov/PHGKB/startPagePubLit.actio*) (Lin et al., 2006) were used, we failed to find any new CRS genes. The reason is that these programs are designed to reduce the number of false positives, aiming at a shorter list of genes, not longer. As these public domain programs did not serve our purposes, we designed our own pipeline via a PUBMED search.

In our approach, we first filtered all PUBMED titles/abstracts by the MeSH term of craniosynostosis and genetic. Then, a list of HUGO approved human gene names was used to find any matching abbreviations in the abstract (two-letter items are ignored). With this raw list of potential gene names, we manually checked each item and the corresponding abstract to determine the appropriateness of the match.

These 17 new genes found through this process were added to the CRS genes list: AXIN2 (Yilmaz et al., 2018), BBS9 (Justice et al., 2012; Sewda et al., 2019), BCOR (O'Byrne et al., 2017) (for Oculo-facio-cardio-dental syndrome, or microphthalmia syndrome), BGLAP (Sowińska-Seidler et al., 2018), COLEC10 (for 3MC syndrome) (Munye et al., 2017), FGFRL1 (Rieckmann et al., 2009) (for Antley-Bixler syndrome), GCK (for Greig cephalopolysyndactyly syndrome) (Zung et al., 2011), LMNA (Sowińska-Seidler et al., 2018), PPP3CA (Mizuguchi et al., 2018), PTH2R (Kim et al., 2015), RAF1 (for Noonan syndrome with multiple lentigines, or leopard syndrome) (Rodríguez et al., 2019), SIX2 (for frontonasal dysplasia syndrome) (Hufnagel et al., 2016), SMURF1, SPRY1, SPRY4 (Timberlake et al., 2016, 2017), TCOF1 (for Treacher Collins syndrome) (Horiuchi et al., 2004), TNFRSF11B (for Juvenile Paget disease)

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(Saki et al., 2013).

There can be various reasons why the above genes are not included in the two collections by Twigg-Wilkie and Goos-Mathijssen. Obviously, if the paper was published after 2018, it would not be in the previous lists. Also, we included gene mutations in rare syndromic cases which include CRS as one of many phenotypes. Note that we did not consider genes causing CRS in animal models (e.g., BMP3 (Schoenebeck et al., 2012), DUSP6 (Li et al., 2007), SNAI1, SNAI2 (Oram and Gridley, 2005)). We also did not consider genes that are differentially expressed in relevant tissues in CRS (e.g., FGF7, SFRP4, VCAM1 (Zhang et al., 2002; Stamper et al., 2011), NELL1 (Zhang et al., 2002), IGF1 (Al-Rekabi et al., 2016)). These genes may still play a role in the etiology of CRS, or, they might be discovered to harbor germline mutations in the future, but are excluded for now to be consistent with the previous lists.

The three compilations ((Twigg and Wilie, 2015), (Goos and Mathijssen, 2019), and here) add up to 113 genes that are genetically associated with craniosynostosis. These 113 craniosynostosis genes are: ABCC9, ADAMTSL4, AHDC1, ALPL, ALX4, ASXL1, ATR, AXIN2, B3GAT3, BBS9, BCOR, BGLAP, BMP2, BRAF, CD96, CDC45, CHST3, COLEC10, COLEC11, CRTAP, CTSK, CYP26B1, DPH1, EFNA4, EFNB1, ERF, ESCO2, FAM20C, FBN1, FGF9, FGFR1, FGFR2, FGFR3, FGFRL1, FLNA, FREM1, FTO, GCK, GLI3, GLIS3, GNAS, GNPTAB, GPC3, HNRNPK, HUWE1, IDS, IDUA, IFT122, IFT140, IFT43, IGF1R, IHH, IL11RA, IL6ST, IRX5, JAG1, KANSL1, KAT6A, KAT6B, KMT2D, KRAS, LMNA, LMX1B, LRP5, MASP1, MED13L, MEGF8, MSX2, NFIA, NTRK2, OSTM1, P4HB, PHEX, POR, PPP1CB, PPP3CA, PTH2R, PTPN11, PTPRD, RAB23, RAF1, RECQL4, RSPRY1, RUNX2, SCARF2, SCN4A, SEC24D, SH3PXD2B, SHOC2, SIX2, SKI, SLC25A24, SMAD6, SMC1A, SMO, SMURF1, SOX6, SPECC1L, SPRY1, SPRY4, STAT3, TCF12, TCOF1, TGFBR1, TGFBR2, TMCO1, TNFRSF11B, TWIST1, WDR19, WDR35, ZEB2, ZIC1, and ZNF462.

More detailed information on the newly added 17 CRS genes are summarized in the Appendix. The gene OSTM1 is labeled as OSTEM1 in (Mahmoud Adel et al., 2013; Goos and Mathijssen, 2019) which is not a HGNC approved name. Also note that BGLAP and LMNA in our new addition are within a 1.26 Mb deletion and each gene was hypothesized as causing the disease (Sowińska-Seidler et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Craniosynostosis genes tend to be intolerant to functional mutations

Sequencing a large number of normal samples allows us to learn which genes tend to have more amino-acid-changing and/or loss-of-function mutations and which genes tend to have less. The former is called "tolerant" and the latter "intolerant". We tested three measures of intolerance with two goals in mind. One is to check whether our CRS gene list as a whole tend to be intolerant to functional mutations. Another is to find out which genes in the list are more tolerant to mutations than others, which flag them as potential false positives.

The RVIS (Residual Variation Intolerance Score) (Petrovski et al., 2013) is derived from the regression of sum of all common (major allele frequency larger than 0.001) functional variants within a gene (Y) over sum of all positions with variant in the gene (X). The residual (difference between the observed Y and the expected value from the regression line), standardized by the standard deviation, is RVIS (Petrovski et al., 2013). Genes with negative (positive) are intolerant (tolerant) to functional mutations. Of 104 CRS genes (out of 113) with RVIS values, 82 are negative (and 19 are positive). The percentage of intolerant CRS genes is 79%, as compared to the baseline percentage of 53% (Fisher's test p-value= 9×10^{-9}). Direct comparison of RVIS in CRS gene list and outside also leads to a significant difference: the t-test p-value is 3.7×10^{-9} and the Mann-Whitney-Wilcoxon test p-value is 1.6×10^{-11}.

The LOEUF (Loss-of-function Observed-over-Expected Upper bound Fraction) is based on a coarse-grained version of the regression of observed loss-of-function variants (Y) over expected (X) (Karczewski et al., 2019), for data from the normal individuals. A large (small) LOEUF value implies a tolerant (intolerant) gene. Of the 113 CRS genes, 95 have LOEUF smaller than 1 (84.1%). This percentage is significantly larger than the baseline proportion (55.5%) (Fisher's test p-value = 1.3×10^{-10}). Direct comparison of LOEUF between CRS genes and the rest leads to the same conclusion (Mann-Whitney-Wilcoxon test p-value is 2.4×10^{-16}; t-test is not used because the distribution of LOEUF is not normal).

The GDI (Gene Damage Index) is calculated by summing all variants' predicted damage impact score according to CADD (Combined Annotation Dependent Depletion) (Kircher et al., 2014) in a normal dataset (Itan et al., 2015). A large (small) GDI implies a tolerant (intolerant) gene. A comparison of the GDI score itself between CRS genes and other genes

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

does not lead to any significant results. In a recent assessment of various variant pathogenicity prediction programs, CADD is not performing well (Niroula and Vihinen, 2019). It may cast doubt on the usefulness of raw GDI score. However, if we use the categorical label on whether a gene has LOW, MEDIUM, or HIGH damage provided by the GDI site, genes showing high level of damaging mutations in normal samples might be potential false positives. According to the assessment from this approach, these genes in the CRS list: FREM1, KANSL1, ALX4, RECQL4, may need further examination. Despite the possibility that some genes in the list may be false positives, as a whole, we show evidence that these genes are intolerant to functional mutations.

Enrichment of craniosynostosis genes in the list of cancer driver genes

A cancer driver gene is defined as those whose mutations "increase net cell growth under the specific microenvironmental conditions that exist in the cell in vivo" (Tokheim et al., 2016), or genes whose mutations "promote tumorigenesis" (Vogelstein et al., 2013). Many cancer driver genes prediction relies on the non-random mutation patterns, instead of mutation frequencies. For example, the 20/20 rule states that for oncogenes, $> 20\%$ of somatic mutations should occur in recurrent positions and to be missense, whereas for tumor suppressing genes, 20% of mutations should be inactivating (loss-of-function, truncating) (Vogelstein et al., 2013).

We obtained the list of 299 cancer driver genes from (Bailey et al., 2018), which is decided by a combination of 26 computer programs. We first tested enrichment of cancer driver genes individually in Twigg-Willie's list, in Goos-Mathijssen's list, and in our newly added list. The results are shown in Table 1. There are 12 (out of 57), 6 (out of 39), 3 (out of 17) CRS genes that are in the cancer driver gene list, for the three collections. The proportions are 21%, 15%, and 18% respectively. On the other hand, the baseline proportion of cancer driver genes can be simply 299/20000 or 1.5%. Clearly, cancer driver genes are greatly enriched in the three CRS gene lists. The overrepresentation test (using Fisher's test) leads to p-values of 6×10^{-11}, 3×10^{-5} and 2×10^{-3}, respectively.

Within Twigg-Wilkie's list, there are 20 CRS genes are labeled as core genes. And 5 of these 20 genes are cancer driver genes (25%). The Goos-Mathijssen's list can be split into

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

two sublists: one with evidence of mutation in multiple persons, and another with mutation observed only on one patient. The cancer driver genes are in 9% and 24% of the two sub-lists. All these proportions are much higher than the 1.5% expected by chance. If we combine the three CRS gene collections (total 113 genes), 21 (or 19%) are cancer driver genes, with the p-value of 1.2×10^{-16}.

To address the question on whether the significant enrichment of cancer driver genes in CRS genes is robust, with respect to the compilation of driver genes itself, we re-ran the analysis using an early cancer driver gene list (125 genes) (Vogelstein et al., 2013). The result is summarized in Table 1 (right). The baseline proportion of this cancer driver genes is $125/20000 = 0.6\%$, whereas the proportion of them in the three CRS gene lists are 8.8%, 7.7%, and 5.9%. The corresponding Fisher's test p-values are 3×10^{-5}, 2×10^{-3}, and 0.1. When all three CRS gene collections are combined, 9 out of 113 (8%) are cancer driver genes, a significant enrichment (p-value= 7×10^{-8}). These results firmly establish the fact that cancer driver genes are over-presented in CRS genes.

Our results can be made more precise by using a more accurate number of human genes (less than 20000), and/or by excluding CRS genes from the comparison group. These changes do not affect the conclusion (result not shown). The 21 cancer driver genes that match the CRS gene list are ASXL1, ATR, AXIN2, BCOR, BRAF, FGFR1, FGFR2, FGFR3, FLNA, GNAS, HUWE1, IL6ST, KANSL1, KMT2D, KRAS, PTPN11, PTPRD, RAF1, SMC1A, TCF12, and TGFBR2.

Enrichment analysis of craniosynostosis genes for all gene-sets in MSigDB

In order to further examine the craniosynostosis-cancer connection, we compared the 113-CRS-gene list with all gene-sets in MSigDB (Liberzon et al., 2015). We checked whether some genes in one list may never have a chance to be present in the second list, due to the two-gene-universe situation (e.g. (Swaminathan and Fury, 2012)). Indeed, MSigDB v6.2 has not updated all gene names to the current standard, with KANSL1 written as KIAA1267, and KMT2D written as MLL2 in MSigDB. We manually corrected these inconsistent gene names. Similarly, non-coding genes in MSigDB may never have a chance to match our CRS gene list

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

which is all protein-coding. Therefore, only protein-coding genes are kept in MSigDB gene-sets in this enrichment analysis. The C1 MSigDB collection is chromosome location based gene clusters, which is of no interest to us and thus not used here.

The Hallmark, C4, C6, and C7 collection do not have any gene-sets to be significantly overlapping with CRS gene list at the family-wide significance p-value of $= 5.7 \times 10^{-7}$. However, when using the collections-specific multiple testing correction, the TGF-beta signaling (Hallmark) ($p = 3.3 \times 10^{-5}$, OR=15.4) is significant. This pathway has already been discussed as being linked to CRS (Opperman et al., 1997; Hunenko et al., 2001) (see Table 2 for matching genes, and Fig.1(D) for the volcano plot (Li, 2012)).

The C2 gene-set collection in MSigDB includes some other popular pathways and gene-sets databases such as KEGG (Kanehisa and Sato, 2000) and REACTOME (Fabregat et al., 2018). Using a stringent criterion of (Bonferroni correction) p-value $< 0.01/n_{GS}$, which is 5.7 $\times 10^{-7}$ if we use $n_{GS} = 17484$ for all gene-sets in MSigDB (used in the enrichment analysis), or 2.1 $\times 10^{-6}$ if we only count the number of gene-sets in C2 ($n_{GS} = 4762$), 3 or 5 gene-sets are family-wise significant. These top five gene-sets are: cancer pathway (KEGG) (p-value (p) $= 3.1 \times 10^{-9}$, odds-ratio (OR) =8.3), FGF pathway (defunct PID (Schaefer et al., 2009)) ($p = 1.1 \times 10^{-7}$, OR=21.6), MAPK signaling pathway (KEGG) ($p = 1.7 \times 10^{-7}$, OR=7.9), FRS2 mediated cascade (REACTOME) ($p = 2.2 \times 10^{-7}$, OR=28), and signaling by FGFR mutants (REACTOME) ($p = 5.7 \times 10^{-7}$, OR=29.8), These results and matching genes are listed in Table 2, and volcano plot (Li, 2012) in Fig.1(A).

The involvement of FGF (fibroblast growth factor), FGFR (fibroblast growth factor receptor), and MAPK/ERK signaling pathways in CRS has long been discussed in the literature (Marie et al., 2005; Shukla et al., 2007; Kim et al., 2015; Kosty and Vogel, 2015; Pfaff et al., 2016; Timberlake et al., 2017). However, our top hit gene-set is KEGG's cancer pathway, and this link between CRS and cancer is to our knowledge rarely discussed in the literature. In Table 3, we list few more cancer pathways, though not family-wide significant, that are significant at $p = 10^{-5}$ level.

The C3 collection in MSigDB defines gene-sets by their shared DNA sequence motif in the promoter region. With a p-value threshold of 5.7 $\times 10^{-7}$ and 1.2 $\times 10^{-5}$ (0.01/836), 1 or 6 gene-sets are family-wise significant. Three of our six enriched gene-sets (AACTTT,

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

TTGTTT, CTTTGT) appear in the top ten enriched C3 gene-sets in SOX10 regulated genes (supplementary Table 5 of (Sun et al., 2014)). Similar to SOX6 (Hagiwara, 2011), one of the CRS genes, SOX10 also plays an important role in embryonic development (Honoré et al., 2003). These results and matching genes are listed in Table 4, and volcano plot in Fig.1(C).

Although AACTTT/AAAGTT is not labeled as a binding motif for a known transcription factor (TF), IRF1 is listed as the binding TF for this motif in (Xie et al., 2005). Among other binding partners, mir21 (Feng and Tsao, 2016; Javanmardi et al., 2017), mir204 (Li et al., 2016), mir211 Mazar et al., (2016), FOXO4 (Liu et al., 2011; Hornsveld et al., 2018), LEF1 (Hovanes et al., 2001; Li et al., 2009), ETS2 (Sizemore et al., 2017; Fry and Inoue, 2018) are all discussed in the context of a cancer.

The C5 collection in MSigDB is defined by the Gene Ontology (GO) terms (Gene Ontology Consortium, 2019). With the number of all MSigDB gene-sets to correct multiple testing, 111 GO gene-sets are significant; with only the number of GO gene-sets ($n_{GS} = 5917$) to correct, 126 GO gene-sets are significant. We highlight two groups of significant GO gene-sets: the first group has higher percentage of genes in the overlapping set (Table 5, for percentage > 0.1), and the second group contains larger gene-sets (Table 6, for gene-set size > 1000 genes). These two groups as marked in the volcano plot (Li, 2012) can be seen at Fig.1(C). The larger-sized gene-sets are located near the inner side of the volcano as these are "common events" (Li et al., 2014), whereas the stronger-signal gene-sets are located at the outer side because these are relatively rare events.

Most family-wide significant GO gene-sets are expected to be related to CRS as they are involved in morphogenesis, development, differentiation, ossification. Once again, we see some significant gene-sets to be related to cell proliferation. A few significantly CRS-associated gene-sets are related to protein binding, protein transport, signal transduction, cell communication, etc.

Enrichment analysis of craniosynostosis genes by Ingenuity pathway analysis

We ran an enrichment analysis of 113 CRS genes by the Ingenuity Pathway Analysis (IPA) (using the Ingenuity Knowledge Base). The top 16 canonical pathways are shown in Fig.2,

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with the p-value cutoff of 1.32E-5. In Fig.2, signaling pathways are marked by blue, and cancer gene-sets are marked by red, with the rest being well known to be related to development and differentiation. The presence of the "molecular mechanisms of cancer" gene-set in the top list once again confirms the link between CRS and cancer.

It is interesting that some IPA cancer genes are not in KEGG cancer pathway list or cancer driver gene list, which lead to some difference between the overlapping results. More specifically, of the three overlapping lists ($n_1 = 21$ for driver genes, $n_2 = 15$ for KEGG cancer genes, and $n_3 = 13$ for IPA molecular mechanism of cancer genes), only four genes appear in all three lists: BRAF, KRAS, RAF1, and TGFBR2. Besides these four genes, there are also pairwise shared genes: AXIN2, FGFR1, FGFR2, and FGFR3 between the driver and KEGG lists, BMP2, SMO, and TGFBR1 between the driver and IPA lists, and ATR, GNAS, and PTPN11 between the KEGG and IPA lists. The remaining genes are unique to specific lists: ASXL1, BCOR, FLNA, HUWE1, IL6ST, KANSL1, KMT2D, PTPRD, SMCIA, and TCF12 for the driver list, FGF9, GLI3, IGF1R, and STAT3 for the KEGG list, and IHH, LRP5, and SMAD6 for the IPA list.

More detailed information of those IPA signaling pathways in Fig.2 are included in Table 7. Among signaling pathways, IPA shares some common results with KEGG/Reactome/MSigDB, such as TGF and FGF, though there are other significant hits, such as STAT3, NF-KB, PTEN, etc. The decision to put certain genes in a pathway, and the definition of a pathway in light of cross-talk among pathways (Attisano and Labbé, 2004; Guo and Wang, 2009; Mendoza et al., 2011), may all contribute to these discrepancies. It justified our use of multiple databases for this analysis.

Besides the IPA canonical pathways, Table 8 shows top-5 hits in other categories: upstream regulators, diseases, molecular/cellular functions, physiological system development and function. Most of these hits are expected due to our knowledge of CRS being a developmental disorder. We do notice cell-signaling as one of the top hits in molecular/cellular functions.

In order to find out which specific CRS genes participate in which signaling pathway, we checked the 15 CRS genes that overlap with the KEGG cancer pathway. Their presence in other MSigDB C2 signaling pathways and IPA signaling pathways is summarized in Table 9. From Table 9, a few "core" genes appear in multiple IPA signaling pathways (FGFR1, FGFR2,

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

FGFR3, KRAS, RAF1, STAT3, TGFBR1, TGFBR2), which other are more specific for one or few signaling pathways. For example, SMO only appears in cardiac hypertrophy and protein kinase A signaling pathways; GLI3 only appears in protein kinase A signaling pathways; BRAF only appears in NF-KB and protein kinase A signaling pathways; etc.

Enrichment analysis of craniosynostosis genes in transcription factors

Transcription factors (TF) can be the end target of a signaling pathway. Due to the enrichment of signaling pathway genes in CRS gene list, we wonder whether TFs are enriched in the list also. There have been previous attempts to examine the connection between TF and diseases (Latchman, 1996; Li et al., 2015a), and it would not be completely unexpected that the CRS gene list contains a higher proportion of TFs. MSigDB does not have a gene-set with a complete list of transcription factors. The closest ones include the gene-ontology set in C5 "GO_SEQUENCE_SPECIFIC_DNA_BINDING" with around 1000 genes, and "GO_REGULATION_OF_TRANSCRIPTION_FROM_RNA_POLYMERASE_II_PROMOTER" with around 1700 genes.

Instead, we used a recently compiled TF list from (Lambert et al., 2018) with 1638 TFs. There are 19 (out of 113, or 16.8%) CRS genes present in the TF list, compared to the baseline TF frequency of 8.2% (1638/20000). The Fisher test of the enrichment has p-value = 0.003. These overlapping genes are: AHDC1, ALX4, ERF, GLI3, GLIS3, IRX5, LMX1B, MSX2, NFIA, RUNX2, SIX2, SKI, SOX6, STAT3, TCF12, TWIST1, ZEB2, ZIC1, and ZNF462.

There is another TF gene list from ten years ago with 1835 TFs (Vaquerizas et al., 2009). The overlap between the old and new TF lists is 1423 genes (87% of the new list, and 78% of the old list). Some CRS genes that are present in the old TF list, e.g., BMP2, SMAD6, are said to not be DNA binding proteins with specific sequence motifs (Lambert et al., 2018). Though we may trust less the TF list from ten years ago than the recent list, if we use the old TF list for enrichment test, it contains 22 CRS genes (22/113=16.5%), vs. the baseline frequency of 9.2% (1835/20000), with a Fisher test p-value of 0.0008.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Methods and Data

Statistical analysis: The Fisher's test for 2-by-2 table is by the R (*www.r-project.org*) function *fisher.test*. Other data manipulation, text parsing, and data plotting tasks are carried out by our own scripts in perl, python, and R.

Gene intolerance databases: Three databases on whether a gene is tolerant to functional mutation are used: (1) Residual Variation Intolerance Score (RVIS): *genic-intolerance.org*. The RVIS values are based on ExAC (Exome Aggregation Consortium) v2 (release 2.0). (2) Loss-of-function Observed/Expected Upper bound Fraction (LOEUF): *gnomad.broadinstitute.org/downloads*, follow the link of "Constraint", then "pLoF metrics by gene TSV". The gene-level metrics are determined from the gnomAD version 2.1.1 (Karczewski et al., 2019). (3) Gene Damage Index (GDI): *lab.rockefeller.edu/casanova/GDI*. The GDI values are based on the healthy individual's whole genome sequence data from the 1000 Genomes Project database (phase 3) (The 1000 Genomes Project Consortium, 2015), and the impact/damage from a variant is determined by the CADD method (Kircher et al., 2014).

Gene-set database: We downloaded the gene-set data from MSigDB (*software.broadinstitute.org/gsea/msigdb/collections.jsp*), version 6.2 (Liberzon et al., 2015). There are 17810 gene-sets in total, and we excluded the C1 collection which contains chromosome location based gene-sets, leaving 17484 gene-sets. We also filter out non-coding genes, according to the annotation from HGNC (*www.genenames.org*), from the gene-sets. The remaining gene-set collections used are: H (Hallmark collection, n_{GS}=50), C2 (curated gene-sets, including KEGG (*www.genome.jp/kegg*), REACTOME (*reactome.org*), BioCarta, Canonical Pathways, Chemical and Genetic Perturbation, n_{GS}=4762), C3 (motif gene-sets, n=836), C4 (cancer gene neighborhoods (Subramanian et al., 2005) and cancer modules (Segal et al., 2004), n_{GS}=858), C5 (Gene Ontology (*geneontology.org*) gene, n_{GS}=5917), C6 (GEO microarray derived cancer signatures n_{GS}=189), and C7 (immunologic signatures (Godic et al., 2016), n_{GS}=4872).

Cancer driver gene lists: The 2018 list with 299 cancer driver genes is downloaded from Supplement Table S1: *https://www.sciencedirect.com/science/article/pii/S009286741830237X#app2* of the paper (Bailey et al., 2018) (doi: 10.1016/j.cell.2018.02.060). The 2013 list with 125 driver

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

genes is downloaded from the Supplementary Table S2A: *www.sciencemag.org/cgi/content/full/339/6127/1546/DC1* in the paper (Vogelstein et al., 2013) (doi: 10.1126/science.1235122).

Ingenuity pathways: Ingenuity Systems is a commercial package (©2000-2019 QIAGEN) with an expert-curated Knowledge Base, including canonical pathways.

Transcription factor lists: The 2018 list of 2764 transcription factors is downloaded from *humantfs.ccbr.utoronto.ca/*, also as a supplement Table S1 in (Lambert et al., 2018) (doi: 10.1016/j.cell.2018.01.029). The 2009 list of transcription factors is downloaded from the supplementary Table 2 of (Vaquerizas et al., 2009) (doi: 10.1038/nrg2538).

Discussions

The statistical evidence linking craniosynostosis and cancer is robust: We have obtained strong statistical evidence (p-value $\sim 10^{-16}$) that craniosynostosis genes and cancer driver genes sets are more overlapping than by chance. The evidence is even more convincing that when craniosynostosis genes list is split into three pieces according to the time when it was compiled, there is an over-representation for each sub-list. Furthermore, even when we use an older cancer driver gene list with half of the genes as the new list, thus smaller sample sizes, the enrichment test is still significant (p-value $\sim 10^{-7}$). The craniosynostosis-cancer connection is further confirmed in two other databases: the KEGG cancer pathway and the IPA molecular mechanism of cancer gene-set, despite the difference between these cancer gene lists.

Redundancy between gene-sets is not enough to carry over an enrichment between craniosynostosis gene list and one gene-set and that of another: It is known that gene-set databases contain redundant gene-sets (Vivar et al., 2013; Stoney et al., 2018). In order to check whether redundancy is the reason that we obtained multiple enrichment gene-sets in Table 2, we examine the self-enrichment between KEGG_PATHWAYS_IN_CANCER and other C2 and C3 MSigDB gene-sets. Fig.3 shows enrichment Fisher's test ($-\log_{10}$) p-value as a function of the C2/C3 gene-set sizes. The points (C2/C3 gene-sets) at the top of the plot exhibit significant overlap with the KEGG cancer pathway gene-set. The gene-sets

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that significantly overlap with craniosynostosis gene list are marked in red.

Indeed, KEGG_MAPK_SIGNALING_PATHWAY shares 82 out of 266 in common with the KEGG cancer gene-set, REACTOME_SIGNALING_BY_FGFR_IN_DISEASE shares 53 out of 122 in common, followed by SIGNALING_BY_FGFR (Reactome): 31/43, FRS2_MEDIATED_CASCADE (Reactome): 29/36, FGF (PID): 30/55, and SHC_MEDIATED_SIGNALING (Reactome): 10/15. For C3 gene-sets, AACTTT motif defined gene-set shares 83/1820 with the KEGG cancer list, once again, indicating a potential link with the cancer. The C2 cancer gene-sets listed in Table 3 are marked in blue in Fig.3, which are all highly correlated with the KEGG cancer gene-set. However, a high enrichment with the KEGG cancer gene-set does not guarantee an enrichment with the craniosynostosis gene list. For example, the top two points in Fig.3(A) are KEGG small-cell lung cancer gene-set and KEGG prostate cancer gene-set. While the prostate cancer gene-set has some overlap with craniosynostosis gene list (6 out of 89, p-value=3×10^{-5}), the small-cell lung cancer gene-set actually has zero overlap with the craniosynostosis gene list.

Signaling pathways as the main cause for craniosynostosis-cancer connection: Our results in Tables 2,7 and Fig.3(A) all indicate the importance of signaling pathways in craniosynostosis, as well as their common link to both craniosynostosis and cancer. This is consistent with previous findings. FGFR2 is the first gene highlighted to be involved in both craniosynostosis and cancer (Pollock et al., 2007). The FGF signaling pathway has been well established to participate both morphogenesis/development/differentiation (Deimling and Drysdale, 2011; Drewer et al. 2016), on one hand, and carcinogenesis, on the other (Ahmed et al., 2012; Fearon et al., 2013; Gallo et al., 2016). The FGF signaling pathway was proposed as a drug target for treating cancers (Lee and Secord, 2014; Porta et al., 2017; Hanrahan et al., 2019). The WNT signaling pathway is associated with craniosynostosis (Timberlake et al., 2016), and is also associated with cancer (Clevers, 2006; Clevers and Nusse, 2012; Zhan et al., 2017). The Hedgehog signaling pathway is discussed in connection to craniosynostosis (Jenkins et al., 2007), and is also associated with cancer (Jiang and Hui, 2008; Skoda et al., 2018).

In Table 9 we can see that all craniosynostosis-KEGG-cancer overlapping genes are components of some signaling pathways (not mentioned in Table 9, AXIN2 is part of the WNT sig-

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

naling pathway, BMP2, GLI3, and SMO are part of the Hedgehog signaling pathway, IGF1R's binding partner, IGF1, is in the P53, mTOR, and PIP3 signaling pathways). Among 16 more craniosynostosis genes that are not in Table 9 but in the cancer driver list or in the IPA molecular mechanisms of cancer list, FLNA, GNAS, IHH, IL6ST, LRP5, PTPN11, PTPRD, and SMAD6 are also in one of the significantly enriched signaling pathways in Tables 2 and 7. The existence of so many three-way overlapping genes (craniosynostosis, cancer, signaling) is evidence that signaling pathways might be behind the craniosynostosis-cancer connection.

Transcription factors and microRNA associated with craniosynostosis: A biological process can be traced from the chemical signal sent from outside the cell, through the cell membrane, through intra-cellular signaling pathways, into the nucleus, to trigger transcription factors, bind to promoters of certain genes, and initiate transcription. From this description, both transcription factors and their targets can be involved in the etiology of a disease. We have already seen that TFs are enriched in craniosynostosis genes. We can also ask the same question on targets of TFs.

Table 4 shows six TF-motif defined gene-sets are enriched in the craniosynostosis gene list. It hints, though does not prove, that FOXO4, LEF1, ETS2, and possibly IRF1, are relevant to craniosynostosis. A similar point can also be made on micro RNA regulation: mir21 and mir204/mar211. Some craniosynostosis-cancer overlapping genes do not appear in any signal pathways, but they are present in some motif-defined significantly enriched gene-sets, e.g. ASXL1, BCOR, TCF12. These results may guide us to investigate further their role in both craniosynostosis and in cancer.

Interestingly, TF genes are also enriched in cancer driver genes. Using the new TF list from (Lambert et al., 2018), 51 out of 299 cancer driver genes from (Bailey et al., 2018) (17%) are TFs, as compared to the baseline TF frequency of 8%. The Fisher's test on enrichment has p-value = 7.8×10^{-7}. The participation of TF in tumorigenesis is widely studied (Darnell, 2002; Harris, 2002; Zheng and Blobel, 2011; Gupta et al., 2019), including the most famous TF p53 (TP53 gene) (Harris, 1993; Staib et al., 2003; Muller and Vousden, 2013). However, when the three gene lists are compared (TF, cancer driver, CRS), only one gene, TCF12 (Lee et al., 2012; Sharma et al., 2013), appears in all lists.

The oncogenesis recapitulating ontogenesis framework: Julius Cohnheim (1839-

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

1884) was perhaps the first person to propose the embryonic origin of cancer (Capp, 2019), influenced by his teacher Rudolf Virchow (1821-1902) (Ackerknecht, 1958). This parallel between normal developmental process in embryo and cancer in adults can be called "oncogenesis recapitulating ontogenesis" (Huang et al., 2009). Note that there is also a perceived parallel between individual development and species evolution, called "ontogeny recapitulating phylogeny" (Swinnerton, 1938; Richardson and Keuck, 2002).

There are multiple similarities between development and cancer, nicely summarized in (Naxerova et al., 2009). The microscope image of some cancer tissues look like undifferentiated immature cells (Rosai et al., 1985). Cell migration during embryonic development is similar to cancer metastasis (Ridley et al., 2003). More recently, expression profiling (transcriptome) of both certain cancer and developmental tissues exhibit similar signatures (Borczuk et al., 2003; Kho et al., 2004; Coulouarn et al., 2005; Hu and Shivdasani, 2005; Liu, 2006; Dekel et al. 2006; Kaiser et al., 2007; Naxerova et al., 2009; Li et al., 2010; Palmer et al., 2012; Soundararajan et al., 2015). There are also many theoretical hypotheses and frameworks building around the idea (Garraway and Sellers, 2006; Ao et al., 2008; Huang et al., 2009; Huang, 2011; Yuan et al., 2017; Zhou et al., 2018; Liu, 2018, 2019) The results in our paper add another path linking development and cancer.

In conclusion, we provide solid evidence that a higher proportion of genes associated with craniosynostosis are also linked to cancer. This observation is consistent with the perspective that adult carcinogenesis (oncogenesis, tumorigenesis) and embryonic ontogenesis (morphogenesis) share certain biological process in common, such as signaling pathways. Transcription factors, as end points of signaling pathways, also play prominent roles in both craniosynostosis and cancers.

Acknowledgement

We would like to thank Marc Symons, Patricia Mongini, Kim Simpfendo, Jan Freudenberg and Ping Ao for discussion and suggestions. SM thanks Dr. Serena McCalla for support and guidance throughout the project, and thanks the financial support from the FIMR Summer Intern Program. WT and AS thank the support from the Robert S Boas Center for Genomics

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and Human Genetics.

Appendix: More information on the 17 craniosynostosis genes

The following table summarizes these information on the extra 17 craniosynostosis genes: (1) gene name and the corresponding OMIM number; (2) chromosome band location (the genes are ordered according to this column); (3) information about mutation, such as whether it's de novo, whether it's non-synonymous (missense) single-nucleotide polymorphism (nsSNP), or copy number variation (CNV), whether it's autosomal dominant (AD) or autosomal recessive (AR) (question mark indicates the information is obtained from OMIM). There are two complex inheritance detection methods using many patients: one is TDT (transmission disequilibrium test) applied to many family trios (for BBS9), another is burden test of rare variance (SPRY1, SPRY4, SMURF1) – these methods are very different from the Mendelian gene detection methods. (4) information about phenotype, such as syndrome name (with OMIM number) for syndromic CRS, and suture location for non-syndromic CRS; (5) number of patients with the mutation described. For the two complex inheritance detection methods, the number of patients is indicated by a non-specific n; (6) the publication in which the gene mutation is studied.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Gene (OMIM)	Cband	Mutation & inheritance pattern	Clinical disorder& suture (OMIM)	NumS	Reference
LMNA (150330)	1q22	de novo. CNV (duplication)	metopic & unilateral lambdoid	1	(Sowińska-Seidler et al., 2018)
BGLAP (112260)	1q22	de novo. CNV (duplication)	metopic & unilateral lambdoid	3	(Sowińska-Seidler et al., 2018)
SIX2 (604994)	2p21	CNV (deletion)	Fronyonasal dysplasia syndrome (136760). sagittal	1	(Hufnagel et al., 2016)
PTH2R (601469)	2q34	inversion	sagittal & metopic	1	(Kim et al., 2015)
RAF1 (164760)	3p25.2	nsSNP	Noonan syndrome with multiple lentigines (151100)	1	(Rodríguez et al., 2019)
FGFRL1 (605830)	4p16.3	AR? frameshift	Antley-Bixler syndrome (207410)	1	(Rieckmann et al., 2009)
PPP3CA (114105)	4q24	de novo nsSNP	metopic (trigonocephaly) plagiocephaly	2	(Mizuguchi et al., 2018)
SPRY1 (602465)	4q28.1	de novo. nsSNP burden	sagittal	5 n	(Timberlake et al., 2016) (Timberlake et al., 2017)
SPRY4 (607984)	5q31.3	de novo. nsSNP burden	sagittal	1 n	(Timberlake et al., 2016) (Timberlake et al., 2017)
TCOF1 (606847)	5q32-q33.1	AD? de novo. stop-gain	Treacher Collins syndrome (154500)	1	(Horiuchi et al., 2004)
BBS9 (615986)	7p14.3	AD. nsSNP GWAS TDT.	coronal, sagittal	1 n	(Sewda et al., 2019) (Justice et al., 2012)
GCK (138079)	7p13	AD? CNV (deletion)	Greig Cephalopolysundactyly syndrome (175700)	1	(Zung et al., 2011)
SMURF1 (605568)	7q22.1	de novo. nsSNP burden	metopic	1 n	(Timberlake et al., 2016) (Timberlake et al., 2017)
TNFRSF11B (602643)	8q24.12		Juvenile Paget's disease (239000) with CRS	1	(Saki et al., 2013)
COLEC10 (607620)	8q24.12	AR. nsSNP	3MC syndrome (248340)	3	(Munye et al., 2017)
AXIN2 (604025)	17q24.1	de novo. frameshift	sagittal	1	(Yilmaz et al., 2018)
BCOR (300485)	Xp11.4	de novo. nsSNP	Oculofaciocardiodental syndrome (300166)	1	(O'Byrne et al., 2017)

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

The 1000 Genomes Project Consortium (2015), A global reference for human genetic variation, *Nature*, 526:68-74.

EH Ackerknecht (1958), Historical notes on cancer, *Med. Hist.*, 2:114-119.

I Ahmad, T Iwata, HY Leung (2012), Mechanisms of FGFR-mediated carcinogenesis, *Biochim. Biophys. Acta, Mol. Cell. Res.*, 1823:850-860.

Z Al-Rekabi, MM Wheeler, A Leonard, AM Fura, I Juhlin, C Frazar, JD Smith, SS Park, JA Gustafson, CM Clarke, ML Cunningham, NJ Sniadecki (2016), Activation of the IGF1 pathway mediates changes in cellular contractility and motility in single-suture craniosynostosis, *J. Cell Sci.*, 129:483-491.

P Ao, D Galas, L Hood, X Zhu (2008), Cancer as robust intrinsic state of endogenous molecular-cellular network shaped by evolution, *Med. Hypotheses*, 70:678-684.

L Attisano and E Labbé (2004), TFGβ and Wnt pathway cross-talk, *Cancer and Metastasis Rev.*, 23:53-61.

MH Bailey, C Tokheim, E Porta-Pardo, S Sengupta, D Bertrand, A Weerasinghe, A Colaprico, MC Wendl, J Kim, B Reardon, et al., (2018), Comprehensive characterization of cancer driver genes and mutations, *Cell*, 173:371-385.

AC Borczuk, L Gorenstein, KL Walter, AA Assaad, L Wang, CA Powell (2003), Non-small-cell lung cancer molecular signatures recapitulate lung developmental pathways, *Am. J. Pathol.* 163:1949-1960.

JR Brewer, P Mazot, P Soriano (2016), Genetic insights into the mechanisms of Fgf signaling, *Gene & Devel.*, 30:751-771.

JP Capp (2019), Cancer stem cells: from historical roots to a new perspective, sIJ. Oncology, 2019:5189232.

F Carinci, F Pezzetti, P Locci, E Becchetti, F Carls, A Avantaggiato, A Becchetti, P Carinci, T Baroni, M Bodo (2005), Apert and Crouzon syndromes: clinical findings, genes and extracellular matrix, *J. Craniofacial Surgery*, 16:361-368.

H Clevers (2006), Wnt/beta-catenin signaling in development and disease, *Cell*, 127:469-480.

H Clevers and R Nusse (2012), Wnt/beta-catenin signaling and disease, *Cell*, 149:1192-1205.

C Coulouarn, C Derambure, G Lefebvre, R Daveau, M Hiron, M Scotte, A Francois, M Daveau, JP Salier (2005), Global gene repression in hepatocellular carcinoma and fetal liver, and suppression of dudulin-2 mRNA as a possible marker for the cirrhosis-to-tumor transition, *J. Hepatol.*, 42:860-869.

JS Darnell Jr. (2002), Transcription factors as targets for cancer therapy, *Nature Rev. Cancer*, 2:740-749.

SJ Deimling and TA Drysdale (2011), Fgf is required to regulate anteriorposterior patterning in the Xenopus lateral plate mesoderm, *Mech. of Devel.*, 128:7-10.

B Dekel, S Metsuyanim, KM Schmidt-Ott, E Fridman, J Jacob-Hirsch, A Simon, J Pinthus, Y Mor, J Barasch, N Amariglio, Y Reisner, N Kaminski, G Rechavi (2006), Multiple imprinted and stemness genes provide a link between normal and tumor progenitor cells of the developing human kidney, *Cancer Res.*, 66:6040-6049.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C Dufour, H Guenou, K Kaabeche, D Bouvard, A Sanjay, PJ Marie (2008), FGFR2-Cbl interaction in lipid rafts triggers attenuation of PI3K/Akt signaling and osteoblast survival, *Bone*, 42:1032-1039.

EL Durham, RN Howie, JJ Cray (2017), Gene/environment interactions in craniosynostosis: a brief review, *Orthod Craniofac Res.*, 20(Suppl 1):811.

A Fabregat, S Jupe, L Matthews, K Sidiropoulos, M Gillespie, P Garapati, R Haw, B Jassal, F Korninger, B May, M Milacic, CD Roca, K Rothfels, C Sevilla, V Shamovsky, S Shorser, T Varusai, G Viteri, J Weiser, G Wu, L Stein, H Hermjakob, P D'Eustachio (2018), The Reactome Pathway Knowledgebase, *Nucleic Acids Res.*, 46(D1):D649-D655.

AE Fearon, CR Gould, RP Grose (2013), FGFR signalling in women's cancers, *Int. J. Biochm. & Cell Biol.*, 45:2832-2842.

YH Feng and CJ Tsao (2016), Emerging role of microRNA-21 in cancer, *Biomed. Rep.*, 5:395-402.

EA Fry and K Inoue (2018), Aberrant expression of ETS1 and ETS2 proteins in cancer, *Cancer Rep. Rev.*, 2:1-10.

LH Gallo, KN Nelson, AN Meyer, DJ Donoghue (2015), Functions of Fibroblast Growth Factor Receptors in cancer defined by novel translocations and mutations, *Cytokine & Growth Factor Rev.*, 26:425-449.

LA Garraway and WR Sellers (2006), Lineage dependency and lineage-survival oncogenes in human cancer, *Nature Rev. Cancer*, 6:593-602.

The Gene Ontology Consortium (2019), The Gene Ontology Resource: 20 years and still GOing strong, *Nucleic Acids Res.*, 47(D1):D330-D338.

J Godec, Y Tan, A Liberzon, P Tamayo, S Bhattacharya, A Butte, JP Mesirov, WN Haining (2016), Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation, *Immunity*, 44:194-206.

JA Goos and IMJ Mathijssen (2019), Genetic causes of craniosynostosis: an update, *Mol. Syndromol.*, 10:6-23.

X Guo and XF Wang (2009), Signaling cross-talk between TGF-β/BMP and other pathways, *Cell Res.*, 19:71-88.

R Gupta, W Li, XJ Yan, J Barrientos, JE Kolitz, SL Allen, K Rai, N Chiorazzi, PKA Mongini (2019), Mechanism for IL-15driven B cell chronic lymphocytic leukemia cycling: roles for AKT and STAT5 in modulating cyclin D2 and DNA damage response proteins, *J. Immunology*, 202:2924-2944.

N Hagiwara (2011), Sox6, jack of all trades: A versatile regulatory protein in vertebrate development, *Devel. Dyn.*, 240:1311-1321.

AJ Hanrahan, G Iyer, DB Solit (2019), Intracellular signaling, in *Abeloff's Clinical Oncology*, 6th edition, ed. JE Niederhuber, JO Armitage, JH Doroshow, MB Kastan, JE Tepper (Elsevier), chapter 2, pp.24-46.

AL Harris (2002), Hypoxia a key regulatory factor in tumour growth, *Nature Rev. Cancer*, 2:38-47.

CC Harris (1993), p53: at the crossroads of molecular carcinogenesis and risk assessment, *Science*, 262:1980-1981.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

SM Honoré, MJ Aybar, R Mayor (2003), Sox10 is required for the early development of the prospective neural crest in Xenopus embryos, *Devel. Biol.*, 260:79-96.

K Horiuchi, T Ariga, H Fujioka, K Kawashima, Y Yamamoto, H Igawa, Y Sakiyama, T Sugihara (2004), Treacher Collins syndrome with craniosynostosis, choanal atresia, and esophageal regurgitation caused by a novel nonsense mutation in TCOF1, *Am. J. Med. Genet. A*, 128:173-175.

M Hornsveld, TB Dansen, PW Derksen, BM Burgering (2018), Re-evaluating the role of FOXOs in cancer, —sl Seminars in Cancer Biol., 50:90-100.

K Hovanes, TW Li, JE Munguia, T Truong, T Milovanovic, J Lawrence Marsh, RF Holcombe, ML Waterman (2001), Beta-catenin-sensitive isoforms of lymphoid enhancer factor-1 are selectively expressed in colon cancer. *Nature Genet.*, 28:5357.

M Hu and RA Shivdasani (2005), Overlapping gene expression in fetal mouse intestine development and human colorectal cancer, *Cancer Res.*, 65:8715-8722.

S Huang (2011), On the intrinsic inevitability of cancer: From foetal to fatal attraction, *Semin. Cancer Biol.*, 21:183-199.

S Huang, I Ernberg, S Kauffman (2009), Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective, *Semin. Cell and Dev. Biol.*, 20:869-876.

RB Hufnagel, SL Zimmerman, LA Krueger, PL Bender, ZM Ahmed, HM Saal (2016), A new frontonasal dysplasia syndrome associated with deletion of the SIX2 gene, *Am. J. Med. Genet. A*, 170:487-491.

O Hunenko, J Karmacharya, G Ong, RE Kirschner (2001), Toward an understanding of nonsyndromic craniosynostosis: altered patterns of TGF-beta receptor and FGF receptor expression induced by intrauterine head constraint, *Ann. Plastic Surg.*, 46:546-554.

AG Hunter and NL Rudd (1976), Craniosynostosis. I. Sagittal synostosis: its genetics and associated clinical findings in 214 patients who lacked involvement of the coronal suture(s), *Teratology*, 14:185-193.

Y Itan, L Shang, B Boisson, E Patin, A Bolze, M Moncada-Vélez, E Scott, MJ Ciancanelli, FG Lafaille, JG Markle, R Martinez-Barricarte, SJ de Jong, XF Kong, P Nitschke, A Belkadi, J Bustamante, A Puel, S Boisson-Dupuis, PD Stenson, JG Gleeson, DN Cooper, L Quintana-Murci, JM Claverie, SY Zhang, L Abel, JL Casanova (2015), The human gene damage index as a gene-level approach to prioritizing exome variants, *Proc. Natl. Acad. Sci.*, 112:13615-13620.

S Javanmardi, MR Aghamaali, SS Abolmaali, S Mohammadi, AM Tamaddon (2017), miR-21, an oncogenic target miRNA for cancer therapy: molecular mechanisms and recent advancements in chemo and radio-resistance, *Curr. Gene Ther.*, 16:375-389.

D Jenkins, D Seelow, FS Jehee, CA Perlyn, LG Alonso, DF Bueno, D Donnai, D Josifova, IM Mathijssen, JE Morton, KH Orstavik, E Sweeney, SA Wall, JL Marsh, P Nurnberg, MR Passos-Bueno, AO Wilkie (2007), RAB23 mutations

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in Carpenter syndrome imply an unexpected role for hedgehog signaling in cranial-suture development and obesity, *Am. J. Hum. Genet.*, 80:1162-1170.

J Jiang and CC Hui (2008), Hedgehog signaling in development and cancer, *Devel. Cancer*, 15:801-812.

G Jürgens (1985), A group of genes controlling the spatial expression of the bithorax complex in Drosophila, *Nature*, 316:153-155.

CM Justice, G Yagnik, Y Kim, I Peter, EW Jabs, M Erazo, X Ye, E Ainehsazan, L Shi, ML Cunningham, et al. (2012), A genome-wide association study identifies susceptibility loci for nonsyndromic sagittal craniosynostosis near BMP2 and within BBS9, *Nature Genet.*, 44:1360-1364.

H Kabbani and TS Raghuveer (2004), Craniosynostosis, *Am. Fam. Physician.*, 69:2863-2870.

S Kaiser, YK Park, JL Franklin, RB Halberg, M Yu, WJ Jessen, J Freudenberg, X Chen, K Haigis, AG Jegga, et al. (2007), Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer, *Genome Biol.*, 8:R131.

M Kanehisa and Y Sato (2000), KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res.*, 28:27-30.

KJ Karczewski, LC Francioli, G Tiao, BB Cummings, J Alföldi, Q Wang, RL Collins, KM Laricchia, A Ganna, DP Birnbaum, et al. (2019), Variation across 141,456 human exomes and genomes reveals the spectrum of loss-offunction intolerance across human protein-coding genes, *bioRxiv* preprint, doi: 10.1101/531210.

P Khatri, M Sirota, AJ Butte (2012), Ten years of pathway analysis: current approaches and outstanding challenges, *PLoS Comp. Biol.*, 8:e1002375.

AT Kho, Q Zhao, Z Cai, AJ Butte, JY Kim, SL Pomeroy, DH Rowitch, IS Kohane (2004), Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers, *Genes & Devel.*, 18:629-640.

J Kim, HH Won, Y Kim, JR Choi, N Yu, KA Lee (2015), Breakpoint mapping by whole genome sequencing identifies PTH2R gene disruption in a patient with midline craniosynostosis and a de novo balanced chromosomal rearrangement, *J. Med. Genet.*, 52:706-709.

SD Kim, G Yagnik , ML Cunningham, J Kim, SA Boyadjiev (2014), MAPK/ERK signaling pathway analysis in primary osteoblasts from patients with nonsyndromic sagittal craniosynostosis, *Cleft Palate-Craniofacial J.*, 51:115-119.

M Kircher, DM Witten, P Jain, BJ O'Roak, GM Cooper, J Shendure (2014), A general framework for estimating the relative pathogenicity of human genetic variants, *Nat Genet.*, 46:310-315.

JM Ko (2016), Genetic syndromes associated with craniosynostosis, *J. Korean Neurosug. Soc.*, 59:187-191.

J Kosty and TW Vogel (2015), Insights into the development of molecular therapies for craniosynostosis, *Neurosurgical Focus*, 38:E2.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

SA Lambert, A Jolma, LF Campitelli, PK Das, Y Yin, M Albu, X Chen, J Taipale, TR Hughes, MT Weirauch (2018), The human transcription factors, *Cell*, 172:650-665.

DS Latchman (1996), Transcription-factor mutations and disease, *New Eng. J. Med.*, 334:28-33.

CC Lee, WS Chen, CC Chen, LL Chen, YS Lin, CS Fan, TS Huang (2012), TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer, *J. Biol. Chem.*, 287:2798-2809.

PS Lee and AA Secord (2014), Targeting molecular pathways in endometrial cancer: A focus on the FGFR pathway, *Cancer Treament Rev.*, 40:507-512.

C Li, DA Scott, E Hatch, X Tian, SL Mansour (2007), Dusp6 (Mkp3) is a negative feedback regulator of FGF-stimulated ERK signaling during mouse development, *Development*, 134:167-176.

T Li, H Pan, R Li (2016), The dual regulatory role of miR-204 in cancer, *Tumour Biol.*, 37:11667-11677.

T Li, B Wan, J Huang, X Zhang (2010), Comparison of gene expression in hepatocellular carcinoma, liver development, and liver regeneration, *Mol. Genet. and Genomics*, 283:485-492.

W Li (2012), Volcano plots in analyzing differential expressions with mRNA microarrays, *J. Bioinf and Comp. Biol.*, 10:1231003.

W Li, J Espinal-Enríquez, KR Simpfendorfer, E Hernaández-Lemus (2015), A survey of disease connections for CD4+ T cell master genes and their directly linked genes, *Comp. Biol. and Chem.*, 59(B):78-90.

W Li, J Freudenberg, M Oswald (2015), Principles for the organization of gene-sets, *Comp. Biol. Chem.*, 59(B):139-149.

W Li, J Freudenberg, YJ Suh, Y Yang (2014), Using volcano plots and regularized-chi statistics in genetic association studies, *Comp. Biol. and Chem.*, 48:77-83.

Y Li, L Wang, M Zhang, J Melamed, X Liu, R Reiter, J Wei, Y Peng, X Zou, A Pellicer, MJ Garabedian, A Ferrari, P Lee (2009), LEF1 in androgen-independent prostate cancer: regulation of androgen receptor expression, prostate cancer growth, and invasion, *Cancer Res.*, 69:3332-3338.

A Liberzon, C Birger, H Thorvaldsdttir, M Ghandi, JP Mesirov, P Tamayo (2015), The Molecular Signatures Database (MSigDB) hallmark gene set collection, *Cell Syst.*, 1:417-425.

BK Lin, M Clyne, M Walsh, O Gomez, W Yu, M Gwinn, MJ Khoury (2006), Tracking the epidemiology of human genes in the literature: the HuGE published literature database, *Am. J. Epidemology*,164:1-4.

H Liu, AT Kho, IS Kohane, Y Sun Y (2006), Predicting survival within the lung cancer histopathological hierarchy using a multi-scale genomic model of development, *PLoS Med.*, 3:e232.

J Liu (2018), The dualistic origin of human tumors, *Seminars in Cancer Biol.*, 53:1-16.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

J Liu (2019), The "life code: A theory that unifies the human life cycle and the origin of human tumors, *Seminars in Cancer Biol.*, in press. doi: 10.1016/j.semcancer.2019.09.005

X Liu, Z Zhang, L Sun, N Chai, S Tang, J Jin, H Hu, Y Nie, X Wang, K Wu, H Jin, D Fan (2011), microRNA-499-5p promotes cellular invasion and tumor metastasis in colorectal cancer by targeting FOXO4 and PDCD4, *Carcinogenesis*, 32:1798-1805.

Y Liu, Y Liang, D Wishart (2015), PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more, *Nucleic Acids Res.*, 43(W1):W535-W542.

J Lokau and C Garbers (2019), Activating mutations of the gp130/JAK/STAT pathway in human diseases, in *Advances in Protein Chemistry and Structural Biology: Intracellular Signalling Proteins*, vol 116, eds. R Donev (Academic Press), chapter 7, pp.283-310.

SN Magge, K Snyder, ABS Sajja, TA DeFreitas, SE Hofherr, RE Broth, RF Keating, GF Rogers (2017), Identical twins discordant for metopic craniosynostosis, *J. Craniofacial Surgery*, 28:14-16.

AH Mahmoud Adel, A Al Jabri Adbullah, F Eissa (2013), Infantile osteopetrosis, craniosynostosis, and Chiari malformation type I with novel OSTEM1 mutation, *J. Pediatr. Neurosci.*, 8:34-37.

PJ Marie, JD Coffin, MM Hurley (2005), FGF and FGFR signaling in chondrodysplasias and craniosynostosis, *J. Cell Biochem.*, 96:888-896.

J Mazar, F Qi, B Lee, J Marchica, S Govindarajan, J Shelley, JL Li, A Ray, RJ Perera (2016), MicroRNA 211 functions as a metabolic switch in human melanoma cells, *Mol. and Cell. Biol.*,36:10901108.

J McClellan and MC King (2010), Genetic heterogeneity in human disease, *Cell*, 141:210-217.

MC Mendaza, E Emrah Jr., J Blenis (2011), The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation, *Trends Biochem. Sci.*, 36:320-328.

T Mizuguchi, M Nakashima, M Kato, N Okamoto, H Kurahashi, N Ekhilevitch, M Shiina, G Nishimura, T Shibata, M Matsuo, et al. (2018), Loss-of-function and gain-of-function mutations in PPP3CA cause two distinct disorders, *Hum. Mol. Genet.*, 27:1421-1433.

A Moenning, R Jäger, A Egert, W Kress, E Wardelmann, H Schorle (2008), Sustained platelet-derived growth factor receptor α signaling in osteoblasts results in craniosynostosis by overactivating the phospholipase C-γ pathway *Mol. and Cell. Biol*, 29:881-891.

M Muenke and AO Wilkie (2014), Craniosynostosis syndromes, in *The Metabolic and Molecular Bases of Inherited Disease*, eds. D Valle, AL Beaudet, B Vogelstein, KW Kinzler, SE Antonarakis, A Ballabio, K Gibson, G Mitchell (McGraw-Hill, New York, NY). *http://ommbid.mhmedical.com/content.aspx?bookid=971&sectionid=62663640*

PAJ Muller and KH Vousden (2013), p53 mutations in cancer, *Nature Cell Biol.*, 15:2-8.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

MM Munye, A Diaz-Font, L Ocaka, ML Henriksen, M Lees, A Brady, D Jenkins, J Morton, SW Hansen, C Bacchelli, PL Beales, V Hernandez-Hernandez (2017), COLEC10 is mutated in 3MC patients and regulates early craniofacial development, *PLoS Genet.*, 13:e1006679.

K Naxerova, CJ Bult, A Peaston, K Fancher, BB Knowles, S Kasif, IS Kohane (2009), Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers, *Genome Biol.*, 9:R108.

A Niroula and M Vihinen (2019), How good are pathogenicity predictors in detecting benign variants? *PLoS Comp. Biol.*, 15:e1006481.

JJ O'Byrne, E Laffan, DJ Murray, W Reardon (2017), Oculofaciocardiodental syndrome with craniosynostosis, temporal hypertrichosis, and deafness, *Am. J. Med. Genet. A*, 173:1374-1377.

LA Opperman, AA Nolen, RC Ogle (1997), TGFβ1, TGFβ2, and TGFβ3 exhibit distinct patterns of expression during cranial suture formation and obliteration in vivo and in vitro, *J. Bone and Mineral Res.*, 12:301-310.

KF Oram and T Gridley (2005), Mutations in snail family genes enhance craniosynostosis of Twist1 haplo-insufficient mice. Implications for Saethre-Chotzen Syndrome, *Genetics*, 170:971-974.

DM Ornitz and PJ Marie (2002), FGF signaling pathways in endochondral and intramembranous bone development and human genetic disease, *Genes & Development*, 16:1446-1465.

NP Palmer, PR Schmid, B Berger, IS Kohane (2012), A gene expression profile of stem cell pluripotentiality and differentiation is conserved across diverse solid and hematopoietic cancers, *Genome Biol.*, 13:R71.

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

MJ Pfaff, K Xue, L Li, MC Horowitz, DM Steinbacher, JVP Eswarakumar (2016), FGFR2c-mediated ERKMAPK activity regulates coronal suture development, *Devel. Biol.*, 415:242-250.

PM Pollock, MG Gartside, LC Dejeza, MA Powell, MA Mallon, Cancer Genome Project, H Davies, M Mohammadi, PA Futreal, MR Stratton, JM Trent, PJ Goodfellow (2007), Frequent activating FGFR2 mutations in endometrial carcinomas parallel germline mutations associated with craniosynostosis and skeletal dysplasia syndromes, *Oncogene*, 26:7158-7162.

R Porta, R Borea, A Coelho, S Khan, A Ara´jo, P Reclusa, T Franchina, N Van Der Steen, P Van Damb, J Ferri, R Sirera, A Naing, D Hong, C Rolfo (2017), FGFR a promising druggable target in cancer: Molecular biology and new drugs, *Critical Rev. Oncology/Hematology*, 113:256-267.

MK Richardson and G Keuck (2002), Haeckel's ABC of evolution and development, *Biol. Rev.*, 77:495-528.

AJ Ridley, MA Schwartz, K Burridge, RA Firtel, MH Ginsberg, G Borisy, JT Parsons, AR Horwitz (2003), Cell migration: integrating signals from front to back, *Science*, 302:1704-1709.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

T Rieckmann, L Zhuang, CE Flück, B Trueb (2009), Characterization of the first FGFRL1 mutation identified in a craniosynostosis patient, *Bioch. Biophys. Acta (BBA) – Mol. Basis Dis.*, 1792:112-121.

F Rodríguez, D Ponce, FJ Berward, B Lopetegui, F Cassorla, M Aracena (2019), RAF1 variant in a patient with Noonan syndrome with multiple lentigines and craniosynostosis, *Am. J. Med. Genet. A*, 179:1598-1602.

J Rosai, EA Saxén, L Woolner (1985), Undifferentiated and poorly differentiated carcinoma, *Semin. Diagn. Pathol.*, 2:123-136.

F Saki, Z Karamizadeh, S Nasirabadi, S Mumm, WH McAlister, MP Whyte (2013), Juvenile paget's disease in an Iranian kindred with vitamin D deficiency and novel homozygous TNFRSF11B mutation, *J. Bone Miner Res.*, 28:1501-1508.

CF Schaefer, K Anthony, S Krupa, J Buchoff, M Day, T Hannay, KH Buetow (2009), PID: the Pathway Interaction Database, *Nucleic Acids Res.*, 37(Database issue):D674-D679.

JJ Schoenebeck, SA Hutchinson, A Byers, HC Beale, B Carrington, DL Faden, M Rimbault, B Decker, JM Kidd, R Sood, AR Boyko, JW Fondon 3rd, RK Wayne, CD Bustamante, B Ciruna, EA Ostrander (2012), Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet.*, 8:e1002849.

E Segal, N Friedman, D Koller, A Regev (2004), A module map showing conditional activity of expression modules in cancer, *Nature Genet.*, 36:1090-1098.

A Sewda, SR White, M Erazo, K Hao, G Garca-Fructuoso, I Fernndez-Rodriguez, Y Heuzé, JT Richtsmeier, PA Romitti, B Reva, EW Jabs, I Peter (2019), Nonsyndromic craniosynostosis: novel coding variants, *Ped. Res.*, 85:463-468.

VP Sharma, AL Fenwick, MS Brockop, SJ McGowan, JAC Goos, AJM Hoogeboom, AF Brady, NO Jeelani, SA Lynch, JB Mulliken, DJ Murray, JM Phipps, E Sweeney, SE Tomkins, LC Wilson, S Bennett, RJ Cornall, J Broxholme, A Kanapin, 500 Whole-Genome Sequences (WGS500) Consortium, D Johnson, SA Wall, PJ van der Spek, IMJ Mathijssen, RE Maxson, SRF Twigg, AOM Wilkie (2013), Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal, *Nature Genet.*, 45:304-307.

V Shashi and TC Hart (2002), Environmental Etiologies of Orofacial Clefting and Craniosynostosis, in *Understanding Craniofacial Anomalies*, eds. MP Mooney, MI Siegel, chapter 8 (Wiley).

V Shukla, X Coumoul, RH Wang, HS Kim, CX Deng (2007), RNA interference and inhibition of MEK-ERK signaling prevent abnormal skeletal phenotypes in a mouse model of craniosynostosis, *Nature Genet.*, 39:1145-1150.

GM Sizemore, JR Pitarresi, S Balakrishnan, MC Ostrowski (2017), The ETS family of oncogenic transcription factors in solid tumours, *Nature Rev. Cancer*, 17:337-351.

AM Skoda, D Simovic, V Karin, V Kardum, S Vranic, L Serman (2018), The role of the Hedgehog signaling pathway in cancer: a comprehensive review, *Bosn. J. Basic Med. Sci.*, 18:8-20.

R Soundararajan, AN Paranjape, V Barsan, JT Chang, SA Mani (2015), A novel embryonic plasticity gene signature that predicts metastatic competence and clinical outcome, *Sci. Rep.*, 5:11766.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A Sowińska-Seidler, EM Olech, M Socha, D Larysz, A Jamsheer (2018), Novel 1q22-q23.1 duplication in a patient with lambdoid and metopic craniosynostosis, muscular hypotonia, and psychomotor retardation, *J. Appl. Genet.*, 59:281-289.

F Staib, SP Hussain, LJ Hofseth, XW Wang, CC Harris (2003), TP53 and liver carcinogenesis, *Human Mut.*, 21:201-216.

BD Stamper, SS Park, RP Beyer, TK Bammler, FM Farin, B Mecham, ML Cunningham (2011), Differential expression of extracellular matrix-mediated pathways in single-suture craniosynostosis, *PLoS ONE*, 6:e26557.

RA Stoney, J-M Schwartz, DL Robertson, G Nenadic (2018), Using set theory to reduce redundancy in pathway sets, *BMC Bioinf.*, 19:386.

A Subramanian, P Tamayo, VK Mootha, S Mukherjee, BL Ebert, MA Gillette, A Paulovich, SL Pomeroy, TR Golub, ES Lander, JP Mesirov (2005), Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci.*, 102:15545-15550.

C Sun, L Wang, S Huang, GJ Heynen, A Prahallad, C Robert, J Haanen, C Blank, J Wesseling, SM Willems, D Zecchin, S Hobor, PK Bajpe, C Lieftink, C Mateus, S Vagner, W Grernrum, I Hofland, A Schlicker, LF Wessels, RL Beijersbergen, A Bardelli, F Di Nicolantonio, AM Eggermont, R Bernards (2014), Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma, *Nature*, 508:118-122.

K Swaminathan and W Fury (2012). Non-hypergeometric overlap probability, U.S. Patent No. US 8,255,167. Washington, DC: U.S. Patent and Trademark Office.

HH Swinnerton (1938), Development and evolution, *Nature*, 142:459-462.

T Takenouchi, Y Sakamoto, T Miwa, C Torii, R Kosaki, K Kishi, T Takahashi, K Kosaki (2014), Severe craniosynostosis with Noonan syndrome phenotype associated with SHOC2 mutation: clinical evidence of crosslink between FGFR and RAS signaling pathways, *Am. J. Med. Genet. A*, 164:2869-2872.

AT Timberlake, J Choi, S Zaidi, Q Lu, C Nelson-Williams, ED Brooks, K Bilguvar, I Tikhonova, S Mane, JF Yang, R Sawh-Martinez, S Persing, EG Zellner, E Loring, C Chuang, A Galm, PW Hashim, DM Steinbacher, ML DiLuna, CC Duncan, KA Pelphrey, H Zhao, JA Persing, RP Lifton (2016), Two locus inheritance of non-syndromic midline craniosynostosis via rare SMAD6 and common BMP2 alleles, *eLIFE*, 5:e20125.

AT Timberlake, CG Furey, J Choi, C Nelson-Williams, Yale Center for Genome Analysis, E Loring, A Galm, KT Kahle, DM Steinbacher, D Larysz, JA Persing, RP Lifton (2017), De novo mutations in inhibitors of Wnt, BMP, and Ras/ERK signaling pathways in non-syndromic midline craniosynostosis, *Proc. Natl. Acad. Sci.*, 114:E7341-E7347.

CJ Tokheim, N Papadopoulos, KW Kinzler, B Vogelstein, R Karchin (2016), Evaluating the evaluation of cancer driver genes, *Proc. Natl. Acad. Sci.*, 113:14330-14335.

SR Twigg and AOM Wilkie (2015), A genetic-pathophysiological framework for craniosynostosis, *Am. J. Hum. Genet.*, 97:359-377.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

JM Vaquerizas, SK Kummerfeld, SA Teichmann, NM Luscombe (2009), A census of human transcription factors: function, expression and evolution, *Nature Rev. Genet.*, 10:252-263.

JC Vivar, P Pemu, R McPherson, S Ghosh (2013), Redundancy control in pathway databases (ReCiPa): an application for improving gene-set enrichment analysis in omics studies and "big data" biology, *Omics*, 17:414-422.

B Vogelstein, N Papadopoulos, VE Velculescu, S Zhou, LA Diaz Jr, KW Kinzler (2013), Cancer genome landscapes, *Science*, 339:1546-1558.

AOM Wilkie (1997), Craniosynostosis: genes and mechanisms, *Hum. Mol. Genet.*, 6:1647-1656.

AOM Wilkie, SF Slaney, M Oldridge, MD Poole, GJ Ashworth, AD Hockley, RD Hayward, DJ David, LJ Pulleyn, P Rutland, S Malcolm, R M Winter, W Reardon (1995), Apert syndrome results from localized mutations of FGFR2 and is allelic with Crouzon syndrome, *Nature Genet.*, 9:165-172.

X Xie, J Lu, EJ Kulbokas, TR Golub, V Mootha, K Lindblad-Toh, ES Lander, M Kellis (2005), Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals, *Nature*, 434:338-345.

E Yilmaz, E Mihci, B Guzel Nur, OM Alper (2018), A novel AXIN2 gene mutation in sagittal synostosis, *Am. J. Med. Genet. A*, 176:1976-1980.

R Yuan, X Zhu, G Wang, S Li, P Ao (2017), Cancer as robust intrinsic state shaped by evolution: a key issues review, *Rep. Prog. Phys.*, 80:042701.

JS Zeiger, TH Beaty, JB Hetmanski, H Wang, AF Scott, L Kasch, G Raymond, EW Jabs, C Vander Kolk (2002), Genetic and environmental risk factors for sagittal craniosynostosis, *J. Craniofacial Surgery*, 13:602-606.

T Zhan, N Rindtorff, M Boutros (2017), Wnt signaling in cancer, *Oncogene*, 36:1461-1473.

X Zhang, S Kuroda, D Carpenter, I Nishimura, C Soo, R Moats, K Iida, E Wisner, FY Hu, S Miao, S Beanes, C Dang, H Vastardis, M Longaker, K Tanizawa, N Kanayama, N Saito, K Ting (2002), Craniosynostosis in transgenic mice overexpressing Nell-1, *J. Clin. Invest.*, 110:861-870.

R Zheng and GA Blobel (2011), GATA transcription factors and cancer, *Genes & Cancer*, 1:1178-1188.

JX Zhou, L Cisneros, T Knijnenburg, K Trachana, P Davies, S Huang (2018), Phylostratigraphic analysis of tumor and developmental transcriptomes reveals relationship between oncogenesis, phylogenesis and ontogenesis, *Converg. Sci. Phys. Oncol.*, 4:025002.

A Zung, E Petek, B BenZeev, T Schwarzbraun, SJ BenYehoshua (2011), MODY type 2 in Greig cephalopolysyndactyly syndrome (GCPS) as part of a contiguous gene deletion syndrome, *Am. J. Med. Genet. A*, 155:2469-2472.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

	n=299 driver				n=125 driver		
CRS gene-set	N	N_{dri} (per)	odds-ratio	Fisher p-value	N_{dri} (per)	OR	pv
Twigg,Wilkie(-2015)	57	12(21%)	17.6	6×10^{-11}	5 (8.8%)	15.3	3×10^{-5}
core genes	20	5 (25%)	22.0	1×10^{-5}			
Goos,Mathijssen(2015-2017)	39	6 (15%)	12.0	3×10^{-5}	3 (7.7%)	13.3	2×10^{-3}
multiple-per	22	2 (9%)	6.6	0.04			
single-per	17	4 (24%)	20.3	1×10^{-4}			
Twigg+Goos	96	18 (19%)	15.2	1×10^{-14}	8 (8.3%)	14.5	2×10^{-7}
Suchir et al (-2019)	17	3 (18%)	14.1	0.002	1 (5.9%)	9.9	0.1
All: Twigg + Goos + Suchir	113	21 (19%)	14.7	1.2×10^{-16}	9 (8%)	13.8	7×10^{-8}

Table 1: Enrichment of cancer driver genes (n=299 set from (Bailey et al., 2018) and n=125 set from (Vogelstein et al., 2013)) in various lists of craniosynostosis genes ((Twigg and Wilie, 2015; Goos and Mathijssen, 2019) and this paper). The 299 cancer driver genes are used to compare with the craniosynostosis genes. for cancer driver genes frequency is 299/20000= 1.5% or 125/20000= 0.6%, respectively in the two cancer driver gene lists, assuming the total number of human genes to be 20000.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

H (hallmark); C2 (curated)

gene-set name	MSigDB-g	gs-size	match	fisher-pv	OR
TGF-beta signaling BMP2, SKI, SMAD6, SMURF1, TGFBR1	h	54	5	3.3E-5	15.4
cancer pathway (KEGG) AXIN2, BMP2, BRAF, FGF9, FGFR1, FGFR2, FGFR3, GLI3, IGF1R, KRAS, RAF1, SMO, STAT3, TGFBR1, TGFBR2	c2	325	15	3.1E-9	8.3
FGF (PID) BGLAP, FGF9, FGFR1, FGFR2, FGFR3, PTPN11, RUNX2	c2	55	7	1.1E-7	21.6
MAPK signaling (KEGG) BRAF, FGF9, FGFR1, FGFR2, FGFR3, FLNA, KRAS, NTRK2, PPP3CA, RAF1, TGFBR1, TGFBR2	c2	266	12	1.7E-7	7.9
FRS2 mediated cascade (Reactome) FGF9, FGFR1, FGFR2, FGFR3, KRAS, RAF1	c2	36	6	2.2E-7	28
signaling by FGFR mutants (Reactome) FGF9, FGFR1, FGFR2, FGFR3, KRAS, STAT3	c2	43	6	5.7E-7	23.5
SHC mediated cascade (Reactome) FGF9, FGFR1, FGFR2, FGFR3, KRAS	c2	28	5	1.8E-6	29.8
signaling by FGFR in disease (Reactome) BRAF, FGF9, FGFR1, FGFR2, FGFR3, KRAS, RAF1, STAT3	c2	122	8	1.5E-6	11.2

Table 2: Gene-sets in MSigDB H and C2 that are enriched with craniosynostosis genes (significant at the family-wide level), and the corresponding overlapping gene names. Gene-set size, overlapping set size, Fisher's test p-value and odds-ratio (OR) are also listed.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

potentially interesting enrichment for cancer-related gene-sets in C2

gene-set name	gs-size	match	fisher-pv	OR
colorectal cancer (KEGG)	62	6	4.1E-6	16.2
AXIN2, BRAF, KRAS, RAF1, TGFBR1, TGFBR2				
pancreatic cancer (KEGG)	70	6	7.8E-6	14.4
BRAF, KRAS, RAF1, STAT3, TGFBR1, TGFBR2				
Greshock cancer copy number	322	11	7.4E-6	5.9
BRAF, FGFR1, FGFR2, FGFR3, GNAS, KAT6B				
KRAS, PTPN11, RECQL4, SMO, TCF12				
melanoma (KEGG)	71	6	8.5E-6	14.2
BRAF, FGF9, FGFR1, IGF1R, KRAS, RAF1				
CML (KEGG)	73	6	9.8E-6	13.8
BRAF, KRAS, PTPN11, RAF1, TGFBR1, TGFBR2				

Table 3: Cancer-related gene-sets in MSigDB C2 that just missed the family-wide significance level, and the corresponding overlapping gene names. Gene-set size, overlapping set size, Fisher's test p-value and odds-ratio (OR) are also listed.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C3 (motif)

gene-set: motif (TF or micro RNA)	gs-size	m	pv	OR
AACTTT (IRF1?) ALX4, AXIN2, BCOR, COLEC10, CYP26B1, EFNA4, ERF, FGF9, FGFR1, FGFR2, GNAS, GPC3, IGF1R, IL6ST IRAX5, JAG1, KIAA1267, KAT6A, MLL2, KRAS, LMNA, MSX2, NFIA, PHEX, PTH2R, RAB23, RSPRY1, SMAD6, SOX6, SPRY1, STAT3, TCF12, ZEB2, ZIC1, ZNF462	1820	35	1E-9	4
ATAAGCT (mir21) HNRNPK, JAG1, PPP3CA, SKI, SOX6, SPRY1, STAT3, TGFBR2	112	8	8.3E-7	12.2
TTGTTT (FOXO4_01) ASXL1,AXIN2,BCOR, CD96, CTSK, DPH1, EFNA4,EFNB1,FGF9,GNAS,GPC3, IGF1R, IL6ST, JAG1, LRP5, MASP1, MSX2, NFIA, NTRK2, PHEX, PPP1CB, PTPRD,RAB23,RUNX2,SHOC2,SMAD6,SOX6,STAT3,TWIST1,ZEB2,ZIC1	1979	31	1.2E-6	3.1
AAAGGGA (mir204/mir211) ALPL,FREM1,GLIS3,MED13L,NTRK2,P4HB,RUNX2,SEC24D,TCF12,TGFBR2	219	10	1.7E-6	7.9
CTTTGT (LEF1_Q2) ATR, AXIN2, BRAF, CD96, EFNA4, EFNB1, FAM20C, FGF9, FGFR3, GLIS3 GPC3,HNRNPK,IGF1R,IRX5,KAT6A,MLL2,KRAS,LRP5,MASP1,PPP1CB RSPRY1, SMAD6, SOX6, STAT3, TCF12, TCOF1, TGFBR1, ZEB2, ZIC1	1892	29	4.5E-6	2.9
RYTTCCTG (ETS2_B) BMP2,BRAF,CD96,COLEC11,CTSK,ERF,FGFR2,HNRNPK,IFT43,IL11RA, KIAA1267,MLL2,LMX1B,LRP5,NFIA,PHEX,PTPN11,TCF12,TGFBR2,ZEB2	1048	20	8.9E-6	3.5

Table 4: Gene-sets in MSigDB C3 (motif) that are enriched with craniosynostosis genes (significant at the family-wide level), and the corresponding overlapping gene names. Gene-set size, overlapping set size(m), Fisher's test p-value (pv) and odds-ratio (OR) are also listed.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C5 (gene ontology), stronger signal (10% genes of gene-set overlap)

gene-set name	gs-size	m	pv	OR
bone dev	155	22	1E-22	27.8
skeletal sys morphog	200	23	8E-22	22.8
bone morphog	79	16	6.3E-19	37.5
cranial ske dev	55	14	6.7E-18	46.2
reg ossification/pos/neg	174/82/67	19/9/7	9.3E-18/5.2E-9/3.8E-7	21.1/18.9/17.7
appendage dev	166	18	8.1E-17	20.8
reg osteoblast diff/pos	109/58	15/9	1.4E-15/3.2E-10	25.2/26.8
reg smoothened sig/pos	62/24	11/5	8.6E-13/9E-7	31.3/34.7
embryonic skeletal sys morphog	93	12	2.3E-12	22.9
embryonic skeletal sys dev	122	13	2.4E-12	19.1
odontogenesis	105	12	8.3E-12	20.9
embryonic cranial skeleton morphog	46	9	5.1E-11	33.8
smad binding	70	10	6.7E-11	24.9
biomineral tissue dev	75	10	1.2E-10	23.2
chondrocyte diff	60	9	4.2E-10	25.9
reg cartilage dev/pos	62/29	9/6	5.5E-10/7.2E-8	25.1/34.8
craniofacial suture morphog	14	6	1.8E-9	75.2
reg chondrocyte diff/pos	45/19	8/5	1.2E-9/3.3E-7	30.4/43.9
trans-mem rec prot kinase/tyrosine	81/64	9/7	4.7E-9/2.9E-7	19.2/18.5
bone mineralization	38	7	1.1E-8	31.3
ciliary tip	43	7	2.4E-8	27.6
reg mesenchymal cell proliferation/pos	34/27	6/6	1.7E-7/4.9E-8	29.7/39.6
embryonic digit morphog	59	7	1.7E-7	20.1
pos reg stem cell proliferation	61	7	2.1E-7	19.5
body morphog	44	6	6.5E-7	22.9
endochondral bone morphog	45	6	7.3E-7	22.4
prot transport along microtubule	27	5	1.5E-6	30.9

Table 5: Partial gene-sets in MSigDB C5 (gene-ontology) that are enriched with craniosynostosis genes (significant at the family-wide level), with number of overlapping genes (m) at least 10% of the gene-set size. Fisher's test p-value (pv) and odds-ratio (OR) are also listed. act: activity; dev: development; diff: differentiation; mem: membrane; morphog: morphogenesis; neg: negative; pos: positive; prot: protein; rec: receptor; reg: regulation; sig: signaling; sys: system;

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C5: gene ontology, large gene-set size (> 1000 genes); or very significant

gene-set name	gs-size	match	fisher-pv	OR
tissue dev	1501	43	4.7E-18	6.8
reg cell diff	1470	41	9.6E-17	6.4
reg multicell org dev	1652	39	1.5E-13	5.2
pos reg dev	1129	32	3.9E-13	5.9
reg tran (RNA polII prom)	1763	39	1.1E-12	4.9
res endogenous stimulus/cellular..	1440/1000	35/25	1.8E-12/31.E-9	5.2/4.8
reg cell proliferation	1478	35	3.7E-12	5
cell res org substance	1833	38	1.7E-11	4.5
pos reg biosynthetic proc	1792	36	1.7E-10	4.2
pos reg gene expression	1719	35	2.3E-10	4.3
cell dev	1406	31	5.1E-10	4.5
pos reg resp stimulus	1878	36	5.9E-10	4
pos reg multicell org proc	1382	30	1.5E-9	4.4
neurogenesis	1388	30	1.7E-9	4.3
pos/neg reg cell communication	1520/1176	31/27	3.3E-9/3.8E-9	4.1/4.5
reg anatomical struture morphog	1011	25	3.8E-9	4.8
reg cell death	1458	29	2.1E-8	3.9
immune sys proc	1927	32	2.1E-7	3.3
skeletal sys dev	453	37	3.1E-30	18.9
org morphog	839	42	1.8E-26	12.1

Table 6: Partial gene-sets in MSigDB C5 (gene-ontology) that are enriched with craniosynostosis genes (significant at the family-wide level), with at least 1000 genes in the gene-set. Number of overlapping genes (m), Fisher's test p-value (pv) and odds-ratio (OR) are also listed. dev: development; diff: differentiation; morphog: morphogenesis; neg: negative; org: organismal; pos: positive; proc: process; prot: protein; prom: promoter; reg: regulation; res: response; sys: system; tran: transcription;

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ingenuity pathway analysis (IPA). Signalingand cancer-related canonical pathways.

name	p-value
STAT3 pathway (12) FGFR1, FGFR2, FGFR3, IGF1R, IL11RA, IL6ST, KRAS, NTRK2, RAF1, STAT3, TGFBR1, TGFBR2	4E-12
NF-KB signaling (12) BMP2, BRAF, FGFR1, FGFR2, FGFR3, IGF1R, KRAS, NTRK2, RAF1, TGFBR1, TGFBR2, TNFRSF11B	1E-10
cardiac hypertrophy signaling (enhanced) (17) FGF9, FGFR1, FGFR2, FGFR3, FGFRL1, GNAS, IGF1R, IL11RA, IL6ST, KRAS, PPP3CA, RAF1, SMO, STAT3, TGFBR1, TGFBR2, TNFRSF11B	4E-10
TGF-beta signaling (9) BMP2, KRAS, RAF1, RUNX2, SKI, SMAD6, SMURF1, TGFBR1, TGFBR2	1E-9
FGF signaling (8) FGF9, FGFR1, FGFR2, FGFR3, FGFRL1, PTPN11, RAF1, STAT3	1E-8
PTEN signaling (9) FGFR1, FGFR2, FGFR3, IGF1R, KRAS, NTRK2, RAF1, TGFBR1, TGFBR2	2E-8
molecular mechanism of cancer (13) ATR, BMP2, BRAF, GNAS, IHH, KRAS, LRP5, PTPN11, RAF1, SMAD6, SMO, TGFBR1, TGFBR2	1E-7
protein kinase A signaling (13) BRAF, FLNA, GLI3, GNAS, IHH, PPP1CB, PPP3CA, PTPN11, PTPRD, RAF1, SMO, TGFBR1, TGFBR2	1E-7
BMP signaling (6) BMP2, KRAS, RAF1, RUNX2, SMAD6, SMURF1	5E-6
colorectal cancer metastasis signaling (9) BRAF, GNAS, IL6ST, KRAS, LRP5, SMO, STAT3, TGFBR1, TGFBR2	6E-6
CNTF signaling (5) IL6ST, KRAS, PTPN11, RAF1, STAT3	1E-5
role of NFAT in cardiac hypertrophy (8) GNAS, IGF1R, IL6ST, KRAS, PPP3CA, RAF1, TGFBR1, TGFBR2	1E-5
GM-CSF signaling (5) KRAS, PPP3CA, PTPN11, RAF1, STAT3	3E-5

Table 7: More information on the IPA canonical pathways in Fig.2 (enrichment p-value $< 3 \times 10^{-5}$) that are either a signaling pathway, or related to cancer: number and names overlapping genes and enrichment p-value.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ingenuity pathway analysis. others.

name	p-value	num.molecules
upstream regulators		
RUNX2	5.9E-14	
beta-glycerophosphoric acid	1.3E-11	
FGF2	3.9E-11	
BMP2	5.5E-11	
beta-estradiol	3.6E-9	
diseases and disorders		
connective tissue disorders	1.8E-8 ∼ 7.1E-71	90
hereditary disorder	1.9E-8 ∼ 7.1E-71	82
organismal injury and abnormalities	2.1E-8 ∼ 7.1E-71	112
developmental disorder	1.9E-8 ∼ 1.8E-43	99
skeletal and muscular disorders	1.8E-8 ∼ 1.8E-43	82
molecular and cellualr functions		
cellular development	2.1E-8 ∼ 8.8E-25	69
cell signaling	1.7E-14 ∼ 1.6E-16	22
cellular growth and proliferation	2.1E-8 ∼ 2.1E-15	60
cell death and survival	1.2E-8 ∼ 8.6E-15	67
gene expression	5E-11 ∼ 3.4E-14	51
physiological system development and function		
organismal development	1.7E-8 ∼ 2.7E-38	88
embryonic development	1.4E-3 ∼ 2.2E-37	75
connective tissue development and function	3E-2 ∼ 5.8E-35	67
skeletal and muscular system development and function	4.9E-2 ∼ 5.8E-35	66
tissue development	2.4E-1 ∼ 5.8E-35	79

Table 8: IPA top five hits in other categories, based on enrichment/association analysis with the craniosynostosis genes: upstream regulators, diseases and disorders, molecular and cellular functions, physiological system development and function.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

gene-set	AXIN2	BMP2	BRAF	FGF9	FGFR1	FGFR2	FGFR3	GLI3	IGF1R	KRAS	RAF1	SMO	STAT3	TGFBR1	TGFBR2
cancer driver gene	y		y		y	y	y			y	y				y
IPA cancer genes		y	y							y	y	y		y	y
(MSigDB C1) MAPK mediate signaling			✓	✓	✓	✓	✓			✓	✓			✓	✓
FRS2 mediated cascade				✓	✓	✓	✓			✓	✓				
signaling FGFR mutants				✓	✓	✓	✓			✓			✓		
SHC mediated signaling				✓	✓	✓	✓			✓					
signaling FGFR in disease			✓	✓	✓	✓	✓			✓	✓		✓		
(IPA) STAT3 pathway					✓	✓	✓		✓	✓	✓		✓	✓	✓
NF-KB signaling		✓	✓		✓	✓	✓		✓	✓			✓	✓	✓
cardiac hypertrophy signaling				✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
TGF-beta signaling		✓								✓	✓			✓	✓
FGF signaling				✓	✓	✓	✓				✓		✓		
PTEN pathway					✓	✓	✓		✓	✓	✓			✓	✓
protein kinase A signaling			✓					✓			✓	✓		✓	✓
BMP signaling		✓								✓	✓				
(MSigDB C3) AACTTT (unknown, IRF?)	✓			✓	✓	✓			✓	✓			✓		
ATAAGCT (mir21)													✓		✓
TTGTTT (FOX04)	✓			✓					✓				✓		
AAAGGGA (mir204/mir211)															✓
CTTTGT (LEF1)	✓		✓	✓			✓		✓	✓			✓	✓	
TYTTCCTG (ETS2)		✓	✓			✓									✓
transcription factor								y					y		

Table 9: Fifteen craniosynostosis genes that are also in KEGG cancer pathway: their presence in cancer driver gene list; in IPA cancer gene list; in other MSigDB C2, C3 gene-sets; in IPA signaling pathways; and in a transcription factor list (Lambert et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

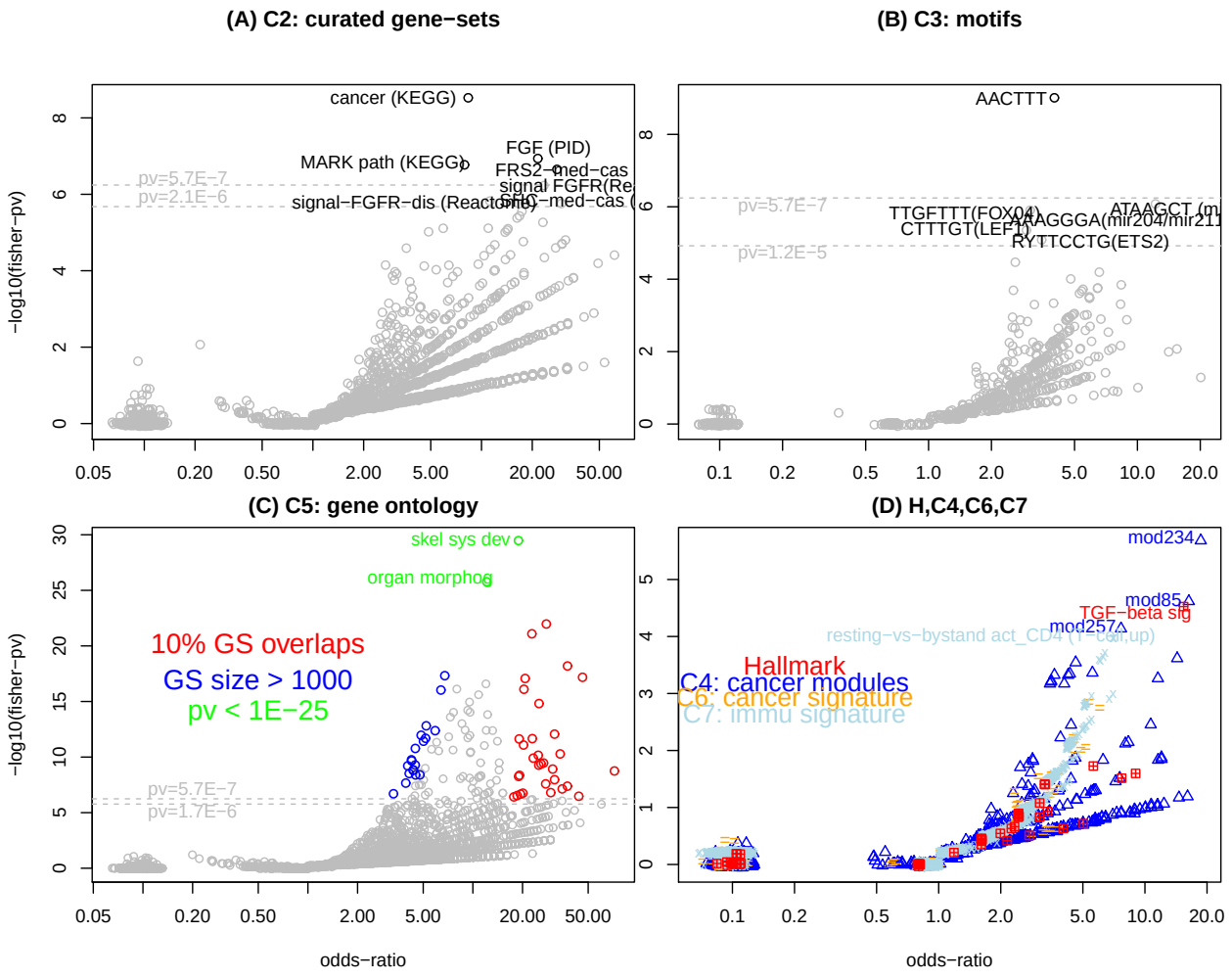

Figure 1: Volcano plot (x: (log) odds-ratio, y: $-\log_{10}$(Fisher p-value) (Li et al., 2014)) for enrichment (over-representation) analysis of craniosynostosis genes in MSigDB gene-sets: (A) curated gene-sets, including KEGG, Reactome, PID (C2); (B) motif based gene-sets (C3); (C) gene ontology gene-sets (C5). Two groups of significantly enriched gene-sets are highlighted: large gene-sets (> 1000 genes) (in blue), and high proportion (> 10%) of overlapping genes (in red); (D) other categories of gene-sets (hallmark, microarray expression derived signatures (C4, C6, C7)).

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

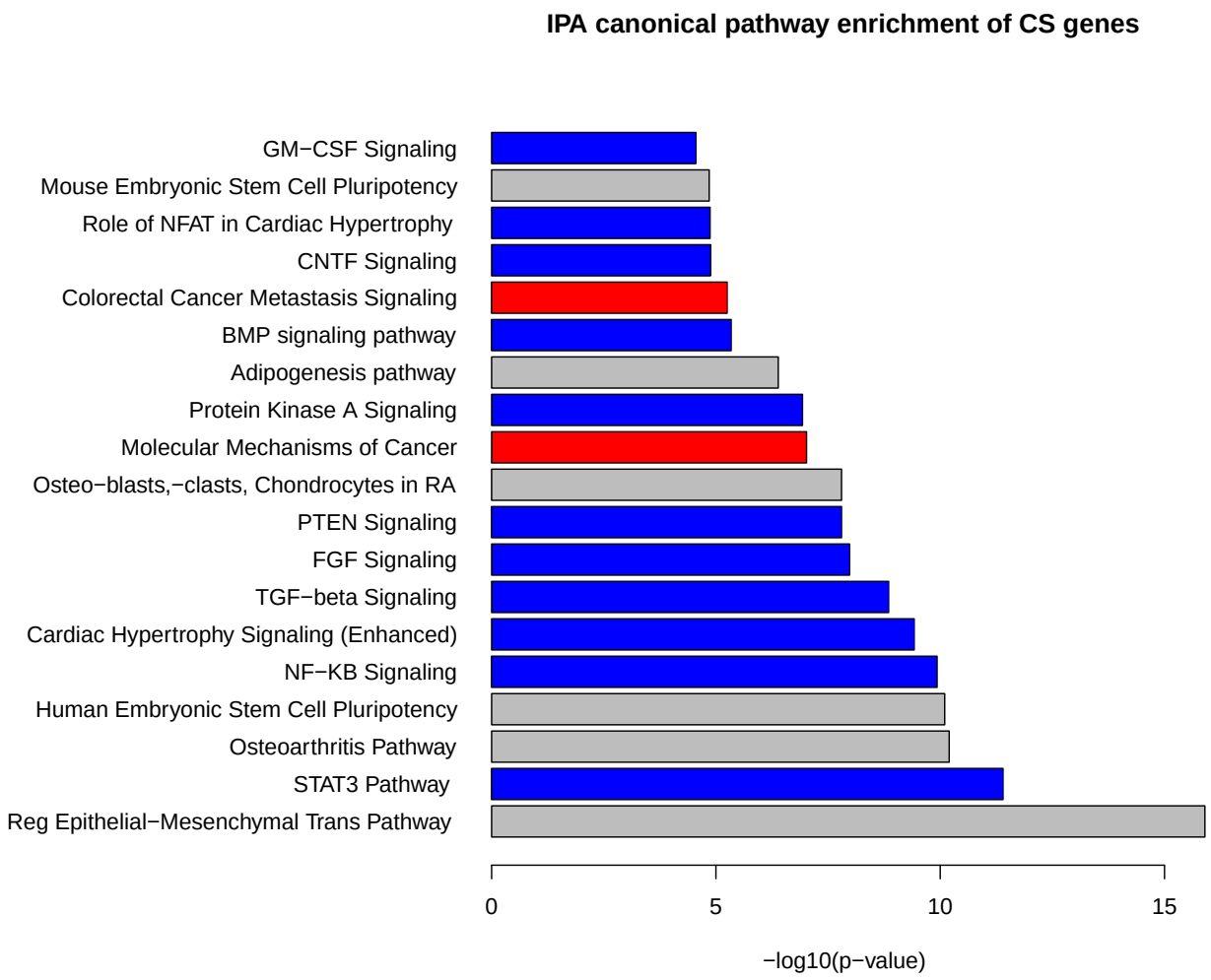

Figure 2: IPA canonical pathways that are significantly enriched with craniosynostosis genes at p-value of 3×10^{-5} level. The cancer related pathways are colored by red, and signaling pathways by blue. The remaining pathways are all related to development and differentiation.

bioRxiv preprint doi: https://doi.org/10.1101/872093; this version posted December 11, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

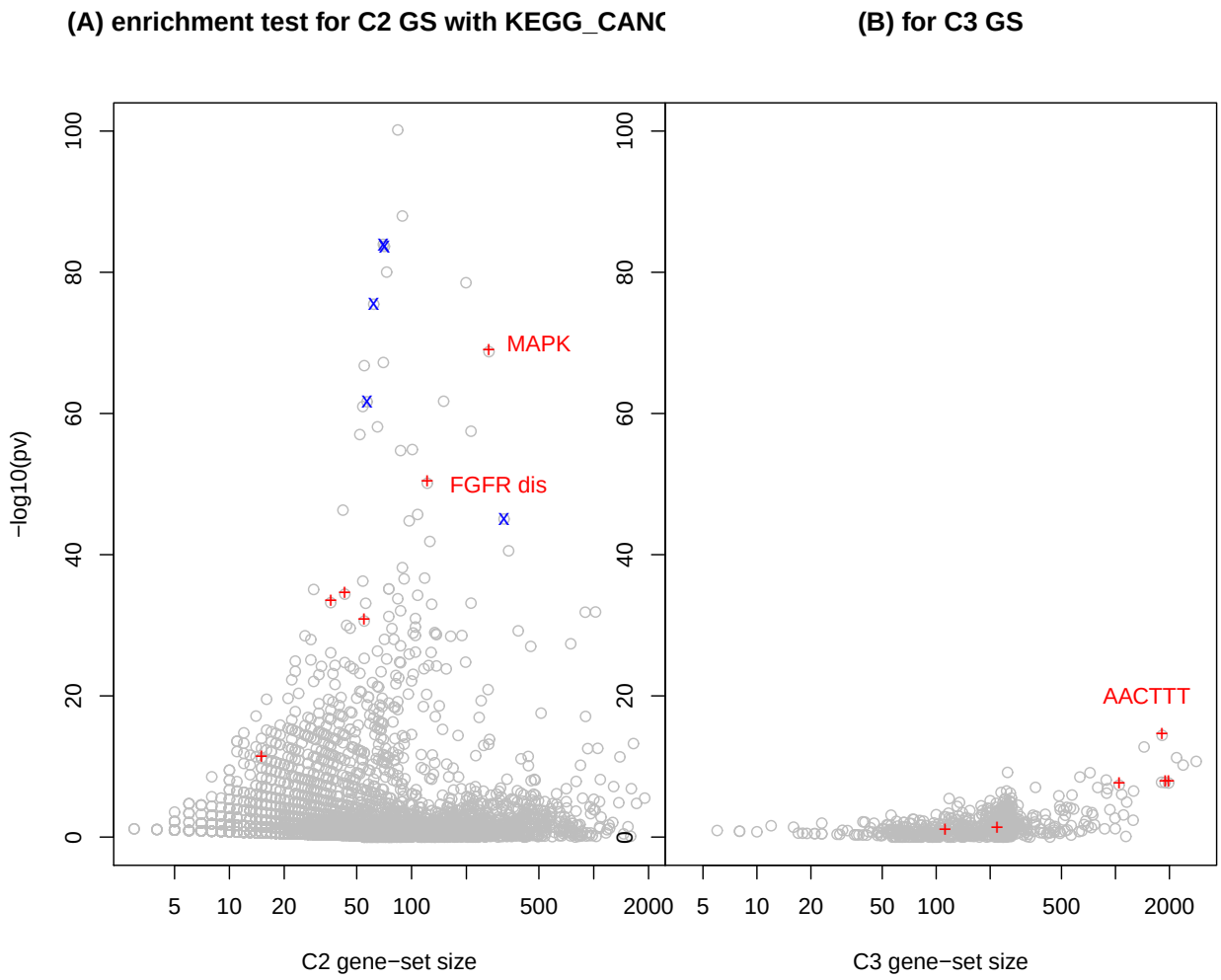

Figure 3: Enrichment of KEGG cancer pathway genes in other MSigDB C2 and C3 gene-sets. The x-axis is the the gene-set size (in log-scale), and the y-axis is the minus log10 of Fisher's test p-value. In (A), the red dots are C2 gene-sets listed in Table 2, and blue dots are the cancer-related C2 gene-sets listed in Table 3. In (B), the red dots are C3 gene-sets listed in Table 4.