CSE 182 Final Report: CCND2

Group 7

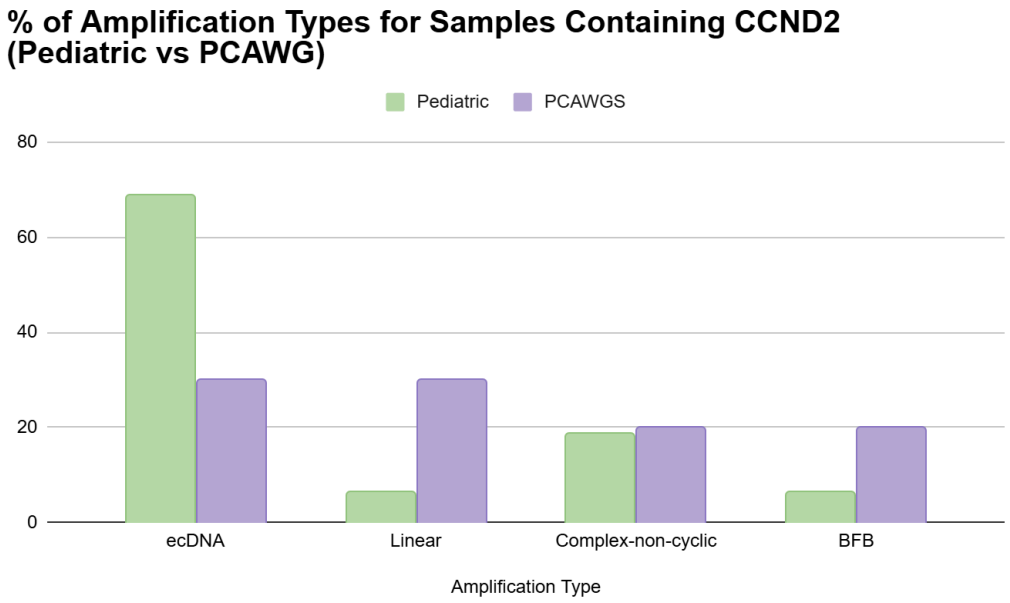
By: Kayla Cooper, Riani Shah, Sameera Agarwala, Dhruv Patravali, Ryan Izadshenas

Part I

CCND2 (Cyclin D2) is a protein involved in managing the cell cycle, specifically the G1/S Phase. CCND2 creates a complex with either CDK4 or CDK6 and phosphorylates protein Rb (retinoblastoma), a tumor suppressor gene. High expression levels of this gene were primarily seen in ovarian and testicular tumors, and mutations in CCND2 are frequently associated with MPPH Syndrome [6], a neurodevelopmental disorder. Knockout studies in mice have demonstrated the importance of CCND2 in germ cell proliferation. [1]

CCND2 is highly conserved throughout multiple taxonomic groups. In order to identify any distant orthologs, we used BLAST to compare CCND2's protein sequence alongside other organisms (using NCBI's non-redundant protein records) [2]. Due to its highly conserved nature, the majority of alignments had >95% identity and similar lengths to the CCND2 sequence. The most distant hit was seen in the Damaraland mole-rat (Fukomys damarensis), with a percent identity of 87.9%, and 8.2% of the alignment being gaps. Conceptually, this makes sense, germ cell proliferation is a specific, complex process with little room for change or improvement. Small changes to account for speciation are expected, but massive changes to the genome structure would not be expected, as then successful reproduction could be jeopardized.

To compare the amplification types between Pediatric and PCAWG samples, we filtered each dataset to only samples containing the CCND2 gene and then counted the feature count for each amplification type. For the pediatric data, there were 16 samples that contained CCND2, and for PCAWG, there were 10 samples. From these samples, we calculated the frequency of each amplification type whenever CCND2 was present. The graph below visualizes these counts.



For pediatric samples, 68.75% of CCND2 amplicons were ecDNA, followed by 18.75% linear amplicons, and 6.25% for both BFB and Complex-non-cyclic amplicons. There is a drastic change in these distributions when compared to the PCAWG samples. ecDNA amplicons significantly decrease to 30%, while linear, complex-non-cyclic, and BFB increase to 30%, 20% and 20%, respectively. In order to explain these differences, we began to research the effects of CCND2 mutations for young children vs adult patients. There is a difference in the pathology of CCND2 mutations based on age. In infants, CCND2 mutations are associated with microcephaly, a neurodevelopmental disorder that can be diagnosed prenatally [3]. MedlinePlus reports that CCND2 mutations are implicated in megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) syndrome, a rare neurodevelopmental disorder marked by brain overgrowth, abnormal cortical development, and sometimes extra digits, further emphasizing the gene's critical role in early brain and limb development. In adults, CCND2 mutations often manifest as ovarian/testicular tumors. There have also been studies that link CCND2 amplification to shorter overall survival of patients with adult-type diffuse gliomas [4]. From this, we speculate that in

embryonic development, CCND2 mainly impacts neurological development. As the organism ages, CCND2 is still active in the neurological system, but plays a more significant role in germ cell proliferation and reproductive processes.

CCND2 is known to form complexes with both CDK4 and CDK6 (separately) in order to regulate the tumor suppressor protein Rb. Because both CCND2 and CDK4/CDK6 are needed for this regulation, these genes are heavily dependent on each other's activity and frequently co-occur. CCND2 is found on chromosome 12, at p13.32, while CDK4 is also on chromosome 12 and is located at q14.1. CDK6 is actually located on chromosome 7, which suggests that CDK4 and CCND2 co-occur and interact more frequently than CDK6 and CCND2.

Part II

From the sample DO29146-SP62110 in the PCAWG dataset, we used the NCBI RefSeq database to retrieve the amino acid sequences of Cyclin D2, which is the protein that CCND2 (our oncogene) codes for. This served as our query for alignment. We then used the BED file for Amplicon 5 from sample DO29146-SP62110 in the PCAWG dataset. This BED file contained the genomic coordinates of the ecDNA that contained the CCND2 gene. We used this to extract the relevant DNA sequence for our alignment, and used it as our database for alignment. For alignment, we used tblastn to align the protein sequence of Cyclin D2 against the ecDNA region defined by our BED coordinates, in order to identify any truncated genes or rearrangements. We used tblastn in particular, as this is a form of blast used for aligning protein sequences with a DNA database, checking alignments for all six reading frames. This was key, because for ecDNA, we often lack knowledge about the correct strand or frame that the gene is in. Using tblastn returns the best matching alignments for all possibilities, which is particularly helpful for

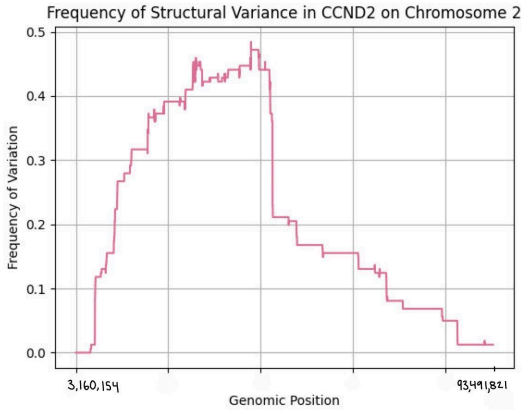
analyzing ecDNA for cancer genomics, where the ecDNA can often be rearranged, doesn't have clear exon annotations, or could only contain partial gene copies. The tblastn alignment results are shown below:

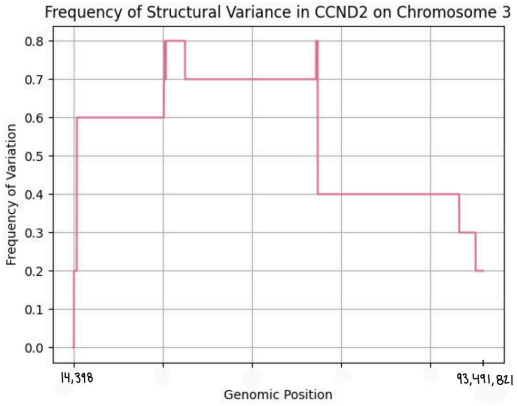
Query ID	Subject ID	% Identity	Alignment Length	Mismatches	Gap Opens	Query Start	Query End	Subject Start	Subject End	E-value	Bit Score
NP_001750.1	chr12:87478344-87499229	100.000	61	0	0	77	137	35104021	3510603	1.55e-33	127
NP_001750.1	chr12:87478344-87499229	96.429	56	2	0	136	191	35131337	3513304	2.45e-29	114
NP_001750.1	chr12:87478344-87499229	88.333	60	7	0	181	240	35231994	3523373	7.63e-26	104
NP_001750.1	chr12:87478344-87499229	100.000	49	0	0	241	289	35342443	3534389	2.71e-24	100
NP_001750.1	chr12:87478344-87499229	100.000	64	0	0	1	64	3508424	3508615	1.99e-21	91.7
NP_001750.1	chr12:87478344-87499229	35.484	31	20	0	182	212	680291	680383	3.7	26.2
NP_001750.1	chr12:87478344-87499229	34.694	49	17	1	92	125	2282849	2282703	5.8	25.8
NP_001750.1	chr12:87478344-87499229	23.404	94	41	3	119	188	1906510	1906770	5.8	25.8

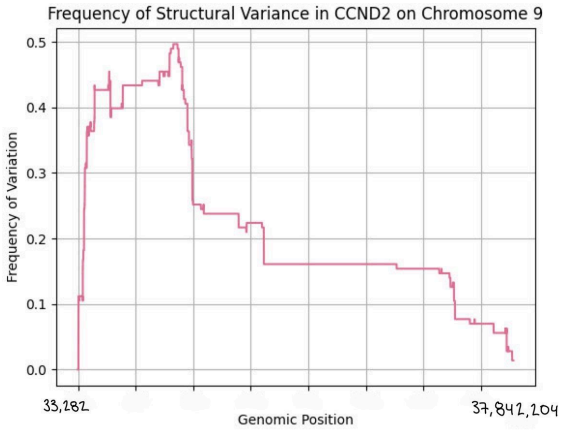
The query sequence (CCND2 protein) is 289 amino acids long. For hits with an E-value less than .05, the average alignment length was 58 amino acids long. From our chart, we can see that the entirety of the query was found if we combine all significant alignments, indicating that there was some transversion or alternative splicing occurring to change the protein structure. Considering that we can find the entirety of our query, we can observe that there is no truncation, however this fragmentation suggests that the gene has been broken up and rearranged during the process of ecDNA formation. This is a well-documented phenomenon — ecDNA often forms through genomic rearrangements, allowing oncogenes to be retained but in a structurally scrambled format.

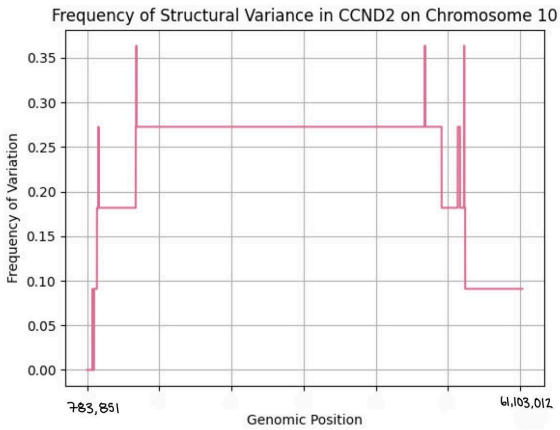
Part III

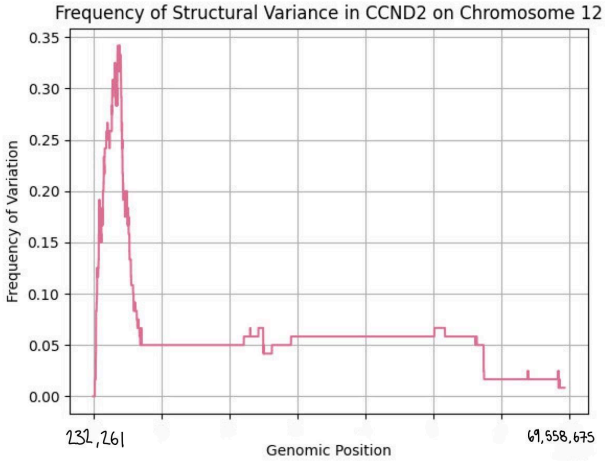
To highlight structural variants in CCND2, we extracted 10 different amplicon graphs where CCND2 was present. Using the discordant edges in each of these graphs, we created a file called discordant.txt which listed all discordant edges throughout all 10 amplicons. From here, we separated each edge depending on which chromosome the variant was located at, and for each chromosome, tallied how frequently that region was structurally variant. For visualization, we created graphs displaying these frequencies. It is important to note that the graphs we created only represented the frequency of variants on a single chromosome, not when the discordant edge spanned multiple chromosomes. Any discordant edges spanning across multiple chromosomes were listed in the file seperate_Discordant.txt. Below are the frequencies of structural variation in CCND2 for chromosomes 12, 10, 9, 3, and 2. For each graph, the 0 index refers to the earliest discordant edge found on that chromosome.


Frequency of Structural Variance in CCND2 on Chromosome 2
Frequency of Variation
0.0
0.1
0.2
0.3
0.4
0.5
3,160,154
93,491,821
Genomic Position


Frequency of Structural Variance in CCND2 on Chromosome 3
Frequency of Variation
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
14,398
93,491,821
Genomic Position


Frequency of Structural Variance in CCND2 on Chromosome 9
Frequency of Variation
0.0
0.1
0.2
0.3
0.4
0.5
33,282
37,842,104
Genomic Position


Frequency of Structural Variance in CCND2 on Chromosome 10
Frequency of Variation
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
783,851
61,103,012
Genomic Position


Frequency of Structural Variance in CCND2 on Chromosome 12
Frequency of Variation
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
232,261
69,658,675
Genomic Position

Chromosomes 3 and 10 have relatively flat distributions, with a preference for mutations within the middle of the gene. Chromosomes 2, 9, and 12 show a preference for variations in the beginning of the gene.

In addition to the intra-chromosomal variation, we identified numerous inter-chromosomal discordant edges that suggest complex genomic rearrangements. The most frequent of these involve chromosomes 17, 12, and 10. These both may play a significant role in the structural reconfiguration of regions of CCND2. From chr17 to chr12, these edges appeared 26 times in samples and frequently connected high-density regions on chromosome 17 to locations in the q-arm of chromosome 12, where CCDN2 is located at 12q13.3. From chr12 to chr10 and vice versa, these linkages emerged 20 and 7 times respectively, indicating bidirectional rearrangements of these chromosomes. To help explain this, we researched proteins that commonly interact with CCND2 and looked at which chromosomes they were located at in order to find any potential relationship. A study done on transcriptional repression in cervical cancer revealed protein EZH2-rich domains were found in the CCND2 loci, indicating that these genes are frequently interacting with each other. EZH2 originates from chromosome 17 and is involved in transcriptional repression, which aligns with CCND2's function. [7]

Resources

All code and files used to create this report are also viewable from our Github Repository.

1. NCBI CCND2 (Gene ID: 894)
    a. CCND2: National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [accessed 2025 May 31]. Available from: https://www.ncbi.nlm.nih.gov/gene?Cmd=DetailsSearch&Term=894
2. BLASTP for Finding Orthologs
    a. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.
3. CCND2 and Microcephaly
    a. Pirozzi F, Lee B, Horsley N, Burkardt DD, Dobyns WB, Graham JM Jr, Dentici ML, Cesario C, Schallner J, Porrmann J, Di Donato N, Sanchez-Lara PA, Mirzaa GM. Proximal variants in CCND2 associated with microcephaly, short stature, and developmental delay: A case series and review of inverse brain growth phenotypes. Am J Med Genet A. 2021 Sep;185(9):2719-2738. doi: 10.1002/ajmg.a.62362. Epub 2021 Jun 4. PMID: 34087052; PMCID: PMC8725575.
4. CCND2 and Adult-type diffuse gliomas
    a. Timothy E Richardson, Jamie M Walker.CCND2 amplification is an independent adverse prognostic factor in IDH-mutant lower-grade astrocytoma. Clin

Neuropathol. 2021; 40: 209-214. doi: 10.5414/NP301354. Pubmed: https://pubmed.ncbi.nlm.nih.gov/33560216/; PMID: 33560216.

5. CCND2 Protein Sequence (RefSeq)
    a. National Library of Medicine (US), National Center for Biotechnology Information; [1988] – . Accession No. NP_001750, G1/S-specific cyclin-D2 [Homo sapiens]; [cited 2025 May 31]. Available from: https://www.ncbi.nlm.nih.gov/protein/4502617
6. MPPH Syndrome
    a. MedlinePlus [Internet]. Bethesda (MD): National Library of Medicine (US); [updated 2020 Jun 24]. CCND2 Gene; [cited 2025 May 31]. Available from: https://medlineplus.gov/genetics/gene/ccnd2/#conditions
7. EZH2 and CCND2 Interactions
    a. Salmerón-Bárcenas EG, Zacapala-Gómez AE, Ortiz-Ortiz J, Torres-Rojas FI, Ávila-López PA. Integrated bioinformatics analysis reveals that EZH2-rich domains promote transcriptional repression in cervical cancer. EXCLI J. 2022 Jun 23;21:852-868. doi: 10.17179/excli2022-5029. PMID: 36172073; PMCID: PMC9489889.